Verifiable Internet for Artificial Intelligence: The Convergence of Crypto, Internet and AI

Pre-publication

Trace Labs OriginTrail Core developers Feedback & Discussion : https://discord.gg/9CvynQ9kYe

"We live in a time of abundant connectivity and alas abundant misinformation. The OriginTrail Decentralized Knowledge Graph (DKG) is an evolving tool for finding the truth in knowledge. In particular, we see knowledge graphs improving the fidelity of artificial intelligence." Dr Bob Metcalfe, Ethernet inventor & Internet pioneer

Abstract

This whitepaper presents a vision for the future of Artificial Intelligence (AI) through the concept of a Verifiable Internet for AI, leveraging the synergies of crypto, internet, and AI technologies. It introduces the Decentralized Knowledge Graph (DKG), a novel approach to ensure the provenance, integrity, and verifiability of information utilized by AI systems. It aims to address the challenges posed by misinformation, data ownership, intellectual property rights, and biases inherent in AI technologies, by synergizing neural and symbolic AI approaches with Web3 technologies. The whitepaper delves into the intricate mechanics of the DKG, how it enhances AI reliability with Knowledge Assets and supports autonomous paranets. The Verifiable Internet for AI is not only a technological advancement but a paradigm shift towards a more transparent, inclusive, and ethical digital landscape. This vision combines blockchain technology with the dynamic capabilities of AI to foster a democratic, ethically guided, and economically viable future for AI applications.

1 Purpose

The sudden rise of Artificial Intelligence (AI) in the mainstream is reshaping all of our interactions with knowledge. It changes how we discover it, how we process it for our advantage, and how we use it to create new knowledge. It is an unprecedented event in the history of civilization, as we receive an unseen boost to cognition capabilities, outperforming any previously available human or machine methods. It allows us to leverage endless amounts of data, identifying patterns and insights that would otherwise remain undiscovered. This enables creating adaptive and continuously self-improving solutions that are tailored to every individual.

As the technology is growing more powerful, its anomalies can be much more devastating. The challenge of misinformation is no longer limited to malicious printing of propaganda. Even abusing social networks for political manipulations may look minuscule compared to a lack of trust in solutions to which we are "outsourcing" our cognition. Systems that we would trust to process large amounts of knowledge and provide us inputs for our actions or even perform certain actions autonomously, have the highest possible requirements for transparency and verifiability. There should be no compromise in designing AI solutions when it comes to data ownership, information provenance, verifiability of information, or bias that would include any censorship-by-design approach. The risk of this revolution not unfolding in an inclusive way is a societal threat of establishing a monopoly on AI.

The present whitepaper discusses the building blocks of the Verifiable Internet for AI. To harness a harmonious development between Web3 fundamentals and rapidly deployed AI systems, we propose practical steps. In this approach, we aim to integrate the core Web3 technologies such as the OriginTrail Decentralized Knowledge Graph (DKG), Ethereum, and Polkadot (more specifically - NeuroWeb) and AI systems (OpenAI, Google, Microsoft, X's GrokAI, and others).. We believe that we can realize the potential of trusted AI by creating a Verifiable Internet for AI that is founded on principles of neutrality, inclusiveness, and usability.

2 Artificial Intelligence and its anomalies

Artificial Intelligence (AI) is a field of computer science that develops and studies intelligent machines. It is the simulation of human intelligence processes by machines, especially computer systems. Generative AI (GenAI) is a type of AI technology that can produce various types of content, such as text, imagery, audio, and other synthetic data. It uses generative neural network models to create new and original content in response to a prompt. It's been widely popularized by large language models (LLMs) such as OpenAI's ChatGPT, Google's Gemini, Meta's LLaMA. The performance of such well and widely-trained LLMs has increased tremendously quickly and allows them to provide intelligent responses to a wide spectrum of topics. The standalone impact of these models on the economy was estimated to be 4.4T USD by McKinsey [1]. In order for GenAI to fulfill its potential and, moreover, unlock new potential for growth, the following anomalies should be addressed.

Hallucinations. As the name suggests, Generative AI is not a deterministic system as its responses to prompts are generated on the basis of the large amounts of data that a model has been trained on. They are, therefore, probabilistic systems that are very good at "guessing" what the most appropriate response to your prompt is. While they might be very good at predicting a high-quality response, there is still a large chance that, for example, the response you receive might not be the correct answer to a question you posed. This is considered a "hallucination", a wrongful prediction of the correct response created by a GenAI system. Hallucinations can be entertaining and even inspiring, however, they are still generally an unwanted response, especially in situations where a precise response is of utmost importance (e.g. education, medicine, supply chain, science, construction ...).

Data ownership and intellectual property. The vast datasets the GenAI models are trained on may include copyrighted material to which the owner has not given specific access/license to use. This opens up questions of (1) the legality of using such data to advance (train) a GenAI model and (2) the intellectual property rights of the generative outputs (e.g. art piece or a text). Things get even more complex with using personal data or business-sensitive data which can have significant privacy concerns. The challenges don't stop there as disassociating copyrighted content from a model can only be done by retraining or fine-tuning which are both resource-intensive processes while data-forgetting techniques yield limited results.

Bias. The GenAI models are created by training on large amounts of data and by doing so the model also adopts biases present in the data. This leads to systemic prejudice appearing in the responses to prompts which creates a wide spectrum of problems - racial, national, gender, or political biases at the societal level or, for example, smaller-scale biases of preferring one provider over another in a fine-tuned model for an enterprise solution. Fighting existing biases in models can, again, be very resource intensive as it also has to be done through retraining or fine-tuning (similar to data disassociating techniques above).

Model collapse. The performance of GenAI, especially LLMs, is crucially reliant on the quality of the training data they consume. The data has to be relevant and diverse so the model can become stronger by training on it. In a scenario where data would be generated by the model itself (i.e. synthetic data), the model would not only stop getting better or plateau, it would actually start to degenerate [2]. Using synthetic data for training leads to reinforcing existing biases and it minimizes the "tails" in the data. Tails are the less-likely but critical parts of the data as they give the LLM outputs a greater variety of responses. Without them, misinterpretation of data starts to happen and the model can slide towards implosion. This threat is becoming much greater as the use of AI becomes more prevalent on the internet (the main source for LLM training data is scraping the internet). If the majority of content on the internet becomes AI generated (synthetic), the value this data has for AI systems training will decrease as it will be more and more dangerous to use it. What will grow in value, on the flip side, is verifiably non-AI generated real-world data.

3 Verifiable Internet for AI

In the near future, AI models will largely dominate the internet both in activity and content production. The disruption created by OpenAI's launch of GenAI large language model ChatGPT has forced all major technology companies to show their hands with the AI developments they made (e.g. Google's launch of Gemini, X's GrokAI or Meta's progress on LLaMA). Furthermore, the open source community has taken great leaps to follow the proprietary models, and with great success, providing useful components to the Verifiable internet for AI.

Text LLMs, however, are by far not the only GenAI models out there. Among others, you can find:

- Image generation (e.g. Stable Diffusion) the ability to create images either from text, image, text and image or image and image inputs,
- Video generation (e.g. OpenAI Sora) the ability to create a short video, up to 60s, with a text prompt
- Multimodal models (e.g. OpenAI CLIP) models that are capable of accepting multiple media as input as well as providing multiple media responses.

The AI-dominated internet future will therefore have people interacting with a plethora of AI models - anything from single-model interactions to agent-driven applications that will be able to carry out autonomous activities tapping into the capabilities of multiple agents. As AI penetration into the mainstream continues, such a world has a dire need to effectively address the above-mentioned anomalies of the current AI systems.

We identify the following key properties of the Verifiable Internet for AI: information provenance, information verifiability, and knowledge-sharing incentives.

Information provenance. Any AI system that we expect to operate at scale in a meaningful way needs to ensure transparency of its information provenance. We need to be able to validate that a response to a particular prompt was created from known sources that we can check. A known source should be, at the very least, equipped with information about the identity of the publisher, the time of publishing, and the related content (knowledge). Basing GenAI responses on known sources is the most effective way to **manage hallucinations** both for AI engineers as well as for knowledge consumers, as now users can always compare the response with the very source it was used to create it. Instead of blindly trusting an AI response, they can get the information "straight from the horse's mouth". The information provenance is also critical in ensuring that **proper intellectual property protection is achieved**. This can be as simple as referencing sources for a response all the way to a more complex implementation of royalties for generative creations based on original data - e.g. a generative digital art piece that was created on the basis of an original artwork. Lastly, transparency achieved by information provenance is also a **powerful tool to address bias**. While the information used can also have a bias in itself, the capability to access and filter sources allows the users to create their own critical thoughts around what is presented to them as a response.

Verifiable information. It's not enough just to find a source, we also need to check that the content our GenAI solution is consuming hasn't been tampered with. As we see GenAI solutions more widely used, they will also interact with a growing spectrum of global knowledge, a lot of which will be sourced from third parties. This in turn significantly increases the complexity of the knowledge network used in producing responses. In such a scenario, we need a cryptographic way of validating that whatever content was used doesn't only have provenance but also has integrity, meaning it hasn't changed since it was created.

Incentivizing high-quality knowledge. Access to shared open knowledge constructed in a collaborative way is mission-critical for the future of AI, especially since non-AI-generated content is expected to be surpassed in size by synthetic, AI-generated content in the coming period. The importance of it has also been highlighted by the Turing Award winner in the field of Deep Learning, Yann LeCun:

"The way you train that (AI) system will have to be crowdsourced ... if you want it to be a repository of all human knowledge, all humans need to contribute to it."

Even though the need may exist, open knowledge is often poised with the "tragedy of the commons" dilemma. The motivation to contribute to open knowledge can be perceived as a less desirable option to growing a knowledge base in private stores. This opens a strong requirement for a two-pronged incentivization model:

- Incentivizing open knowledge creation incentivizing publishing of open knowledge, especially one in the public domain, turns the tragedy of common considerations into a race for capturing incentives. If you do not add open knowledge another actor will.
- Monetizing proprietary knowledge allowing business models to develop on the basis of owning a piece of data or knowledge that is not in the public domain (e.g. anonymized personal data or company's know-how).

Both cases benefit greatly from a system where there is a way to determine the quality of knowledge which can be used as a yardstick measure for incentives distribution.

4 Ensuring knowledge provenance in Verifiable Internet for AI

Given the above requirements, we propose an effective way of establishing a new paradigm for the Verifiable Internet for AI, which employs multiple decentralized systems in a **Decentralized Retrieval Augmented Generation (dRAG) framework**.

The term Retrieval Augmented Generation (RAG) was coined by Patrick Lewis in a 2020 paper [3] and it represents a technique for enhancing the accuracy and reliability of GenAI models with facts fetched from external sources. This allows AI solutions to dynamically fetch relevant information before the generation process, enhancing the accuracy of responses by limiting the generation to re-working the retrieved inputs.

The dRAG advances the model by organizing external sources in a Decentralized Knowledge Graph (DKG)

while introducing incentives to grow a global, crowdsourced network of knowledge made available for AI models to use. The framework enables a hybrid AI system that brings together **neural** (e.g. LLMs) **and symbolic AI** (e.g. Knowledge Graph) methodologies. Contrary to using a solely neural AI approach which is based on vector embedding representations, a symbolic AI approach enhances it with the strength of Knowledge Graphs by introducing a basis in the symbolic representation.



Figure 1: The interplay between neural (LLMs) and symbolic (KGs) AI methodologies. Source: <u>https://arxiv.org/pdf/2306.08302.pdf</u> [4]

dRAG is, therefore, a framework that allows AI solutions to tap into the strengths of both paradigms: the powerful learning and generalization capabilities of neural networks, and the precise, rule-based processing of symbolic AI. It

operates on two core components - (1) the DKG paranets and (2) AI models. The dRAG applications framework is entirely compatible with the existing techniques, tools, and RAG frameworks and supports all major data formats.



Figure 2: dRAG framework Application architecture

4.1 The Decentralized Knowledge Graph (DKG)

The OriginTrail Decentralized Knowledge Graph (DKG) is an open, permissionless decentralized network of knowledge graph nodes integrated with multiple blockchains. It is a core component of the verifiable internet infrastructure enabling AI solutions to leverage its discoverability, verifiability, and ownership capability. Its core resources are Knowledge Assets, organized in DKG Paranets together with knowledge services.



Figure 3: Three- layer conceptual architecture of the Decentralized Knowledge Graph

4.1.1. Enhancing AI reliability with Knowledge Assets

A Knowledge Asset is the primary resource in the DKG that AI systems can consume. They represent an ownable container of knowledge in multiple formats, such as knowledge graph triples, vector embeddings, images, etc. Knowledge Assets can be created, updated, and transferred by their owners with a single transaction using DKG protocol-defined operations. There are three core properties to each Knowledge Asset - ownership, discoverability, and verifiability.

Ownership. Each Knowledge Asset has an owner, implemented with a non-fungible token (NFT) issued on the blockchain. An owner can manage their Knowledge Assets, such as updating their content, changing content from privately or publicly shared, or transferring ownership. **This allows knowledge to become an asset class**. As ownership is implemented on the blockchain level, it enables seamless monetization, including the buying or selling of Knowledge Assets by AI models and agents.

Discoverability. Knowledge Assets are inherently discoverable due to their content structure and availability in the DKG. This makes DKG a **global, decentralized knowledge base with censorship-free discoverability**. Knowledge can be discovered through multiple methods, such as dereferencing, Knowledge Graph queries, vector similarity search, etc. According to the principles of linked data, Knowledge Assets can form meaningful connections - just as websites are linked via URLs (Uniform Resource Locators), a Knowledge Asset can reference the content of other Knowledge Assets using UALs (Uniform Asset Locators) and URLs, making them interoperable with the existing World Wide Web and Internet infrastructure. Through the DKG, these connections are **enriched with symbolic AI context which makes it perfect for use in combination with neural AI techniques**.

Verifiability. Knowledge Assets contain Merkle-tree-based cryptographic proofs of knowledge state (digests) stored on the blockchain. As Knowledge Assets evolve through updates, each proof is recorded, making all Knowledge Asset operations transparent and auditable on the blockchain. Verifiability is supported on a granular level of content (such as knowledge inclusion proofs), as well as on the level of the entire Knowledge Asset. Knowledge Assets follow the W3C Verifiable Credentials data model [5], where the DKG implements the verifiable data registry function, and verifiable presentations can be created from knowledge inside the Knowledge Assets. **This enables AI** systems to filter out any content where verifiability cannot be established prior to consumption (in dRAG, training, etc).



Figure 4: Knowledge Asset core elements

Knowledge Assets are the trusted, referenceable resource that AI systems are able to utilize in the dRAG framework. The level of granularity of what structured knowledge is included in the Knowledge Asset is, therefore, dependent on what a use case is required to reference. It could span from a single statement or an immutability proof to an entire database. Groups of Knowledge Assets can be used to form autonomously operated para-networks or paranets.

4.1.2. Autonomous paranets

Para-networks or paranets are **autonomously operated units**, **owned by its community** in the DKG. In paranets, we find assemblies of Knowledge Assets driving use cases with associated paranet-specific AI services and an incentivization model to reward knowledge miners fueling its growth.

Each paranet contains a set of:

- **Knowledge assets**, which include expected attributes that knowledge miners have to conform to in order for the Knowledge Assets to be included in the paranet (e.g. containing knowledge about a particular topic, data structured according to defined ontology, etc).
- AI services such as dRAG interfaces, AI agents, smart contracts, data oracles, etc.
- **Incentivization model** specifying the rules under which growth activities in the paranet are rewarded, such as knowledge mining and paranet-specific AI services.
- A supported blockchain on which the paranet is running and assembling Knowledge Assets,

The characteristics of a paranet, including its knowledge asset parameters and how services are provisioned, are all defined by the paranet operator, which can be an individual, an organization, or a Decentralized Autonomous Organization (DAO). Paranets together form the DKG, leveraging the common underlying network infrastructure. Given the DKG is a permissionless system, anyone can initiate a paranet.



Figure 5: Paranets as a collection of Knowledge Assets

Paranets provide a powerful substrate for AI systems, leveraging **network effects** of verifiable inputs from multiple sources to receive accurate answers through dRAG, allowing it to gather information from the (1) graph of public knowledge and (2) privately held knowledge in relevant Knowledge Assets that it has access to.

Paranets can form around topics of different nature and size, for example:

- Industry 4.0
- Decentralized Science
- Sustainability
- Public company reports
- LLM training data
- Metaverse
- Video content networks
- Social media
- Art discovery networks
- Prediction markets
- Sports and betting
- Entertainment

While some of the examples are inspired by the ongoing adoption seen on the DKG, the list is by no means exhaustive and serves as a showcase of the variety and inclusiveness of different autonomous paranets that DKG can support and make available to AI systems. All paranets are together part of the DKG and can be queried and accessed in combination.

5 Incentivising Verifiable Internet for AI on NeuroWeb

The dRAG framework offers a strong approach to addressing many of the key challenges of AI, however, the value of using dRAG grows exponentially with the amount of knowledge AI tools can access in the DKG. **To bend the adoption curve steeper, incentivization mechanisms are a key element**. The incentivization of Verifiable Internet for AI is implemented by allowing paranets operators to **leverage NeuroWeb incentives to fuel the paranet**

growth. The size of emissions and proposed rules to drive knowledge mining on a particular paranet are decided on with decentralized governance. The core activity of creating knowledge on paranets to earn incentives is Knowledge Mining.

5.1 Knowledge Mining

When launching an autonomous paranet, the paranet operators define the key characteristics. One of the most important ones is the incentivization mechanism for knowledge mining on their paranet. Such mechanisms may include conditions like topic relevance, ontology compliance, size limitations, and many others. In addition to knowledge miners, incentivization mechanisms may also reward service providers. Services can range from simple ontology compliance checks to AI-powered chat applications.

A paranet can be proposed and operated by anyone and they can be in any size that the operators feel is relevant - they can be functional both as narrow and wide. Paranet operators compete to obtain the rewards for their paranets with other paranets through NeuroWeb governance voting. Each operator creates a reward proposal specifying the reward emissions they require, the knowledge mining mechanism, incentivized services, and other key elements that operators or the voting NeuroWeb community may find to be important.



Figure 6: Incentives flow in autonomous paranet

In this way, there is an ongoing motivation to publish knowledge on the DKG which is tailored to the relevant domains, increasing the value of DKG for those building dRAG solutions.

6 AI - powered Autonomous DKG

The combination of the DKG, dRAG, and Verifiable Internet for AI incentives creates a structure that supports the constant growth of global knowledge. Knowledge miners will start to mine knowledge in a particular paranet by gathering data from relevant sources using manual and automated techniques. However, once more knowledge becomes available in a paranet, the possibility of autonomous knowledge mining is unlocked. As high-quality, annotated, ontology-fit data is present in a paranet, we can run **deductive reasoning** which is a logical process where new knowledge is deterministically made from existing knowledge based on ontology rules and connections/relationships between knowledge. The autonomous knowledge graph equally offers opportunities for **inductive reasoning** where new knowledge can be created in a probabilistic manner, by using AI to identify patterns and regularities in existing knowledge and create predictions on the best new knowledge.

An important body of work advancing this field is being created with foundation models for knowledge graph reasoning [6] which merge the foundation models used in AI systems onto a knowledge graph structure. This introduces the possibility of universal and transferable graph representations which would allow an AI system [7] to infer and populate any knowledge graph using its foundation model and vice-versa, get fine-tuned by consuming graph representations from existing knowledge graphs. Particularly powerful at such predictive tasks are the previously mentioned Graph Neural Networks (GNNs), since this is one of their key use cases.

The autonomous Decentralized Knowledge Graph will effectively use both types of reasoning (logical and probabilistic) to enrich and expand the DKG, driving a better understanding of relationships and entities, and making more informed predictions and recommendations to autonomously create new knowledge.

7 Economics of Verifiable Internet for AI

The Verifiable Internet fosters a cohesive approach towards the vision of more inclusive and democratic AI with two basic concepts driving its unit economics:

- 1. Al-ready Knowledge Assets
- 2. Knowledge paranets

Each of these units is supported with a tokenomics design ensuring the OriginTrail DKG network is enhanced in security, resilience, and usability. The incentive structure enacted on the NeuroWeb blockchain ensures that activities contributing to the DKG network growth receive attractive rewards and create further opportunities for building knowledge as a foundation for Verifiable Internet for AI.

7.1. TRAC utility token driving Verifiable Internet for AI infrastructure

Knowledge Assets are the most basic unit economics concept discussed as part of the Verifiable Internet for AI. Each **Knowledge Asset generated consumes a certain unit of service** within the OriginTrail DKG which is provisioned in a permissionless way by node runners and delegated stakers. Every time knowledge publishers add a new Knowledge Asset to the DKG, **they use TRAC to compensate for the use of the DKG network services** and use a gas token for transactions. The gas token depends on the blockchain the DKG is being used for publishing (e.g. NEURO on NeuroWeb, DAI on Gnosis, ETH on Ethereum, etc).

The nodes in the DKG network compete to collect publishing fees provided by knowledge publishers. They compete on three core elements of (1) providing adequate service (storage and availability of published knowledge) for paranets, (2) the amount of TRAC stake on their node, and (3) node paranet address required for efficient distribution of knowledge and rewards across the network. The nodes with the highest scores according to the criteria are eligible to collect their portion of rewards throughout the duration of the period a Knowledge Asset has been published for.

Since TRAC staked is a critical component to the node's success, node operators can also **allow other TRAC holders to delegate their TRAC** to their node stake. In exchange for their delegated stake, node operators share a portion of the node rewards with delegators, according to the relative size of their delegations. TRAC stake also has a vital security role as it ensures that the DKG nodes are performing their services adequately. If they act maliciously, their stake can get slashed.

7.2. NEURO native token of NeuroWeb - incentivizing Verifiable Internet for AI

Knowledge paranets, as explained in previous chapters, are a collection of Knowledge Assets on a particular topic and associated services. Paranets are proposed by **paranet operators**, an entity that also defines the core rules of the paranet. In addition to the technical rules (e.g. ontologies, data processing rules, service requirements, etc), paranet operators may also define the paranet incentive system. The operators may choose to incentivize **knowledge mining** (publishing of new Knowledge Assets to their paranet), **knowledge validation services** (ensuring a required level of data quality), **knowledge access services** (e.g. running an AI discoverability service, public interfaces, or AI agent service) or other paranet specific services.

Once defined, the paranet proposal for incentives can be submitted to a governance vote towards the **DKG** incentives pool on the NeuroWeb blockchain. The NEURO token community then acts as a (de)central(ized) bank and decides which areas of the knowledge economy will create the most positive effects for the ecosystem and vote on them receiving the proposed amount of NEURO emission for running their incentive structure. Important to highlight is that NEURO incentives can also be issued for paranets running on other blockchains by providing verifiable proof of activity. The NEURO rewards, however, have to always be collected on NeuroWeb as that is where they are issued.

In addition to being used for voting on paranet incentives, NEURO has utility as the **native asset of the NeuroWeb blockchain on Polkadot**. By being a native asset, NEURO is used as gas for transactions, for running and staking on parachain collators, and for governance voting on NeuroWeb.

7.3 Token economics overview

Growing and using the Verifiable Internet for AI via the OriginTrail DKG means always interacting with at least two decentralized networks, the DKG, and a corresponding blockchain. For this reason, using the infrastructure always requires at least one more utility token besides TRAC. Below is an overview of key actions you can perform on the two layers and the corresponding asset you would have to use to execute them.

	TRAC	NEURO	Other assets
Knowledge Assets			
Publishing Knowledge Assets	Compensating for use of DKG.	Gas token for txs if using NeuroWeb.	Gas tokens for txs (e.g. DAI, ETH) if using other chains
Running the DKG infrastructure	(Delegated) staking on DKG nodes.	Gas token for txs if using NeuroWeb.	Gas tokens for txs (e.g. DAI, ETH) if using other chains
Knowledge Paranets			
Incentivizing Knowledge mining	Used when publishing Knowledge Assets	Providing incentives for publishers.	Not applicable
Incentivizing Knowledge services (e.g. knowledge validation, discoverability)	Not applicable	Providing incentives for service providers.	Not applicable
NeuroWebAI			
Blockchain use	Not applicable	Gas token for txs on NeuroWeb.	Not applicable
Collators system	Not applicable	Running and staking on collators.	Not applicable
Governance	Not applicable	Voting on updates and treasury and paranets growth incentives for knowledge mining.	Not applicable

Table 1: Overview of key actions and corresponding assets for two layers

8 Conclusion

The exploration of the Verifiable Internet for Artificial Intelligence requires us to look into the details of three converging technologies of AI, Internet, and blockchain. The fusion of these cutting-edge domains promises to reshape our digital landscape, creating unprecedented opportunities while addressing critical challenges.

This paper introduces an AI ecosystem that is not only powerful and dynamic but also trustworthy and transparent. The OriginTrail Decentralized Knowledge Graph (DKG), with Knowledge Assets as its primary resource, is a pivotal innovation in this context, offering a robust framework for ensuring the ownership, discoverability, and verifiability of information utilized by AI systems through the dRAG framework. Autonomous paranets, characterized by their self-operating and community-owned nature, offer a modular and scalable approach to organizing Knowledge Assets and growing the DKG.

The incentives for continued growth of Verifiable Internet for AI span across multiple networks. Different assets are used according to their respective designs - TRAC for operating, securing, and using the DKG infrastructure and NEURO for incentivizing growth in addition to the utility as the native NeuroWeb blockchain token.

The convergence of AI, the Internet, and crypto as embodied in the Verifiable Internet for AI represents a transformative step towards a more transparent, equitable, and intelligent digital world.

9 References

[1] Chui M, Roberts R, Yee L., Hazan Z., Singla A., Smaje K., Sukharevsky A, Zemmel R, (2023) The economic potential of generative AI: The next productivity frontier. URL: <u>https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-AI-the-n ext-productivity-frontier</u>

[2] Shumailov, I., Shumaylov, Z., Zhao, Y., Gal, Y., Papernot, N., Anderson, R. (2023) The curse of Recursion: Training on generated data makes models forget. New York, May 31., URL: <u>https://arxiv.org/abs/2305.17493</u>

[3] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S. and Kiela, D. (2021) Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. New York, April 12., URL: <u>https://arxiv.org/abs/2305.17493v2</u>

[4] Pan, S., Luo, L., Wang, Y., Chen, C., Wang, J., Wu, X., Fellow. (2024) Unifying Large Language Models and Knowledge Graphs: A Roadmap. Jan 25., URL: <u>https://arxiv.org/pdf/2306.08302.pdf</u>

[5] Sporny M., Longley D., Chadwick D., Steele O., Verifiable Credentials Data Model v2.0 URL: https://www.w3.org/TR/vc-data-model-2.0/

[6] Galkin M., Yuan X., Mostsafa H., Tang J., Zhu Z., (2023) Towards Foundation Models for Knowledge Graph Reasoning., URL: <u>https://arxiv.org/pdf/2310.04562.pdf</u>

[7] LeCun, Y. (2022) A Path Towards Autonomous Machine Intelligence. New York, June 27., URL: https://openreview.net/pdf?id=BZ5a1r-kVsf