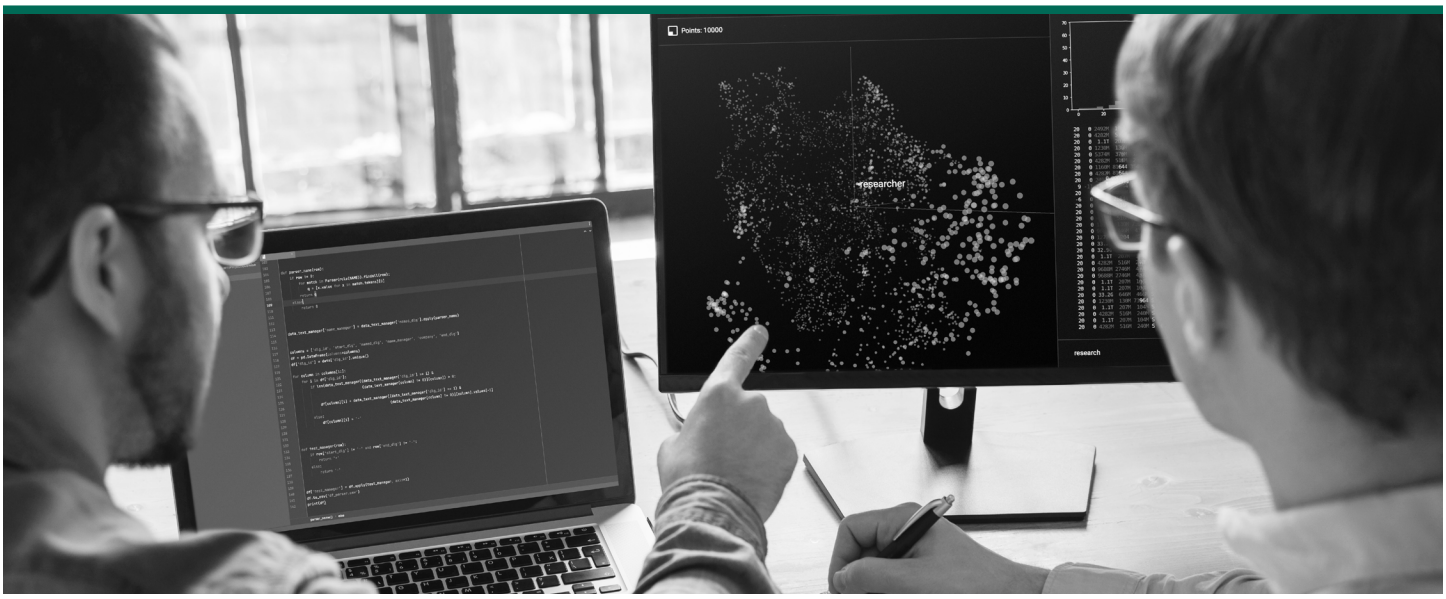
FORTRA

WHITE PAPER

# Inside Fortra's Data Science:

## Advanced Email Threat Detection with Machine Learning

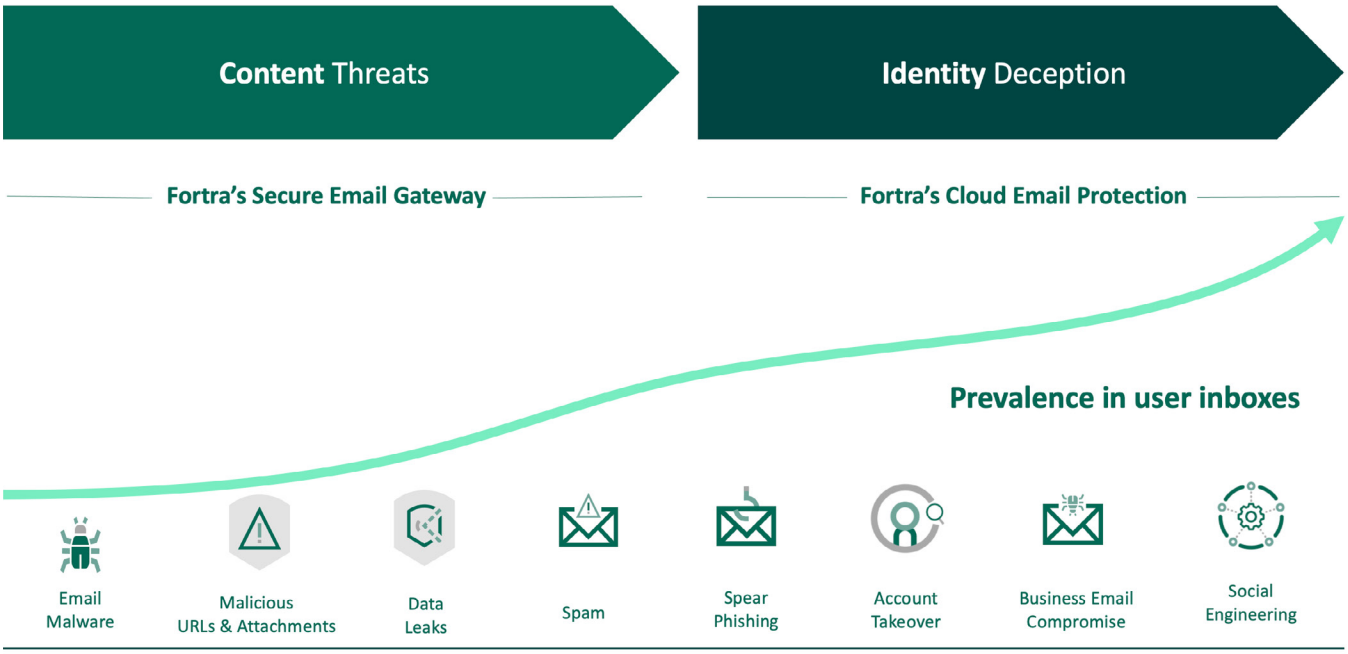



There has been a fundamental shift in the email threat landscape as attackers have moved beyond trying to deceive the email environment to deceiving human beings. These modern attacks leverage impersonation techniques, where the attacker sends a message that appears to come from a known identity—an individual, organization, or consumer brand—that is inherently trusted by the recipient.

This shift is clearly illustrated by the kinds of threats that are currently reaching enterprise user inboxes. In-depth analysis of threats observed in inboxes found that more than 98% of threats getting past enterprise email security controls are impersonation threats like Business Email Compromise (BEC) and credential theft phishing lures. While they can consistently detect malware payloads, Secure Email Gateways and "baked-in" cloud email security add-ons do not reliably stop impersonation tactics and social engineering threats.

In this whitepaper, we will explain how Fortra's Cloud Email Protection leverages a powerful combination of machine learning techniques to accurately detect advanced threats that legacy security controls miss.

## Spanning the Advanced Email Attack Landscape




**BEC threats found in corporate inboxes increased 43% from 2021 to 2022**

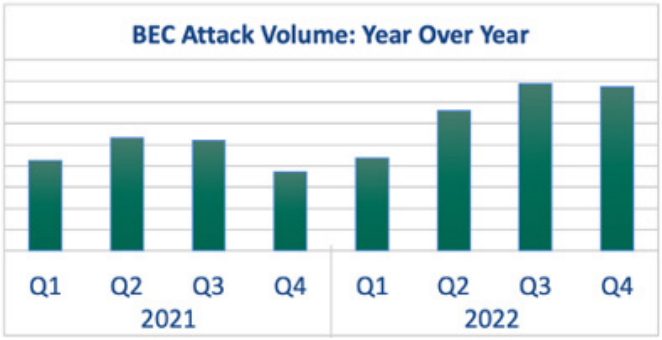



Figure 1: As recently revealed in CyberEdge Group's 2023 Cyberthreat Defense Report, BEC attack volume has risen the last three quarters year over year from 2021 to 2022.

# Why Is Impersonation Impervious to Traditional Email Blocking Methods?

## Manipulating Identity Markers to Trick Recipients

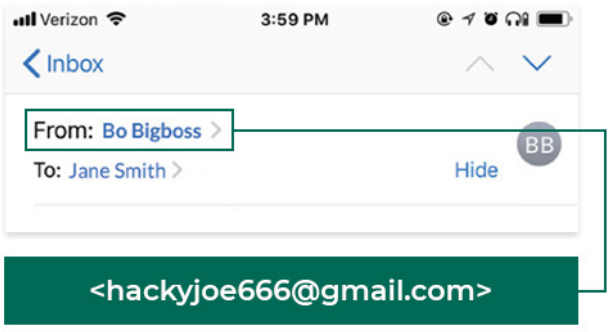
Email impersonation relies on manipulating components of an email message to exactly match or bear similarity to identity markers in a real message sent from a trusted source "From" header, the Subject header, and the body of the message.

Of these, the "From" header is the most commonly recognized marker, as it is displayed prominently in email clients. It is also the identity marker that is most prevalently abused since the sender of a message can specify any value for it. The Subject header and body can contain identity markers, such as words, phrases, brand names, logos, URLs, and narrative structures. These are often secondary to those in the "From" header and primarily serve to support, rather than define, the perceived sending identity for a message.

The "From" header is generally made up of two parts: a display name that is the suggested display label for an email client and an email address, which has a local part and a domain. For example, the "From" header – "Bo Bigboss" <hackyjoe666@gmail.com> – has a display name of "Bo Bigboss," a local part of "hackyjoe666," and a domain of "gmail.com."

Since, as shown below, many email clients show only the display name in certain views, Display Name attacks are the most common form of identity deception. Attackers often insert the identity of a trusted individual (such as the name of an executive of the targeted company) or a trusted brand (such as the name of the bank used by the targeted individual) into the display name. Since common consumer mail services, such as Gmail and Yahoo, allow a user to specify any value in the display name, this type of attack is simple and cheap to stage.

## Display Name Attack Example



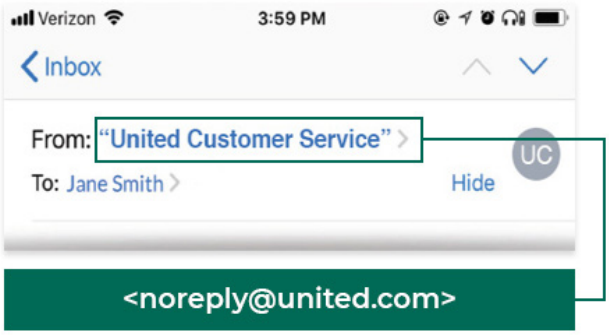

In addition to manipulating the display name, an attacker may also use the actual email address of the impersonated identity in the "From" header, such as "United Customer Service" <noreply@united.com>. This type of attack, known as a Domain Spoofing attack, does not require compromising the account or the servers of the impersonated identity, but instead exploits the security holes in the underlying email protocols. Attackers often use public cloud infrastructure or third-party email sending services that do not verify domain ownership to send such attacks. Email authentication standards, such as Domain-based Message Authentication, Reporting, and Conformance, or DMARC, can be used by a domain owner to prevent spoofing, but unfortunately this is still not widely adopted by popular brands or Fortune 500 companies. In fact, Fortra's latest Email Fraud & Identity Deception Trends: The State of DMARC Enforcement report cited that **2 in 3 Fortune 500 companies** were vulnerable to being impersonated in phishing scams targeting their customers, partners, investors, and consumers.

In fact, even with setting up the various email authentication protocols available—including DMARC, Sender Policy Framework (SPF), and Domain Keys Identified Mail (DKIM)—attackers, or more accurately, untrustworthy sender identities can successfully deceive the intended recipient(s) by registering and using look-alike domains. These often use homoglyphs or characters that appear similar to the original characters in the impersonated domain.

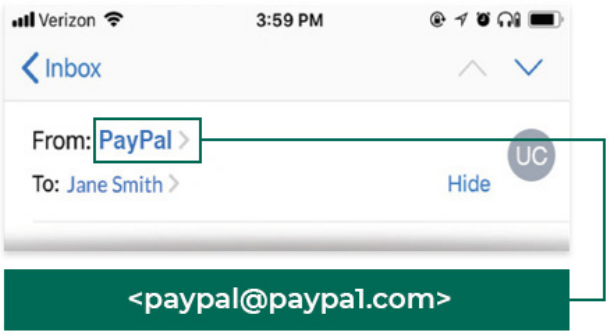
Attackers can use rendering similarities by exploiting specific fonts and styles that are used in popular email clients or by using characters from another script in the Unicode set, such as Cyrillic in the "From" header. Real examples of these are "Dropbox" <notifications@dr0pb0x.com> – where the "o"s in the domain name are actually the Cyrillic character, "O", but an email client will render the version that looks exactly like the impersonated domain. Or, "PayPal" written with the numeral "1" instead of an alpha "l" in the domain of the email like this: paypal@paypa1.com. Another variation of this is when attackers register additional words to the domain name in order to send an advanced fee request or bogus invoice like acme-payments.com, or invoices-acme.com instead of the actual domain: @acmecorporation.com.

## Domain Spoofing Look-alike Domain Attack Examples

### Display Name Attack Example




### Display Name Attack Example




Finally, the most insidious form of identity deception can take place when the attacker has compromised the email account or server of the identity they are impersonating. This type of attack, known as Account Takeover, while low in volume, is generally the hardest to detect since it leverages the identity markers, infrastructure, and many of the behavioral characteristics of legitimate messages coming from that identity.

While the various forms of impersonation attacks may differ in prevalence and sophistication, they have some similarities. First, they manipulate the perception of the recipient, convincing them that the message was sent by an identity with which they are familiar. Secondly, they exploit the recipient's trust in the sender's identity to convince the recipient to take the intended action or disclose information they inadvertently assume is safe. Security awareness training and phishing simulations can help a recipient detect some of these attacks, but the burden of detection can't fall only on the individual as the quality and volume of identity deception attacks increase. Even the most savvy users experience moments of weakness and make mistakes that can lead to major security incidents.

## Data Science: Behind-the-Scenes, Yet Ahead of the Curve

The advanced data science behind Fortra's Cloud Email Protection is an advanced system of machine learning (ML) models that work together to accurately detect impersonation and social engineering techniques used in messages. Fortra's Data Science team is comprised of data scientists with extensive experience in practical applications of modern ML and AI technologies. Also, by partnering with Fortra's storied roster of email impersonation and threat intelligence researchers spanning Agari, Clearswift, PhishLabs, and other Email Security solutions, our customers uniquely benefit from a comprehensive wealth of email security domain expertise.

This robust team has developed high-performing models that consider parts of messages and contextual data both individually and collectively. These models are combined using ensemble learning techniques that relate them to one another to consider all threat characteristics and patterns. Together, they maximize decision confidence and ensure accuracy. To best explain how the data science works, it is important to break down the parts of a message that are analyzed and how the machine learning models are applied.

## The Parts & Parcel of Email Parsing: Email Header Data

Here is a breakdown of the various components embedded in email header data:

**From**

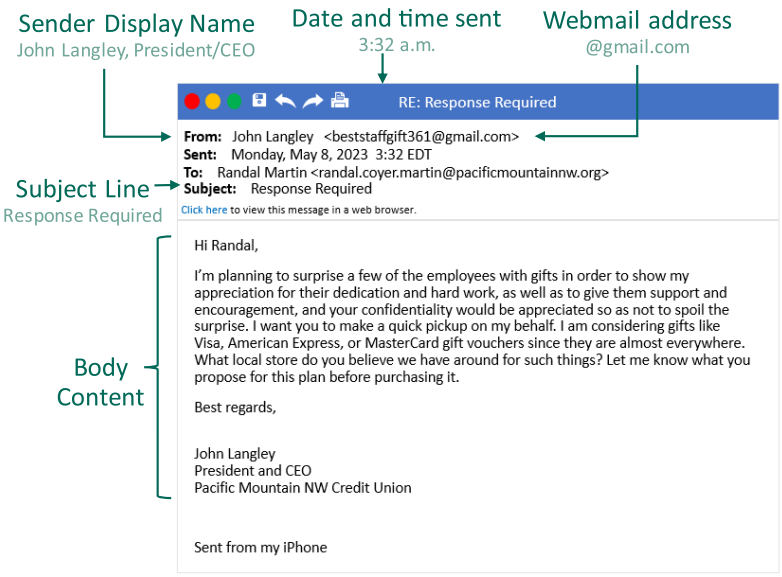
The "From" field, or Sender Display Name, is who the message is sent from [header.from]. This field can be easily forged and is why we suggest ensuring you only open email that has been authenticated by SPF and DKIM.

**To**

This displays who the message is addressed to, but may not contain the recipient's email address.

**Subject**

The subject line is generally the high-level topic of the message being sent as created by the sender.




**Local Part**

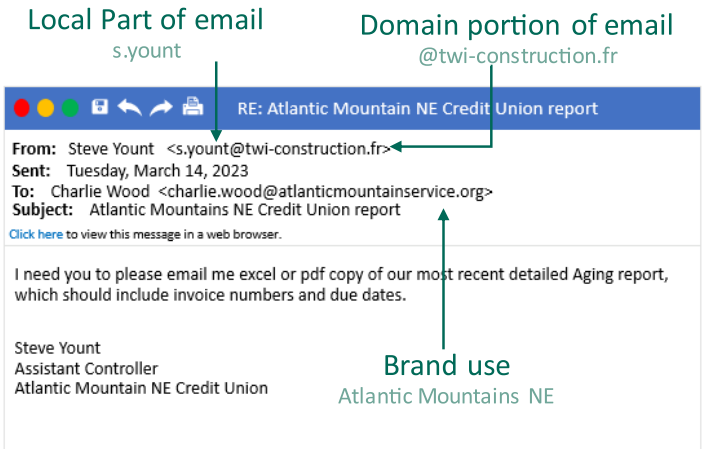
Also called the email prefix, it is the unique identifier or username that comes before the @ symbol in an email address (i.e., the person or entity within an organization that the email is sent directly to).

**Email Domain**

The email address that follows in brackets after the "From" field shows the actual email domain from which the email was sent.

**Brand Use**

This is a technique used by threat actors that includes the sending company's name with the goal of legitimizing the email communication.




In addition to the header data, the machine learning models also consider other email characteristics and data, including, but not limited to:

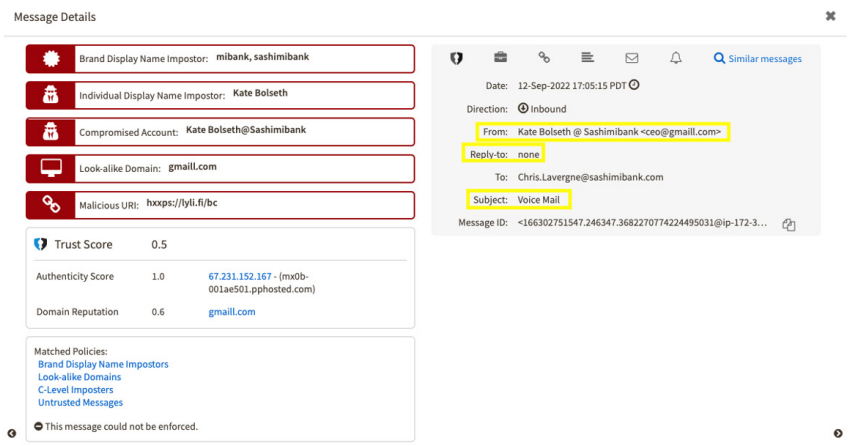
- The sender's infrastructure
- The number of days IP address has been used to send from on behalf of domain
- The number of emails sent using the same "Local Part" of email
- The number of emails using the same Display Name
- Intent of the email (examine nature of content, like Subject Line)
- Matching Address Group and Display Name
- Character text used in the Local Part of email (Latin vs. Cyrillic)
- SPF/DKIM/DMARC records
- And others

Once all of this data is evaluated and scored, Cloud Email Protection produces Reputation and Authenticity scores.

This basically determines if the email was part of a previously unidentified targeted attack, or if it was simply an anomaly. For example, if there was an email that received a very low Authenticity score but a high Reputational score, the models would suggest that there was likely a spoof of a highly reputed domain. The models will then output a "final" Trust Score based on the unique combination between the two model scores on a scale of 0 – 10–where generally a score of 0 -1 is untrustworthy; a 1.1 - 5 rating is suspicioius, and a >5 rating signifies a trusted communication.

**Real Example**

This malicious email attempts to impersonate a CEO. While the sender's name and the company name look legitimate, if you look at the fields highlighted in yellow there are indicators of impersonation, specifically:




1) The "From" address which has <ceo@gmaill.com> with Gmail's sending domain name being misspelled, which is actually a look-alike domain;

2) The "From" address also has "Kate Bolseth" as the sender identity from Sashimibank, however she is the CEO of Fortra, not Sashimibank;

3) The Reply-To address of "none";

4) The subject line of "Voice Mail", representing a fraudulent vishing attempt employing a very generic subject.

This message passed SPF authentication due to relaxed SPF alignment in the domain's DMARC policy. However, this message was given an overall trust score of 0.5 by Cloud Email Protection because it has matched multiple policies associated with impersonation at the bottom, including Brand Display Name Imposters (Sashimibank), Look-alike Domains (gmaill), C-Level Imposters (i.e. Kate), and thus, a label of "Untrusted Message".

## Applying Machine Learning to Message Data

Fortra's Cloud Email Protection combines a multitude of models that interpret, analyze, and assign individual scores to each message component. For example, the table below lists several of these models and the parts that they score:

It's important to understand how the models behind Cloud Email Protection data science work. Three machine learning paradigms are used to parse and analyze inbound email data:

Figure 2: *All* above in the Parts column indicates that the scores are aggregated to determine attacks like Webmail scams, targeted email campaigns, and combine them to assign an overall risk score.

| Model Repository | | |
|---|---|---|
| **Parts** | **Attack Type** | **Model** |
| Domain | Malicious domain | Domain Reputation |
| Sending Infrastructure | Sender impersonation | Authenticity |
| Full header from, subject line | Brand imposter | BDNI |
| Full header from, address groups | Individual imposter | IDNI |
| Subject line | Approach attempt, scam | Subject Line (types, suspiciousness) |
| URLs | Malicious URLs | URL Classifier |
| *All* | Webmail scams | Webmail Catcher |
| *All* | Multiple emails, one actor | Campaign Detection |
| *All* | - | Overall Risk Score |
| *All* | Services | Service Abuse Detector (SAD) |
| *All* | Spam | Spam Account Protector (SAP) |

**1) Feature-Engineered**

In feature-engineered machine learning models, domain experts define the individual measurable properties (e.g., the features) of input data that are instrumental in making predictions or performing tasks. These models excel in scenarios, like email security, where domain expertise plays a vital role in understanding the problem. Also, expert-designed features provide greater interpretability, meaning it can be easier to analyze and improve the model's performance. With deep expertise in email trust as co-founders of the DMARC standard for email authentication back in 2012, Fortra's experts have applied their combined decades of knowledge and these tried and true models to create hundreds of features that are considered by the feature-engineered models within Cloud Email Protection.

**2) Neural Networks**

Neural networks are used to automatically learn complex patterns and features from data. This allows them to handle a greater diversity of data, which is useful in areas like email security where a broad range of file types, images, text, etc. are encountered. In some scenarios, they can capture subtle patterns and nuances in the data that would be missed by feature-engineered models.

**3) Large Language Models**

Large Language Models (LLMs) are designed to understand language in a human-like fashion. They are trained on massive amounts of text data to learn patterns, relationships, and contextual information in language, and then making judgments about them – this makes them ideal for natural language processing (NLP) tasks.

In practice, Fortra's Data Science team uses a pre-train and fine-tune paradigm, adapting pre-trained models to various downstream tasks, such as:

- Analyzing specific text-based email components, such as subject lines and body content
- Task-specific functions, such as determining message type or level of suspiciousness
- Identifying groupings or clusters in email data
- Creating data lakes with labeled data that can be used by other models
- Ingesting feeds of useful data–like lists of suspicious domains and IP address groups

A good illustration of Cloud Email Protection employing LLMs is the analysis of email subject line intent. Instead of parsing features when an email is received, the subject line is encoded and then analyzed to determine the type of subject line and determine its level of suspiciousness. The Subject Type model ascertains what the language resembles the most (e.g., an automated service notification, a marketing email, a scam, an approach attempt, etc.). Then, the Subject Suspiciousness model determines if the subject phrase is suspicious, unsuspicious, or perhaps just overly aggressive in its marketing message or jargon. The outputs of these two models are then analyzed by an ensemble model that considers both determinations to provide an overall measurement of the subject line's intent.

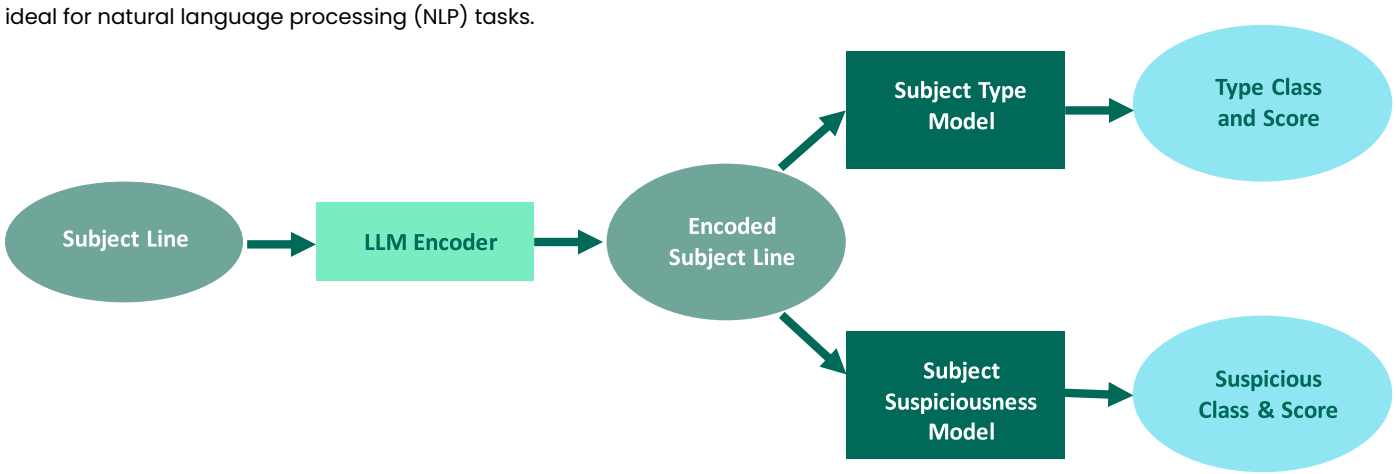



Figure 3: A visualization of the Subject Line attribute using LLM pre-training to assign a score based on its Type/Category & Suspiciousness levels.

Every machine learning paradigm has strengths and weaknesses. Also, the science of machine learning continues to advance and create opportunities to perform tasks more accurately and efficiently. For these reasons, it is critical to understand the task objective and select the ideal machine learning paradigm to perform the task well. This is especially true for applications of machine learning that can impact the operation of business processes, which could potentially bottleneck the flow of inbound and outbound communications.

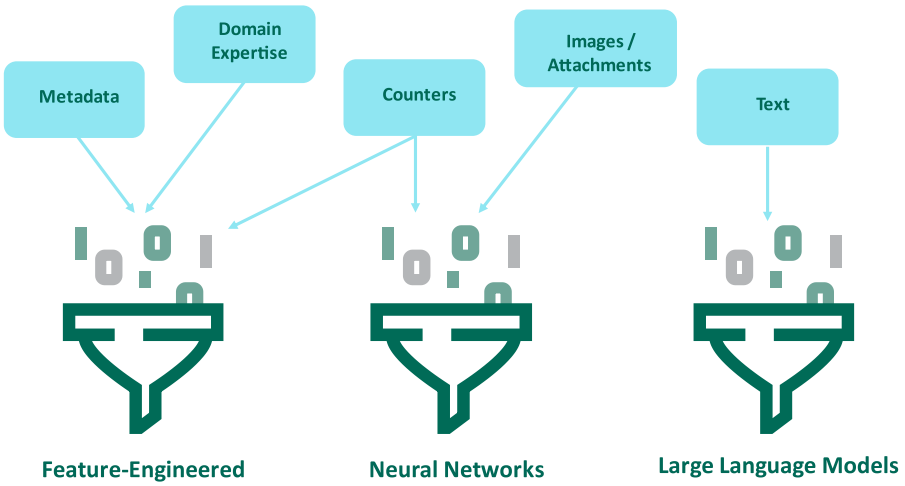
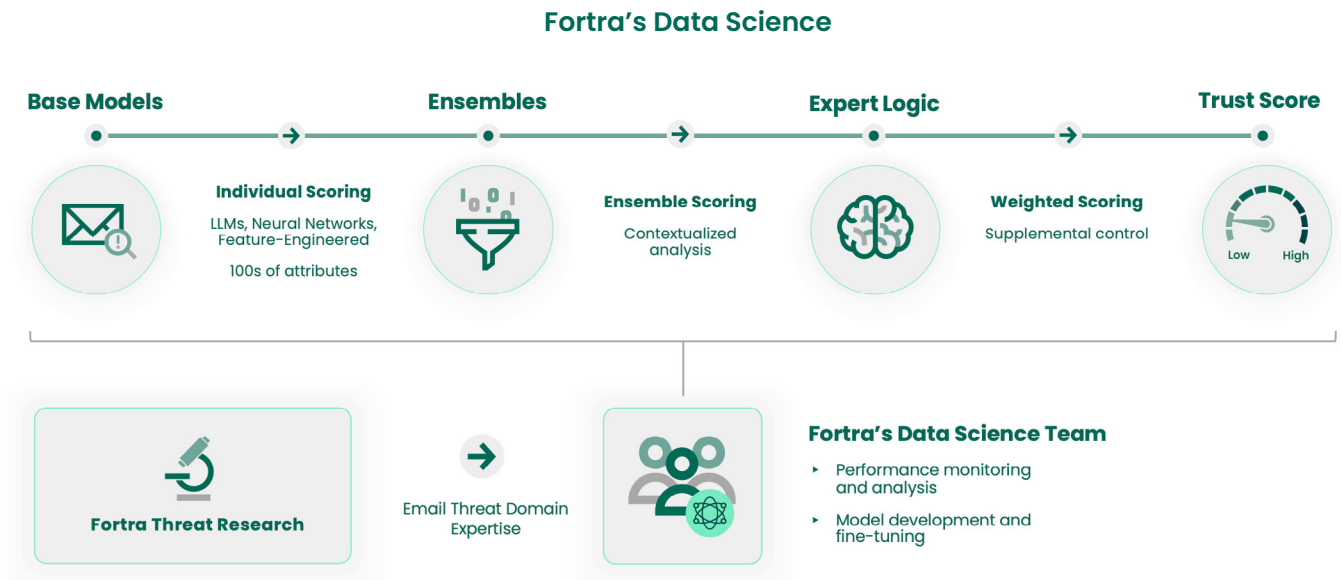



Figure 4: Each ML paradigm has strengths and weaknesses. It is important to utilize the ML paradigm that is best suited for the task.

## Does Logic Come into Play?

AI and ML technologies, including those used in Cloud Email Protection, are a major leap forward. They provide unique insights into data and are a powerful way to find threats that would otherwise go undetected. That said, no artificial intelligence technology gets it right 100% of the time. Even the most sophisticated AI technologies require logic that supplements and provides boundaries. This is especially true for applications in business-critical production systems like email.

This is where expert logic from Fortra's Data Science team comes into play. Our data scientists have a deep understanding of how Fortra's models operate individually, and as parts of the whole. They also constantly monitor and evaluate performance. This team develops and maintains a control layer of expert logic that supplements Fortra's machine learning models while ensuring optimum performance.



Customers of Fortra's Cloud Email Protection can also take advantage of the service's Continuous Detection and Response (CDR) module to apply logic to messages in parallel to machine learning processing. CDR enables several use cases, such as:

- Automatic scanning for threat indicators (from Fortra's Threat Intelligence and 3rd party sources)
- Configuring granular policies for email threat response, such as automatic inbox search and quarantine
- Setting custom policies unique to the customer's email environment

It is also important to share that the architecture of Cloud Email Protection is purposefully designed to be responsive to feedback from internal data science assessments, emerging threat research, and customer input. Multiple modules and points of control provide flexibility in how feedback is incorporated and make it possible to pragmatically fine-tune performance.

## Conclusion

AI is no longer the future of email security; it is the present. Traditional email defenses struggle to detect impersonation and social engineering, allowing BEC and credential theft attacks to reach user inboxes at an alarming rate. Stopping these threats at enterprise scale demands the use of data science to determine whether messages should be trusted. But accomplishing this requires exceptional expertise in modern email threats and data science.

Fortra is at the forefront of this endeavor, driving state-of-the-art data science with cutting-edge research on emerging email threats. We know how modern email threats work inside and out, and we use that insight to design robust models that can detect these threats in the most challenging enterprise environments. With Fortra's Cloud Email Protection, organizations can finally put a stop to known and unknown advanced email threats.

**REQUEST A DEMO**

FORTRA®

Fortra.com

**About Fortra**

Fortra is a cybersecurity company like no other. We're creating a simpler, stronger future for our customers. Our trusted experts and portfolio of integrated, scalable solutions bring balance and control to organizations around the world. We're the positive changemakers and your relentless ally to provide peace of mind through every step of your cybersecurity journey. Learn more at fortra.com.