A Hybrid, Transparent Trust and Risk Assessment Framework for Cryptocurrency Exchanges

Bongani Mawhayi[1,] Johnny Botha[2] and Louise Leenen[1,3]
[1]University of the Western Cape, Cape Town, South Africa
[2]Council for Scientific and Industrial Research, Pretoria, South Africa
[3]The Centre for Artificial Intelligence Research, Cape Town, South Africa

4013109@myuwc.ac.za
jbotha1@csir.co.za
lleenen@uwc.ac.za

Abstract : Cryptocurrency exchanges act as critical intermediaries within the digital asset ecosystem, yet users currently rely on largely opaque, platform-defined trust scores to assess their reliability and risk. Existing industry frameworks, notably those produced by CoinGecko and CoinMarketCap, provide useful signals related to liquidity and volume integrity but suffer from limited transparency, fixed weighting schemes, and the absence of sentiment-based assessment. This paper presents HTREx (Hybrid Trust and Risk Evaluation framework for exchanges), a semi-automated, modular trust and risk assessment framework for cryptocurrency exchanges that addresses these limitations. The framework integrates five dimensions of exchange integrity: user sentiment, regulatory compliance, technical security, transparency, and incident history. Sentiment is quantified using transformer-based natural language processing applied to user-generated content from mobile application reviews and online forums. Compliance is assessed through structured extraction of regulatory and operational disclosures from Terms of Service documents using large language models. Security, transparency, and incident history are evaluated through a combination of publicly verifiable indicators, third-party assessments, and a recency-weighted incident scoring model. All components are normalised and aggregated into a composite score using user-adjustable weights, enabling personalised risk prioritisation while retaining a defensible default configuration for comparative analysis. The framework is demonstrated using four prominent exchanges—Kraken, Coinbase, Binance, and Uniswap—highlighting clear differences between centralised and decentralised platforms and illustrating how sentiment, compliance, and historical incidents materially influence overall trust assessments. The results suggest that transparent, extensible, and user-configurable scoring models can provide a more interpretable and context-sensitive evaluation of exchange risk than existing monolithic trust scores, with direct relevance for both retail and institutional participants.

Keywords: Cryptocurrency exchanges, Trust assessment, Risk scoring, Sentiment analysis

1. Introduction

Cryptocurrency exchanges constitute the primary access points to digital asset markets, enabling users to buy, sell, trade, and custody cryptocurrencies. As critical infrastructure within the blockchain ecosystem, exchanges face persistent challenges related to uneven regulation, heterogeneous operational standards, and recurring security and governance failures. High-profile incidents, including exchange hacks, prolonged service outages, and regulatory enforcement actions, demonstrate that exchange-related risk extends beyond asset price volatility and directly affects user funds, data security, and market integrity (Le Pennec, et al., 2021) (Cong, et al., 2022). To support user decision-making, industry platforms have introduced exchange ranking and trust-scoring mechanisms. However, these systems have been criticised for methodological volatility and opaque aggregation practices, raising concerns regarding consistency and interpretability (Fitzpatrick, 2019). Vidal-Tomás (2022) further highlights the absence of standardised methods for aggregating exchange data across coin ranking platforms.

CoinGecko and CoinMarketCap, the two most widely used cryptocurrency data aggregators, provide composite exchange scores derived from metrics such as liquidity, trading volume, web traffic, and selected qualitative indicators (Ong, 2020) (CoinMarketCap, 2024). Although these frameworks help reduce information asymmetry in cryptocurrency markets, they exhibit several limitations. Key aspects of their methodologies remain partially opaque or proprietary, particularly in the case of CoinMarketCap, limiting reproducibility and independent scrutiny (Fenton, 2020). In addition, both platforms rely on fixed weighting schemes that assume a homogeneous user risk profile and do not explicitly incorporate behavioural sentiment as a component of exchange trust, despite evidence that public perception and user experience influence platform adoption and risk exposure (Garcia, et al., 2014) (Broder, 2025; Wang, et al., 2022). Trust and risk in cryptocurrency exchanges are inherently multi-dimensional and context dependent, with different users prioritising factors such as security, compliance, transparency, liquidity, and usability. Static trust scores may therefore obscure trade-offs between exchange risk dimensions and misrepresent risk for specific user profiles (Simon, 2023).

This paper addresses these limitations through HTREx (Hybrid Trust and Risk Evaluation Framework for Cryptocurrency Exchanges), a transparent, modular trust and risk assessment framework. HTREx integrates five dimensions of exchange integrity: user sentiment, regulatory compliance, technical security, transparency and liquidity, and incident history. Sentiment is quantified using transformer-based natural language processing applied to user-generated content, while compliance indicators are extracted from Terms of Service documents using large language models. Security, transparency, and incident history are assessed using publicly verifiable indicators and a recency-weighted incident model. Component scores are normalised and aggregated through a user-configurable weighting mechanism, enabling personalised risk prioritisation while retaining a default benchmark configuration.

The contribution of this paper is threefold. First, it formalises sentiment as a measurable dimension of exchange trust. Second, it demonstrates how large language models can be applied to systematically extract compliance-related indicators from unstructured legal documentation (Hutto & Gilbert, 2014) (Barbieri, et al., 2020). Third, it introduces a configurable aggregation model that bridges the gap between opaque platform-defined rankings and user-specific risk assessment.

2. Methodology

This study proposes HTREx, a modular framework for cryptocurrency exchange trust and risk assessment. HTREx builds on existing industry approaches while addressing limitations related to static weighting schemes and the lack of sentiment integration(Ong, 2020) (CoinMarketCap, 2024). The framework decomposes exchange trustworthiness into five analytically distinct components and supports user-adjustable weighting to reflect different risk priorities (CoinGecko, 2023).

2.1 Conceptual Structure of the Framework

HTREx models exchange trust as a composite function of five components: Sentiment (S), Compliance (C), Security (Sec), Transparency (T), and Incident History (I). Each component is independently assessed using data sources and techniques appropriate to its underlying risk dimension. Component scores are normalised to a common scale in the interval [0,1], where higher values indicate lower perceived risk and higher trust.

The final trust score, T_{Trust}, for a given exchange is computed as a weighted linear aggregation of these components:

$$T_{\text{Trust}} = \sum_{k=1}^{5} w_k X_k$$

where $X_k \in \{S_{Sentiment}, S_{Compliance}, S_{Security}, S_{Transparency}, S_{Incident}\}$ denotes the normalised component scores, and $w = [w_1, w_2, w_3, w_4, w_5]^T$ is the weight vector, with w_k representing the weight assigned to the kth component. The weights satisfy the constraint:

$$\sum_{k=1}^{5} w_k = 1\,, w_k \geq 0\ \forall k$$

Unlike many industry rankings, the weighting vector is user-adjustable to reflect risk tolerance, regulatory context, or use case. A default configuration is also provided for benchmarking and comparative analysis.

2.2 Data Collection Strategy

HTREx adopts a hybrid data collection strategy combining automated and manual methods. Automated collection is used for sentiment and liquidity-related metrics through public APIs and aggregator services, while manual collection supports compliance, security disclosures, and incident history derived from Terms of Service documents, audit reports, and public reporting. All collected data are converted into structured JSON representations for downstream processing. This approach reflects the fragmented nature of exchange-related data and the absence of a single comprehensive source for exchange risk assessment (Simon, 2023)

2.3 Component Normalisation and Aggregation

Each component score is derived using a method tailored to its data type and risk characteristics but is ultimately normalised to the same numeric range. This design choice ensures that no single component dominates the composite score by virtue of scale rather than substantive importance. Binary indicators (e.g. presence or absence of specific compliance disclosures) are encoded as 0 or 1, while continuous indicators (e.g. sentiment polarity or bid–ask spread proxies) are linearly normalised. The modularity of HTREx allows components to be extended, replaced, or recalibrated without affecting the overall structure. For example, additional sentiment sources or alternative security metrics may be incorporated as data availability improves. Similarly, the weighting mechanism supports scenario-based analysis by enabling systematic exploration of how trust rankings change under different user priorities.

2.4 Scope and Exchange Selection

To demonstrate the applicability of HTREx, four exchanges were selected for evaluation: Kraken, Coinbase, Binance, and Uniswap. The selection reflects diversity in operational models, regulatory posture, and market structure. Kraken, Coinbase, and Binance represent large custodial exchanges, while Uniswap represents a decentralised non-custodial exchange. This enables comparative analysis across centralised and decentralised paradigms and illustrates how differences in compliance, security, and incident exposure influence trust assessment. The framework is intended to demonstrate transparent, component-based evaluation rather than provide definitive exchange rankings.

3. Sentiment and Behavioural Risk Assessment

User sentiment represents a behavioural dimension of trust that is largely absent from existing exchange ranking frameworks, despite evidence that public perception and collective experience influence platform adoption, reputational risk, and market behaviour (Garcia, et al., 2014) (Wang, et al., 2022). In this framework, sentiment is treated as a measurable proxy for user satisfaction and perceived reliability, complementing technical and regulatory indicators rather than replacing them.

3.1 Data Sources

Sentiment data were collected from mobile application reviews on the Google Play Store and discussion posts from Reddit. These sources capture complementary perspectives: app reviews reflect direct user experience with exchange functionality, while Reddit discussions capture broader community reactions to outages, regulation, and security incidents. Garcia, et al (2014)) showed that such informal online content can reveal early signals of operational risk and shifts in user confidence. For each exchange, a fixed number of recent reviews and discussion posts were collected to ensure comparability. Textual data were pre-processed through lowercasing, URL removal, and language filtering.

3.2 Sentiment Classification Method

Sentiment classification was performed using a RoBERTa-based transformer model fine-tuned for short, informal text (Barbieri *et al.*, 2020). Transformer models were selected over rule-based approaches such as VADER due to their ability to capture contextual nuance, sarcasm, and mixed sentiment common in cryptocurrency discourse (Hutto & Gilbert, 2014) (SuperAnnotate, 2023). For each text item, the model produced a continuous sentiment score representing positive, neutral, or negative sentiment. Individual predictions were aggregated at source level and subsequently combined into a single exchange-level sentiment score.

3.3 Platform-Level Sentiment Aggregation

To combine sentiment information across multiple sources, a weighted aggregation approach is employed. The final sentiment score is defined as:

$$S_{\text{Sentiment}} = \sum_{k=1}^{n} \alpha_k \cdot P^{(k)}, \alpha_k \in [0,1], \sum_{k=1}^{n} \alpha_k = 1$$

where $P^{(k)}$denotes the aggregated sentiment score derived from source k, and α_k represents the relative importance assigned to that source, and there is a total of *n* sources. The framework can accommodate additional sentiment sources without altering its overall structure. The resulting sentiment score is linearly normalised to the interval [0,1], where higher values indicate more positive user perception and lower behavioural risk.

3.4 Limitations

Sentiment scores revealed meaningful variation in user perception that did not always align with market-centric rankings. However, sentiment analysis captures expressed opinion rather than objective risk and may be influenced by short-term events, highly vocal user groups, or domain-specific language use. Sentiment should therefore be interpreted as a complementary behavioural signal rather than a standalone trust indicator.

4. Compliance and Regulatory Assessment

The compliance component of HTREx evaluates regulatory disclosures, licensing claims, and the implementation of know-your-customer (KYC) and anti-money laundering (AML) controls. Regulatory compliance represents a core dimension of exchange trust, particularly for custodial platforms operating across multiple jurisdictions. Compliance indicators provide insight into financial crime prevention, regulatory accountability, and user protection, but such information is often embedded within lengthy and unstructured legal documents. To address this challenge, HTREx extracts structured compliance indicators from exchange Terms of Service (ToS) documents. The prioritisation of KYC and AML measures aligns with international guidance identifying financial crime prevention as a central component of virtual asset oversight (Financial Action Task Force, 2025).

4.1 Compliance Indicators and Scoring Model

Compliance is operationalised through three core dimensions: financial crime prevention, regulatory accountability, and user protection. These dimensions are represented through a set of binary indicators reflecting widely recognised regulatory and governance practices. Specifically, financial crime prevention is captured through the explicit presence of Know Your Customer (KYC) and Anti-Money Laundering (AML) measures. Regulatory accountability is assessed through disclosure of operating jurisdictions and references to regulatory authorities. User protection is represented through references to data protection obligations and explicit statements regarding user fund protection.

Each indicator is encoded as being either present (1) or absent (0). Indicators within each dimension are averaged and combined into a single compliance score using a weighted formulation:

$$S_{\text{Compliance}} = 0.4 \cdot \frac{\text{KYC} + \text{AML}}{2} + 0.3 \cdot \frac{\text{Regulators Mentioned} + \text{Jurisdictions}}{2} + 0.3 \cdot \frac{\text{Data Protection} + \text{User Fund Protection}}{2}$$

This weighting prioritises financial crime prevention, reflecting its central role in international regulatory frameworks, while retaining visibility of governance transparency and user safeguards. The resulting compliance score is normalised to the interval [0,1], with higher values indicating stronger observable regulatory alignment.

4.2 Automated Extraction Using Large Language Models

To extract compliance indicators at scale, a GPT-family large language model was used to analyse exchange ToS documents. Large language models were selected due to their ability to interpret complex legal language and perform structured information extraction without task-specific retraining (Kumar, 2025). A constrained output schema ensured consistent extraction of predefined compliance indicators for integration into the scoring pipeline. Although LLM-based extraction reduces manual effort, potential errors such as omission or misinterpretation of legal clauses remain possible. Limited human oversight was therefore applied to high-impact indicators related to regulatory accountability and user fund protection.

4.3 Empirical Compliance Outcomes

The application of the compliance component to the evaluated exchanges revealed clear structural differences between centralised and decentralised platforms. Custodial exchanges operating within regulated jurisdictions achieved high compliance scores, reflecting explicit disclosure of KYC, AML, and data protection practices. In contrast, the decentralised exchange exhibited lower compliance scores, primarily due to the absence of formal identity verification requirements and explicit user fund protection mechanisms, despite partial acknowledgment of regulatory considerations. These findings underscore the value of treating compliance as a distinct trust dimension rather than conflating it with market activity or liquidity. The findings also illustrate how exchange architecture and custody models influence observable compliance characteristics—an effect that is often obscured in aggregate exchange rankings (Ong, 2020) (CoinMarketCap, 2024).

5. Security, Transparency and Incident History

While sentiment and compliance capture behavioural and regulatory aspects of exchange trust, technical security, operational transparency, and historical performance provide critical insight into an exchange's capacity to safeguard user assets over time. These dimensions represent both preventative controls and realised risk and are therefore treated as distinct but complementary components within HTREx.

5.1 Technical Security Assessment

Technical security reflects an exchange's ability to protect its infrastructure and user assets against unauthorised access, data compromise, and systemic failure. Security maturity is assessed using a set of commonly disclosed indicators that reflect prevailing industry practices and custodial risk management norms. These indicators include the availability of two-factor authentication (2FA), the existence of a public bug bounty programme, evidence of cold wallet asset storage, secure communication standards as indicated by SSL/TLS configuration, the presence of third-party security audits, and disclosure of internal security practices. Indicators are encoded either as binary variables (present or absent) or as normalised continuous values, depending on the nature of the metric. The overall security score is computed as the arithmetic mean of all available indicators:

$$S_{\text{Security}} = \frac{\text{2FA} + \text{Bug Bounty} + \text{Cold Wallet Storage} + \text{SSL/TLS} + \text{Audit} + \text{Security Disclosures}}{N}$$

where N denotes the number of security indicators considered. This formulation ensures comparability across exchanges while accommodating partial disclosure or data availability. Although the model does not capture all aspects of an exchange's internal security posture, it provides a transparent and reproducible approximation suitable for cross-platform analysis.

5.2 Transparency

Transparency reflects the degree to which an exchange provides verifiable information regarding its operations, reserves, and market behaviour. Rather than relying on region-specific trust badges or proprietary labels, transparency in this framework is operationalised through globally observable indicators: public proof-of-reserves disclosures, accessibility of audit reporting, and observable market liquidity.

The transparency score is calculated as:

$$S_{\text{Transparency}} = \frac{\text{Proof of Reserves} + \text{Audit Reporting} + \text{Bid–Ask Spread}}{N}$$

where all indicators are encoded as binary or normalised values and N represents the number of transparency indicators. Liquidity is proxied through bid–ask spread normalisation, with narrower spreads indicating higher liquidity, improved price discovery, and greater operational transparency. This approach aligns with methodologies employed by major exchange aggregators while retaining methodological transparency and reproducibility (Ong, 2023). Transparency indicators within exchange ranking systems frequently incorporate liquidity-related measures, including trading volume and bid–ask spread. Bid–ask spread is widely recognised as a practical indicator of market liquidity and execution efficiency, particularly for active traders and short-term market participants, as it reflects transaction cost and market depth (Cryptopedia, 2025). However, such indicators provide limited insight into custodial risk, governance practices, or regulatory exposure, underscoring the need to interpret liquidity-based trust signals with caution. The growing emphasis on proof-of-reserves disclosures reflects broader industry efforts to restore confidence following a series of high-profile custodial failures. While such disclosures improve transparency regarding asset backing, their implementation remains uneven and lacks consistent standardisation across exchanges, limiting their comparability as standalone trust indicators (CoinGecko, 2023).

5.3 Incident History

Historical incidents provide insight into realised operational and governance risk that cannot be inferred from preventative controls alone. Security breaches, regulatory enforcement actions, and prolonged service disruptions may have lasting effects on trust, even where current controls appear robust. To capture both severity and recency, the framework employs a decay-weighted incident scoring model, where each incident is weighted by its impact and reduced over time according to an exponential decay function $\text{f(t)} = e^{-kt}$. This ensures that recent, severe events influence trust more strongly than older or minor incidents. The approach is adapted from standard risk management practices and provides a transparent and interpretable mechanism for incorporating both the severity and temporal relevance of historical evidence.

The incident score is defined as:

$$S_{\text{Incident}} = 1 - \frac{\sum_{i=1}^{n}(w_i \cdot f(t_i))}{W_{\text{max}}}$$

where w_i represents the severity weight of incident i, t_i denotes the time elapsed since the incident occurred, and $W_{\max}$ is a normalisation constant ensuring that the resulting score lies within the interval $[0,1]$. Temporal decay is modelled using an exponential function:

$$f(t_i) = e^{-kt}$$

where k controls the rate at which the influence of historical incidents diminishes over time. This formulation reduces the impact of older incidents while preserving the significance of recent failures, enabling the framework to distinguish between short-term operational risk and long-term reputational recovery.

Together, the security, transparency, and incident components provide a balanced view of preventative controls, observable openness, and historical outcomes. Centralised custodial exchanges generally score higher on security due to formal access controls and custody practices, while decentralised platforms often exhibit greater transparency but face structural limitations in compliance and incident mitigation. Importantly, the incident component functions as a corrective mechanism, ensuring that strong present-day controls do not fully offset a history of severe operational failures. By maintaining separation between these dimensions rather than collapsing them into a single opaque metric, HTREx enables a nuanced interpretation of trust and risk trade-offs. This separation also supports scenario-based weighting, allowing users to prioritise preventative security, transparency, or historical performance depending on their risk tolerance and time horizon.

6. Composite Trust Scoring

This section presents the aggregation of component scores into a composite trust value and discusses the empirical outcomes observed across the selected exchanges. The objective is not to establish definitive rankings, but to demonstrate how a transparent, component-based framework produces interpretable and context-sensitive trust assessments that differ meaningfully from existing industry scores.

6.1 Trust Score Aggregation

The final trust score is computed as a weighted aggregation of the five component scores introduced in Sections 3–5. This formulation reflects the multi-dimensional nature of exchange trust while allowing prioritisation to vary according to user preferences and risk tolerance.

The composite trust score is defined as:

$$T_{\text{Trust}} = w_1 S_{\text{Sentiment}} + w_2 S_{\text{Compliance}} + w_3 S_{\text{Security}} + w_4 S_{\text{Transparency}} + w_5 S_{\text{Incident}}$$

subject to the constraint:

$$\sum_{j=1}^{5} w_j = 1$$

where $S_{\text{Sentiment}}, S_{\text{Compliance}}, S_{\text{Security}}, S_{\text{Transparency}}$, and S_{Incident} represent the normalised component scores, and w_j denotes the corresponding weights. To enable baseline comparison, a default weighting configuration was defined based on the relative importance of regulatory compliance, asset protection, and user experience as reflected in prior exchange assessment practices (Ong, 2020):

$$w = \{w_1 = 0.20,\ w_2 = 0.25,\ w_3 = 0.25,\ w_4 = 0.15,\ w_5 = 0.15\}$$

This configuration places greatest emphasis on compliance and technical security while retaining behavioural sentiment and historical performance as meaningful contributors. The weighting scheme is not claimed to be optimal; rather, it serves as a defensible reference point for comparative analysis and illustrates the flexibility of HTREx.

6.2 Exchange Level Results

Table 1 presents the component scores, composite trust values, and resulting relative rankings for the evaluated exchanges under the default weighting configuration defined in Section 6.1. The table provides a compact empirical summary of the framework's output and illustrates how individual trust dimensions contribute to overall platform assessment.

Table 1: Component Scores, Composite Trust Scores, and Relative Rankings

Platform	Sentiment $S_{\text{Sentiment}}$	Compliance $S_{\text{Compliance}}$	Security S_{Security}	Transparency $S_{\text{Transparency}}$	Incident S_{Incident}	Final Risk/Trust T_{trust}	Ranking of the Platforms
Kraken	0.158	1	0.8333	0.943	0.329	0.735	1
Coinbase	0.0915	1	0.8333	0.448	0.558	0.646	2
Binance	0.439	0.85	0.8333	0.445	0.796	0.617	3
Uniswap	0.2125	0.45	0.6667	0.543	0.996	0.510	4

The results highlight clear structural distinctions between centralised and decentralised platforms. Kraken achieved the highest composite trust score, driven primarily by comprehensive compliance disclosures and high transparency, despite comparatively low sentiment and a moderate incident score.

Coinbase ranked second, reflecting similarly strong compliance and security scores, but received a lower transparency score and weaker sentiment during the evaluation period. Binance exhibited a contrasting profile, combining high sentiment and a strong incident recovery score with comparatively lower compliance and transparency. These trade-offs resulted in a lower composite trust value under the default weighting configuration despite strong market presence. Uniswap, as a decentralised exchange, displayed a markedly different trust structure. Its near-maximal incident score reflects the absence of major custodial breaches, while its transparency score exceeds that of some centralised platforms. However, lower compliance and security scores—stemming from its non-custodial architecture and limited regulatory alignment—reduced its overall trust score under the default configuration. It is important to emphasise that the rankings presented in Table 1 are relative and configuration dependent. They should be interpreted as illustrative outputs of the framework rather than definitive or universal statements regarding exchange trustworthiness.

6.3 Sensitivity to Weight Adjustment

A key advantage of the proposed framework is its support for sensitivity analysis through explicit weight adjustment. Increasing the weight assigned to compliance and security amplifies the advantage of regulated custodial exchanges, while increased emphasis on transparency or sentiment narrows this gap and, in some cases, alters the relative ordering of platforms. This behaviour demonstrates that trust is not a fixed or objective property, but a context-dependent judgement shaped by user priorities, regulatory exposure, and risk tolerance.

To illustrate this effect more concretely, Table 2 summarises a set of alternative weighting profiles and the resulting changes in relative exchange rankings. All scenarios presented in Table 2 use identical underlying component scores; only the weighting configuration is modified. The default profile reflects a balanced assessment, while the alternative profiles place increased emphasis on security, compliance, or transparency and sentiment, respectively.

Table 2: Sensitivity of exchange rankings under alternative weighting profiles

Weighting profile	Sentiment w_1	Compliance w_2	Security w_3	Transparency w_4	Incident w_5	Resulting ranking
Default (baseline)	0.20	0.25	0.25	0.15	0.15	Kraken > Coinbase > Binance > Uniswap
Security focused	0.10	0.15	0.40	0.10	0.25	Binance > Coinbase > Kraken > Uniswap
Compliance focused	0.10	0.35	0.25	0.20	0.10	Binance > Coinbase > Kraken > Uniswap
Transparency/Sentiment focused	0.30	0.10	0.20	0.30	0.10	Kraken > Binance > Uniswap > Coinbase

The results shown in Table 2 confirm that no single exchange is universally optimal. Custodial exchanges such as Kraken and Coinbase consistently perform well when compliance is prioritised, while Binance improves its relative standing under security- or sentiment-focused configurations. Uniswap performs comparatively better in scenarios that de-emphasise compliance and custodial guarantees. These shifts reinforce the importance of exposing weighting assumptions and stand in contrast to proprietary exchange trust scores, which typically embed such assumptions implicitly. By making the consequences of weighting choices explicit, as demonstrated in Table 2, the framework supports informed decision-making and reduces the risk of misplaced reliance on single-score rankings—an issue that remains largely unaddressed in existing exchange trust metrics.

When benchmarked against industry trust indicators, the proposed framework produces broadly aligned results for highly regulated exchanges but diverges in cases where sentiment, transparency, or incident history plays a significant role. Concerns have also been raised regarding the preferential treatment of certain trading pairs and exchanges within ranking systems, further complicating the interpretation of aggregate trust scores and reinforcing the limitations of market-centric metrics (Thompson, 2020). These divergences highlight the explanatory value of a multi-dimensional and configurable trust assessment approach. These divergences highlight the limitations of market-centric metrics and demonstrate the added explanatory value of incorporating sentiment, compliance transparency, and historical performance. Importantly, the framework is not intended to replace existing data aggregators. Instead, it complements them by offering a transparent, configurable alternative that prioritises methodological openness and interpretability. This positioning aligns with growing calls for greater transparency in cryptocurrency risk assessment and exchange evaluation methodologies (Simon, 2023).

7. Discussion

The results demonstrate that a modular, transparent approach to exchange trust assessment can reveal dimensions of risk that are obscured by existing monolithic scores. By explicitly separating sentiment, compliance, security, transparency, and incident history, the framework supports more nuanced interpretation and enables users to align trust evaluations with their own risk priorities. This section discusses the broader implications of these findings and outlines key limitations.

7.1 Interpretability and Practical Relevance

A central contribution of this work lies in its emphasis on interpretability. Each component score is derived from observable indicators and clearly defined assumptions, reducing reliance on implicit or proprietary methodologies. This is particularly relevant in the cryptocurrency domain, where trust is frequently inferred from market activity rather than substantiated operational or regulatory evidence. The ability to adjust component weights further reinforces the view that trust is context-dependent rather than absolute. From a practical perspective, the framework is applicable to both retail and institutional users. Retail participants may prioritise sentiment and usability, while institutional actors may emphasise compliance and security. By accommodating both perspectives within a single structure, the framework addresses a gap between user-centric perception and institutionally defined risk criteria.

Several limitations must be acknowledged. First, data availability and quality vary significantly across exchanges. Some indicators rely on self-disclosed information, which may be incomplete or selectively reported. Although this limitation is inherent to public-data-driven assessments, it underscores the need for cautious interpretation of results. Second, sentiment analysis is sensitive to short-term events and platform-specific user behaviour. While transformer-based models mitigate some limitations of earlier approaches, they cannot fully resolve issues such as sarcasm, coordinated review manipulation, or sudden sentiment shocks following high-profile incidents. Consequently, sentiment should be interpreted as a complementary signal rather than a definitive measure of trust. Third, the use of large language models for compliance extraction introduces potential risks related to misinterpretation or hallucination. Although constrained output schemas and human oversight reduce these risks, the approach remains probabilistic rather than deterministic. Future work could explore hybrid rule-based and model-based extraction to further improve reliability.

The framework was evaluated on a limited set of exchanges selected to illustrate structural diversity rather than market coverage. As such, the empirical results should not be generalised to all platforms without further validation. Nonetheless, the framework itself is exchange-agnostic and can be extended to additional platforms as data availability permits. The study also focuses on exchange-level trust rather than asset-specific or transaction-level risk. Integrating on-chain analytics or smart contract verification could enhance the framework's applicability to decentralised finance contexts, but such extensions fall outside the scope of the present work.

8. Conclusion

This paper presented HTREx, a semi-automated, modular framework for assessing trust and risk in cryptocurrency exchanges. By integrating behavioural sentiment, regulatory compliance, technical security, transparency, and incident history into a transparent and user-configurable scoring model, HTREx addresses key limitations of existing industry trust scores. The empirical application of the framework to four prominent exchanges demonstrated that meaningful differences in trust assessment emerge when behavioural and regulatory dimensions are explicitly considered. Rather than proposing a definitive ranking, HTREx emphasises interpretability and adaptability, recognising that trust is inherently context sensitive. As cryptocurrency markets continue to mature and regulatory scrutiny intensifies, such transparent and extensible assessment models may play an important role in supporting informed participation and reducing reliance on opaque metrics.

Acknowledgements

This work is based on the research supported in part by the National Research Foundation of South Africa (Reference Number: RCHDI241217296940)

AI Declaration: Artificial intelligence tools were used in a limited and transparent manner during the development of this paper. A LLM from the GPT family was employed to support the automated extraction of compliance indicators from exchange ToS documents, using a constrained output schema. AI generated outputs were only used as intermediate structured inputs, with post processing and limited human verification. The entire paper was written by the authors and no tools were used to produce any part of the paper.

Ethics Declaration: Ethical Clearance was not required for this research.

References

Barbieri, F., Camacho-Collados, J., Espinosa Anke, L. & Neves, L., 2020. TweetEval: Unified Benchmark and Comparative Evaluation for Tweet Classification. *Findings of the Association for Computational Linguistics: EMNLP 2020,* pp. 1644-1650.

Broder, F., 2025. *Social Sentiment Analysis: Strategies, Techniques and Insights.* [Online] Available at: https://www.revuze.it/blog/social-sentiment-analysis-strategies-techniques-and-insights/ [Accessed 10 August 2025].

CoinGecko, 2023. *Methodology.* [Online] Available at: https://www.coingecko.com/en/methodology [Accessed 10 April 2025].

CoinMarketCap, 2024. *Exchange Ranking.* [Online] Available at: https://support.coinmarketcap.com/hc/en-us/articles/360052030111-Exchange-Ranking [Accessed 21 June 2025].

Cong, L. W., Li, X., Tang, K. & Yang, Y., 2022. *Crypto Wash Trading,* Cambridge, MA: National Bureau of Economic Research.

Cryptopedia, 2025. *Cryptocurrency Pairs: Trading & How They Work.* [Online] Available at: https://www.gemini.com/en-SG/cryptopedia/what-are-cryptocurrency-pairs [Accessed 10 June 2025].

Fenton, A., 2020. *CoinMarketCap Initially Rejected the Metric That Now Puts Binance at No.1.* [Online] Available at: https://cointelegraph.com/news/the-vc-who-argued-death-spiral-scenario-may-be-right-after-all [Accessed 20 June 2025].

Financial Action Task Force, 2025. *Guidance on Financial Inclusion and Anti-Money Laundering and Terrorist Financing Measures,* Paris: Financial Action Task Force.

Fitzpatrick, L., 2019. *CoinMarketCap Algorithm Changed: Is The Ranking Flawed?.* [Online] Available at: https://www.forbes.com/sites/lukefitzpatrick/2019/08/20/coinmarketcap-algorithm-changed-is-the-ranking-flawed/[Accessed 4 July 2025].

Garcia, D., Tessone, C. J., Mavrodiev, P. & Perony, N., 2014. The digital traces of bubbles: feedback cycles between socio-economic signals in the Bitcoin economy. *Journal of the royal society interface,* 11(99).

Hutto, C. & Gilbert, E., 2014. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media,* 8(1), pp. 216-225.

Kumar, R., 2025. *Day 5: Most Used LLMs in 2025 — Who's Leading the AI Revolution?.* [Online] Available at: https://medium.com/@rajukumardalimss/day-5-most-used-llms-in-2025-whos-leading-the-ai-revolution-12956f983301[Accessed 15 May 2025].

Le Pennec, G., Fiedler, I. & Ante, L., 2021. Wash trading at cryptocurrency exchanges. *Finance Research Letters,* 43(1544-6123), p. 101982.

Ong, B., 2020. *Trust Score Explained.* [Online] Available at: https://blog.coingecko.com/trust-score-explained/ [Accessed 10 June 2025].

Ong, B., 2023. *Trust Score 3.0 – Proof of Reserves (Assets & Liabilities).* [Online] Available at: https://blog.coingecko.com/trust-score-3-proof-of-reserves-asset-liabilities/ [Accessed 12 June 2025].

Simon, 2023. *CoinGecko or CoinMarketCap (Which is Best?).* [Online] Available at: https://hellocrypto.com/article/coinmarketcap-vs-coingecko/ [Accessed 7 June 2025].

SuperAnnotate, 2023. *Sentiment analysis explained 2024.* [Online] Available at: https://www.superannotate.com/blog/sentiment-analysis-explained [Accessed 20 June 2025].

Thompson, P., 2020. *CoinMarketCap updated system appears to rank trading pairs in Binance favor.* [Online] Available at: https://coingeek.com/coinmarketcap-updated-system-appears-to-rank-trading-pairs-in-binance-favor/ [Accessed 22 June 2025].

Vidal-Tomás, D., 2022. Which cryptocurrency data sources should scholars use?. *International Review of Financial Analysis,* Volume 81.

Wang, Q. Zhang, W, Li, J, Mai, F & Ma, Z., 2022. Effect of online review sentiment on product sales: The moderating role of review credibility perception. *Computers in Human Behavior,* Volume 133.