Uncovering the Trust Signals Supporting Telegram's Cybercrime Economy

Roy Ricaldi*, Tina Marjanov†, Luca Allodi*, Alice Hutchings†

*Eindhoven University of Technology, Department of Mathematics and Computer Science
{r.j.ricaldi.saavedra, l.allodi}@tue.nl

†University of Cambridge, Department of Computer Science and Technology
{tina.marjanov, alice.hutchings}@cl.cam.ac.uk

***Abstract*—Telegram has become a central hub for cybercriminal activity, favored for its perceived privacy, user anonymity, ease of use, and the many features it offers. Unlike traditional markets on underground forums Telegram lacks many structural elements of trust, such as stable identities and reputation within a community. This raises important questions about whether and how trust is built in these newer, more fluid marketplace environments. In our work, we characterize the Telegram cybercrime ecosystem by identifying key market segments and developing a framework of trust-building mechanisms that support trade within those segments. We apply this framework at scale across 1,116,071 messages from 167 Telegram cybercriminal communities. Our analysis shows that although trust signals are fewer than on forums and are often sparsely distributed, cybercriminals on Telegram still actively signal trust using various strategies, from proof-of-delivery and vouching messages to pinned rules and automated bots. To estimate how frequently these signals are actually encountered by users, we implement a Monte Carlo simulation that models cybercriminal browsing behavior across different market segments. Our results reveal that users in different segments are exposed to different levels and type of trust signaling, and that exposure varies significantly with time. Together, our findings suggest that Telegram differs substantially from cybercriminal forums in supporting cybercriminal activities, offering a fragmented but evolving economic ecosystem for threat actors to operate in.**

***Index Terms*—Cybercrime, Telegram, Trust, Illicit Markets**

I. Introduction

Telegram is an increasingly popular platform for cybercriminal activity [1]. Compared to traditional forum-based cybercriminal marketplaces, Telegram is easily accessible and there are generally low barriers to joining communities (e.g. registration). Telegram comes with a reputation for private communications that has, allegedly, made it an increasingly popular choice for certain types of cybercriminal activities. These include the sale of stolen data, malware, hacking tools, and other illegal services.

On the other hand, immediate availability and ease of use are not the only important mechanisms for trade. For the Telegram ecosystem to credibly support trade of illegal products, much of the same foundational problems from other ecosystems, such as the Internet Relay Chat [2] and, more recently, cybercrime forums [3] apply. These problems undermine the ability of the market to operate efficiently under conditions of information asymmetry, that is conditions wherein information is not equally available to all parties involved. These issues can be so severe as to lead to market collapse when all 'good' traders end up being pushed out of the market by the 'bad' ones.

A market populated by 'bad' cybercriminal actors is not necessarily an 'interesting' one from a cybersecurity perspective. Data traded on there would probably be outdated or fake; malware could be ineffective, easily detectable; digital services for malware distribution could underperform, being unreliable, or simply scams. In other words, a 'bad' criminal market would have inefficiencies affecting the criminals, and thus hardly have any impact on the real world. Cybercrime offenders looking for effective and innovative attack capabilities would be forced to look elsewhere. Consequently, cyber threat intelligence derived from there may inflate sales pitch numbers for cybersecurity products, but not deliver indicators of compromise (IoCs) or other information of actual relevance to security operations and decisions.

The establishment of trust among actors remains one of the cornerstones to address market inefficiencies leading to these issues. In cybercriminal markets, trust is essential. Participants often operate anonymously and cannot rely on formal systems to resolve disputes or verify identities. As a result, they use various signals to show that they are reliable and to reduce the risk of scams. Prior work has evaluated how trust is established in forum cybercrime communities. However, there is no existing evaluation of how trust can be established in cybercriminal communities on Telegram, which lack many of the formal feedback loops available on traditional cybercrime forums.

This study focuses on how cybercriminals build trust in Telegram communities. We adopt the perspective of *signaling theory*, a dominant theory within the field of Economics and generally regarded as the leading theory explaining how actors mitigate information asymmetries in trade and governance settings [4], [5]. Specifically, we analyze how sellers or community administrators on Telegram signal trust to their users, and how exposed these signals are in the otherwise rather unstructured environment Telegram offers. We examine how these signals vary across different segments of the cybercrime market economy on Telegram.

979-8-3315-8969-1/25/$31.00 ©2025 IEEE

We make the following contributions:

- We map Telegram's cybercrime economy structure with the identification of six distinct market segments.
- We characterize how trust is signaled and varies across market segments on Telegram, exposing both the mechanisms used and the inconsistencies in their application.
- We provide user-centric insights into trust signal exposure through simulated interactions in cybercriminal Telegram communities.

We provide an overview of related work and outline the research gaps in §II. We outline our methodology in §III, leading to identified trust signals and the results of our Monte Carlo simulation in §IV. We discuss our findings in §V and conclude in §VI.

II. Related work and background

A. Conceptualizing cybercrime

Boundaries between what is and is not cybercrime can be quite vague. To define the scope of cybercrime for this research, we draw upon Dupont and Whelan's [6] definition of crimes that "*can only be committed using a computer, computer networks, or other form of information communications technology [...] and are primarily directed against computers or network resources*". Further, cybercrime refers to economically motivated illicit activities conducted through or against digital systems. It encompasses acts that involve unauthorized access, manipulation, or misuse of digital information, systems, or services for personal or financial gain. It also includes acts that facilitate, conceal, or educate others in such practices. This definition is limited to criminal behaviors that are market-oriented, excluding politically driven, state-sponsored, or ideologically motivated activities.

B. Trust signaling in underground forums

Signaling theory, rooted in economics, addresses how individuals navigate issues caused by asymmetric information such as adverse selection, which arises when the quality of goods is unknown before a transaction, and moral hazard, which occurs when one party carries the bulk of consequences, caused by the actions of the other party they are unable to observe. In a criminology context, Gambetta differentiates between 'signals', deliberate, often costly actions to demonstrate commitment or credibility, and 'signs', involuntary and difficult-to-alter characteristics (e.g., accent, social affiliation) that non-deliberately convey information about a party's type [5], [7]. This framework has been applied to analyze criminal performance and signaling in illicit markets, such as the carding underworld [8]. Participants in high-risk and anonymous environments rely on signals such as reputation to reduce information asymmetry and convey credibility, as effective signals are costly or difficult to fake [3], [9], [10]. While not perfect [11], these signals keep the markets from devolving into markets for lemons. By strategically using such signals, criminals can foster trust, mitigate perceived risks, and promote repeat transactions despite the inherently untrustworthy setting.[1]

Several studies documented the existence and use of various trust mechanisms on underground forums and dedicated dark net marketplaces. The trust mechanisms can be broadly grouped into two categories, namely market- and individual-level mechanisms. Market-level mechanisms are provided and often mandated centrally by market organizers. Trust-building mechanisms provided by markets include sponsoring or vouching for new market members and membership payments [12], optional or mandatory use of marketplace contracts [13], reputation [14], feedback and experience tracking [15], dispute resolution [16], verification of goods [17], escrow [18], and moderation [9]. Individual-level mechanisms are implemented by market participants at their discretion. They include the use of jargon [19], participating in smaller [20] and more public [21] trades, or providing free samples [22].

Campobasso et al. [3] provide and evaluation of fundamental economic issues affecting illicit markets, and a comprehensive Forum Trust Framework documenting the related mitigation mechanisms. A synthesis of the key mechanisms is presented in Table I. Reputation systems and the ability to file complaints (here synthesized under 'reputation') and transparent customer feedback help users assess trustworthiness before transacting ('transparency'), while features like escrow ('payment security') and presence of rules and enforcement mechanisms ('moderation') protect parties post-agreement. Mechanisms addressing these issues work collectively to reduce risk, level information asymmetries, and align incentives between sellers and buyers. However, it is unclear how these mechanisms translate into Telegram.

TABLE I: Forum trust framework

	Key Issue	Description
Adverse Selection	Reputation	Mechanisms that allow users to evaluate others' past behavior and reliability before transacting. Helps mitigate the risk of engaging with bad actors.
	Transparency	Mechanisms that promote visible and auditable accountability, such as public transaction logs, offering customer support, or samples. Supports credibility and deters post-agreement misconduct.
Moral Hazard	Payment Security	Systems that ensure a buyer or seller upholds their end of a transaction, such as escrow services or third-party intermediaries.
	Moderation	Community-based or administrator-led enforcement of rules that discourage deceptive behavior. Can include banning, shaming, or other forms of visible punishment that alter incentives.

C. Gap in our understanding of Telegram cybercrime economy

Although much research has focused on the operational structures of traditional underground forums, particularly in relation to trust signals, the mechanisms by which trust is built on Telegram remain underexplored.

At the same time, the cybercrime economy on Telegram is on the rise [1], [23]. Research indicates the platform has

[1]While participants in underground illicit markets might take steps to protect themselves against law enforcement action or the platform themselves, in this paper we are interested in steps taken by cybercriminals to protect themselves from negative trade outcomes.

become a key venue for illicit market activities. Telegram's illicit markets mirror e-commerce platforms, with vendors engaging in effort-intensive operations such as marketing, secure trading, and reputation management [24]. Roy et al. [1] conducted the first large-scale analysis of cybercriminal channels on Telegram. The authors identified 339 communities engaged in at least one of the five cybercriminal activities: credential compromise, pirated software, blackhat resources, pirated media, and social media manipulation. Their work highlights the strategies used by Telegram community operators to attract users and monetize operations.

However, the trust signals, reputation systems, and market security measures fundamental to online criminal networks have not been examined, nor has the frequency or prevalence to which these can be found. This information is critical to assess the maturity of Telegram cybercrime groups and channels as criminal venues and threat sources. This is particularly important given the highly volatile and dynamic nature of communities on Telegram.

A well-founded view of the ecosystem is impossible without a characterization of the trust mechanisms underpinning it. Critically, combined with the vastness of its ecosystem (and the criminal ecosystem it complements), this makes it especially challenging to gauge how a participant to these communities is *exposed* to these signals, a mechanism that is essential for any signal to be effective: if signals are present, but exceedingly sparse or rare to find [4], they are void and play no role in addressing market inefficiencies. Our research seeks to fill this gap by explicitly investigating the trust-building strategies employed within Telegram groups and channels, evaluating their distribution across market segments, and the extent to which a potential attacker is exposed to them when considering Telegram as a source to acquire specific threat capabilities. To address these gaps, we explore the following research questions:

- **RQ1:** What market segments do we detect in the Telegram cybercriminal ecosystem?
- **RQ2:** What trust signals do cybercriminals use in Telegram communities?
- **RQ3:** How are trust signals distributed across different market segments in illicit Telegram communities?
- **RQ4:** How frequently are users exposed to these trust signals when searching for illicit offerings on Telegram?

III. Methodology

Our methodology develops over three main steps: (i) data collection, (ii) developing a coding framework, and (iii) applying our framework on the data. Figure 1 provides a visual overview of the methodology and highlights which RQs are primarily addressed by each step. Our empirical analysis uses data collected from Telegram channels and groups and serves two primary purposes. First, we use a subsample of this data to develop two coding frameworks, one to identify market segments (RQ1) and the other to analyze trust-building practices (RQ2). Using qualitative coding, we iteratively annotate randomly selected Telegram communities and messages, refining each framework whenever new market segments or trust signals emerge. This process continues until saturation is reached, defined as reviewing 10 additional communities without identifying new market segments, and 100 additional messages without identifying new trust signals. Upon reaching saturation, we compile the final codebooks, which serve as structured annotation guides for each framework. Next, we scale the annotation process using DeepSeek-V3, chosen due to overall performance and accuracy compared to other LLMs tested. [2] We make a DeepSeek-V3 based classifier to automate data labeling using the codebooks finalized during the framework development stage as the prompts. Finally, we use Monte Carlo methods to simulate the exposure to trust signals that a cybercriminal would have across market segments.

A. Scope and assumptions

The purpose of this method is to understand how and how often trust is signaled and encountered in illicit Telegram communities. In this study we aim to explore how trust is operationalized rather than how successful it is in leading to trade. As we currently have no way to verify the success or veracity of these trust signals, in this work we focus on evaluating the prevalence of trust signals across different Telegram criminal communities. In line with findings in the extant literature discussed in subsection II-B, we assume that users evaluate at least partially the communities they find based on visible trust signals, and use these as factors for decision making on which communities to join, especially when the user lacks experience in the ecosystem. Our work therefore aims to capture the perception of users entering the Telegram space may have of different communities: these users are at an early stage in the cybercriminal process, exploring the ecosystem to find trustworthy communities to join to buy and/or sell cybercrime-related products and/or services.

B. Data collection

Our analysis is performed on a manually vetted set of communities exhibiting cybercriminal activity as outlined in §II-A. Communities are included if they show recent activity at the time of the data collection, defined as having at least 10 messages in the last 30 days, and if the majority of messages are advertisements for cybercrime offerings. Illicit offerings include phishing, hacking services, malware distribution, stolen data trading, ransomware services, botnet rentals, digital piracy, credential stuffing, exploit kits, or cryptojacking services, but exclude crimes such as child exploitation, domestic abuse, and drug dealing. We determine the presence of cybercriminal activity by examining the 10 most recent messages from a candidate community. Two authors independently marked each community as cybercrime or not with perfect inter-rater agreement.

We collect messages, replies and basic community information using a custom-built Telegram scraper. We initially

[2]We tested BERT, LLAMA, GPT-4o, and DeepSeek-V3 for our classification task. The last two were successful, and we opted for DeepSeek-V3 due to overall general performance and accuracy.

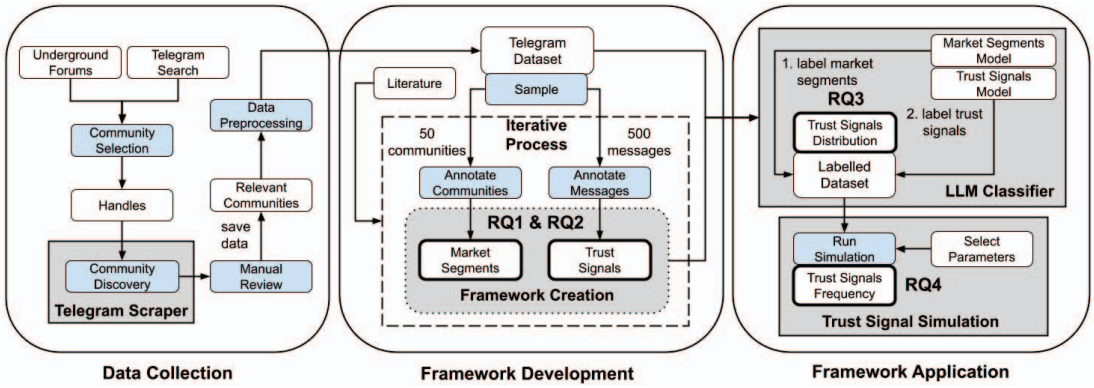


Fig. 1: Pictorial overview of the research methodology

scrape 30 groups and channels mentioned in three prominent cybercrime forums and found on Telegram using keywords via the community search function. This ensures a strong link between the cybercrime (forum) ecosystem and the Telegram channels we start our investigation from. To further expand our set of communities we perform snowballing on groups and channel handles present in messages from that initial dataset. This process includes private Telegram communities which do not have public handles and can only be joined by clicking on a private link circulated by the administrators.

Our final dataset contains 1,116,071 messages from 167 Telegram communities. Data collection is performed between November 2024 and March 2025, but includes messages going back to January 2023. We remove metadata and only retain the essential for understanding conversations. Appendix Table A2 details all scraped data variables.

To facilitate automated analysis, we convert text to lowercase and remove emojis, hyperlinks, non-standard Unicode characters, and newline characters from the messages. We also remove rows with strips of leading/trailing spaces and replace multiple spaces with a single space. We retain currency signs, alphanumeric characters, and punctuation.

C. Development of frameworks

To develop the trust-building and market segment frameworks we develop codebooks specific to each tasks. To maintain consistency, the two tasks are procedurally identical and are each executed independently by two researchers.

To build the market segments framework (RQ1), we sample 50 random cybercriminal communities and 30 messages from each. A market segment is defined as a group of offerings with a common illicit product or service. We use an iterative process: new segments are added when offerings do not fit existing ones, and definitions are refined as needed. Iteration stops when no new segments appear in ten consecutive communities. A communities' primary offering is identified based on the majority of messages and what is being marketed.[3]

To build the trust signals framework (RQ2), we sample 500 messages using stratified sampling to ensure representation from all 167 communities. We develop the framework starting from the trust signals identified in [3] and use this as a lens to identify trust signals on Telegram. Coders independently annotate messages to identify corresponding trust signals. As new patterns emerge, the framework is updated. This process continues until saturation is reached, defined as 100 consecutive messages requiring no further updates, after which the final trust signal codebook is established.

The annotations had high inter-annotator agreement.[4] Researchers met to discuss conflicts in their labeling and update a common codebook for each of the classification tasks. All codebooks can be found in the Appendix.

D. Frameworks application

We scale up the classification using a fine-tuned large language model (LLM). We evaluate the performance of two open-source LLMs (GPT-4o and DeepSeek-V3) on our annotation tasks.[5] We test both models on the classification of market segments and trust-building mechanisms, providing each with the same input data, codebook summaries, and task-specific prompts (see Appendix). DeepSeek-V3 was selected for both tasks, as it performed on par with or better than GPT-4o, while being significantly more cost- and time-efficient. DeepSeek-V3 achieved an F1-score of 0.871 (GPT-4o: 0.663) and Cohen's Kappa of 0.843 (GPT-4o: 0.685) for market segmentation, and an F1-score of 0.940 (GPT-4o: 0.974) and Cohen's Kappa of 0.760 (GPT-4o: 0.871) for trust signal classification. A full breakdown of results is provided in the Appendix.

E. Trust exposure simulation

To evaluate the (relative) frequency with which users are exposed to trust signals across specific market segments (RQ4), we use a Monte Carlo simulation.

We simulate 10,000 unique users, each assigned to a randomly chosen market segment. We then vary their browsing

[3]Offerings within communities are typically stable and thematic. Over 90% of communities had more than 75% of sampled messages referring to a single category of product or service.

[4]Cohen's Kappa was 0.90 for the Telegram channel labeling task (6 categories, 50 items, 4 disagreements), and 0.98 for the message-level labeling task (11 categories, 500 items, 9 disagreements).

[5]Prior research suggests that open-source LLMs can match or exceed human annotation accuracy [25], with GPT-4o and DeepSeek-V3 outperforming models like Gemini and LLaMA [26].

model based on Campobasso and Allodi [27], who suggest that users looking for products to buy on criminal platform only consider the most recent ones, and evaluate product characteristics before making a decision. Each simulated user is first assigned a random market segment and a visiting date within the data period (Jan 2023-Mar 2025). Each user then visits a randomized number of Telegram communities (three to five) relevant to their assigned market segment. For each community, the simulated user samples a randomized number (between 10 and 50) of the most recent messages relative to their assigned visiting date, emulating what they would realistically see upon visiting the community at a given moment in time.[6] The simulation ensures that the content viewed was available within the assigned time period (January 2023 to March 2025), preserving the temporal integrity of the generated scenarios.

For each community selected by a simulated user, we count the total occurrences of trust signals the user is exposed to. Trust signals that are always visible to the user, such as those contained within pinned messages and community description, are counted only once per visit to avoid double-counting. Following the simulation, we compute descriptive statistics, including the mean, median, standard deviation, and confidence interval of trust signals encountered both globally and per market segment.

F. Ethical considerations

The research is carried out as part of a program that received ethical review and approval from the ethics committee of the Department of Mathematics and Computer Science of Eindhoven University of Technology, under ERB approval no. ERB2021MCS1. The datasets used have been collected from publicly available channels, or channels advertised on public forums. As such, we expect users to be aware that their posts are publicly visible. Due to the scale and nature of the channels informed consent cannot be gained from all members of the forum. However, under the British Society of Criminology's Ethics Statement [28], informed consent is not required for research into online communities where the data is publicly available, and the research outputs focus on collective rather than individual behavior. Our analysis is performed collectively and as such no individual can be de-anonymized. The examples of posts provided in the paper have been paraphrased to remove attribution to their authors.

IV. Trust signals on Telegram

Our dataset includes 167 groups. 24 groups, containing approximately 26,000 messages, are private and require an invitation or special link for access, and 143 groups are publicly discoverable and account for 1.09 million messages. At the time of data collection, we also captured 65 pinned messages and 133 group descriptions. On average, each group contains 6,600 messages, but the median is much lower at 200, reflecting a highly skewed distribution. While most groups have fewer than 1,000 messages, a small number contain tens or even hundreds of thousands.

Similarly, participant counts vary widely, with a total of 24.4 million members across all groups; the median number of participants per group is 10.1k, while the mean is significantly higher at 150k. This indicates that there are a few exceptionally large and influential groups. There is a positive correlation between the number of messages and participants in a group (Spearman's correlation: $r_s = 0.423, p < 0.001, N = 160$).[7] Figure 2 provides an overview of community distribution over message and participant counts.

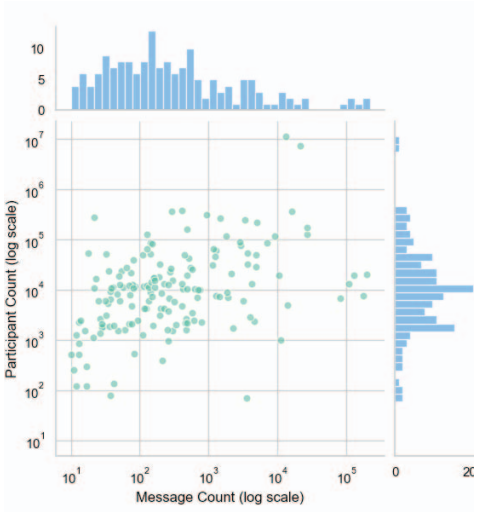


Fig. 2: Joint distribution of message and participant counts

A. RQ1: Market segments on Telegram

Table II reports the six market segments identified in this study alongside a description of their activities and a frequency count. Most communities we find advertise tools or services to commit fraud (55). Communities selling personal data (36) and cyberattacks (35) are also prominent in our data collection. Infrastructure (11) and tutorials (7) are the least prominent.

TABLE II: Market segment framework

Market Segment	Comms.	Messages	Description
Cyberattacks	35	220,363	Malicious tools enabling unauthorized access, disruption, or system damage
Infrastructure	11	209,803	Services and resources enabling or anonymising cybercriminal activity
Digital piracy	23	64,295	Distribution or support for illicit digital content
Fraud tools	55	474,231	Schemes involving financial deception and payment system abuse
Personal data	36	114,574	Trade in stolen or leaked personal information
Tutorials	7	32,805	Educational content aimed at teaching individuals how to perform cybercriminal activities
Total	**167**	**1,116,071**	

[6]We note that the parametrization is only meant to keep the model within realistic boundaries of user actions. We have run the simulation with different options in this parameter space, and results remain qualitatively stable.

[7]For 7 of the 167 communities we scraped, the participant count was unavailable.

B. RQ2: Trust signals on Telegram

Table III provides an overview of the trust signals identified

TABLE III: Trust signal framework

Key Issue	Trust Signal	Description
Reputation	1. Vouching system	Users vouch for each other to build credibility.
	2. Scam reporting	Mechanism to report and flag fraudulent activities.
Payment Security	3. Escrow service	Holds funds until both parties fulfill their obligations.
	4. Digital wallets	Secure online payment methods.
	5. Cryptocurrencies	Use of blockchain-based currencies for secure transactions.
	6. Automation	Automated bots to facilitate secure payments.
Moderation	7. Clear rules	Well-defined rules to ensure fair trading practices.
Transparency	8. Proof of delivery	Verification of goods/services delivery.
	9. Customer support	Assistance provided to resolve issues and queries.
	10. Free Samples/Trials	Offering samples or trials to build trust.
	11. Warranty	Guarantee of product/service quality.

in the studied communities. We provide quotes of messages talking about each trust signal below.

a) Reputation:

1. Vouching System is a method in which members of a group or channel endorse or vouch for other users, supporting their authenticity and trustworthiness. In the absence of a similar reputation system as that found in traditional forums [3], more immediate vouching messages are used to support user credibility on Telegram. Most vouching focuses on positive experience in a trade.

> *"Vouch for @username, quality products and an overall high quality person to work with."*

This system may provide an important trust signal in the creation of a trusted network of users at the community and platform level. Vouching is generally implemented with messages either posted to the group, or forwarded to the channel. Some users may also get verified by trusted members of the community. To further substantiate their support, some users choose to share the economic earnings they achieved.

> *"I'm vouching for this channel owner, he made me earn 3k in just 5 mins after I met him."*
>
> *"This dude claiming to hack iclouds with imei. Can someone here vouch for this guy?"*

2. Scam Reporting consists of warnings posted by users. Since users on Telegram do not have an elaborate scheme to report scams such as those on forums, they use more immediate methods. In Telegram groups, communities for multi-way communication where all users can interact sending messages, scams can be reported by anyone, and often describe their own personal negative experience.

> *"He is a scammer he sells the account and then changes its password, he expects money from your first account."*

It is a common practice of users in groups to provide context through their scam report, signaling other users to trust instead. This displays some sense of community within Telegram groups.

> *"@user1 is a well-known scammer in the hacking community, and he is back! Buy from @user2."*

In Telegram channels, communities for unilateral communication, only administrators can report scams to the larger audience. Accusations are sometimes accompanied by proof of the scam, in the form of screenshots or forwarded messages. When a user is exposed for scamming, they are usually removed from the community. Scam reports may be employed to maintain the integrity of the community, possibly establishing the perceived legitimacy of users and transactions.

b) Payment Security:

3. Escrow Service, sometimes also known as a middleman (MM), is a payment method in which a third party temporarily holds the money during a transaction until the buyer confirms receipt of the products or services. As seen in traditional forums [3], escrow services are used to ensure that both the seller and the buyer keep their word in the deal, reducing the risk of being scammed and enhancing trust. Users on Telegram value the use of escrow and ask for it as a payment option when requesting offerings. Cybercriminals mention escrow services to build trust and attract customers.

> *"Rats, grabbers, and all the malware you need. Escrow available with any trusted admin."*
>
> *"I am looking for someone with good stealer, dm me and let me know your rates and services. Escrow please or verified sellers only!"*

4. Automation is achieved through Telegram bots, often used to secure and automate the payment process. Automation allows potential buyers to interact directly with a Telegram bot to view available products, check their prices, or otherwise explore the catalog, and make payments.

> *"Netflix accounts reloaded in our bot, get Netflix premium account pay bot."*

This mechanism increases the predictability and velocity of the payment process, allowing users to receive immediate feedback on the outcomes of the transaction and access to the desired service.

> *"With our easy-to-use bot, you can make your purchases quickly and securely using your ton wallet. Purchases will be delivered in only 5 minutes!"*

5. Digital Wallets are centralized financial platforms that allow users to store and transfer money electronically. In the context of illicit Telegram ecosystems, services like PayPal, CashApp, and Venmo are sometimes used to facilitate quick and familiar transactions, while providing an additional layer of (perceived) privacy, compared to traditional banking. Their integration with conventional payment systems suggests an effort to reduce friction for buyers, particularly those less familiar with decentralized payment methods [29].

> *"All payments done in PayPal, dm your order."*

6. Cryptocurrencies are decentralized digital assets that enable peer-to-peer transactions without intermediaries. Their use in illicit Telegram markets may signal trust by implying a level of operational security, or familiarity with cybercriminal norms [30].

"Start spamming 100% guaranteed results, all crypto payment accepted "

The reliance on cryptocurrencies can suggest alignment with practices that prioritize privacy, which may be reassuring to certain buyer communities [31].

c) Moderation:

7. Clear Rules are guidelines set by the administrator or sellers and are used to regulate user behaviors and interactions. In the context of illicit activities, these rules are most often seen as pinned messages and cover various aspects, such as allowed behavior or the permitted types of transaction.

"Rules: No spamming, no link posting, no ads, no porn, no adverts, join us today to share more ideals on building spamming tools."

Some can be descriptive and specific, while others are open to interpretation. Warnings are also given to enforce these rules.

"You have one warning left, then you will be banned permanently - this is your last warning and you have been warned before - you must read the first rule of group and adhere to that."

d) Transparency:

8. Proof of Delivery is provided in some marketplaces by administrators or vendors by way of evidence that a product or service has been successfully delivered before payment is released or to support dispute resolution. Proof of delivery can take various forms, including screenshots of completed transactions, timestamps, system logs, or offering to show proof.

"Live calls running, proof available for every call."

Its purpose is to build trust between buyers and sellers, prevent fraud, and demonstrate that the vendor has met their obligation.

9. Customer Support is offered to provide assistance or solve problems customers may be facing. It is often referred to in the advertisement of the product or service as an added value to the offering.

"Get all phishing tools for good price, extra fast delivery and customer support available."

10. Free Samples/Trials are limited versions or small quantities of products or services provided at no cost. This allows potential buyers to test the quality and functionality before committing to a purchase, and vendors to attract buyers by showcasing the veracity and quality of their offerings.

"A collection of new databases free samples available. I want serious people only."

Free samples or trials on Telegram are used to build trust and demonstrate the effectiveness of the product, encouraging users to make a full purchase.

11. Warranties are guarantees provided by the vendor that promise that the product or service will perform as advertised or that any issues will be rectified.

"Offering high quality gmail domains, umlimited stock, warranty 1 day."

Warranties on Telegram can include promises of lifetime, refunds, replacements, or technical support if the product fails to meet expectations. This mechanism helps build trust by ensuring that buyers will be reimbursed if problems arise.

C. RQ3: Distribution of trust signals

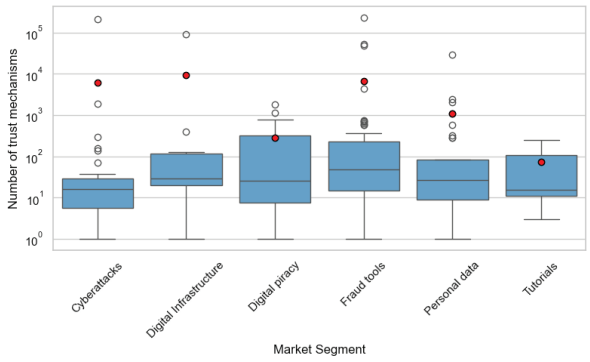


Fig. 3: Boxplot of trust signals within market segments

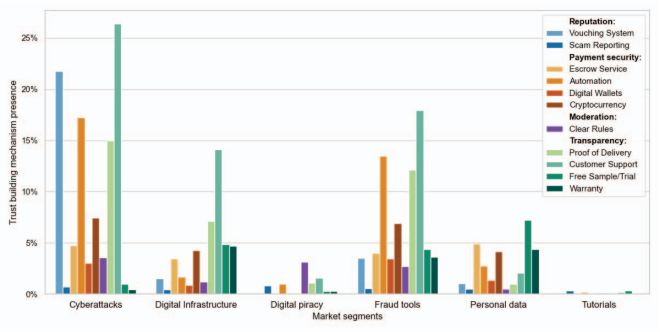
Figure 3 provides a breakdown of trust signals across market segments. The boxplot shows the aggregated number of trust signals mentioned in each community we study. The median numbers of trust signals are relatively uniform and move between 10 and 100 per community. We also plot the mean of each segment in red, which shows that on average digital piracy and tutorials lag severely behind the other market segments. The disparity between median and average number of trust signals is largely driven by the imbalanced number of messages across the observed communities.

Figure 4 provides an overview of the distribution of individual trust signals (top) as well as the broader trust signal groups (bottom) across market segments. The bar graphs are normalized by the number of messages within each segment. We show the distribution of trust signals across market segments separately for public and private communities.

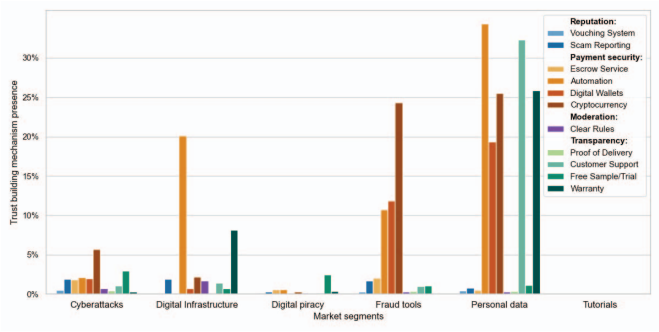
Across the board, sellers most often provide customer support and proof of delivery, while relying on bots for predictable and automated exchange of goods and services. Additionally, cryptocurrency is a universal trust tool across markets, enabling secure, pseudonymous payments. On the other hand, we find little evidence of widespread scam reporting.

More broadly, two groups of trust building mechanisms stand out: payment security and transparency. Especially in private groups, the two groups represent the vast majority of trust signals observed. Among the public groups, the diversity of trust groups is a bit higher, where reputation and moderation also appear, albeit in much smaller numbers compared to the leading groups.

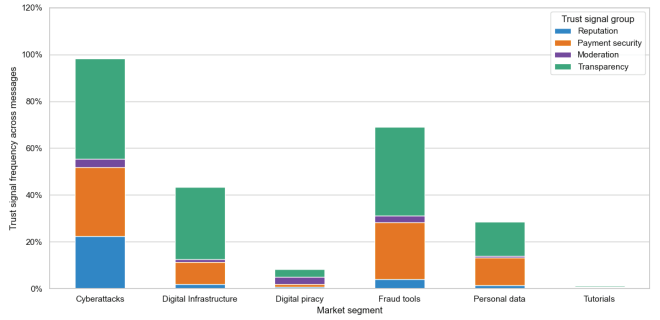
When we zoom in on individual market segments, several patterns emerge. The following subsection provides a detailed analysis of trust mechanism distributions both across and within market segments.



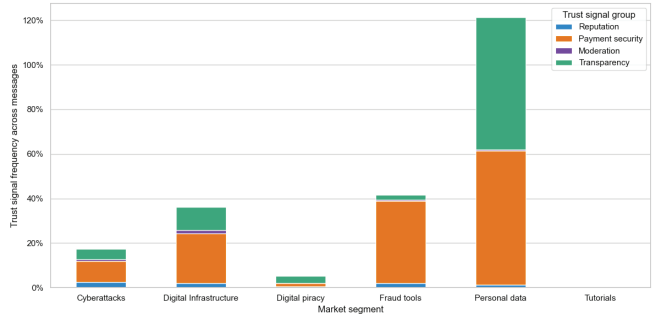
(a) Public communities only



(b) Private communities only



(c) Public communities only



(d) Private communities only

A single message can contain multiple trust signals. Note the different y-axes. Public communities are on the left, private communities are on the right. Individual trust signals are on the top, trust signal groups are on the bottom.

Fig. 4: Distribution of trust signals across market segments

1) Trust mechanisms across market segments: In public communities, there are notable differences across market segments in how often trust mechanisms are present. We find that cyberattacks and fraud tools market segments have the highest presence of trust mechanisms, closely followed by digital infrastructure and personal data market segments. In private communities, personal data and fraud tools lead in number of trust signals.

On the other hand, messages in digital piracy and tutorials market segments almost entirely lack any trust mechanisms both in public and private communities. One explanation for this pattern could be the nature of the market segment and the associated monetary risk. The segments with a higher proportion of trust mechanisms are also the segments where products and services are exchanged for money. This contrasts with the two segments with the lowest proportion of trust mechanisms, where the majority of goods are exchanged for free. To test this, we manually classify a small subset of randomly selected messages from both public and private communities ($N = 300$) from each market segment. We aim to identify the prevalence of paid-for offerings compared to freely available ones. We consider a message to contain a paid-for offering when it mentions a price (even if the price is not explicitly stated, e.g. 'low price') or advertises a professional service or a good.

We find suggestive evidence that messages selling illicit products contain trust signals much more frequently than messages offering goods for free. Approximately half of the manually classified messages offer some paid-for service/good. Among those, we find on average 0.70 trust signals per message, whereas we only find on average 0.09 trust signals in the remaining half of the messages.

As tutorials and piracy market segments are among the market segments with the lowest proportion of messages selling something (4% and 2% of messages, respectively), this might explain the low presence of trust signals. On the contrary, fraud (88%) and infrastructure (82%) market segments contain the majority of messages advertising a paid-for service/good, followed by personal data (68%) and cyberattacks (52%).

2) Trust mechanisms within market segments: The cyberattacks segment has the highest concentration of trust-building mechanisms overall. Customer support (25%) is especially dominant, followed by vouching systems (21%) and Telegram bots (17%). This suggests cyberattack vendors heavily rely on professionalization and customer reassurance. The fraud tools segment shows a strong presence of Telegram bots (13%), customer support (17%), and proof of delivery (13%). Cryptocurrency and escrow services are also moderately represented. Here, trust seems to be built on efficient automation and transactional safeguards.

In the digital infrastructure segment customer support (14%) and proof of delivery (8%) stand out, followed by cryptocurrency and Telegram bots. In the personal data segment sellers attempt to gain trust by providing free samples, a mechanism

less prominent in other market segments. Finally, we see minimal use of trust mechanisms in the digital piracy market segment, and virtually no trust mechanisms in tutorials. As previously discussed, this could indicate a less commercialized or lower-risk market, requiring fewer trust reinforcements. Additionally, tutorials might be freely shared, low-stakes, or community-driven, where trust is less critical.

Among private communities, the personal data segment has the highest concentration of trust signals. Among them, automation, customer support, warranty and cryptocurrencies are the most prevalent. Payment security trust signals, specifically cryptocurrency, digital wallets and automation are also popular in the fraud tools segment. Additionally, Automation is by far the most prevalent trust signal in the digital infrastructure segment. Compared to public communities, we see much less transparency trust signals and more focus on payment security.

3) Trust signals beyond messages: While assuming that participants read all messages may be far fetched, it is reasonable to assume that they read group descriptions and pinned messages. Of the 167 communities, 133 had group descriptions and 65 had a pinned message at the time of scraping. We find that descriptions typically do not contain trust signals and instead provide information about the channel's topics (e.g.*"'We crack custom software"* or *"Crypto drainer service"*), information about the channel owner and links to any affiliated channels (e.g. *"DM @username"*, *"Owner: @username"*, *"Backup channel: @channel"*) or disclaimers (e.g. *"For educational purposes only"*). Observationally, trust signals are more present in pinned messages that often contain updates, rules and instructions on how to conduct business. However, we are unable to quantitatively confirm our observation due to small sample size of pinned messages.

D. RQ4: Trust exposure simulation results

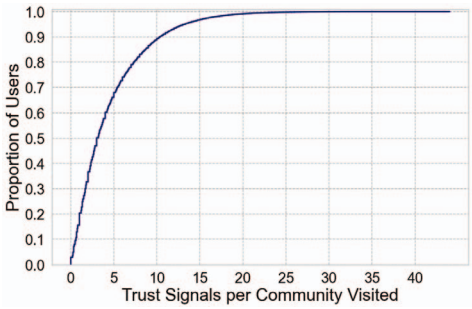


Fig. 5: Distribution of trust signals seen per community

1) Overall user exposure to trust signals: Figure 5 shows the proportion of simulated users who encounter a given number of trust signals across Telegram communities. The steep initial rise indicates that most users encounter at most only a few trust signals; around 70% of users encounter 5 or fewer (out of, on average, 120 seen messages). The curve flattens as the number of trust signals increases, with nearly all users seeing at least one and over 90% encountering fewer than 10. This highlights the uneven amount of trust signals that a user experiences in each community they visit. The average number of trust signals seen per community visited overall is 5.22 ($SD = 4.67$).

2) User exposure to trust signals by market segment: Figure 6 provides a description by market segment. We find

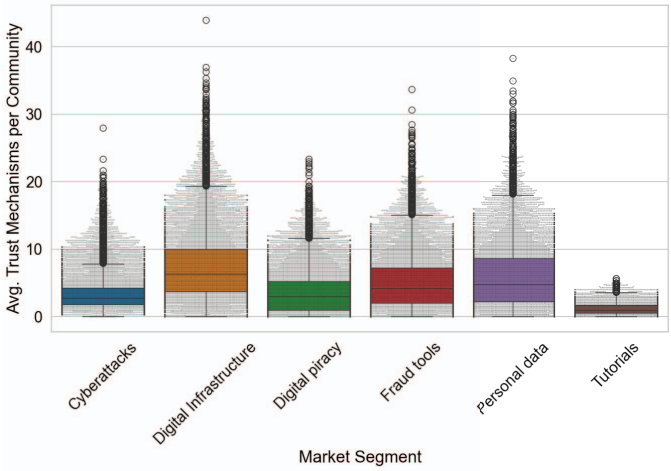


Fig. 6: Simulated user trust signals per segment

that simulated users browsing digital infrastructure and personal data markets experience the highest average trust signal exposure, at 7.43 ($median = 6.33$) and 5.99 ($median = 4.8$) respectively. In contrast, users seeking content in tutorials and digital piracy communities encounter far fewer trust signals, 1.16 ($median = 1$) and 3.68 ($median = 3$) on average, suggesting these segments may rely less on trust signaling, potentially due to lower perceived risk or different transactional norms.

A Wilcoxon rank-sum test confirms that the difference in trust exposure between digital infrastructure and tutorials is statistically significant ($p < 0.001$), as is the difference between fraud tools and personal data ($p < 0.001$), despite their similar appearance in Figure 2. Similarly, although cyberattack communities appear active in aggregate bar plots, users navigating these spaces encounter significantly fewer trust signals than those exploring digital infrastructure communities. The average exposure for cyberattacks is 3.56 ($median = 2.75$) trust signals per community, compared to 7.43 ($median = 6.33$) in digital infrastructure. A Wilcoxon rank-sum test confirms this difference is statistically significant ($p < 0.001$), reinforcing the idea that users in infrastructure markets experience a markedly richer trust environment.

These findings highlight that trust exposure varies across Telegram's cybercrime ecosystem. Professionalized markets tend to display more visible trust signals, while others leave users navigating greater uncertainty. This heterogeneity underscores the need to analyze trust as it is experienced along user pathways, shaped by both market segment and entry point. Shifts in trust signaling may indicate increased cyber-operations, changes in governance, or shifts in clientele that alter a community's risk profile.

3) User exposure to trust signals per type and market segment: To assess the user experience of trust signaling type across market segments, we analyze the mean per-message exposure to 11 trust signals, disaggregated by market segment. Figure 7 reveals differences not apparent in aggregate

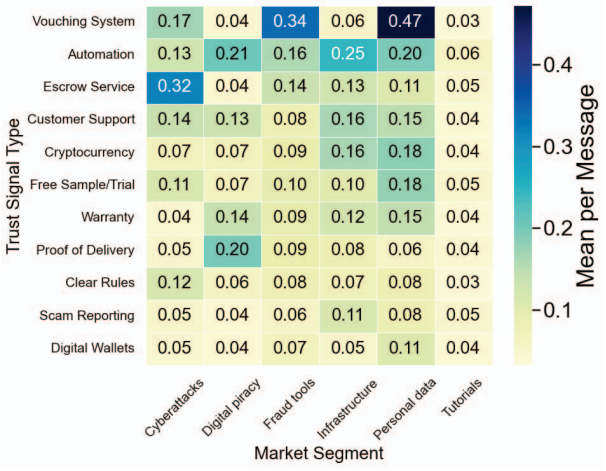


Fig. 7: Exposure to trust signals by market segment

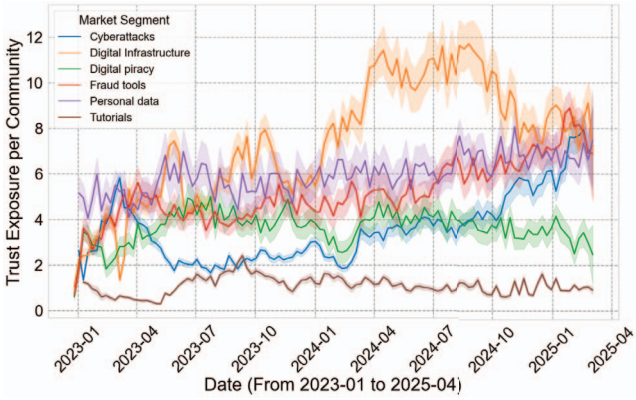
message-level plots. For instance, the average exposure to vouching Systems in cyberattacks was 0.17 (*median=0.16*), compared to 0.47 (*median=0.18*) in personal data. This occured despite messages with vouching systems being more prominent in cyberattacks communities than in personal data communities in Figure 2. Similar patterns were seen for customer support (0.14 compared to 0.15) and Telegram bots (0.13 compared to 0.2). These differences suggest that some trust signals in cyberattacks communities may be concentrated within a small subset of messages, rather than broadly encountered by users. Cyberattacks was the market segment that had the highest count of messages with vouching. Upon further exploration of why it had a lower exposure in the simulation, we found that more than 99% of vouching signals in cyberattacks communities were concentrated in only one community, while the other market segments had Vouching signals better spread out through their communities. This could explain why the users browsing multiple communities of cyberattacks found vouching less often than expected.

The comparison between fraud tools and personal data markets further complicates assumptions based on aggregate trends. While earlier bar plots suggested that fraud tools communities were more trust-rich overall, simulated user-level exposure revealed higher typical exposure in personal data. For example, exposure to cryptocurrency signals was 0.09 (*median=0.06*) in fraud tools and 0.19 (*median=0.16*) in personal data; for warranty, 0.09 (*median=0.06*) compared to 0.15 (*median=0.15*); and for automation, 0.16 (*median=0.11*) compared to 0.2 (*median=0.17*). These differences point to personal data communities offering denser and more consistent signaling to users, despite appearing less prominent in aggregate views.

Some trust signals also showed high exposure but only in select markets. Escrow services had an exposure rate of 0.32 (*median=0.17*) in cyberattacks, while vouching systems reached 0.47 (*median=0.12*) in personal data and 0.34 (*median=0.21*) in fraud tools. In contrast, these same signals were nearly absent elsewhere, and display a lower median, which indicates quite a lot of variation on these signals per user experience. Automation was particularly common in digital infrastructure with 0.25 (*median=0.2*), and was also the signal most frequently encountered in both digital piracy with 0.2, *median=0.2*) and tutorials with 0.06, (*median=0.05*), suggesting a potential segment-specific signaling norm. Rather than being evenly distributed, the exposure to trust signals varies significantly by market and signal type. These results underscore the importance of evaluating trust signaling from the perspective of individual users, rather than relying solely on total signal volume or aggregate message counts.

4) Evolution of users exposure to trust signals over time: To capture how the trust signaling ecosystem evolves from the user's perspective during the lifetime of the communities, we provide a detailed view of trends in encountered trust signals per period of time.

As shown in Figure 8, trust signals vary significantly across market segments, shaping the user experience in distinct ways. Users engaging with digital infrastructure communities encounter the highest concentration of trust signals. Between April and October of 2024, these communities reach a peak in terms of user exposure to trust signals, averaging around 10 per (simulated) user visit. Those interested in personal data or fraud tools encounter a different but still evolving landscape. Between 2024 and early 2025, users in these segments would have noticed a steady rise in trust mechanisms, reaching six to seven signals per community by the end of the period, suggesting growing norms around building trust.



Each timeseries is augmented with 95% Confidence Interval

Fig. 8: Weekly average of trust exposure per community

Users focused on cyberattacks content experience a more gradual shift. Trust signals increase from about two to nearly six, reflecting a slow but noticeable change in these communities. Meanwhile, those engaging with digital piracy and tutorials see little change. These environments remain relatively

static, with only three to four trust signals on average for digital piracy and one for Tutorials.

Figure 9 reveals how exposure to individual trust signals varies across market segments and over time. Vouching systems peak in fraud tools markets between mid-2023 and early 2024, with minimal activity outside this window. After 2024, digital infrastructure communities see a sharp rise in automation and customer support signals—despite earlier users encountering far fewer signals, even in communities that ultimately show the highest aggregate trust exposure. This highlights a key pattern: Telegram communities evolve dynamically in how they signal trust. Aggregate message views obscure temporal fluctuations. The seasonal appearance of signals, such as Proof of Delivery and Warranty in Digital Piracy, shows that trust cues emerge in waves, leading to user experiences that vary significantly by time of entry. This temporal variability calls into question the maturity and standardization of trade practices. Trust is not underpinned by a stable signaling infrastructure but is deployed opportunistically, shaped by product type, vendor competition, and buyer feedback. Signals are not constant; their use reflects shifting needs to establish credibility.

Post-2024 users in digital infrastructure communities encounter the most trust-rich environments, with over 10 signals per community. Earlier users see far fewer. While fraud tools and personal data markets appear similar in aggregate plots, (simulated) users report fraud tools as slightly denser and more volatile, marked by bursts of vouching and automation. Ultimately, trust signal exposure is uneven across the cybercrime economy, shaped by both market segment and user timing. This creates unpredictable trust landscapes where some users face robust signaling while others navigate with minimal cues—unlike the more stable, centrally controlled environments of cybercriminal forums [3], [32].

V. Discussion

A. Implications for Research

Telegram is rapidly emerging as a platform for trading and sharing illicit content, expanding the cybercrime ecosystem beyond traditional forums. However, its low entry barriers and fragmented structure challenge the formation of stable, rule-governed markets. This instability undermines trust, an essential component in sustaining illicit trade, unlike what has been documented in cybercrime forums, where trust systems are more robust and structured.

Our findings suggest avenues for future research, particularly work examining the dynamics of trust signaling in fluid, decentralized platforms like Telegram. While channels do attempt to signal trust, these signals vary widely across and within market segments. Some, like fraud tools and digital infrastructure, display a large volume of trust signals, but when normalized by message count, their signal density is low. Conversely, cyberattacks show the highest density, indicating that a larger share of messages actively promote trust. This variation across segments aligns with prior work on cybercriminal markets [3], and calls for a more granular understanding of how trust emerges in different illicit economies. Notably, some segments, such as cyberattacks and fraud tools, appear more mature, frequently using signals like proof of delivery and customer support. Reputation via vouching is prominent only in cyberattacks, while negative feedback is rare across the board, limiting the capacity of these markets to correct information asymmetries.

An insight for research is that trust signaling on Telegram is not static. Many channels, particularly in cyberattacks and digital infrastructure, exhibit seasonality and evolving signaling patterns. This temporal variability raises questions about whether such fluctuations are product-specific or driven by broader market conditions. It highlights the limitations of relying on snapshot analyses and points to the need for longitudinal research designs to avoid mischaracterizing the trustworthiness or relevance of these communities.

Moreover, our simulation of 10,000 users searching for communities to join in specific market segments reveals another layer of complexity. Although cyberattacks show the highest overall vouching, simulated users encountered them more often in the fraud tools segment. This suggests that user experiences are highly variable and may not align with aggregate statistics. Future research should explore how users perceive trustworthiness and make decisions in such dynamic and uncertain environments. It should examine the veracity and effectiveness of specific trust signals used on Telegram, investigating whether these can be linked to specific actors or groups, thus enhancing attribution of high-value actors in the ecosystem. Such work could also assess the economic or operational cost these signals impose on cybercriminals, making a distinction between signals that are expensive to genuinely earn versus those that are costly merely to claim or convincingly fake. Understanding these dynamics can help distinguish between superficial markers of trust and those that impose real friction or risk, thus informing more robust strategies for platform intervention and user study.

Finally, our opportunistic identification of private channels, present in all segments except tutorials, opens further research avenues. These private, invite-only groups emphasize payment security over reputation or transparency, suggesting the presence of preexisting social capital, possibly imported from forums. These warrant a different line of investigation, especially around the interplay between closed networks and open platforms.

B. Implications for Practice

From a practical perspective, particularly for cybersecurity monitoring, threat intelligence, and law enforcement, our findings shed light on how trust operates differently on Telegram than on forums, and what that means for intervention. The key insight from our work is that prioritization and evaluation of the information captured in Telegram channels and groups is fundamental to obtain effective and actionable insights: not all criminal groups and channels are the same. We find that key structural limitations of Telegram restrict the development of trust mechanisms found in forums. Campobasso et al. [3]

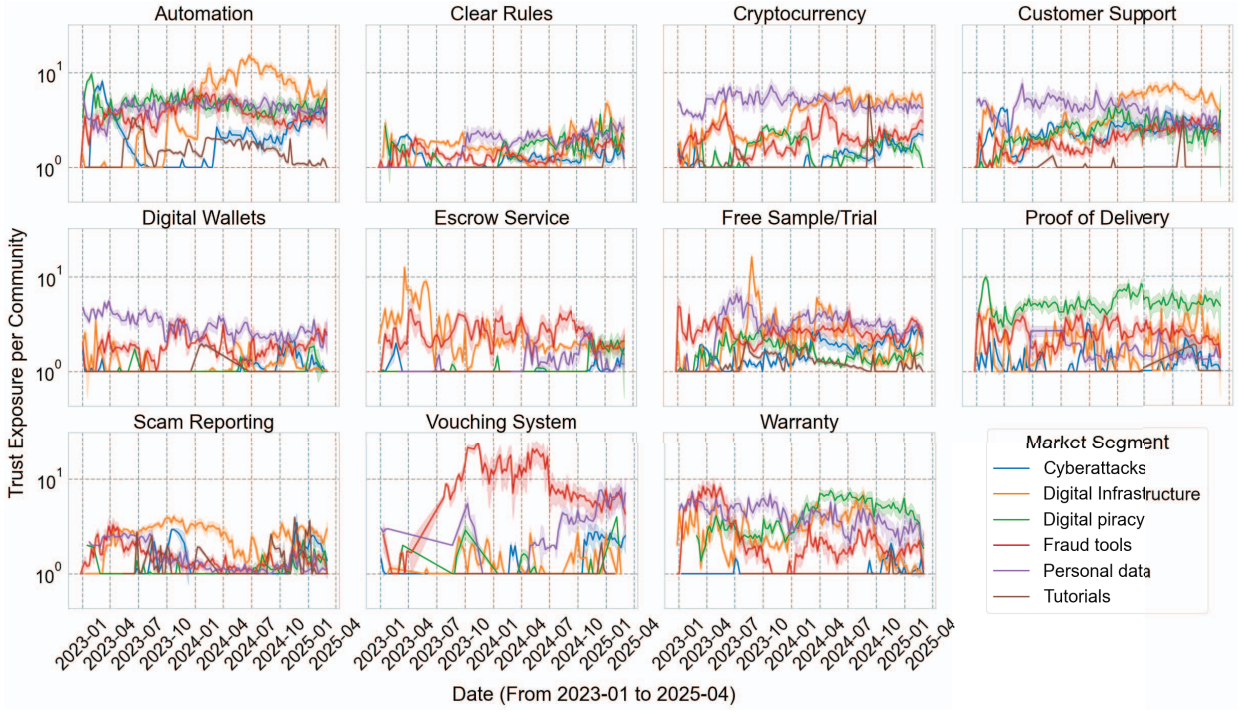


Fig. 9: Weekly Trust Exposure per Community by Trust Signal

identified 28 trust mechanisms in forums, while we observed only 11 on Telegram. Mechanisms like third-party seller verification and controlled access are harder to implement in a broadcast-based platform where typical forum governance structures are absent. This limits Telegram's ability to replicate structured, rule-governed marketplaces, where market-level trust mechanisms ensure effective functioning. Instead, on Telegram, the community itself is more involved in self-policing, using individual-level trust mechanisms.

Understanding these constraints is essential for designing more effective detection and disruption strategies. For example, the high density of trust signals in cyberattack-related channels can offer leverage points for law enforcement, as these communities rely more visibly on public reputation building. More trust may also indicate a higher maturity of the community, where products, services, and vendors are of greater value. In such cases, trust signals can become useful proxies for identifying which communities warrant closer monitoring. Conversely, segments with a lower density of trust signaling may require different approaches, as trade dynamics in these segments may be fundamentally different from elsewhere.

Furthermore, the seasonal and volatile nature of trust signaling means that timing matters. A community may appear trustworthy at one point in time and become dormant or fraudulent shortly after. This has major implications for how cybercrime intelligence is gathered and interpreted: real-time and repeated monitoring becomes essential.

Our comparison with other studies, such as Roy et al. [1], illustrates the importance of source selection in ecosystem mapping. Their use of Telemtr.io, a general-purpose Telegram catalog, leads to different segmentation than our method based on seeding from criminal forums. This variation underscores the need for replicability and transparent methodology when characterizing such platforms.

Finally, by applying a game-theoretic lens to Telegram markets, we can begin to quantify the interplay between actors and communities. Trust signaling can be seen as a strategic move within a constrained signaling environment. Some actors invest more in reputation, while others exploit low-cost entry to engage in quick and opportunistic behavior. These dynamics could be formalized and modeled, aiding both academic understanding and applied efforts to predict or disrupt cybercriminal operations.

C. Limitations

This study has several limitations. First, it is difficult to determine whether a given feature genuinely functions as a trust mechanism or how effective it is in practice. Although we document elements that may signal trust, we cannot confirm their role or impact. Our analysis is therefore limited to identifying potential trust indicators, as informed by signaling theory and related work, without assessing their actual effectiveness within the ecosystem. Second, the representativeness of our dataset is a concern. Although we selected communities from multiple sources, including different forums, Telegram sections, and keyword-based searches, the sample may not capture the full range of behaviors, actors, or contexts. As with any observational study, sampling bias is possible, and our findings should be interpreted as reflective of Telegram communities linked to criminal forums, not the entire ecosystem. Finally, our approach may miss the trust-building mechanisms that

occur outside of the observed message content. Vouching, reviews, or scam reports can happen in private chats, dedicated channels, or external platforms. Trust may also carry over from prior interactions or from other communities beyond our view. Future research could address this by comparing Telegram communities discovered via dark web forums with those found directly on the platform.

D. Key Takeaways and Future Directions

From this work, we derive three key takeaways: (1) **trust signaling on Telegram is fragmented and highly inconsistent across communities**, especially when contrasted with the structured governance and stable identity systems observed in forums; (2) **trust on Telegram relies heavily on payment and transparency-related assurances**, such as automation with bots, cryptocurrencies, proof-of-delivery, and offering customer support, rather than on reputation or moderation; and (3) **user exposure to trust signals is shaped more by temporal dynamics than by aggregate signal availability**, highlighting the fluid and unstable character of Telegram markets. These findings suggest several avenues for future work, including examining trust signal success, whether specific mechanisms on Telegram measurably reduce scams, assessing which signals most effectively sustain repeat transactions, and studying how cross-platform actors port reputation from underground forums into Telegram ecosystems. Ultimately, operationalizing trust signal monitoring as an indicator of market professionalization and emerging threats is the goal.

VI. Conclusion

In response to **RQ1**, we identified six distinct market segments, each with varying levels of maturity and trust-related behaviors. Addressing **RQ2**, we found that Telegram actors employ a variety of trust signals, including transparency mechanisms, automation, customer support, and more rarely, voucher systems. For **RQ3**, our analysis shows that the distribution of these signals is highly uneven between segments and even more so across individual channels, with some segments (e.g. cyberattacks, digital infrastructure) exhibiting more dense and consistent trust signaling than others (e.g., tutorials, piracy). Regarding **RQ4**, our simulation of user experience reveals that users are generally exposed to only a small subset of these trust signals, and this exposure varies significantly depending on the market segment and channel structure. Additionally, individual-level mechanisms play a more important role compared to forums with more market-level mechanisms. In general, our findings display the fragmented and dynamic nature of cybercriminal trade on Telegram, highlighting the limitations of static or one-size-fits-all monitoring approaches and the need for more adaptive and context-sensitive strategies in both research, cyber threat intelligence, and law enforcement.

Acknowledgements

Part of this study is funded by the INTERSECT project, Grant No. NWA.1162.18.301, funded by NWO and by the CATRIN project, Grant No. NWA.1215.18.003. This work is partly supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, grant No 949127, and by an unrestricted gift from Meta. Tina Marjanov is supported by King's College, Cambridge and the Cambridge Trust. We thank Jack Hughes and Konstantinos Ioannidis of the University of Cambridge for their support during this project, as well as the Cambridge Cybercrime Centre, for providing the data used in this study.

References

[1] S. S. Roy, E. P. Vafa, K. Khanmohamaddi, and S. Nilizadeh, "DarkGram: A large-scale analysis of cybercriminal activity channels on Telegram," in *Proceedings of the USENIX Security Symposium*. USENIX Association, 2025, pp. 4839–4858. [Online]. Available: https://www.usenix.org/conference/usenixsecurity25/presentation/roy

[2] C. Herley and D. Florêncio, "Nobody sells gold for the price of silver: Dishonesty, uncertainty and the underground economy," in *Economics of Information Security and Privacy*, T. Moore, D. Pym, and C. Ioannidis, Eds. Springer US, 2010, pp. 33–53. [Online]. Available: https://doi.org/10.1007/978-1-4419-6967-5_3

[3] M. Campobasso, R. Rădulescu, S. Brons, and L. Allodi, "You can tell a cybercriminal by the company they keep: A framework to infer the relevance of underground communities to the threat landscape," in *The 22nd Workshop on the Economics of Information Security (WEIS'23)*. Geneva, Switzerland: WEIS, 2023, pp. 1–17. [Online]. Available: https://arxiv.org/pdf/2306.05898

[4] C. von Deimling, M. Eßig, and A. H. Glas, "Signalling theory," in *Handbook of Theories for Purchasing, Supply Chain and Management Research*. Edward Elgar Publishing, 2022, pp. 445–470. [Online]. Available: https://www.elgaronline.com/edcollchap/book/9781839104503/book-part-9781839104503-33.xml

[5] D. Gambetta, *Trust: Making and Breaking Cooperative Relations*, D. Gambetta, Ed. Blackwell, 1988.

[6] B. Dupont and C. Whelan, "Enhancing relationships between criminology and cybersecurity," *Journal of criminology*, vol. 54, no. 1, pp. 76–92, 2021. [Online]. Available: https://journals.sagepub.com/doi/full/10.1177/00048658211003925

[7] D. Gambetta, *Codes of the Underworld: How Criminals Communicate*, 1st ed. Princeton University Press, 2009. [Online]. Available: https://doi.org/10.1515/9781400833610

[8] D. Décary-Hétu and A. Leppänen, "Criminals and signals: An assessment of criminal performance in the carding underworld," *Security Journal*, vol. 29, no. 3, pp. 442–460, 2016. [Online]. Available: https://link.springer.com/article/10.1057/sj.2013.39

[9] J. Lusthaus, "Trust in the world of cybercrime," *Global crime*, vol. 13, no. 2, pp. 71–94, 2012. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/17440572.2012.674183

[10] L. Allodi, R. Ricaldi, J. Wientjes, and A. Radu, "Where is dmitry going? framing 'migratory' decisions in the criminal underground," 2024. [Online]. Available: https://arxiv.org/abs/2411.16291

[11] T. Tsuchiya, A. Cuevas, and N. Christin, "Identifying risky vendors in cryptocurrency P2P marketplaces," in *Proceedings of the ACM Web Conference 2024*. ACM, 2024, pp. 99–110. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3589334.3645475

[12] M. R. J. Soudijn and B. C. H. T. Zegers, "Cybercrime and virtual offender convergence settings," *Trends in Organized Crime*, vol. 15, no. 2-3, pp. 111–129, 2012. [Online]. Available: http://link.springer.com/10.1007/s12117-012-9159-z

[13] B. Dupont, A.-M. Côté, J.-I. Boutin, and J. Fernandez, "Darkode: Recruitment patterns and transactional features of "the most dangerous cybercrime forum in the world"," *American Behavioral Scientist*, vol. 61, no. 11, pp. 1219–1243, 2017. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0002764217734263

[14] A. Cuevas and N. Christin, "Does online anonymous market vendor reputation matter?" in *Proceedings of the USENIX Security Symposium*. USENIX Association, 2024, pp. 4641–4656. [Online]. Available: https://www.usenix.org/conference/usenixsecurity24/presentation/cuevas

[15] M. Yip, C. Webber, and N. Shadbolt, "Trust among cybercriminals? carding forums, uncertainty and implications for policing," pp. 108–131, 2017. [Online]. Available: https://eprints.soton.ac.uk/348014/

[16] D. Georgoulias, J. M. Pedersen, M. Falch, and E. Vasilomanolakis, "A qualitative mapping of darkweb marketplaces," in *Proceedings of the APWG Symposium on Electronic Crime Research*. IEEE, 2021, pp. 1–15. [Online]. Available: https://doi.org/10.1109/eCrime54498.2021.9738766

[17] A. Hutchings and T. J. Holt, "A crime script analysis of the online stolen data market," *British Journal of Criminology*, vol. 55, no. 3, pp. 596–614, 2015. [Online]. Available: https://doi.org/10.1093/bjc/azu106

[18] T. J. Holt and E. Lampke, "Exploring stolen data markets online: Products and market forces," *Criminal Justice Studies*, vol. 23, no. 1, pp. 33–50, 2010. [Online]. Available: https://doi.org/10.1080/14786011003634415

[19] J. Hughes, A. Caines, and A. Hutchings, "Argot as a trust signal: Slang, jargon & reputation on a large cybercrime forum," in *Workshop on the Economics of Information Security*. Geneva, Switzerland: WEIS, 2023, pp. 1–11. [Online]. Available: https://weis2023.econinfosec.org/wp-content/uploads/sites/11/2023/06/weis23-hughes.pdf

[20] A. V. Vu, J. Hughes, I. Pete, B. Collier, Y. T. Chua, I. Shumailov, and A. Hutchings, "Turning up the dial: The evolution of a cybercrime market through set-up, stable, and covid-19 eras," in *Proceedings of the ACM Internet Measurement Conference*. ACM, 2020, pp. 551–566. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3419394.3423636

[21] T. Marjanov, K. Ioannidis, T. Hyndman, N. Seyedzadeh, and A. Hutchings, "Breaking the ice: Using transparency to overcome the cold start problem in an underground market," in *Workshop on the Economics of Information Security*. Dallas, TX, USA: WEIS, 2024, pp. 1–12. [Online]. Available: https://www.repository.cam.ac.uk/items/a7c91f75-e0d0-4628-ac1a-823858e06824

[22] T. J. Holt, O. Smirnova, and A. Hutchings, "Examining signals of trust in criminal markets online," *Journal of Cybersecurity*, vol. 2, no. 2, pp. 137–145, 2016.

[23] T. Marjanov and A. Hutchings, "SoK: Digging into the digital underworld of stolen data markets," in *2025 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2025, pp. 1–18. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/11023368

[24] T. Garkava, A. Moneva, and E. R. Leukfeldt, "Stolen data markets on telegram: a crime script analysis and situational crime prevention measures," *Trends in Organized Crime*, pp. 1–25, 2024. [Online]. Available: https://link.springer.com/article/10.1007/s12117-024-09532-6

[25] M. Alizadeh, M. Kubli, Z. Samei, S. Dehghani, M. Zahedivafa, J. D. Bermeo, M. Korobeynikova, and F. Gilardi, "Open-source llms for text annotation: a practical guide for model setting and fine-tuning," *Journal of Computational Social Science*, vol. 8, p. 17, 2024. [Online]. Available: https://www.zora.uzh.ch/id/eprint/266592/1/s42001_024_00345_9.pdf

[26] T. Gao, J. Jin, Z. T. Ke, and G. Moryoussef, "A comparison of deepseek and other llms," arXiv, Tech. Rep., 2025. [Online]. Available: https://arxiv.org/abs/2502.03688

[27] M. Campobasso and L. Allodi, "Know your cybercriminal: Evaluating attacker preferences by measuring profile sales on an active, leading criminal market for user impersonation at scale," in *Proceedings of the USENIX Security Symposium*. USENIX Association, 2023, pp. 553–570. [Online]. Available: https://www.usenix.org/system/files/usenixsecurity23-campobasso.pdf

[28] British Society of Criminology, "Statement of ethics," https://www.britsoccrim.org/ethics/, 2015.

[29] U. D. of Homeland Security, "Combatting illicit activity utilizing financial technologies and cryptocurrencies," Department of Homeland Security, Tech. Rep., 2022. [Online]. Available: https://publicintelligence.net/dhs-combatting-illicit-cryptocurrency-activity-phase-1/

[30] D. Laferrière and D. Décary-Hétu, "Examining the uncharted dark web: Trust signalling on single vendor shops," *Deviant Behavior*, vol. 44, no. 1, pp. 37–56, 2023. [Online]. Available: https://doi.org/10.1080/01639625.2021.2011479

[31] E. B. Sasson, A. Chiesa, C. Garman, M. Green, I. Miers, E. Tromer, and M. Virza, "Zerocash: Decentralized Anonymous Payments from Bitcoin," in *2014 IEEE symposium on security and privacy*. IEEE, 2014, pp. 459–474. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6956581

[32] M. Yip, N. Shadbolt, and C. Webber, "Why forums?: an empirical analysis into the facilitating factors of carding forums," in *Proceedings of the Annual ACM Web Science Conference*. ACM, 2013, pp. 453–462. [Online]. Available: https://dl.acm.org/doi/10.1145/2464464.2464524

APPENDIX

A. Initial Telegram community selection

This appendix provides more information on the initial selection of forums. Table A1a provides the thread categories of criminal forums we searched for references to Telegram channels, and Table A1b provides the list of keywords we used in the search function of Telegram.

TABLE A1: Cybercrime across forums and Telegram

(a) Forum sections by cybercrime category

Cybercrime Category	Common Forum Sections
Carding & Fraud	Carding, CVV/Dumps, Fullz, Bank Drops, Fraud Techniques, Gift Carding.
Malware & Tools	Malware Development, Crypters, Stealers, Loaders, RATs, Exploits, Botnets.
Phishing & Identity Theft	Phishing Kits, Email Templates, Social Engineering, Spoofing Tools, Scam Pages, Identity Theft.
Access Services	RDP Sales, Web Shells, Domain Access, Admin Panel Hacks, Database Access, VPN/Proxy Rentals.
Marketplaces & Shops	Verified Sellers, Escrow Services, Market Listings, Shops, Wholesalers, Bulk Logs Sales.
Leaks & Data Dumps	Database Leaks, Combo Lists, Credentials Dumps, Account Dumps, Private Logs.

(b) Telegram keywords by cybercrime category

Cybercrime Category	Telegram Search Keywords
Carding & Fraud	carding, cvv, fullz, dumps, cashout, bin, bank logs, PayPal logs.
Malware & Tools	botnet, stealer, crypter, RAT, loader, spreader, malware, malware-as-a-service, keylogger.
Phishing & Identity Theft	phishing kit, otp bot, email bomber, sms flooder, spoof call bot, identity dump, fake login page.
Access Services	RDP access, shell access, root access, bot shop, credentials access, hacked admin panel.
Marketplaces & Shops	hack shop, dump shop, card shop, logs shop, dark market.
Leaks & Data Dumps	combo list, database leak, breach dump, credentials, private logs, leaked db.

B. Overview of scraped data

This appendix documents the raw data as scraped. Table A2 provides the full list of scraped data variables we recorded,

TABLE A2: Overview of scraped data variables

(a) Community-level variables

Field	Description
`community id`	Unique identifier for the Telegram community
`name`	Name of the community
`handle`	Public handle of the community, if available
`about`	Description or bio of the community
`type`	Whether the community is public or private
`participants count`	Number of participants in the community
`pinned message id`	ID of the pinned message in the community

(b) Message-level variables

Field	Description
`message id`	Unique message identifier within a community
`community id`	Unique identifier for the Telegram community
`user id`	Unique identifier of message sender
`username`	Username of the sender, if available
`timestamp`	Time when the message was sent
`content`	The text content of the message
`number of views`	Number of times the message was viewed
`number of forwards`	Number of times the message was forwarded
`reaction list`	List of message reactions (e.g., likes, emojis)
`type`	Type of message (text, image, video, etc.)

both on community and on message level.

C. Full list of market offerings

This section documents the full list of illicit products from each market segment as seen in Table A3.

TABLE A3: Market segment offerings

Market Segment	Offerings
Cyberattacks	Malware-as-a-service, obfuscation tools, exploit kits, hacking services, remote access trojans, crypters and packers, DDoS botnets, credential stealers, C2 panel rentals, brute-force tools.
Personal data	Combo lists (email:password), fullz packages (SSNs, DOBs, addresses), bank login dumps, stolen accounts, credit card dumps, subscription service logins (Netflix, Spotify, etc.).
Fraud	Cashout services, carding tools, crypto scams, gift card reselling, fake store receipts, fake invoices, drop services, refunding services, online store manipulation scripts.
Infrastructure	RDP/VPN rentals, bulletproof hosting, rotating proxy services, SMTP servers for spam, exploit servers, SMS/OTP bypass services.
Tutorials	Carding tutorials, phishing kit walkthroughs, PDF eBooks on refunding, Telegram mentorship programs, malware deployment guides, step-by-step fraud plans.
Piracy	Cracked software, license key generators, game bypass tools, pirated movies/music, torrent invites, streaming platform hacks.

D. LLM performance

This appendix documents the performance of DeepSeek-V3 and GPT-4o for the two annotation tasks. Table A4a and Table A4b show the performance of the two LLMS for the market segment and the trust signal annotation task respectively.

TABLE A4: DeepSeek-V3 and GPT-4o performance

(a) Telegram market segment labeling

Label	Human Count	DeepSeek-V3				GPT-4o			
		Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
Cyberattacks	11	0.818	1.000	0.818	0.900	0.636	0.778	0.636	0.700
Digital Infrastructure	4	0.750	0.750	0.750	0.750	0.500	0.500	0.500	0.500
Digital Piracy	3	1.000	1.000	1.000	1.000	1.000	0.500	1.000	0.667
Fraud Tools	10	1.000	0.714	1.000	0.833	0.700	0.583	0.700	0.636
Personal Data	19	0.895	1.000	0.895	0.944	0.947	1.000	0.947	0.973
Tutorials	3	0.667	1.000	0.667	0.800	0.333	1.000	0.333	0.500
Overall	-	**0.88**	**0.911**	**0.855**	**0.871**	**0.76**	**0.727**	**0.686**	**0.663**
Cohen's Kappa	-	**0.843**	-	-	-	**0.685**	-	-	-

(b) Telegram trust signal labeling

Label	Human Count	DeepSeek-V3				GPT-4o			
		Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
Customer Support	13	1.000	0.684	1.000	0.813	0.923	1.000	0.923	0.960
Scam Reporting	4	0.500	0.500	0.500	0.500	0.500	0.667	0.500	0.571
Free Sample/Trial	22	0.955	0.913	0.955	0.933	0.955	0.840	0.955	0.894
Escrow Service	18	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Warranty	5	0.800	1.000	0.800	0.889	1.000	1.000	1.000	1.000
Automation	7	0.714	0.500	0.714	0.588	1.000	0.778	1.000	0.875
Digital Wallets	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Nothing	431	0.968	0.983	0.968	0.975	0.984	0.998	0.984	0.991
Proof of Delivery	5	0.600	0.375	0.600	0.462	0.800	0.800	0.800	0.800
Clear Rules	2	0.500	1.000	0.500	0.667	1.000	0.500	1.000	0.667
Vouching System	3	0.667	1.000	0.667	0.800	1.000	1.000	1.000	1.000
Cryptocurrency	9	0.889	0.500	0.889	0.640	1.000	0.750	1.000	0.857
Overall	-	**0.937**	**0.852**	**0.952**	**0.940**	**0.966**	**0.901**	**0.977**	**0.974**
Cohen's Kappa	-	**0.760**	-	-	-	**0.871**	-	-	-

E. Further details on the Monte Carlo simulations

Table A5 lists all the variables recorded during the Monte Carlo simulations. Table A6 provides a breakdown of community and message level statistics of user exposure to trust signals across market segments. Those are the raw results of the Monte Carlo simulation exercise that were used to produce the figures to answer RQ4. Table A7 breaks down individual trust signal statistics as encountered by simulated agents, and Table A9 further breaks down those statistics across market segments.

TABLE A5: Recorded variables during simulation

Variable	Purpose
`cybercriminal_id`	Unique identifier generated for each simulated attacker.
`market_segment`	Indicates the specific illicit market segment randomly assigned.
`community_id`	Identifies which Telegram community was visited.
`n_views`	Total number of views for sampled messages in a community; used as a proxy for visibility and engagement.
`n_reactions`	Total reactions on sampled messages; indicates user interaction and community activity levels.
`ts_count`	Total number of trust signals encountered in a single community visit.
`unique_ts_count`	Number of distinct trust signal types seen in the community; reflects diversity of trust strategies.
`[mechanism name]`	Individual columns for each trust-building mechanism to record how many times each was encountered in the sampled messages.

TABLE A6: Trust signal exposure statistics by market segment

(a) Community-level statistics

Market segment	Mean	Median	Std. Dev.	95% CI
Cyberattacks	3.56	2.75	2.70	0.75 - 11.22
Digital Infrastructure	7.43	6.33	5.22	0.33 - 20.50
Digital Piracy	3.68	3.00	3.29	0.00 - 12.00
Fraud Tools	5.19	4.20	4.11	0.33 - 15.67
Personal Data	5.99	4.80	4.82	0.33 - 18.00
Tutorials	1.16	1.00	0.89	0.00 - 3.25
Overall	4.49	3.00	4.28	0.00 - 16.00

(b) Message-level statistics

Market segment	Mean	Median	Std. Dev.	95% CI
Cyberattacks	0.17	0.14	0.10	0.05 - 0.43
Digital Infrastructure	0.31	0.28	0.18	0.02 - 0.72
Digital Piracy	0.15	0.13	0.12	0.00 - 0.44
Fraud Tools	0.21	0.18	0.14	0.02 - 0.53
Personal Data	0.28	0.25	0.19	0.02 - 0.72
Tutorials	0.05	0.04	0.03	0.00 - 0.12
Overall	0.19	0.15	0.16	0.00 - 0.60

TABLE A7: Trust signal aggregate exposure statistics

Trust signal	Mean	Median	Std. Dev.	95% CI
Clear Rules	0.07	0.04	0.09	0.02 – 0.33
Cryptocurrency	0.06	0.04	0.05	0.02 – 0.18
Customer Support	0.07	0.05	0.06	0.02 – 0.21
Digital Wallets	0.04	0.04	0.03	0.02 – 0.11
Escrow Service	0.12	0.05	0.20	0.02 – 1.00
Free Sample/Trial	0.07	0.05	0.09	0.02 – 0.27
Proof of Delivery	0.05	0.04	0.07	0.02 – 0.13
Scam Reporting	0.05	0.04	0.04	0.02 – 0.13
Automation	0.07	0.05	0.07	0.02 – 0.25
Vouching System	0.10	0.04	0.15	0.02 – 0.50
Warranty	0.05	0.04	0.04	0.02 – 0.14
Overall	0.07	0.04	0.09	0.02 – 0.25

F. Raw trust signals across market segments

Table A8 breaks down the full distribution of trust signals across all market segments.

TABLE A8: Distribution of trust signals across market segments

Trust Signals	Cyberattacks		Digital Infrastructure		Digital Piracy		Fraud Tools		Personal Data		Tutorials	
	Private	Public	Private	Public	Private	Public	Private	Public	Private	Public	Private	Public
Clear Rules	82	7,370	8	2,535	19	1,668	4	12,846	6	570	-	32
Customer Support	127	54,938	7	29,568	17	855	15	84,764	554	2,338	-	49
Digital Wallets	234	6,243	3	1,858	7	15	179	16,332	331	1,505	-	16
Escrow Service	226	9,923	-	7,243	60	8	33	18,935	10	5,539	-	4
Free Sample/Trial	353	2,047	3	10,152	264	162	16	20,629	19	8,163	-	106
Proof of Delivery	53	31,218	-	14,937	11	568	5	57,240	6	1,112	-	18
Scam Reporting	227	1,507	8	911	33	429	25	2,516	13	552	-	101
Automation	259	35,932	86	3,564	62	536	167	63,665	588	3,127	-	62
Vouching System	63	45,365	-	3,188	17	19	4	16,573	7	1,141	-	6
Warranty	38	947	34	9,854	40	134	-	17,069	442	4,939	-	9
Total trust messages	1,662	195,490	149	83,810	530	4,394	448	310,569	1,976	28,986	-	403
Total messages	12,732	207,631	483	209,320	10,649	53,646	1,843	472,388	3,495	111,079	-	32,805
Communities	14	21	2	9	2	21	3	52	3	33	-	7

TABLE A9: Trust signal statistics by market segments

Trust Signal	**Cyberattacks**				**Infrastructure**				**Digital Piracy**			
	Mean	**Median**	**SD**	**95% CI**	**Mean**	**Median**	**SD**	**95% CI**	**Mean**	**Median**	**SD**	**95% CI**
Clear Rules	0.07	0.06	0.03	0.07–0.07	0.07	0.07	0.03	0.07–0.07	0.07	0.04	0.09	0.02–0.33
Cryptocurrency	0.06	0.04	0.05	0.06–0.06	0.16	0.14	0.11	0.16–0.16	0.06	0.04	0.05	0.02–0.18
Customer Support	0.12	0.08	0.12	0.11–0.12	0.16	0.14	0.11	0.16–0.17	0.07	0.05	0.06	0.02–0.21
Digital Wallets	0.05	0.04	0.04	0.05–0.05	0.05	0.05	0.04	0.05–0.05	0.04	0.04	0.03	0.02–0.11
Escrow Service	0.15	0.11	0.13	0.14–0.15	0.13	0.11	0.11	0.12–0.13	0.12	0.05	0.20	0.02–1.00
Free Sample/Trial	0.12	0.06	0.14	0.11–0.12	0.10	0.06	0.12	0.10–0.10	0.07	0.05	0.09	0.02–0.27
Proof of Delivery	0.11	0.08	0.09	0.11–0.11	0.08	0.06	0.08	0.08–0.08	0.05	0.04	0.07	0.02–0.13
Scam Reporting	0.08	0.06	0.06	0.08–0.08	0.11	0.11	0.08	0.11–0.12	0.05	0.04	0.04	0.02–0.13
Automation	0.19	0.15	0.15	0.18–0.19	0.25	0.20	0.17	0.24–0.25	0.07	0.05	0.07	0.02–0.25
Vouching System	0.35	0.18	0.31	0.34–0.36	0.06	0.05	0.05	0.05–0.06	0.10	0.04	0.15	0.02–0.50
Warranty	0.11	0.10	0.07	0.11–0.11	0.12	0.10	0.09	0.12–0.12	0.05	0.04	0.04	0.02–0.14
Overall	0.07	0.04	0.09	0.02–0.25	0.07	0.04	0.09	0.02–0.25	0.07	0.04	0.09	0.02–0.33

Trust Signal	**Fraud Tools**				**Personal Data**				**Tutorials**			
	Mean	**Median**	**SD**	**95% CI**	**Mean**	**Median**	**SD**	**95% CI**	**Mean**	**Median**	**SD**	**95% CI**
Clear Rules	0.08	0.05	0.09	0.08–0.08	0.08	0.08	0.04	0.08–0.08	0.03	0.03	0.03	0.03–0.03
Cryptocurrency	0.09	0.06	0.10	0.08–0.09	0.18	0.16	0.12	0.18–0.19	0.04	0.03	0.04	0.04–0.04
Customer Support	0.08	0.07	0.07	0.08–0.08	0.15	0.10	0.13	0.15–0.15	0.04	0.03	0.04	0.04–0.04
Digital Wallets	0.07	0.06	0.06	0.07–0.07	0.11	0.10	0.08	0.11–0.11	0.04	0.03	0.04	0.04–0.04
Escrow Service	0.14	0.11	0.13	0.13–0.14	0.11	0.10	0.09	0.10–0.11	0.05	0.04	0.05	0.05–0.05
Free Sample/Trial	0.10	0.07	0.10	0.10–0.10	0.18	0.10	0.17	0.17–0.18	0.05	0.04	0.06	0.05–0.05
Proof of Delivery	0.09	0.07	0.08	0.09–0.09	0.06	0.05	0.05	0.06–0.06	0.04	0.03	0.04	0.04–0.04
Scam Reporting	0.06	0.04	0.07	0.06–0.06	0.08	0.05	0.07	0.08–0.08	0.05	0.04	0.05	0.05–0.05
Automation	0.16	0.11	0.14	0.16–0.16	0.20	0.17	0.13	0.19–0.20	0.06	0.05	0.06	0.06–0.06
Vouching System	0.34	0.21	0.32	0.33–0.35	0.47	0.12	0.41	0.45–0.49	0.03	0.03	0.03	0.03–0.03
Warranty	0.09	0.06	0.08	0.09–0.10	0.15	0.15	0.09	0.14–0.15	0.04	0.03	0.04	0.04–0.04

G. Annotation codebooks

This appendix describes the annotation tasks for market segments and trust signals in full detail.

1) Market segments annotation task: For the market segment annotation task, the procedure is as follows.

Market segments:

- Cyberattacks: Malware, DDoS services, hacking and spamming tools (e.g., exploit kits, phishing kits).
- Personal Data: Stolen credentials, payment/identity data (e.g., credit card dumps, SSNs).
- Fraud Tools: Tools for financial fraud or deception (e.g., fake ID generators, carding guides).
- Digital Infrastructure: Services supporting cybercrime (e.g., bulletproof hosting, VPNs, proxies).
- Tutorials: Educational material on cybercrime techniques (e.g., malware creation, phishing guides).
- Digital Piracy: Unauthorized digital content (e.g., cracked software, pirated media).

Annotation Procedure:

- Review group/channel metadata (description, pinned posts, recent messages).
- Determine the dominant activity.
- Assign the most appropriate segment label.
- Add notes if the case is ambiguous or requires clarification.

Annotation Format:

- `Group/Channel ID`
- `Segment Label` (one of the categories above)
- `Notes` (optional)

Based on the results on the manual annotation, we developed the following prompt for the LLMs.

Task: You are a cybersecurity expert categorizing Telegram channels into distinct cybercriminal market segments. Each channel belongs to one of the categories. Your task is to analyze the messages from a single channel and classify them into one of the categories listed below, based on its primary activity.

Market Segmentation Categories

- Cyberattacks: Malware distribution, DDoS services, hacking tools, spamming tools, system compromise activities.
- Personal data: Trading or sharing stolen personal data (credentials, payment info, identity documents).
- Fraud tools: Tools and methods for financial fraud or deception (fake ID generators, social media growth, ban/unban services, carding, scamming scripts).
- Digital Infrastructure: Services/resources supporting cybercriminal activities (bulletproof hosting, VPNs, proxies, botnets).
- Tutorials: Tutorials, guides, or instructions for learning cybercriminal techniques (hacking, malware creation, security bypassing).
- Digital piracy: Unauthorized distribution or sale of copyrighted content (cracked software, pirated movies, music, eBooks).

Your classification should reflect the primary purpose or most consistent theme of the channel, not isolated or occasional content.

2) Trust-building mechanisms annotation task: For the trust-building mechanism annotation task, the procedure is as follows.

Trust-building mechanism categories:

- Vouching System: Endorsements from known members or links to vouch groups.
- Scam Reporting: Public reports or scam lists warning of fraudulent actors.
- Escrow Service: Neutral third-party holds payment until both sides fulfill terms.
- Digital Wallets: Use of traditional or fintech wallets (e.g., PayPal, Zelle, Revolut).
- Cryptocurrency: Crypto-based payments (e.g., BTC, XMR, ETH); includes wallets like MetaMask, Trust Wallet.
- Automation: Bots used for payments, transactions, or access automation.
- Clear Rules/Guidelines: Pinned/group descriptions outlining platform rules.
- Proof of Delivery: Screenshots, tracking, or customer testimonials.
- Customer Support: Mention of after-sale support or availability to help or answer questions about a product or service.
- Free Samples/Trials: Provides a free sample or trial of product or service.
- Warranty: Offering guarantees about the product or service to increase credibility.

Annotation Procedure:

- Read each message in context.
- Identify any trust-building mechanisms from the list above (multiple allowed).
- Link the mechanisms to the relevant market segment.
- Add notes if context is ambiguous or novel mechanisms are observed.

Annotation Format:

- `Group/Channel ID`
- `Message ID`
- `Trust-Building Mechanisms` (multiple allowed)
- `Associated Market Segment` (e.g., Fraud Tools, Tutorials)
- `Notes` (optional)

Based on the results on the manual annotation, we developed the following prompt for the LLMs.

Task: You are analyzing messages from cybercriminal markets on Telegram to identify the presence of trust-building mechanisms. Each message may reflect one, multiple, or none of these mechanisms. Your task is to assign the appropriate trust-building mechanism labels to each message using the categories below. If no trust-building signal is present, return only *Nothing*.

`Trust-Building Mechanism Categories:`

- Vouching System: Endorsements or social proof from within the community.
- Scam Reporting: Public disclosure of fraudulent actors or activities.
- Escrow Service: Mention the use of escrow or a neutral third party to ensure transaction security.
- Digital Wallets: Indication of payment through conventional or digital wallet systems.
- Cryptocurrency: Use of decentralized digital currencies for secure transactions.
- Automation: Automation of services or transactions through bots within Telegram.
- Clear Rules/Guidelines: Presence of formalized rules governing behavior or exchanges.
- Proof of Delivery: Evidence shared to verify that a transaction or service was completed.
- Customer Support: Offering assistance or communication channels for buyer reassurance.
- Free Samples/Trials: Provision of products or services at no cost to demonstrate legitimacy.
- Warranty: Statements guaranteeing product reliability or refund/replacement policies.

Return all labels that apply to the message, separated by commas. If no label is applicable, return only *Nothing*.