Trustworthy Governance of Agentic Societies Based on DePIN and VLA

XIAOLONG LIANG[1], RUI QIN[1], JUANJUAN LI[1], and THALES S. W. THESEUS[2]
[1]Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[2]Parallel Intelligence DeSci Center, Obuda University, Budapest, Hungary

Corresponding author: Juanjuan Li (e-mail: juanjuan.li@ia.ac.cn).

This work was supported by The National Natural Science Foundation of China (62103411), and the Science and Technology Development Fund, Macao SAR (0093/2023/RIA2, 0050/2020/A1).

ABSTRACT While Decentralized Autonomous Organizations (DAOs) and Artificial Intelligence are reshaping the governance of academic societies, reliably integrating on-chain decisions with off-chain physical activities remains a critical challenge. The fundamental bottleneck is the difficulty of reliably integrating real-world execution outcomes into the digital decision-making loop. To address this, we propose an endogenous contribution evaluation framework integrating Decentralized Physical Infrastructure Networks (DePIN) and Vision-Language-Action (VLA) models. This approach maps physical entities to on-chain decentralized identities. By leveraging VLA edge nodes to analyze multimodal behavioral data collected via DePIN, the system autonomously generates a verifiable Proof of Real-World Contribution (PoRWC). This proof subsequently drives on-chain incentive distribution through a reputation-weighted consensus mechanism. Consequently, this framework establishes an endogenously trustworthy closed loop from physical processes to digital governance. We demonstrate its feasibility and scalability through a case study of the Chinese Association of Automation (CAA), providing a robust engineering path for the parallel governance of modern academic societies.

INDEX TERMS Digital CAA, Vision-Language-Action Model, AI Agents.

I. INTRODUCTION

In the era of fierce global technological competition and industrial transformation, academic and scientific societies play a pivotal role in linking academic research, industrial practice, and talent cultivation. However, traditional organizational governance models are facing severe challenges, often constrained by long decision-making chains and low cross-departmental collaboration efficiency. To address these bottlenecks, the integration of smart contracts [3], [4], Decentralized Autonomous Organizations (DAOs) [5], [6], and Artificial Intelligence (AI) represented by large models [7] has emerged as a promising paradigm for digital transformation. Building upon this trend, the concept of "Digital CAA" was proposed [1], [2], aiming to construct a parallel governance system that digitally empowers physical entities through virtual-real interaction.

The key to unlocking the full potential of such parallel governance lies in constructing a trustworthy closed-loop feedback mechanism between digital on-chain decisions and physical off-chain activities. The core value creation of academic societies fundamentally relies on operations in the physical world, such as hosting academic conferences, matching technological transformations, and implementing industry standards [8]–[10]. However, the evaluation of these physical contributions relies heavily on an exogenous closed-loop model: the implementation results of physical activities require centralized, human-in-the-loop interventions to be uploaded on-chain and embedded into the digital decision-making process. This fundamentally creates a fragile and untrustworthy oracle link [11]–[13]. Passive dependence on exogenous data not only lacks endogenous verifiability but also introduces unintentional human operational errors and opens vectors for malicious attacks, such as data forgery and exaggerated contributions.

To bridge this physical-digital divide, this paper proposes an endogenous contribution evaluation framework integrating Decentralized Physical Infrastructure Networks (DePIN) [14], [15] and Vision-Language-Action (VLA) models [16]–[18]. Within this framework, DePIN serves as a trustworthy data foundation, utilizing token economics to incentivize distributed physical hardware to collect verifiable, high-fidelity real-time data streams from the physical world. Concurrently, VLA acts as an intelligent analysis engine capable of understanding visual inputs, conducting language reasoning, and autonomously planning actions. By deploying VLA on De-

PIN edge nodes, the system can reliably process raw multimodal data streams and translate complex physical activities in real-time into standardized Proof of Real-World Contribution (PoRWC) credentials. This establishes an endogenously trustworthy closed loop from physical processes to on-chain incentive distribution. Furthermore, we design a distributed identity management system to semantically map physical devices, personnel, and organizations to on-chain identities. By dynamically updating time-sensitive qualifications, this system ensures the completeness of contribution traceability and the precision of attribution.

The organization of this paper is as follows: Section II introduces background knowledge and preliminaries; Section III presents the architecture of the digital-physical integrated CAA; Section IV proposes the VLA-based actual CAA contribution evaluation method; Section V builds the DePIN-based digital CAA identity system; Section VI takes the CAA Member Representative Assembly as an example to elaborate on the trustworthy governance mechanism of physical-digital CAA collaboration, demonstrating how the proposed mechanism assists actual CAA governance; Section VII summarizes the full text and looks forward to future development directions.

II. BACKGROUND AND PRELIMINARIES

A. BLOCKCHAIN AND DECENTRALIZED TECHNOLOGIES

The concept of Decentralized Autonomous Organizations (DAOs) emerged from the convergence of blockchain technology and smart contracts. In 2016, "The DAO" launched on Ethereum as the first large-scale decentralized venture fund, raising substantial capital through token sales [24]. However, a reentrancy vulnerability in its smart contract was subsequently exploited, resulting in massive asset theft and triggering Ethereum's controversial hard fork [25]. This seminal event exposed fundamental challenges in smart contract security, raising critical questions regarding the trade-offs between code immutability and governance intervention in decentralized systems.

Subsequent research on DAO governance mechanisms has identified persistent structural limitations in token-based voting systems [5], [6]. Empirical studies reveal that governance power tends to concentrate among large token holders, voter participation remains critically low, and token-weighted voting often fails to align with the domain expertise required for informed decision-making [26]. In response to these challenges, the framework of True Autonomous Organizations and Operations (TAO) was proposed [27], [28]. TAO redefines the objectives of decentralized governance through the principles of being Trustable, Reliable, Usable, and Efficient & Effective. Unlike traditional DAOs that prioritize structural decentralization as an end in itself, TAO introduces multi-dimensional reputation systems, hybrid consensus mechanisms, and value-sensitive design principles, thereby shifting the focus toward trustworthy governance outcomes [29].

Identity management constitutes a foundational component of DAO governance, with direct implications for member authentication, voting rights allocation, and reputation tracking. Traditional identity verification relies on Know Your Customer (KYC) procedures, wherein centralized authorities collect, validate, and store personal data. While KYC fulfills regulatory compliance and mitigates specific fraud vectors, it introduces inherent limitations: privacy vulnerabilities from centralized data repositories, single points of failure in verification infrastructure, and a strict dependency on third-party custodians [30]. These limitations are particularly problematic in decentralized governance, where pseudonymous participation and cross-organizational identity portability are essential.

Decentralized Identifiers (DID), standardized as a W3C Recommendation in 2022 [31], represent a paradigm shift toward self-sovereign identity management. In contrast to KYC-based systems where identity is controlled by service providers, DID empowers individuals to create and manage their identifiers without reliance on central authorities. Within DAO contexts, DID addresses four critical challenges [32]: (1) Sybil attack resistance through cryptographically verifiable uniqueness; (2) privacy preservation via the selective disclosure of verifiable credentials; (3) cross-organizational interoperability, enabling a single DID to function seamlessly across multiple DAOs; and (4) reputation portability, allowing on-chain credentials and contribution records to transfer across governance contexts. Integrating DID with DAO architectures thus establishes the identity foundation necessary for trustworthy, privacy-preserving governance.

Extending blockchain coordination mechanisms to physical domains, Decentralized Physical Infrastructure Networks (DePIN) have emerged as a paradigm for organizing physical resource contributions through token-based incentives [15]. DePIN leverages blockchain to align incentives, coordinating distributed participants who contribute physical resources such as sensors, wireless nodes, and storage devices. Successful implementations—including Helium (decentralized wireless networks), Filecoin (distributed storage), and Hivemapper (crowdsourced mapping)—demonstrate the viability of decentralized coordination for physical infrastructure provisioning [33]. It is worth noting that DePIN research also encompasses hardware-level security mechanisms, including Trusted Execution Environments (TEEs), hardware attestation protocols, and tamper-evident device provisioning, which collectively mitigate risks associated with compromised end nodes such as cameras or microphones. While the present work focuses on the governance methodology and contribution evaluation framework, these hardware security guarantees provided by the DePIN ecosystem serve as a complementary trust anchor for the physical data collection layer.

B. ARTIFICIAL INTELLIGENCE AND MULTIMODAL MODELS

The foundation of modern pre-trained language models traces back to Word2Vec, introduced by Mikolov et al. in 2013 [34]. This work demonstrated that distributed word representations learned from large text corpora could efficiently capture semantic relationships through skip-gram and continuous bag-of-words architectures. It established the paradigm of learning transferable representations from unlabeled data, inspiring subsequent advances in neural language modeling.

The concept of contextualized word representations was significantly advanced by ELMo [35], which employed bidirectional LSTMs to learn deep contextual features, and subsequently by BERT [36], which introduced masked language modeling and next sentence prediction to pre-train deep bidirectional Transformers. Concurrently, GPT [37] pioneered the generative pre-training approach using Transformer decoders with autoregressive language modeling, proving that pre-training on large unlabeled corpora followed by fine-tuning could achieve strong performance across diverse downstream tasks.

The scaling of language models later revealed emergent capabilities. GPT-3 [38] demonstrated that models of sufficient scale could perform few-shot learning without gradient updates, generalizing to novel tasks strictly through in-context examples. This scaling paradigm continues to drive the development of increasingly capable LLMs exhibiting sophisticated language understanding, generation, and reasoning abilities.

Vision-Language Models (VLMs) extended this pre-training paradigm to multimodal understanding. CLIP [39] enabled zero-shot image classification by aligning visual and textual representations through contrastive learning on web-scale image-text pairs, while subsequent models achieved even more sophisticated visual-language reasoning [40], [41]. Building upon these advances, Vision-Language-Action (VLA) models [16] integrate visual perception, language understanding, and action generation for embodied AI applications. For instance, RT-2 [17] demonstrated that VLMs could be transformed into VLA models by expressing robotic actions as text tokens, enabling web-scale knowledge transfer to robotic control. OpenVLA [18] further advanced this direction as an open-source model pre-trained on diverse robot demonstrations, enabling generalizable manipulation across multiple embodiments. In our framework, VLA models are deployed at edge nodes to analyze human activities and generate structured contribution evaluations.

III. ARCHITECTURE OF DIGITAL-PHYSICAL CAA

This section proposes a governance closed-loop architecture for the digital-physical CAA, which constructs a proactive data acquisition mechanism through physical world contribution evaluation technology based on VLA agents, bridging the gap between on-chain decisions and core activities in the physical world. As shown in Figure 1, this closed-loop architecture consists of three core modules: on-chain decision-making, off-chain execution, and physical world contribution evaluation.

The on-chain decision module is the core of consensus formation and contracted expression of digital CAA governance, mainly consisting of a series of smart contracts such as identity contracts, node control contracts, decision contracts, and contribution contracts. The digital CAA on-chain governance layer follows the basic process of proposal, voting, decision, and execution [19], [20]. Any community member of the digital CAA can initiate a governance proposal and collect feedback through discussions in the off-chain community. Subsequently, the proposal is authorized by members holding specific identity credentials for a formal proposal, and is reviewed and voted upon via the decision contract to generate a decision document or be abolished. If the proposal passes, the on-chain decision layer will simultaneously generate a text decision to guide contributors in the real world and task information to guide physical world contribution observation nodes. The former is conveyed to off-chain contributors through electronic signature documents, etc., while the latter is handed over to off-chain task control nodes via automated programs.

The final decision results generated by the group decision-making of the digital CAA in the on-chain decision module are used to guide the actual task execution of the real CAA in the off-chain execution module. The off-chain execution module involves the distributed execution and feedback of on-chain decisions in the physical world. This stage relies on incentive mechanisms to drive the advancement of off-chain tasks, aiming to establish execution specifications and collaboration mechanisms through standardized contracts, forming a positive feedback loop constrained by a reward and punishment mechanism [21]. In this process, off-chain staff receive corresponding remuneration based on their input in the task.

The physical contribution evaluation module utilizes task control nodes deployed in the cloud and task execution nodes deployed on edge devices to evaluate the contributions of human activities in the actual CAA. This module is key to achieving digital-physical integrated contribution evaluation and is the main logical layer for the digital CAA to proactively acquire and verify physical world data. When the on-chain decision generates task prompt information, this task is sent to the task control agent node, which decomposes the task into subtasks oriented towards different DePIN devices, and delegates observation tasks based on the capabilities and reputations of the edge nodes. After receiving the subtask, the edge node calls the corresponding sensing device to proactively observe the physical activities of the off-chain execution layer, and transforms unstructured physical phenomena into structured contribution data that conforms to a quantifiable format. The task evaluation node in the cloud will collect the feedback information from the edge nodes and calculate the physical world activity contribution score. Subsequently, the off-chain program automatically calls the

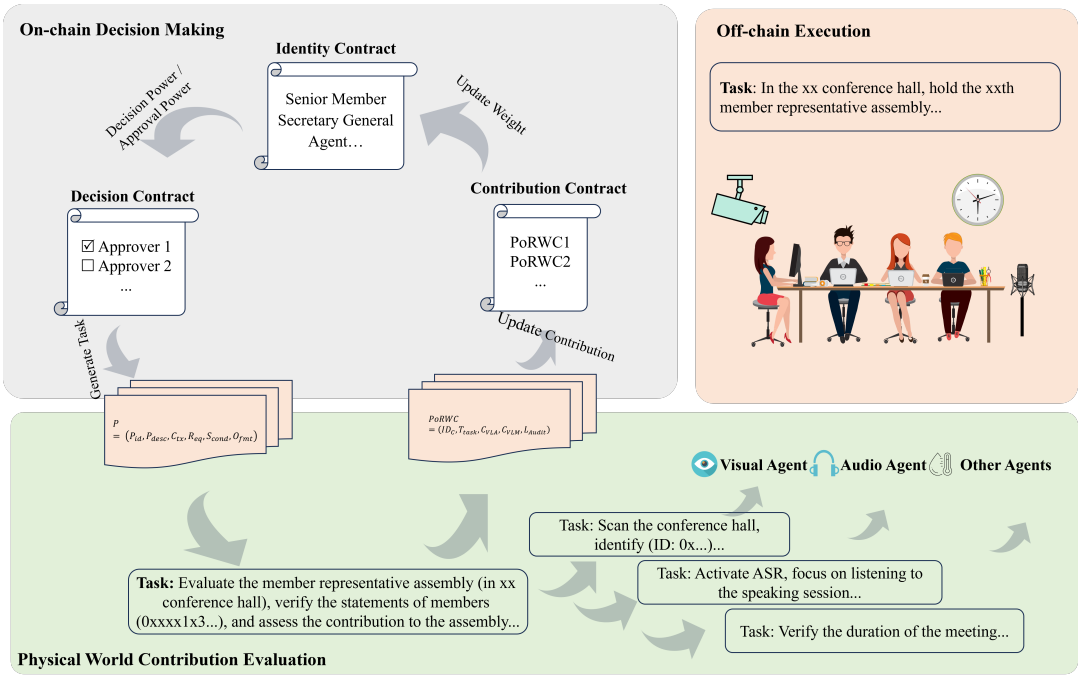

FIGURE 1: Trusted governance closed-loop architecture of digital-physical CAA. The framework comprises three interconnected modules: the on-chain decision module (left), which handles proposal submission, voting, and smart contract execution; the off-chain execution module (center), where physical-world tasks are carried out by distributed contributors; and the physical contribution evaluation module (right), where DePIN-connected edge nodes equipped with VLA models observe, analyze, and quantify real-world contributions into verifiable on-chain credentials.

contribution contract to generate a complete proof of real-world contribution for the activity.

Next, the two core technologies in this closed-loop architecture will be elaborated in detail: the task contribution evaluation method in the actual CAA and the trustworthy identity system in the digital CAA.

IV. VLA-BASED ACTUAL CAA CONTRIBUTION EVALUATION METHOD

A. TASK EXECUTION OF ACTUAL CAA

The activity contribution evaluation task of the actual CAA can be represented by a structured tuple P, where each tuple corresponds to an entity's task evaluation description:

$$P = (P_{id}, P_{desc}, C_{tx}, R_{eq}, S_{cond}, O_{fmt}) \tag{1}$$

Where P_{id} is the unique identifier of the task. P_{desc} is the natural language description of the task, used to guide observation nodes to conduct proactive observation. C_{tx} is the spatio-temporal context of task execution, containing event information, address, time window, etc.. R_{eq} is the dependency and permission for task execution, containing the assigned identity token of each observation node. S_{cond} is the declarative condition for task success. O_{fmt} is the output format that observation nodes must follow, as shown in Figure 2. The execution of physical world contribution tasks is divided into two phases: task planning and task delegation.

First, the task control node will decompose task P into one or more logical subtasks s_{sub} that can be completed using available tools; this process will be automatically executed by large language models. The planning function $Plan$ can be defined as:

$$s_{sub} = Plan(P, A_{available}) \tag{2}$$

Where P is the top-level task tuple received by the task planning agent, and $A_{available}$ represents the current set of available edge nodes. Then, the subtask s_{sub} is dispatched to one or more edge nodes bound with DePIN devices, namely:

$$A_{s_{sub}} = Delegate(s_{sub}, A_{available}) \tag{3}$$

Where $A_{s_{sub}}$ is the set of VLA agents selected to execute this subtask, which is a subset of all VLA agents qualified and capable of executing the subtask, i.e., $A_{s_{sub}} \subseteq A_{available}$.

B. CONTRIBUTION EVALUATION AND RECORDING BASED ON PORWC

The PoRWC is constructed based on the calculation results of the subtasks. PoRWC is the core to achieving trustworthy quantification of physical contributions, and its formalized structure is as follows:

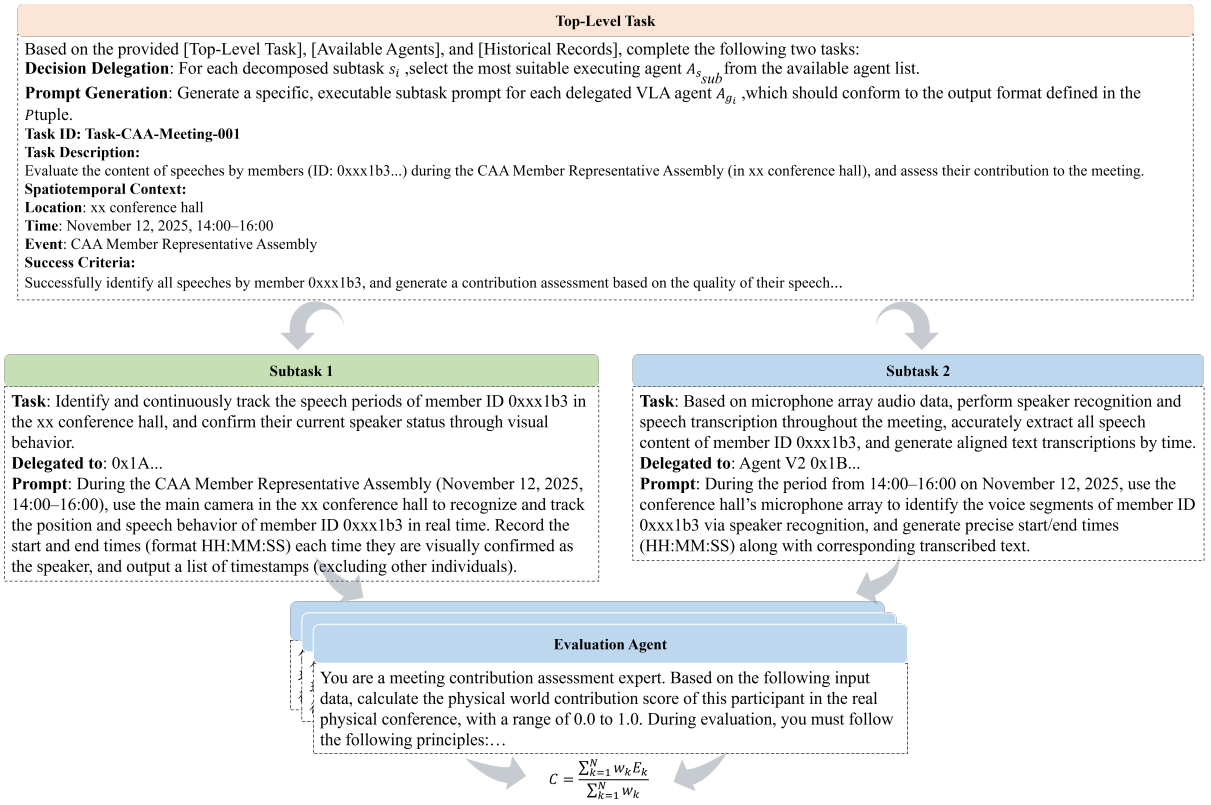

FIGURE 2: Contribution assessment process of real-world CAA activities. The upper portion illustrates the task decomposition from the top-level task tuple P into subtasks s_{sub} by the task control node. The lower portion shows the execution pipeline at edge nodes, where DePIN sensors collect multimodal data, VLA agents process the data into structured credentials, and cloud-based VLM evaluators independently score the contributions to produce the final PoRWC.

$$PoRWC = (ID_C, T_{task}, C_{VLA}, C_{VLM}, L_{Audit}) \quad (4)$$

Where ID_C represents the core identifier of the credential, containing information such as the unique identifier of the physical world contribution credential $PoRWC_{id}$, the associated task identifier P_{id}, and the creation timestamp t, used for indexing and tracing the credential. T_{task} is the description information from the evaluation task P, containing the original task description and the decomposed subtask description information s_{sub}. C_{VLA} represents the execution result of the VLA agent deployed at the edge node, containing only output information in a standard format. C_{VLM} identifies the execution result of the task evaluation agent deployed at the cloud node, containing the evaluation value E_k given by each evaluation agent k. L_{Audit} is the audit flag bit, identifying whether the credential has been manually audited or has exceeded the audit period. PoRWC records the core information of the physical world contribution evaluation process; in the actual calculation of contribution, it can be quantified by the following formula:

$$C = Eval(PoRWC) = \frac{\sum_{k=1}^{N} w_k E_k}{\sum_{k=1}^{N} w_k} \quad (5)$$

Where the reputation weight of node k is w_k, and its contribution evaluation result is E_k. It should be noted that this framework provides a general-purpose weighted aggregation method for contribution quantification; the specific calibration of weights w_k is intended to be application-dependent and configurable according to the governance requirements of different organizations. In practice, organizations can define weight assignment policies based on factors such as node historical accuracy, hardware certification level, and domain relevance, allowing the framework to adapt to diverse governance scenarios without prescribing a fixed calibration strategy.

Since the evaluation results of the models are not always reliable, this mechanism introduces a contribution audit mechanism. The core of this mechanism is to introduce a mandatory audit window period after the contribution calculation is completed, which has a fixed duration T_{audit}. During this period, any network member holding sufficient reputation values can raise a dispute challenge against the evaluation consensus C. The challenger needs to stake a reputation value no less than R_{stake} to submit a structured objection statement, clarifying the points of dispute and attaching supporting data. The system will automatically verify the challenger's staked amount and current reputation value

$R_{challenger}$; if both meet the threshold requirements, the decentralized arbitration process is officially initiated.

In the arbitration phase, existing decentralized arbitration protocols (such as Kleros [22]) can be adopted, where a randomized jury conducts on-chain deliberation on the original PoRWC records, VLA output data, and challenge evidence. If the ruling supports the original evaluation result, the reputation value staked by the challenger is deducted proportionally and distributed to the affected edge nodes; if the ruling overturns the original result, it triggers a state rollback of the PoRWC, corrects the contribution value C, and returns the reputation value staked by the challenger, while simultaneously applying a punitive decay to the reputation value of the original VLM evaluation node based on the degree of consensus deviation.

The final arbitration result will atomically update the on-chain credential status, appending the audit flag bit L_{audit} to the PoRWC structure to ensure the full lifecycle traceability of the physical world contribution records.

C. RELIABILITY AND LIMITATIONS OF VLA-BASED EVALUATION

While VLA models provide a promising approach for automated contribution evaluation, their deployment in governance systems requires careful consideration of potential failure cases and inherent biases. First, VLA models may produce inaccurate assessments under adverse environmental conditions. For example, poor lighting, occlusion, or unusual camera angles at event venues can degrade the quality of visual perception, leading to incorrect identity verification or behavioral analysis results. Second, current VLA models may exhibit biases inherited from their training data. If the training corpus underrepresents certain demographic groups or cultural contexts, the model could systematically undervalue contributions from those populations. Third, adversarial manipulation remains a concern: participants could potentially game the evaluation system through deliberate behavioral patterns designed to inflate contribution scores without substantive participation.

To mitigate these risks, the proposed framework incorporates several safeguards. The multi-evaluator consensus mechanism in Equation (5) reduces the impact of any single erroneous assessment by aggregating scores from multiple independent VLM evaluation nodes with different analytical focuses. The mandatory audit window with staked dispute challenges provides a human-in-the-loop correction mechanism that can override automated evaluations when errors are detected. Furthermore, the reputation decay mechanism penalizes evaluation nodes that produce consistently disputed results, creating a self-correcting feedback loop. Nevertheless, these mitigations do not eliminate all risks, and future work should investigate formal robustness guarantees, bias auditing procedures for deployed VLA models, and the development of domain-specific fine-tuning strategies to improve evaluation accuracy in academic governance scenarios.

V. DEPIN-BASED DIGITAL CAA IDENTITY SYSTEM

When the digital CAA conducts governance, it needs to establish an accurate cognition of the capabilities of execution individuals, execution agents, and DePIN devices. This cognition of complex capabilities usually originates from expert knowledge verification in real society and the entity's own historical behavioral data. However, traditional identity management methods are difficult to carry such complex dynamic descriptive information. Therefore, this paper adopts a hybrid identity management system that integrates semantic web technology and DID technology. Its core idea is to establish an anchoring mechanism between real-world trust and digital trust through a distributed authentication network, storing core anchoring information and identity relationships on-chain, while placing rich semantic description information in off-chain network graphs, thereby realizing the tokenized mapping of physical world trust relationships. The overall architecture of this identity mechanism is shown in Figure 3.

To enhance the semantic expression capability and data reliability of the identity system, this method constructs a decentralized identity system based on the W3C DID standard, uses non-fungible tokens to mint a unique and persistent identifier for each entity, and is extended to support RDF semantic triple modeling. The core data structure is shown in Figure 4. Through semantic web technology, this method aggregates discrete RDF triples into queryable and inferable on-chain capability graphs. Compared to flat key-value storage, the graph structure can flexibly express composite capabilities, cross-entity authorization relationships, and capability decay models, significantly improving the machine readability and composability of identity declarations.

To further enhance tamper resistance and data availability, this scheme introduces bidirectional hash anchoring. The `contentHash` field of each RDF triple not only covers the off-chain capability description file but also recursively includes the hash values of all referenced subgraph data, forming a hash chain. This ensures that any modification to a fragment of the off-chain capability graph will trigger the invalidation of the on-chain fingerprint. In the off-chain description, the `tokenId` field is compulsorily declared to explicitly bind the ID of its on-chain NFT credential. During an audit, the verification contract executes a cross-check, namely: `keccak256(Id || offChainData) == contentHash`, to ensure that off-chain data and on-chain credentials strictly correspond, preventing off-chain data replacement attacks.

The reputation value is automatically calculated by on-chain contracts, reflecting the reliability of the entity's decision-making history and the maturity of its exercise of rights, serving as a non-consumable accumulated credit capital. Its growth depends on the verifiable value created by the entity in physical world activities and the corresponding on-chain contribution credentials. Let the standardized contribution of entity N_i verified by the community at time t be

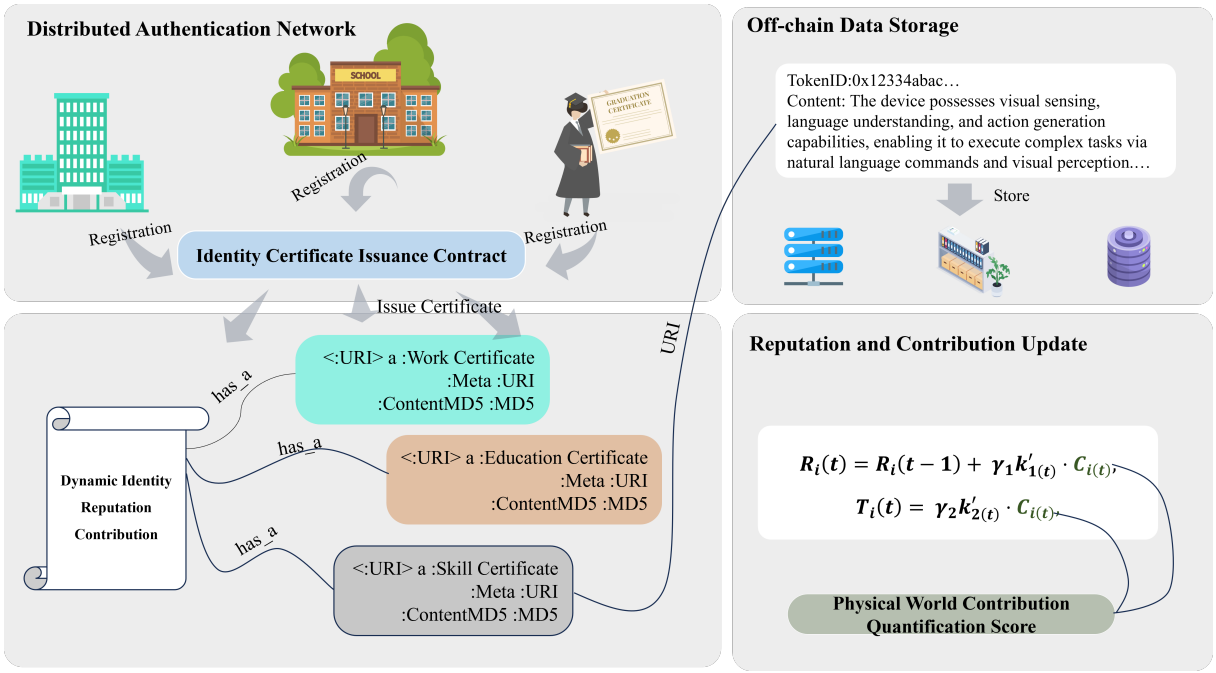

FIGURE 3: Hybrid identity architecture integrating on-chain DID anchoring with off-chain semantic descriptions. The on-chain layer stores core identity credentials as non-fungible tokens with hash-based integrity anchors, while the off-chain layer maintains rich RDF-based capability graphs that describe entity attributes, qualifications, and cross-entity authorization relationships. Bidirectional hash anchoring ensures tamper resistance between the two layers.

```
struct RDFTriple {
    address subject;        // Described resource address
    string predicate;       // Relationship description
    uint256 object;         // Target address or value
    bytes32 contentHash;    // Hash value of the off - chain content
    bytes signature;        // Publisher signature
}
```

FIGURE 4: Core data structure of the DID-based identity token. Each identity token contains the entity DID, role type, capability description hash (`contentHash`), and metadata fields. The `contentHash` field links to off-chain RDF semantic triples that encode detailed capability and qualification information, enabling machine-readable identity verification and cross-entity interoperability.

$C_i(t)$, its reputation value $R_i(t)$ is dynamically calculated as:

$$R_i(t) = R_i(t-1) + \gamma_1 k'_1(t) \cdot C_i(t) \tag{6}$$

Where $k'_1(t)$ is the scaling factor of reputation, and γ_1 is the global reputation weight coefficient.

The contribution value, on the other hand, is a consumable equity token within the organization, representing the quota of services or resources that can be redeemed in the current period. Unlike the accumulative nature of reputation, contribution values are deducted upon use. Its calculation method is:

$$T_i(t) = T_i(t-1) + \gamma_2 k'_2(t) \cdot C_i(t) \tag{7}$$

Where $k'_2(t)$ is the contribution conversion rate, and γ_2 is the incentive adjustment coefficient. Contribution values are reset or rolled over periodically to ensure the timeliness of resource flows.

VI. TRUSTWORTHY GOVERNANCE MECHANISM OF PHYSICAL-DIGITAL CAA COLLABORATION

This paper takes the CAA Member Representative Assembly as a case study to demonstrate how the DePIN and VLA-based endogenous trustworthy governance mechanism achieves the perception, quantification, and dynamic incentivization of complex physical governance activities. Rather than conducting large-scale empirical experiments, this case study adopts a deductive approach to illustrate how each component of the proposed framework operates in a concrete governance scenario. By tracing the complete workflow from task generation through contribution evaluation to on-chain credential minting, we demonstrate the internal consistency and operational feasibility of the framework, showing how the theoretical constructs translate into a coherent governance process.

During the convening of the Member Representative Assembly, distributed sensing devices such as smart cameras, microphone arrays, and environmental sensors deployed at the venue relying on the DePIN network form a trustworthy

data collection network covering the entire venue. Each type of device is bound to an on-chain identity via DID, and its hardware fingerprint, firmware version, and calibration certificate are stored as RDF triples in the capability graph. For example, the 4K camera `did:depin:camera-001` located in the main venue records its basic parameters such as resolution and field of view, and anchors its detailed off-chain description through `contentHash`; an example is shown in Figure 5.

On this basis, the VLA model is deployed at the edge nodes of DePIN, endowing them with closed-loop capabilities of perception, understanding, and action. When the on-chain decision contract generates the meeting observation task $P_{congress}$, the task control node automatically triggers the task planning function:

$$s_{sub} = Plan(P_{congress}, A_{available}) \tag{8}$$

Where $P_{congress}$ contains the task description P_{desc} to conduct contribution evaluation for the entire process of the Member Representative Assembly.

The task delegation function $Delegate(s_{sub}, A_{available})$ matches edge nodes based on their reputation values w_k and special capability graphs. For example, the subtask s_{speak} needs to be delegated to the edge node set $A_{s_{speak}}$ equipped with advanced microphone arrays and emotion analysis VLA models.

For the subtask s_{attend}, when an off-chain member `did:caa:member-0521` enters the main venue, the DePIN camera `did:depin:camera-001`, controlled by the VL agent node, captures their facial image, forms an original data stream Data, and conducts preliminary analysis, mainly including:

- **Identity verification:** Calling the facial recognition model to match the biometric hash in the on-chain DID document to confirm identity consistency.
- **Spatio-temporal anchoring:** Combining the camera intrinsic calibration matrix and the venue GIS map to calculate their coordinate position in the venue.
- **Behavioral analysis:** Through eye tracking, counting the proportion of time their gaze is focused on the podium.

The above analysis results are encapsulated into an edge execution credential in a standard format. Subsequently, this execution credential is broadcast to multiple task evaluation nodes, and each node independently calculates the evaluation value. For example, in this case, three evaluation nodes are as follows:

- `did:caa:vlm-evaluator-01`: Focuses on legal compliance, verifying whether the representative has the qualification to attend the meeting.
- `did:caa:vlm-evaluator-02`: Focuses on behavioral semantic analysis, judging whether the substantive participation threshold is met.
- `did:caa:vlm-evaluator-03`: Focuses on data integrity, verifying that the signature matches the camera hardware fingerprint.

Figure 6 shows an example of possible evaluation results.

After the evaluation results are generated, the off-chain aggregation program calls the physical world contribution credential minting function of the contribution contract, and through the oracle mechanism, passes in data such as task ID, edge credentials, and the cloud evaluation set, whereby the attendance credential for `did:caa:member-0521` is automatically minted.

Simultaneously, the on-chain contribution contract automatically executes the quantification formula:

$$C = \frac{\sum_{k=1}^{N} w_k E_k}{\sum_{k=1}^{N} w_k} \approx 0.9408 \tag{9}$$

This contribution value will trigger the update of the reputation and contribution value for `did:caa:member-0521`:

$$R_{0521}(t) = R_{0521}(t-1) + \gamma_1 k_1'(t) \cdot 0.9408 \tag{10}$$

$$T_{0521}(t) = T_{0521}(t-1) + \gamma_2 k_2'(t) \cdot 0.9408 \tag{11}$$

The edge node `did:depin:edge-007` will also experience a corresponding growth in reputation due to providing high-quality C_{VLA}. If there is no dispute in subsequent audits, the reputation of each VLM evaluation node will receive a minor growth based on their participation degree.

If there is no staked challenge within 72 hours, it is automatically marked as audited, the credential status is frozen, and the representative's final attendance contribution value is permanently deposited and mapped to the current consumable contribution tokens. Similar processes will occur synchronously for behavioral contributions such as hosting, speaking, and voting.

VII. CONCLUSION

Aiming at the problem of lacking a trustworthy closed-loop feedback mechanism between on-chain decisions and physical world activities in the governance of digital-physical integrated parallel CAA, this paper proposes an endogenous contribution evaluation method integrating DePIN and VLA. Furthermore, taking the CAA Member Representative Assembly as an empirical scenario, this study explores the feasibility of this framework in realizing contribution quantification, process traceability, and incentive automation in complex academic governance activities. This study expands the boundaries of digital CAA governance at the theoretical level, incorporating human actions in the physical world into the scope of on-chain consensus, and provides an engineering implementation path for its practical application [23]. In summary, the digital CAA proposed in previous studies mainly solved the problems of artificial system modeling and infrastructure design, while the trustworthy governance method based on DePIN and VLA proposed in this paper

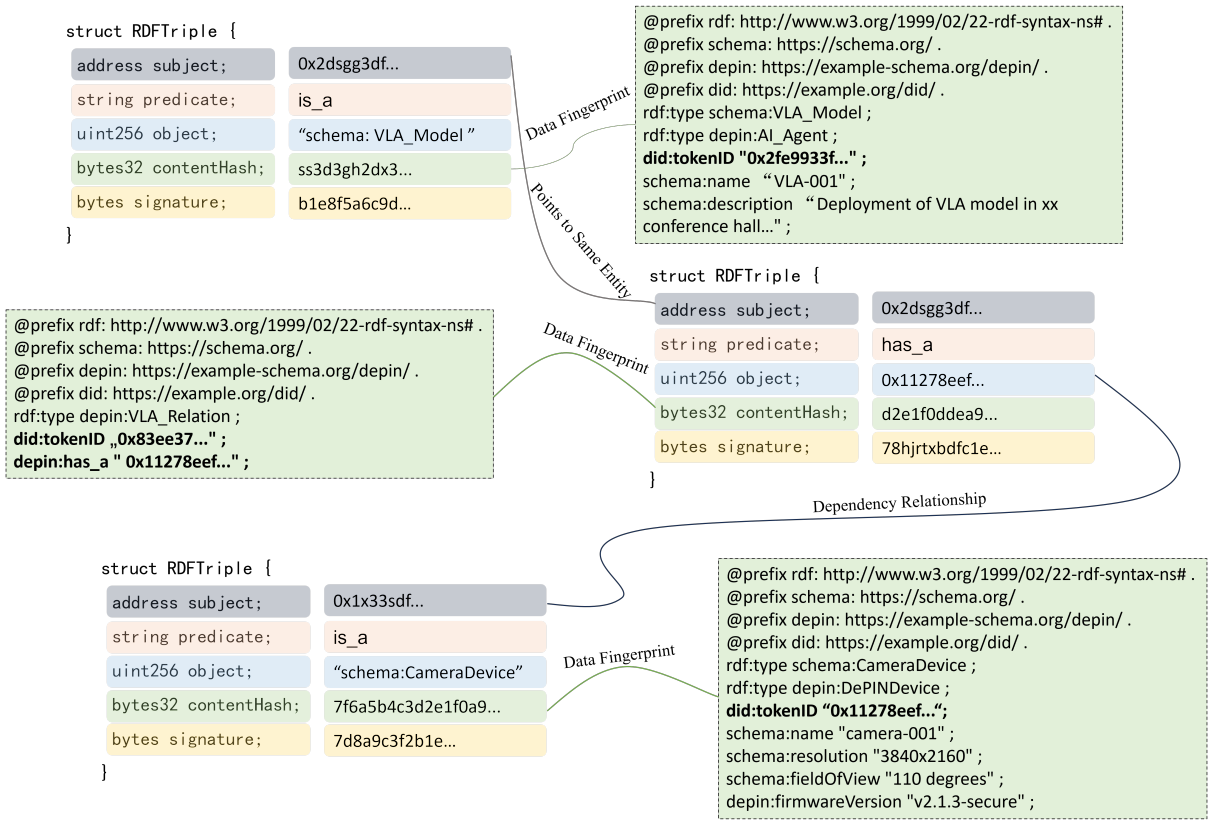

FIGURE 5: An illustrative example of DePIN device and member identity registration in the CAA Member Representative Assembly scenario. The left portion shows the on-chain identity token for a 4K camera (`did:depin:camera-001`) with its hardware fingerprint and capability hash, while the right portion displays the corresponding off-chain RDF description detailing resolution, field of view, and firmware version.

```
{
  "evaluator_did": "did:caa:vlm-01",
  "reputation_weight": 120.5,
  "evaluation_score": 1.0,
},
{
  "evaluator_did": "did:caa:vlm-02",
  "reputation_weight": 150.0,
  "evaluation_score": 0.85,
},
{
  "evaluator_did": "did:caa:vlm-03",
  "reputation_weight": 110.0,
  "evaluation_score": 1.0,
}
```

FIGURE 6: Example of multi-evaluator scoring results for the attendance contribution of member `did:caa:member-0521`. Three VLM evaluation nodes independently assess the contribution from different perspectives: legal compliance (vlm-01), behavioral semantics (vlm-02), and data integrity (vlm-03). Each evaluator provides a reputation-weighted score that is aggregated into the final contribution value via Equation (5).

mainly focuses on solving the problem of forming a trustworthy feedback closed loop in digital-physical integration; they respectively serve the artificial system and parallel execution links in the parallel CAA paradigm.

The contribution evaluation method proposed in this paper provides a human-machine hybrid incentive approach, where VLA-based autonomous assessment is complemented by human-initiated dispute mechanisms and decentralized arbitration, ensuring that the governance process balances automation efficiency with human oversight. While existing oracle solutions and decentralized governance mechanisms address the general problem of bridging on-chain and off-chain data, the focus of this work is distinct: we propose a method that can directly evaluate physical-world contributions for governance purposes, rather than relying on generic data feeds. Formal analysis of incentive compatibility, security guarantees, and comparative evaluation against existing contribution and reputation systems constitute important directions for future research.

However, current research still faces challenges in multiple aspects. Firstly, in the construction of the digital-physical integrated parallel CAA, although we have successfully incorporated off-chain activities into on-chain consensus and built a trustworthy verification closed loop, there is still a

lack of a computational experiment framework for large-scale social computing and counterfactual deduction based on these high-fidelity on-chain behavioral data, limiting the quantitative analysis capability of the long-term evolution laws of complex governance strategies. Secondly, in terms of data privacy and governance effectiveness, on-site perception involves biometric and behavioral data, posing a risk of privacy leakage, and the current model reasoning, on-chain storage, and verification overheads are high. The security and privacy implications of technologies such as facial recognition and gaze tracking in governance contexts are particularly significant, and future work will systematically address privacy-preserving computation techniques, such as federated learning and differential privacy, to ensure that sensitive biometric data is processed locally at edge nodes without exposing raw data to the network. Finally, regarding ecological expansion and governance ethics, further research is needed on governance protocols for cross-organizational contribution mutual recognition, scalable evaluation models for more complex physical scenarios, and the definition of rights and responsibilities and value alignment models of AI agents in governance.

REFERENCES

[1] Ying Kou, "Chinese association of automation: building a first-class society to help scientific and technological innovation," Society, no. 8, pp. 24-27, 2023.

[2] N. Zhang, X. L. Liang, S. T. Guan, et al., "The DAO-based digital CAA: a novel digital transformation paradigm for scientific and technological associations," Chinese Journal of Intelligent Science and Technology, vol. 7, no. 1, pp. 30-40, 2025.

[3] V. Buterin, "A next-generation smart contract and decentralized application platform," White Paper, vol. 3, no. 37, pp. 2-1, 2014.

[4] S. Wang, L. W. Ouyang, Y. Yuan, X. C. Ni, X. Han, and F.-Y. Wang, "Blockchain-enabled smart contracts: architecture, applications, and future trends," IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 49, no. 11, pp. 2266-2277, 2019.

[5] J. J. Li, R. Qin, W. W. Ding, G. Wang, T. Wang, and F.-Y. Wang, "A new framework for Web3-powered decentralized autonomous organizations and operations," Acta Automatica Sinica, vol. 49, no. 5, pp. 985-998, 2023.

[6] A. Wright, "The rise of decentralized autonomous organizations: opportunities and challenges," Stanford Journal of Blockchain Law & Policy, vol. 4, p. 1, 2020.

[7] J. J. Li, X. G. Liang, R. Qin, and F.-Y. Wang, "True autonomous organizations and operations for Web3," Journal of Cyber-Physical-Social Intelligence, vol. 2, no. 1, pp. 1-8, 2023.

[8] H. Gao, "Research on phenomenon, cause and solution in development of Chinese scientific and technical community," Shandong University, 2007.

[9] H.-Y. Yu, "China's associations have tremendous potential for 'star-level' development," Science and Technology Daily, 2022-06-27(2).

[10] F.-R. Meng, T.-X. Ma, M. Yuan, et al., "International comparison of the development status of world-class scientific and technological associations: a perspective based on practical dynamics," Science & Technology Review, vol. 39, no. 10, pp. 80-89, 2021.

[11] H. Soodeh and S. Hossein, "The role of agentic AI in shaping a smart future: a systematic review," Array, vol. 26, 100399, 2025.

[12] G. Caldarelli, "Understanding the blockchain oracle problem: a call for action," Information, vol. 11, no. 11, p. 509, 2020.

[13] A. Pasdar, Y. Lee, and Z. Dong, "Connect API with blockchain: a survey on blockchain oracle implementation," ACM Computing Surveys, vol. 55, no. 10, pp. 1-39, 2023.

[14] G. Caldarelli, "Real-world blockchain applications under the lens of the oracle problem. A systematic literature review," 2020 IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD), pp. 1-6, 2020.

[15] Z. Lin, T. Wang, L. Shi, et al., "Decentralized physical infrastructure networks (DePIN): challenges and opportunities," IEEE Network, vol. 39, no. 2, pp. 91-99, 2025.

[16] A. Brohan, N. Brown, J. Carbajal, et al., "RT-2: vision-language-action models: transferring web-scale knowledge to robotic control," arXiv:2307.15818, 2023.

[17] M. J. Kim, K. Pertsch, S. Karamcheti, et al., "Openvla: an open-source vision-language-action model," arXiv preprint arXiv:2406.09246, 2024.

[18] W. Guan, Q. Hu, A. Li, et al., "Efficient vision-language-action models for embodied manipulation: a systematic survey," arXiv preprint arXiv:2510.17111, 2025.

[19] T. Börgers, An introduction to the theory of mechanism design, Oxford University Press, 2015.

[20] X. L. Liang, R. Qin, J. J. Li, et al., "The engineering of circular causality for specialization and design of complex systems: Cad2CAS and CasCAD2," Frontiers of Information Technology & Electronic Engineering, vol. 25, no. 2, pp. 323-332, 2024.

[21] D. Bergemann and S. Morris, "Robust mechanism design," Econometrica, pp. 1771-1813, 2005.

[22] L. Bergolla, K. Seif, and C. Eken, "Kleros: a socio-legal case study of decentralized justice & blockchain arbitration," Ohio State Journal on Dispute Resolution, vol. 37, p. 55, 2022.

[23] R. Qin, W. W. Ding, J. J. Li, S. T. Guan, G. Wang, Y. Ren, and Z. Qu, "Web3-based decentralized autonomous organizations and operations: architectures, models, and mechanisms," IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 53, no. 4, pp. 2073-2082, 2023.

[24] S. Du, "The DAO hack explained: A brief history," Medium, 2018. [Online]. Available: https://medium.com/@samdu/the-dao-hack-explained-a-brief-history-9f5e8a1b4c4f

[25] P. Daian, "Analysis of the DAO exploit," Hacking, Distributed, 2016. [Online]. Available: https://hackingdistributed.com/2016/06/18/analysis-of-the-dao-exploit/

[26] L. W. Cong, Y. Li, and N. Wang, "Token-based platform finance," Journal of Financial Economics, vol. 144, no. 3, pp. 972-991, 2022.

[27] J. J. Li, X. L. Liang, R. Qin, and F.-Y. Wang, "True autonomous organizations and operations for Web3," Journal of Cyber-Physical-Social Intelligence, vol. 2, no. 1, pp. 1-8, 2023.

[28] J. J. Li, X. L. Liang, R. Qin, and F. Wang, "From DAO to TAO: Finding the essence of decentralization," in Proc. IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2023, pp. 1-4.

[29] J. J. Li and F.-Y. Wang, "The TAO of blockchain intelligence for intelligent Web 3.0," IEEE/CAA Journal of Automatica Sinica, vol. 10, no. 12, pp. 2183-2186, 2023.

[30] D. W. Chadwick, G. D. B. C. C. Callas, et al., "Improved identity management with verifiable credentials and FIDO," IEEE Communications Standards Magazine, vol. 3, no. 4, pp. 46-53, 2019.

[31] W3C, "Decentralized identifiers (DIDs) v1.0," W3C Recommendation, 2022. [Online]. Available: https://www.w3.org/TR/did-core/

[32] M. Mühle, A. Grüner, T. Gayvoronskaya, and C. Meinel, "A survey on essential components of a self-sovereign identity," Computer Science Review, vol. 30, pp. 80-86, 2018.

[33] M. C. Ballandies, H. Wang, A. C. Law, et al., "A taxonomy for blockchain-based decentralized physical infrastructure networks (DePIN)," in Proc. IEEE World Forum on Internet of Things (WF-IoT), 2023, pp. 1-7.

[34] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in Proc. International Conference on Learning Representations (ICLR), 2013.

[35] M. Peters, M. Neumann, M. Iyyer, et al., "Deep contextualized word representations," in Proc. Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2018, pp. 2227-2237.

[36] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2019, pp. 4171-4186.

[37] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," OpenAI Technical Report, 2018.

[38] T. Brown, B. Mann, N. Ryder, et al., "Language models are few-shot learners," in Advances in Neural Information Processing Systems (NeurIPS), 2020, pp. 1877-1901.

[39] A. Radford, J. W. Kim, C. Hallacy, et al., "Learning transferable visual models from natural language supervision," in Proc. International Conference on Machine Learning (ICML), 2021, pp. 8748-8763.

[40] J. Li, D. Li, C. Xiong, and S. Hoi, "BLIP: Bootstrapping language-image pre-training for unified vision-language understanding and generation," in Proc. International Conference on Machine Learning (ICML), 2022, pp. 12888-12900.
[41] H. Liu, C. Li, Y. Wu, and Y. J. Lee, "Visual instruction tuning," in Advances in Neural Information Processing Systems (NeurIPS), 2023, pp. 34892-34916.

XIAOLONG LIANG received the B.S. and M.S. degrees in computer science and technology from Shandong University, Jinan, Shandong, China, in 2011 and 2016, respectively. He received the Ph.D. degree from the Faculty of Innovation Engineering, Macau University of Science and Technology, Macau, China, in 2024. His main interests include parallel intelligence, parallel governance, blockchain, knowledge graph, and DAO.

RUI QIN received the Ph.D. degree in computer application technology from University of Chinese Academy of Sciences, Beijing, China, in 2016. She is currently an associate professor with The State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China. Her research interests include blockchain, DAO and parallel management.

JUANJUAN LI received her M.S. degree in economics from Renmin University of China and Ph.D. degree in philosophy from Beijing Institute of Technology, Beijing, China. Currently she is an associate professor with The State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China. Her research interests include blockchain, DAO and parallel management.

THALES S. W. THESEUS is a Senior Researcher with the Parallel Intelligence DeSci Center, Obuda University, Budapest, Hungary. His research interests include parallel art and music for education and intelligent living and working environments.