Research Proposal
Integrating AI and Blockchain: Developing AI Standards for Cardano

UnboundedMarket Team

October 18, 2024

Contents

1 Motivation

Recent advancements in Artificial Intelligence (AI) have led numerous projects on Cardano to integrate AI for several use cases, such as data analytics, domain-specific chatbots, and identifying bugs in smart contracts. These integrations highlight AI's potential to enhance blockchain applications, providing greater efficiency, accuracy, and automation. Specifically, in a recent study, Bhumichai et al. [2] have identified fourteen key features where AI and blockchain benefit each other, including data security and privacy, data encryption, data sharing, decentralized intelligent systems, efficiency, automated decision systems, collective decision making, scalability, system security, transparency, sustainability, device cooperation, and mining hardware design.

A major challenge in using AI models in the Cardano ecosystem is the absence of standardized frameworks to guide this integration. Without clear standards, developing efficient, secure, and interoperable AI solutions on Cardano remains limited.

In this project, we aim to address these gaps and define four foundational pillars essential for establishing standardized AI integration within Cardano (see Figure 1): Accessibility via an API, Certification Procedure, Benchmarking Procedure, and Consensus Mechanisms. These pillars are interrelated, each influencing and reinforcing the others to create a cohesive framework for AI integration (see Section 6 for a discussion). With these pillars, we seek to establish a framework that accelerates the development and integration of AI solutions on the Cardano blockchain.

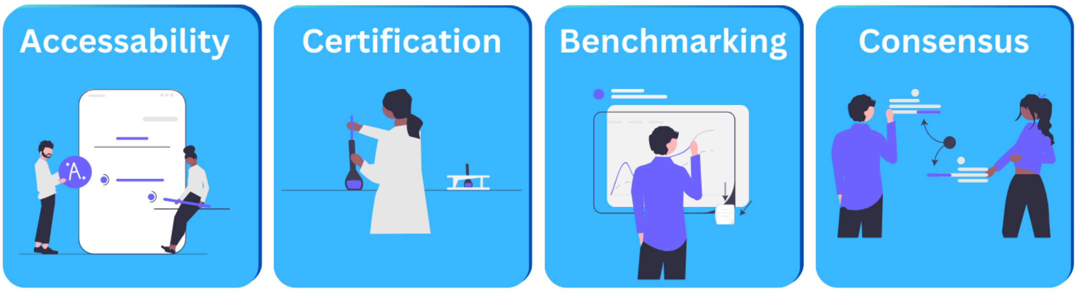

Figure 1: Four foundational pillars for standardized AI integration within Cardano.

2 Accessibility & API Framework

Integrating AI models into blockchain ecosystems like Cardano requires easy accessibility of these models. We approach this problem by establishing standards for Application Programming Interfaces (APIs). APIs play a central role in modern software development by allowing different software components to communicate and share data [8, 11]. They define a set of rules and protocols enabling interaction between different systems, such as software libraries, web services, or microservices. A well-designed API framework is important to allow for easy usability and interaction [4]. Specifically, carefully defining the requirements of the API framework enables easy interactions between AI models and decentralized applications (dApps) and, thereby, lowers the barriers to integration.

Thus, we start by formulating a set of essential requirements for such an API framework: Foremost, the API must support **standardized communication** protocols, ensuring reliable interactions between AI services and dApps. Among the available architectural styles, REST (Representational State Transfer) [8] is particularly suitable due to its widespread adoption, simplicity, and compatibility with web-based services. REST's stateless nature, which leverages standard HTTP methods for requests, enhances both the scalability and maintainability of the API framework. Over the years, developers and researchers have established a set of guidelines and best practices for REST APIs [24, 5, 25]. Despite this, many existing APIs only implement a subset of these principles, often leading to inconsistencies or limitations [21]. To create an intuitive API, it is also important to align its structure with that of established APIs from AI providers. To do so, we analyze existing standards used by leading AI providers, such as OpenAI[1], Google Vertex AI [2], Meta[3], and Hugging Face[4]. This analysis enables us to incorporate features contributing to the API's robustness, adaptability, and ease of use while maintaining compatibility with industry standards. A second essential requirement is **scalability and efficiency**. In blockchain environments, the API must handle high volumes of requests across distributed nodes with minimal latency. It should support efficient load balancing and asynchronous processing to manage increased network traffic. Moreover, appropriate **documentation and developer support** are essential accessibility components, often not properly addressed. Previous studies show that well-designed APIs increase developer productivity [15, 19]. However, common APIs often lack adequate documentation, which hinders adoption and integration [6]. To address this, the API must include extensive documentation, along with tools and resources that allow for easy implementation.

Note that the primary interaction between the API and blockchain is to ensure proper on-chain recording of certifications and benchmarks, as outlined in Section 3 and Section 4. However, we will revisit the API requirements in the next milestone, where we provide a more detailed specification.

3 Certification

Certification establishes trust and accountability in various domains, ensuring that products, services, or processes meet predefined standards and requirements. For AI models,

[1] https://platform.openai.com/docs/api-reference/introduction
[2] https://cloud.google.com/vertex-ai/docs/reference/rest
[3] https://wit.ai/docs
[4] https://huggingface.co/docs/api-inference/index

certification involves verifying that they meet established functional criteria, which promotes transparency, an important criterion for ethical and responsible use of AI [9, 13]. We propose a specific form of certification that focuses on creating standardized documentation for AI models in the form of model cards. Model cards serve as a documentation method that includes essential information about the model's capabilities, limitations, training data, and intended use cases [18]. A specific example of such is provided by IBM's Watson.ai model cards[5]. These model cards offer an in-depth overview of a model's parameters and usage guidelines, serving as a valuable reference for users and developers.

Our proposed certification specifications include a set of fixed parameters and a set of optional parameters: The set of fixed parameters that are required for every model card, ensuring consistency across different models. These parameters include **usage instructions** detailing how to implement and interact with the model, information on any **associated costs**, the model's size indicated by the **number of parameters**, and **token or usage limits**. Additionally, details about **instruction tuning**—whether the model has been fine-tuned with instruction-following data—will be included, along with a description of the **model architecture** (e.g., Transformer-based [31], recurrent neural network [26]). The **licensing** terms under which the model is released are also a critical component of the certification. Optional parameters provide additional insights that may be relevant for specific models or use cases. These may include information on **training data sources**, **performance benchmarks** on standard datasets, and other potential ethical considerations. These optional parameters allow for more detailed information on the model.

Integrating the certification process within a blockchain environment such as Cardano provides additional benefits, such as enhanced transparency, immutability, and decentralized verification [see, e.g., 33, 17, 22, inter alia]. Storing model certifications on the blockchain allows for real-time updates and version tracking, which is essential for maintaining up-to-date information as models are refined or retrained. This is particularly relevant in AI applications where models evolve rapidly, necessitating frequent updates to certification records. Blockchain's inherent properties of traceability and auditability thus support continuous monitoring of model compliance with evolving standards and regulations. This form of certification process provides greater transparency and comparability across different models and lays the foundation for continuous improvements in model benchmarking (see Section 4).

4 Benchmarking

Here, we discuss developing robust benchmarking methods to evaluate specific aspects of model performance, such as robustness, efficiency, and scalability. These benchmarking

[5] https://dataplatform.cloud.ibm.com/docs/content/wsj/analyze-data/fm-models.html?context=wx

are designed to enable comparative performance indicators, helping stakeholders select appropriate AI models for specific applications. Developing methods to evaluate AI performance in a blockchain environment, ensuring that AI models are robust, efficient, and scalable.

AI model benchmarks consist of a set of tasks designed to evaluate the capabilities of AI models, typically focusing on a specific domain. Well-known examples are ImageNet [7], which evaluates visual recognition performance, and GLUE (General Language Understanding Evaluation) [32], which assesses language understanding tasks. These benchmarks provide standardized datasets and evaluation protocols that facilitate model comparison and drive progress in specific AI research areas. However, conventional benchmarks often have significant limitations. They tend to focus narrowly on static datasets and specific tasks, which may not fully capture the complexity and variability of real-world scenarios [23]. To address this problem, researchers have started to create combined test suites, which combine a wide range of benchmarks. Examples of large language models are the Language Model Evaluation Harness [10, 3] or HELM (Holistic Evaluation of Language Models) [16].

However, recent trends, such as those seen with OpenAI, have shifted away from open-source practices, with many of the best-performing models on public evaluation suites being closed-source. In these cases, key details such as model architecture, parameter counts, and training data are kept undisclosed. This lack of transparency prevents reliable and consistent evaluation, making it difficult for stakeholders to verify model performance claims or reproduce results. We argue that combining blockchain technology with AI benchmarking can significantly improve both the reliability and transparency of the evaluation process, especially when combined with certification, e.g., via model cards, as discussed in Section 3. Specifically, by jointly storing model cards and benchmarking outcomes on-chain, stakeholders can ensure that the records of an AI model's performance, training data, and other relevant characteristics are tamper-proof and publicly verifiable.

Thus, to summarize the requirements for benchmarking AI models in a blockchain environment, we require methods to incorporate **diverse evaluation tasks**, ensuring broad applicability across domains while addressing the transparency issues posed by closed-source models, by **jointly storing model cards and evaluation results on-chain**.

5 Consensus

Here we discuss methods to establish a form of consensus between AI multiple AI models. Consensus mechanisms are a fundamental component of blockchains and ensure that distributed participants reach agreement on a unified version of the ledger. Traditionally, algorithms such as Proof of Work (PoW) [1, 20] and Proof of Stake (PoS) [14] maintain the integrity of the blockchain.

In the context of AI, blockchain can enable consensus among different models and their outputs. Recent research has explored ensemble-based strategies to enhance the performance of large language models (LLMs) [12, 34], which involve combining multiple models to solve tasks, leveraging their collective strengths to improve accuracy, robustness, and generalization.

Blockchain adds to ensemble strategies by providing a decentralized, immutable ledger to record model predictions, which ensures transparency and traceability throughout the decision-making process [27, 28]. Additional benefits can be obtained by combining ensemble strategies with AI model certification and benchmarking, as discussed Section 3 and Section 4. Verified information about a model and its performance on specific tasks can be used to weigh its contributions within the ensemble, increasing the overall accuracy and reliability. In a similar way, certification and benchmarking could help identify and mitigate the influence of faulty or malicious models. This is important for managing Byzantine faults in decentralized AI systems, where models fail or act maliciously. Resilience to such faults is necessary to maintain system integrity and reliability. Stroble et al. [29, 30] highlight the use of blockchain technology to enhance Byzantine fault tolerance, underscoring the importance of robust consensus mechanisms capable of handling adversarial behaviors in distributed AI models.

Thus, achieving effective consensus in decentralized AI systems requires **transparency and traceability of models and model outputs**, **resilience against faulty or malicious participants**, and **integration of verified model performance data** to weight the contributions of individual models within ensemble-like strategies.

6 Discussion

In this section, we revisit and discuss the requirements outlined in the previous sections, beginning with a summary in Section 6.1 and then addressing the challenges in Section 6.2.

6.1 Summary

For accessibility, we have defined standardization, scalability, and documentation as our central requirements. A standardized API framework allows for a smooth interaction between AI models and decentralized applications, while scalability guarantees that the system can handle increasing demand efficiently. Extensive documentation supports both developers and users in understanding how to implement and use AI models within the Cardano ecosystem.

Our key requirements for certification are transparency and verifiability. We have proposed a certification process by creating model cards that document essential information, such as model architecture, training data, and performance metrics, ensuring that models meet predefined standards. Storing certification data on-chain increases transparency by enabling version tracking and decentralized verification.

Reliable benchmarking is important for evaluating and comparing AI models. By jointly storing benchmarking results alongside certification, we ensure that performance evaluations are transparent, immutable, and verifiable. This integration allows stakeholders to assess AI models confidently, knowing that the benchmark data and certification records are secure and tamper-proof. Lastly, for consensus mechanisms, we propose that ensemble techniques, combined with on-chain certification and benchmarking, provide a resilient framework for validating AI outputs. These mechanisms ensure that only accurate and trustworthy results are accepted, promoting trust and reliability in decentralized AI environments. While we have established requirements for each of these four pillars, their successful implementation requires careful consideration of various challenges, which we discuss next.

6.2 Challenges

Foremost, **scalability of the API** is an important challenge. As the number of users and AI models increases, the standardized API framework must handle growing demand without compromising performance. Ensuring low latency and high throughput in a decentralized environment requires efficient network protocols and optimized infrastructure. Second, **defining comprehensive certification standards in the model cards** presents a significant challenge. The model cards must include all essential information to reflect the AI models and their training processes accurately. **Finding ways to store benchmarking results on-chain** is critical for transparency and verifiability. Determining the optimal method for recording benchmarking data involves balancing the need for detailed performance metrics with storage efficiency and cost considerations. Finally, **achieving robust consensus** remains a core challenge. Consensus mechanisms must be resilient against faulty or malicious participants, providing Byzantine fault tolerance to maintain system integrity and reliability.

7 Conclusion

In this research proposal, we have outlined our framework for integrating artificial intelligence (AI) into the Cardano blockchain ecosystem by defining four foundational pillars: accessibility, certification, benchmarking, and consensus mechanisms. By defining clear requirements for each pillar and acknowledging the challenges ahead, we lay the groundwork for developing a decentralized AI infrastructure that is widely accessible, transparent, and robust. Our framework is designed to enhance trust and productivity in AI systems within the Cardano blockchain, offering clear, actionable guidelines.

References

[1] Adam Back. Hashcash - a denial of service counter-measure. 2002.

[2] Dhanasak Bhumichai, Christos Smiliotopoulos, Ryan Benton, Georgios Kambourakis, and Dimitrios Damopoulos. The convergence of artificial intelligence and blockchain: The state of play and the road ahead. *Information*, 15(5), 2024.

[3] Stella Biderman, Hailey Schoelkopf, Lintang Sutawika, Leo Gao, Jonathan Tow, Baber Abbasi, Alham Fikri Aji, Pawan Sasanka Ammanamanchi, Sidney Black, Jordan Clive, Anthony DiPofi, Julen Etxaniz, Benjamin Fattori, Jessica Zosa Forde, Charles Foster, Jeffrey Hsu, Mimansa Jaiswal, Wilson Y. Lee, Haonan Li, Charles Lovering, Niklas Muennighoff, Ellie Pavlick, Jason Phang, Aviya Skowron, Samson Tan, Xiangru Tang, Kevin A. Wang, Genta Indra Winata, François Yvon, and Andy Zou. Lessons from the trenches on reproducible evaluation of language models, 2024.

[4] Joshua Bloch. How to design a good api and why it matters. In *Proc. 21st ACM SIGPLAN Conference (OOPSLA)*, pages 506–507, Portland, Oregon, 2006.

[5] Frederik Bülthoff and Maria Maleshkova. Restful or restless-current state of today's top web apis. volume 1165, pages 42–51, 05 2014.

[6] Alex Cummaudo, Rajesh Vasa, John Grundy, and Mohamed Abdelrazek. Requirements of api documentation: A case study into computer vision services. *IEEE Transactions on Software Engineering*, 48(6):2010–2027, 2022.

[7] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009.

[8] Roy T. Fielding. Architectural styles and the design of network-based software architectures"; doctoral dissertation. 2000.

[9] Luciano Floridi and Josh Cowls. A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review*, 1(1), jul 1 2019. https://hdsr.mitpress.mit.edu/pub/l0jsh9d1.

[10] Leo Gao, Jonathan Tow, Baber Abbasi, Stella Biderman, Sid Black, Anthony DiPofi, Charles Foster, Laurence Golding, Jeffrey Hsu, Alain Le Noac'h, Haonan Li, Kyle McDonell, Niklas Muennighoff, Chris Ociepa, Jason Phang, Laria Reynolds, Hailey Schoelkopf, Aviya Skowron, Lintang Sutawika, Eric Tang, Anish Thite, Ben Wang, Kevin Wang, and Andy Zou. A framework for few-shot language model evaluation, 07 2024.

[11] Daniel Jacobson, Greg Brail, and Dan Woods. *APIs: A Strategy Guide*. O'Reilly Media, Inc., 2011.

[12] Dongfu Jiang, Xiang Ren, and Bill Yuchen Lin. Llm-blender: Ensembling large language models with pairwise comparison and generative fusion. In *Proceedings of the 61th Annual Meeting of the Association for Computational Linguistics (ACL 2023)*, 2023.

[13] Anna Jobin, Marcello Ienca, and Effy Vayena. The global landscape of ai ethics guidelines. *Nature Machine Intelligence*, 1:389 – 399, 2019.

[14] Sunny King and Scott Nadal. Ppcoin: Peer-to-peer crypto-currency with proof-of-stake, 2012.

[15] Timothy C. Lethbridge, Susan Elliott Sim, and Janice Singer. Studying software engineers: Data collection techniques for software field studies. *Empirical Software Engineering*, 10(3):311–341, 2005.

[16] Percy Liang, Rishi Bommasani, Tony Lee, Dimitris Tsipras, Dilara Soylu, Michihiro Yasunaga, Yian Zhang, Deepak Narayanan, Yuhuai Wu, Ananya Kumar, Benjamin Newman, Binhang Yuan, Bobby Yan, Ce Zhang, Christian Alexander Cosgrove, Christopher D Manning, Christopher Re, Diana Acosta-Navas, Drew Arad Hudson, Eric Zelikman, Esin Durmus, Faisal Ladhak, Frieda Rong, Hongyu Ren, Huaxiu Yao, Jue WANG, Keshav Santhanam, Laurel Orr, Lucia Zheng, Mert Yuksekgonul, Mirac Suzgun, Nathan Kim, Neel Guha, Niladri S. Chatterji, Omar Khattab, Peter Henderson, Qian Huang, Ryan Andrew Chi, Sang Michael Xie, Shibani Santurkar, Surya Ganguli, Tatsunori Hashimoto, Thomas Icard, Tianyi Zhang, Vishrav Chaudhary, William Wang, Xuechen Li, Yifan Mai, Yuhui Zhang, and Yuta Koreeda. Holistic evaluation of language models. *Transactions on Machine Learning Research*, 2023. Featured Certification, Expert Certification.

[17] D S V Madala, Mahabir Prasad Jhanwar, and Anupam Chattopadhyay. Certificate transparency using blockchain. In *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 71–80, 2018.

[18] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. Model cards for model reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19, page 220–229, New York, NY, USA, 2019. Association for Computing Machinery.

[19] Brad A. Myers, Amy J. Ko, Thomas D. LaToza, and YoungSeok Yoon. Programmers are users too: Human-centered methods for improving programming tools. *Computer*, 49(7):44–52, 2016.

[20] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. *Cryptography Mailing list at https://metzdowd.com*, 03 2009.

[21] Andy Neumann, Nuno Laranjeiro, and Jorge Bernardino. An analysis of public rest web service apis. *IEEE Transactions on Services Computing*, 14(4):957–970, 2021.

[22] Shuyi Pu and Jasmine Siu Lee Lam. The benefits of blockchain for digital certificates: A multiple case study analysis. *Technology in Society*, 72:102176, 2023.

[23] Deborah Raji, Emily Denton, Emily M. Bender, Alex Hanna, and Amandalynne Paullada. Ai and the everything in the whole wide world benchmark. In J. Vanschoren and S. Yeung, editors, *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, volume 1, 2021.

[24] Dominik Renzel, Patrick Schlebusch, and Ralf Klamma. Today's top "restful" services and why they are not restful. In X. Sean Wang, Isabel Cruz, Alex Delis, and Guangyan Huang, editors, *Web Information Systems Engineering - WISE 2012*, pages 354–367, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[25] Carlos Rodriguez, Marcos Baez, Florian Daniel, Fabio Casati, Juan Trabucco, Luigi Canali, and Gianraffaele Percannella. Rest apis: A large-scale analysis of compliance with principles and best practices. pages 21–39, 06 2016.

[26] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986.

[27] Khaled Salah, M. Habib Ur Rehman, Nishara Nizamuddin, and Ala Al-Fuqaha. Blockchain for ai: Review and open research challenges. *IEEE Access*, 7:10127–10149, 2019.

[28] Muhammad Shafay, Raja Wasim Ahmad, Khaled Salah, Ibrar Yaqoob, Raja Jayaraman, and Mohammed A. Omar. Blockchain for deep learning: review and open challenges. *Cluster Computing*, 26:197 – 221, 2021.

[29] Volker Strobel, Eduardo Castelló Ferrer, and Marco Dorigo. Managing byzantine robots via blockchain technology in a swarm robotics collective decision making scenario. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, AAMAS '18, page 541–549, Richland, SC, 2018. International Foundation for Autonomous Agents and Multiagent Systems.

[30] Volker Strobel, Eduardo Castelló Ferrer, and Marco Dorigo. Blockchain technology secures robot swarms: A comparison of consensus protocols and their resilience to byzantine robots. *Frontiers in Robotics and AI*, 7, 2020.

[31] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[32] Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel Bowman. GLUE: A multi-task benchmark and analysis platform for natural language understanding. In Tal Linzen, Grzegorz Chrupała, and Afra Alishahi, editors,

Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP, pages 353–355, Brussels, Belgium, November 2018. Association for Computational Linguistics.

[33] Xiwei Xu, Cesare Pautasso, Liming Zhu, Vincent Gramoli, Alexander Ponomarev, An Binh Tran, and Shiping Chen. The blockchain as a software connector. In *2016 13th Working IEEE/IFIP Conference on Software Architecture (WICSA)*, pages 182–191, 2016.

[34] Fuzhao Xue, Zian Zheng, Yao Fu, Jinjie Ni, Zangwei Zheng, Wangchunshu Zhou, and Yang You. Openmoe: An early effort on open mixture-of-experts language models. *arXiv preprint arXiv:2402.01739*, 2024.