# Beyond the Fiat Vault: Evaluating Machine-Readable Blockchain Disclosures for Systemic Safety, Operational Security, and MEV Conflict Mitigation.

Imane El Imami[1,2], Jeonghoon Oh[1], Jason Meyers[3], Gerard Rod Brennan[1], Miklos Vasarhelyi[1], Alexander Sannella[1], Thomas Egan[3], Abdelkader El Alaoui[4], Bassma Guermah[5],
Said Ouatik El Alaoui[2]

[1]*Rutgers Continuous Auditing and Reporting Lab (CARLab), Rutgers Business School, Rutgers University, New Jersey, United States*
[2]*Engineering Sciences Laboratory, National School of Applied Sciences, Ibn Tofail University, Kenitra, Morocco*
[3]*Auditchain Labs AG*
[4]*Rabat Business School, International University of Rabat, Morocco*
[5]*TICLab Research Laboratory, International University of Rabat, Rabat 11103, Morocco*

*imane.elimami@uit.ac.ma*
*ie138@scarletmail.rutgers.edu*
*jo610@scarletmail.rutgers.edu*
*jm@auditchain.com*

**Abstract**

The Guiding and Establishing National Innovation for U.S. Stablecoins (GENIUS) Act, establishes the first comprehensive federal framework, stipulating that the stability of digital assets, such as stablecoins, is primarily determined by the adequacy of their backing fiat reserves. This paper challenges this structural misconception which we term the "Illusion of the Vault", by demonstrating that digital asset stability is in fact a joint reserve-and-infrastructure problem. Even fully backed digital assets become functionally inaccessible to users if the underlying blockchain infrastructure (Base Layer and Layer 2) fails due to network congestion, cross-chain bridge exploits, smart contract vulnerabilities, predatory third-party Maximal Extractable Value (MEV) front-running and transaction reordering attacks. To address this critical regulatory gap, this paper proposes a comprehensive, machine-readable 148 input question, 125 output disclosure element Blockchain Network Participation (BNP) disclosure taxonomy formatted in XBRL-JSON. To empirically validate this taxonomy we deployed a multi-agent AI methodology utilizing three independent pre-trained Large Language Models (Claude Sonnet 4.6, Gemini 3.1, and GPT-5). The empirical consensus, rigorously verified by robust Gwet's AC1 inter-model agreement scores, overwhelmingly rejected the minimized framework. The findings mathematically confirm that robust infrastructure and MEV conflict disclosures are non-deferrable priorities; without them, traditional fiat-reserve audits remain systematically insufficient to ensure true operational security and investor protection.

**Keywords:** Digital Asset, GENIUS Act, Stablecoins, Blockchain Network Participation, LLMs, Maximal Extractable Value, XBRL, Continuous Auditing, Illusion of the Vault, Cross-Chain Bridges

# 1 Introduction

The digital asset ecosystem has reached a regulatory turning point where regulators will begin to define the rules for the industry. The Guiding and Establishing National Innovation for U.S. Stablecoins Act, often known as the GENIUS Act, is the first federal law to regulate the creation and issuance of payment stablecoins in the United States (GENIUS Act, Pub. L. No. 119-27, codified at 12 U.S.C. §§ 5901–5919, enacted July 18, 2025)[1]. The GENIUS Act is now being implemented, with the Federal Deposit Insurance Corporation (FDIC), the Office of the Comptroller of the Currency (OCC), and related agencies working to establish rules to implement the law. However, the GENIUS Actis intended to operate in alignmentwith other supplementary federal laws addressing digital assets, such as the Digital Asset Market Clarity Act (CLARITY Act). While the combined effect of these statutes establishes a broad congressional mandate for the robust oversight of digital assets, the law explicitly delegates the operational determination of reporting requirements to regulatory agencies. Consequently, fulfilling this legislative mandate and safeguarding the public use of digital assets relies entirely on these agencies exercising their delegated authority to implement structured, transparent, and verifiable data architectures.

This paper has a broad scope. The proposed and tested disclosure framework crucially applies to any digital asset issuer (e.g., banks, fintech companies, a protocol development organization, or another type of entity that utilizes its own validators for a blockchain network or enters into an agreement with a third-party service provider). However, what matters is not how the issuer of the digital asset is chartered, but how dependent the issuer is on the blockchain network infrastructure. Any entity whose digital asset's functionality, redemption capability, or par-value maintenance depends on the continued operation of blockchain network infrastructure primarily faces the same category of Layer 0 and Layer 2 risks, regardless of its regulatory classification. Those risks are best described as the "Illusion of the Vault."

Traditional financial regulation, and specifically the GENIUS Act's reserve-adequacy framework, assumes that the stability of a digital asset is primarily determined by the quality and availability of its backing reserves. Therefore, if a stablecoin has sufficient 1:1 fiat currency backing in the vault, undergoes regular independent audits, and maintains sufficient high-quality liquidity to meet withdrawal requests, it is considered safe. This assumption is not only incomplete; it is demonstrably wrong in the context of stablecoins, simply because such assets cannot be forced to behave like traditional bank deposits. A stablecoin, or any tokenized instrument, is a programmable payment obligation that can only be redeemed when executed via smart contract logic on a functioning blockchain network. If the digital pipes in the blockchain infrastructure freeze or are compromised due to network congestion in processing transactions, a cross-chain bridge hack, or a group of validator/miner operators extracting as much value as possible from users (MEV) to maximize their earnings, the dollars in the vault do not matter, given that they are functionally inaccessible to the token holder. The vault may be full, but the redemption mechanism is dysfunctional, hence the peg is gone.

While this is not hypothetical, Aronoff et al. 2026 provide evidence-based analysis demonstrating that even compliance with the reserve requirements specified in the GENIUS Act will not guarantee that a stablecoin maintains its price stability if problems arise in operating the blockchain network. The March 2023 depegging of USDC following the collapse of Silicon Valley Bank demonstrated that the most conservatively reserved stablecoin in the market at that time lost its peg within hours due to the failure of the institutional architecture supporting its reserve custody, not due to any deficiency in the reserves themselves(Diop et al. 2024). Similarly,

[1] https://www.govinfo.gov/app/details/PLAW-119publ27

(Belenkov et al. 2025) documented that approximately $600 million was destroyed in 2022 as a direct result of a validator key compromise on the Axie Infinity Ronin bridge. Furthermore, an estimation of $550-$650 million was extracted from users of the Ethereum network alone by miners and validators through various forms of MEV, namely, front-running, transaction censorship, and sandwich attacks, describing these actions as illegal front-running by brokers in traditional markets (Auer et al. 2022). Each of these examples represents failures that would be completely invisible to regulators relying on reserve monitoring. In response to these market challenges, the FDIC issued a proposed rulemaking (RIN 3064-AG20)[2] outlining application requirements for stablecoin issuers. However, rather than requiring standardized data, the proposed rule only broadly requests qualitative information regarding "recordkeeping, reconciliation, and transaction processing policies and procedures, including both on- and off-chain procedures." It completely fails to mandate explicit, standardized metrics concerning validator participation, network infrastructure characteristics, or smart contract operational risks. This critical omission leaves a massive supervisory blind spot; relying merely on policy descriptions fails to capture the technical infrastructure vulnerabilities that actually trigger digital asset depegging.

This paper proposes and empirically tests a Blockchain Network Participation (BNP) Disclosure Taxonomy, presented in machine-readable XBRL-JSON format, to resolve this regulatory gap. The framework was developed iteratively through two formal regulatory submissions: an initial comment proposing a 136-question, 107-element architecture (Auditchain Labs AG, Jason Meyers Comment to FDIC RIN 3064–AG20, February 9, 2026), followed by a supplemental submission that expanded the framework to account for third-party risks (Supplemental Comment, Auditchain Labs AG, Jason Meyers Comment to FDIC RIN 3064–AG20, March 9, 2026). The final, comprehensive taxonomy analyzed in this study contains 148 input questions mapped to 125 output XBRL disclosure elements across eight disclosure domains (Blockchain Protocols Utilized, Direct Network Participation, Affiliate Network Participation, Concentration & Systemic Risk, Operational Controls & Policies, Multi-Network Disclosure, Affiliate Disclosure Table, and Third-Party Service Providers). The taxonomy enforces an equal input/equal output relationship. Every disclosure question corresponds exactly to one XBRL element, and vice versa. Therefore, based upon this taxonomy's design characteristics, every omission will be programmatically identifiable without requiring a regulator to intervene, thereby fulfilling the anti-evasion mandate in 12 U.S.C. § 5903(h)(1)[3]. The purpose of this BNP disclosure taxonomy is to ensure that investors can make informed decisions about digital asset products. The framework proposed here could be applied broadly to any regulated digital asset issuer subject to the GENIUS Act and the CLARITY Act, or to any analogous State-regulated stablecoin regime.

To rigorously validate the necessity of the comprehensive 148 input question, 125 output element taxonomy, our research team constructed and evaluated a competitive scientific hypothesis: a minimized, 18-disclosure phased alternative. This hypothetical model, which is adapted from informal structural concepts circulated among working group members, tests the theoretical proposition that early regulatory adoption could be facilitated by limiting initial implementation to 18 basic input questions while indefinitely deferring the remaining 130. However, it is critical to recognize that regulatory agencies historically deploy comprehensive taxonomy architectures upon enactment; they generally utilize phased implementation only to manage the population of adopting filers, not to fragment the underlying data model itself. Throughout this research, we evaluate this minimized alternative not as an official regulatory proposal, but as a methodological baseline. Section 4.1 details how this minimized approach theoretically functions, while subsequent sections present empirical evidence demonstrating that restricting

[2] https://www.fdic.gov/federal-register-publications/comments-rin-3064-ag20-december-19-2025
[3] https://www.law.cornell.edu/uscode/text/12/5903

the taxonomy to 18 input questions completely eliminates regulators' ability to achieve effective oversight.

To test whether our proposed taxonomy adequately protects investors we employed a multi-agent AI methodology in which three Large Language Models (LLMs): Claude Sonnet 4.6, Gemini 3.1, and GPT-5 were independently tested to determine whether they agreed on classifications among all 148 taxonomy elements with respect to GENIUS Act's three primary supervisory mandates: i) investor safety ii) operational soundness iii) avoiding potential conflicts of interests particularly those associated with MEV attacks. The global inter-model Gwet AC1 correlation coefficients across the full 148 input question, 125 output element data set showed a significant level of agreement, ranging from 0.653 to 0.696, confirming that "High-Priority" classifications are analytically robust across significantly disparate AI architectures. Collectively, none of the three models classified more than five taxonomy elements as Low-Priority, severely undercutting the viability of the 18-element minimalist approach. Throughout this paper, the term "Comprehensive Taxonomy" refers to the full 148 input question, 125 output element disclosure proposal, and the term "minimized alternative" refers to the 18-disclosures hypothetical baseline.

Our paper includes three research questions (RQs):

- RQ1: Do digital asset stabilities rely on the operational integrity of blockchain infrastructure in addition to reserve quality across all types of issuers, and what are the mechanisms of failure?
- RQ2: Is our empirically validated 148 input question, 125 output element BNP Disclosure Framework justified as a regulatory floor under GENIUS Act's safety, soundness, and anti-evasion mandates?
- RQ3: Will minimizing our disclosure framework to 18 input questions create significant supervisory blind spots concerning bridge custody and third-party MEV extractions that deny consumers structural information necessary for informed use decision-making?

The remainder of this paper will proceed as follows: Section 2 synthesizes the existing literature across three streams: financial contagion arising from unstable run dynamics in stablecoins, infrastructure vulnerabilities, and MEV risks. Section 3 presents the dimensional structure of the 148 input question, 125 output element BNP taxonomy. Section 4 explains our multi-agent AI methodology for testing our proposed taxonomy and provide an objective account detailing the rationale for minimizing phased approaches; Section 5, presents our empirical results and findings. Section 6 includes an advanced discussion with limitations and future research directions, and finally, we present our conclusions in Section 7.

## 2 Literature Review

Our paper's central argument draws upon research across three academic fields, namely, financial economics, systems security, and accounting information systems, which typically operate independently and do not overlap in terms of methodology or regulatory implications. Researchers in financial economics have modeled the dynamics of runs and contagion without addressing how these dynamics occur through the blockchain infrastructure required for redemptions. In the current literature, computer researchers have detailed the risks of MEV extraction and bridge exploits, but have not referenced a regulatory standard to which their work should contribute. Accounting researchers have developed ongoing auditing architectures suitable for high-speed data environments, primarily in scenarios where the underlying transactional rail is present. While these research streams tackle structurally related phenomena, they

systematically overlook the joint problem our paper addresses. Read together, they reveal not three parallel research streams but a single integrated mandate for continuous, machine-readable disclosure of blockchain infrastructure.

### 2.1 The Limitations of Reserve-Centric Stability: The Illusion of the Vault

The predominant regulatory framework governing the stability of digital assets (the GENIUS Act) relies upon a deposit insurance analogy. If there are adequate amounts of assets to support a stablecoin's peg, and if independent attestation verifies the adequacy of the backing assets, then it is considered to be stable. However, the financial economics literature regarding the fragility of stablecoins has demonstrated that this analogy is flawed in ways that reserve attestations fail to capture. Catalini and Gortari (2021) established fundamental theoretical results on the conditions under which peg defenses become infinitely expensive once redemption expectations deteriorate, regardless of the quality of the reserves. Their model shows that stablecoin stability is self-referential: the peg will remain intact so long as participants expect it will. While it establishes the necessity of adequate reserves to support a stablecoin, it also highlights that reserves are merely a necessary condition for stability, but not a sufficient one.

Numerous empirical studies have validated this theoretical prediction. (Diop et al. 2024) using Machine Learning models, (e.g., Neural Networks, Gradient boosting, Random Forest, Ada boosting) on daily cryptocurrency data covering October 2022 – November 2023 to show that USD Coin which is the most conservatively reserved and transparently managed stablecoin at the time, was the most vulnerable and susceptible instrument to devaluing in response to the March 2023 Silicon Valley Bank (SVB) stress event. As such, the depegging of USDC was not due to any lack of reserves; rather, it was due to the collapse of the institutional architecture that supported reserve custody, namely, a banking counterpart whose balance sheet was inextricably tied to the backing reserves of USDC. Furthermore, the causal mechanism responsible for the depegging was located outside of the vault, thereby being invisible to reserve attestations, and resulted in catastrophic consequences. Similarly, Wen and Lau (2025) stress-tested a hybrid monetary architecture using parameters corresponding to the same 2023 SVB event, formally stating that par-value maintenance requires a stable redeeming rail in addition to adequate reserves. Wu and Liu (2026) empirically demonstrate that during periods of extreme stress, direct volatility channels emerge between the U.S. Dollar Index and Bitcoin that bypass stablecoin intermediation, using Quantile Vector Autoregressions across eight major stablecoins between 2021 and 2025.

Adrian et al. (2025) analyzing the threat posed by rapid scale-up in stablecoin usage under the GENIUS Act to global financial stability in an International Monetary Fund analysis concludes that this systemic financial risk is beyond the capacity of reserve frameworks to detect. Azar et al. (2024) wrote for the Federal Reserve Bank of New York to demonstrate that stablecoins perform maturity and liquidity transformations akin to shadow banks, yet without providing prudent safeguards afforded to regulated banks. The authors' characterization supports the need for bank-equivalent disclosure requirements to ensure safety and soundness. Gross and Senner (2026) describe a feedback loop between surge-redemption activity and fire-sale pressure on bonds, arising from the blockage of redemptions, as self-amplifying contagion mechanisms whose severity is determined by both the reserve assets backing the stablecoin and the blockchain infrastructure's capacity to process redemption requests.

Thus, the collective body of evidence from the above-referenced financial economics literature clearly indicates that reserve requirements are insufficient and that redemption mechanisms (their speed, reliability, and structural integrity) are independent determinants of stability. Yet, this literature views redemption throughput solely as an abstract financial concept and therefore

as a model parameter rather than as a function of node counts on blockchain platforms, uptime commitments made by validators, architecture of bridges linking chains, and dependencies in the execution of smart contracts. Eichengreen et al. (2025) quantitatively demonstrated that the risk of stablecoin devaluation exceeded 200 basis points during the Terra-Luna crisis, which is a risk that occurred over days rather than weeks. Clearly, a disclosure system calibrated to periodic reserve audits cannot detect risks evolving at blockchain speeds. MacDonald and Zhao (2023) added an additional layer of complexity: auditing price stability independently ignores the aggregation of leverage and liquidity mismatches within decentralized infrastructure, which can create conditions for future contagion.

### 2.2 Base Layer and Layer 2 Infrastructure & Systemic Vulnerabilities

Substantial bodies of research in computer science and systems security have documented numerous failure modes associated with blockchain infrastructure used to execute digital asset transactions. Research in this area is structurally invisible to financial regulators drafting regulatory frameworks for digital assets, resulting in measurable dollar losses. For example, neither the GENIUS Act nor the proposed application requirements under RIN 3064-AG20 referenced by the FDIC require disclosure of bridge custody topology, validator key management, or smart contract execution dependencies across chains. Nonetheless, asset losses attributable to these vulnerabilities currently exceed the thresholds that are anticipated to be protected against under reserve-only regulatory frameworks.

The empirical record concerning cross-chain bridge vulnerability is extensive and unambiguously negative. Notland et al. (2026) reviewed 64 bridge implementations and cataloged 31 exploits over a three-year period (2021–2023). The authors identified 13 architectural features linked to eight distinct categories of inherent vulnerabilities, including design flaws that constitute properties of bridge architectures rather than characteristics of individual defective implementations. Estimated losses from bridge exploits totaled $1.5- $2 billion in 2022 alone. Zhang et al. (2023) developed a systematic taxonomy for 12 attack surfaces concentrated around two structural weaknesses: 1) centralized smart contract permission structures; and 2) off-chain verification dependencies establishing trust assumptions incompatible with public blockchain security models. The trust model embodied in these bridge architectures structurally equates to counterparty risk managed by banking regulation; however exists completely outside financial supervisory oversight.

Belenkov et al. (2025) described a canonical case: the Axie Infinity Ronin Bridge Exploit of 2022, in which private keys belonging to validators were compromised, enabling unauthorized withdrawals totaling approximately $600 million. Compromising validator key management represented one category of third-party infrastructure risk that no reserve attestation could detect. Liao et al. (2024) developed a static analysis framework, “SmartAxe,” that identifies a class of exploits, Cross-Chain Vulnerabilities (CCVs), that occur specifically at interfaces between chains and are undetectable by single-chain audit methodologies. Their analysis found 232 previously unidentified CCVs across 129 production bridge contracts, representing approximately $1.885 million in immediately at-risk digital assets. Temporally, periodic audits (whether performed monthly as required by attestation schedules included in the GENIUS Act or annually as mandated by conventional financial reporting cycles) operate at temporal resolutions orders of magnitude less frequent than exploit cycles enabled by these vulnerabilities.

Aronoff et al. (2026), have provided the definitive academic synthesis linking this body of research on infrastructure security with financial regulatory frameworks. Analyzing the reserve provisions in the GENIUS Act relative to operational requirements for supporting blockchain transaction rails, they demonstrate that par-value stability depends on the operational reliabil-

ity of blockchain-based transaction rails and advocate for an integrated approach encompassing financial-market infrastructure, prudent regulation, and software governance. While the paper identified a regulatory gap with clarity, it did not provide a disclosure vehicle to address it. Our 148 input question, 125 output element BNP Taxonomy is that disclosure vehicle. The financial stability literature identifies systemic risk categories; however, the infrastructure security literature describes failure modes associated therewith; neither produces disclosure architectures capable of making those failure modes observable to supervisors prior to their realization.

### 2.3 Shadow Intermediation and Maximal Extractable Value: The MEV Conflict

The third group of research studies upon which this paper builds its research agenda holds an especially worrisome regulatory position, as it lies at the intersection of the law regulating the manipulation of financial markets and the architectural design of blockchain infrastructure. Maximal Extractable Value (MEV) refers to the profit miners/validators earn from network users by strategically reordering or including/excluding transactions within a block. It is both a financial phenomenon, manifested in the repeated extraction of value from participants in a market, and a technological phenomenon, resulting from the right to propose a block, mempool visibility, and sequence-order algorithms for transactions. The MEVs resemble the illegal front-running carried out by brokers in traditional markets: when miners/validators recognize a pending, large transaction that will cause significant changes in the price of the relevant market, they can place a comparable transaction immediately prior to the triggering trade, thereby taking advantage of the price movements at the expense of the original user (Auer et al. 2022). The authors characterize this as an "invisible" tax on market participants — i.e., capturing both the mechanics of extracting value, and the broader public harms resulting from system-wide exposure. Estimates suggest that MEV has resulted in $550- $650 million in lost value on the Ethereum network alone since 2020. They further suggest that regulatory agencies worldwide should determine whether MEVs constitute illegal behavior, noting that in many jurisdictions, activities similar to front-running are prohibited.

The game-theoretic structure of MEV was formalized by Kulkarni et al. (2023), who show, in a constant-function market-maker setting, that the price impact of sandwich attacks increases by O(log n) with the number of trades initiated by users. This is not a sporadic or correctible threat; it is a persistent structural tax increasing proportionally with growing transaction volumes— exactly as we would anticipate when GENIUS Act-compliant stablecoins grow to widespread use as mainstream payment settlement instruments. Mancino et al. (2025) have provided the most detailed empirical documentation available thus far: examining 220,993 blocks and 36,015,340 transactions from January 2024 on Ethereum, they demonstrate system-wide ordering of transactions that correlates with communication between users and block builders, and also demonstrate hidden payments made directly to block builders for preferentially placing their favored transactions within a block. While their description of an emergent shadow economy among validators is a factual empirical observation — not mere rhetoric — it clearly describes the economic structure that the GENIUS Act's anti-evasion authority at 12 U.S.C. § 5903(h)(1)[4] was enacted to address, yet which no existing disclosure framework is designed to detect.

Materwala et al. (2025) conducted an exhaustive review of MEV literature in Decentralized Finance (DeFi) and identified the core structural implications for regulatory design: mempool transparency is an inherent characteristic of public blockchain networks, and therefore, MEV is

[4] 12 U.S. Code § 5903 - Requirements for issuing payment stablecoins (2026). en. url: https://www.law.cornell.edu/uscode/text/12/5903

exploitable based on structural factors, rather than being mitigated through conduct regulations alone. Any third-party actor with mempool access – meaning they can see pending transactions prior to public confirmation – possesses the technical capabilities to engage in front-running, back-running, and sandwich attacks against all user-initiated transactions on the network, including redeeming claims. Manipulating the order in which transactions are sequenced poses threats and harms network integrity across both Layer-1 and Layer-2 architectures, making migration to alternative networks impractical for supervisors (Alipanahloo et al. 2024). While centralized solutions to mitigate MEV exist, such as the "Flashbots" relay, they create single points of failure that concentrate control over block creation in opaque intermediaries lacking any present-day disclosure obligations and potentially worsen rather than reduce supervisory problems (Sinai and In 2024). Aramonte et al. (2021) provided the systemic framing: the decentralization illusion prevalent in DeFi governance frameworks means that concentration risks facilitating MEV extractions are not accidental features of specific protocols but structural characteristics of the DeFi ecosystem as presently constructed.

Broker front-running in traditional markets is regulated through a disclosure and surveillance infrastructure: trade reporting, conflict-of-interest disclosures, best execution requirements, and other disclosures that render the relationships between brokers and users executing transactions economically transparent to regulators (Auer et al. 2022). There is presently no similar disclosure infrastructure concerning stablecoin issuers and validators that produce blocks on which those issuers' tokens settle. A validator contracting with a stablecoin issuer to produce blocks on behalf of that issuer possesses a structural mathematical characteristic incentive to extract value from users of that issuer's stablecoin, regardless of any formal legal relationship between them. The potential mechanism for flow-back of MEV revenue to digital asset investors, disguised as compensation for miners' and validator services, is precisely the type of yield evasion activity prohibited under 12 U.S.C. § 5903(a)(11). The TSC Industries materiality standard (426 U.S. 438, 1976)[5] is satisfied as well settled law. Not one of the MEV studies reviewed above contains a reference to the disclosure framework that a prudent regulator would need to detect these arrangements. That disclosure framework is detailed in Section 8 of the 148 input question, 125 output element BNP taxonomy.

### 2.4 The Regulatory Gap and Continuous Auditing Solutions

As noted previously, while each of these three areas represents independent problems in finance regulation, together they represent a common issue, namely, reserve adequacy, infrastructure integrity, and managing conflicts related to MEV, which no current disclosure framework collectively addresses. What each series lacks is not empirical support for these issues but a disclosure architecture capable of exposing its content to supervisors at temporal resolutions commensurate with how blockchain risks emerge.

Theoretical foundations of continuous auditing were laid down by Vasarhelyi and Halper (2018), proposing automated continuous monitoring models for online transaction processing systems in which audit procedures are applied to transactional data as it passes through systems, as opposed to historical samples obtained during periodic audits. The eXtensible Business Reporting Language (XBRL) provides a standardized machine-readable format through which continuous auditing frameworks can be employed for large-scale regulatory disclosures. The Securities and Exchange Commission's (SEC) phased XBRL adoption mandate for financial reporting, starting in 2009[6], created an institutional infrastructure for machine-readable

[5] See TSC Industries, Inc. v. Northway, Inc., 426 U.S. 438, 449 (1976) (information is material if there is "a substantial likelihood that a reasonable [investor] would consider it important" in making a decision) https://supreme.justia.com/cases/federal/us/426/438/

[6] https://www.sec.gov/rules-regulations/2009/01/interactive-data-improve-financial-reporting

financial data: structured tags that enable automated ingestion, comparison, and anomaly detection without manual re-keying of reported values. First-wave mandatory XBRL disclosures significantly reduce market-wide information asymmetry, directly contributing to increased information efficiency in stablecoin markets, where such asymmetry currently exists due to the lack of any disclosure requirement regarding the blockchain risks faced by issuers/users. Blocks are produced every twelve seconds on Ethereum; Bridge transactions occur across chains within minutes; MEV extraction events are observable milliseconds in the mempool prior to confirming blocks. Monthly attestations required under GENIUS Act requirements are five to six orders of magnitude slower than risk events they are attempting to monitor. The implication is not merely that periodic audits are less informative than continuous audits; it is that periodic audits are functionally identical in timing with monitoring fires after the building burns down.

# 3 Design and Architecture of The Comprehensive BNP Framework

The literature review demonstrates that stablecoin stability is a problem of reserves and infrastructure, that MEV constitutes a legally identifiable form of market manipulation with no comparable disclosure mechanisms, and that the periodic auditing interval is fundamentally incompatible with the velocity of blockchain risks. While early regulatory frameworks, such as the European Union's Markets in Crypto-Assets (MiCA) regulation, pioneered the use of machine-readable crypto-asset disclosures, and regimes like Dubai's VARA established operational guidelines, neither system structurally mitigated evasion. Furthermore, the emerging U.S. regulatory landscape remains too nascent to definitively assess. Consequently, prior frameworks have failed to supply a disclosure mechanism capable of closing all three monitoring gaps simultaneously. This section outlines a methodology designed to achieve enhanced monitoring: the BNP disclosure taxonomy, which was developed by Auditchain Labs AG; a Swiss-based provider of web3 and AI disclosure automation infrastructure, and subsequently became the subject of a research collaboration between the Rutgers Continuous Auditing and Reporting Laboratory (CARLab) at Rutgers Business School and Auditchain Labs AG[7]. The BNP framework is a structurally validated architecture applicable to any digital asset issuer operating on public blockchain infrastructure. Its development reflects a rigorous iterative process submitted to the FDIC in response to proposed application requirements under RIN 3064-AG20. Initially proposed as a 136-question, 107-element architecture, the taxonomy was subsequently supplemented to explicitly capture the risks of third-party Blockchain-as-a-Service (BaaS) provider arrangements. This expansion resulted in the final, comprehensive 148 input question, 125 output element taxonomy analyzed in this study.

For the purposes of this paper, the BNP Disclosure Taxonomy is defined as follows: a machine-readable, 148 input question, 125 output disclosure element regulatory reporting framework, formatted in XBRL-JSON, that requires digital asset issuers to disclose the operational characteristics, concentration risks, affiliate relationships, third-party service provider dependencies, and conflict-of-interest exposures arising from their participation in blockchain network infrastructure. The taxonomy is organized across eight disclosure domains and enforces an equal-input/equal-output architecture in which every input disclosure question maps to exactly one XBRL output element, ensuring that any omission is programmatically detectable without examiner intervention.

[7] https://auditchain.com/

### 3.1 Architectural Principles: Equal Inputs, Equal Outputs, and the Completeness Constraint

The BNP taxonomy was designed to implement a one-to-one correspondence between disclosure questions and XBRL concepts used to describe them: i.e., each XBRL output element corresponds to a group of input questions (see Figure 1). This correspondence shows how the taxonomy converts the disclosure obligation into a supervisory tool. Any XBRL instance document that lacks an element will be programmatically flagged as incomplete: the validator identifies every expected element and flags any omission automatically (selectively disclosing positive information while withholding negative information), without examiner involvement.

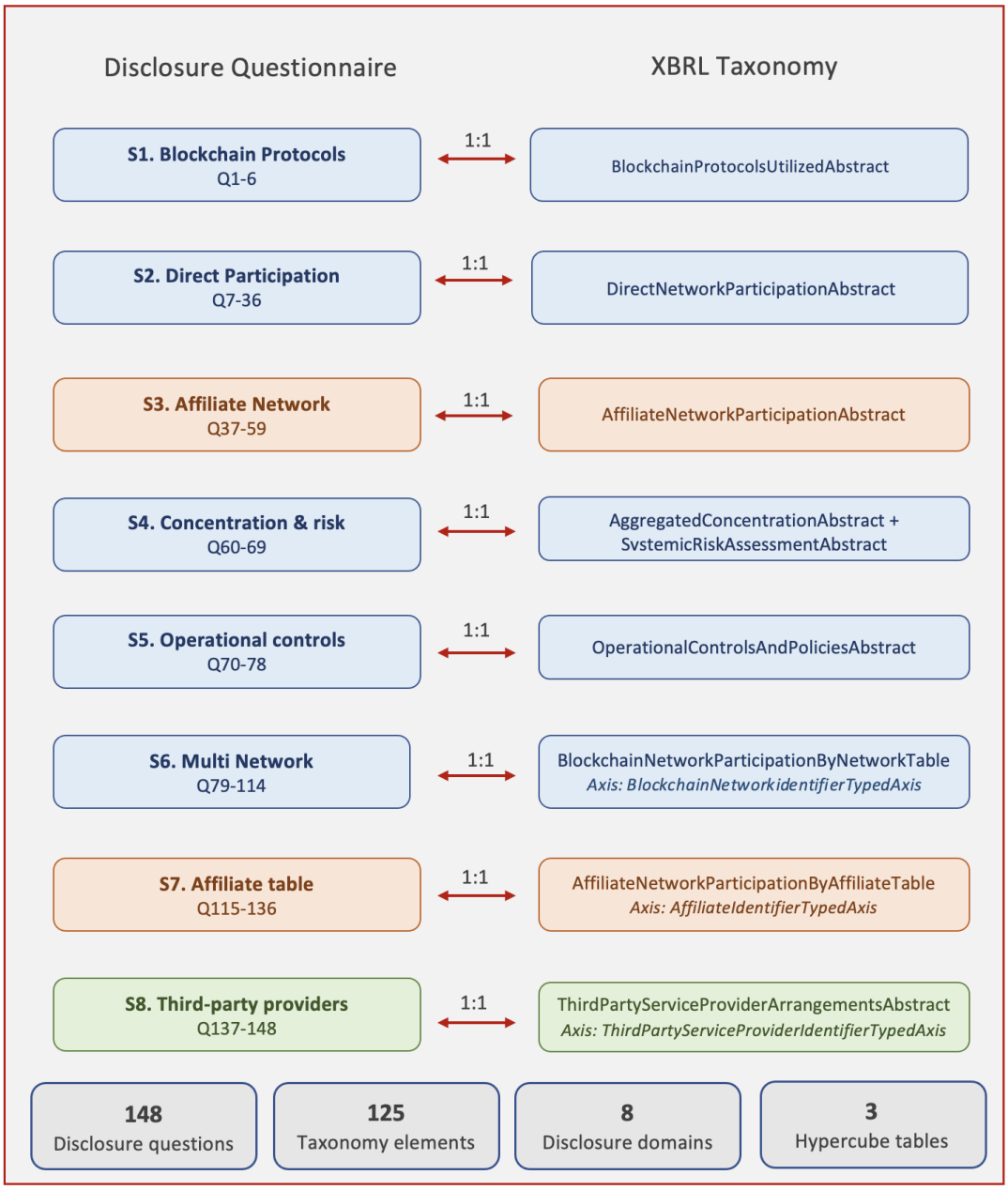


Figure 1: BNP Disclosure Framework

The 148 input questions are grouped into eight disclosure domains. Each domain targets a specific aspect of the GENIUS Act mandates and addresses distinct supervisory blind

spots left unaddressed. Crucially, the structural scalability of this taxonomy is enforced via three-dimensional axes: the "BlockchainNetworkIdentifierTypedAxis," the "AffiliateIdentifierTypedAxis," and the "ThirdPartyServiceProviderIdentifierTypedAxis." In XBRL architecture, a "typed axis" is a dynamic table feature that allows regulators to specify the taxonomy in a manner that permits the issuer to disclose as many Blockchains, Affiliates, and Third Parties as required [8] . This feature programmatically duplicates the core reporting elements and all associated concepts for every new entity the issuer reports, ensuring the framework scales perfectly regardless of the issuer's operational complexity.

By leveraging these typed axes, the BNP framework shapes the first regulatory disclosure instrument that is designed to address the three failure modes documented in the literature review: reserve-and-infrastructure joint stability, validator operational risk, cross-chain bridge, and conflict of interest (including MEV) between issuers and their network participants. Table 1 below provides an explanation of the conceptual logic underlying each of the 8 domains and their corresponding regulatory implications.

Table 1: Domain Mapping to GENIUS Act Mandates and Supervisory Functions

| **Section** | **Questions per section** | **Domain** | **Primary Mandate** | **Regulatory Function and Supervisory Blind Spot Addressed** |
|---|---|---|---|---|
| S1 | 6 | Blockchain Protocols Utilized | Safety | Identifies the attack surface baseline. It specifies the name and type of the blockchain protocol used, the consensus mechanism (PoW/PoS), and the smart contract deployment parameters and address. |
| S2 | 30 | Direct Network Participation | Safety / Soundness | Shapes the issuer's direct exposure to blockchain networks across five separate sub-domains, namely: basic participation information, network governance participation, network share and concentration metrics, operational dependencies, and financial performance. It addresses whether the issuer itself is a cause or a potential source of the extraction risk (MEV) and network control, which the taxonomy is designed to detect. |
| S3 | 23 | Affiliate Network Participation | Conflicts / Anti-Evasion | Extends network participation mapping to affiliate perimeter: affiliate identification and participation details, services provided, and conflicts of interest analysis. Q.54–55 probe whether an affiliate can influence the ordering of transactions and/or whether it has access to mempool data to see the pending ones before they are publicly confirmed. |

*Continued on next page...*

[8] The structural design of the BNP Taxonomy can be found in https://raw.rutgers.edu/bnp-disclosure-taxonomy

*(Continued from previous page)*

| Section | Questions per section | Domain | Primary Mandate | Regulatory Function and Supervisory Blind Spot Addressed |
|---|---|---|---|---|
| S4 | 10 | Aggregated Concentration Analysis & Systemic Risk | Safety / Soundness | Computes combined issuer and affiliate network control as a single consolidated control metric to see if it could influence consensus outcomes. The systemic risk sub-section (Q64-69) applies the Illusion of the Vault test directly and requires the issuer to affirm whether congestion on networks would simultaneously impair both its blockchain operations and ability to fulfill redemption commitments to holders of tokens. |
| S5 | 9 | Operational Controls and Policies | Soundness / Anti-Evasion | Conducts governance audit for BNP activities: it examines whether the issuer is subject to internal audit, it reviews, gives a board-level oversight, information barriers between issuer management and network operators, transaction-ordering logging, and active monitoring for any MEV-related misconduct in the network operations, like sandwich or front-running attacks. |
| S6 | 36 | Multi-Network Disclosure (Per-Network Hypercube) | Safety / Soundness | Replicates the entire protocol/participation disclosure schema across each blockchain network on which the issuer operates infrastructure or has deployed smart contracts, thereby eliminating the potential for aggregation masking (using the typed dimension axis "BlockchainNetworkIdentifierTypedAxis") that could conceal multiple-chain risk concentrations by disclosing only at the issuer level. |
| S7 | 22 | Affiliate Disclosure Table (Per-Affiliate Hypercube) | Conflicts / Anti-Evasion | Replicates the entire affiliate disclosure schema per individual affiliate entity using the AffiliateIdentifierTypedAxis. Due to their dimensional nature, both domains provide a structural basis for detecting entities that are not reported (an affiliate not identified; a network not listed), creating a quantitative gap in aggregate concentration metrics within domain 4. |
| S8 | 12 | Third-Party Service Provider Arrangements | Conflicts / Anti-Evasion | Requires structured disclosure per non-affiliate third party (using the typed dimension axis "ThirdPartyServiceProviderIdentifierTypedAxis") providing blockchain network participation services, including yield facilitation arrangements, fee structures, reserve-to-yield relationships, and conflicts of interest & mitigation. |

### 3.2 Validation Architecture and Evasion Detection

The three typed axis tables extend the completeness constraint to the entity level. Each network, affiliate, and third-party service provider reported by an issuer will include all of the relevant disclosure elements for each respective entity. This is enforced via the type-oriented dimension architecture; an issuer that states it participates in six different blockchain networks, yet only completes data for three, will create an identifiable structural inconsistency and mismatch that can be detected automatically via cross-validation against on-chain data within the XBRL instance itself. To mathematically validate the architecture, Quantitative relationships are enforced through calculation linkbases, to enforce arithmetic consistency, which ensure that concentration percentages sum to 100% (See Equation 1) and that total network activity accurately reconciles across the direct/affiliate/third-party dimensions. If an issuer's aggregate network activity exceeds the sum of the three single dimensions disclosed, discrepancies will be flagged without examiner intervention. Temporal validation enforces period-over-period consistency and ensures that there are no material changes occurring during reporting periods that should trigger some interim disclosures.

$$\text{Aggregate Network Activity} = \text{DRA} + \text{ARA} + \text{TPRA} = 100\% \tag{1}$$

where:
DRA: Direct Reported Activities
ARA: Affiliate Reported Activities
TPRA: Third-Party Reports Activities

As such, these three validation stages constitute a fully automated, continuous-auditing execution pipeline that translates XBRL instance documents submitted by issuers into a real-time supervisory assertion (Figure 2). Structural validation: validates the data type integrity of submissions prior to inclusion in the supervisory database. Arithmetic validation: verifies that quantitative relationships exist within and among dimensional instances at the moment of ingestion. Temporal validation continuously identifies material changes to infrastructure relative to prior-period baselines. Although the pipeline does not require any examiner involvement to generate its primary output (i.e., completeness flags, arithmetic discrepancies, and interim disclosure triggers), an examiner still needs to review any flagged filings before making a final decision on their regulatory status.

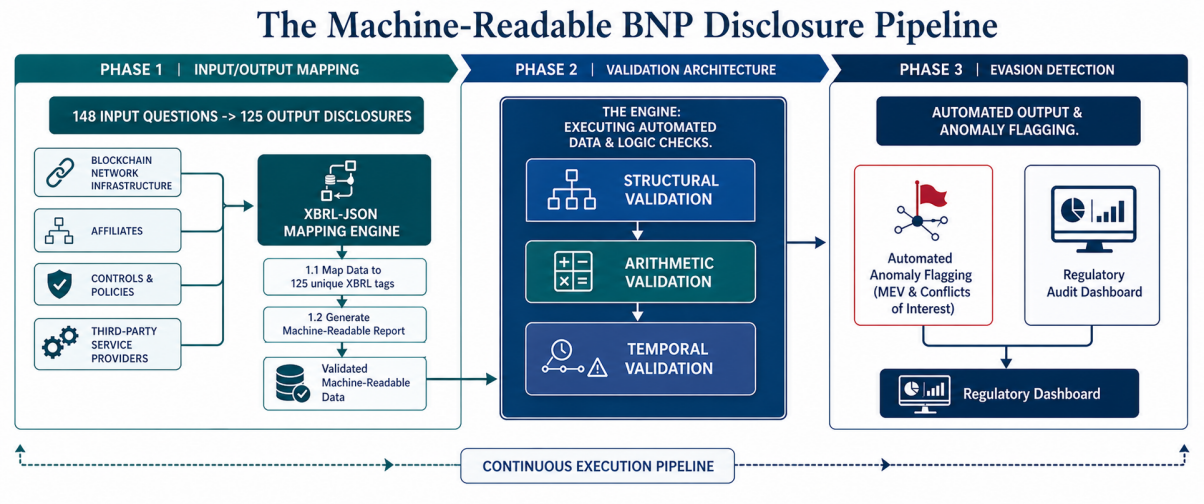


Figure 2: The Machine-Readable BNP disclosure Pipeline

Section 8 (Q137–Q148), is a novel category added in the supplemental submission. It discloses the relationship between the issuer and non-affiliate third parties that provide BNP services. The evasion mechanism targeted by this category operates as follows: a stablecoin issuer subject to categories one through seven may achieve compliance with those categories by divesting from direct and affiliate network participation, while contracting with an unrelated third-party validator to provide network participation services. Issuers' responses in sections two and three will accurately reflect zero direct and affiliate participation. However, they might depend on the continued operation of contracted validators' infrastructure, while financial arrangements between the issuer and third-party validators remained invisible to supervisors.

More critically, these arrangements create the specific evasion pathway that § 5903(a)(11)'s yield prohibition was written to prevent. Stablecoin issuers are precluded from providing any form of yield to holders because they hold, use or retain the stablecoin. In the absence of formal BNP structured disclosure of third party service providers, an issuer could disguise prohibited yield payments as compensation for validator services—structuring payments that flow from the issuer to a third-party validator (ostensibly for network participation services) and then back to token holders as network participation rewards, creating a circularity of payment that the yield-prohibition is intended to prevent, yet which would be undetectable through narrative-based disclosure.

# 4 Methodology

The methodology challenge addressed in this study is unique in the regulatory accounting literature. The question here is not whether the 148 input questions result in 125 output elements of the BNP are individually acceptable. The analytical basis for each is detailed in sections 2 and 4. Rather, the question is whether systematically removing 130 input questions will create material supervisory blind spots under the GENIUS Act's three main mandates. This represents a large-scale, structured evaluation task: 148 components, 3 sets of criteria, and a formal legal standard (the GENIUS Act's substantive provisions at 12 U.S.C. §5903 – 5905)[9] that will determine their classification. Moreover, it represents precisely the type of problem for which multi-agent Generative Artificial Intelligence (AI) methodology is particularly well-suited: parallel, independent evaluation of a large set of elements against a formally defined set of criteria, with disagreements among agents indicating analytical contestability and consensus among agents indicating robustness.

## 4.1 The Competitive Scientific Hypothesis: An Objective Analysis of a Minimized Phased Alternative

Our study approaches the concept of a minimized, phased disclosure framework exactly as it deserves to be treated, as a competitive scientific hypothesis. To rigorously test the necessity of the comprehensive taxonomy, our research team constructed an 18-element minimalist alternative which is derived from informal documents and prompts circulated among industry working group members, to evaluate the theoretical boundaries of regulatory reporting capacity.
This hypothetical 18-element phased approach is best understood as a constraint satisfaction problem. Proponents of this strategy seek to maximize two mutually exclusive objectives: comprehensive risk coverage (which would favor a larger set of disclosure elements) and early regulatory adoption (which would favor a smaller, administratively feasible set). The phased

[9] 12 U.S. Code § 5905 - Supervision and enforcement with respect to Federal qualified payment stablecoin issuers and subsidiaries of insured depository institutions (2026). en url: https://www.law.cornell.edu/uscode/text/12/5905

approach resolves this dilemma by accepting the latter objective as a tight constraint and treating the former as an aspirational goal. The underlying administrative theory is that regulatory capacity, namely the Federal Financial Institutions Examination Council (FFIEC) and principal stablecoin regulators, to process, validate, and assess structured Digital asset disclosures, is a limited resource for regulatory agencies in the short run. For example, introducing the full 148 input question, 125 output element taxonomy at once could result in institutional overload: too many new data fields, too many new validation processes to develop, and too much technical complexity for examiner training to handle prior to the GENIUS act taking effect. The 18-element Phase I is, therefore, a minimum viable disclosure set adjusted for what regulators can usefully implement now. The remaining 130 elements are postponed until the regulatory institution develops sufficient capacity.

Proponents of this alternative also claim that, while the GENIUS Act does not explicitly require reporting of Blockchain Network Participation (BNP) as a separately designated regulatory category, it does require reporting of reserve, attestation, and operational risk management requirements. However, none of the phrases, Blockchain Network Participation, validator uptime, hash power, or smart contract address, appear in the statute itself. From this perspective, inclusion of base layer and Layer 2 metrics such as hash power concentration, consensus mechanisms, and validator slots, requiring complex disclosures in Phase I, constitutes regulatory overreach beyond what the statute immediately requires and should instead be considered for implementation in Phase III after notice-and-comment rulemaking processes have addressed blockchain infrastructure disclosure as a separate named regulatory obligation. The 18-element alternative is thus a principle-based example of regulatory minimalism: report what the statute clearly requires in Phase I, and allow the rulemaking process to address the more technical issues in later phases.

Proponents of this alternative correctly state that administrative capacity is a limiting factor, that regulatory adoption depends on feasibility, and that the GENIUS Act's language does not contain a reference to "Blockchain Network Participation." While these statements are accurate, the phased alternative approach fails to account for the possibility that the 130 deferred elements won't cause serious supervisory shortcomings. Empirical evidence presented in Section 4 indicates that this assumption is incorrect, not because it reflects a regulatory preference, but because it reflects analytical consensus among three separate AI models, with intermodel Spearman rank correlation values ranging from 86% to 91%. Even though the minimized phased alternative may be easier for regulators in the short run, multi-agent AI analysis shows that it eliminates the most important mathematical disclosures from the Taxonomy. Delaying consideration of BNP elements indefinitely into a future Phase III means that regulators and the public remain entirely unaware of systemic base layer and Layer 2 risks (e.g., MEV attacks on liquidity pools, bridge hacks, congestion bottlenecks during redemptions) that provokes financial contagion effects, documented as as persistent, measurable, and systemically inherent to public blockchain operation in the current literature (Auer et al. 2022; Durachman and Rahman 2024; Aronoff et al. 2026).

### 4.2 Multi-Agent AI Experimental Design and Model Selection

Three different LLMs were used as independent adversaries modeling regulatory analysts: Sonnet 4.6 (Anthropic Claude), Google Gemini 3.1, and OpenAI GPT-5. Each model was assigned the same regulator identity and had access to the same information package across four steps. First, we provided the agents with the BNP Disclosure Questionnaire, the BNP XBRL Taxonomy hierarchical overview including all concrete elements, and the documents detailing the competing 18-element proposal and its Phase III deferral rationale. At no point during the evaluation sequence did any model know what others had analyzed. Selecting three distinct ar-

chitectures, rather than multiple runs on a single architecture, serves a particular methodological purpose. Robustness of analysis, evidenced by convergence across models trained on vastly different data sets with disparate architectural parameters and by Reinforcement Learning from Human Feedback (RLHF) alignment, provides stronger evidence than repeating samples from a single model.

### 4.3 The Four-Stage Sequential Prompting Framework

The evaluation was conducted using a four-stage, sequential prompts framework (Figure 3), in which each stage built on outputs from previous stages. Specifically, this mirrored the sequentially implemented steps an examining agency would take in practice: establishing legal context and background; stress-testing proposals by evaluating competing alternatives against specific criteria; and classifying disclosures independently at the elemental level and revising those classifications based on newly introduced technical research.

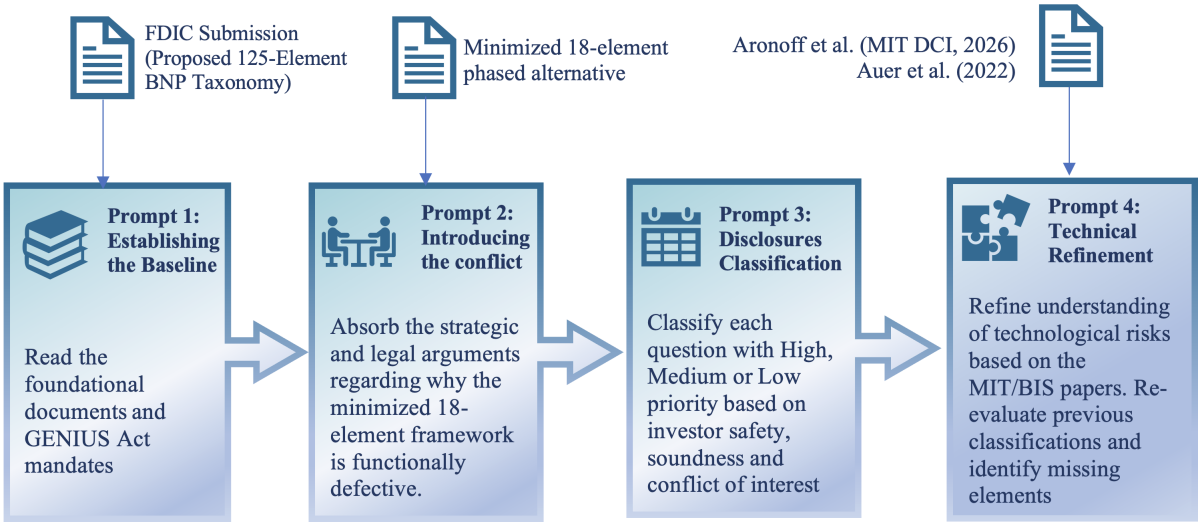


Figure 3: Methodology roadmap: Four-stage multi-agent stress test

**Prompt 1: Regulatory Context Establishment.** Each model was delegated the role of a continuous audit expert working at a university research lab. The subject area was stated as evaluating the structural disclosures made by digital asset issuers to assess whether they pose threats to investors regarding safety/security/conflict-of-interest avoidance. Models were directed to read word by word, and every footnote cited a reference to law, statute, and case contained in both original and supplementary FDIC comment letters. The requested output was a confirmation statement that summarized how the taxonomy framework updates addressed each of the three research areas (safety, security, and conflict of interest), thereby establishing the legal/regulatory context governing all subsequent classifications.

**Prompt 2: Stress Testing Competitive Proposal.** In this stage, each model was provided with the proposed framework outlining the details of the competing minimized 18-element and the associated Phase III deferral rationale. Models were instructed to identify the strongest arguments supporting this approach prior to beginning element-based classification. The models were specifically conditioned to impose stricter evidential burdens before any further priority classification. Adversarial conditioning, as represented in Prompt 2, addressed construct validity concerns by demonstrating that High-Priority rated disclosures were formed after each model was exposed to competing arguments. Prioritization

based upon High-Priority ratings, therefore, reflected judgments surviving opposition from competing arguments rather than judgments formed absent such competition.

**Prompt 3: Independent Element-Level Classification.** The AI agent models were then assigned to classify every single element in the BNP Disclosure Questionnaire as either High-Medium-Low Priority based upon three district research pillars: (1) Investor Safety; (2) Operational Soundness; and (3) Conflict-of-Interest prevention, especially MEV attack vector risks. Each model generated structured tables containing one- or two-sentence justifications for each classification, without considering the outcomes of other models. Explicit identification of MEV as part of Conflict-of-Interest Evaluation Criteria represented a significant methodological concern, assuring that affiliate mempool access and third-party validator elements were explicitly assessed for MEV-related risks rather than receiving generic conflict-of-interest ratings.

**Prompt 4: Refining Classifications Based Upon Current Literature Technical Findings.** In the final stage, the models were supplied with two papers: "The Hidden Plumbing of Stablecoins," by Aronoff et al. (2026), and "Miners as Intermediaries: Extractable Value and Market Manipulation in Crypto and DeFi" by Auer et al. (2022), and were requested to explain how these technical developments affected either modified or validated previously developed classifications. Inclusion of the BIS paper at this juncture was intended to translate the conflict-of-interest risk associated with MEV into an enforceable regulatory problem akin to illegal front-running, as the authors described. The BIS document enabled each model to rely on institutional regulatory authority to enhance the high-priority status of MEV-related elements, e.g., section 8 third-party service provider arrangements, solely on account of conflict-of-interest/public-harm grounds rather than solely on technical merit. Elements classified differently or whose justifications improved between prompt stages 3 and 4 are noted in findings as elements whose risk importance is understated by purely technical analysis alone and necessitates institutional regulatory framing provided by BIS /MIT papers.

# 5 Empirical Findings And Taxonomy Validation

Section 4.1 describes the collective AI consensus regarding the 148 BNP Taxonomy questions, demonstrating that the majority of the models reject the notion that the comprehensive Taxonomy can be reduced to a safe minimum of 18 elements. Section 4.2 explains the three critical blind spots created by the minimized framework with respect to the three models' classifications of each component and relevant statute. Section 4.3 shows evidence of how BIS and MIT papers affected the calibration phase of prompt 4 and thereby strengthened Section 8 classifications.

## 5.1 Inter-Model Convergence: Scientific Proof That the Taxonomy Cannot Be Safely Compressed

This central empirical question is whether or not the BNP Taxonomy can be reduced to 18 elements without material deficiencies and supervisory disadvantages. As illustrated in Figure 4, the collective evidence from the three independently deployed LLMs indicate that such a reduction is unlikely, due to the high/priority confirmation when it comes to section-level subtotals. Analysis of the multi-agent stress test indicates almost universal empirical support for the need for a complete 148 input question, 125 output element BNP taxonomy. The three AI agents all rejected the notion of a reduced disclosure framework and classified over 80% of the 148 input-125 output as "high priority" in order to meet the statute-based mandates of the GENIUS Act. Claude assigned High-priority status to 127 (85.81%) out of the 148 inputs,

whereas Gemini and GPT-5 assigned High-priority status to 103 (69.6%) and 102 (68.9%) respectively. All other input questions were either primarily classified as Medium Priority (which serve as critical contextual anchors for auditability) or Low Priority (receiving less than four percent allocations from all the models). The qualitative explanations of each of the models clearly explain why delays in disclosing this information will create blind spots in supervision systems. The three primary risk pillars used to justify their classifications of high priority were:

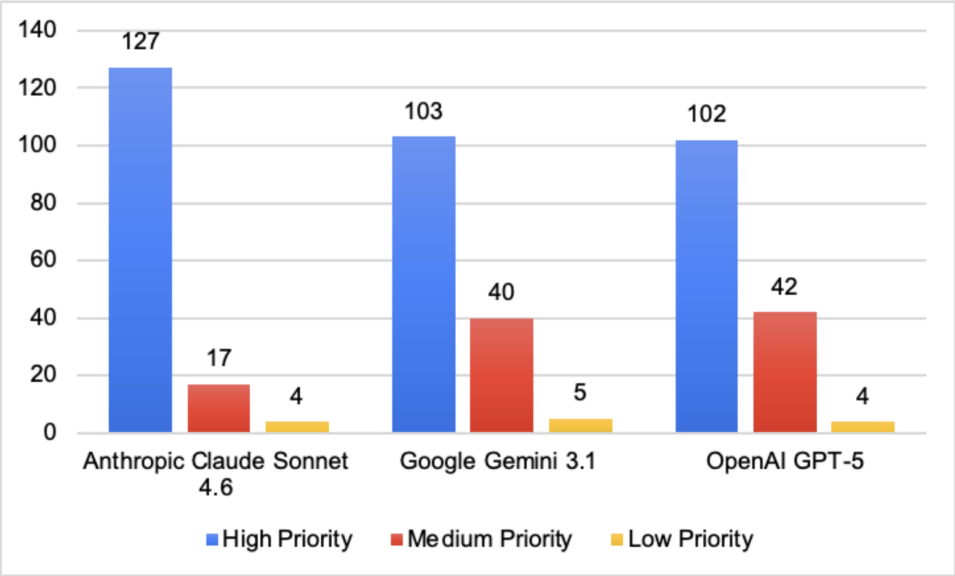


Figure 4: Priority classification of the proposed 148-input questions across three multi-agent AI

- **Systemic Security and Base Layer/Layer 2 Infrastructure:** Elements that describe the base layer and Layer 2 technical architecture were considered non-negotiable. For example, Question 3 -which identifies the specific consensus mechanism used by the network- was designated as high priority across all three models because it defines how network security is implemented. If Proof-of-Work is used, security is based on hash power, whereas if Proof-of-Stake is used, security is based on capital stake. and therefore this defines the technical conditions which govern transaction ordering abuses. Likewise, Question 5, which requires issuers to disclose their smart contract addresses, was deemed mandatory because these addresses are the only means to verify an asset's technical status in near real time. The lack of transparency prevents operational control narratives from being verified.

- **Conflicts of Interest and MEV Attacks:** The models showed significant awareness of hidden financial architectures and shadow intermediation. For instance, the models recognized that when issuers or their vendors act as gatekeepers (validators/miners) of a network, they have independent motivations to reorder, censor, or front-run transactions in order to maximize MEVs. They thereby uniformly classified Question 7, which concerns whether issuers operate directly or indirectly through affiliates on blockchain infrastructure, as a high priority. Moreover, the models identified questions about the reserve-to-yield relationship (See Questions 144 & 145 in Table 2) as important for detecting self-reinforcing insolvency spirals in which reserve assets are excessively encumbered to allow issuers to generate illicit yields under the guise of service fees.

- **Investor Protection (Safety) and Third Party Mitigation:** The LLM models placed a heavy emphasis on disclosures relating to third-party service providers and corresponding mitigation measures ( See Question 148). Outsourcing blockchain functions creates opportunities

for hidden self-dealing and concentration failures. The models determined that contractual and operational controls designed to mitigate third-party MEV practices (i.e., anti-front-running commitments) are the best empirical evidence of compliance against evasion, and thus, support investor safety.

### 5.1.1 Inter-Model Agreement and Consensus Validation

We used Gwet's AC1 in order to determine the degree of agreement between the models regarding the taxonomy sections. When evaluating regulatory compliance and safety-critical frameworks, data distributions typically exhibit a high degree of skewness; when multiple evaluators correctly and uniformly classify a strict regulatory requirement (i.e., continuously assign 'high' ratings), common correlation metrics such as Spearman or Cohen's kappa (Wongpakaran et al. 2013) either fail or significantly distort the true level of consensus. Gwet's AC1, which was introduced in 2001, is specifically designed to measure robust inter-rater reliability, especially in cases where ratings show high prevalence, by using a pooled average marginal probability for chance agreement (Gwet 2008); Gwet 2014). Table 2 depicts the pairwise AC1 values, illustrating convergence among the three LLMs across the eight BNP domain areas. The aggregated inter-model reliability across the entire 148 input question, 125 output element dataset indicates a significant level of agreement, with total AC1 scores indicating a very close range of 0.696 for Claude versus Gemini; 0.653 for Gemini versus GPT, and 0.695 for Claude versus GPT. These results confirm that, regardless of the training architectures, the independent AI agents produced a similar ranking of the taxonomy's regulatory materiality.

Table 2: Inter-model agreement correlation using Gwet AC1

| **Section** | **Disclosures (N)** | **Claude - Gemini** | **Claude - GPT** | **Gemini - GPT** |
|---|---|---|---|---|
| 1. Protocols | 6 | 0.543 | 0.534 | 0.040 |
| 2. Direct Participation | 30 | 0.570 | 0.614 | 0.686 |
| 3. Affiliates | 23 | 0.590 | **0.730** | 0.575 |
| 4. Concentration | 10 | **0.756** | **1.000** | **0.756** |
| 5. Controls | 9 | 0.660 | **1.000** | 0.660 |
| 6. Multi-Network | 36 | 0.631 | 0.517 | 0.659 |
| 7. Affiliate Table | 22 | **0.901** | **0.715** | 0.566 |
| 8. Third-Party | 12 | **0.803** | 0.680 | 0.627 |
| **Total** | **148** | **0.696** | **0.695** | **0.653** |

Further analysis into domain-specific reliability confirmed that the reported disclosures are extremely crucial. For instance, in Sections 4 (Aggregated Concentration Analysis) and 5 (Operational Controls and Policies), the models yielded the highest levels of agreement, achieving a nearly-perfect AC1 value of 1.000 between Claude and GPT-5. A similarly high level of reliability was observed in Section 7 (Affiliate Network Participation Table), with a near-unity AC1 value of 0.901 between Claude and Gemini, along with Section 8 (Third-Party Service Providers Arrangement), with AC1 scores ranging from 0.627 to 0.803.

These statistics support a regulatory fact: Even the most conservatively parameterized models tested in this adversarial experiment concurred that base layer and Layer 2 concentration,

affiliate exposure, and MEV conflicts are major sources of systemic risk. Therefore, these levels of inter-model agreement conclusively prove mathematically that the full-dimensional architecture represented by the 148 input question, 125 output element taxonomy is necessary to avoid immediate systemic contagion risks.

# 6 Discussion, Limitations, and Future Directions

The empirical consensus we documented in Section 5 addresses the paper's core normative question. But the implications of these findings go far beyond the choices between the 148 input question taxonomy and the 18 input minimized alternative inputs. More profoundly, the multi-agency stress test shows that digital asset management cannot take place using the same disclosure analogues used in traditional finance (e.g., narrative disclosure, periodic attestation statements, and point-in-time reserve certificates). These instruments rely on a stable institutional relationship in which the audit entity's operational reality changes more slowly than its reporting cycle. On the other hand, public blockchain technology fundamentally and permanently violates that assumption; That is because MEV attacks techniques evolve based on regulators' interest, bridge custody architectures change, consensus mechanisms fork, and validator sets reconstitute. The BNP taxonomy encodes supervisory requirements as machine-readable XBRL-JSON elements that map directly to the issuer's operational statements. Therefore, while improving on traditional disclosure practices, it implements a fundamentally new supervisory paradigm.

We call the supervisory artifact created by this architecture a "Regulatory Digital Twin", which is a continually-updated machine-readable representation of a stablecoin issuer's blockchain operational reality that corresponds exactly to the issuer's own submitted information. In engineering terms, a digital twin is a live computational model of a physical system that enables outlier detection and state monitoring without human intervention. In the blockchain context, when a stablecoin issuer submits a BNP instance document, the equal-input/equal-output architecture, the dimensional validation structure, and the temporal assertion set together form a computable model of an issuer's network control posture, conflict-of-interest exposure, and infrastructure dependency profile. This paper represents a paradigm shift rather than an incremental enhancement in financial reporting. It addresses the transition from compliance-as-narrative to compliance-as-computation, and from auditing-by-inspection to real-time state monitoring. As previously mentioned, there are clear implications for systemic risk resulting from this transition. The systemic risk implications of this transition are direct. The literature review established that the financial contagion mechanism in stablecoin runs is not reserve insufficiency but a redemption rail failure; a cascade in which a blockchain infrastructure disruption, a bridge exploit, or a MEV attack surge renders reserve assets functionally inaccessible before an auditor observes the evolving distress.

However, it is critical to acknowledge the boundaries of isolated disclosure frameworks. The BNP taxonomy alone cannot absolutely prove intentional omission, particularly if sophisticated actors deliberately silo network activities off-balance-sheet or obscure them within complex crypto-derivative markets. Ultimately, identifying true omission anomalies requires triangulating these machine-readable BNP disclosures with overarching financial reporting standards such as United States Generally Accepted Accounting Principles (GAAP) or International Financial Reporting Standards (IFRS) to detect instances where revenue from unreported chains manifests incongruously in other accounting concepts. The empirical execution of this comprehensive, cross-disciplinary evasion detection, (e.g., layering BNP validation against traditional financial audits), represents the core focus of CARLab's next stage of research.

The systemic implementation of the BNP framework is strongly supported by the safety and soundness provisions of Federal Banking laws. Information regarding base layer and Layer 2 infrastructure vulnerabilities, bridge exposures, and MEV conflicts explicitly meets the standard of materiality established in TSC Industries, Inc. v. Northway, Inc. (426 U.S. 438, 1976)[10], which holds that information is material if there is a substantial likelihood that a reasonable stakeholder would consider it important. Consequently, the primary challenge to implementation is not a lack of legal or technological justification, but rather achieving structural consensus among disparate regulatory bodies (e.g., the FDIC, OCC, NCUA, Federal Reserve, and SEC) under emerging legislative frameworks like the GENIUS and CLARITY Acts. Future policy research must explore the development of unified, inter-agency data-sharing architectures that allow multiple regulatory bodies to ingest, validate, and analyze standardized BNP XBRL filings concurrently, thereby moving the taxonomy from a theoretical standard to an active supervisory utility.

# 7 Conclusion

Our research has demonstrated, both theoretically and empirically, that the GENIUS Act's reserve-adequacy provisions constitute a necessary but insufficient component of stablecoin regulation. Stablecoins operating on public blockchains do not exist as isolated systems; their operational stability depends on a distributed network of validators, miners, and bridge operators, with no obligation to disclose their operations or financial ties to the issuer. These actors have every incentive to extract as much value as possible from the system (via MEV techniques), which can lead to systemic instability before even a scheduled audit cycle can identify the emerging stress.

The 148 input question, 125 output element BNP Disclosure Taxonomy addresses material disclosure deficiency through a machine-readable, structurally validated architecture that simultaneously closes three monitoring gaps: (i) the gap between reserve quality and redemption-rail integrity; (ii) the gap between financial-statement auditing and blockchain infrastructure security; and (iii) the gap between conflict-of-interest disclosure and MEV attack detection. To empirically validate this, we conditioned three independent pre-trained AI models (Claude Sonnet 4.6, Gemini 3.1, and GPT-5) against the strongest arguments for a minimized 18-disclosure alternative. All three models converged on a definitive conclusion: omitting any of the four primary domains (Base Layer and Layer 2 infrastructure, affiliate concentration, MEV conflicts, or third-party arrangements), creates severe monitoring blind spots. This satisfies the strict materiality standard established in TSC Industries, Inc. v. Northway, Inc. (426 U.S. 438, 1976) as well settled law. The Gwet's AC1 inter-model agreement scores ranging from 0.653 to 0.696, and the near-perfect domain-level agreement in Sections 4 and 5 between Claude and GPT, confirm that this conclusion is analytically robust across significantly disparate model architectures rather than an artifact of any single model's training priors.

The BNP Taxonomy represents a paradigm shift in regulatory oversight that extends well beyond stablecoin regulation. It's a new category of disclosure information that in the aggregate, becomes the basis for on-chain systemic risk measurement and economic activity driven by the on-chain migration of the global capital markets. This proposed framework is the first regulatory disclosure infrastructure to apply these principles to the substrate of digital asset infrastructure. Future research will determine whether the velocity and scale required by public blockchain infrastructure requires the operationalization of continuous auditing theory. The theoretical case and the empirical validation are now established. What remains is implementation.

[10] https://supreme.justia.com/cases/federal/us/426/438/

**Acknowledgments:** We extend our thanks to the entire team at the CARLab of Rutgers Business School as well as the National Center for Scientific and Technical Research in Rabat. Special thanks to our Auditchain colleagues for their constant support.

**Conflict of interest:** The authors declare no conflict of interest.

## References

Adrian, Tobias et al. (2025). "Understanding Stablecoins". en. In: URL: https://www.imf.org/-/media/files/publications/dp/2025/english/usea.pdf.

Alipanahloo, Zeinab, Abdelhakim Senhaji Hafid, and Kaiwen Zhang (2024). "Maximum Extractable Value (MEV) Mitigation Approaches in Ethereum and Layer-2 Chains: A Comprehensive Survey". In: *IEEE Access* 12, pp. 185212–185231. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2024.3514375. URL: https://ieeexplore.ieee.org/abstract/document/10787131.

Aramonte, Sirio, Wenqian Huang, and Andreas Schrimpf (2021). "DeFi risks and the decentralisation illusion". en. In.

Aronoff, Daniel et al. (2026). *The Hidden Plumbing of Stablecoins: Financial and Technological Risks in the GENIUS Act Era*. arXiv:2604.17167 [econ]. DOI: 10.48550/arXiv.2604.17167. URL: http://arxiv.org/abs/2604.17167.

Auer, Raphael, Jon Frost, and Jose María Vidal Pastor (2022). "Miners as intermediaries: extractable value and market manipulation in crypto and DeFi". en. In: *BIS Bulletins*. Number: 58. URL: https://ideas.repec.org//p/bis/bisblt/58.html.

Azar, Pablo D. et al. (2024). *The Financial Stability Implications of Digital Assets*. en. SSRN Scholarly Paper. Rochester, NY. DOI: 10.2139/ssrn.5029118. URL: https://papers.ssrn.com/abstract=5029118.

Belenkov, Nikita et al. (2025). *SoK: A Review of Cross-Chain Bridge Hacks in 2023*. arXiv:2501.03423 [cs]. DOI: 10.48550/arXiv.2501.03423. URL: http://arxiv.org/abs/2501.03423.

Catalini, Christian and Alonso de Gortari (2021). *On the Economic Design of Stablecoins*. en. SSRN Scholarly Paper. Rochester, NY. DOI: 10.2139/ssrn.3899499. URL: https://papers.ssrn.com/abstract=3899499.

Diop, Papa Ousseynou, Julien Chevallier, and Bilel Sanhaji (2024). "Collapse of Silicon Valley Bank and USDC Depegging: A Machine Learning Experiment". en. In: *FinTech* 3.4, pp. 569–590. ISSN: 2674-1032. DOI: 10.3390/fintech3040030. URL: https://www.mdpi.com/2674-1032/3/4/30.

Durachman, Yusuf and Abdul Wahab Abdul Rahman (2024). "Blockchain and the Evolution of Decentralized Finance Navigating Growth and Vulnerabilities". en. In: *Journal of Current Research in Blockchain* 1.3, pp. 166–177. ISSN: 3048-1430. DOI: 10.47738/jcrb.v1i3.20. URL: https://jcrb.net/index.php/Journal/article/view/20.

Eichengreen, Barry, My T. Nguyen, and Ganesh Viswanath-Natraj (2025). "Stablecoin devaluation risk". en. In: *The European Journal of Finance* 31.11, pp. 1469–1496. ISSN: 1351-847X, 1466-4364. DOI: 10.1080/1351847X.2025.2505757. URL: https://www.tandfonline.com/doi/full/10.1080/1351847X.2025.2505757.

El Imami, Imane et al. (2026). *BNP Disclosure Framework — Rutgers CARLab*. URL: https://raw.rutgers.edu/bnp-disclosure-taxonomy.

Gross, Marco and Richard Senner (2026). "From Par to Pressure: Liquidity, Redemptions, and Fire Sales with a Systemic Stablecoin". en. In.

Gwet, Kilem L (2008)). "INTRARATER RELIABILITY". en. In.

— (2014). *Handbook of Inter-Rater Reliability, 4th Edition: The Definitive Guide to Measuring The Extent of Agreement Among Raters*. en. Google-Books-ID: fac9BQAAQBAJ. Advanced Analytics, LLC. ISBN: 978-0-9708062-8-4.

Kulkarni, Kshitij, Theo Diamandis, and Tarun Chitra (2023). *Towards a Theory of Maximal Extractable Value I: Constant Function Market Makers*. arXiv:2207.11835 [cs]. DOI: 10.48550/arXiv.2207.11835. URL: http://arxiv.org/abs/2207.11835.

Liao, Zeqin et al. (2024). "SmartAxe: Detecting Cross-Chain Vulnerabilities in Bridge Smart Contracts via Fine-Grained Static Analysis". In: *Proc. ACM Softw. Eng.* 1.FSE, 12:249–12:270. DOI: 10.1145/3643738. URL: https://dl.acm.org/doi/10.1145/3643738.

MacDonald, Cameron and Laura Zhao (2023). "Stablecoins and Their Risks to Financial Stability". en. In: *SSRN Electronic Journal*. ISSN: 1556-5068. DOI: 10.2139/ssrn.4466522. URL: https://www.ssrn.com/abstract=4466522.

Mancino, Davide et al. (2025). "Decentralization or Favoritism? An Analysis of Ethereum Transactions and Maximal Extractable Value Strategies". In: *2025 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*. ISSN: 2832-8906, pp. 1–9. DOI: 10.1109/ICBC64466.2025.11114553. URL: https://ieeexplore.ieee.org/abstract/document/11114553.

Materwala, Huned et al. (2025). "Maximal Extractable Value in Decentralized Finance: Taxonomy, Detection, and Mitigation". In: *IEEE Transactions on Services Computing* 18.6, pp. 4386–4407. ISSN: 1939-1374. DOI: 10.1109/TSC.2025.3620604. URL: https://ieeexplore.ieee.org/abstract/document/11202291.

Notland, Jakob Svennevik et al. (2026). "SoK: cross-chain bridging architectural design flaws and mitigations". In: *Blockchain: Research and Applications* 7.1, p. 100315. ISSN: 2096-7209. DOI: 10.1016/j.bcra.2025.100315. URL: https://www.sciencedirect.com/science/article/pii/S2096720925000429.

Sinai, Nday Kabulo and Hoh Peter In (2024). "Q-RTOP: Quantum-Secure Random Transaction Ordering Protocol for Mitigating Maximal Extractable Value Attacks in Blockchains With a Priority Gas-Fee Policy". In: *IEEE Access* 12, pp. 10036–10046. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2024.3351830. URL: https://ieeexplore.ieee.org/abstract/document/10384878.

United States: National Archives and Records Administration: Office of the Federal Register (2025). *An act to provide for the regulation of payment stablecoins, and for other purposes*. eng. Accession Number: PLAW-119publ27, PLAW-119publ27 Source: DGPO, DGPO. URL: https://www.govinfo.gov/app/details/PLAW-119publ27.

Vasarhelyi, Miklos A. and Fern B. Halper (2018). "The Continuous Audit of Online Systems1". In: *Continuous Auditing: Theory and Application*. Ed. by David Y. Chan, Victoria Chiu, and Miklos A. Vasarhelyi. Emerald Publishing Limited, p. 0. ISBN: 978-1-78743-414-1. DOI: 10.1108/978-1-78743-413-420181004. URL: https://doi.org/10.1108/978-1-78743-413-420181004.

Wen, Hongzhe and R. S. M. Lau (2025). *A Risk Mitigation Model of Monetary Ecosystem with Stablecoins*. arXiv:2510.10469 [q-fin]. DOI: 10.48550/arXiv.2510.10469. URL: http://arxiv.org/abs/2510.10469.

Wongpakaran, Nahathai et al. (2013). "A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples". en. In: *BMC Medical Research Methodology* 13.1, p. 61. ISSN: 1471-2288. DOI: 10.1186/1471-2288-13-61. URL: https://doi.org/10.1186/1471-2288-13-61.

Wu, Wenbin and Can Liu (2026). *Stability Anchors and Risk Amplifiers: Tail Spillovers Across Stablecoin Designs*. arXiv:2602.18820 [econ]. DOI: 10.48550/arXiv.2602.18820. URL: http://arxiv.org/abs/2602.18820.

Zhang, Mengya et al. (2023). *SoK: Security of Cross-chain Bridges: Attack Surfaces, Defenses, and Open Problems*. arXiv:2312.12573 [cs]. DOI: 10.48550/arXiv.2312.12573. URL: http://arxiv.org/abs/2312.12573.