BforeAI

Predicting The Unseen

An In-Depth Analysis of BforeAI's Predictive Technology

Powered by

DNSFilter®

Predicting the Unseen: An In-Depth Analysis of Bfore.Ai's Predictive Technology

Abstract

Traditional cybersecurity strategies are reactive by responding to threats after they breach a system. In contrast, predictive analytics enables a proactive approach with the identification of potential threats before they materialize and breach a system. This research explores the efficacy of Bfore.Ai's predictive analytics-driven cyber defense in the domain-based threat intelligence landscape by analyzing a curated sample of domains provided by Bfore.Ai with other sources. The findings reveal that Bfore.Ai's domain feed is highly unique and demonstrates remarkable speed in identifying threats, outperforming other vendors in 90% of cases and identifying threats before others by more than three weeks in 50% of instances. Additionally, an analysis against the Tranco List suggests a very small potential for false positives within Bfore.Ai's sample feed, emphasizing Bfore.Ai's accurate and reliable threat identification capabilities. Overall, this research demonstrates the effectiveness of Bfore.Ai's predictive analytics-driven cyber defense, enhancing the domain-based threat intelligence landscape and contributing to more robust cybersecurity measures. Bfore.Ai's technology offers promising solutions for organizations seeking to strengthen their security posture against evolving cyber threats.

1. Introduction

A new generation of cyber defense has emerged, enabled by the power of predictive analytics. Traditional cyber defense strategies have often relied on reactive approaches, responding to cyber threats after they have already breached a system or caused damage. However, with the integration of predictive analytics, organizations can now shift towards a proactive stance, better equipped to anticipate, and prevent potential cyber threats before they materialize.[1]

Predictive analytics leverages advanced algorithms and data analysis techniques to identify patterns, trends, and anomalies within vast amounts of historical and real-time data. This includes information about network traffic, user behavior, system logs, and other relevant data sources. By analyzing this data, cybersecurity professionals can identify potential threats, vulnerabilities, or indicators of compromise long before an actual attack occurs.

Bfore.Ai uses predictive technology to evaluate and identify malicious infrastructures before an attack is launched by the threat actor. Their technology uses behavioral analytics wherein billions of data points are collected daily and scored based on previously known "good" and identified "bad" behaviors, to match

[1] Husák, M., Bartoš, V., Sokol, P., & Gajdoš, A. (2021). Predictive methods in cyber defense: Current experience and research challenges. *Future Generation Computer Systems*, *115*, 517-530.

malicious behaviors in infrastructure set-up with previously known malicious behaviors.

The key advantages of using predictive analytics in cyber defense generally include, early threat detection, unique threat identification, and reduced false positives.[2] To determine whether these advantages of predictive analytics are reflected in Bfore.Ai's technology, a cybersecurity company that provides DNS filtering and threat protection services, DNSFilter, created a benchmark analysis report to assess the efficacy of Bfore.Ai's domain categorization and threat identification capabilities.

The primary objectives of this research are to evaluate the uniqueness and overlap of Bfore.Ai's domain feed with other security feeds, assess the speed of Bfore.Ai's threat identification in comparison to other security vendors, and analyze the potential for false positives. Additionally, we delve into customer query overlap to determine the efficacy of Bfore.Ai's threat identification in real-time scenarios. The remainder of this paper will outline the methodologies employed for data analysis, and the key findings derived from the investigation.

2. Research Methodology

Data Collection: The research study was based on the dataset provided by Bfore.Ai, comprising over 1.8 million Fully Qualified Domain Names collected over a span of 6 months, ranging from November 24, 2022, to May 24, 2023. DNSFilter performed an extensive analysis specifically focusing on indicators within Bfore.Ai's sample that possessed a threat score surpassing the threshold of 0.7.[3] This amounted to: 1,817,530 Fully Qualified Domain Names (FQDNs).

Feed Overlap and Speed Evaluation: To assess the uniqueness and overlap of Bfore.Ai's domain feed, DNSFilter conducted an in-depth comparison with other security feeds. The analysis focused on identifying the percentage of domains that appeared in multiple feeds, thus providing insights into the distinctiveness of Bfore.Ai's dataset. Furthermore, the speed of Bfore.Ai's threat identification was evaluated by comparing the identification time of malicious domains in Bfore.Ai's feed with that of other security vendors. The analysis sought to determine the efficiency and promptness of Bfore.Ai's threat identification capabilities in comparison to its competitors.

Tranco List Analysis: To investigate potential false positives in Bfore.Ai's domain feed, DNS Filter performed a comparative analysis using the Tranco List. The Tranco List maintains a compilation of top websites based on popularity and traffic, making it a valuable reference to assess the legitimacy of domains. Domains from Bfore.Ai's feed that appeared in the Tranco top 1 million List were considered as potential candidates for false positives, and their proportion was calculated.

Customer Query Overlap: To gauge the effectiveness of Bfore.Ai's threat identification in real-world scenarios, DNS Filter selected a representative subset of 20,000 FQDNs from Bfore.Ai's feed. This

[2] Pirc, J., DeSanto, D., Davison, I., & Gragido, W. (2016). *Threat forecasting: Leveraging big data for predictive analysis*. Syngress.

[3] At Bfore.Ai domains are given a threat score determined by their association or above are considered malicious.

subset spanned from April 1st to May 1st and was analyzed for query occurrences during the comparative timeframe from February 1st to June 1st. The study aimed to identify the percentage of FQDNs that were queried during this period, and to determine the proportion of FQDNs that were first identified as malicious by Bfore.Ai before being observed in DNS Filter's query dataset.

QTypes and RCodes Analysis: To gain insights into the nature and response patterns of malicious domains, DNSFilter analyzed the observed QTypes and RCodes from the same sample size of 20,000 FQDNs. The study aimed to identify the prevalent types of DNS queries used for these domains and the status of DNS responses they generated.

By employing these methodologies, this research study aims to provide a comprehensive assessment of Bfore.Ai's domain feed, threat identification capabilities, and potential false positives, contributing to the advancement of domain-based threat intelligence and cybersecurity measures.

3. Research Findings

3.1. Feed overlap and discovery time

Overlap Analysis: Based on DNS filter's assessment, it was determined that approximately 2.65% of the domains within Bfore.Ai's domain feed exhibited an overlap with domains sourced from other security feeds as presented in Figure 1. Consequently, a notable 97.35% of the domains derived from Bfore.Ai's feed were distinct and not classified by alternative vendors as malicious entities.

*Figure SEQ Figure * ARABIC 1: Feed Overlap*

Speed Evaluation: Regarding the speed performance, DNSFilter ascertained that Bfore.Ai demonstrated noteworthy efficiency in identifying malicious domains in comparison to other security vendors. Specifically, Bfore.Ai's identification process surpassed other vendors' performance in 90% of the instances examined as presented in Figure 2. Additionally, in 50% of the analyzed cases, Bfore.Ai showcased a substantial lead of more than three weeks in identifying threats ahead of other security vendors.

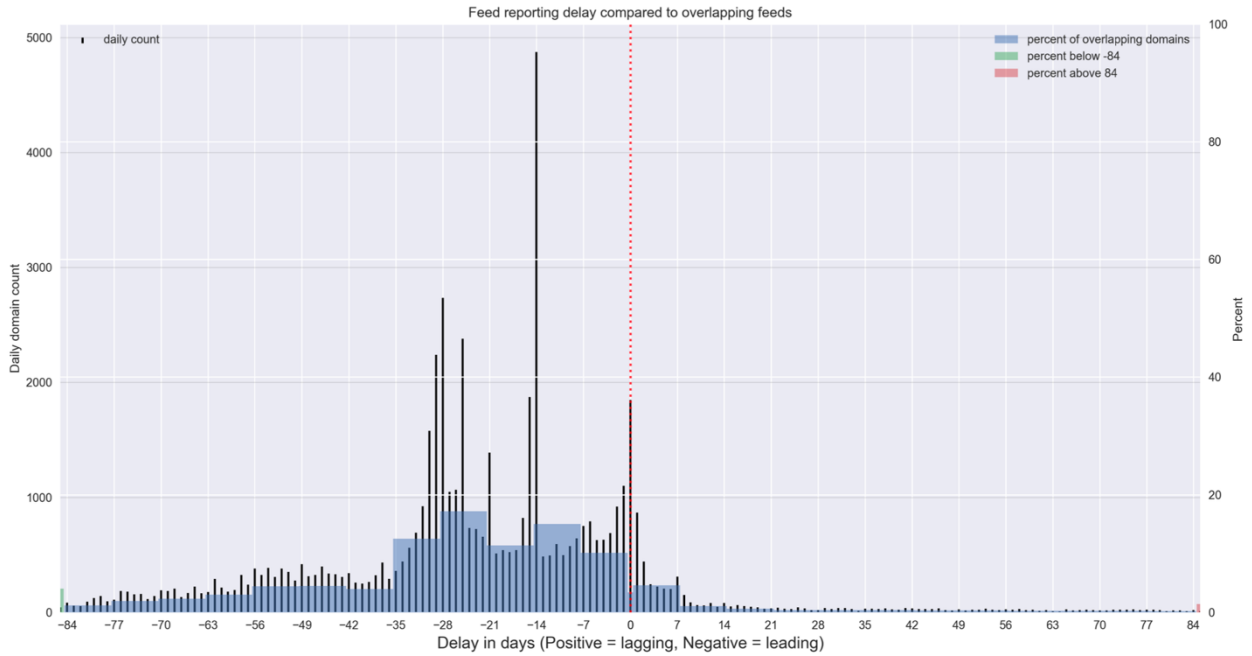


Figure SEQ Figure * ARABIC 2: Speed Overlap

3.2. Tranco List

In the pursuit of assessing potential false positives within Bfore.Ai's domain feed, DNSFilter conducted a comparative analysis against the Tranco List, a well-known project maintaining a compilation of the most popular and heavily trafficked websites on the Internet. The ranking within the Tranco List is established using the available rankings from (currently) five providers as our source data.

Within this investigation, DNSFilter scrutinized the overlap between Bfore.Ai's feed of 1.8 million domains and the domains listed within the Tranco top 1 million List.

The rationale behind this approach stems from the presumption that malicious domains are less likely to be represented on the Tranco List, as the list predominantly comprises legitimate websites.

The findings of this examination indicated that a total of 3318 domains from Bfore.Ai's feed, out of the 1.8 million domains, were identified within the Tranco top 1 million List. Consequently, these 3318 domains emerge as potential candidates for false positives. This implies that although Bfore.Ai marked these

domains as malicious, the presence of these domains on the Tranco List suggests a higher probability that they are indeed legitimate websites.

From a quantitative perspective, the 3318 domains identified within the Tranco top 1 million List account for approximately 0.18% of the domains initially marked as malicious within Bfore.Ai's feed. The low percentage of overlap (0.18%) indicates that the likelihood of Bfore.Ai's feed containing false positives is minimal. This is a positive validation of Bfore.Ai's predictive model low error rate. This accuracy is crucial in the domain-based threat intelligence landscape, as false positives can cause unnecessary disruptions impacting businesses.

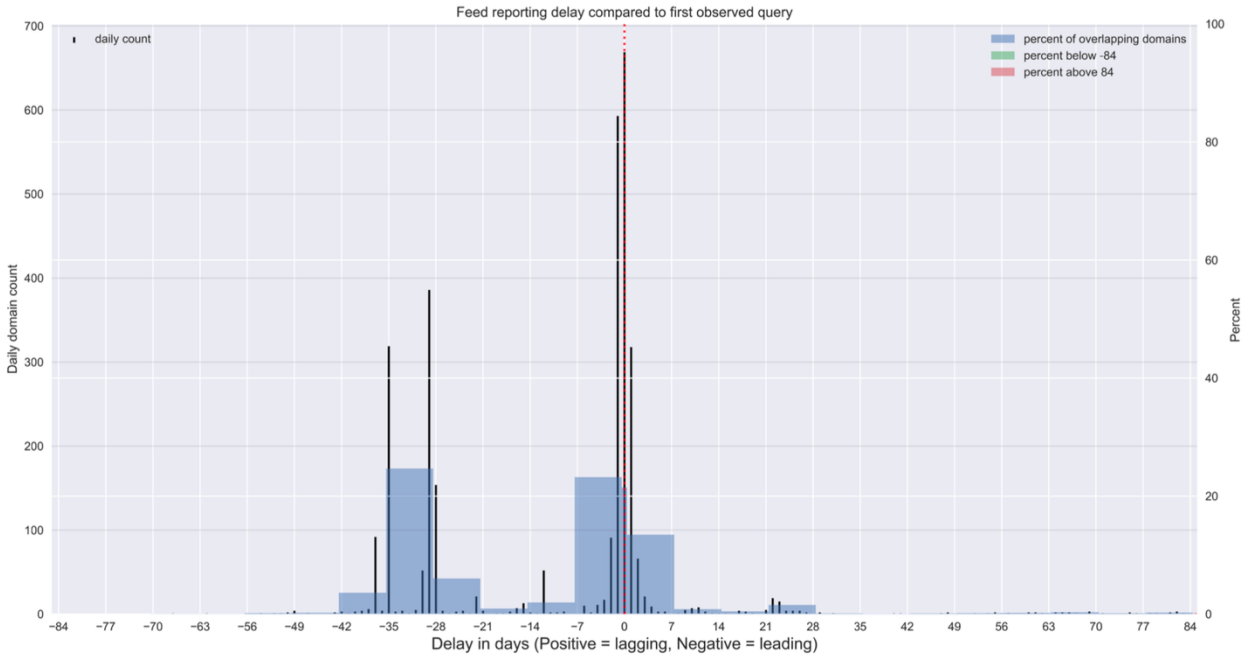
3.3. Customer query overlap analysis

In order to conduct a comprehensive evaluation, DNS Filter selected a representative subset of 20,000 Fully Qualified Domain Names (FQDNs) from Bfore.Ai's larger feed of 1.8 million domains. This subsample was drawn from the period spanning from April 1st to May 1st. A comparative analysis was performed by juxtaposing these FQDNs with DNSFilter's query dataset encompassing the timeframe from February 1st to June 1st. The results of this investigation are as presented in Figure 3 and revealed the following key findings:

Query Occurrence: Among the selected 20,000 FQDNs, approximately 15.6% (equivalent to 3124 FQDNs) were subjected to queries during the designated time period. This indicates that these particular future malicious domains were visited by DNSFilter customers during the comparative timeframe spanning from February 1st to June 1st proving the relevance to the customer base.

First Observation Prevalence: Further analysis uncovered that a notable 60.8% of the FQDNs had been previously reported and identified as malicious by Bfore.Ai before their initial observation in DNSFilter's dataset. In essence, Bfore.Ai demonstrated a proactive stance in identifying these malicious domains before DNSFilter's query dataset even registered their presence.

*Figure SEQ Figure * ARABIC 3: Query Overlap Comparison*



3.4. Query Types and Response Codes

Based on the examination of the same sample size, comprising 20,000 Fully Qualified Domain Names (FQDNs), DNS Filter conducted an analysis of observed QTypes and RCodes to gain insights into the nature of these malicious domains and their responses. QTypes represent numerical codes that elucidate the type of DNS queries employed, thereby shedding light on how these malicious domains are being utilized. On the other hand, RCodes consist of numerical codes that signify the status or outcome of the DNS queries and responses.

Observed QTypes: The analysis revealed the following distribution of Qtypes also presented in Figure 4:

A (Address): 62.68%
A, HTTPS: 35.18%
A, AAAA (IPv6 Address): 1.09%
A, AAAA, HTTPS: 0.61%
HTTPS: 0.10%
SRV (Service): 0.06%
DS, DNSKEY, A: 0.03%
MX (Mail Exchange), A, NS (Name Server), DS, AAAA, HTTPS: 0.03%
HTTPS, NS, A: 0.03%
AAAA, DS, HTTPS, Reserved, A: 0.03%
AAAA (IPv6 Address): 0.03%
TXT (Text): 0.03%
MX, A: 0.03%
NS, A, AAAA, DS, HTTPS: 0.03%
NS, MX, A, TXT: 0.03%
HTTPS, A, AAAA: 0.03%
DS, A, HTTPS, AAAA: 0.03%
A, TXT: 0.03%

The majority of these malicious domains (totaling 99.56%) were predominantly utilized to attract web traffic, as indicated by the high frequency of A (Address) records among the observed QTypes.

Observed RCodes: The RCodes distribution was as follows:

NOERROR (No Error): 97.57%
null: 0.74%
NOERROR, NXDOMAIN (Non-Existent Domain): 0.48%
NXDOMAIN: 0.32%
NOERROR, null, NXDOMAIN: 0.09%

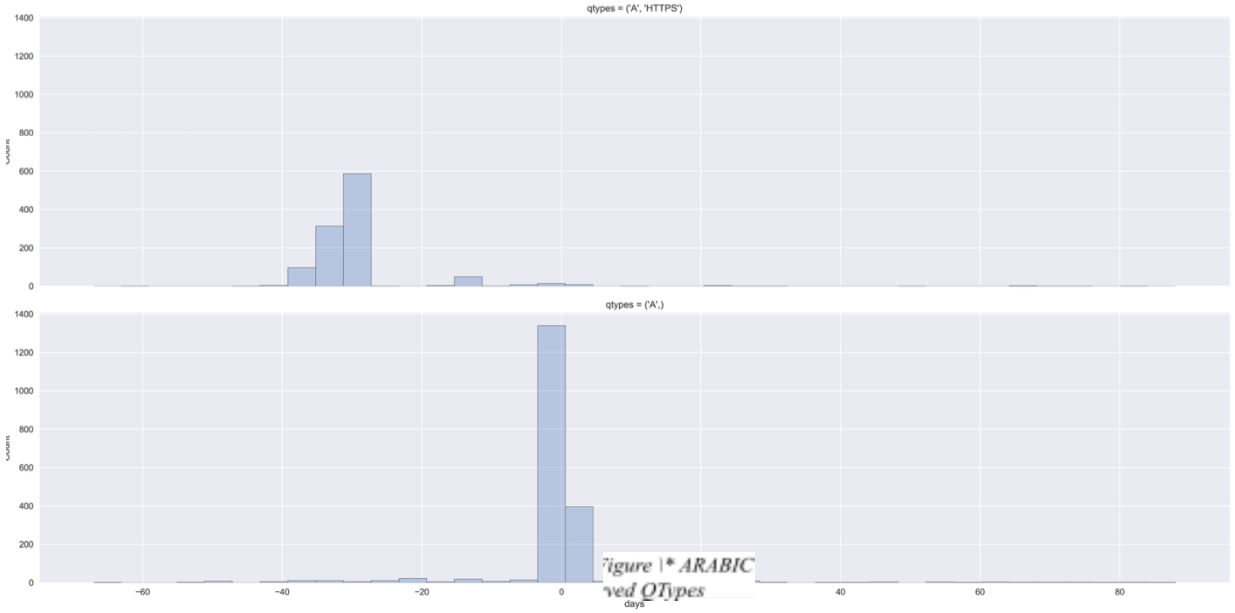


Figure 1 ARABIC
ved QTypes

Notably, the vast majority of domains from Bfore.Ai's feed of 20,000 domains (97.57%) returned a NOERROR status code. This denotes that the DNS queries were successfully executed, and the requested information for the domains was found without any errors or issues.

These findings provide valuable insights into the operational aspects of the malicious domains within the sample dataset. The prevalence of A records and the high percentage of NOERROR responses indicate that these domains are actively seeking web traffic and successfully resolving DNS queries without encountering errors, thereby potentially facilitating their malicious activities.

---

4. Discussion

The investigation revealed valuable insights concerning feed overlap, discovery time, Tranco List analysis, and observed QTypes and RCodes.

Regarding feed overlap, a mere 2.65% of domains in Bfore.Ai's feed exhibited an overlap with other security feeds. This finding indicates that Bfore.Ai's domain feed is highly unique, with a significant majority (97.35%) of domains not identified as malicious by other security vendors. This showcases the distinctiveness and effectiveness of Bfore.Ai's domain curation process. This uniqueness suggests that Bfore.Ai's domain curation process is effective in identifying and categorizing malicious domains that might go undetected by other security solutions. This benefit is crucial because having a highly distinct domain feed means Bfore.Ai can identify threats that other security solutions might miss, providing a more comprehensive and robust defense against cyber threats.

The speed evaluation revealed that Bfore.Ai demonstrated exceptional efficiency in identifying malicious domains. In 90% of cases, Bfore.Ai outperformed other security vendors in identifying threats, and in 50% of instances, these threats were identified more than three weeks ahead of others. This highlights Bfore.Ai's robust threat identification capabilities, indicating its potential to provide early warnings for emerging threats. Bfore.Ai's technology maintains the ability to rapidly identify potential threats, allowing organizations to respond and mitigate attacks at an early stage. Early attack identification is crucial as it eliminates the time threat actors have to dwell within a system, reducing the potential damage and minimizing the associated costs of remediation.

The comparative analysis against the Tranco List highlights a remarkable characteristic of Bfore.Ai's feed, with only a small subset of domains (0.18%) overlapping with the Tranco top 1 million List. The low percentage of overlap indicates that the likelihood of Bfore.Ai's feed containing false positives is minimal and that they boast accurate threat detection capabilities. The findings underscore the reliability of Bfore.Ai's domain classification methodology, reducing the chances of wasting valuable resources and time on investigating non-malicious domains. The benefits of

this low potential for false positives are significant. First and foremost, it ensures that organizations relying on Bfore.Ai's cyber defense solutions can have a high level of confidence in the accuracy of threat identification. This means they can prioritize and respond to real threats more efficiently, focusing their resources on genuine security risks and mitigating potential cyberattacks before they cause harm. Moreover, the low false positive rate reduces the chances of wasting time and resources investigating non-malicious domains. Efficiently distinguishing between genuine threats and false positives is essential to maintain smooth operations without unnecessary interruptions to legitimate services.

In the context of customer query overlap, DNS Filter's analysis of 20,000 FQDNs from Bfore.Ai's feed revealed that 15.6% of them had been queried. This confirms the malicious nature of these domains as they were marked as such during the comparative timeframe and validated by the activity queried by users giving the feed a high degree of relevance.

Additionally, a substantial portion (60.8%) of the FQDNs was first identified as malicious by Bfore.Ai before DNS Filter's dataset recorded their presence. This once again highlights the ability of Bfore.Ai's technology to rapidly identify potential threats and demonstrates their potential to provide early warnings for emerging threats. By staying ahead of the curve, Bfore.Ai enables organizations to proactively prepare and defend against new and evolving cyber threats. This benefit is invaluable in today's rapidly changing threat landscape, where new attack techniques and malware variants constantly emerge.

Lastly, the analysis of observed QTypes and RCodes demonstrated that the majority of the malicious domains were used to attract traffic through multiple sources, as evident from the prevalence of A records (62.68%) and the high percentage of NOERROR responses (97.57%). Overall, the combination of the high prevalence of A records and the high percentage of NOERROR responses indicates that these malicious domains are actively seeking visitors and are well-prepared to engage in potentially harmful activities. The primary objective of these domains is likely to attract web traffic from unsuspecting users and execute malicious actions, such as distributing malware, stealing sensitive information, conducting phishing attacks, or engaging in other fraudulent activities.

5. Conclusions

The research conducted in this study aimed to assess the effectiveness of Bfore.Ai's predictive analytics-driven cyber defense in the domain-based threat intelligence landscape. This research was conducted to understand whether the advantages of predictive security, including early threat detection, unique threat identification, and reduced false positives, also apply to Bfore.Ai and their predictive technology.

By analyzing a curated sample of over 1.8 million FQDNs collected over six months, the study delved into key aspects such as feed overlap, speed evaluation, Tranco List analysis, customer query overlap, and observed QTypes and RCodes.

First, in terms of early threat detection, in 90% of cases, Bfore.Ai outperformed other security vendors in

identifying threats, often identifying threats more than three weeks ahead of others. The research also revealed that a significant portion of the queried domains (60.8%) was first identified as malicious by Bfore.Ai before being observed in DNS Filter's dataset. This ability to stay ahead of the curve enables organizations to proactively prepare and defend against new and evolving cyber threats, which is invaluable in today's rapidly changing threat landscape.

Second, the analysis showcased Bfore.Ai’s unique threat identification capabilities. 97.35% of the domains derived from Bfore.Ai's feed were distinct and not classified by alternative vendors as malicious entities. This uniqueness indicates the distinctiveness and effectiveness of Bfore.Ai's domain curation process. By having a highly distinct domain feed, Bfore.Ai can identify threats that other security solutions might miss, providing a more comprehensive and robust defense against cyber threats.

Finally, Bfore.Ai demonstrated a low false positive rate, with a mere 0.18% of their identified malicious domains showing any likelihood of being false positives. This indicates that the likelihood of Bfore.Ai's feed containing false positives is minimal with accurate threat identification capabilities. Additionally, the low false positive rate ensures that organizations relying on Bfore.Ai's cyber defense solutions can have high confidence in the accuracy of threat identification, allowing them to prioritize and respond to real threats more efficiently.

As a result, this research confirms that Bfore.Ai's predictive technology indeed provides the advantages of predictive security, enhancing the domain-based threat intelligence landscape and contributing to more robust cybersecurity measures. Bfore.Ai's predictive analytics-driven approach has ushered in a new generation of cyber defense, empowering organizations with proactive threat identification capabilities.