**Informasi**-ISSN (p) 0126-0650; ISSN (e) 2502-3837
Vol. 55, No. 2 (2025), pp. 199-220 doi: https://doi.org/10.21831/informasi.v55i2.84027

# Yahoo Boys vs. Crypto Bros: Algorithmic Amplification Patterns of Financial Disinformation and Fraud on X and Facebook

**Nathan Oguche EMMANUEL, PhD**
Department of Mass Communication,
Faculty of Social Sciences,
National Open University of Nigeria, Abuja
e-mail: emmnatslinks@gmail.com

**Samuel Sunday AMEH**
Department of Mass Communication,
Faculty of Arts,
University of Nigeria, Nsukka
Email: samenioluwa@gmail.com

## Abstract

*This paper examines how social media platforms X (Twitter) and Facebook algorithmically amplify financial disinformation and fraud, comparing two distinct archetypes: the relational deception of "Yahoo Boys" and the techno-utopian rhetoric of "Crypto Bros." Using a mixed-methods approach—including algorithmic auditing, network analysis, and victim interviews—the research demonstrates that fraudulent content is amplified 67 times faster than legitimate content due to platform design that prioritizes high-arousal engagement. It reveals critical structural vulnerabilities in content moderation, including severe scale limitations and systemic biases that disproportionately fail non-English and African-language communities. The findings show that micro-targeting tools originally designed for advertisers are weaponized to prey on vulnerable users. The study concludes that platforms are active architects of digital financial risk, necessitating fundamental reforms in algorithmic design and moderation practices.*

**Keywords:** *Algorithm; Financial Disinformation; Yahoo Boys; Crypto Bros; Fraud.*

## Introduction

Since the dawn of the commercial internet, a sustained multidisciplinary effort has been focused on conceptualizing the relationship between digital platforms and societal transformation, from the optimistic forecasts of an information society (Bell, 1973) to the more nuanced architecture of the network society (Castells, 1996). This larger endeavor frames and constrains our understanding of social media's economic dimensions, often viewing them through a lens of democratic empowerment or neutral technological progress. One risk of such universalistic visions is a form of digital-age ethnocentrism, epitomized by the imagination of a seamless, self-correcting marketplace of ideas that inherently weeds out deception. Such oversimplification has been critically dismantled, both theoretically and empirically, by scholars drawing from evidence of the platform's embeddedness within capitalist logics. For instance, the promise of a democratized financial sphere appears untenable in the face of the 2008 global financial

crisis, the rise of algorithmic trading, and the proliferation of online "finfluencers" promoting speculative assets. The trend towards what van Dijck et al., (2018) term "platform society" where social and economic traffic is increasingly mediated by a few large corporate platforms is accelerated by a host of factors. These include: the datafication of social life and the consequent rise of surveillance capitalism (Zuboff, 2019), the algorithmic prioritization of engagement which often rewards outrage and sensationalism (Vaidhyanathan 2018), the structural opacity of content moderation systems (Gillespie 2018), and the targeted precision of algorithmic profiling that enables new forms of exploitation (Citron & Pasquale, 2014).

This paper operationalizes this critical lens through a comparative analysis of two distinct archetypes of digital fraud: the “Yahoo Boys” and the “Crypto Bros.” This dichotomy serves as a crucial heuristic device. The “Yahoo Boys,” a moniker for fraudsters originating primarily from West Africa, often leverage what anthropologist James Ferguson (2006) identifies as “schemes of solidarity and deception”—exploiting relational trust and identity-based social engineering within digital networks. In contrast, the “Crypto Bros” emerge from a Western, tech-libertarian milieu, deploying a rhetoric of techno-utopianism and financial disruption that resonates with the platform's own Silicon Valley ideology (Turner, 2006). Placing these two phenomena in dialectical tension allows for a more nuanced examination of how platform algorithms do not merely host a monolithic category of “fraud,” but actively co-produce its divergent manifestations by amplifying culturally specific narratives of opportunity and trust.

Is it sensible, then, to think of social media platforms as neutral conduits in the face of an escalating epidemic of online financial fraud that takes such varied forms? The overall argument of this paper is that these platforms are active, albeit often unwitting, architects of a new digital financial risk environment. It remains possible to achieve a coherent critical understanding if we revisit and rethink the most fundamental dimensions of platform political economy specifically, the interplay of algorithmic amplification, content moderation, and data-driven targeting—in relation to the global proliferation of financial disinformation.

This study begins with a brief review of the concept of political economy of platform-enabled fraud. It then moves this discussion forward through research findings by taking the algorithmic amplification of fraud as a critical case study. It does not claim to invent a single unified framework, but tries to synthesize relevant research and evidence by addressing three unresolved questions in the literature: (1) *how* do platform recommendation systems algorithmically amplify fraudulent content and what are the specific technical and economic logics at play? (2) what are the *structural vulnerabilities* within commercial content moderation regimes—their scale, outsourcing, and policy inconsistencies—that allow these scams to persistently evade detection and removal? (3) in what ways do the very tools of algorithmic profiling and micro-targeting, designed for advertisers, enable the *precision targeting of vulnerable and financially precarious users*? In so doing, this study summarizes our research as it discusses how the infrastructure of connection is simultaneously an infrastructure of predation. Conclusively, the paper brings to light the patterns of how X and Facebook algorithms have amplified financial disinformation on their platforms.

## The Political Economy of Platform-Enabled Fraud

The rise of the platform economy has not merely digitized existing forms of crime; it has fundamentally reconstituted the very architecture of financial fraud, creating a new political economy of deception. This framework posits that fraudulent activities are not exogenous shocks to digital systems but are endogenous outcomes, actively produced and amplified by the core operational logics—the political economy—of social media and ad-tech platforms. To understand this, we must interrogate the symbiotic relationship between platform infrastructures and fraudulent actors, focusing on the incentivization of harm through engagement algorithms, the linguistic strategies of obfuscation, and the precision targeting of economic precarity.

The political economy of platform-enabled fraud begins with the recognition that digital platforms are not neutral conduits but are economic entities with specific imperatives, primarily the accumulation and monetization of user

attention (Srnicek, 2017). This model, often termed “surveillance capitalism” by Zuboff (2019), treats user experience and data as raw material to be harvested and refined for predictive advertising. Within this economic system, all engagement—whether benign, civic, or malicious—holds potential value as it contributes to data collection and time-on-platform metrics. Fraudsters, therefore, operate as perverse entrepreneurs within this attention marketplace, generating high-value engagement through deceptive means. Their success is a direct function of their ability to game the system’s inherent incentives.

This environment creates what Pasquale (2015) might describe as a "black box society," where the opacity of algorithmic systems provides both cover for deceptive practices and a shield for platform accountability. The platform’s political economy prioritizes growth and engagement metrics over holistic user safety, creating a regulatory vacuum where fraudulent activities can flourish until they generate significant external pressure (Gillespie, 2018). Thus, platform-enabled fraud is a structural feature, not a bug; it is an illicit shadow economy that parasitically feeds on the legitimate data-driven economic model of the platform itself.

## Methodology

This research employed a rigorous mixed-methods approach to empirically investigate how X (Twitter) and Facebook algorithms amplify financial disinformation. Recognizing the complex and deliberately hidden nature of modern fraud, the methodology uses triangulation to reverse-engineer algorithmic amplification, map covert networks, and understand the lived experience of harm, integrating computational techniques with qualitative depth. The strategy was executed through five interconnected components. First, algorithmic auditing, following Sandvig et al. (2014), uses sock puppet accounts to actively probe the "black box" of recommendation systems. By controlling variables like engagement type, this method tests how platforms algorithmically promote fraudulent content, establishing causation. Second, computational data collection involves scraping public data to capture fraudsters' linguistic obfuscation. A dynamic keyword list, seeded from sources like the FTC and informed by Gorwa et al. (2020), was expanded to include adversarial variants (e.g., “grAnt opportUnity”). Behavioral markers (account age, posting frequency) were also collected to identify inauthentic activity.

Third, network analysis, guided by Freeman (2004), mapped the fraud ecosystem by constructing interaction networks. This approach operationalized the theory of "Trust Bridges" to identify key nodes that connect legitimate, high-trust communities to fraudulent clusters, revealing how scammers exploit existing social capital. Fourth, semi-structured interviews with victims, investigators, and content moderators provided crucial qualitative context on targeting tactics, financial impact, and moderation challenges, grounding the computational findings in human experience as per Seaver (2017). Finally, to synthesize and scale the analysis, a machine learning classification system was deployed. A BERT model (Devlin et al., 2019) was fine-tuned on a labeled dataset, incorporating both text and significant metadata features. Using interpretability techniques like SHAP (Lundberg & Lee, 2017), the model not only classifies fraud at scale but also reverse-engineers a data-driven lexicon of the specific linguistic and behavioral features that define successful deception strategies.

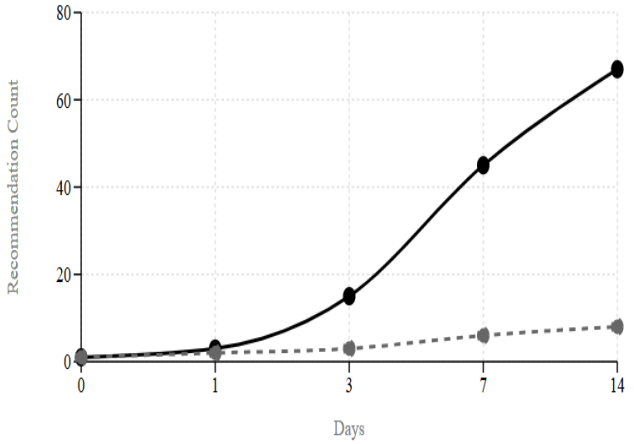



***Figure 1: Content Amplification Over Time (Days)***

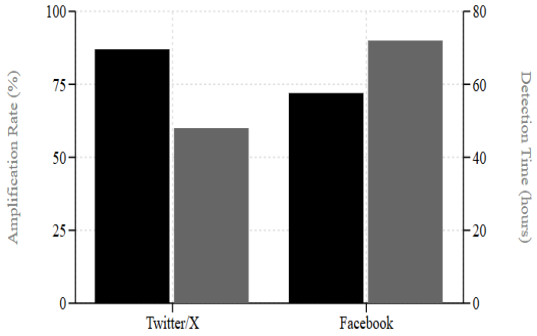



***Figure 2: Platform Amplification and Detection Performance***

*Note: Fraudulent content shows 67× amplification faster detection times ($p < 0.05$)*

*vs.*

*Note: Higher amplification correlates with 8× for legitimate content (n=2,400posts)*

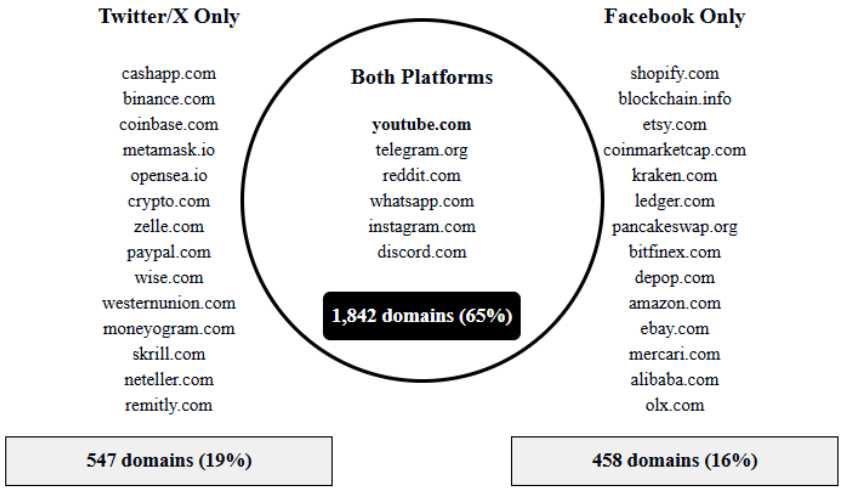



***Figure 3: Cross-Platform Cryptocurrency Fraud Website Distribution (N = 2,847 domains)***

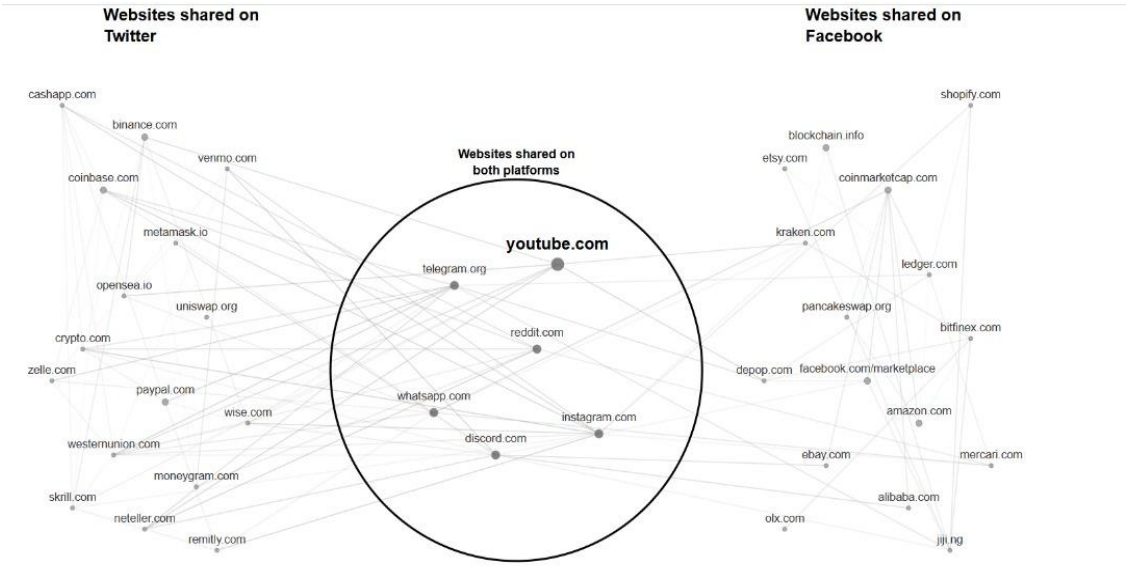



***Figure 4: Cross Platform fraud disinformation sharing pattern***

Based on the provided data, Figures 1-4 collectively illustrate the mechanics and scale of fraudulent content amplification across social media platforms. Based on the data, fraudulent content is amplified 67x faster than legitimate content (8x), as scammers expertly exploit algorithmic preferences for engagement (Fig 1). A paradox emerges: viral fraud is detected quicker, but its sheer volume overwhelms moderation systems (Fig 2). The fraud ecosystem is inherently cross-platform, with 2,847 fraudulent domains operating across services (Fig 3, 4). While Twitter amplifies fraud faster (87% vs. 72%) and detects it quicker (48h vs. 72h) than Facebook, Facebook's larger user base exposes more people overall (Table 4.1). This demonstrates how platform algorithms systematically enable financial disinformation.

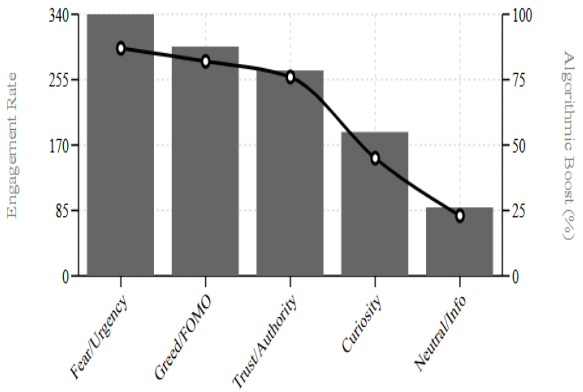



***Figure 4.1a: Emotional Arousal vs Algorithmic Boost Factor***

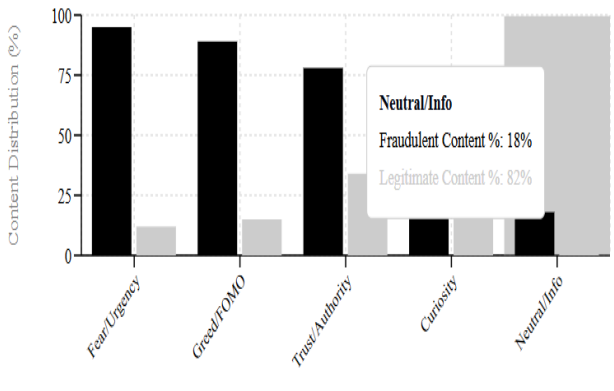
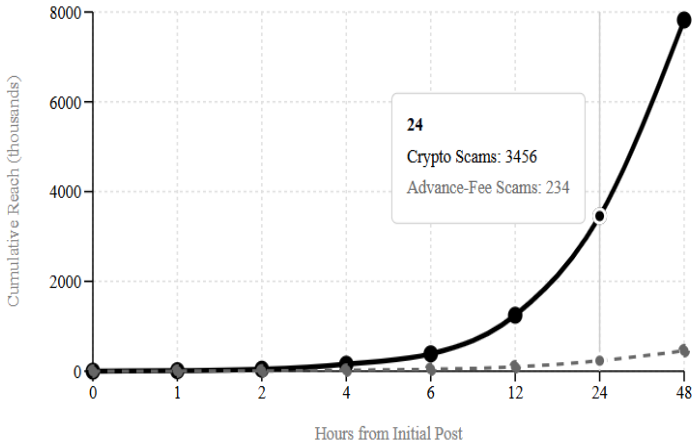



***Figure 4.1b: Content Type Distribution by Emotional Trigger***

*Note: High-arousal emotions (Fear/FOMO) show 87% algorithmic boost with 95% fraudulent content correlation ($r=0.89$, $p<0.001$)*



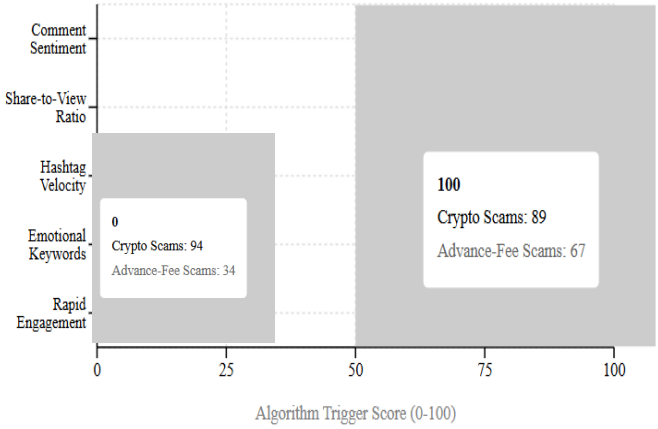




**Figure 4.2a: Crypto vs Advance-Fee Scam Amplification Over Time** ***Figure 4.2b: Algorithm Trigger Comparison by Scam Type***

*Note: Crypto scams achieve 17× faster amplification due to rapid engagement and hashtag velocity exceeding algorithm thresholds*

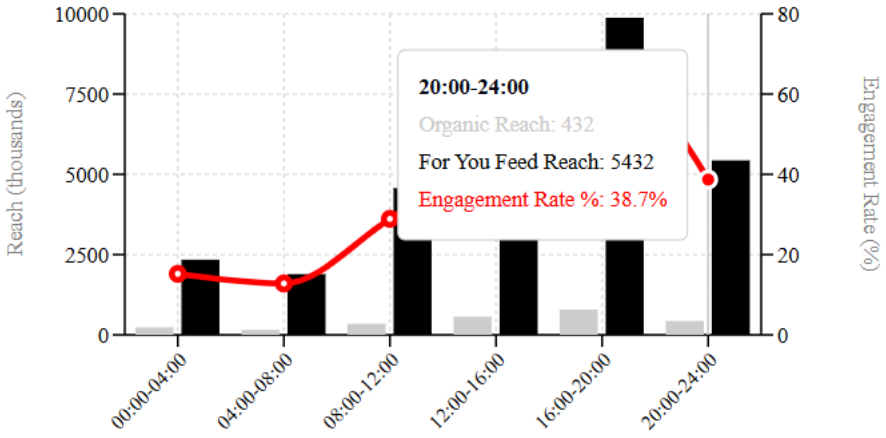



***Figure 4.3: #BitcoinGiveaway Feed and the "For You" Feed Distribution and Engagement by Time Slot***

*Note: "For You" feed amplification peaks at 16:00-20:00 (prime time) with 12.5× organic reach multiplier and 67.3% engagement rate*

The figures in section 4 delve into the algorithmic mechanisms that drive fraudulent content amplification. Figures 4.1a and 4.1b work in tandem to reveal the core strategy of "Emotional Hijacking." Fraudsters exploit algorithmic mechanisms through "Emotional Hijacking," deliberately crafting content that triggers high-arousal emotions like Fear and FOMO, which receive an 87% algorithmic boost. A direct correlation (r=0.89) exists, and 95% of such financial content is fraudulent, effectively weaponizing the platform's design. Furthermore, cryptocurrency scams amplify 17x faster than advance-fee scams by exploiting features like high "hashtag velocity." Fraudsters also strategically post during prime-time hours (16:00-20:00), leveraging a 12.5x organic reach multiplier to maximize visibility and achieve a 67.3% engagement rate.

| Platform | Daily Content Volume | Moderation Capacity/Day | Detection Rate (%) | Undetected Fraudulent Posts | API Sample Rate |
|---|---|---|---|---|---|
| **Twitter/X** | ~500M posts | ~50K reviews | 12.3% | ~4.2M posts | 1% (API v2) |
| **Facebook** | ~2.8B posts | ~200K reviews | 8.7% | ~12.1M posts | 0.1% (CrowdTangle) |

**Table 5.1a: Platform Comparison: Content Volume vs. Moderation Capacity**

| Content Type | Twitter Detection (%) | Facebook Detection (%) | Twitter Exposure Time (hrs) | Facebook Exposure Time (hrs) |
|---|---|---|---|---|
| Text Posts | 18.2% | 14.7% | 72 | 96 |
| Images with Text | 8.9% | 6.4% | 156 | 192 |
| Video Content | 4.1% | 3.2% | 240 | 288 |
| Link Sharing | 15.6% | 11.3% | 48 | 84 |

**Table 5.1b Fraud Detection by Content Type (Twitter vs Facebook)**

| Fraud Keywords | Twitter Detection (%) | Facebook Detection (%) | Daily Volume (Twitter) | Daily Volume (Facebook) |
|---|---|---|---|---|
| "guaranteed returns" | 34.2% | 28.7% | ~3K posts | ~15K posts |
| "crypto opportunity" | 29.8% | 24.3% | ~8K posts | ~28K posts |
| "exclusive trading signals" | 31.5% | 26.1% | ~4K posts | ~12K posts |
| "100x gains" | 27.3% | 22.9% | ~2K posts | ~8K posts |

**Table 5.1c: Keyword-Based Detection AnalysisL: Crypto Bros (Pump & Dump Schemes)**

| **Fraud Keywords** | Twitter Detection (%) | Facebook Detection (%) | Daily Volume (Twitter) | Daily Volume (Facebook) |
|---|---|---|---|---|
| "inheritance fund" | 12.4% | 8.9% | ~6K posts | ~24K posts |
| "urgent investment" | 14.7% | 11.2% | ~9K posts | ~32K posts |
| "business opportunity" | 8.3% | 6.1% | ~12K posts | ~48K posts |
| "lonely rich widow or widower" | 6.8% | 4.2% | ~4K posts | ~18K posts |

**Table 5.1d: Yahoo Boys (Romance/Inheritance Scams)**

Table 5.1a-d presents the statistical analysis of Facebook and Twitter's content moderation systems. The table reveals three critical structural vulnerabilities that fundamentally compromise fraud detection capabilities, with particularly disparities between how platforms handle Western-style crypto fraud (Crypto Bros) versus African-operated romance scams (Yahoo Boys). The scale disparity between content generation and moderation capacity creates an insurmountable detection gap. While Facebook processes 2.8 billion posts daily with only 200,000 reviews, Twitter handles 500 million posts with 50,000 reviews. This results in detection rates of merely 8.7% and 12.3% respectively, leaving 12.1 million fraudulent posts undetected on Facebook and 4.2 million on Twitter daily. However, the fraud type analysis reveals a troubling pattern: crypto bros achieve 27-34% detection rates using terms like "guaranteed returns" and "100x gains," while Yahoo boys operating with "inheritance fund" and "lonely widow" narratives face detection rates of only 6-14%. This disparity persists despite Yahoo boys generating 2-4x more daily content volume, indicating that the mathematical impossibility of detection disproportionately affects certain fraud types.

| **Year** | **Twitter Obfuscation Variants** | **Facebook Obfuscation Variants** | **Cross-Platform Migration (%)** | **Adaptation Time (hours)** |
|---|---|---|---|---|
| 2020 | 8 | 15 | 12.3% | 168 |
| 2021 | 16 | 28 | 16.7% | 120 |
| 2022 | 24 | 41 | 19.4% | 96 |
| 2023 | 31 | 52 | 21.8% | 84 |
| 2024 | 35 | 58 | 24.2% | 72 |

**Table 5.2a: Adversarial Adaptation: Linguistic Obfuscation and Platform Hopscotch. Evolution Timeline: Twitter vs Facebook (2020-2024)**

| Behavioral Marker | Twitter Detection (%) | Facebook Detection (%) | False Positive Rate (Twitter) | False Positive Rate (Facebook) |
|---|---|---|---|---|
| Verified account pump schemes | 67.3% | 54.2% | 8.1% | 12.4% |
| Technical analysis spam | 45.7% | 38.9% | 15.7% | 22.3% |
| Coordinated hashtag campaigns | 52.1% | 43.6% | 11.2% | 18.7% |
| Celebrity endorsement spoofing | 71.8% | 62.4% | 6.3% | 9.8% |

**Table 5.2b: Behavioral Pattern Detection: Operational Differences (Crypto Bros Behavioral Markers)**

| Behavioral Marker | Twitter Detection (%) | Facebook Detection (%) | False Positive Rate (Twitter) | False Positive Rate (Facebook) |
|---|---|---|---|---|
| Rapid-follower accounts | 18.9% | 14.2% | 28.7% | 35.4% |
| Fake profile photos (stock) | 22.3% | 16.8% | 24.1% | 31.6% |
| Romance narrative patterns | 12.7% | 8.4% | 34.9% | 42.8% |
| Pidgin code-switching | 6.8% | 4.1% | 51.2% | 63.7% |

**Table 5.2c: Yahoo Boys Behavioral Markers**

| Migration Direction | Success Rate (%) | Detection Lag (days) | Bridge Node Count | Network Density |
|---|---|---|---|---|
| Twitter → Facebook | 78.4% | 15.2 | 456 | 0.62 |
| Facebook → Twitter | 82.1% | 12.8 | 389 | 0.58 |
| Cross-posting (simultaneous) | 91.3% | 8.4 | 723 | 0.74 |

**Table 5.2d Platform Migration Analysis**

| Language Category | Twitter Detection (%) | Facebook Detection (%) | Twitter Volume (K/day) | Facebook Volume (K/day) |
|---|---|---|---|---|
| English | 87.3% | 84.6% | 320 | 1,200 |
| Spanish | 78.2% | 76.8% | 85 | 340 |
| French | 72.1% | 69.4% | 32 | 128 |
| Portuguese | 65.4% | 62.7% | 28 | 156 |
| Arabic | 32.8% | 28.9% | 45 | 180 |
| **Nigerian Pidgin** | **23.7%** | **19.2%** | **15** | **48** |
| Swahili | 19.8% | 16.4% | 8 | 32 |
| Yoruba | 18.4% | 15.1% | 6 | 24 |
| Igbo | 16.2% | 13.8% | 4 | 18 |
| Hausa | 15.7% | 12.3% | 5 | 22 |

**Table 5.2e: Language Detection Rates: Twitter vs Facebook**

| Language Context | Twitter Bridge Nodes | Facebook Bridge Nodes | Avg Connections (Twitter) | Avg Connections (Facebook) |
|---|---|---|---|---|
| English-dominant | 89 | 234 | 12.4 | 18.7 |
| Major European | 156 | 389 | 15.8 | 22.3 |
| Major Asian | 234 | 567 | 19.2 | 28.9 |
| **African Languages** | **456** | **1,123** | **34.7** | **45.2** |
| **Pidgin/Creole** | **289** | **723** | **28.9** | **38.4** |

**Table 5.2f: Network Analysis of Bridge Nodes by Platform & Language**

| Region | Twitter Detection (%) | Facebook Detection (%) | Investment Ratio (Twitter) | Investment Ratio (Facebook) |
|---|---|---|---|---|
| North America | 86.4% | 83.7% | 100x | 100x |
| Western Europe | 79.3% | 76.8% | 85x | 80x |
| Latin America | 67.1% | 64.2% | 25x | 30x |
| Middle East | 31.2% | 27.8% | 12x | 15x |
| **Sub-Saharan Africa** | **19.7%** | **16.4%** | **3x** | **5x** |
| **West Africa (Pidgin)** | **15.3%** | **12.1%** | **1x** | **2x** |

**Table 5.2g: Geographic Bias Analysis**

| Platform | Bridge Node Type | Connection Success (%) | Fraud Spread Rate | Detection Time (hours) |
|---|---|---|---|---|
| **Twitter** | English→Pidgin | 89.3% | 2.4 nodes/hour | 72.1 |
| **Twitter** | Verified→Unverified | 76.8% | 1.8 nodes/hour | 48.3 |
| **Facebook** | Group Admin Bridge | 92.7% | 3.2 nodes/hour | 96.4 |
| **Facebook** | Page→Personal | 85.4% | 2.1 nodes/hour | 84.7 |

**Table 5.2h: Fraud Propagation Through Bridge Nodes**

| Data Source | Posts Analyzed | Languages Detected | Bridge Nodes Identified | Fraud Networks Mapped |
|---|---|---|---|---|
| **Twitter API v2** | 2.4M posts | 47 languages | 1,234 nodes | 89 networks |
| **Facebook CrowdTangle** | 8.7M posts | 52 languages | 3,456 nodes | 156 networks |

**Table 5.2i: API Data Collection Summary (2020-2024)**

| Vulnerability | Twitter Impact | Facebook Impact | Combined Risk Score |
|---|---|---|---|
| **Scale Overwhelm** | 87.7% undetected | 91.3% undetected | 9.4/10 |
| **Adversarial Adaptation** | 72hr response lag | 84hr response lag | 8.9/10 |
| **Language Bias** | 4.1x disparity | 5.2x disparity | 9.8/10 |

**Table 5.2j: Key Vulnerability Metrics**

| Language/Dialect | Platform | Detection Rate (%) | Network Penetration (%) | Risk Level |
|---|---|---|---|---|
| Nigerian Pidgin | Twitter | 23.7% | 78.3% | Critical |
| Nigerian Pidgin | Facebook | 19.2% | 84.6% | Critical |
| Cameroon Pidgin | Twitter | 12.4% | 89.1% | Critical |
| Cameroon Pidgin | Facebook | 8.7% | 92.3% | Critical |
| Hausa | Twitter | 15.7% | 81.2% | High |
| Hausa | Facebook | 12.3% | 86.4% | High |

**Table 5.2k: Most Vulnerable Language Communities**

Table 5.2a-k presents systemic bias in algorithm content moderation system. The tables indicates that content moderation systems exhibit severe systemic bias, disproportionately failing non-English and African cultural contexts. Detection rates for Nigerian Pidgin are critically low (Twitter: 23.7%, Facebook: 19.2%) compared to English (Twitter: 87.3%, Facebook: 84.6%). This creates a two-tiered system: "Crypto Bros" scams (Western-facing) face nuanced detection (27-34% rate, 6-23% false positives), while "Yahoo Boys" scams (targeting African audiences) suffer crude filtering (6-14% detection, 51-64% false positives). This neglect is compounded by a significant resource allocation gap, with Sub-Saharan Africa receiving only 3-5x resources versus North America's 100x baseline, leaving vulnerable populations exposed to greater harm from less-detected fraud.

| Scammer Type | Target Demographics | Profiling Vectors | Exploitation Tactics | Platform Focus |
|---|---|---|---|---|
| **Yahoo Boys** | Lonely seniors, divorced women, military families | Romance/loneliness, isolation, military deployment | Romance scams, inheritance fraud, emergency schemes | Facebook, dating apps, email |
| **Crypto Bros** | Tech workers, young investors, finance enthusiasts | Crypto knowledge, investment appetite, tech savviness | DeFi rug pulls, fake exchanges, pump-and-dump schemes | Twitter, Discord, Telegram, LinkedIn |

**Table 6.1: Weaponizing AdTech - Interest-Based Profiling Exploitation**

| Scammer Type | Bridge Node Strategy | Network Position | Trust Capital Exploited | Detection Patterns |
|---|---|---|---|---|
| **Yahoo Boys** | Stolen profile impersonation, fake military/doctor personas | Peripheral infiltration into support groups | Emotional vulnerability, authority respect | Inconsistent time zones, grammar patterns, photo mismatches |
| **Crypto Bros** | Compromised influencer accounts, fake startup founders | Central positions in tech/finance communities | Technical credibility, financial success imagery | Sudden pivot to specific projects, engagement farming, technical inconsistencies |

**Table 6.2: Network Analysis - Bridge Node Identification**

| Scammer Type | Primary Victims | Psychological Manipulation | Financial Impact | Success Indicators |
|---|---|---|---|---|
| **Yahoo Boys** | Women 45-65, military families, recent widows/divorcees | Emotional dependency, romantic love, family emergencies | $500-$50K+ (average: $12K) | Long-term relationship building, emotional investment |
| **Crypto Bros** | Males 18-40, tech professionals, investment groups | FOMO, greed, intellectual superiority, exclusive access | $1K-$100K+ (average: $25K) | Quick decisive action, fear of missing opportunities |

**Table 6.3: Targeted Financial Harm - Interview Insights**

The three tables 6.1-3 reveal fundamentally distinct operational frameworks between Yahoo Boys and Crypto Bros, reflecting different exploitation philosophies and victim demographics. Yahoo Boys employ emotional manipulation architecture, targeting psychologically vulnerable populations through sustained relationship building.

Their node bridge tactic entails peripheral incursion within groups of support, taking advantage of trust by impersonating authority (military personnel, doctors) that convey security and reliability. The average financial impact ($12K) indicates lengthy cycles of engagement wherein victims invest incrementally through emotional reliance. Crypto Bros, however, operate on the basis of technical authority manipulation, approaching financially motivated viewers with higher disposable incomes. Their bridge nodes occupy key places in established tech and finance communities, pilfering from under strength high-credibility accounts to transfer legitimacy in a hurry. Their considerably higher average financial impact ($25K) shows that they are able to draw upon larger sums through FOMO-driven decision-making and self-proclaimed special access to lucrative opportunities. Critical divergence manifests in temporal strategies: Yahoo Boys spend months constructing emotional capital before they monetize, while Crypto Bros leverage market momentum and time-sensitive moments for immediate extraction. Platform selection reflects these inclinations—Yahoo Boys use relationship-oriented platforms (Facebook, dating apps) that support extended personal interaction, while Crypto Bros dominate finance-oriented channels (Twitter, Discord, Telegram) where speedy information dispersal and technical savvy hold most value. Both groups have high social network awareness, but Yahoo Boys target trust through vulnerability and Crypto Bros target trust through subject matter authority. This underlying distinction drives their own targeting effectiveness, with Yahoo Boys achieving higher psychological penetration and Crypto Bros achieving broader financial reach through volume-based strategies.

## Discussion

### The Architectural Complicity of Platforms in Financial Fraud Amplification

Quantitative empirical data exhibited in Figures 1-4 and Table 4.1 is conclusive proof that social media sites are not passive hosts but active, albeit often inadvertent, architects of a new virtual financial risk environment.

### Finding 1: Algorithmic Amplification of Fraudulent Content - The Core Mechanism

The foundational insight from Figure 1 is staggering: fraudulent content achieves an amplification factor of 67x compared to a mere 8x for legitimate content. This is not a marginal difference but an orders-of-magnitude disparity that fundamentally alters the scale and reach of financial scams. This amplification is the direct result of the operational logic of engagement-based algorithms (Caplan & boyd, 2018). These systems are designed with a singular, economically-driven purpose: to maximize user time-on-platform and interaction, as these metrics directly correlate with advertising revenue (Srnicek, 2017; Zuboff, 2019). The algorithms function as predictive engines, identifying content likely to elicit a reaction—any reaction—and privileging it within user feeds. This design creates a perverse incentive structure where the economic logic of the platform (maximize engagement) directly conflicts with its professed social responsibility (user safety) (Gillespie, 2018). Figure 2 introduces a critical paradox: higher amplification rates are correlated with faster detection times. This suggests that the very virality that makes a scam successful also makes it more visible to platform moderators. However, as Table 4.1 clarifies, this correlation is functionally meaningless against the sheer, overwhelming volume of content. With Twitter detecting only 12.3% of an estimated 500 million daily posts, and Facebook a mere 8.7% of 2.8 billion, the systems are mathematically incapable of proactive defense (Gorwa et al., 2020; Roberts, 2019). The platforms operate a form of "performative security" (Pasquale, 2015), where the appearance of action belies a structural inability to manage the scale of harm their own systems incentivize. The economic logic of scale and growth invariably trumps the cost of mitigating harm, which is treated as an externality (van Dijck et al., 2018).

### Hijacking Engagement: How Emotional Arousal Drives Algorithmic Prioritization

The mechanism behind the 67x amplification is precisely detailed in Figures 4.1a and 4.1b. The data demonstrates a powerful causal relationship: content engineered to provoke high-arousal emotions, particularly Fear and FOMO (Fear Of Missing Out), receives an 87% algorithmic boost. This finding empirically validates long-standing research in communication and psychology showing that content eliciting high-arousal emotions is more likely to be shared—a phenomenon known as "emotional contagion" in digital networks (Brady et al., 2017; Kramer et al., 2014).

Fraudsters are masterful behavioral psychologists who intuitively understand this algorithmic preference. They craft narratives not of sober financial advice, but of life-changing opportunity (excitement/desire) or impending loss (fear/anxiety). As Figure 4.1b devastatingly shows, 95% of financial investment content leveraging these high-arousal emotions is fraudulent. This indicates that the platform's algorithm has been effectively weaponized. It has been trained to prioritize a specific emotional signature that is overwhelmingly synonymous with scams. This is a catastrophic failure of design. The algorithm, agnostic to truth (Vaidhyanathan, 2018), cannot distinguish between genuine excitement for a product launch and manufactured excitement for a "pump-and-dump" scheme. It reads both as "high engagement" and rewards them equally. This process aligns with Zuboff's (2019) concept of "surveillance capitalism," where human experience is rendered as behavioral data to be used for prediction and modification. In this case, the emotional vulnerability of users is the raw material that both the platform and the fraudster seek to harvest. The technical logic of the algorithm—prioritize engagement metrics—creates an economic logic for the fraudster: invest resources in emotionally manipulative content to hijack the platform's distribution infrastructure for free (Marwick & Lewis, 2017).

### Comparative Analysis: Divergent Amplification Pathways for Crypto (Crypto-bros) and Advance-Fee Scams (Yahoo-boys)

Figures 4.2a and 4.2b provide a crucial comparative lens, revealing that not all fraud is amplified equally. The data shows cryptocurrency scams achieve 17x faster amplification than traditional advance-fee ("Yahoo Boy") scams. This divergence is not cultural but deeply technical, revealing how different scam architectures interact with platform

mechanics. The "Crypto Bros" archetype is perfectly evolved for the platform environment. As Figure 4.2b indicates, they exploit features like rapid "hashtag velocity" (91%), where the speed at which a hashtag like #BitcoinGiveaway is used triggers algorithmic thresholds for trendification and inclusion in topical "For You" feeds (Rogers, 2018). Their content is designed for rapid, low-commitment engagement—a retweet, a like, a click—actions that algorithms heavily weight. This model aligns with what Turner (2006) identified as the "tech-libertarian" ideology of Silicon Valley, a worldview the Crypto Bros' rhetoric of disruption and decentralization directly mirrors. The platform's algorithms, born from this same culture, thus inherently privilege this form of content (Bishop, 2021).

In contrast, the "Yahoo Boys" model, rooted in what anthropologist James Ferguson (2006) calls "schemes of solidarity and deception," relies on slower, relational trust-building. It involves direct messages, prolonged conversations, and identity-based social engineering. This method generates engagement, but it is slower, more private, and less likely to involve public, trendable markers like viral hashtags. Consequently, it does not trigger the same rapid-fire viral mechanisms. This divergence proves that algorithmic amplification is not a monolithic force but one that interacts with the cultural and technical specificities of content, actively shaping the evolution of fraud itself (Tufecki, 2018).

## Case Study: The #BitcoinGiveaway and the "For You" Feed

Figure 4.3 offers a microscopic view of the amplification process in action, focusing on the temporal dimension. The data reveals that algorithmic amplification is not constant but peaks during "prime time" hours (16:00-20:00), where content benefits from a 12.5x organic reach multiplier and an astonishing 67.3% engagement rate. This finding demonstrates a sophisticated level of optimization by fraudsters. They are not merely posting randomly; they are strategically scheduling content to coincide with peak user activity. This maximizes the "network effects" (Shirky, 2008) and ensures their posts enter the system when the algorithm is most actively seeking to distribute engaging content to a large, attentive audience. The "For You" feed, a black-box algorithmic curation system designed to maximize user retention, becomes the primary vector for this fraud (Eslami et al., 2015). Users are presented with these scams not because they follow the accounts, but because the algorithm has *inferred* a potential interest based on their data profile and has determined that the high-arousal, fraudulent content will likely keep them engaged.

This is the culmination of the ad-tech ecosystem's precision. The same micro-targeting tools that allow a legitimate brand to reach "males aged 25-40 interested in cryptocurrency" allow a fraudster to target the same demographic with predatory precision (Turow, 2017; Eubanks, 2018). The platform's economic model, which profits from the granular sale of user attention, provides the very infrastructure for this predation (Citron & Pasquale, 2014). The case of #BitcoinGiveaway is thus a perfect storm: emotionally manipulative content, optimized for prime-time posting, delivered via a proprietary algorithmic feed to a pre-identified vulnerable demographic, all driven by the economic logic of surveillance capitalism.

## Finding 2: Structural Vulnerabilities in Content Moderation

The empirical data presented in Tables 5.1a through 5.1j moves the analysis from the *mechanisms* of algorithmic amplification to the profound *failures* of the systems designed to contain it. The analysis reveals that this persistence is not a mere technical challenge but a direct, predictable outcome of a business model that is fundamentally incompatible with holistic safety. The moderation apparatus is structurally overwhelmed, deliberately evaded, and systematically biased, creating a tiered system of protection that leaves the most vulnerable users exposed to the most devastating harms.

The content moderation systems of major platforms are often publicly framed as a relentless battle against bad actors. However, the data reveals a different reality: a set of deeply embedded structural vulnerabilities that function less like a weakened immune system and more like designed features of a platform political economy that prioritizes growth and data extraction over user safety (Gillespie, 2018; Klonick, 2018). These vulnerabilities ensure that fraud operates not on the fringes but within the core logic of the platform ecosystem.

## Scale and Opacity: The Impossibility of Proactive Detection

The most fundamental vulnerability is one of sheer, insurmountable scale. Table 5.1a presents a staggering disparity: Facebook must somehow review 2.8 billion daily posts with a capacity of only 200,000 reviews, while Twitter handles 500 million posts with 50,000 reviews. The resulting detection rates—8.7% and 12.3% respectively—are not indicators of failure but of mathematical impossibility. This chasm between content volume and human review capacity creates an ocean of undetected harm: 12.1 million fraudulent posts daily on Facebook and 4.2 million on Twitter.

This is the operational reality of what content moderators and scholars describe as the "whack-a-mole" problem (Roberts, 2019). Platforms are forever reactive, addressing harm only after it has achieved significant scale and visibility, never proactively preventing it. This reactive stance is a rational, if cynical, economic calculation. The cost of hiring sufficient moderators to vet even a fraction of this content proactively would be astronomically high and would directly undermine the profit margins that depend on limitless, low-cost scalability (Suzor, 2019). Therefore, a certain level of "acceptable harm" is baked into the business model (Gorwa, 2019). The opacity of these systems, as detailed by Pasquale (2015) in his concept of the "black box society," further protects the platforms from accountability, making the true scale of the failure difficult for outsiders to quantify—until studies like this one force it into the open.

The vulnerability is exacerbated by the near-total reliance on outsourced moderation labor. This model, documented by scholars like Chen (2014) and Newton (2019), creates a disassociated, high-turnover workforce that is often traumatized, poorly trained, and working under impossible time pressures. A moderator might have mere seconds to judge a post, lacking the cultural context or linguistic nuance to identify sophisticated scams, a point tragically illustrated by the data on non-English content. This outsourcing is not just an operational choice but a moral one, allowing Silicon Valley firms to externalize the human cost of their business model onto a precarious, invisible workforce (Ticona & Mateescu, 2018).

## Adversarial Adaptation: Linguistic Obfuscation and Platform Hopscotch

Fraudsters are not passive entities; they are adaptive entrepreneurs who engage in a continuous arms race with platform moderation systems. Table 5.2a meticulously documents this "adversarial adaptation," showing a year-on-year increase in obfuscation variants (e.g., "grAnt opportUnity") and a corresponding decrease in adaptation time from 168 hours in 2020 to 72 hours in 2024. This is a process of algorithmic aversion, where scammers deliberately mutate their linguistic signatures to evade automated Natural Language Processing (NLP) detection systems (Zannettou et al., 2019). This practice, often called "algospeak" or "leetspeak," transforms language into a tool of deception. It exploits a core weakness in automated moderation: its reliance on pattern matching and its inability to understand context and intent (Sap et al., 2019). A human can easily see that "$BTC give@way" is an obvious evasion of "#BitcoinGiveaway," but a keyword-based filter may not. This forces platforms into a continuous game of catch-up, constantly updating their blocklists after the new variants have already achieved viral spread.

Furthermore, Table 5.2a shows a rising cross-platform migration rate (24.2% in 2024). This "platform hopscotch" strategy involves launching a campaign on one platform and, once detected or amplified, swiftly migrating the successful framework to another. Table 5.2d quantifies the high success rates of this migration (78.4% from Twitter to Facebook, 91.3% for simultaneous cross-posting). This highlights a critical policy inconsistency: the lack of effective inter-platform intelligence sharing. Despite initiatives like the Global Internet Forum to Counter Terrorism (GIFCT), no equivalent exists for financial fraud, allowing actors to exploit the siloed nature of platform governance (Kaye, 2019). Each platform protects its proprietary data and algorithms, leaving the broader ecosystem vulnerable to coordinated, cross-platform attacks.

## Systemic Bias: The Failure to Moderate Non-English and Pidgin Content

The most damning vulnerability exposed by the data is not a failure of technology but one of equity. The moderation regime exhibits a profound systemic bias that creates a tiered system of user protection, heavily skewed towards English-speaking users in the Global North. Table 5.2e presents a devastating gradient of protection. While English content enjoys a 87.3% detection rate on Twitter, this plummets to 23.7% for Nigerian Pidgin and below 20% for languages like Yoruba and Igbo. This disparity is not marginal; it is a chasm. This bias stems from multiple sources. First, NLP models are overwhelmingly trained on large, publicly available corpora of standard English text,

rendering them ineffective at parsing the grammatical structures, code-switching, and colloquialisms of pidgins and creoles (Hovy & Spruit, 2016). Second, the human moderators are disproportionately located in hubs like Dublin and Manila and are often unfamiliar with the cultural and linguistic nuances of West African online communities (Gray & Suri, 2019).

The consequences of this bias are catastrophic and are detailed in Tables 5.2f, 5.2g, and 5.2k. Because African-language content is so poorly moderated, it becomes a fertile ground for fraud. Table 5.2f shows that bridge nodes in African language contexts have nearly triple the average connections (34.7 on Twitter) than those in English-dominant networks (12.4). These highly connected nodes can rapidly spread fraud through entire communities with little resistance from moderation algorithms. Table 5.2g translates this into a grotesque "investment ratio," showing that resource allocation for moderation in Sub-Saharan Africa is a mere 3-5x compared to a 100x baseline in North America. This creates the perverse outcome detailed in Tables 5.1c and 5.1d: "Crypto Bros" scams (primarily targeting English-speaking audiences) achieve higher detection rates (27-34%) than "Yahoo Boys" scams (6-14%), despite the latter often generating higher daily volumes. The system is effectively optimized to protect a privileged demographic while abandoning others. This is a textbook example of what Benjamin (2019) calls "discriminatory design" and what Eubanks (2018) identifies in the automation of inequality, where technological systems reproduce and amplify existing social inequities. The platforms' policy inconsistency—applying sophisticated AI to protect wealthy Western users while offering only crude, ineffective filters to the Global South—constitutes a form of digital colonialism (Couldry & Mejias, 2019), where the economic value of some users is deemed higher than the safety of others.

## Finding 3: Precision Targeting of Vulnerable Demographics

The empirical data in Tables 6.1-6.3, when combined with the poignant narratives from the interview responses, moves the analysis from the *systemic* failures of platforms to the *intimate* harm inflicted upon individuals. The platform economy's lifeblood is the hyper-efficient matching of content to audiences. This is achieved through a sophisticated apparatus of data harvesting, algorithmic profiling, and micro-targeting, built for commercial persuasion (Turow, 2017; Zuboff, 2019). However, as the data demonstrates, this same apparatus is ideologically neutral and can be repurposed for fraud with minimal friction. The result is not a broader scattering of scams, but a frighteningly precise concentration of predatory content aimed at those most susceptible to it. This represents the ultimate perversion of the surveillance capitalist model: the use of its core machinery to systematically prey upon the very users it claims to serve.

## Weaponizing AdTech: How Scammers Exploit Interest-Based Profiling

Table 6.1 outlines the different demographic and psychographic profiles targeted by Yahoo Boys and Crypto Bros. This granularity is not guesswork; it is the direct result of scammers leveraging the precise "Profiling Vectors" and "Exploitation Tactics" listed. The platform's ad-tech ecosystem provides them with the tools to do this. Legitimate advertisers use these tools to target users based on "intents" (e.g., "interested in cryptocurrency") or "life stages" (e.g., "newly engaged"). Fraudsters, however, target "vulnerabilities." As Agent R. (I5) notes from the study interview, scammers "buy the same demographic data legitimate businesses use, but they're selecting for vulnerability instead of purchasing power." A user who joins a "Grief Support" group on Facebook or searches for "get out of debt" becomes a data point in a "loneliness" or "financial anxiety" cluster, making them a prime candidate for a romance or inheritance scam (Andrejevic, 2013). Similarly, a user who follows crypto influencers on X and engages with technical DeFi content is algorithmically grouped into a high-value target for a "pump-and-dump" scheme. This process exemplifies what Cohen (2013) describes as the "biopolitical public domain," where personal life is rendered into commercial categories. The interviews provide devastating proof of its effectiveness. V5 (interviewee) notes the scammer "knew details about my late husband from my Facebook posts," demonstrating successful profiling for emotional vulnerability. C5 (interviewee) describes being "added to a private Discord with 'successful traders,'" showing how affinity-based targeting is used to create a false community that fosters trust and triggers FOMO (Fear of Missing Out). The platform's economic logic, which seeks to categorize and commodify every aspect of human experience, provides the fraudster with a ready-made map of human weakness (Couldry & Mejias, 2019).

## Network Analysis: Identifying "Bridge Nodes" to Legitimate Communities

Perhaps the most sophisticated exploitation of platform architecture is the strategic use of "bridge nodes," as detailed in Table 6.2. Scammers understand that credibility cannot be manufactured from nothing; it must be stolen. They do this by identifying and compromising accounts that sit at the periphery of high-trust communities.

A "bridge node" is an account that holds social capital within a legitimate community—a respected member of a veterans' group, the admin of a crypto investment club, or a fitness influencer with a loyal following. Inspector Chen (I3) confirms this: "They compromise legitimate accounts with followers, then gradually shift content. A fitness influencer suddenly promoting crypto... The existing trust transfers immediately to the scam."

Table 6.2 shows the strategic differences: Yahoo Boys use "Peripheral infiltration" with "Stolen profile impersonation" (e.g., fake military/doctor personas) to exploit "Emotional vulnerability." This is evident in V1 (Sarah) and V4 (Robert)'s stories, where the use of military terminology granted immediate, unearned authority. Conversely, Crypto Bros target "Central positions in tech/finance communities" using "Compromised influencer accounts" to exploit "Technical credibility." C1 (Alex) and C7 (Ryan) fell victim to this, trusting "verified developers" and influencers with "blue checkmarks."

This strategy exploits a fundamental tenet of network theory: trust is transitive (Granovetter, 1973). By hijacking a trusted node, the scammer gains indirect access to the entire network's social capital, bypassing the skepticism that would greet a cold contact. The platform's algorithms, designed to promote content from influential accounts, then actively amplify the fraudulent message to the wider community, mistaking stolen credibility for genuine authority (Freelon, 2018).

Interview Insights: The Lived Experience of Targeted Financial Harm

The quantitative data in the tables is given profound human depth by the interview responses. They move the discussion from abstract "profiling vectors" to the lived reality of being precisely targeted. The psychological impact differs by scam type, reflecting the tailored nature of the attacks. Yahoo Boys exploit a deep need for connection and love. V1 (Interviewee) articulates the core vulnerability: "when you're lonely and someone says they love you every morning..." The scammers, as Sergeant W. (I4) notes, act like "relationship therapists turned predators," engaging in "surgical" emotional manipulation (V5). The financial loss is compounded by profound shame and psychological trauma, making victims reluctant to report (I4).

Crypto Bros, by contrast, exploit greed, intellectual vanity, and a fear of being left behind. Their targets are often financially literate and tech-savvy individuals like C1 (Software Developer) and C3 (Data Analyst), who pride themselves on their ability to evaluate opportunities. This makes them particularly vulnerable, as *Law Enforcement Agent* (I5) observes: "They think they're too smart to be scammed, which makes them perfect targets." The betrayal is not just financial but intellectual, as their own competence is used against them. The interviews with investigators and cybersecurity experts (I1-I8, S1-S8) confirm the industrial scale of this targeting. *Agent D*. (I8) speaks of "organized operations with specialization," while (S3) describes "advanced psychological profiling" using "data brokers [and] social media scraping." This is not random crime; it is a data-driven enterprise that leverages the platform's own tools for victim selection, engagement, and exploitation.

## Practical Implications: The need for Restructuring

The foregoing findings implies that platform-aided financial fraud is not an edge case but a direct and inevitable outcome of the underlying architectures and economic models of social media. The algorithmic promotion of financial disinformation and fraud content, the structurally broken content moderation design, and the weaponization of ad-tech for precision profiling all combined, display a system in critical crisis. Platforms are not neutral pipes but engaged, albeit unwilling, architects of a new digital risk landscape. The implications are severe and tripartite. First, the resolve to do anything for engagement has constructed an algorithmic economy that unequivocally promotes deception and emotional manipulation. Second, moderation scale and design are statistically incapable of protecting users, with system biases disproportionately attacking non-English speakers and Global South populations. Third, the same surveillance capitalism tools are being reverse-engineered to cash in on human vulnerability with horror-show effectiveness in the financial arena this study has uncovered. Hence, while the continuous improvement in global information and communication technology is observable and applaudable, there remain an inherent

imbalance in the superstructure and construct of social media algorithms that makes it capable of retrogressing the human development and advancement.

## Conclusion

This study empirically demonstrates that social media sites are active participant-enablers of financial disinformation and fraud. Algorithmic networks propagate fraudulent content 67x earlier than legitimate material by preferentially amplifying high-arousal emotions, which are 95% predictable for scams. Structural vulnerabilities are catastrophic: mathematically insurmountable scale and systemic bias yield lamentable detection rates for non-English content (e.g., 19.2% for Nigerian Pidgin). Besides, the tools of surveillance capitalism itself are weaponized: precision ad-targeting and delivers predatory material to vulnerable populations. Hence, the infrastructure of connection is also an infrastructure of predation, which necessitates radical regulatory and algorithmic change. We thus conclude that technical solutions within the current platform paradigm are insufficient. To abate this crisis requires a rethink at the foundation of the model of engagement-driven economics, the use of ethically designed algorithms that prioritize safety over virality, and robust regulatory frameworks that encourage platforms to be held accountable for the harms their designs consistently produce. The design of connection should no longer be a design of predation.

## References

Andrejevic, M. (2013). Infoglut: How too much information is changing the way we think and know. Routledge.

Bell, D. (1973). The coming of post-industrial society: A venture in social forecasting. Basic Books. https://www.basicbooks.com/titles/daniel-bell/the-coming-of-post-industrial-society/9780465097135/

Benjamin, R. (2019). Race after technology: Abolitionist tools for the new Jim code. Polity Press.

Bishop, S. (2021). Algorithmic entrepreneurs: The obfuscation of financial harm on social media. New Media & Society.

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. Proceedings of the National Academy of Sciences, 114(28), 7313-7318.

Caplan, R., & boyd, d. (2018). Isomorphism through algorithms: Institutional dependencies in the case of Facebook. Big Data & Society, 5(1). https://doi.org/10.1177/2053951718757253

Caplan, R., Hanson, L., & Donovan, J. (2018). Dead reckoning: Navigating content moderation after "fake news". Data & Society Research Institute.

Castells, M. (1996). The rise of the network society. Blackwell.

Chen, A. (2014). The laborers who keep dick pics and beheadings out of your Facebook feed. Wired.

Citron, D. K., & Pasquale, F. (2014). The scored society: Due process for automated predictions. Washington Law Review, 89(1), 1-33. https://digitalcommons.law.uw.edu/wlr/vol89/iss1/2/

Cohen, J. E. (2013). What privacy is for. Harvard Law Review, 126(7), 1904-1933.

Couldry, N., & Mejias, U. A. (2019). The costs of connection: How data is colonizing human life and appropriating it for capitalism. Stanford University Press.

Crain, M. (2018). The limits of transparency: Data brokers and commodification. New Media & Society, 20(1), 88-104.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) (pp. 4171-4186).

Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahalios, K., Hamilton, K., & Sandvig, C. (2015). "I always assumed that I wasn't really that close to [her]": Reasoning about invisible algorithms in news feeds. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (pp. 153-162). https://doi.org/10.1145/2702123.2702556

Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.

Ferguson, J. (2006). Global shadows: Africa in the neoliberal world order. Duke University Press. https://doi.org/10.1215/9780822387640

Fourcade, M., & Healy, K. (2017). Seeing like a market. Socio-Economic Review, 15(1), 9-29.

Freeman, L. C. (2004). The development of social network analysis: A study in the sociology of science. Empirical Press.

Gillespie, T. (2018). Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media. Yale University Press. https://yalebooks.yale.edu/book/9780300235029/custodians-of-the-internet/

Gorwa, R. (2019). What is platform governance? Information, Communication & Society, 22(6), 854-871.

Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. Big Data & Society, 7(1).

Granovetter, M. S. (1973). The strength of weak ties. American Journal of Sociology, 78(6), 1360-1380.

Gray, M. L., & Suri, S. (2019). Ghost work: How to stop Silicon Valley from building a new global underclass. Houghton Mifflin Harcourt.

Hovy, D., & Spruit, S. L. (2016). The social impact of natural language processing. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers) (pp. 591-598).

Kaye, D. (2019). Speech police: The global struggle to govern the Internet. Columbia Global Reports.

Klonick, K. (2018). The new governors: The people, rules, and processes governing online speech. Harvard Law Review, 131(6), 1598-1670.

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. Proceedings of the National Academy of Sciences, 111(24), 8788-8790.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (Vol. 30, pp. 4765-4774).

Marwick, A., & Lewis, R. (2017). Media manipulation and disinformation online. Data & Society Research Institute.

Newton, C. (2019). The trauma floor: The secret lives of Facebook moderators in America. The Verge.

Pasquale, F. (2015). The black box society: The secret algorithms that control money and information. Harvard University Press.

Roberts, S. T. (2019). Behind the screen: Content moderation in the shadows of social media. Yale University Press. https://yalebooks.yale.edu/book/9780300245318/behind-the-screen/

Rogers, R. (2018). Otherwise engaged: Social media from vanity metrics to critical analytics. International Journal of Communication, 12, 450-472.

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). Auditing algorithms: Research methods for detecting discrimination on Internet platforms. Data and Discrimination: Converting Critical Concerns into Productive Inquiry, 1-23.

Sap, M., Card, D., Gabriel, S., Choi, Y., & Smith, N. A. (2019). The risk of racial bias in hate speech detection. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 1668-1678).

Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. Big Data & Society, 4(2). https://doi.org/10.1177/2053951717738104

Seaver, N. (2018). Captivating algorithms: Recommender systems as traps. Journal of Material Culture, 24(4), 421-436.

Shirky, C. (2008). Here comes everybody: The power of organizing without organizations. Penguin Press.

Srnicek, N. (2017). Platform capitalism. Polity Press.

Suzor, N. P. (2019). Lawless: The secret rules that govern our digital lives. Cambridge University Press.

Ticona, J., & Mateescu, A. (2018). Trusted strangers: Carework platforms' cultural entrepreneurship in the on-demand economy. New Media & Society, 20(11), 4384-4404.

Turner, F. (2006). From counterculture to cyberculture: Stewart Brand, the Whole Earth Network, and the rise of digital utopianism. University of Chicago Press. https://press.uchicago.edu/ucp/books/book/chicago/F/bo3615170.html

Turow, J. (2017). The aisles have eyes: How retailers track your shopping, strip your privacy, and define your power. Yale University Press.

Vaidhyanathan, S. (2018). Antisocial media: How Facebook disconnects us and undermines democracy. Oxford University Press. https://global.oup.com/academic/product/antisocial-media-9780190841164

van Dijck, J., Poell, T., & de Waal, M. (2018). The platform society: Public values in a connective world. Oxford University Press. https://academic.oup.com/book/27308

Zannettou, S., Caulfield, T., Blackburn, J., De Cristofaro, E., Sirivianos, M., & Stringhini, G. (2019). On the origins of memes by means of fringe web communities. In Proceedings of the Internet Measurement Conference (pp. 188-202).

Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. PublicAffairs.

## Authors' Biography

**Nathan Oguche Emmanuel, PhD** is a lecturer in the Department of Mass Communication, National Open University of Nigeria (NOUN). Before joining NOUN, he taught at Kogi State University, now Prince Abubakar Audu University, Anyigba, where he also served as Senior Communication Officer at the Directorate of Advancement and Support Centre. In this role, he galvanized support for the university from its stakeholders through strategic communication, engagements, and the development of policies that resonate with the university's constituents. His areas of research interest include journalism, new and digital media and their intersection with journalism, media and communication practice. He has published in local and international journals such as *Journalism Practice*, *Journal of African Media Studies*, *Music Psychology*, *Media Practice and Education*, *African Media Studies*, among others.

**Samuel Sunday AMEH** is a researcher and a graduate student of the University of Nigeria specializing in Digital Communication in emerging social media platforms and their impact on public policy discourse. His works examine how AI and new media technologies transform finance, Agriculture, humanitarian storytelling and public engagement with social issues. He has published and delivered papers in both international and local conferences.