

Louvain School of Management

AI-Driven Whitepaper Analysis and ICO Performance: Predicting Fundraising Success, Token Trading, and Post-ICO Returns with Large Language Models

Authors: G. Collin and T. Warscotte
Supervisor: Pr. James Thewissen Academic
year 2025-2026
Master of Management

DECLARATION

The declaration below is **mandatory and must appear on the first page (after the front page) of your master's thesis.**

STATEMENTS ON YOUR USE OF GENERATIVE AI

Confirm the following statements:

Statements on responsible use of generative AI	Your answer
The content of my/our master's thesis is my/our original output, even if I/we used generative AI to help prepare it.	☒ Yes ☐ No
I/we master the theories, tools and methods that I/we use for the thesis.	☒ Yes ☐ No
I/we verified that the output presented in the thesis is accurate and valid, and that it contains no hallucination, no false reference, etc.	☒ Yes ☐ No
I/we acknowledge that I/we master the final output and that I am/we are able to explain it and answer questions about it.	☒ Yes ☐ No
I/we made sure not to provide generative AI tools with access to confidential data or information	☒ Yes ☐ No

SIGNATURE

Sign your declaration:

	Author last name, first name(s)	Date	Signature
1	Warscotte Thomas	May 20th 2026	
2	Collin Guillaume	May 20th 2026	

Table of content

1. Abstract :	6
2. Introduction	7
3. Literature Review	8
3.1. ICO performance and the multidimensional definition of success	8
3.2. Whitepapers as disclosure documents under information asymmetry	9
3.3. AI-based prediction in ICO markets and comparison with related studies	10
3.4. LLMs and fundraising prediction in entrepreneurial finance	11
3.5. Research gap and contribution	13
3.6. Hypothesis	14
H1a: AI-based prediction and token trading success	14
H1b: AI-based prediction and fundraising success	14
H2: AI-based selection and buy-and-hold returns	15
H3: Model performance and investment return	15
4. Methodology	16
4.1 Data and sample construction	16
4.2. Whitepaper analysis pipeline	17
4.2.1 Model selection	17
4.2.2 Prompt standardisation	17
4.2.3 Data collection protocol	18
4.3. Evaluation of classification performance	19
4.4. Token trading success regressions	20
4.5. Fundraising success regressions	21
4.6. Buy-and-hold return regressions	21
4.7. AI-selected portfolios and return comparison	22
4.8. Model comparison	23
4.9. Methodological limitations	23
5. Results Analysis	24
5.1. Can AI Classify ICO Success?	24
5.1.1 Overall Accuracy: A Misleading Metric	24
5.1.2 Recall and Precision: The Core Performance Trade-Off	25
5.1.3 Logistic Regression: Does AI Add Independent Predictive Information?	26
5.2. Fundraising Signal: Does AI Prediction Explain the Amount Raised?	27
5.2.1 OLS Regression Results: $\log(\text{AmountRaised})$	27
5.2.2 Control Variable Analysis	28
5.2.3 Interpreting the AI-Fundraising Relationship	29
5.3. Financial Signal: Does AI Prediction Translate into Returns?	30

5.3.1	Portfolio Returns and Statistical Tests: An Integrated Overview.....	30
5.3.2	OLS Regression on BHR: Controlling for ICO-Level Covariates.....	31
5.3.3	The Paradox of Model Size: A Non-Linear Relationship	32
6.	Discussion	33
6.1.	Answering the Research Question	33
6.2.	Practical Implications for Investors	35
6.3.	Limitations.....	36
6.4	Avenues for Future Research.....	37
7.	Conclusion.....	38
8.	References	40
9.	Appendix.....	42
	Appendix A – Confusion Matrices and Prediction Accuracy	42
	Table A1. ChatGPT-5.4 prediction matrix.....	42
	Table A2. GPT-Mini prediction matrix.....	42
	Table A3. GPT-Nano prediction matrix	42
	Table A4. Mistral Medium 3 prediction matrix	42
	Table A5. Prediction accuracy and success prediction rates	43
	Appendix B – Logistic Regression Results (ICO Success Prediction)	44
	Table B1. Logistic regression – ChatGPT-5.4.....	45
	Table B2. Logistic regression – GPT-Mini.....	47
	Table B3. Logistic regression – GPT-Nano	49
	Table B4. Logistic regression – Mistral Medium 3	51
	Table B5. Logistic regression model statistics.....	52
	Appendix C – Comparative Model Quality	53
	Table C1. Comparative classification and logistic model quality	53
	Appendix D – Amount Raised Regressions.....	54
	Table D1. AmountRaised regression – ChatGPT-5.4	55
	Table D2. AmountRaised regression – GPT-Mini	57
	Table D3. AmountRaised regression – GPT-Nano.....	59
	Table D4. AmountRaised regression – Mistral Medium 3.....	61
	Table D5. AmountRaised residuals and model statistics.....	63
	Appendix E – BHR Regression Results	64
	Table E1. AbnormalBHR regression – ChatGPT-5.4	65
	Table E2. AbnormalBHR regression – GPT-Mini	67
	Table E3. AbnormalBHR regression – GPT-Nano.....	69
	Table E4. AbnormalBHR regression – Mistral Medium 3.....	71
	Appendix F – Portfolio Comparison	73

Table F1. Portfolio summary statistics including Mistral	73
Table F2. Welch t-tests against the Real Success benchmark.....	74

1. Abstract :

The thesis explores whether artificial intelligence language models can forecast success and future financial performance with the information presented in whitepapers for ICOs. Initial Coin Offerings are also known to suffer from severe information asymmetry, weak disclosure requirements, and limited investor protection, which is why whitepapers are an important but challenging source of information for investors. A novel dataset containing 1,202 ICOs is employed and four large language models are compared for this paper, ChatGPT-5.4, ChatGPT-5.4 Mini, ChatGPT-5.4 Nano, and Mistral Medium 3.

The analysis involves three dimensions of ICO performance: token trading success, fundraising success, and abnormal buy-and-hold returns. The outputs indicate that AI-generated predictions represent essential information about token trading success and amount raised. Mistral presents the best classification performance, with the highest recall and the lowest AIC, but it also produces the largest number of false positives. Each of the four models is significantly correlated with increased fundraising amount; thus, we conclude that LLMs are able to capture whitepaper signals pertaining to investor behaviour in ICO phases.

However, the implications are narrower for financial performance. Only GPT-Nano has statistically significant positive association with abnormal BHR, but such results on a very small portfolio can only be used cautiously. Broadly, the research indicates that AI language models are able to recognize signals associated with ICO attractiveness and market viability, but are more unclear as to their prognostications for longer-term investment performance.

2. Introduction

ICOs (Initial Coin Offerings) is a popular type of business finance in the blockchain era. This enables project teams to raise capital directly from investors by issuing digital tokens, and thereby cuts out intermediaries like banks, venture capitalists, and underwriters. While such a lack of intermediation opens up new possibilities for new financing avenues with the blockchain technology being pursued at infancy stage, it also raises information asymmetries and opens up uncertainty for investors.

Whitepapers have a pivotal role in all this. These papers are frequently the central channels through which ICO projects communicate their technology, business model, tokenomics, roadmap, team, and market opportunities. But since these documents are generally voluntary, unaudited, and inadequately standardized, assessing their usefulness is also highly difficult. A whitepaper could signal a meaningful signal of project quality but could also be seen as persuasive or promotional in nature more than delivering long-term benefit.

Recent developments in Artificial Intelligence offer a new way to examine this problem. Large language models are able to process complex textual documents and may identify predictive signals in ICO whitepapers that are difficult to capture through traditional variables or manual analysis.

However, it remains unclear whether these models mainly detect short-term fundraising attractiveness, or whether they can also identify projects that later become tradable and generate stronger post-listing performance.

This thesis therefore investigates the following research question: Can AI language models predict ICO fundraising success and future financial performance from whitepapers? The study addresses this question by analysing a dataset of 1,202 ICOs and assessing the performance of 4 language models: ChatGPT-5.4, ChatGPT-5.4 Mini, ChatGPT-5.4 Nano, and Mistral Medium 3. We will test predicting capabilities in terms of three types of ICO performance: token trading success, amount raised, and abnormal buy-and-hold returns.

The contribution of the thesis is at two scales. First, it tests and validates general-purpose large language models directly in ICO whitepapers in terms of their performance, rather than using either conventional machine-learning techniques or manually generated textual indicators. Next, it distinguishes between different dimensions of ICO performance. Through this, we can discuss whether AI models capture such points as fundraising attractiveness, market viability, long-term financial value or some combination of these three elements.

The remainder of the thesis is structured as follows. Section 3 reviews the relevant literature and develops the hypotheses. Section 4 presents the data and methodology. Section 5 reports the empirical results. Section 6 discusses the findings, limitations, and avenues for future research. Section 7 concludes.

3. Literature Review

3.1. ICO performance and the multidimensional definition of success

Initial Coin Offerings (ICOs) are a decentralized form of entrepreneurial finance consisting of blockchain-based projects issuing digital tokens directly to investors. Token offerings allow project teams to raise capital without relying on banks, venture capitalists, or underwriters, which makes them different from traditional equity funding. Although this disintermediated structure lowers the barriers for early-stage start-ups, it also creates real informational and regulatory risks.

ICO markets have poor disclosure rules, limited protection for individual investors, and a high information asymmetry between project organizers and investors (Zetzsche et al., 2017). Token-based fundraising is therefore a high-uncertainty context for investors. They frequently need to weigh early-stage companies with little operating history, no audited or publicly available financial statements, and very sophisticated business models. Momtaz (2024) conducts a search-theoretical analysis of the ICO market and argues that distributed ledger technology reduces entry barriers, raises market granularity, and makes valuable projects harder to identify. Consequently, capital allocation in ICO markets relies extremely upon the interpretation of observable project-level signals.

As a result, a large part of the literature examines which factors explain fundraising success in token offerings. Adhami, Giudici, and Martinazzi (2018) show that project characteristics, source-code availability, presales, and token utility play a role in fundraising success. Campino, Brochado and Rosa (2022) additionally demonstrate that human capital, social media activity, cryptocurrency market conditions, and the project's proximity to markets with strong financial and human capital are associated with the success of the campaign. These studies show that investor decision-making in token sales depends on project-level, market-level, and communication-related signals.

ICO success is not consistently defined across the literature. Some studies use fundraising metrics like the amount raised, the achievement of the soft cap, or the amount raised as a proportion of the hard cap. For example, Roosenboom, van der Kolk, and de Jong (2020) employ various successful funding indicators such as soft-cap attainment, capital raised relative to the hard cap, and the logarithm of the amount raised. Furthermore, Xu (2021) considers soft-cap achievement to be a vital success factor. Other studies use market-related outcomes, such as token tradability, exchange listing, or continued trading after the offering. Amsden and Schweizer (2018) clearly define token or coin tradability as a key criterion when determining ICO success. Campino et al. (2022) also emphasize the significance of secondary-market trading for token projects. These definitions show that ICO success is frequently understood both as the ability to raise capital and as the ability to produce a tradable token after the campaign.

But there is a fundamental tension at the heart of this multiplicity of definitions. Fundraising success measures how well a project attracts investors during the token sale, but this does not mean that the token will be a good investment after listing. A successful fundraising campaign is therefore not

necessarily a successful investment. More specifically, Aslan, Şensoy, and Akdeniz (2023) draw attention to this difference when distinguishing ICO success from post-ICO performance. They demonstrate that ratings, projected length of sale, token-sale share, number of experts and team size are explanatory of fundraising outcomes, whereas offer price and market sentiment shape post-ICO performance over a longer time frame. Lyandres, Palazzo, and Rabetti (2022) also address both ICO financing success and post-ICO operating and financial performance, showing that early and longer-term outcomes need to be analysed separately.

Overall, the literature indicates that ICO performance is a multidimensional concept. Fundraising success is generally measured by the total amount raised or by the achievement of funding thresholds such as the soft cap or hard cap (Roosenboom et al., 2020; Xu, 2021). Market success is commonly translated through token tradability or exchange listing (Amsden and Schweizer, 2018; Campino et al., 2022). But those dimensions only tell part of the story, because they do not fully indicate whether the token creates value for investors after listing. For that reason, post-ICO financial performance is also measured through return-based indicators, such as post-listing returns or buy-and-hold returns (Aslan et al., 2023; Lyandres et al., 2022).

3.2. Whitepapers as disclosure documents under information asymmetry

Because token offerings often operate without traditional intermediaries and without traditional regulatory standards for disclosure, the whitepaper is central to the investment process (Zetzsche et al., 2017; Bourveau et al., 2022). It often serves as the main document in which a project presents its technological architecture, tokenomics, roadmap, business plan, team, and use of funds. But unlike IPO prospectuses, whitepapers are generally voluntary, unaudited, and weakly standardized. This makes them very important for investors, but also difficult to assess.

Research on voluntary information provision supports the importance of disclosure in ICO markets. Bourveau et al. (2022) show that voluntary disclosure and information intermediaries support the functioning of ICOs as an alternative capital market. Thus, in a context of weak regulation and high information asymmetry, disclosed project information plays a critical role in investor decision-making. This means that whitepapers function as key disclosure documents, even if their reliability is not guaranteed.

Empirical research also shows that whitepaper characteristics contain information related to ICO outcomes. Samieifar and Baur (2021) show that whitepaper length is positively related to the amount raised and to the chance of successfully completing the ICO. Thewissen et al. (2022) use topic modeling on 5,210 ICO whitepapers from offerings completed between August 2015 and June 2020. They identify 30 topic categories and show that the thematic content improves the identification of successful projects. Florysiak and Schandlbauer (2022) further support the idea that whitepapers contain informative signals by studying whitepaper informativeness and the role of ICO analysts.

However, the kind of value provided by whitepapers remains ambiguous. While their predictive value may measure the quality of actual projects, it may also uncover persuasive disclosure practices, promotional stories, or purely document sophistication (Bourveau et al., 2022; Florysiak and Schandlbauer, 2022; Thewissen et al., 2022). This difference is important since whitepapers can be especially useful at the fundraising stage, when investors rely heavily on project disclosure to make decisions. Once tokens are listed, their performance is more affected by execution, liquidity, adoption, exchange listing, and market sentiment (Thewissen et al., 2022; Aslan et al., 2023; Lyandres et al., 2022).

This provides a methodological opportunity for LLM-based analysis. Conventional NLP and machine learning models can learn structured features of the data from whitepapers, but large language models enable semantic content to be processed directly. One could therefore question to what extent large language models can analyse whitepapers in a way that produces meaningful predictions about the outcomes of ICO fundraising and future financial performance.

3.3. AI-based prediction in ICO markets and comparison with related studies

Recent work has explored whether artificial intelligence, machine learning, and natural language processing can improve the prediction of ICO outcomes from structured project data and unstructured textual information (Meoli and Vismara, 2022; Thewissen et al., 2022; Liu et al., 2025; Karimov and Wójcik, 2021). AI prediction is especially relevant for token-offering markets, where investors face uncertainty, weak standardization in project disclosures, and heterogeneous project details. AI models can handle large datasets, identify nonlinear relationships, and extract predictive signals from both structured and unstructured sources.

Meoli and Vismara (2022) make one of the closest contributions to the current research. Based on 383 ICOs launched between August 2014 and December 2019, they show that machine learning improves forecast accuracy compared with standard Logit models, increasing accuracy from 54.3% to 72.8%. They also highlight the important role played by the structure of ICO whitepapers in explaining success. Their contribution is valuable because it shows that algorithmic models can improve the prediction of fundraising success in token-offering markets.

But their study is focused mainly on the fundraising stage. It tells us whether a project is successful at raising capital relative to its funding objective. It does not look at whether the token remains traded or delivers positive returns after listing. In addition, their methodology is based on standard machine-learning approaches using constructed variables. In contrast, the current study compares large language models and assesses whether they can extract predictive signals directly from the content of the whitepaper.

Thewissen et al. (2022) also show that whitepaper text contains predictive information by using topic modeling to identify themes linked to successful projects. However, their approach remains topic-

based and focuses mainly on ICO success. Liu, Sheng, and Wang (2025) provide another important contribution by building a Tech Index from cryptocurrency whitepapers to capture technological sophistication. Their research indicates that technology-related whitepaper signals are associated with early success and higher first-day listing prices, while also suggesting that investors may overreact to technical signals, since high-Tech-Index projects tend to earn lower subsequent returns.

The focus of this study is different from Liu, Sheng, and Wang (2025). Their research views technological sophistication as a valuation driver. Rather than building a technology-based index, this study adopts a broader LLM-based prediction framework that analyses the whitepaper as a whole rather than isolating a single technological dimension.

Other papers apply the use of AI to similar but distinct prediction problems: Basu et al. (2021) investigate human decision-making in AI-augmented ICO evaluation, while Karimov and Wójcik (2021) use machine learning to detect scam ICOs based on ex-ante characteristics. Although these studies highlight the relevance of AI-based screening in ICO markets, they do not directly test whether large language models are able to predict fundraising success or future financial performance from whitepapers.

Overall, these studies show that AI-based methods can extract useful information from ICO-related data, including whitepaper structure, thematic content, technological signals, and fraud-related indicators (Meoli and Vismara, 2022; Thewissen et al., 2022; Liu et al., 2025; Karimov and Wójcik, 2021). Yet, most research has focused on traditional ML, topic modeling, technological sophistication, decision support, or fraud detection. There is a key tension here: predictive accuracy does not always imply economic explanation. This raises the possibility that AI models capture investor persuasion, attention, disclosure style, or technological narratives rather than real long-term project quality (Thewissen et al., 2022; Liu et al., 2025; Aslan et al., 2023). A model can identify which projects will attract capital without necessarily knowing which tokens will remain traded or generate value after listing. The question is not only whether AI can predict ICO success, but what kind of performance it can predict.

3.4. LLMs and fundraising prediction in entrepreneurial finance

The relevance of LLM-based prediction is supported by the broader literature on entrepreneurial finance. Token offerings are different from conventional venture capital and crowdfunding because they consist of blockchain-based assets and secondary-market trading. But they have a similar underlying problem as other kinds of entrepreneurial fundraising: investors must evaluate young start-ups in turbulent circumstances with scarce history and considerable asymmetry of information.

Machine learning has gained significant traction in venture capital, including startup screening and investment decisions. Ross, Das, Sciro, and Raza (2021) develop CapitalVX, a machine-learning model for predicting the outcomes of startups, i.e., IPO, acquisition, failure, remaining private, and

follow-on funding. This is significant for token offerings, where ICO investors face a similar screening challenge: they must evaluate early-stage entities whose long-term success is uncertain and whose level of disclosure available to outsiders is often restricted to public-facing disclosures.

Previous research has also shifted from conventional machine learning toward language-model- based prediction. Maarouf, Feuerriegel, and Pröllochs (2025) develop a fused large language model to predict startup success based on structured startup information and textual self-descriptions from Crunchbase profiles. The importance of their study lies in the fact that textual descriptions of innovation and business models contain predictive information, and that language models can be used as decision-support tools for investors. This provides a direct methodological rationale for applying large language models to ICO whitepapers, which are also investor-facing textual documents.

Crowdfunding work provides a second pertinent comparison. Like token offerings, public-facing documents and narratives are critical to drawing a wide investor base and this is where crowdfunding campaigns come to play. Kaminski and Hopp (2020) present a neural network and natural language processing method for predicting crowdfunding outcomes by text, speech, and video metadata. Elitzur et al. (2024) similarly use several machine-learning techniques, for example: boosted trees, random forests, shallow neural networks, and deep neural networks to predict crowdfunding success. These studies suggest that textual and narrative signals could be predictive for entrepreneurial fundraising contexts.

However, evidence from crowdfunding, venture capital, and startup prediction cannot be transferred mechanically to token offerings, because ICOs combine a fundraising stage with a secondary-market phase in which tokens can be traded and priced after listing (Aslan et al., 2023; Lyandres et al., 2022). This makes the prediction problem more complex. The model does not only need to assess whether investors will fund the project. It also needs to help evaluate whether the token will be traded and whether it will generate returns after listing.

The broader financial machine-learning literature also points to the premise that AI-based investment outcomes will likely be better based on selection. DeMiguel et al. (2023), for example, show that machine-learning approaches can be used to choose tradable long-only mutual fund portfolios with significant abnormal performance. But this relationship is not automatic. Dessain (2022) cautions that stronger return-prediction models do not always outperform buy-and-hold strategies once their economic value is rigorously evaluated. In addition, Li et al. (2023) demonstrate that more advanced language models like GPT-4 achieve superior results on financial text analysis tasks, a finding that supports the hypothesis that model sophistication may matter.

Collectively, these studies enhance the methodological foundations of the present study, but they also strongly emphasize the specificity of token offerings. If language models can use textual self-descriptions to predict startup success, large language models may also derive useful insights from

ICO whitepapers. At the same time, evidence from financial machine learning suggests that better prediction or better model sophistication does not automatically translate into superior investment performance.

3.5. Research gap and contribution

Therefore, these streams of research do not fully answer the question addressed in this thesis. Studies on token offerings identify significant factors that influence fundraising and post-listing outcomes, but they tend to rely more on structured project variables than on AI-driven interpretation of whitepaper content. Whitepaper research shows that textual disclosure matters, but it focuses mainly on early fundraising success or on specific textual aspects. AI research indicates that machine learning has predictive value, but it does not fully assess the value of large language models applied to whitepapers for predicting both fundraising success and future financial performance.

The closest papers help clarify this gap. Meoli and Vismara (2022) find that machine learning can forecast ICO success and that whitepaper-related data have predictive power, but their analysis is mainly based on fundraising success and uses traditional machine-learning methods. Thewissen et al. (2022) show that whitepaper topics improve the identification of successful projects, but also that whitepaper informativeness decreases after listing. Liu, Sheng, and Wang (2025) argue that technology-based whitepaper indicators matter for early and later market outcomes, but their analysis focuses on technological sophistication rather than on the broader predictive capability of large language models. Basu et al. (2021) and Karimov and Wójcik (2021) also show the usefulness of AI for ICO assessment and screening, but their dependent variables and research objectives differ from those of this paper.

The contribution of this study should therefore be understood as a refinement of the existing literature rather than as the first application of AI to ICO prediction. Previous studies have already shown that machine learning and natural language processing can extract predictive signals from ICO data. However, direct empirical evidence remains limited on how different large language models use ICO whitepapers to predict both fundraising success and subsequent financial performance.

In this study, these outcomes are measured through amount raised, token traded status, and buy- and-hold returns. This distinction is important because, even though these parameters are all related to ICO success, they do not capture the same dimension of performance. The amount raised shows how well the project was funded during the token sale. Token traded status indicates whether the project reaches a minimum level of market viability after the offering. Buy-and-hold returns indicate the financial performance experienced by investors after the token is listed. A whitepaper could be attractive enough to convince investors without generating strong post-listing returns. On the other hand, some projects might raise lower levels of capital but perform better later on.

By separating these outcomes, this study assesses whether large language models capture short- term

investor persuasion, market viability, long-term financial value, or a combination of these dimensions. Empirically, it adds two elements to the existing literature. First, it moves beyond traditional machine-learning and topic-modeling approaches by evaluating general-purpose large language models directly on ICO whitepapers. Second, it distinguishes between different dimensions of ICO performance, assessing whether signals extracted from whitepapers are mainly related to fundraising attractiveness, post-offering market viability, or long-term token returns.

3.6. Hypothesis

Based on the theoretical framework and prior literature review, the hypothesis section is structured around three main blocks regarding the relationship between artificial intelligence and ICO outcomes. The first block concerns AI-based success predictability and is divided into two sub- hypotheses, because ICO success is measured in this study through both fundraising success and token trading success. The second block concerns AI-based selection and buy-and-hold returns.

The third block concerns the relationship between model performance and investment returns.

Prior literature suggests that ICO outcomes are at least partly predictable from available information. Meoli and Vismara (2022) show that machine-learning techniques can forecast ICO success and that whitepaper-related information has predictive value. Aslan et al. (2023) further show that ICO success and post-ICO performance are linked to observable characteristics such as token-sale design, team structure, ratings, and market-related factors.

H1a: AI-based prediction and token trading success

The first dimension of ICO success is token trading success. Several studies measure ICO success through token tradability, exchange listing, or continued trading after the offering. A project may raise capital during the ICO, but this does not necessarily mean that its token will become tradable or remain active on the market. If whitepapers contain signals about project credibility, technical feasibility, token design, or market potential, AI models may also be able to predict whether the token will be traded.

H0: AI-based predictions do not significantly predict token trading success, measured by token traded status.

H1a: AI-based predictions significantly predict token trading success, measured by token traded status.

H1b: AI-based prediction and fundraising success

The second dimension of ICO success is fundraising success. In the literature, fundraising success is often measured through the amount raised, the achievement of the soft cap, or the amount raised relative to the hard cap. Since whitepapers contain information that investors use during the fundraising stage, AI models may be able to extract signals that are associated with the amount raised.

H0: AI-based predictions do not significantly predict ICO fundraising success, measured by the amount raised.

H1b: AI-based predictions significantly predict ICO fundraising success, measured by the amount raised.

H2: AI-based selection and buy-and-hold returns

Second, the literature also supports the hypothesis that AI-based selection may lead to superior investment outcomes. Although this evidence is stronger in the broader financial literature than in the ICO literature specifically, prior studies show that machine-learning methods can identify portfolios with significant out-of-sample abnormal performance. DeMiguel et al. (2023), for example, show that machine-learning techniques can help select tradable long-only mutual fund portfolios that earn significant annual alpha net of costs. This supports the logic behind the second hypothesis: if AI improves asset selection by extracting useful predictive signals, it may also improve realized buy-and-hold returns compared to the benchmark portfolio.

H0: AI-based selection of ICOs does not lead to a significantly higher Buy-and-Hold Return (BHR) compared to the benchmark portfolio.

H2: AI-based selection of ICOs leads to a significantly higher Buy-and-Hold Return (BHR) compared to the benchmark portfolio.

H3: Model performance and investment return

Third, the relationship between model performance and investment returns appears theoretically plausible but should be approached with caution. More advanced language models may be better able to process complex financial and technical documents. Li et al. (2023), for example, show that GPT-4 performs better than earlier ChatGPT models across several financial text-analysis tasks. This suggests that model performance may matter when extracting information from complex financial texts.

However, even better predictive performance does not automatically mean better investment results. Dessain (2022) demonstrates that common performance metrics used to evaluate return-prediction models can be misleading and that the economic value of such models must be assessed carefully. Hence, whether better-performing AI models generate better investment outcomes remains an empirical question.

H0: Better-performing AI models do not generate significantly better investment results than lower-performing AI models.

H3: Better-performing AI models generate significantly better investment results than lower-performing AI models.

4. Methodology

4.1 Data and sample construction

The empirical part is based on an original dataset combining ICO-level information, whitepaper data, AI-generated predictions, and market outcome variables. Each observation is linked to an Initial Coin Offering where a whitepaper was accessible in plain-text format. This is a condition needed since the analysis depends on large language models to process the content of whitepapers.

After cleaning and matching the whitepapers with ICO-level and market outcome data, the final classification sample contains 1,202 ICOs. The four models are evaluated on this same sample: ChatGPT-5.4, ChatGPT-5.4 Mini, ChatGPT-5.4 Nano, and Mistral Medium 3. The effective number of observations is lower in some regression models because observations with missing values in at least one variable included in the specification are automatically excluded.

The analysis focuses on three main outcome variables. First, `TokenTradedHistorical` is a binary variable equal to 1 if the token issued during the ICO was subsequently listed and traded on an exchange, and 0 otherwise. This variable captures post-offering market viability. Second, `AmountRaised` measures fundraising success. Since ICO fundraising amounts are highly skewed, the main specification uses the logarithmic transformation of this variable, labelled `log_AmountRaised`. Third, `AbnormalBHR` measures abnormal buy-and-hold returns and is used to assess future financial performance. Following prior research, BHR is computed over a 180-day period after token issuance. Since the return analysis requires market price data, the effective sample size for this part of the analysis is lower than for the classification and fundraising tests.

The buy-and-hold return can be expressed as:

$$BHR_{i,180} = \frac{P_{i,180} - P_{i,0}}{P_{i,0}}$$

where $P_{i,0}$ is the initial token price and $P_{i,180}$ is the token price 180 days after token issuance. In the empirical analysis, the dependent variable used for financial performance is *AbnormalBHR*, which corresponds to the abnormal 180-day buy-and-hold return available in the dataset.

The dataset also includes several control variables describing each ICO at the time of the offering. These include whitepaper-related variables, such as *WP_Readability*, *WP_Sentiment*, and *WP_Pages*; project-level variables, such as *HardCap*, *SoftCap*, *PreICO*, *MinInvest*, *NumbCurr*, *Fiat*, *TokenDist*, and *Bonus*; legal and geographical variables, such as *TaxHaven* and *USRestrict*; quality and visibility proxies, such as *Rating*, *SocialMedia*, *Video*, and *Team*; and contextual variables, such as *Year*, *Eth*, *WhitelistKYC*, and *ESG* where available.

4.2. Whitepaper analysis pipeline

4.2.1 Model selection

The study uses four large language models to generate predictions from ICO whitepapers. Three of them are ChatGPT models: *ChatGPT-5.4*, *ChatGPT-5.4 Mini*, and *ChatGPT-5.4 Nano*. These models represent different cost and capacity levels within the same model family. ChatGPT-5.4 is treated as the highest-capacity and most expensive ChatGPT model in the analysis, while ChatGPT-5.4 Mini and ChatGPT-5.4 Nano represent lower-cost and lower-capacity alternatives.

The fourth model is *Mistral Medium 3*. This AI tool is not a GPT model and does not belong to OpenAI's model family. It is included as an external LLM benchmark in order to compare the ChatGPT models with a language model from another provider. This design makes it possible to compare performance within one model family, while also testing whether the observed predictive patterns extend to a non-OpenAI model.

For readability, the empirical tables refer to Mistral Medium 3 simply as *Mistral*.

4.2.2 Prompt standardisation

To ensure comparability across models and observations, the same prompt was applied to every whitepaper for each model. The prompt instructed the model to act as an expert in blockchain fundraising and ICO analysis. Each whitepaper was evaluated according to seven criteria: the clarity and robustness of the value proposition, team credibility, tokenomics, market size and realism, roadmap and technical feasibility, competitive differentiation, and the overall writing quality and professionalism of the document.

Each model was required to return a structured output containing three elements: a binary prediction, a

confidence score, and a short justification. The binary prediction takes the value 1 when the model predicts that the ICO will be successful, and 0 otherwise. The confidence score ranges from 0 to 1. The textual justification is kept for interpretation, but the empirical analysis mainly relies on the binary prediction.

For each ICO, the binary prediction generated by each model is stored as a separate variable:

ChatGPT, *GPT_Mini*, *GPT_Nano*, and *Mistral*.

The same binary success prediction is then used across the three empirical tests. This allows the analysis to examine whether a general AI assessment of ICO success is associated with fundraising outcomes, token traded status, and post-listing financial performance.

4.2.3 Data collection protocol

All whitepapers were converted into plain-text files and matched with the corresponding ICO-level observations using the available project identifier. The same set of 1,202 whitepapers was run four times: once by ChatGPT-5.4, once by ChatGPT-5.4 Mini, once by ChatGPT-5.4 Nano, and once by Mistral Medium 3.

For each model, the prediction, confidence score, reasoning text, and potential error flag were added to the master dataset. When a response was missing, malformed, or could not be parsed, the observation was flagged and excluded from the relevant analysis.

The pipeline was implemented in Python for text processing, API calls, and data manipulation. The final dataset was then exported for statistical analysis in R.

4.3. Evaluation of classification performance

The first step of the empirical analysis evaluates the raw classification performance of each model. For the binary success outcome, each model's prediction is compared with the observed value of `TokenTradedHistorical`.

The confusion matrix is composed of four elements:

- True positives (TP): ICOs predicted as successful and actually traded.
- True negatives (TN): ICOs predicted as failures and not traded.
- False positives (FP): ICOs predicted as successful but not traded.
- False negatives (FN): ICOs predicted as failures but actually traded.

Based on these values, the following metrics are computed:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Accuracy measures the overall proportion of correct classifications in the predictions. However, because the sample is imbalanced and failures represent the majority of observations, accuracy alone may be misleading: a model can obtain a relatively high accuracy simply by correctly predicting many failures. Precision and recall therefore provide a more informative view of the model's ability to identify successful ICOs. Precision measures the reliability of positive predictions, while recall measures the model's ability to detect actual successful ICOs. The F1- score combines precision and recall into a single metric, which is useful for comparing models when the positive class is less frequent.

4.4. Token trading success regressions

To test whether AI-based predictions are associated with post-offering market viability, the first set of regressions uses *TokenTradedHistorical* as the dependent variable. Since this variable is binary, logistic regressions are estimated.

The model is specified as follows:

$$P(\text{TokenTradedHistorical}_i = 1) = \Lambda(\alpha + \beta \text{PredictionAI}_i + \gamma X_i)$$

where Λ denotes the logistic function. *PredictionAI_i* is again replaced successively by the prediction generated by each model, and *X_i* represents the vector of available control variables included in each specification.

A positive and statistically significant coefficient means the model's prediction is significantly related to the likelihood that the token will be traded post-ICO. This specification tests whether the AI prediction adds information beyond observable project-level and whitepaper-related characteristics.

Model fit is assessed using the Akaike Information Criterion (AIC) and McFadden's pseudo-*R*². The AIC is defined as:

$$AIC = 2k - 2\ln(L)$$

where *k* is the number of estimated parameters and *L* is the likelihood of the model. The AIC is interpreted only in relative terms: a lower AIC indicates a better balance between model fit and model complexity when comparing models estimated on the same dependent variable.

McFadden's pseudo-*R*² is computed as:

$$Pseudo - R^2 = 1 - \frac{Residual\ deviance}{Null\ deviance}$$

4.5. Fundraising success regressions

To test whether AI-based predictions are associated with fundraising success, the second set of regressions uses $\log_AmountRaised$ as the dependent variable. The logarithmic transformation is used because the distribution of ICO fundraising amounts is highly skewed.

The model is specified as follows:

$$\log(AmountRaised_i) = \alpha + \beta PredictionAI_i + \gamma X_i + \varepsilon_i$$

where $PredictionAI_i$ corresponds successively to the binary prediction generated by ChatGPT-5.4, ChatGPT-5.4 Mini, ChatGPT-5.4 Nano, and Mistral Medium 3. X_i represents the vector of available control variables included in each specification.

A positive and statistically significant β coefficient shows that ICOs predicted as successful by the model tend to raise more capital than ICOs predicted as unsuccessful, once controlling for observable ICO characteristics. This specification tests whether AI language models extract information from whitepapers that is relevant for fundraising success.

4.6. Buy-and-hold return regressions

The third set of regressions examines whether AI-based predictions are associated with future financial performance. The dependent variable is $AbnormalBHR$, which measures abnormal buy- and-hold returns over a 180-day period after token issuance. This variable is used to assess whether ICOs predicted as successful by the models also generate superior post-issuance financial performance.

The model is specified as follows:

$$AbnormalBHR_{i,180} = \alpha + \beta PredictionAI_i + \gamma X_i + \varepsilon_i$$

where $AbnormalBHR_{i,180}$ represents the abnormal 180-day buy-and-hold return of token i ,

$PredictionAI_i$ represents the binary prediction generated by each model, and X_i represents the vector of available control variables included in each specification.

A positive and statistically significant β coefficient would indicate that ICOs predicted as successful by the model generate higher abnormal buy-and-hold returns after token issuance. This test is important because statistical prediction and economic value are not necessarily the same. A model may predict whether an ICO reaches a certain success threshold without necessarily identifying tokens that generate superior returns for investors.

4.7. AI-selected portfolios and return comparison

In addition to the return regressions, the study constructs AI-selected portfolios. For each model, all ICOs predicted as successful are grouped into a model-specific portfolio. This produces four AI-selected portfolios: ChatGPT-5.4, ChatGPT-5.4 Mini, ChatGPT-5.4 Nano, and Mistral Medium 3.

The benchmark portfolio consists of ICOs that were actually successful according to token traded status and for which BHR data are available. This benchmark represents ICOs that reached a minimum level of market viability and for which post-issuance financial performance can be measured.

The return difference for each model is calculated as follows:

$$\Delta BHR_m = \underline{BHR}_{AI,m,180} - \underline{BHR}_{Benchmark,180}$$

where m represents each model, $\underline{BHR}_{AI,m,180}$ is the average 180-day BHR of the portfolio selected by model m , and $\underline{BHR}_{Benchmark,180}$ is the average 180-day BHR of the benchmark portfolio.

Welch two-sample t-tests are applied to determine whether the mean BHR of each AI-selected portfolio differs significantly from the benchmark portfolio. The test is run because portfolio sizes and variances differ across models. The Welch t-statistic is defined as follows:

$$t = \frac{\underline{BHR}_{AI,m,180} - \underline{BHR}_{Benchmark,180}}{\sqrt{\frac{s_{AI,m}^2}{n_{AI,m}} + \frac{s_{Benchmark}^2}{n_{Benchmark}}}}$$

where s^2 denotes the sample variance and n the number of observations in each portfolio. This analysis tests whether AI-based selection is associated with superior realized returns.

However, the benchmark must be interpreted carefully because it is based on an ex-post definition of success. It is therefore not a fully realistic investment benchmark available to investors at the ICO stage.

4.8. Model comparison

The final step of the empirical analysis compares the performance of the four models across the different tests.

First, classification performance is compared using accuracy, precision, recall, F1-score, and confusion matrices. Second, the token trading regressions are compared using AIC and pseudo- R^2 . Third, the fundraising regressions are compared using adjusted R^2 and the statistical significance of the AI prediction coefficient. Fourth, investment performance is compared using the average BHR of the AI-selected portfolios and the Welch t-tests against the benchmark.

This enables the analysis to determine whether either the larger or more sophisticated models lead to better empirical results. It also helps the analysis to examine whether better classification performance translates into better investment outcomes.

Since Mistral Medium 3 is not part of the ChatGPT model family, differences between Mistral Medium 3 and the ChatGPT models cannot be interpreted only as differences in model size.

Performance differences may therefore reflect model capacity, but also differences in training data, architecture, or model design.

4.9. Methodological limitations

Several methodological limitations should be acknowledged, as they affect the scope and reliability of the analysis.

First, the results are sensitive to prompt design. The predictions produced by the models depend on the wording in the prompts, the criteria in the evaluation, and the required output format. Thus, a different prompt structure can result in different classifications and may change comparative performance of the models.

Second, the models operate largely as black boxes. Although the reasoning field provides a short textual justification for each prediction, this explanation should not be interpreted as a complete or verifiable account of the model's decision process. It is therefore not possible to determine with certainty which specific elements of the whitepaper drove the classification.

Third, there is a potential risk of data leakage. Since several ICOs in the sample occurred before the training cutoffs of the models, the models may have been exposed during training to information about some of the projects, their subsequent development, or their market performance. This could artificially increase the apparent predictive accuracy of the models, especially for better-known ICOs.

Fourth, the sample is temporally concentrated in the 2017–2018 ICO boom and its immediate aftermath. This period was characterized by exceptional market exuberance, limited regulatory oversight, and retail investor behavior that may not be representative of current or future token markets. The findings may therefore not generalize to more recent forms of token financing, such as

Security Token Offerings, Initial Exchange Offerings, or decentralized token launches, which operate under different institutional and regulatory conditions.

Fifth, none of the AI models used in this study were fine-tuned on financial or ICO-specific data. The predictions generated by ChatGPT-5.4, GPT-Mini, GPT-Nano, and Mistral reflect the general language understanding of each model rather than domain-adapted financial judgement. It is therefore plausible that models fine-tuned on a labelled corpus of ICO whitepapers with known outcomes could outperform the general-purpose models evaluated here, both in terms of classification accuracy and financial relevance.

Finally, the analysis treats every model's prediction as a binary variable, hence discarding potentially important insights about confidence or level of conviction. A model tentatively categorizing a project as a likely success may not provide the same signal as a model that does so with high confidence, but this distinction is lost in the present coding scheme. Future studies with probabilistic outputs or confidence scores, where available, could take advantage of this extra dimension of the AI signal.

5. Results Analysis

5.1. Can AI Classify ICO Success?

In this section we investigate whether AI language models can classify ICOs as likely successes or failures based only on the content from the white papers. In this case, an ICO is considered successful if the token was traded on exchange platforms after the fundraising. Three dimensions of predictive performance are evaluated: raw classification accuracy, precision-recall trade-offs, and the statistical contribution of AI predictions within a multivariate logistic regression framework.

5.1.1 Overall Accuracy: A Misleading Metric

Overall classification may seem promising. ChatGPT-5.4 correctly classifies 74.7% of ICOs; GPT-Mini 71.6%, GPT-Nano 68.5%, Mistral 74.8%. However, this indicator is misleading here because of the imbalance in the dataset: out of 1,202 observations, only 388 ICOs (32%) are actual successes, meaning 68% are failures. The full tables are available in Appendix A.

Under these conditions, a trivial classifier predicting failure for every case would already reach 68% accuracy without any analytical value. The marginal gains observed across all models cannot be interpreted as evidence of strong predictive power. A better assessment requires decomposing accuracy into its components through the confusion matrix framework, as presented in Table 1 below.

Table 1 — Confusion matrices across all models (actual vs. predicted classifications)

Model	TN (Pred. Fail / Act. Fail)	FP (Pred. Success / Act. Fail)	FN (Pred. Fail / Act. Success)	TP (Pred. Success / Act. Success)	Total N
-------	-----------------------------	--------------------------------	--------------------------------	-----------------------------------	---------

ChatGPT-5.4	778	36	268	120	1,202
GPT-Mini	803	11	330	58	1,202
GPT-Nano	810	4	375	13	1,202
Mistral	711	103	198	190	1,202

Note: TN = True Negative, FP = False Positive, FN = False Negative, TP = True Positive. Success is defined as TokenTradedHistorical = 1.

These matrices reveal the root cause of the misleading accuracy figures: the four models over-predict failures. GPT-Nano, for instance, predicts success for only 17 ICOs in total, correctly identifying a mere 13 actual successes while missing 375. This extreme conservatism inflates accuracy by correctly classifying the abundant failure class, while generating almost no information about the success class. Mistral on the other hand, is more permissive in its decision making, predicting 293 successes and correctly identifying 190 of them.

5.1.2 Recall and Precision: The Core Performance Trade-Off

From an investor's perspective, the critical question is not how often the model is globally correct, but rather how often it successfully identifies ICOs that will genuinely succeed. This corresponds to the recall metric defined as the proportion of actual successes that the model correctly identifies. Alongside recall, precision captures the reliability of positive predictions: among predicted success, which percentage truly succeeds. Table 2 summarizes the key classification metrics across the four models.

Table 2 — Classification performance metrics across models (binary ICO success prediction)

Model	Overall Accuracy	Recall (Sensitivity)	Precision	Predicted Successes	F1-Score (approx.)
Mistral	74.8%	49.0%	64.8%	293	0.558
ChatGPT-5.4	74.7%	30.9%	76.9%	156	0.441
GPT-Mini	71.6%	14.9%	84.1%	69	0.253
GPT-Nano	68.5%	3.4%	76.5%	17	0.065

Note: Recall = $TP / (TP + FN)$; Precision = $TP / (TP + FP)$. F1-Score approximated as harmonic mean of recall and precision. All metrics computed on the full sample ($N \approx 1,202$).

The results show a large tradeoff between recall and precision amongst models. Mistral achieves the highest recall overall, correctly identifying 190 out of the 388 actual successes by a huge margin (49.0%). It is thus the most useful model for an investor looking at broad coverage of the opportunity set. But at the cost of precision: 103 of the 293 predicted successes are false positives, 35.2%, so Mistral's portfolio would include a sizable share of failed projects.

ChatGPT-5.4 is intermediate, identifying 120 successes (30.9% recall) with high precision (76.9%). In smaller models, the GPT-Mini predicts just 69 successes at 84.1% precision, and GPT-Nano limits itself to 17 predictions, of which 13 are correct (76.5% precision). Their recall figures are almost nothing, 14.9% and 3.4%, respectively, indicating that these models are very conservative and usually miss practically all possibilities to obtain a return on investment.

A key observation is that the model size, considering the hierarchy GPT-5.4 > GPT-Mini > GPT-Nano, does not linearly translate into better classification performance for the positive class. Mistral, which is architecturally distinct from the GPT models, outperforms all three GPT variants on recall despite having a lower AIC-based goodness-of-fit rank for ChatGPT (discussed in Section 1.3).

5.1.3 Logistic Regression: Does AI Add Independent Predictive Information?

Beyond classification metrics, the main empirical question of this research is whether AI predictions contain information about ICO outcomes that is not already captured by the observable characteristics. The AI prediction variable is statistically significant at the 0.1% threshold in both the ChatGPT-5.4 model ($\beta = 2.030$, $z = 8.863$, $p < 2 \times 10^{-16}$) and the Mistral model ($\beta = 1.683$, $z = 10.133$, $p < 2 \times 10^{-16}$). GPT-Mini ($\beta = 2.339$, $p < 0.001$) and GPT-Nano ($\beta = 1.492$, $p < 0.05$) are similarly significant. These results establish that AI adds genuine incremental predictive power rather than

simply proxying for variables already present in the model. The full tables are available in Appendix B.

Among the control variables, Rating, Year, and TaxHaven emerge as consistently significant across all specifications, consistent with the existing literature. It is also noticeable that whitepaper structural characteristics (WP_Pages, WP_Readability) are not significant, implying that the raw format of the document carries little information beyond its content.

To compare model fit across specifications, the Akaike Information Criterion (AIC) is employed.

Table 3 presents the AIC values and shows differences among models. The full tables are available in Appendix C.

Table 3 — Model fit comparison: AIC values and differences across AI specifications

Model	AIC	Δ AIC vs. Mistral	Pseudo-R ²
Mistral	1,224.4	—	—
ChatGPT-5.4	1,245.7	+21.3	0.176
GPT-Mini	1,286.5	+62.1	0.148
GPT-Nano	1,330.3	+105.9	0.118

Note: Δ AIC computed relative to the best-fitting model (Mistral). Pseudo-R² approximated using the McFadden formula: $1 - (\text{Residual deviance} / \text{Null deviance})$.

Mistral has the lowest AIC (1,224.4), indicating the best balance between goodness-of-fit and parsimony among the four specifications. The differences relative to Mistral are substantial: +21.3 for ChatGPT-5.4, +62.1 for GPT-Mini, and +105.9 for GPT-Nano. According to Burnham and Anderson (2002), a difference in AIC greater than 10 constitutes strong evidence against the model with the higher AIC. All three GPT variants therefore fall significantly short of Mistral.

However, it is important to note that AIC does not directly reflect the financial utility of a model's predictions. As will be demonstrated in Section 2, a model that achieves the best statistical fit does not necessarily generate the highest investment returns.

5.2. Fundraising Signal: Does AI Prediction Explain the Amount Raised?

This section looks into the total capital raised during the ICO (AmountRaised) and asks whether AI model predictions carry an independent signal about fundraising volumes.

5.2.1 OLS Regression Results: $\log(\text{AmountRaised})$

Four OLS regressions are estimated, one per AI model, Table 4 presents the core results for the AI

prediction variable across all four specifications. The full tables are available in Appendix D.

Table 4 — OLS regression results: effect of AI prediction on $\log(\text{AmountRaised})$

Model	β (AI var.)	Std. Err.	t-value	p-value	Sig.	Implied Premium ($e^{\beta} - 1$)
ChatGPT-5.4	1.275	0.152	8.401	$< 2 \times 10^{-16}$	***	+258% vs. rejected
GPT-Mini	1.277	0.219	5.837	6.87×10^{-9}	***	+259%
GPT-Nano	1.197	0.427	2.805	0.005	**	+231%
Mistral	0.987	0.121	8.182	7.13×10^{-16}	***	+168%

Note: $N = 1,202$. Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Implied premium computed as $(\exp(\beta) - 1) \times 100\%$, representing the percentage difference in AmountRaised between AI-predicted successes and failures, holding all other variables constant. Full regression outputs including all controls available in the appendix.

The four AI predictors are significant at the 0.01 level, t-statistics ranging from 2.805 (GPT-Nano, $p = 0.005$) to 8.401 (ChatGPT-5.4, $p < 2 \times 10^{-16}$). Premia fundraising-implicit, calculated as $(e^{\beta} - 1)$, are economically meaningful on the whole, namely ICOs positively rated by ChatGPT-5.4 pulled in around 258% more than rival badly assessed ICOs; GPT-Mini-predicted gains brought in 259% more; GPT-Nano 231% more; and Mistral 168% more. These differences correspond to a near tripling in fundraising volumes for AI selected projects compared to the rest, provided that the properties are all observable.

And it is also noticeable how the standard errors for GPT-Nano (0.427) are greater than those for ChatGPT-5.4 (0.152) and Mistral (0.121), attributed to the relatively few successes issued by GPT-Nano (17 ICOs vs 156 ChatGPT-5.4 and 293 for Mistral). While the coefficient remains significant, the broad confidence interval also implies uncertainty in estimates and cautions against treating the GPT-Nano premium as a precisely observed one.

5.2.2 Control Variable Analysis

The log specification reveals a richer and more coherent pattern of control variable significance than the level specification. Table 5 summarizes the significant control variables across all four models, with coefficient signs and significance levels.

Table 5 — Significant control variables in $\log(\text{AmountRaised})$ specifications (all four models)

Control Variable	ChatGPT-5.4	GPT-Mini	GPT-Nano	Mistral
Year	*** ($\beta=-0.381$)	*** ($\beta=-0.443$)	*** ($\beta=-0.485$)	*** ($\beta=-0.391$)
Rating	*** ($\beta=+0.246$)	*** ($\beta=+0.292$)	*** ($\beta=+0.343$)	** ($\beta=+0.238$)
Team	*** ($\beta=+0.210$)	*** ($\beta=+0.220$)	*** ($\beta=+0.243$)	*** ($\beta=+0.220$)
TaxHaven	** ($\beta=+0.154$)	** ($\beta=+0.164$)	** ($\beta=+0.167$)	* ($\beta=+0.127$)
SocialMedia	** ($\beta=-0.186$)	** ($\beta=-0.216$)	*** ($\beta=-0.239$)	** ($\beta=-0.214$)
WP_Readability	** ($\beta=+0.134$)	** ($\beta=+0.133$)	** ($\beta=+0.134$)	* ($\beta=+0.125$)
Fiat	* ($\beta=+0.300$)	* ($\beta=+0.301$)	* ($\beta=+0.288$)	** ($\beta=+0.340$)
WhitelistKYC	* ($\beta=+0.129$)	* ($\beta=+0.141$)	* ($\beta=+0.140$)	* ($\beta=+0.132$)

Note: Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Coefficients (β) represent log-point effects, interpretable as approximate percentage changes in *AmountRaised* for a unit increase in the variable (or for binary variables, the proportional difference between categories). Direction of effect noted by sign of β .

Eight control variables emerge as consistently significant across all specifications, providing a coherent and theoretically interpretable picture of the determinants of ICO fundraising success.

Among those variables, only Year and Social Media Presence have a negative impact on the raised amount. (β ranging respectively from -0.381 to -0.485 and -0.186 to -0.239).

All the other relevant variables: Rating, Team, TaxHaven, White Paper Readability, Fiat, and Whitelist KYC all have a significant and positive impact on the funding

5.2.3 Interpreting the AI–Fundraising Relationship

The consistently large and statistically significant fundraising premium associated with AI-positive assessment raise a fundamental question: do these estimates reflect a genuine causal effect of whitepaper quality on investor behavior, or is it that better-resourced issuers produce both higher-quality whitepapers and raise more capital through channels unrelated to the whitepaper itself?

Some properties of the results also provide evidence for causal quality-signal interpretation. First, the AI prediction variable remains significant after conditioning on Rating, indicating that the AI captures content that the rating does not fully reflect. Second, the significant White Paper Readability coefficient provides independent evidence that the textual qualities of the whitepapers predict fundraising outcomes independently. Third, the significance of the AI coefficient is invariant across all four model architectures, which would not exist if the variable were just proxying a single omitted

feature associated with issuer resources.

However, documentation endogeneity is not excluded on the basis of the available evidence alone. A purely persuasive causal identification strategy would have to incorporate an instrumental variable that impacts whitepaper quality apart from issuer financial readiness which is not the focus of this analysis in the present study. Results should be thus interpreted as establishing a robust and economically meaningful conditional association between AI-assessed whitepaper quality and fundraising outcomes, with the caveat that the specific causal mechanism is an open empirical question.

5.3. Financial Signal: Does AI Prediction Translate into Returns?

Showing AI models can classify ICO outcomes with statistical significance is necessary but not enough for professional investment. The most relevant point is to measure if AI-selected portfolios generate superior financial returns than a benchmark. This section addresses that question by analysing the Abnormal Buy-and-Hold Return (BHR) after 6 months, of portfolios constructed on the basis of each model's positive predictions, benchmarked against a portfolio of all ICOs that were successful. Three analytical steps are followed: a comparison of portfolios' descriptive statistics, a formal two-sample t-test framework, and a multivariable OLS regression to isolate the marginal contribution of AI models.

5.3.1 Portfolio Returns and Statistical Tests: An Integrated Overview

The BHR measure captures the cumulative holding return of a token from its first trading day, offering a long-horizon performance metric that reflects the full post-ICO price trajectory. A portfolio is created for each AI model by including all ICOs that the model predicted as successes, regardless of whether they were actually successful. An important caveat must be noted at the outset: the return analysis is conducted on a smaller effective sample than the classification analysis, because post-listing return data are not available for all 1,202 ICOs. The benchmark portfolio consists of actual token-trading successes, while each AI-selected portfolio consists of ICOs predicted as successful by the corresponding model. Welch two-sample t-tests are performed to formally examine whether each AI portfolio outperforms the benchmark. Table 6 combines the descriptive portfolio statistics with the inferential t-test results to provide a concise and structured summary of each model's economic magnitude as well as statistical reliability.

Portfolio	N	Mean BHR	Std. Dev.	Interpretation	Sig. vs. Benchmark
Real Successes (benchmark)	388	−0.345	2.865	Baseline portfolio	—
ChatGPT-5.4	156	−0.289	1.864	Marginally better	No (p = 0.819)
GPT-Mini	69	+0.093	1.826	Positive return	No (p = 0.150)
GPT-Nano	17	+1.249	2.695	Strong outperformance	Yes* (p = 0.057)
Mistral	293	−0.266	2.453	Marginally better	No (p = 0.756)

*Note: BHR = Buy-and-Hold Return from first trading day; expressed as raw returns (e.g., +1.249 = +124.9%). Benchmark BHR = mean BHR of all TokenTradedHistorical = 1 observations (N = 388). Welch two-sample t-tests; H_0 : mean BHR of AI portfolio = mean BHR of benchmark. * Significant at the 10% level. Spread = Mean BHR (AI) – Benchmark BHR.*

The first observation is that even the benchmark portfolio of verified successes yields a negative mean BHR of −0.345, implying an average holding underperformance of approximately 34.5%.

Against this backdrop, each AI portfolio produces a mean BHR that exceeds the benchmark, in two cases by a substantial margin. ChatGPT-5.4 yields a mean BHR of −0.289, a marginal improvement of 5.6 percentage points over the benchmark that is economically small and statistically negligible (p = 0.819). Mistral produces a slight improvement of 7.9 percentage points (mean BHR = −0.266, p = 0.756). GPT-Mini generates a better return than the benchmark (+0.093), representing a spread of 43.8 percentage points over it, but is not significant, as its p-value is 0.15.

GPT-Nano stands apart from all other models. Its 17-ICO portfolio achieves a mean BHR of +1.249 and is the only model that outperformed the benchmark with statistical significance (p = 0.057, significant at the 10% level). This result is remarkable not only in its magnitude but in its origin: GPT-Nano is the weakest classifier by every conventional statistical metric, yet it delivers the strongest financial outcome. The confidence interval for this comparison [−0.058, +3.247] is wide, which can be attributed to the small sample size; moreover, certain individual outliers in the ICOs may have a disproportionate influence on the mean. Despite these limitations, the consistency of this result gives it credibility that goes further than a simple sampling artifact.

5.3.2 OLS Regression on BHR: Controlling for ICO-Level Covariates

The portfolio comparison above does not control for the possibility that systematic differences in ICO characteristics drive the observed return differentials. To address this, regressions are estimated with AbnormalBHR as the dependent variable. Table 7 presents the coefficient, standard error, t- value, p- value, and adjusted R² of each specification. The full tables are available in Appendix E.

Table 7 — OLS regression results: coefficient on AI prediction variable (dependent variable: AbnormalBHR)

Model	Coefficient (β)	Std. Error	t-value	p-value	Adj. R ²
ChatGPT-5.4	0.151	0.389	0.387	0.699	−1.24%
GPT-Mini	0.586	0.484	1.211	0.227	−0.72%
GPT-Nano	1.910	0.879	2.172	0.031	~0.00%
Mistral	0.014	0.376	0.037	0.971	−1.30%

Note: $N = 277$ observations with available AbnormalBHR data and non-missing control variables across all specifications. Only the AI variable coefficient is reported; full regression outputs available in the appendix.

Adjusted R² of the full model reported. * Significant at the 5% level.

The results of the regression confirm the portfolio-level results while providing some nuance. ChatGPT-5.4 ($\beta = 0.151$, $p = 0.699$), Mistral ($\beta = 0.014$, $p = 0.971$), and GPT-Mini ($\beta = 0.586$, $p = 0.227$) all produce coefficients that are statistically indistinguishable from zero after employing ICO-level controls. This suggests that whatever return advantage these models seem to provide at the portfolio level is largely reflected in the properties of the ICOs that they choose, not the signal from the AI. The adjusted R² values for these three specifications also fall slightly negative, indicating that the models contribute no meaningful explanation of BHR beyond the control variables. GPT-Nano again is the only decisive exception. The coefficient is positive ($\beta = 1.910$), economically significant at the 5% level ($t = 2.172$, $p = 0.031$) even after adjusting for the entire set of observable ICO attributes. This provides further support for an analysis that GPT-Nano's selection captures a real, non-redundant signal about future financial performance. Importantly, HardCap was the only control variable to achieve consistent significance across the four specifications ($\beta \approx 1.73 - 1.76$, $p \approx 0.025 - 0.027$), indicating ICOs with credible fundraising targets independently attract higher long-run token returns.

5.3.3 The Paradox of Model Size: A Non-Linear Relationship

Together, the portfolio and regression results expose a counterintuitive pattern: the relationship between model size and financial performance is non-linear, and reverses direction relative to the classification performance rankings established in Section 5.1.

As said section, model size is positively correlated with statistical fit: ChatGPT-5.4 achieves the best AIC among GPT variants, Mistral outperforms all on recall and overall fit, and GPT-Nano ranks last on all of the classification measures. But when the metric of performance becomes portfolio returns,

this ordering inverts altogether: GPT-Nano provides the only statistically significant financial outperformance, whilst ChatGPT-5.4 and Mistral provide no significant alpha despite their superior statistical characteristics. The full tables are available in Appendix F.

This paradox may be due to three mechanisms: To begin with, extreme selectivity acts as a quality filter: with only 17 positive predictions each GPT-Nano selection indicates a very strong whitepaper signal. When the model does give a favorable recommendation, it only does so for projects with characteristics so unambiguously favorable that the model's small processing power will not rationalize a negative verdict. In contrast, ChatGPT-5.4's 156 positive predictions necessarily contain many marginal cases that dilute portfolio-level returns. Second, larger models may have overconfidence calibration bias: because processing more nuanced bits of information makes them better than smaller models, they label reasonable signals to be sufficient bases for a positive prediction, decreasing the effective decision threshold below the level that would be optimal for financial screening. Third, with only 17 observations, the GPT-Nano result suffers from the existence of exceptional outliers, and the wide confidence interval warns against treating the outperformance as a stable, repeatable investment strategy without out-of-sample validation.

Irrespective of the specific mechanism, the finding is a new empirical contribution: in the context of ICO whitepaper evaluation, statistically more rigorous is not necessarily the most financially valuable AI model. Thus model selection should be decision-maker driven, where a specific objective is achieved, rather than by statistical fit metrics alone.

6. Discussion

6.1. Answering the Research Question

The core research question of this thesis is as follows: Can an AI language model accurately predict the success of ICO and future financial performance? The results gathered across three analytical areas produce systematic and distinct responses.

H1a: AI-based predictions and token trading success: H0 can be rejected

Regarding binary ICO success, yes. At an ordinary significance level, all four models produce variable prediction that is statistically significant in multivariate logistic regressions. This relevance derives from a broader set of observable ICO characteristics for which the AI signal is also truly incremental, thus making it more than simply a proxy for known predictors. Mistral has the best overall statistical fit (lowest AIC = 1,224.4), while also achieving the highest recall (49.0%) suggesting it is the best classifier. In terms of statistical significance, the three GPT variants each rank in descending order, with GPT-Nano falling behind as a classifier.

H1b: AI-based predictions and fundraising success: H0 can be rejected

Relating to fundraising volumes, the response is also positive, though it is more consistent across models. All the four AI models generate an alpha-scale fundraising premium of between \$6.19 million (Mistral) and \$9.34 million (GPT-Mini) per positively assessed ICO, after controlling for observable characteristics. This result is calculated on the full sample of 1202 observations mitigating survivorship concerns and is reliable at the inclusion of issuer rating, team quality, and whitepaper content quality metrics. It yields independent evidence that AI whitepaper assessments reflect dimensions of project quality that are consequential in real capital markets.

H2: AI-based selection and buy-and-hold returns: H0 can be partially rejected

The answer is partial and model-specific for future financial performance. GPT-Nano generation of a portfolio return is only statistically significant above the benchmark of verified successes ($t = 2.081$, $p = 0.057$ at portfolio level and $\beta = 1.910$, $p = 0.031$ in the multivariate OLS). ChatGPT-5.4, GPT-Mini, and Mistral in all cases, demonstrate directionally positive, but statistically insignificant BHR differentials. The financial results of GPT-Nano must be approached with caution because of small size of the portfolio ($N = 17$). The qualified conclusion can thus be drawn that AI models are able to yield investment applicable signals with respect to whitepaper content, but only if in a high-selectivity mode of operation that does not apply to wide dissemination.

H3: Model performance and investment returns: H0 cannot be rejected

The relative performances of the models are compared based on the major metrics in this analysis in Table 9. Their color code represents relative ranking in each dimension: green refers to a good one, yellow an average one, and red a lesser than or no significant outcome.

Table 9 — Comparative performance of AI models across all evaluation dimensions

Performance Dimension	ChatGPT-5.4	GPT-Mini	GPT-Nano	Mistral
Statistical Fit (AIC)	1,245.7 (2nd)	1,286.5 (3rd)	1,330.3 (4th)	1,224.4 (1st ✓)
Recall on Successes	30.9% (2nd)	14.9% (3rd)	3.4% (4th)	49.0% (1st ✓)
Precision	76.9% (2nd)	84.1% (1st ✓)	76.5% (3rd)	64.8% (4th)
Portfolio BH R (Mean)	−0.289 (n.s.)	+0.093 (n.s.)	+1.249 (sig.*) ✓	−0.266 (n.s.)
AmountRaised Premium	+\$8.73M (***)	+\$9.34M (***) ✓	+\$7.19M (*)	+\$6.19M (***)

*Note: Statistical significance codes: *** $p < 0.001$, * $p < 0.05$ (10% for BHR). AIC: lower is better. Recall and Precision computed on binary ICO success classification. BHR mean expressed as raw return; sig. = significant at 10% level. AmountRaised premium expressed in USD, net of controls.*

The synthesis table makes the central paradox of this study visually apparent. No single model dominates across all five dimensions. Mistral leads on statistical classification criteria (AIC, recall) but generates no significant BHR outperformance and the smallest fundraising premium. GPT-Nano is the weakest classifier by every statistical measure yet is the sole model to deliver statistically significant financial returns. GPT-Mini achieves the highest precision and the largest fundraising premium but falls short on recall and financial returns. ChatGPT-5.4 occupies a balanced intermediate position: second-best on most classification metrics, a large fundraising premium, but negligible BHR alpha.

This multi-dimensional picture challenges the common assumption that a more capable model, as measured by standard statistical benchmarks, is unambiguously preferable for financial applications. The evidence instead supports a contingent view: model selection should be driven by the specific objective of the decision-maker. A regulator or academic researcher prioritizing broad screening and statistical reliability should prefer Mistral. An institutional investor seeking a concentrated, high-conviction portfolio of early-stage token opportunities may find GPT-Nano's extreme selectivity more valuable. An issuer seeking to benchmark the likely fundraising impact of whitepaper quality will find consistent and large effects across all four models, with GPT-Mini offering the largest estimated premium.

6.2. Practical Implications for Investors

The results bear forward-looking implications for professional investors working in early-stage token markets. So first of all, AI whitepaper screening is an explicitly value-additive step in the due diligence exercise, even ignoring BHR results, the sustained fundraising premium that comes with AI-positive assessments means these models find projects that the aggregate market rewards by way of larger capital commitments. It serves as indirect verification for the signal being of value in a real-world setting. Second, the model choice should be adjusted based on the portfolio strategy. Investors pursuing small holdings in concentrated, high-conviction strategies should look to high-selectivity models like GPT-Nano, which finds a small number of ICOs statistically delivering superior expected returns. Investors interested in diversified early-stage token strategies, such as that of this publication, where coverage breadth is prized, should take Mistral's higher recall. Critically, these two objectives cannot be best optimized with just one model, and therefore the AI tool should not be recognized as the universal screening tool that is to be used as part of the broader, strategy-aware due diligence framework. Third, investors should note the non-linear relationship between model quality and financial performance explained in this study. The deployment of the most advanced or largest AI

model is not always the best to maximize investment return; it may be the confidence level at which the model finds a project is a probable success rather than the underlying architecture of the model that matters. This implies that either fine-tuning AI models on domain-based financial data and verifying decision thresholds from historical ICO samples could be a successful path for institutional investors aiming to implement AI-based screening on a massive scale.

6.3. Limitations

There are a number of limitations to the present study that need to be recognized since they condition the strength of the conclusions that can be drawn and help to delineate the boundaries of analyses. The biggest limitation is survivorship bias in the BHR analysis. Buy-and-Hold Return data are only available in around 277 of the 1,202 ICOs in the full sample, and this truncation is non-random: tokens that never made it to a secondary market, that stopped trading soon after listing, or whose price data was not systematically recorded are disproportionately concentrated among the worst-performing projects. The mean BHR figures reported for all portfolios are thus possibly upward biased compared to what an investor operating real time would have gotten, since they tacitly condition on a project's viability to the level of data already available. Although this concern is applicable equally across all portfolios (and not a single benchmark is ruled out) the cross-portfolio comparisons would not be mechanically invalidated, it would restrict the external validity of the absolute BHR figures. Third, although the GPT-Nano financial performance result is statistically significant, the entire portfolio was based on only 17 ICOs. At this sample size, a small number of outlier observations at this sample size have the potential to have disproportionate influence on the mean BHR, and the outcome might not be generalizable on a single independent out-of-sample dataset. The large confidence interval for the GPT-Nano t-test $[-0.058, +3.247]$ indicates this uncertainty and should temper strong causal inference. Fourth, none of our AI models in this project were fine-tuned on financial or ICO-specific data. The predictions created by ChatGPT-5.4, GPT-Mini, GPT-Nano and Mistral reflect the understanding of each model at its base in broader corpora rather than through domain-specific financial judgment. Since we had fine-tuned models on a corpus of ICO whitepapers with known outcomes, it could be quite fair to say that their classification and financial relevance would be way better than those of the general applicability models examined here. Lastly, we consider the prediction of each AI model as a binary variable and fail to incorporate data that would reveal how confident or conviction the model holds. A model which tentatively classifies a project as a success may be fundamentally less accurate than a model which categorises an implementation with high confidence; but the binary coding scheme blurs this division. Where available, future research can utilize this additional dimension in the AI signal to generate probabilistic outputs.

6.4 Avenues for Future Research

The results of this study unlock a number of fruitful avenues for further empirical and methodological developments. An immediate application seems to be moving from the analysis into more modern token offering formats. The shift from ICOs to IEOs (Initial Exchange Offerings), STOs (Security Token Offerings), and decentralised IDOs (Initial DEX Offerings) has significantly changed the disclosure landscape. Indeed, whether AI whitepaper assessment is still valid as a predictor in these novel formats is an open empirical question which may bear a practical application.

Second, the paradoxical nature of model size observed here requires further scrutiny. In particular, it might be worthwhile in the future to investigate if this finding reflects true differences in quality signals given by variable model sizes versus decision threshold calibration. Incremental changes in the confidence threshold at which a model is supposed to produce a correct prediction, accompanied by traceability of predicted precision-recall and return trade-offs as outcome, are needed to understand mechanisms with more detail.

Third, we could address the documentation endogeneity of the study using instrumental variable approaches. A valid tool would have to impact whitepaper quality independently of issuer resources- for example, the availability of AI writing assistance tools at the time of issuance that are not necessarily associated with whitepaper provision. Although tools like these are hard to quantify in practice, the methodological challenge is worth tackling in order to determine causality in the AI-fundraising link.

Finally, if using AI whitepaper assessments with some other data sources helps produce a more robust composite signal than if an AI's input alone is considered, future work might explore this question too. The R^2 values are low across specifications in this study, implying that while informative, whitepaper content is still capturing only a fraction of variance in ICO outcomes, and a multi-modal approach could enhance performance by enabling more accurate results when trying to predict ICO scores.

7. Conclusion

This thesis investigated the possibility of AI language models predicting ICO success through whitepapers, as well whether this prediction capability also has bearing on their future financial performance. We analysed a dataset of 1,202 ICOs and compared data between four models: ChatGPT-5.4, ChatGPT-5.4 Mini, ChatGPT-5.4 Nano, Mistral Medium 3. The purpose was not just to assess whether these models could label ICOs as successful or unsuccessful; also to understand what kind of success could be captured by them. As a result, the study drew a clear distinction between token trading success, fundraising success and abnormal buy-and-hold returns.

The results show that AI predictions contain helpful information regarding ICO prospects, but their relevance varies across performance measures. For token trading success, all four models give statistically significant predictive signals in the logistic regression analysis. Due to its lowest AIC and highest recall, Mistral gives the best results on the main classification criteria. This makes it a good, broad screening approach because it detects more successful ICOs than the other models. But in return, more false positives are introduced. In sum, Mistral identifies more successful projects than the other models, but also selects more projects that do not turn out to be viable. This already demonstrates that classification performance cannot be evaluated in a one-dimensional way.

The results in terms of fundraising success are even more visible. Positive predictions are tightly related to increased capital raised within the context of all four AI models. ICOs that are seen positively (positive) by the models, even once the visible features of the project are controlled, are, on average, raised more capital than is seen negatively by the models. This suggests that language models extract signals from whitepapers relevant to investor behavior at the fundraising stage. But this result is not proof of causation. A stronger whitepaper may appeal to investors, but better- resourced projects may also create stronger whitepapers and raise more capital through channels the model doesn't capture.

The evidence is less clear in relation to post-listing financial results. GPT-Nano is the only model demonstrating a statistically significant and strong positive association with abnormal buy-and-hold returns. This is one of the most fascinating insights of the analysis, because, per traditional classification norms, GPT-Nano is the weakest model. Because its portfolio is composed of just 17 ICOs, a high rate of extreme observations might strongly sway the result. Therefore, the thesis does not provide strong evidence that AI models generally predict superior investment performance.

What it does suggest is more nuanced: highly selective AI predictions may contain financially relevant information, but this result requires further testing on larger and independent samples.

One of the main takeaways from the study is that economic value should not be confused with predictive accuracy. Mistral is the strongest classifier, but it does not generate significant BHR outperformance. GPT-Nano, on the other hand, performs poorly as a classifier but delivers the

strongest return-based result. And this mismatch suggests the best statistical model is not the best investment model. Therefore, we have to evaluate the performance of the model with respect to the purpose that is being pursued: broad screening, fundraising prediction, token trading prediction, or investment selection.

Accordingly, the main contribution of this thesis is to identify that LLM-driven whitepaper examination should indeed work in conjunction with ICO evaluation and it should expose the shortcomings of treating ICO success as a single concept. Success in fundraising, success in token trading, and financial performance after listing don't say the same thing. A whitepaper could potentially help a project raise capital or become tradable while not necessarily demonstrating that a token will create value for investors. The fact that the benchmark portfolio averaged a negative BHR shows this up well: even ICOs that reached the secondary market failed to produce on average positive returns for investors.

There are several limitations that need to be taken into account in view of these findings. The results are likely contingent on the prompt, which different prompt designs may lead to different classifications. The models remain somewhat opaque in part also, because their textual reasons can hardly be regarded as full explanations of how they arrived at a decision. Moreover, certain ICOs that appear in the sample may have been part of models' training data, posing a risk of data leakage. The BHR analysis is also done from a very small effective sample and the dataset is heavily concentrated around the ICO boom period. Future research may thus incorporate the same method with more recent token offering formats, explore alternative prompting strategies, employ probabilistic confidence scores, and link whitepaper analysis with further market or project-level data.

In summary, the thesis demonstrates that AI-based language models can extract meaningful signals from ICO whitepapers, specifically for predicting the attractiveness of fundraising and the success of token trading. Where they are useful in forecasting long-term financial performance is less certain. The larger point is that AI can help with ICO evaluation, but should not be considered as a one-size-fits-all investment tool. Its utility depends, of course, on the targeted performance dimension and the manner in which predictive outputs influence decision-making.

8. References

- Adhami, S., Giudici, G., & Martinazzi, S. (2018). *Why do businesses go crypto? An empirical analysis of Initial Coin Offerings*. *Journal of Economics and Business*, 100, 64–75. DOI: 10.1016/j.jeconbus.2018.04.001.
- Amsden, R., & Schweizer, D. (2018). *Are Blockchain Crowdsales the New “Gold Rush”? Success Determinants of Initial Coin Offerings*. SSRN Electronic Journal. DOI: 10.2139/ssrn.3163849.
- Aslan, A., Şensoy, A., & Akdeniz, L. (2023). *Determinants of ICO success and post-ICO performance*. *Borsa Istanbul Review*, 23(1), 217–239. DOI: 10.1016/j.bir.2022.10.005.
- Basu, S., Garimella, A., Han, W., & Dennis, A. R. (2021). *Human decision making in AI augmented systems: Evidence from the Initial Coin Offering market*. In T. X. Bui (Ed.), *Proceedings of the 54th Annual Hawaii International Conference on System Sciences*, 176–185. DOI: 10.24251/HICSS.2021.020.
- Bourveau, T., De George, E. T., Ellahie, A., & Macciocchi, D. (2022). *The role of disclosure and information intermediaries in an unregulated capital market: Evidence from Initial Coin Offerings*. *Journal of Accounting Research*, 60(1), 129–167. DOI: 10.1111/1475-679X.12404.
- Campino, J., Brochado, A., & Rosa, Á. (2022). *Initial coin offerings (ICOs): Why do they succeed?* *Financial Innovation*, 8, Article 17. DOI: 10.1186/s40854-021-00317-2.
- DeMiguel, V., Gil-Bazo, J., Nogales, F. J., & Santos, A. A. P. (2023). *Machine learning and fund characteristics help to select mutual funds with positive alpha*. *Journal of Financial Economics*, 150(3), Article 103737. DOI: 10.1016/j.jfineco.2023.103737.
- Dessain, J. (2022). *Machine learning models predicting returns: Why most popular performance metrics are misleading and proposal for an efficient metric*. *Expert Systems with Applications*, 199, Article 116970. DOI: 10.1016/j.eswa.2022.116970.
- Elitzur, R., Katz, N., Muttath, P., & Soberman, D. (2024). *The power of machine learning methods to predict crowdfunding success: Accounting for complex relationships efficiently*. *Journal of Business Venturing Design*, 3, Article 100022. DOI: 10.1016/j.jbvd.2024.100022.
- Florysiak, D., & Schandlbauer, A. (2022). *Experts or charlatans? ICO analysts and white paper informativeness*. *Journal of Banking & Finance*, 139, Article 106476. DOI: 10.1016/j.jbankfin.2022.106476.
- Kaminski, J. C., & Hopp, C. (2020). *Predicting outcomes in crowdfunding campaigns with textual, visual, and linguistic signals*. *Small Business Economics*, 55(3), 627–649. DOI: 10.1007/s11187-019-00218-w.
- Karimov, B., & Wójcik, P. (2021). *Identification of scams in Initial Coin Offerings with machine learning*. *Frontiers in Artificial Intelligence*, 4, Article 718450. DOI: 10.3389/frai.2021.718450.

- Li, X., Chan, S., Zhu, X., Pei, Y., Ma, Z., Liu, X., & Shah, S. (2023). *Are ChatGPT and GPT-4 general-purpose solvers for financial text analytics? A study on several typical tasks*. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: Industry Track, 408–422. DOI: 10.18653/v1/2023.emnlp-industry.39.
- Liu, Y., Sheng, J., & Wang, W. (2025). *How Do Investors Value Technology in Cryptocurrency? Evidence from Textual Analysis*. SSRN Working Paper No. 3577208. DOI: 10.2139/ssrn.3577208.
- Lyandres, E., Palazzo, B., & Rabetti, D. (2022). *Initial Coin Offering (ICO) success and post-ICO performance*. Management Science, 68(12), 8658–8679. DOI: 10.1287/mnsc.2022.4312.
- Maarouf, A., Feuerriegel, S., & Pröllochs, N. (2025). *A fused large language model for predicting startup success*. European Journal of Operational Research, 322(1), 198–214. DOI: 10.1016/j.ejor.2024.09.011.
- Meoli, M., & Vismara, S. (2022). *Machine-learning forecasting of successful ICOs*. Journal of Economics and Business, 121, Article 106071. DOI: 10.1016/j.jeconbus.2022.106071.
- Momtaz, P. P. (2024). *Decentralized finance (DeFi) markets for startups: Search frictions, intermediation, and the efficiency of the ICO market*. Small Business Economics, 63, 1415–1447. DOI: 10.1007/s11187-024-00886-3.
- Roosenboom, P., van der Kolk, T., & de Jong, A. (2020). *What determines success in Initial Coin Offerings?* Venture Capital, 22(2), 161–183. DOI: 10.1080/13691066.2020.1741127.
- Ross, G., Das, S. R., Sciro, D., & Raza, H. (2021). *CapitalVX: A machine learning model for startup selection and exit prediction*. The Journal of Finance and Data Science, 7, 94–114. DOI: 10.1016/j.jfds.2021.04.001.
- Samieifar, S., & Baur, D. G. (2021). *Read me if you can! An analysis of ICO white papers*. Finance Research Letters, 38, Article 101427. DOI: 10.1016/j.frl.2020.101427.
- Thewissen, J., Shrestha, P., Torsin, W., & Pastwa, A. M. (2022). *Unpacking the black box of ICO white papers: A topic modeling approach*. Journal of Corporate Finance, 75, Article 102225. DOI: 10.1016/j.jcorpfin.2022.102225.
- Xu, W. (2021). *Prediction of initial coin offering success based on team knowledge and expert ratings*. Decision Support Systems, 146, Article 113574. DOI: 10.1016/j.dss.2021.113574.
- Zetsche, D. A., Buckley, R. P., Arner, D. W., & Föhr, L. (2017). *The ICO Gold Rush: It's a Scam, It's a Bubble, It's a Super Challenge for Regulators*. SSRN Electronic Journal. DOI: 10.2139/ssrn.3072298.

9. Appendix

Appendix A – Confusion Matrices and Prediction Accuracy

Success is defined as $\text{TokenTradedHistorical} = 1$. TN = true negative, FN = false negative, FP = false positive, TP = true positive.

Table A1. ChatGPT-5.4 prediction matrix

Predicted / Actual	Actual Failure (0)	Actual Success (1)
Predicted Failure (0)	778	268
Predicted Success (1)	36	120

Table A2. GPT-Mini prediction matrix

Predicted / Actual	Actual Failure (0)	Actual Success (1)
Predicted Failure (0)	803	330
Predicted Success (1)	11	58

Table A3. GPT-Nano prediction matrix

Predicted / Actual	Actual Failure (0)	Actual Success (1)
Predicted Failure (0)	810	375
Predicted Success (1)	4	13

Table A4. Mistral Medium 3 prediction matrix

Predicted / Actual	Actual Failure (0)	Actual Success (1)
Predicted Failure (0)	711	198
Predicted Success (1)	103	190

Table A5. Prediction accuracy and success prediction rates

Model	Total	Accuracy	Accuracy (%)	Actual success prediction rate / Recall	Recall (%)	Precision (%)	F1-score (%)
ChatGPT-5.4	1202	0.7470881864	74.71%	0.3092783505	30.93%	76.92%	44.12%
GPT-Mini	1202	0.7163061564	71.63%	0.1494845361	14.95%	84.06%	25.38%
GPT-Nano	1202	0.6846921797	68.47%	0.0335051546	3.35%	76.47%	6.42%
Mistral Medium 3	1202	0.7495840266	74.96%	0.4896907216	48.97%	64.85%	55.80%

For ChatGPT-5.4, GPT-Mini, and GPT-Nano, the accuracy and actual success prediction rates reproduce the source output. Mistral metrics are computed directly from the reported confusion matrix.

Appendix B – Logistic Regression Results (ICO Success Prediction)

Specification: $\text{TokenTradedHistorical} \sim \text{AI prediction} + \text{WP_Readability} + \text{WP_Sentiment} + \text{WP_Pages} + \text{Year} + \text{TaxHaven} + \text{USRestrict} + \text{Rating} + \text{SocialMedia} + \text{Video} + \text{Eth} + \text{WhitelistKYC} + \text{Team} + \text{TokenDist} + \text{MinInvest} + \text{NumbCurr} + \text{Fiat} + \text{PreICO} + \text{HardCap} + \text{SoftCap} + \text{ESG} + \text{Bonus}$; binomial family.

*Significance codes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; . $p < 0.10$.*

Table B1. Logistic regression – ChatGPT-5.4

Variable	Estimate	Std. Error	z value	p-value	Sig.
(Intercept)	1.863e+03	3.526e+02	5.282	1.27e-07	***
ChatGPT	2.030e+00	2.290e-01	8.863	< 2e-16	***
WP_Readability	4.375e-02	7.669e-02	0.570	0.568354	
WP_Sentiment	-2.114e+01	1.110e+01	-1.905	0.056813	.
WP_Pages	1.013e-02	8.882e-02	0.114	0.909204	
Year	-9.234e-01	1.748e-01	-5.283	1.27e-07	***
TaxHaven	2.405e-01	7.180e-02	3.350	0.000808	***
USRestrict	1.068e-01	7.838e-02	1.363	0.172892	
Rating	5.557e-01	1.120e-01	4.963	6.96e-07	***
SocialMedia	2.992e-03	1.068e-01	0.028	0.977664	
Video	1.313e-01	1.078e-01	1.217	0.223472	
Eth	-1.073e-01	8.222e-02	-1.305	0.191979	
WhitelistKYC	-8.360e-02	8.659e-02	-0.965	0.334315	
Team	1.775e-01	7.975e-02	2.226	0.025997	*
TokenDist	1.329e-01	1.210e-01	1.098	0.272049	
MinInvest	-1.716e-01	1.529e-01	-1.122	0.261753	

Variable	Estimate	Std. Error	z value	p-value	Sig.
NumbCurr	9.481e-03	3.798e-02	0.250	0.802868	
Fiat	-9.905e-02	1.788e-01	-0.554	0.579678	
PreICO	-2.366e-01	1.677e-01	-1.411	0.158135	
HardCap	1.882e-01	3.094e-01	0.608	0.543067	
SoftCap	-2.061e-01	2.082e-01	-0.990	0.322232	
ESG	-3.090e+00	1.778e+00	-1.738	0.082229	.
Bonus	-2.149e-01	1.760e-01	-1.221	0.222200	

Table B2. Logistic regression – GPT-Mini

Variable	Estimate	Std. Error	z value	p-value	Sig.
(Intercept)	2.037e+03	3.478e+02	5.856	4.75e-09	***
GPT_Mini1	2.339e+00	3.839e-01	6.092	1.11e-09	***
WP_Readability	2.637e-02	7.524e-02	0.350	0.725991	
WP_Sentiment	-1.656e+01	1.086e+01	-1.525	0.127291	
WP_Pages	5.817e-03	8.652e-02	0.067	0.946393	
Year	-1.010e+00	1.724e-01	-5.856	4.75e-09	***
TaxHaven	2.547e-01	7.026e-02	3.625	0.000289	***
USRestrict	9.845e-02	7.680e-02	1.282	0.199886	
Rating	6.160e-01	1.095e-01	5.623	1.88e-08	***
SocialMedia	-5.525e-02	1.033e-01	-0.535	0.592711	
Video	1.611e-01	1.069e-01	1.507	0.131823	
Eth	-9.633e-02	8.041e-02	-1.198	0.230898	
WhitelistKYC	-5.496e-02	8.427e-02	-0.652	0.514310	
Team	1.789e-01	7.640e-02	2.341	0.019217	*
TokenDist	1.629e-01	1.195e-01	1.362	0.173090	
MinInvest	-1.870e-01	1.500e-01	-1.247	0.212448	

Variable	Estimate	Std. Error	z value	p-value	Sig.
NumbCurr	1.351e-02	3.673e-02	0.368	0.713033	
Fiat	-1.081e-01	1.742e-01	-0.621	0.534783	
PreICO	-2.622e-01	1.639e-01	-1.600	0.109670	
HardCap	1.184e-01	2.975e-01	0.398	0.690660	
SoftCap	-2.448e-01	2.030e-01	-1.205	0.228011	
ESG	-1.989e+00	1.725e+00	-1.153	0.248751	
Bonus	-2.604e-01	1.721e-01	-1.513	0.130352	

Table B3. Logistic regression – GPT-Nano

Variable	Estimate	Std. Error	z value	p-value	Sig.
(Intercept)	2.086e+03	3.421e+02	6.099	1.07e-09	***
GPT_Nano1	1.492e+00	6.156e-01	2.424	0.015358	*
WP_Readability	3.278e-02	7.362e-02	0.445	0.656130	
WP_Sentiment	-1.963e+01	1.063e+01	-1.847	0.064758	.
WP_Pages	7.528e-03	8.597e-02	0.088	0.930226	
Year	-1.034e+00	1.696e-01	-6.098	1.07e-09	***
TaxHaven	2.428e-01	6.868e-02	3.535	0.000408	***
USRestrict	8.740e-02	7.519e-02	1.162	0.245056	
Rating	6.729e-01	1.067e-01	6.308	2.83e-10	***
SocialMedia	-9.236e-02	1.009e-01	-0.915	0.360199	
Video	1.377e-01	1.037e-01	1.328	0.184109	
Eth	-8.969e-02	7.904e-02	-1.135	0.256468	
WhitelistKYC	-4.933e-02	8.260e-02	-0.597	0.550402	
Team	2.058e-01	7.394e-02	2.783	0.005384	**
TokenDist	1.621e-01	1.170e-01	1.386	0.165855	
MinInvest	-2.149e-01	1.465e-01	-1.467	0.142365	

Variable	Estimate	Std. Error	z value	p-value	Sig.
NumbCurr	8.592e-03	3.652e-02	0.235	0.814006	
Fiat	-1.200e-01	1.711e-01	-0.701	0.483012	
PreICO	-3.372e-01	1.596e-01	-2.113	0.034641	*
HardCap	1.608e-01	2.945e-01	0.546	0.585028	
SoftCap	-2.341e-01	1.995e-01	-1.173	0.240680	
ESG	-1.469e+00	1.684e+00	-0.872	0.383065	
Bonus	-3.382e-01	1.666e-01	-2.030	0.042386	*

Table B4. Logistic regression – Mistral Medium 3

Variable	Estimate	Std. Error	z value	p-value	Sig.
(Intercept)	1.791e+03	3.520e+02	5.087	3.63e-07	***
Mistral1	1.683e+00	1.661e-01	10.133	< 2e-16	***
WP_Readability	4.448e-02	7.839e-02	0.567	0.5704	
WP_Sentiment	-2.219e+01	1.120e+01	-1.981	0.0476	*
WP_Pages	-2.204e-02	9.112e-02	-0.242	0.8089	
Year	-8.878e-01	1.745e-01	-5.088	3.61e-07	***
TaxHaven	1.783e-01	7.294e-02	2.444	0.0145	*
USRestrict	8.629e-02	7.964e-02	1.083	0.2786	
Rating	5.464e-01	1.131e-01	4.833	1.35e-06	***
SocialMedia	-6.595e-02	1.072e-01	-0.615	0.5385	
Video	1.113e-01	1.084e-01	1.027	0.3045	
Eth	-1.017e-01	8.222e-02	-1.237	0.2161	
WhitelistKYC	-6.897e-02	8.699e-02	-0.793	0.4278	
Team	1.824e-01	7.964e-02	2.290	0.0220	*
TokenDist	1.443e-01	1.216e-01	1.187	0.2353	
MinInvest	-2.053e-01	1.546e-01	-1.328	0.1842	

Variable	Estimate	Std. Error	z value	p-value	Sig.
NumbCurr	1.194e-02	3.896e-02	0.307	0.7592	
Fiat	9.646e-04	1.812e-01	0.005	0.9958	
PreICO	-2.128e-01	1.696e-01	-1.255	0.2096	
HardCap	1.855e-01	3.072e-01	0.604	0.5460	
SoftCap	-2.657e-01	2.091e-01	-1.270	0.2040	
ESG	-3.723e+00	1.820e+00	-2.046	0.0408	*
Bonus	-2.815e-01	1.766e-01	-1.594	0.1109	

Table B5. Logistic regression model statistics

Statistic	ChatGPT-5.4	GPT-Mini	GPT-Nano	Mistral Medium 3
Null deviance	1455.6 on 1165 df	1455.6 on 1165 df	1455.6 on 1165 df	1449.5 on 1157 df
Residual deviance	1199.7 on 1143 df	1240.5 on 1143 df	1284.3 on 1143 df	1178.4 on 1135 df
Missing observations	50	50	50	44
AIC	1245.7	1286.5	1330.3	1224.4

Appendix C – Comparative Model Quality

Table C1. Comparative classification and logistic model quality

Model	TN	FN	FP	TP	Pred. success	Accuracy	Precision	Recall	F1	AIC	Null deviance	Residual deviance	Missing obs.
ChatGPT-5.4	778	268	36	120	156	74.71%	76.92%	30.93%	44.12%	1245.7	1455.6 on 1165 df	1199.7 on 1143 df	50
GPT-Mini	803	330	11	58	69	71.63%	84.06%	14.95%	25.38%	1286.5	1455.6 on 1165 df	1240.5 on 1143 df	50
GPT-Nano	810	375	4	13	17	68.47%	76.47%	3.35%	6.42%	1330.3	1455.6 on 1165 df	1284.3 on 1143 df	50
Mistral Medium 3	711	198	103	190	293	74.96%	64.85%	48.97%	55.80%	1224.4	1449.5 on 1157 df	1178.4 on 1135 df	44

Appendix D – Amount Raised Regressions

Specification: AmountRaised $\sim$ AI prediction + WP_Readability + WP_Sentiment + WP_Pages + Year + TaxHaven + USRestrict + Rating + SocialMedia + Video + Eth + WhitelistKYC + Team + TokenDist + MinInvest + NumbCurr + Fiat + PreICO + HardCap + SoftCap + Bonus; OLS regression.

Significance codes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; . $p < 0.10$.

Table D1. AmountRaised regression – ChatGPT-5.4

Variable	Estimate	Std. Error	t value	p-value	Sig.
(Intercept)	2038538362	1399873468	1.456	0.14559	
ChatGPT1	8733649	1005094	8.689	< 2e-16	***
WP_Readability	261701	327893	0.798	0.42495	
WP_Sentiment	-124397140	49857905	-2.495	0.01273	*
WP_Pages	-6288	345979	-0.018	0.98550	
Year	-1006788	693868	-1.451	0.14705	
TaxHaven	903765	327118	2.763	0.00582	**
USRestrict	-185908	348742	-0.533	0.59408	
Rating	909612	490822	1.853	0.06410	.
SocialMedia	-1445986	458723	-3.152	0.00166	**
Video	271890	441035	0.616	0.53769	
Eth	-270713	370228	-0.731	0.46480	
WhitelistKYC	583688	375064	1.556	0.11992	
Team	728479	348909	2.088	0.03702	*
TokenDist	-390001	457721	-0.852	0.39436	
MinInvest	-455845	683910	-0.667	0.50520	

Variable	Estimate	Std. Error	t value	p-value	Sig.
NumbCurr	259049	178755	1.449	0.14755	
Fiat	1021618	794546	1.286	0.19877	
PreICO	242434	749181	0.324	0.74630	
HardCap	701960	1319157	0.532	0.59474	
SoftCap	-650384	938241	-0.693	0.48832	
Bonus	-196159	781742	-0.251	0.80192	

Table D2. AmountRaised regression – GPT-Mini

Variable	Estimate	Std. Error	t value	p-value	Sig.
(Intercept)	2818502063	1412302415	1.996	0.046197	*
GPT_Mini1	9337716	1446784	6.454	1.58e-10	***
WP_Readability	262395	332414	0.789	0.430058	
WP_Sentiment	-109257046	50612169	-2.159	0.031073	*
WP_Pages	55084	350608	0.157	0.875186	
Year	-1392748	700040	-1.990	0.046873	*
TaxHaven	968469	331454	2.922	0.003545	**
USRestrict	-176427	353579	-0.499	0.617889	
Rating	1195802	494954	2.416	0.015843	*
SocialMedia	-1633373	463824	-3.522	0.000445	***
Video	372673	447044	0.834	0.404652	
Eth	-204110	375296	-0.544	0.586636	
WhitelistKYC	663201	379947	1.746	0.081155	.
Team	786242	353975	2.221	0.026527	*
TokenDist	-271199	463623	-0.585	0.558689	
MinInvest	-547548	693086	-0.790	0.429677	

Variable	Estimate	Std. Error	t value	p-value	Sig.
NumbCurr	260771	181234	1.439	0.150454	
Fiat	1035354	805555	1.285	0.198950	
PreICO	60626	758766	0.080	0.936330	
HardCap	499639	1337293	0.374	0.708754	
SoftCap	-759959	951247	-0.799	0.424504	
Bonus	-541513	790079	-0.685	0.493231	

Table D3. AmountRaised regression – GPT-Nano

Variable	Estimate	Std. Error	t value	p-value	Sig.
(Intercept)	3527340143	1430377957	2.466	0.013803	*
GPT_Nano1	7193663	2831953	2.540	0.011206	*
WP_Readability	256590	337546	0.760	0.447309	
WP_Sentiment	-122336541	51309361	-2.384	0.017268	*
WP_Pages	114538	356145	0.322	0.747808	
Year	-1743623	709006	-2.459	0.014065	*
TaxHaven	993333	336287	2.954	0.003201	**
USRestrict	-214303	358717	-0.597	0.550345	
Rating	1573068	498394	3.156	0.001638	**
SocialMedia	-1814826	469592	-3.865	0.000117	***
Video	322730	453594	0.711	0.476918	
Eth	-240120	380782	-0.631	0.528425	
WhitelistKYC	670638	386354	1.736	0.082857	.
Team	966303	358040	2.699	0.007056	**
TokenDist	-242039	470448	-0.514	0.607008	
MinInvest	-657951	702990	-0.936	0.349499	

Variable	Estimate	Std. Error	t value	p-value	Sig.
NumbCurr	240374	183873	1.307	0.191370	
Fiat	935381	817184	1.145	0.252589	
PreICO	-202167	768671	-0.263	0.792590	
HardCap	675216	1356996	0.498	0.618872	
SoftCap	-758082	966005	-0.785	0.432751	
Bonus	-959330	798546	-1.201	0.229857	

Table D4. AmountRaised regression – Mistral Medium 3

Variable	Estimate	Std. Error	t value	p-value	Sig.
(Intercept)	2315024785	1407442916	1.645	0.100267	
Mistral1	6186252	802984	7.704	2.77e-14	***
WP_Readability	203655	329980	0.617	0.537238	
WP_Sentiment	-124342245	50180426	-2.478	0.013354	*
WP_Pages	-207416	349973	-0.593	0.553519	
Year	-1143837	697619	-1.640	0.101347	
TaxHaven	741765	330690	2.243	0.025075	*
USRestrict	-223789	350959	-0.638	0.523825	
Rating	914742	495488	1.846	0.065120	.
SocialMedia	-1656699	459986	-3.602	0.000329	***
Video	225222	444059	0.507	0.612117	
Eth	-229335	372594	-0.616	0.538338	
WhitelistKYC	615699	377399	1.631	0.103066	
Team	816607	350514	2.330	0.019987	*
TokenDist	-291059	460342	-0.632	0.527334	
MinInvest	-515128	688173	-0.749	0.454280	

Variable	Estimate	Std. Error	t value	p-value	Sig.
NumbCurr	262521	179929	1.459	0.144823	
Fiat	1261249	800879	1.575	0.115562	
PreICO	267937	754772	0.355	0.722659	
HardCap	825668	1327932	0.622	0.534213	
SoftCap	-836861	944625	-0.886	0.375841	
Bonus	-573559	783207	-0.732	0.464117	

Table D5. AmountRaised residuals and model statistics

Statistic	ChatGPT-5.4	GPT-Mini	GPT-Nano	Mistral Medium 3
Residuals: Min	-19469937	-20814596	-15252699	-16570816
Residuals: 1Q	-6314861	-6690968	-7026163	-6332651
Residuals: Median	-3702865	-3848368	-3964762	-3582575
Residuals: 3Q	2385982	2551038	2980120	2411061
Residuals: Max	51281913	51388328	51850243	51141687
Residual SE	11080000 on 1186 df	11230000 on 1186 df	11390000 on 1186 df	11150000 on 1186 df
R-squared	0.1153	0.09094	0.0641	0.1039
Adjusted R-squared	0.09967	0.07484	0.04753	0.08799
F-statistic	7.363 on 21 and 1186 df	5.65 on 21 and 1186 df	3.868 on 21 and 1186 df	6.545 on 21 and 1186 df
p-value	< 2.2e-16	9.001e-15	1.101e-08	< 2.2e-16

Appendix E – BHR Regression Results

Specification: AbnormalBHR $\sim$ AI prediction + ESG + WP_Readability + WP_Sentiment + WP_Pages + Year + TaxHaven + USRestrict + Rating + SocialMedia + Video + Eth + WhitelistKYC + Team + TokenDist + MinInvest + NumbCurr + Fiat + PreICO + HardCap + SoftCap + Bonus; OLS regression.

*Significance codes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; . $p < 0.10$. Residual quantiles are included where they appeared in the source output.*

Table E1. AbnormalBHR regression – ChatGPT-5.4

Variable	Estimate	Std. Error	t value	p-value	Sig.
(Intercept)	824.93125	725.19568	1.138	0.2564	
ChatGPT1	0.15063	0.38890	0.387	0.6988	
ESG	-3.26051	4.50343	-0.724	0.4697	
WP_Readability	-0.15695	0.20924	-0.750	0.4539	
WP_Sentiment	4.54391	27.82794	0.163	0.8704	
WP_Pages	0.20481	0.20540	0.997	0.3196	
Year	-0.40945	0.35949	-1.139	0.2558	
TaxHaven	0.27813	0.18387	1.513	0.1316	
USRestrict	0.13806	0.20132	0.686	0.4935	
Rating	-0.06008	0.25721	-0.234	0.8155	
SocialMedia	0.29705	0.29529	1.006	0.3154	
Video	0.03036	0.31057	0.098	0.9222	
Eth	0.12952	0.20857	0.621	0.5352	
WhitelistKYC	0.05588	0.21128	0.264	0.7916	
Team	-0.18198	0.16040	-1.134	0.2577	
TokenDist	-0.06877	0.43386	-0.158	0.8742	

Variable	Estimate	Std. Error	t value	p-value	Sig.
MinInvest	-0.02007	0.38302	-0.052	0.9583	
NumbCurr	-0.09010	0.10406	-0.866	0.3874	
Fiat	-0.39845	0.49468	-0.805	0.4213	
PreICO	-0.25285	0.40422	-0.626	0.5322	
HardCap	1.75943	0.78381	2.245	0.0256	*
SoftCap	-0.59663	0.53321	-1.119	0.2642	
Bonus	0.60845	0.43631	1.395	0.1644	

Table E2. AbnormalBHR regression – GPT-Mini

Variable	Estimate	Std. Error	t value	p-value	Sig.
(Intercept)	7.498e+02	7.224e+02	1.038	0.3003	
GPT_Mini1	5.857e-01	4.839e-01	1.211	0.2272	
ESG	-3.199e+00	4.470e+00	-0.716	0.4749	
WP_Readability	-1.445e-01	2.090e-01	-0.692	0.4898	
WP_Sentiment	8.113e+00	2.793e+01	0.291	0.7716	
WP_Pages	2.069e-01	2.042e-01	1.014	0.3118	
Year	-3.722e-01	3.581e-01	-1.039	0.2996	
TaxHaven	2.863e-01	1.833e-01	1.563	0.1194	
USRestrict	1.549e-01	2.013e-01	0.769	0.4425	
Rating	-7.149e-02	2.543e-01	-0.281	0.7788	
SocialMedia	3.018e-01	2.927e-01	1.031	0.3035	
Video	2.582e-02	3.096e-01	0.083	0.9336	
Eth	1.268e-01	2.080e-01	0.610	0.5427	
WhitelistKYC	3.591e-02	2.114e-01	0.170	0.8652	
Team	-1.910e-01	1.588e-01	-1.203	0.2302	
TokenDist	-8.059e-02	4.329e-01	-0.186	0.8525	

Variable	Estimate	Std. Error	t value	p-value	Sig.
MinInvest	8.415e-04	3.824e-01	0.002	0.9982	
NumbCurr	-8.545e-02	1.038e-01	-0.823	0.4113	
Fiat	-3.618e-01	4.944e-01	-0.732	0.4649	
PreICO	-2.465e-01	4.028e-01	-0.612	0.5412	
HardCap	1.722e+00	7.822e-01	2.202	0.0286	*
SoftCap	-6.489e-01	5.328e-01	-1.218	0.2244	
Bonus	6.511e-01	4.368e-01	1.491	0.1373	

Table E3. AbnormalBHR regression – GPT-Nano

Variable	Estimate	Std. Error	t value	p-value	Sig.
(Intercept)	693.433730	716.043701	0.968	0.3338	
GPT_Nano1	1.910215	0.879490	2.172	0.0308	*
ESG	-3.697912	4.449871	-0.831	0.4067	
WP_Readability	-0.104229	0.208824	-0.499	0.6181	
WP_Sentiment	11.425521	27.768471	0.411	0.6811	
WP_Pages	0.237264	0.203545	1.166	0.2448	
Year	-0.344268	0.354958	-0.970	0.3330	
TaxHaven	0.285114	0.181852	1.568	0.1182	
USRestrict	0.160570	0.199761	0.804	0.4223	
Rating	-0.009853	0.252195	-0.039	0.9689	
SocialMedia	0.290699	0.290473	1.001	0.3179	
Video	0.027688	0.307694	0.090	0.9284	
Eth	0.122003	0.206734	0.590	0.5556	
WhitelistKYC	-0.011948	0.211726	-0.056	0.9550	
Team	-0.212668	0.158193	-1.344	0.1800	
TokenDist	-0.070779	0.430002	-0.165	0.8694	

Variable	Estimate	Std. Error	t value	p-value	Sig.
MinInvest	0.061182	0.381466	0.160	0.8727	
NumbCurr	-0.085523	0.103149	-0.829	0.4078	
Fiat	-0.366562	0.490280	-0.748	0.4554	
PreICO	-0.243532	0.400144	-0.609	0.5433	
HardCap	1.729181	0.776917	2.226	0.0269	*
SoftCap	-0.708029	0.530366	-1.335	0.1831	
Bonus	0.612003	0.432185	1.416	0.1580	

Table E4. AbnormalBHR regression – Mistral Medium 3

Variable	Estimate	Std. Error	t value	p-value	Sig.
(Intercept)	861.35134	722.36930	1.192	0.234	
Mistral1	0.01386	0.37604	0.037	0.971	
ESG	-3.10178	4.50422	-0.689	0.492	
WP_Readability	-0.15676	0.21010	-0.746	0.456	
WP_Sentiment	4.35433	27.83173	0.156	0.876	
WP_Pages	0.19716	0.20479	0.963	0.337	
Year	-0.42748	0.35808	-1.194	0.234	
TaxHaven	0.27255	0.18390	1.482	0.140	
USRestrict	0.13573	0.20143	0.674	0.501	
Rating	-0.04532	0.25565	-0.177	0.859	
SocialMedia	0.28326	0.29322	0.966	0.335	
Video	0.02684	0.31055	0.086	0.931	
Eth	0.12933	0.20894	0.619	0.536	
WhitelistKYC	0.05996	0.21136	0.284	0.777	
Team	-0.17258	0.15859	-1.088	0.278	
TokenDist	-0.06739	0.43400	-0.155	0.877	

Variable	Estimate	Std. Error	t value	p-value	Sig.
MinInvest	-0.02097	0.38366	-0.055	0.956	
NumbCurr	-0.08966	0.10409	-0.861	0.390	
Fiat	-0.40279	0.50064	-0.805	0.422	
PreICO	-0.25997	0.40529	-0.641	0.522	
HardCap	1.75716	0.78448	2.240	0.026	*
SoftCap	-0.60463	0.53377	-1.133	0.258	
Bonus	0.60309	0.43666	1.381	0.168	

Appendix F – Portfolio Comparison

Table F1. Portfolio summary statistics including Mistral

Portfolio	N	Mean_BHR	SD_BHR
Real Success	388	-0.34535939	2.865337
ChatGPT	156	-0.28940961	1.864138
GPT-Mini	69	0.09264313	1.826476
GPT-Nano	17	1.24913740	2.695499
Mistral	293	-0.26587907	2.452551

Table F2. Welch t-tests against the Real Success benchmark

Comparison	Mean_AI	Mean_Real	t-statistic	df	p-value	95% CI lower	95% CI upper	Significant at 10%
ChatGPT vs Real	-0.28940961	-0.3453594	0.2289719	300.27	0.81904661	-0.4249098	0.5368094	FALSE
GPT-Mini vs Real	0.09264313	-0.3453594	1.4514808	105.09	0.14962489	-0.1603311	1.0363362	FALSE
GPT-Nano vs Real	1.24913740	-0.3453594	2.0811237	13.236	0.05738157	-0.05771556	3.24670914	TRUE
Mistral vs Real	-0.26587907	-0.3453594	0.3106093	374.75	0.75627034	-0.4236705	0.5826311	FALSE

