Tokenizing Tomorrow: Introducing the Global Tokenized Startup Financing (GTSF) Dataset for Entrepreneurial Finance

Yuan Sun, Che Zhang, Jiajun Xin*[1]

Keywords: tokenization, entrepreneurial finance, green innovation, model context protocol, digital assets.

1. Introduction

The motivation for this research stems from the author's extensive work on green innovation financing and the rapidly accelerating trend of carbon trading mechanisms moving on-chain. Chernenko, Lerner, and Zeng (2021) document how "Founder-friendly" venture financings and nontraditional venture investors have both flourished over the past decade, yet the emergence of tokenization represents an entirely new paradigm that existing data infrastructure cannot adequately capture. As demonstrated by recent developments in 2025, the tokenization of carbon credits in Canada operates within a complex regulatory landscape that is still evolving (Osler, Hoskin & Harcourt LLP, 2025), reflecting a global phenomenon that is fundamentally reshaping environmental finance.

Beyond its growing popularity, tokenization is particularly consequential for startups, whose characteristics often amplify traditional financing frictions. Early-stage ventures typically lack tangible collateral, face high cash-flow volatility, and operate with extreme information opacity. Tokenization directly mitigates these issues by enabling fractionalized ownership that introduces secondary-market liquidity, embedding programmable governance rules that reduce information asymmetry, and broadening global investor access to previously localized and illiquid opportunities. In this sense, tokenization is a fundamental contracting technology that reshapes how entrepreneurial capital is mobilized. The Global Tokenized Startup Financing (GTSF) dataset addresses this critical gap by providing the first systematic documentation of

[1]Yuan Sun (Corresponding Arthor), Assistant professor at Shanghai University of Finance and Economics, School of Digital Economics, China, sunyuan89574@gmail.com; Che Zhang, Research assistant professor at Tsinghua University, School of Economics and Management, China, zhangche@sem.tsinghua.edu.cn; Jiajun Xin, Senior doctoral student at Hong Kong University of Science and Technology, School of Engineering, China, jxin@cse.ust.hk. All errors and omissions are solely ours.

tokenized startup financing, with an initial focus on green innovation companies that are pioneering new models of sustainable entrepreneurship.

The GTSF dataset represents a foundational contribution to establish tokenized entrepreneurial finance as a distinct research area. According to recent industry data, in 2025, more than 60% of new carbon credit platforms are using blockchain, with most focused on agriculture and forestry projects (F6S, 2025), demonstrating the rapid mainstreaming of tokenization technology. Our initial dataset will document approximately 1,000 companies across 15 major tokenization ecosystems, including the United States, Singapore, China, India, United Kingdom, Canada, South Korea, Hong Kong SAR, Vietnam, Israel, Australia, Germany, Brazil, Japan, and other emerging markets, covering the period from 2021 to 2025. As noted by Brickken (2025), tokenization allows startups to raise capital efficiently while offering investors greater accessibility and liquidity, fundamentally changing the dynamics of early-stage financing.

2. The Inadequacy of Existing Research Infrastructure

The study of entrepreneurial finance has long been constrained by fundamental data limitations that have only intensified with the emergence of blockchain-based financing mechanisms. Kaplan and Strömberg (2003) highlight the opacity of venture capital contracts and the difficulty of obtaining comprehensive deal terms, while Gompers and Lerner (2000) document the systematic biases in databases that only capture successful investments. The problem of missing data is particularly acute for early-stage financing, where Hellmann and Puri (2002) note that many angel investments and seed rounds go entirely unrecorded in traditional databases.

These data limitations have profound implications for our understanding of entrepreneurial finance. Ewens, Gorbenko, and Korteweg (2022) demonstrate that selection bias in venture capital databases can lead to incorrect inferences about the relationship between contract terms and startup outcomes. Similarly, Abuzov, Gornall, and Strebulaev (2025) show how information disclosure concerns shape venture capital behavior in ways that remain hidden from traditional data sources. The geographic concentration of venture capital documented by Chen and Ewens (2025) may partly reflect data availability rather than true market structure.

The emergence of tokenized financing has exposed additional limitations in traditional data infrastructure. Existing blockchain research has largely focused on cryptocurrency trading and decentralized finance applications rather than entrepreneurial financing. Lehar and Parlour (2025) provide detailed analysis of the Uniswap decentralized exchange mechanism, demonstrating the rich data available from blockchain systems, yet their focus remains on secondary market trading rather than primary capital formation. Similarly, Cong, Li, and Wang (2021) develop theoretical models of token economies but lack comprehensive empirical data to test

their predictions. As Li, Mann, and Goldstein (2025) note in their work on digital tokens and platform building, the integration of tokenization with traditional financing mechanisms requires new data infrastructure to capture these hybrid models.

3. AI-Powered Data Construction Methodology

The GTSF dataset employs cutting-edge artificial intelligence techniques, particularly autonomous agent tracking systems, to construct comprehensive profiles of tokenized startup ecosystems. Our methodology deploys sophisticated AI agents that continuously monitor, analyze, and synthesize information across multiple dimensions of the tokenization ecosystem, creating a living dataset that evolves in real-time as the market develops.

At the core of our approach is a multi-agent architecture where specialized AI agents track different aspects of each tokenized startup's ecosystem. Token tracking agents monitor blockchain networks to identify token creation events, track holder distributions, and analyze transaction patterns using advanced pattern recognition to distinguish genuine startup tokens from other cryptocurrency activities. Social ecosystem agents simultaneously monitor Telegram groups, Discord servers, Twitter discussions, Reddit threads, and GitHub repositories, using natural language processing to extract sentiment, identify key developments, and measure community engagement levels.

Our architecture also integrates decentralized finance (DeFi) primitives and smart contract analytics. DeFi integration allows agents to track how startup-issued tokens are embedded into liquidity pools, staking contracts, or lending protocols, distinguishing genuine financing tokens from purely speculative instruments. Smart contract parsing further enables the extraction of contractual features such as vesting schedules, governance rights, or revenue-sharing clauses, thereby bridging on-chain code with the contract-theory literature in entrepreneurial finance. This comprehensive approach captures not just token issuance but the entire lifecycle of tokenized financing instruments within the broader DeFi ecosystem.

The ecosystem mapping capabilities of our AI system represent a particular innovation in data construction. By analyzing the flow of tokens between wallets combined with social media data and public announcements, these agents construct detailed network maps showing how knowledge, capital, and talent flow through the tokenized startup ecosystem. This network perspective is essential for understanding spillover effects and the role of serial entrepreneurs and investors in the tokenization economy. Our agents also employ advanced validation techniques, with cross-referencing agents continuously verifying information across multiple sources and machine learning models identifying potential data anomalies that might indicate fraudulent projects or market manipulation.

4. Model Context Protocol (MCP) Enabled Multi-Modal Data Access Infrastructure

The GTSF dataset will be accessible through three complementary channels, with a particular emphasis on our innovative Model Context Protocol (MCP) implementation. First, we provide a traditional downloadable dataset for standard econometric analyses. Second, we offer a web portal for interactive exploration of the tokenization ecosystem with network visualizations and real-time metrics. Third, and most innovatively, we implement comprehensive MCP services that enable natural language interaction with the dataset through advanced AI assistants.

The MCP service represents a paradigm shift in research data interaction. Researchers can ask complex questions like "Which green technology startups in Southeast Asia successfully raised funds through tokenization in 2024?" and the system interprets these queries, performs appropriate statistical analyses, and presents results with proper econometric context. The MCP maintains awareness of the research conversation, suggesting related analyses, identifying potential confounding variables, and recommending robustness checks based on best practices in the entrepreneurial finance literature. This integration of real-time blockchain data with historical patterns through conversational interaction dramatically accelerates the research process while maintaining methodological rigor, making sophisticated tokenization research accessible to scholars regardless of their blockchain expertise.

5. Green Innovation Focus and Anticipated Impact

Our initial focus on green innovation companies reflects both the author's expertise in this domain and the particular importance of understanding how tokenization affects sustainable entrepreneurship. According to recent market developments, tokenization of carbon credits represents a paradigm shift in sustainable finance, leveraging blockchain technology to enhance transparency, liquidity, and accessibility in carbon markets (Tech.eu, 2025). Green technology startups face unique financing challenges, including long development cycles, capital intensity, and difficulty communicating complex technical value propositions to traditional investors. Tokenization potentially addresses these challenges by enabling global investor participation and creating liquid markets for patient capital.

The comprehensive nature of the GTSF dataset will enable research addressing fundamental questions about the transformation of entrepreneurial finance. Researchers will be able to investigate whether tokenization truly democratizes access to startup investment or merely shifts risks to unsophisticated investors. The dataset will facilitate studies of optimal token design, examining how different mechanisms for vesting, governance, and utility features affect both fundraising success and long-

term value creation. Furthermore, the rich ecosystem data will enable investigation of network effects, knowledge spillovers, and the role of community governance in innovation trajectories.

By establishing tokenized entrepreneurial finance as a distinct research area with proper data infrastructure, the GTSF dataset will catalyze a new stream of academic research while informing policy discussions about the regulation of digital assets and the future of innovation financing. Evidence from the GTSF dataset will help determine whether tokenization can play a meaningful role in financing the transition to a sustainable economy. As recent developments show, major financial institutions are already embracing this transformation, with DBS, Franklin Templeton and Ripple joining forces to roll out tokenized trading and lending services for institutional investors (StartupNews, 2025), signaling the mainstream adoption of tokenized financial infrastructure.

References

Abuzov, R., Gornall, W., & Strebulaev, I. A. (2025). The value of privacy and the choice of limited partners by venture capitalists. *Journal of Finance*, 80(4), 2156-2198.

Chen, J., & Ewens, M. (2025). Venture capital and startup agglomeration. *Journal of Finance*, 80(5), 2387-2423.

Chernenko, S., Lerner, J., & Zeng, Y. (2021). Mutual funds as venture capitalists? Evidence from unicorns. *Review of Financial Studies*, 34(5), 2362-2410.

Cong, L. W., Li, Y., & Wang, N. (2021). Tokenomics: Dynamic adoption and valuation. *Review of Financial Studies*, 34(3), 1105-1155.

Ewens, M., Gorbenko, A., & Korteweg, A. (2022). Venture capital contracts. *Journal of Financial Economics*, 143(1), 131-158.

Gompers, P., & Lerner, J. (2000). Money chasing deals? The impact of fund inflows on private equity valuations. *Journal of Financial Economics*, 55(2), 281-325.

Hellmann, T., & Puri, M. (2002). Venture capital and the professionalization of start-up firms: Empirical evidence. *Journal of Finance*, 57(1), 169-197.

Kaplan, S. N., & Strömberg, P. (2003). Financial contracting theory meets the real world: An empirical analysis of venture capital contracts. *Review of Economic Studies*, 70(2), 281-315.

Lehar, A., & Parlour, C. (2025). Decentralized exchange: The Uniswap automated market maker. *Journal of Finance*, 80(2), 891-936.

Li, J., Mann, W., & Goldstein, I. (2025). Digital tokens and platform building. *Review of Financial Studies*, 38(7), 1921-1954.

Brickken. (July 2025). The official 2025 tokenization report. https://www.brickken.com/en/startups-tokenization

F6S. (July 2025). 100 top tokenization companies - July 2025. https://www.f6s.com/companies/tokenization/mo

Osler, Hoskin & Harcourt LLP. (February 2025). Tokenized carbon credits: How blockchain is revolutionizing carbon markets. https://www.osler.com/en/insights/updates/tokenized-carbon-credits-blockchain-revolutionizing-markets/

StartupNews. (September 2025). DBS, Franklin, Ripple launch tokenized lending on XRP ledger. https://startupnews.fyi/2025/09/18/dbs-franklin-ripple-launch-tokenized-lending-on-xrp-ledger/

Tech.eu. (May 2025). Will tokenization go mainstream as founders seek alternatives to traditional funding? https://tech.eu/2025/05/09/will-tokenization-go-mainstream-as-founders-seek-alternatives-to-traditional-funding/