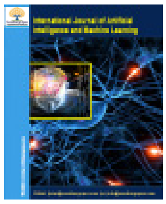
https://doi.org/10.51483/IJAIML.4.2.2024.52-66 ISSN: 2789-2557

SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/



Research Paper Open Access

Decentralized Governance to Optimize Human Output Datasets for AI Learning

Wulf A. Kaal[1*]

[1]Professor, School of Law, University of St. Thomas, 1000 LaSalle Avenue, MSL 400, Minneapolis, MN 55403, United States. E-mail: wulfkaal@stthomas.edu

Article Info

Volume 4, Issue 2, July 2024
Received : 16 March 2024
Accepted : 20 June 2024
Published : 05 July 2024
doi: 10.51483/IJAIML.4.2.2024.52-66

Abstract

The evolution of AI depends on upgradable quality datasets. Data is the foundation on which AI algorithms learn and make predictions. High-quality, diverse, and labeled datasets are crucial for training AI models effectively. The availability of quality data plays a significant role in determining the success and impact of AI in disrupted industries. The AI Learning Ecosystem (ALE) facilitates a micro task ecosystem for AI learning. ALE uses its proven and tested decentralized governance ecosystem to provide high-quality diverse datasets for AI learning via gamified micro-task work. Through its testing environment in the industry-leading Code Review DAO (CRDAO), ALE distinguishes itself from competitors through unparalleled decentralized governance optimization that minimizes micro-task work duplication in centralized systems and allows gamified micro-task work to scale high-quality diverse datasets for AI learning.

Keywords: *Artificial intelligence, Large language models, Dataset, Micro task work, Gamification, Quality controls, Decentralized autonomous organization, Token models, Crypto currencies, Feedback effects, Emerging technology, Tokens, Blockchain, Distributed ledger technology, Code assurances*

1. Introduction

1.1. AI for Human Evolution

Artificial Intelligence (AI) epitomizes human evolution. Based on findings associated with human brain functionality, deep learning in AI systems is a method of producing AI components that can process data similar to how the human brain would process data. This supports the AI's recognition of complex patterns in datasets, including in pictures, text elements, and sounds. The analysis of such datasets in deep learning allows the AI to generate insights, analysis, and predictions for a variety of human generated outcomes. Through deep learning AI can automate sequences of tasks that otherwise would require human intelligence. Examples may include the analysis and causal recognition of images or transcribing a sound file into text, among others.

* *Corresponding author: Wulf A. Kaal*, Professor, School of Law, University of St. Thomas, 1000 LaSalle Avenue, MSL 400, Minneapolis, MN 55403, United States. E-mail: wulfkaal@stthomas.edu

In 2024, AI deep learning is possible through so called Large language models (LLM). LLMs are trained on large datasets. For this purpose, neural networks, through their encoders and decoders extract human level meanings from a sequence of data including text and create associated relationships between words and phrases in the text that is contained in the large datasets.

So-called transformer neural network architecture enables the utilization of large datasets with up to hundreds of billions of parameters. While transformer neural network architecture can utilize huge datasets that are often generated from internet data, with often over 50 billion web pages, and Wikipedia, with its ever growing amount of pages, data quality continues to be a huge issue for LLMs.

1.2. AI Business Model Disruption and Scaling

Deep learning AI has unparalleled use cases that allow an industry-wide scaling of new business models built on AI, including in electric vehicles, legal, administrative, manufacturing, aerospace, electronics, medical research, and many other industries and fields.

AI has the potential to disrupt existing business models and enable the creation of new and innovative businesses in several ways:

1. **Automation and efficiency:** AI can automate repetitive and mundane tasks, leading to increased efficiency and cost savings. This disruption can eliminate the need for certain job roles or transform them, allowing businesses to allocate resources more effectively and focus on higher-value activities.
2. **Enhanced decision-making:** AI algorithms can analyze vast amounts of data quickly and accurately, providing valuable insights for decision-making. This disruption empowers businesses to make data-driven decisions, optimize processes, and improve outcomes.
3. **Personalization and customer experience:** AI enables businesses to personalize their products, services, and customer experiences based on individual preferences and behavior. By leveraging AI algorithms, businesses can create tailored recommendations, targeted marketing campaigns, and personalized interactions, enhancing customer satisfaction and loyalty.
4. **Predictive analytics:** AI algorithms can predict future trends, customer behavior, and market dynamics by analyzing historical data. This disruption allows businesses to anticipate demand, optimize inventory, mitigate risks, and make proactive decisions, giving them a competitive advantage.
5. **New business models and services:** AI opens doors to entirely new business models and services. For example, AI-powered platforms can connect consumers and providers directly, disrupting traditional intermediaries. Additionally, AI can enable the creation of innovative products and services that were not possible before, such as autonomous vehicles, virtual assistants, or personalized healthcare solutions.
6. **Scalability and scalability:** AI systems can scale efficiently and handle large volumes of data and interactions, making it easier for businesses to expand their operations rapidly. This disruption allows businesses to reach broader markets and serve a larger customer base without significant infrastructure investments.

1.3. Pain Point Data Quality

The growth and success of AI are highly dependent on quality data sets. Data is the fuel that powers AI algorithms, and without high-quality, relevant, and diverse data, the performance and accuracy of AI systems can be compromised.

While it is important to note that data quality alone is not sufficient for AI growth and its disruption and reinvention of business models, the algorithms, computing power, and expertise of developers also play significant roles. Nonetheless, access to quality data sets is a fundamental requirement for the development and advancement of AI technologies.

The key reasons why high quality data sets are crucial for the AI evolution include but are not limited to the following:

1. **Training AI models:** AI systems learn from data through a process called training. During training, AI algorithms analyze and extract patterns from large datasets to make predictions or perform specific

tasks. The quality of the training data directly impacts the AI model's ability to learn and make accurate predictions.

2. **Bias reduction:** High-quality data sets help in reducing bias within AI systems. Biased data can lead to biased outcomes, perpetuating social or cultural prejudices. By ensuring diverse and representative data sets, AI developers can work towards minimizing bias and creating fairer AI systems.
3. **Generalization and adaptability:** Quality data sets help AI models generalize their learning to new situations and adapt to changing environments. A diverse and comprehensive dataset allows AI systems to encounter a wide range of scenarios, leading to better performance in real-world applications.
4. **Robustness and reliability:** AI models trained on high-quality data sets tend to be more robust and reliable. They can handle edge cases, outliers, and unexpected inputs with greater accuracy, improving the overall performance and user experience.

In 2024, AI development companies are beginning to use smaller datasets[1] for the development of LLMs. This is done in an effort to increase model accuracy and prevent overfitting in machine learning models, which, in turn, allows the new machine learning models to produce better results on unseen data. However, with small datasets in LLMs, the risk of overfitting also rises, especially with complex models. Therefore, LLM developers have to turn to regularization in an effort to address overfitting of the model with the training data.

Because of the trend towards smaller datasets in LLMs, data quality is a crucial pain point in smaller datasets. In addition to balancing and normalizing the data, LLM developers focus on adequate model validation through techniques including dropouts in neural networks, pruning decision trees, and cross-validation. All of these efforts are devoted to generating better training datasets as the quality of LLM training data can break LLM results or improve them significantly.

1.4. Total Addressable Market

Industries across the board can be significantly affected by the availability of high-quality AI training data sets. Some industries may experience a more pronounced impact due to the nature of their operations and the potential for AI-driven transformations.

1.4.1. Disrupted Industries

Estimating the total addressable market (TAM) for industries impacted by technological advancements, particularly artificial intelligence (AI), is a multifaceted task. It requires analyzing various sectors, each with its own market size and growth trajectory. The dynamic nature of AI's influence complicates the effort to pinpoint an exact market value. Nonetheless, the aggregate potential market size is undeniably vast.

To contextualize the TAM for industries undergoing disruption, consider the following sectors:

1.4.1.1. Healthcare

AI applications in healthcare, such as disease diagnosis, drug discovery, personalized medicine, and patient monitoring, depend on high-quality data sets for accuracy. These applications are poised to enhance medical decision-making and patient care significantly. According to the World Health Organization, the global healthcare market was estimated at $8.45 trillion in 2020 and is projected to grow to $11.91 trillion by 2025, highlighting AI's critical role in this sector's evolution.[2]

1.4.1.2. Finance

The finance sector relies on data for functions including risk assessment, fraud detection, and algorithmic trading. AI models that process this data can uncover patterns and predict market movements, thereby facilitating smarter investment strategies and customer service improvements. The global financial services market's value stood at about $19.7 trillion in 2020, with AI expected to transform banking, insurance, and asset management.[3]

---

1 https://towardsdatascience.com/is-small-data-the-next-big-thing-in-data-science-9acc7f24907f

2 "Global Spending on Health: A World in Transition," World Health Organization, 2021.

3 "Global Financial Markets: An Overview," International Monetary Fund, 2020.

1.4.1.3. Retail and E-commerce

In retail and e-commerce, AI-driven tools for personalization, demand forecasting, and inventory management rely on detailed data analysis. These tools enable businesses to tailor their offerings and optimize operations. The global retail market's valuation exceeded $22 trillion in 2020, and AI's contributions are set to redefine the industry by boosting growth and enhancing the consumer experience.[4]

1.4.1.4. Manufacturing and Logistics

AI enhances manufacturing and logistics through improved efficiency, quality control, and supply chain management. The global manufacturing market was valued at around $39.6 trillion in 2020. AI applications in this sector promise significant cost reductions and performance enhancements.[5]

1.4.1.5. Transportation

AI's potential in transportation, including autonomous vehicles and traffic management, is considerable. The global transportation services market was valued at approximately $4.3 trillion in 2020. As AI technologies such as autonomous driving and route optimization advance, the TAM for AI in transportation is expected to witness substantial growth.[6]

1.4.1.6. Customer Service

The transformation of customer service by AI, through chatbots and virtual assistants, underscores the technology's value in enhancing customer interactions. The customer service sector, valued at about $55 billion in 2020, will see its TAM for AI expand as businesses increasingly adopt these solutions to improve efficiency and customer satisfaction.[7]

These sectors exemplify the significant, albeit challenging to quantify, TAM for AI across industries. The ongoing integration of AI into these areas promises not only to drive market growth but also to catalyze profound changes in how services are delivered and consumed worldwide.

The impact of high-quality AI training data sets is not limited to these above mentioned industries. AI has the potential to disrupt and transform various sectors, including education, energy, agriculture, entertainment, and more. The availability of quality data sets is a catalyst for AI-driven advancements and innovation across industries.

Please note that these TAM figures are approximate and subject to various factors such as market dynamics, technological advancements, and global economic conditions. The growth potential of AI in these industries is significant and will continue to evolve as the technology matures and adoption increases.

However, acquiring and curating high-quality, diverse, and labeled datasets can be a challenge. Industries that can access large volumes of structured and labeled data, such as finance and healthcare, may have a head start in leveraging AI. On the other hand, industries like transportation and customer service may face hurdles in obtaining quality data due to privacy concerns, data fragmentation, and the need for extensive labeling efforts. Overcoming these challenges and ensuring access to quality data is essential for maximizing the potential TAM of AI in these sectors.

Despite these challenges, the TAM for AI in transportation and customer service is substantial. As AI technologies mature, data availability improves, and companies invest in AI-driven solutions, the TAM for these industries is expected to increase significantly. The combined TAM of all disrupted industries, including healthcare, finance, retail, manufacturing, logistics, transportation, and customer service, is likely to be in the hundreds of trillions of dollars, making it a highly lucrative market for AI-driven innovation. Adding the TAM of other affected sectors like education, energy, agriculture, and entertainment, the total addressable market for all disrupted industries combined is in the higher hundreds of trillions of dollars.

---

4 "The State of the Global Retail Market," World Retail Congress, 2020.

5 "Global Manufacturing Outlook," United Nations Industrial Development Organization, 2021.

6 "Global Transportation Market Analysis," International Transport Forum, 2020.

7 "Global Customer Service Market Report," Customer Service Institute, 2020.

1.4.2. Data Providers

Determining the exact TAM for data providers who produce high-quality diverse datasets for AI learning is challenging as it depends on various factors such as market demand, pricing, and competition. However, it is safe to say that the TAM for such data providers is significant and growing rapidly.

As AI continues to advance and become more prevalent across industries, the need for quality training data becomes paramount. Data providers that can deliver labeled, diverse, and reliable datasets to train AI models effectively will have a significant market opportunity.

The TAM for data providers in AI learning extends across multiple sectors, including healthcare, finance, retail, manufacturing, logistics, transportation, customer service, education, energy, agriculture, and entertainment, among others. Each of these industries requires specific datasets tailored to their unique needs and use cases.

Furthermore, with the increasing adoption of AI by both large enterprises and smaller businesses, the demand for high-quality data is expected to surge. Startups, established companies, and even AI service providers will rely on data providers to access the necessary datasets for their AI initiatives.

Considering the potential TAM of the disrupted industries mentioned earlier, which is estimated to be in the hundreds of trillions of dollars, the TAM for data providers who produce high-quality diverse datasets for AI learning is likely to be substantial. As AI becomes more integral to business operations and decision-making processes, the importance of quality training data will only grow, further expanding the TAM for data providers in this space.

2. Gamification to Enhance AI Data Quality

The AI Learning Ecosystem (ALE) uses its proven and tested decentralized governance ecosystem to provide high-quality diverse datasets for AI learning via micro-task work.

Through its testing environment in the existing Code Review DAO, ALE distinguishes itself from competitors through industry-leading decentralized governance optimization that minimizes micro-task work duplication in centralized systems and allows micro task work to scale for AI dataset generation.

2.1. Freelance Micro Task Market Growth

In the current freelancer market, leading centralized platforms Fiverr, Taskrabbit, Upwork, Crowdflower and Mechanical Turk have their workforces distributed throughout the world. Micro Tasks freelance work is becoming increasingly important in the freelance market.

Micro tasks are defined as small tasks that require human judgment, can be completed by humans independently over the internet, and are part of a larger unified project. Because of the necessity of human judgment that cannot currently be replaced by machines/computers, micro tasks enable organizations to build products or create outcomes that cannot be synthesized by machines/computers alone. For example, the Chinese government uses 2 million micro task workers to aid in censoring the internet. Internet companies, such as Google, Facebook, Twitter, Ebay, and LinkedIn optimize their production ready solutions and enhance their AI model training with micro task workers. Large scale distributions of machine learning researchers, among others, gather structured labeled data for artificial intelligence (AI) training purposes through the use of micro task workers.

Billions of micro tasks are completed each year and the demand for micro task workers is increasing consistently. In an attempt to capitalize on the growing demand for micro task work, Amazon created Amazon Mechanical Turk (MTurk) in 2005. MTurk is an online marketplace that allows requesters to pay workers for performing micro tasks online, thus crowdsourcing data collection. The World Bank concluded in a 2015 report, that the largest crowdsourced data collection platforms Amazon Mechanical Turk and CrowdFlower, a venture backed company that raised over $58 million and is focused on enriching data used for AI, would quadruple their revenue from 2013 to 2016.

The increasing demand for micro task workers that came initially from data scientists and other academics is further increased by the growth of artificial intelligence (AI) and the increasing scope and scale of AI applications in a broad range of industries. Fortune 500 tech companies including Amazon, Apple, Google Alphabet, Twitter, and Facebook use AI and machine learning to improve their services and cut costs, thus increasing profitability exponentially. According to some estimates, Fortune 500 tech companies spent between $20 and $30 billion on the development and enhancement of their AI systems in 2016. More recent data for 2017 suggests that this trend is continuing and increasing. AI is already playing a significant role in consumer expectations and tech companies have the best possible set of incentives to keep investing in AI to fulfill such expectations and develop new products at marginal cost.

2.2. Micro Task Work for AI Learning

Micro task work plays a crucial role in producing high-quality and diverse datasets for AI learning. It involves breaking down complex tasks into smaller, more manageable tasks that can be completed by a large number of human workers. These workers, often referred to as crowdworkers or micro taskers, perform these tasks in exchange for monetary compensation.

The following examples show how micro task work contributes to the creation of high-quality datasets for AI learning:

1. **Data Annotation:** Micro task work is commonly used for data annotation tasks, where crowdworkers label and annotate data according to specific criteria. For example, in image recognition, crowdworkers may annotate objects or draw bounding boxes around them to train AI models. This annotation process helps create labeled datasets that serve as ground truth for training AI algorithms.
2. **Data Validation:** Crowdworkers also play a critical role in validating the quality and accuracy of labeled data. They can review and verify annotations made by other workers, ensuring consistency and reducing errors. This validation step helps maintain the integrity and reliability of the dataset.
3. **Data Augmentation:** Micro task work can be utilized to generate diverse data by creating variations or augmentations of existing datasets. Crowdworkers can perform tasks like image manipulation, text paraphrasing, or audio synthesis to increase the diversity of the training data. This helps AI models generalize better and perform well on a wider range of real-world scenarios.
4. **Data Cleaning:** Crowdworkers can assist in cleaning and refining datasets by identifying and rectifying errors, removing duplicates, or standardizing data formats. This ensures the dataset is of high quality, reducing noise and improving the performance of AI models during training and inference.

Micro task work platforms, such as Amazon Mechanical Turk, Figure Eight (now Appen), and Scale AI, provide the infrastructure to distribute these tasks to a large pool of workers, manage their contributions, and ensure quality control.

Overall, micro task work enables the efficient and scalable production of high-quality, diverse datasets for AI learning. It leverages human intelligence to handle tasks that are challenging for current AI systems and contributes to the continuous improvement and advancement of AI technologies.

2.3. Shortcomings in Legacy Micro Task Market

The evolution, improvements, and growth of AI is correlated with the evolution, improvements, and growth in micro task work. AI uses supervised and unsupervised as well as reinforcement machine learning. Because unsupervised and reinforcement learning are much more complex than supervised learning, supervised learning is to date more common and more relied upon for AI development. While this may change as unsupervised and reinforcement learning evolve, currently, supervised learning depends on labeled data that is produced via micro task work. The mapping function of the supervised AI learning process necessitates the analysis of labeled input variables x and corresponding output variables y. In the supervised learning training phase, the AI neural network examines the training dataset of labeled x input data to learn to classify the input data idealistically. The higher the quality and quantity of such labeled datasets the better the AI neural network's learning algorithm during the supervised training process. Accordingly, the evolution, improvements, and growth of AI is correlated with the evolution, improvements, and growth in micro tasks work. But alas, microtasks platform systems are subject to significant limitations that inhibit the evolution of AI.

2.3.1. Lack of Scaling Solutions in Centralized Micro Task Market

2.3.1.1. Overpricing

The existing centralized marketplaces for micro task work cannot adequately fulfill the increasing demand for high quality micro task work for AI labeled training datasets. First and foremost, the cost structure for micro task work in centralized systems that necessitate intermediation results in significant overpricing without benefiting the micro task workers directly. The cost structure sub optimality can be traced back to several factors. All too human shortcomings of micro task works, such as limited attention span, irrationality, and inaccuracies result in verification requirements for micro task work. However, manual verification of micro task work is subject to the same human limitations.

2.3.1.2. Duplication of Work

In an attempt to ensure quality of results and minimize the impact of the human limitations of their workers, requesters in centralized micro task structures set up teams of up to 15 workers to perform the same task in an effort to form a consensus. The multiplication of work inherent in this process significantly increases the cost of micro task work. Requiring requesters to pay proportionally (e.g., up to 15 times) for work per project results in waste. The necessity of multiplication of work also subjects micro task workers to lower rates and lack of payment increases. Moreover, because unmanaged centralized micro task platforms do not supply consumer interfaces needed to accomplish specific tasks, requesters of micro task work are forced to either build their own tools or pay large fees to startups that hope to capture the enterprise market. Both options are necessary in current centralized systems but also result in underutilization of resources.

2.3.1.3. No Access for Workers

Most importantly, the circa 38% of the labor pool that is unbanked but skilled does not currently have access to the centralized micro task marketplaces. Without a bank account, workers cannot contribute and profit from the existing centralized micro task marketplace. Furthermore, even for those in the labor pool who do have bank accounts, micro-task work is often associated with practical problems such as high fees of intermediary financial institutions, lost or otherwise affected payments, lost checks, among other issues. Finally, micro workers in centralized systems are faced with invasive, privacy challenging, time consuming, and unclear signup and approval processes that create market entry barriers for micro task workers.

2.4. Gamification to Enhance Micro Task Work

The gamification of micro task work can have several effects on the process of producing high-quality, diverse datasets for AI learning. Gamification refers to the application of game design principles and mechanics to non-game contexts, such as micro task work.

Gamification can impact the process of producing high-quality, diverse datasets for AI learning in the following ways:

1. **Increased Engagement:** Gamification techniques, like adding points, levels, leaderboards, or rewards, all of which may be utilized exclusively within the platform, can enhance the engagement of crowdworkers. By introducing a competitive or rewarding element, gamification motivates workers to actively participate and complete tasks more efficiently. This increased engagement can lead to higher-quality outputs as workers are more invested in the process.

2. **Quality Control:** Gamification can be utilized to improve data quality through mechanisms like consensus-based voting or peer review. Workers can review and rate each other's contributions, earning points or recognition for accurate and consistent work. This approach helps identify and resolve discrepancies, ensuring better quality control in the dataset creation process.

3. **Skill Development:** Gamification can enable skill development among crowdworkers. By providing challenges or levels that progressively increase in difficulty, workers can enhance their annotation, validation, or cleaning skills over time. This skill development can result in improved accuracy and efficiency in dataset creation.

4. **Crowdworker Retention:** Gamification elements like badges, achievements, or virtual currencies, all of which may be utilized exclusively within the platform, can enhance crowdworker retention. By recognizing

and rewarding workers for their contributions, gamification fosters a sense of accomplishment and encourages them to continue participating in micro task work. This retention is crucial for maintaining a consistent pool of experienced and reliable workers for dataset creation.

5. **Scalability and Speed:** Gamification can help expedite the dataset creation process by encouraging workers to complete tasks more quickly. By introducing time-based challenges or offering bonuses for timely completion, gamification can increase the speed at which high-quality datasets are generated, enabling scalability for large-scale AI projects.

It's important to note that while gamification can positively impact engagement and productivity, proper design and implementation in combination with gamification logic of reputation governance are crucial. Through the combination of decentralized governance and gamification of micro task work ALE platform intends to ensure that gamification techniques do not compromise the quality and accuracy of the dataset. Balancing the gamification elements with appropriate quality control measures via decentralized governance is essential to maintain the integrity of the dataset creation process.

3. Case Study - Code Review Platform

The Code Review Platform is operated by an affiliated entity and served as a testing environment in the past. It is accessible under the following URL: https://crdao.ossa.dev/

3.1. *Community Audit for Code Review*

Scaling the Code Review Platform operations as a proof point is a key objective for ALE. The Code Review Platform uses several key reputation related metrics, as illustrated below, to scale its operations over time.

To enable the public to participate in and contribute to the Code Review Platform, ALE is using a gamified reputation system. Any user can access and contribute to the Code Review Platform and earn reputation scores on the platform. While the code reviews themselves are relegated to the expert members of the Code Review Platform, the public can support the ultimate code reviews via gamified engagements.

Mechanical Turk (as defined above) functionality is a core application of the Code Review Platform Protocol and the Code Review Platform ecosystem. Other use cases and applications will evolve on the Code Review Platform over time. The Code Review Platform architecture will evolve in any setting that supports the ethical advancement of the crypto evolution.

3.1.1. Growth Potential of the Smart Contract Industry

The smart contract industry has enormous growth potential. Different measures help assess that growth including the CAGR (compounded annual growth rate) used as well as the current valuation of the smart contract industry. Estimates of the smart contract industry's future value by 2032 range from 1 to 2.5 billion dollars.[8]

---

[8] Varying group estimates are listed below from low to high.Group Estimates for the Smart Contract Industry's Value by 2032:

1. SNS Insider: $1 billion, at a CAGR of 24.2%

1.1. Smart Contracts Market Size, SNS Insider, https://www.snsinsider.com/reports/smart-contracts-market-1542 (last visited Jan. 4, 2024).

2. Verified Market Research: $1.2 billion, at a CAGR of 26.4%

2.1. Smart Contracts Market Size and Forecast, Verified Market Research, https://www.verifiedmarketresearch.com/product/smart-contracts-market/ (last visited Jan. 4, 2024).

3. Valuates Reports: $1.4 billion, at a CAGR of 24.2%

3.1. *Global Smart Contracts Market Research Report*, Valuates Reports, https://reports.valuates.com/market-reports/QYRE-Auto-31L1599/global-smart-contracts (last visited Jan. 4, 2024).

4. Acumen Research & Consulting: $1.417 billion, at a CAGR of 22.8%

4.1. Smart Contracts Market Size: Global Industry, Share, Analysis, Trends and Forecast 2023 – 2032, ACUMEN RESEARCH & CONSULTING, https://www.acumenresearchandconsulting.com/smart-contracts-market (last visited Jan. 4, 2024).

5. Future Market Insights: $1.5 billion, at a CAGR of 23.5%

5.1. Smart Contracts Market Outlook (2022 to 2032), Future Market Insights, https://www.futuremarketinsights.com/reports/smart-contracts-market (last visited Jan. 4, 2024).

6. Allied Market Research: $2.5 billion, at a CAGR of 29.6%

6.1. Smart Contracts Market Research, 2032, ALLIED MARKET RESEARCH, https://www.alliedmarketresearch.com/smart-contracts-market-A144098 (last visited Jan. 4, 2024).

3.1.2. Smart Contract Vulnerabilities Undermine Industry Growth

The growth of the smart contract industry across industry vectors is affected by the attack vectors pertaining to smart contracts. Smart contract bugs can result in financial loss, reputational loss, increased smart contract costs, legal issues, and, in extreme cases, destruction of the smart contract.[9]

While no clear data exists to estimate how much smart contract attack vectors will limit the growth of the smart contract industry, the smart contracts industry is still estimated to grow to several billion dollars in the next decade—*despite* the cost of security audits and financial losses from bad actors exploiting smart contracts bugs. Estimates on financial loss attributable to security breaches and attach vectors of smart contracts range from hundreds of millions to billions of dollars. For example, between 2016 and 2018, seven cyber security incidents occurred in Ethereum smart contracts resulting in financial losses of over $289 million.[10] In 2021 alone total financial loss from smart contract bugs was estimated at $680 million.[11] Some estimates put the current global financial loss due to smart contract vulnerabilities over 6 billion dollars.[12]

3.1.3. Solutions for Smart Contract Vulnerability

A diverse set of proposed solutions has emerged to address smart contract vulnerabilities issues and reduce the overall number of attacks. One potential solution is the use of a smart contract compiler.[13] For example, one such compiler is HCC, which automatically inserts security hardening checks at the source-code level.[14] HCC develops a code property graph (CPG) to model control-flows and data-flows of a given smart contract. Due to the CPG notation, HCC can be applied to various smart contract platforms and programming languages. HCC developers have demonstrated it efficiently mitigates reentrancy and integer bugs.[15] They also show how to integrate HCC within other blockchain platforms such as Hyper ledger Fabric. Their evaluation on 10k real-world contracts demonstrates that HCC is highly practical.[16]

Altering the methodology of bug classification and vulnerability analysis provides another promising approach. In one study, researchers propose two new vulnerability classes: distributed system protocol (DSP) and distributed system resource management (DRM).[17]

3.1.4. AI-Driven Solutions

The integration of blockchain technology and AI has immense promise.[18] AI technology appears best suited to

9 See David Balaban, *Navigating The Security Challenges of Smart Contracts*, Forbes (Feb. 11, 2023, 6:33 AM), https://www.forbes.com/sites/davidbalaban/2023/02/11/navigating-the-security-challenges-of-smart-contracts/?sh=14006afd4992. Majd Soud, Grischa Liebel & Mohammad Hamdaqa, *PrAIoritize: Learning to Prioritize Smart Contract Bugs and Vulnerabilities* (Working Paper), https://arxiv.org/pdf/2308.11082.pdf. Sherman Lee, *Blockchain Smart Contracts: More Trouble Than They Are Worth?* Forbes (Jul 10, 2018,11:38 PM), https://www.forbes.com/sites/shermanlee/2018/07/10/blockchain-smart-contracts-more-trouble-than-they-are-worth/?sh=74b5654523a6. Haozhe Zhou, *Amin Milani Fard & Adetokunbo Makanju, The State of Ethereum Smart Contracts Security*, 2 Journal of Cybersecurity & Privacy 358 (2022). Hanting Chu et al., A Survey on Smart Contract Vulnerabilities: Data Sources, Detection, and Repair, 196 INFORMATION & SOFTWARE TECHNOLOGY 1 (2023), https://www.sciencedirect.com/science/article/pii/S0950584923000757. MacKenzie Sigalos, Bug Puts $162 Million up for Grabs, Says Founder of DeFi Platform Compound, MSNBC (Oct. 3, 2021, 2:41 PM), https://www.cnbc.com/2021/10/03/162-million-up-for-grabs-after-bug-in-defi-protocol-compound-.html. Tamer Abdelaziz & Aquinas Hobor, Smart Learning to Find Dumb Contracts, https://www.usenix.org/system/files/usenixsecurity23-abdelaziz.pdf. Fabio Gritti *et al.*, *Confusum Contractum: Confused Deputy Vulnerabilities in Ethereum Smart Contracts*, https://www.usenix.org/system/files/usenixsecurity23-gritti.pdf. In one study, researchers found that 127 high-impact attacks were responsible for financial losses totaling $2.3 billion. Stefanos Chaliasos et al., *Smart Contract and DeFi Security: Insights from Tool Evaluations and Practitioner Surveys* (Working Paper), https://www.doc.ic.ac.uk/~livshits/papers/pdf/icse24.pdf.

10 Ayman Alkhalifah *et al.*, A Mechanism to Detect and Prevent Ethereum Blockchain Smart Contract Reentrancy Attacks, 3 *Frontiers in Computer Science*, 1 (2021), https://www.frontiersin.org/articles/10.3389/fcomp.2021.598780/full.

11 Thomas Claburn, *Smart Contract Developers Not Really Focused on Security: Who Knew?* Reporter (Apr. 26, 2022), https://www.theregister.com/2022/04/26/smart_contract_losses/.

12 Stefanos Chaliasos *et al.*, *Smart Contract and DeFi Security: Insights from Tool Evaluations and Practitioner Surveys* (Working Paper), https://www.doc.ic.ac.uk/~livshits/papers/pdf/icse24.pdf.

13 Jens-Rene Giesen *et al.*, Practical Mitigation of Smart Contract Bugs (Working Paper), https://arxiv.org/pdf/2203.00364.pdf.

14 *Id*.

15 *Id*.

16 *Id*.

17 Wesley Dingman *et al.*, Defects and Vulnerabilities in Smart Contracts, a Classification Using the NIST Bugs Framework, 73 *International Journal of Networked & Distributed Computing,* 121 (2019), https://www.atlantis-press.com/journals/ijndc/125913574/view?ref=metastate.

18 Rashi Saxena, E. Gayathri and Lalitha Surya Kumari, Semantic Analysis of Blockchain Intelligence with Proposed Agenda for Future Issues, 14, *International Journal of System Assurance Engineering & Management*, 34 (2023), https://link.springer.com/article/10.1007/s13198-023-01862-y.

eliminating bugs through enhanced bug tracking and bug audits. A recent review of over 100 research papers revealed that integrating the two technologies results increases the security, efficiency, and productivity of the applications.[19]

Most bug tracking is done manually by software engineers, which impairs bug triaging.[20] To address this problem, researchers propose PrAIoritize; an automated approach for predicting smart contract bug priorities that assist software engineers in prioritizing highly urgent bug reports.[21]

Enhanced smart contract auditing can be accomplished through deep learning techniques. In a recent study, researchers trained three deep models for detecting vulnerabilities in smart contract: Optimized-CodeBERT, Optimized-LSTM, and Optimized-CNN.[22] Experimental results show that Optimized-CodeBERT model surpasses other methods, achieving an F1-score of 93.53%.[23] To precisely extract vulnerability features, they acquired segments of vulnerable code functions to retain critical vulnerability features. Using the CodeBERT pre-training model for data preprocessing, the authors could capture the syntax and semantics of the code more accurately. The authors evaluated its performance using the SolidiFI-benchmark dataset, which consists of 9369 vulnerable contracts injected with vulnerabilities from seven different types.

In another study, researchers trained artificial neural networks (ANN), long-short term memory (LSTM), and gated recurrent unit models (GRU) and compared their accuracy, precision, recall, and receiver operating characteristic (ROC) curve values.[24] The network was trained on an open Google Big Query dataset with 7000 samples. Their results demonstrated that the LSTM model outperforms ANN and GRU.[25] Lastly, AI technology has the potential to generally improve smart contract security.[26]

3.1.5. Market for Code Reviews

The market for code reviews in 2024 is dominated by centralized market participants. The existing code review industry is subject to significant cost, inefficiencies, barriers to entry, and lack of assurances for code review job posters, among other lack of client services for code reviews. In 2024, the code review industry is dominated by several key players who can charge exorbitant and monopoly-like prices. Despite these high prices, the code review process is subject to significant flaws.

First and foremost, the selection process for the reviewer is an ongoing challenge for the existing code review process in legacy code review firms. The more hierarchical the code review process is, the lower is the quality of the reviewed code. It makes intuitive sense that the more developers review a given code set, the higher the code quality may turn out to be. However, in the legacy review process, the first reviewer within the hierarchical structure of the code review process often gets the highest priority and is often merely followed with minor upgrades by follow-on reviewers. The collective of reviewers is also not incentivized to find flaws in the code to optimize code as a work product of the collective. Rather, it is often seen as the work product of the initial reviewer with minor input from follow-up reviewers. Moreover, the more people review the code with comments that ask for clarification, the more likely it becomes that the code becomes simpler and clearer, which in turn typically increases code quality. However, that is not possible in hierarchical review processes. The hierarchical approach to code reviews undermines long-term participation with opinions from the edges of the reviewer spectrum because those reviewers either have no access or are in no position to help review the code. In other words, the more hierarchical the code review process and the more barriers to entry, the lower the quality of the code.

---

[19] *Id*.

[20] Majd Soud, Grischa Liebel and Mohammad Hamdaqa, PrAIoritize: Learning to Prioritize Smart Contract Bugs and Vulnerabilities (Working Paper), https://arxiv.org/pdf/2308.11082.pdf.

[21] Xueyan Tang, Yuying Du, Alan Lai, Ze Zhang and Lingzhi Shi, Deep Learning based Solution for Smart Contract Vulnerabilities Detection, 13, *Scientific Reports,* 1 (2023), https://www.nature.com/articles/s41598-023-47219-0.

[22] *Id*.

[23] *Id*.

[24] Rajesh Gupta *et al.*, Deep Learning-based Malicious Smart Contract Detection Scheme for Internet of Things Environment, 97, *Computers & Electrical Engineering,* 1 (2022), https://www.sciencedirect.com/science/article/pii/S004579062100519X.

[25] *Id*.

[26] Moez Krichen, *Strengthening the Security of Smart Contracts through the Power of Artificial Intelligence*, 12 Computers 107 (2023), https://www.mdpi.com/2073-431X/12/5/107.

3.1.6. Single Points of Failure

In the existing legacy code review process, the views of the reviewer and the intent of the code author are often at odds with each other without any crowd control. Because the code reviewer may wish to impose their logic on the code author, the code author may be required to rewrite code over and over even though the core functionality of the code is sound and dangerous issues were controlled for. This can be highly time-consuming and inefficient. It also calls the overall role of the code review process into question. Instead, a code review should focus on the functionality of the code and on keeping mistaken, badly constructed, and dangerous code out.

If the single author of a code review has missed something and the follow-on reviewer focused entirely on the first reviewer's concerns, the code review has a higher risk of lack of accuracy. Crowd wisdom is one way to correct possible myopia and single points of failure through standard legacy code review.

3.1.7. Timing

Depending on the setting of the code review, code reviews in legacy systems can last weeks and months. This can be exacerbated by market conditions in the digital asset market. These significant delays can impact development and may require complete rewriting of contracts because the underlying protocol may have upgraded core libraries during the code review.

3.1.8. Current Players in the Field

Even though most code review firms help clients who hope to decentralize different parts of the industry and capitalize on efficiencies created by decentralizing legacy systems, the code review industry is mainly dominated by a few players.

The current market dynamics dominated by the top 5 audit firms also create high barriers to entry for new players to enter into the code review market.

Given these downsides in the existing code review market, it is a bit ironic that one of the strongest forms of exploitation and centralized economies of scale are being created in a market that is, from the outset, supposed to help support the decentralization of disparate industries.

3.1.9. Overpricing

As a result of the centralization of the industry, most code reviews are significantly overpriced. Clients pretty much pay any price to get the stamp of approval from one of the top 5 audit firms.

3.1.10. No Controls

The centralized power also undermines attempts by other industry players to create internal or external controls on the quality of code reviews. As a result, the public has no or very weak control over the quality of code review services it receives. Job posters cannot afford to look for better-priced code reviews and are forced into considerable pricing to obtain market acceptance of their products. In turn, the centralization of the market undermines any form of downward price pressure.

Because of the power of the limited players over the overall market and the process of the code review and its outputs, the quality of code review is often suboptimal.

Moreover, there is little or no recourse for clients in cases in which the code proved to be flawed even after functionality and quality review.

3.2. Code Review DAO

The Code Review DAO (CRDAO) is tackling many of the issues that afflict the modern code review market.

The CRDAO provides a new code review platform, which facilitates a decentralized community-driven code review process that utilizes a bidding process on code reviews to drive prices down (the "Code Review Platform"). It provides open access for code reviews from anyone who qualifies—not just members of the few code review firms. At the same time, it provides full incentivization for community code reviews through its decentralized governance framework.

Given its universal access and price discovery methodology (through a public bidding process), the Code Review Platform creates low barriers to entry in the code review market. Anyone can join the Code Review Platform by submitting high-quality code reviews through the Code Review Platform portal.

The Code Review Platform also uses a community policing and audit methodology for code reviews. This is enabled by the Code Review DAO[27] (CRDAO) governance model. Moreover, the CRDAO governance and policing functions ensure less duplication of code reviews. Code Review Platform maintains strong incentives for community-driven code review audits.

The Code Review Platform can function as a first-round code review or multi-round code review with different code review teams. In summary, job posters receive low-priced, high-quality code reviews with community validation of reviews.

3.2.1. Feedback Loops

Code Review Platform provides an early feedback loop on code reviews for the developer community at a much lower price than the traditional code review and audit market.

The CRDAO crowd controls filter out idiosyncratic code reviewer preferences. Code reviews in legacy code review firms are often highly subjective which leads to rather suboptimal outcomes without crowd controls. No single developer may agree on a given set of code and its intended functionality and quality in achieving the coded objectives. This can be attributed to different programming languages with different styles and unique and often idiosyncratic preferences. Instead, the ALE mandates that reviews are subject to crowd review and policing votes by the CRDAO collective. Accordingly, code reviewers are less likely to engage in highly idiosyncratic reviews as they would need to fear slashing of rep token scores and loss of standing in the community.

3.2.2. Price Discovery

The existing code review market does not provide publicly transparent pricing of code review services. As such, the existing market arguably harms the public for the benefit of the few market players and its clients. In the existing system, the unilateral pricing is not disclosed because both the client and code reviewer may not benefit from public scrutiny of the prices.

The Code Review Platform uses a unique and fully transparent price discovery mechanism for code reviews. The Code Review Platform's price discovery mechanism works as follows:

- Job poster posts the job on the Code Review Platform portal.
- DAO internal - Job price discovery.
  - Internal DAO bids are collected in a table format next to the job posting terms and show the bids up and down on the terms 1. Time +-2. Price - +3. Reputation score range of the bidder (actual reputation of bidder is within the range), 4. Reputation stake of the bidder.
    1. At the end of the internal bidding period—the job poster reviews all the bids and selects the winner and prices are revealed.
    2. If the job poster does not select a winner after a grace period, it automatically goes to public bidding and prices are never revealed.
    3. Bidding is not public during the bidding process and (fully anonymized through reputation score ranges).
    4. All bids on the post and winner will become public at the end of the selection when the job poster has picked the winner.
- Public - Job price discovery
  - If no bids for *n* days internally in the CRDAO community member pool, the job post will be made public automatically for public bidding - public bidding follows these parameters.

[27] DAOs are short for Decentralized Autonomous Organizations. https://en.wikipedia.org/wiki/Decentralized_autonomous_organization#:~:text=Decentralized%20autonomous%20organizations%20are%20typified,dissemination%20of%20a%20distributed%20database.

1. External bidder posts DoS[28] fee to get on boarded (admin approval).
2. External bids are collected in a table format next to the job posting terms and show the bids up and down on the terms 1. Time +- 2. Price - +.
3. At the end of the external bidding period - the job poster reviews all the bids and selects the winner.
4. If the job poster does not select a winner after a grace period, it automatically ends the bidding process with no winner.
5. Bidding is not public during the bidding process and (anonymized if preferred by bidder).
6. All bids on the post and winner will become public at the end of the selection when the job poster has picked the winner.

- If the job poster accepts a bid (DAO internal or public)—the job poster now has to post a deposit for the value of the agreed code review job into a smart contract.

This price discovery mechanism serves a key public service function in that it enables full-price transparency for consumers based on the visibility of the internal and external bids on job posts. This price transparency is unique and unprecedented in a market.

Price discovery is a key public services function because, without public pricing, consumers cannot realistically select the service provider that provides the highest value to the job poster. The lack of transparency enables insider deals to the detriment of the clients, who are forced into the price a group of firms dictates.

3.2.3. Standards

The Code Review Platform also fills the void left in legacy reviews without common standards. The Code Review Platform creates a compendium of reviews and through it a common standard for code reviews that are otherwise lacking in legacy code review environments. Should a collective review code under a common set of standards, the standards help guide both the reviewers and the collective to come to a common form of expectations on the applied functionality and quality outcomes for the code.

3.2.4. Speed

The feedback provided by the ALE for the developer community enables risk-taking for dev teams who wish to move quickly through their governance and upgrade process, which in turn enables accelerated growth and scaling of experimentation.

3.2.5. Code Testing

The code reviews provided by the Code Review Platform provide a first instantiation of flat hierarchy-driven decentralized peer-reviewed code reviews. All testing performed by the Code Review Platform follows the core standards established for the CRDAO community. All testing and standards are subject to constant review and experimentation and are continually, dynamically, and evolutionarily updated in a constant feedback loop between all constituents, that is between the CRDAO member, the job posters, and public bidders and applicants for CRDAO community member status.

3.2.6. Community Audit

The audit starts with the community discussion of the code review, followed by an informal vote on the code. The community that bids for the posted jobs and then reviews the submitted job and votes on it after the forum discussion is constituted by the evolving community membership of the Code Review Platform. The informal vote shows all CRDAO members the collective wisdom as applied to the work product examined in the code review. Once the entire community knows how each member feels about the code review examined, the CRDAO now votes in a formal vote in which the reputation tokens[29] staked are at risk. This sequence of votes provides

---

[28] Denial of Service: https://en.wikipedia.org/wiki/Denial-of-service_attack

[29] "Reputation Tokens" are non-transferable tokens that cannot be valued and represent the reputation a member has within the community. They simply mirror a scoreboard. Each member of the CRDAO holds such reputation tokens to measure the merit and quality of their input into the community, presenting the member's unique and individual reputation status according to certain criteria established by the CRDAO. At the time of this writing, the criteria for reputation token allocation to CRDAO members revolved around the respective members' ability to contribute to the CRDAO code reviews, either as a job performer or as part of the community audit.

job posters with significant assurances that the code examined and the Code Review Platform report on the code adheres to the highest standards of quality available.

3.2.7. Code Review Platform Code Review Process

The code review process of the Code Review Platform revolves around CRDAO community engagement which minimizes issues of lack of crowd controls, lowers time requirements for code reviews, lowers prices of code reviews, increases developer participation, and increases overall feedback.

4. AI Learning Ecosystem Platform

4.1. Scaling the Code Review Platform with Micro Task Work

Via the Code Review Platform as a proof point, ALE is tapping into this rising micro task market for AI training data by providing users with gamified access to code reviews. Key examples that set precedent for the gamification include Axie Infinity and other scaling Ethereum games that attained worldwide audiences.

4.2. Removing Cost of Micro Task Work Duplication

Centralized micro task work requires duplication to ensure quality. All too human shortcomings of micro task works, such as limited attention span, irrationality, and inaccuracies result in verification requirements for micro task work. However, manual verification of micro task work is subject to the same human limitations. In an attempt to ensure quality of results and minimize the impact of the human limitations of their workers, requesters in centralized micro task structures set up teams of up to 15 workers to perform the same task in an effort to form a consensus.[30] The multiplication of work inherent in this process significantly increases the cost of micro task work. Requiring requesters to pay proportionally (e.g., up to 15 times) for work per project results in waste. The necessity of multiplication of work also subjects micro task workers to lower rates and lack of payment increases.

By contrast, in the ALE platform, the community organization software enables a readily available indicator of how reliable a worker or requester is on the ALE Platform. The micro task worker's reputation score is a measure of the worker's history of completing micro tasks on the ALE Platform qualitatively accurate, efficient, and consistent. The requester's reputation score is a measure of the requester's history of interacting with micro task workers on the ALE network. The ALE reputation score is formed and linked to the respective network participant's wallet address.

Market factors balance the equilibrium of supply and demand of micro work on the ALE Platform based on the workers and requesters reputation scores. If requesters have a lower reputation score, workers become less likely to accept requesters' offers. In turn, low reputation scores for micro task workers result in a lower likelihood of retention for micro task work on the ALE Platform. Requesters can select workers based on their reputation score, giving workers an incentive to keep the reputation scores high by performing micro tasks with high accuracy and efficiency. The reputation score mechanism and the building of reputation on the ALE Platform allows workers to graduate to the privilege of being a verifier. The reputation score mechanism helps discern malicious actors and simple mistakes. It also protects workers and verifiers from fraudulent requesters and sub optimally designed requests.

In summary, the market factors and market dynamics related to ALE Platform reputation scores enable a lowering of the cost of duplication as compared to centralized micro task work. If a reliable and high reputation score worker completes the tasks, the duplication may be brought from 15 to 5 or less in the decentralized code review setup. This enables unprecedented scaling of micro task work.

4.3. Transaction Cost Minimization of Micro Task Work

The ALE Platform removes transaction costs associated with micro task work. Unlike centralized mechanical turk platforms that require an existing banking relationship to receive account transfers for otherwise unbanked micro task workers, the ALE Platform operates entirely through crypto transactions.

---

[30] Neeraj Kumar, *Effective Use of Amazon Mechanical Turk (MTurk); Tips and techniques for better usage of Amazon Mechanical Turk for researchers*, Neeraj Kumar (May 2013, updated May 8, 2014), http://neerajkumar.org/writings/mturk/; Rory O'Reilly, *How the Gems Protocol Reduces Consensus by Redundancy*, GEMS (Nov. 27, 2017), https://blog.gems.org/how-the-gems-protocol-reduces-consensus-by-redundancy- b151de80ecb8.

4.4. Gamification of Micro Task Work

The gamification of micro task work can have several effects on the process of producing high-quality, diverse datasets for AI learning via the ALE Platform.

1. **Increased Engagement:** Gamification techniques of the ALE Platform include a point system, levels, leaderboards, and rewards, all of which may be utilized exclusively within the platform, that are allocated via reputation scores in the community. This system of reputation scoring in the ALE Platform enhances the engagement of crowdworkers in the ALE Platform. By introducing the reputation scores as a competitive and rewarding element, the ALE Platoform's gamification motivates workers to actively participate and complete tasks more efficiently and more responsively. This increased engagement can lead to higher-quality outputs as workers are more invested in the process. Workers gain reputation scores and are rewarded with a part of any incoming compensation (stablecoins or other major crypto currencies paid by job posters) pro rata to their reputation scores. This incentivizes them to care about the reputation score and engage with more rigor and due diligence in the micro task work. If they don't engage with the care required, they sacrifice their spot in the rankings of reputation scores which affects their participation in the job fee distribution, which is paid out pro rata to the respective reputation scores.
2. **Quality Control:** Gamification in the ALE Platform improves data quality through the consensus-based voting on work products of individuals. Workers in the ALE Platform review and rate each other's contributions, earning reputation points for accurate and consistent work. This approach helps identify and resolve discrepancies, build consensus, and ensures better quality control in the dataset creation process.
3. **Skill Development:** Gamification in the ALE Platform is focused on expertise building and thus enables skill development among crowdworkers. By providing challenges of work sets in levels that progressively increase in difficulty, workers can enhance their annotation, validation, and cleaning skills over time. All skill development is directly traceable through the reputation score of each worker. This skill development can result in improved accuracy and efficiency in dataset creation.
4. **Crowdworker Retention:** Gamification elements on the ALE Platform like badges, achievement awards, and virtual currency payments that are all related to the reputation score of each worker can enhance crowdworker retention. By recognizing and rewarding workers on the ALE Platform for their contributions, gamification fosters a sense of accomplishment and encourages workers to continue participating in micro task work. This retention is crucial for maintaining a consistent pool of experienced and reliable workers for dataset creation. Attrition rates can fall through the gamification design inherent in the ALE Platform.
5. **Scalability and Speed:** Gamification on the Code Review Platform can help expedite the dataset creation process by encouraging workers to complete tasks more quickly. This is done in an effort to maintain their reputation scores. By introducing time-based challenges in providing the work and by offering bonuses for timely completion, gamification Code Review Platform can increase the speed at which high-quality datasets are generated, enabling scalability for large-scale AI projects.

5. Conclusion

The ALE Platform emphasizes the design and implementation of its proven decentralized governance logic in combination with the gamification logic inherent in its reputation governance. Through the combination of decentralized governance and gamification of micro task work, ALE Platform ensures that gamification techniques do not compromise the quality and accuracy of the dataset. Balancing the gamification elements with appropriate quality control measures via decentralized governance is essential to maintain the integrity of the dataset creation process.

Cite this article as: Wulf A. Kaal (2024). Decentralized Governance to Optimize Human Output Datasets for AI Learning. *International Journal of Artificial Intelligence and Machine Learning,* 4(2), 52-66. doi: 10.51483/IJAIML.4.2.2024.52-66.