Computers & Security 158 (2025) 104630

Contents lists available at ScienceDirect

Computers & Security

journal homepage: www.elsevier.com/locate/cose

Business email compromise: A systematic review of understanding, detection, and challenges

Amirah Almutairi [a,b,*], BooJoong Kang [b], Nawfal Alhashimy [b]

[a] Department of Computer Science, Shaqra University, Shaqra, Riyadh, 11961, Riyadh Region, Saudi Arabia
[b] School of Electronics and Computer Science, The University of Southampton, Highfield Campus, Southampton, SO17 1BJ, Hampshire, United Kingdom

ARTICLE INFO

Keywords:
Business Email Compromise (BEC)
Systematic literature review
Email threat
BEC detection

ABSTRACT

Business Email Compromise (BEC) is a widespread fraud targeting businesses and individuals to obtain financial benefits and gain access to highly sensitive data. BEC fraud significantly impacts almost all organizations worldwide, resulting in substantial losses. Despite its prevalence, there is a shortage of research on understanding and protecting against this fraud. Consequently, this paper aims to survey existing BEC detection techniques. It first provides an overview of the methods and strategies used by attackers in BEC schemes. It also reviews existing BEC detection and prevention techniques, including both technical and non-technical solutions. The strengths of each technique are objectively discussed, and their limitations are critically analyzed. Finally, this study offers a thorough set of current challenges in BEC detection and outlines future research directions, providing valuable guidance for improving security measures against BEC fraud.

Contents

* Corresponding author at: School of Electronics and Computer Science, The University of Southampton, Highfield Campus, Southampton, SO17 1BJ, Hampshire, United Kingdom.
E-mail addresses: amirah@su.edu.sa (A. Almutairi), b.kang@soton.ac.uk (B. Kang), nawfal@soton.ac.uk (N. Alhashimy).

https://doi.org/10.1016/j.cose.2025.104630
Received 14 January 2025; Received in revised form 6 June 2025; Accepted 7 August 2025
Available online 20 August 2025

1. Introduction

Despite its significant financial impact and escalating frequency, Business Email Compromise (BEC) remains underexplored in cybersecurity research. BEC is a specialized form of email fraud in which attackers pose as trusted entities — such as executives, suppliers, or clients — to manipulate victims into making unauthorized payments or revealing sensitive information. According to the Federal Bureau of Investigation (FBI) [FBI2023], global financial losses attributed to BEC have surpassed USD $8 billion from 2021 to March 2023 alone. As shown in Fig. 1, these losses — second only to investment-related cybercrimes — have risen yearly.

The complexity and rapid evolution of BEC methods, particularly spear-phishing and executive spoofing, make them challenging for existing detection and prevention tools, which often focus on general phishing threats. Consequently, many nuanced facets of BEC — such as its reliance on social engineering rather than overt malicious attachments — remain insufficiently addressed. This gap hinders the development of targeted, high-accuracy countermeasures tailored to BEC's unique characteristics.

Business Email Compromise (BEC) continues to rank among the most financially damaging and operationally sophisticated forms of cybercrime. Reports from Microsoft, IBM, and the UK's National Cyber Security Centre (NCSC) consistently identify BEC as a top-tier threat due to its reliance on targeted deception rather than technical exploits (Microsoft Threat Intelligence, 2023; I.B.M. Security, 2023a; National Cyber Security Centre (NCSC), 2023). As cybercriminals continue to refine their tactics, it is imperative to develop advanced detection strategies that are specifically tailored to the unique characteristics of BEC fraud.

1.1. Motivation

Despite the enormous economic losses and the wide range of targeted companies, research on protecting against this fraud remains limited compared to other cybersecurity threats. While several studies have explored phishing and general email fraud, fewer have specifically focused on BEC's unique characteristics. This gap is particularly evident in the lack of comprehensive evaluations of BEC-specific detection techniques, as noted by Almutairi et al. (2024), Mansfield-Devine (2016), Zweighaft (2017).

In contrast to prior reviews, which either focus narrowly on regional case studies (Papathanasiou et al., 2024) or treat BEC as a subset of generic phishing (Atlam and Oluwatimilehin, 2022), this study provides a task-focused synthesis of BEC detection methods, encompassing both technical (e.g., rule-based filters, ML-driven classifiers) and non-technical (e.g., training, policy-level) approaches. We further introduce a structured taxonomy that categorizes countermeasures based on their operational scope, whether algorithmic or organizational, offering a more comprehensive and application-oriented overview than previous work. As a result, it supports and encourages ongoing efforts to develop more robust and adaptable solutions.

Compared to prior literature reviews that treat BEC either peripherally as a subtype of phishing or focus narrowly on local contexts, this study takes a task-centric and threat-informed approach. It uniquely maps BEC-specific attack strategies to detection methods, incorporating both technical and organizational layers. By aligning our taxonomy with real-world tactics and gaps observed in case evidence, this review provide a structured synthesis grounded in operational failure points. This approach offers practical relevance for security practitioners and fills a key gap overlooked by reviews that generalize across broader email threats without addressing BEC's evolving deception strategies or deployment challenges.

1.2. Research question and objectives

Research Question: What are the effective preventive techniques for Business Email Compromise (BEC) fraud in the current literature?

This paper aims *to systematically review and analyze the existing literature on Business Email Compromise (BEC) fraud*, with a particular focus on detection strategies, prevention methods, and associated research challenges. The review is guided by the following objectives:

- **Objective 1** Consolidate existing knowledge on BEC fraud: Summarize and synthesize definitions, characteristics, and operational mechanisms of Business Email Compromise as documented in the literature.
- **Objective 2** Identify detection and prevention techniques: Evaluate the range of technical and organizational strategies used to detect and prevent BEC attacks, highlighting strengths, limitations, and deployment challenges.
- **Objective 3** Review datasets used in BEC research: Examine the types, sources, and quality of datasets employed in existing studies, including discussions on availability, realism, and representativeness.
- **Objective 4** Identify research gaps and future directions: Analyze the limitations of current approaches and propose directions for future work to advance the field of BEC detection and mitigation.

Main contribution

The primary contribution of this review is a comprehensive synthesis of peer-reviewed research on Business Email Compromise (BEC) detection and prevention. The paper offers a consolidated reference

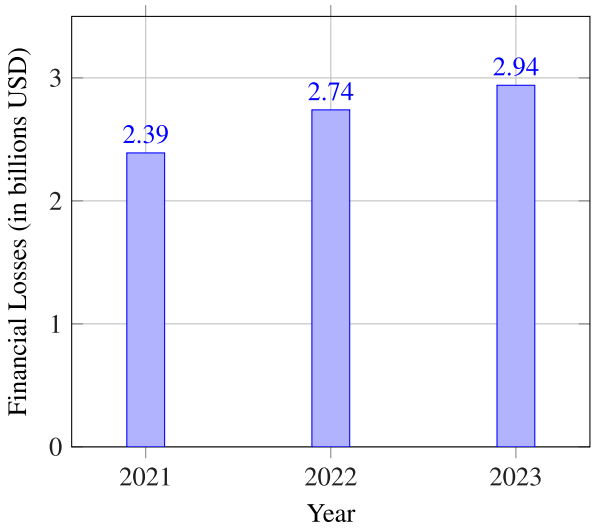

Fig. 1. Financial losses due to BEC from 2021 to 2023.

point for academics and practitioners by identifying current solutions, unresolved challenges, and future opportunities in applying AI and NLP techniques to this under explored threat domain. The contributions can be summarized as follows:

- A structured overview and classification of BEC detection approaches, including traditional rule-based techniques, machine learning models, and NLP-based solutions. This includes comparison across input features, detection strategies, model types, and evaluation setups.
- A critical assessment of 30 peer-reviewed studies published between 2007 and 2024, covering diverse methods and application contexts in the cybersecurity and email fraud detection landscape.
- A conceptual taxonomy that organizes BEC attack strategies and mitigation approaches across both technical and organizational layers, supporting structured analysis of detection failures and response gaps.
- A comparative evaluation of performance metrics, dataset usage, and deployment considerations, with attention to scalability, adversarial robustness, and limitations in current evaluation practices.
- A summary of open challenges and under-addressed research questions, along with a future research agenda that includes directions for robust feature design, dataset generalizability, and hybrid AI–human workflows.

1.3. Organization and reading map

This study is organized to provide a structured and comprehensive exploration of Business Email Compromise (BEC) fraud. The paper is divided into the following sections:

Section 1 introduces the study by outlining its objectives, scope, and relevance within the broader context of cybersecurity, with a particular focus on the critical threat posed by BEC fraud. This section sets the stage for the subsequent review of the literature.

Section 2 builds upon the introduction by critically examining existing research on BEC fraud, highlighting their contributions and identifying gaps that this study aims to address.

Section 3 describes the overarching methodology employed, explaining how the study's structure aligns with its research objectives. It provides a foundation for understanding the systematic review and subsequent analyses.

Section 4 uses the findings from the systematic review to present an in-depth discussion of BEC fraud, exploring its mechanisms, operational methods, and global impact.

Section 5 builds on the understanding of BEC mechanisms by analyzing state-of-the-art detection techniques and countermeasures. This section critically evaluates existing methods and highlights areas for improvement.

Section 6 complements the discussion of detection techniques by describing the dataset used in this study. It provides detailed information about the dataset's composition and relevance to the analysis presented in earlier sections.

Section 7 synthesizes insights from the preceding sections to discuss the current and emerging challenges in combating BEC fraud. This section bridges the gap between the limitations of existing solutions and the need for innovative strategies.

Section 8 builds upon the identified challenges to propose actionable recommendations for future research. It emphasizes the importance of advancing detection and mitigation strategies to address the evolving nature of BEC fraud.

Section 9 concludes the paper by summarizing the key findings from all sections and discussing their implications for both academia and industry. It reinforces the study's contribution to the field and outlines opportunities for further exploration.

In this survey, we examine the full range of BEC detection approaches found in the literature, including global, regional, and industry-specific methods. Having introduced the foundational context and objectives of this study, the next section examines existing literature on Business Email Compromise (BEC), identifying critical gaps and guiding our investigation.

2. Related work

Given the rapid evolution and adaptive nature of Business Email Compromise (BEC) threats, it is essential to continually consolidate knowledge through detailed systematic reviews. In today's fast-changing cyber landscape, relying on outdated or fragmented analyses risks overlooking emerging attacker methodologies, novel detection technologies, and critical non-technical factors that influence the effectiveness of BEC countermeasures. Although previous systematic reviews have provided valuable foundational insights, they often offer a narrow focus that limits their broader applicability.

Several existing reviews concentrate on specific contexts or technical aspects. For example, Papathanasiou et al. (2024) primarily examine cybersecurity practices within Greece and the EU regulatory framework. While their findings are insightful, the region-specific focus restricts the generalizability of their conclusions to other geographical or regulatory contexts. Similarly, Ogwo-Ude (2023) analyzes medium and large-scale firms in the USA, addressing financial and regulatory challenges that are particular to that environment. This focus, however, tends to underrepresent smaller enterprises and diverse cultural settings, which may face unique BEC threats and require different mitigation strategies.

Furthermore, Atlam and Oluwatimilehin (2022) provide a detailed review of machine learning (ML)-based detection methods, including analyses of algorithms, detection features, and datasets. Their work, however, concentrates exclusively on technical ML methodologies and does not adequately address the broader non-technical dimensions — such as employee training, awareness initiatives, organizational policies, and human verification practices — that are crucial for a comprehensive understanding of BEC detection. This omission contributes to a fragmented picture of the current state of BEC research.

To overcome these limitations, our systematic review diverges from prior studies by:

- Providing comprehensive coverage of both technical approaches (e.g., machine learning, NLP) and non-technical approaches (e.g., training, organizational policies) in BEC detection and prevention.

Table 1
Limitations of recent review papers on business email compromise.

Review article	Focus and findings
Papathanasiou et al. (2024)	**Focus:** BEC fraud in Greece within EU regulatory frameworks. **Findings:** Emphasizes regional cybersecurity countermeasures and recommendations. **Limitations:** Region-specific insights limit global applicability.
Ogwo-Ude (2023)	**Focus:** Financial and regulatory impacts on SMEs in the USA. **Findings:** Discusses tailored mitigation strategies for SMEs. **Limitations:** Limited applicability outside SMEs and the US context.
Atlam and Oluwatimilehin (2022)	**Focus:** ML-based phishing and spear-phishing detection methods. **Findings:** Comprehensive technical overview of detection techniques. **Limitations:** Omits detailed exploration of non-technical factors essential for targeted BEC detection.

- Proposing a structured conceptual framework that categorizes BEC threats, detection techniques, and research gaps based on a critical evaluation and synthesis of the current literature.

Table 1 summarizes recent related studies, explicitly highlighting their specific limitations and illustrating how our work addresses these identified gaps.

Thus, explicitly consolidating existing knowledge into an updated, comprehensive framework equips researchers and industry practitioners with nuanced, actionable insights, clearly supporting ongoing and future research initiatives. In contrast to the descriptive tendencies of previous reviews, our synthesis is informed by analytical principles from threat modeling and studies on adversarial behavior. Rather than categorizing methods at face value, we evaluate detection strategies through the lens of their alignment with attacker objectives, operational constraints, and real-world failure points. This enables a more critical and comparative synthesis that not only catalogues existing work but also exposes tensions, contradictions, and blind spots that undermine BEC countermeasures in practice. In the next section, we outline the systematic methodology we employed to fill these research gaps.

3. Methodology

This systematic literature review (SLR) follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework and established guidelines for systematic reviews in cybersecurity research (Kitchenham and Charters, 2007). This approach ensures a rigorous, transparent, and replicable process for identifying, selecting, and analyzing relevant studies.

The methodology, illustrated in Fig. 2, is further detailed in the following sections.

3.1. Search strategy and selection process

In line with PRISMA guidelines and established protocols for systematic reviews in cybersecurity and software engineering (Kitchenham and Charters, 2007), we conducted a comprehensive literature search across leading digital libraries, including *Web of Science, Scopus, IEEE Xplore, ACM Digital Library, SpringerLink, ScienceDirect (Elsevier), and Taylor & Francis*. The search covered publications from 2007 through February 2024.

A Boolean search query was constructed to maximize coverage and retrieval efficiency. The final query combined core terms related to

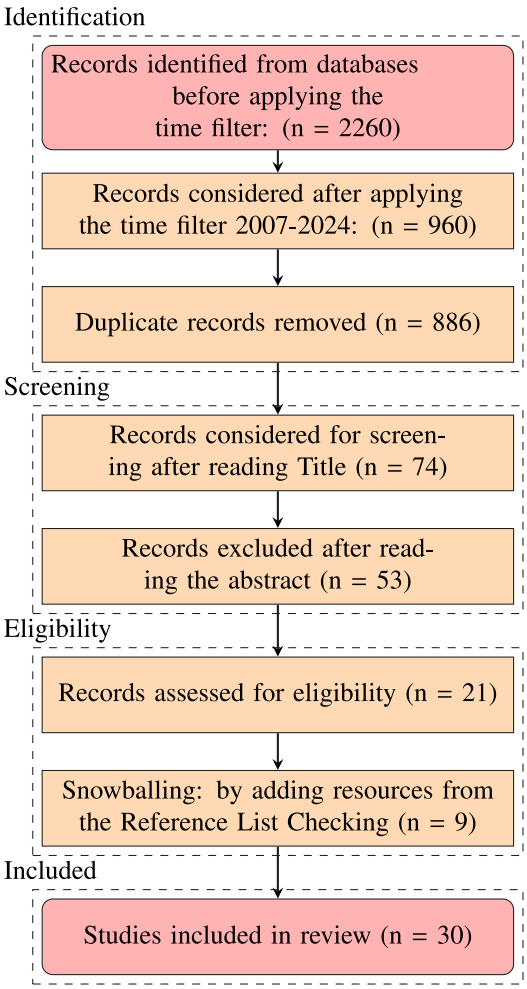

Fig. 2. PRISMA workflow SLR Methodology.

Business Email Compromise (BEC) and its detection and prevention, as follows:

(‘‘Business Email Compromise’’ OR ‘‘BEC’’ OR ‘‘email fraud’’ OR ‘‘CEO fraud’’ OR ‘‘impersonation fraud’’) AND

(‘‘detection’’ OR ‘‘prevention’’ OR ‘‘machine learning’’ OR ‘‘NLP’’ OR ‘‘social engineering’’ OR ‘‘cybersecurity’’)

Only peer-reviewed journal articles and conference proceedings written in English were considered. Preprints and grey literature were excluded not only to ensure peer-reviewed methodological rigor but also to mitigate risks of overrepresenting speculative or preliminary findings in a rapidly evolving threat landscape. Given the legal and organizational implications of BEC countermeasures, prioritizing validated sources enhances the practical relevance of our synthesis. Future reviews may revisit this choice as the field matures.

3.1.1. Screening and corpus selection

The initial search yielded over 2,000 records across all databases. These were first de-duplicated using EndNote (version 20), after which a systematic two-stage screening protocol was applied. In the first stage, two researchers independently reviewed the titles and abstracts using predefined inclusion and exclusion criteria. In the second stage, full-text screening was conducted to assess eligibility more thoroughly. Any discrepancies at either stage were resolved through discussion or, if needed, third-party arbitration.

This screening process resulted in a final corpus of 30 peer-reviewed studies. The selected studies span a range of methodological paradigms — including rule-based approaches, machine learning models, and hybrid detection frameworks — as well as varied organizational contexts and evaluation strategies. This curated set provides a diverse yet methodologically coherent basis for in-depth synthesis of BEC-specific detection and prevention techniques. Generalizability considerations and limitations are further addressed in the Discussion section.

3.2. Study selection criteria

The study selection process was conducted in two stages: (i) initial screening of titles and abstracts, followed by (ii) full-text review. To validate consistency, two reviewers independently screened a random subset of articles ($n = 30$), resolving disagreements via consensus or third-party arbitration.

Inclusion Criteria. Studies were included if they:

- Focused specifically on Business Email Compromise (BEC) detection, prevention, or threat analysis.
- Employed a formal methodology or technical implementation (e.g., experimental study, case analysis, or algorithmic evaluation).
- Reported quantitative evaluation metrics (e.g., accuracy, precision, recall, F1-score, or false-positive rate).
- Were published in peer-reviewed journals or conference proceedings.
- Included empirical validation using real-world or publicly available datasets.

Exclusion Criteria. Studies were excluded if they:

- Addressed general phishing or email security without explicit reference to BEC-specific attack vectors or defences.
- Lacked empirical validation or performance evaluation.
- Were not published in English.
- Presented narrative reviews or opinion pieces without original technical contributions or systematic synthesis.

3.3. Quality assessment and consistency check

To ensure methodological rigor and the credibility of included studies, A predefined set of quality assessment criteria was used to evaluate each study:

- Clarity of research objectives and hypotheses.
- Methodological transparency, including dataset description and reproducibility.
- Completeness of experimental setup, evaluation methods, and performance metrics.
- Relevance to BEC detection, prevention, or threat mitigation.
- Peer-reviewed status (journal/conference).

Each selected study was assessed for methodological rigor using a binary scoring system (1 = meets criteria, 0 = does not meet criteria), with a higher total score indicating stronger rigor. To validate the consistency of the quality assessment process, an inter-rater reliability check was conducted. Two independent reviewers evaluated a randomly selected subset of 15 studies. Cohen's kappa (κ) was calculated, resulting in $\kappa = 0.89$, indicating a high level of agreement. All studies scoring below 3 out of 5 were flagged during the synthesis stage. These studies were included for completeness but treated with interpretive caution—particularly when drawing conclusions about methodological effectiveness or empirical claims. During synthesis, lower-quality studies were not excluded outright but were deprioritized in drawing thematic inferences, especially when evidence was contradictory. For example, studies with minimal dataset transparency or inconsistent metric reporting were used cautiously in performance discussions and omitted from informing taxonomy criteria.

3.4. Data extraction and analysis techniques

Data extraction was performed independently by two reviewers using a structured extraction form. The extracted data included:

- Study metadata (author, year, publication venue, country).
- Methodology details (detection approach, algorithm, framework, or analytical model used).
- Dataset characteristics (type, size, source, publicly available or proprietary).
- Evaluation metrics (accuracy, precision, recall, F1-score, false-positive rate).
- Reported challenges, limitations, and future research directions.

Data synthesis was performed using thematic analysis (Braun and Clarke, 2006) to systematically identify common patterns, methodologies, and research gaps across the selected studies. Given the significant methodological heterogeneity among the reviewed studies, a meta-analysis was deemed inappropriate. Instead, a narrative synthesis approach was adopted, focusing on qualitative comparison and classification of methodologies.

With our systematic approach established, we now turn to the core phenomenon under study Business Email Compromise highlighting its mechanisms, strategies, and global ramifications.

4. Business email compromise

Business Email Compromise (BEC) is a sophisticated form of cyber-enabled fraud that primarily exploits email-based trust relationships to achieve financial theft or sensitive data exfiltration. Unlike conventional phishing, BEC typically omits malware or suspicious links, relying instead on psychological manipulation, impersonation, and business process exploitation. This section presents a critical synthesis of Business Email Compromise (BEC), including its definitions, global impact, evolving attack strategies, and fraud anatomy, contributing to **Objective 1** by grounding the taxonomy in real-world dynamics.

4.1. Impact of BEC by continent

North America: In North America, particularly in the United States, BEC has resulted in substantial financial losses. According to the FBI's Internet Crime Complaint Centre (IC3), BEC scams led to losses exceeding $1.8 billion in 2020 alone. The increase in remote work during the COVID-19 pandemic has exacerbated the vulnerability of businesses to BEC attacks (Abnormal Security, 2023b; Alder, 2023).

Europe: In Europe, countries like Greece have seen a notable rise in BEC incidents. The Greek landscape of cyberattacks has been extensively analyzed, highlighting the effectiveness of countermeasures in this region as well as the challenges faced. The European Union's NIS Directives play a crucial role in shaping the regulatory framework to combat such cybercrimes (Bitdefender, 2023).

Asia: Asian countries have also been significantly affected by BEC. For instance, in Japan, BEC-related financial losses have increased, emphasizing the need for robust cybersecurity measures. Cybercriminals often exploit vulnerabilities in financial institutions to redirect funds fraudulently (Kelly, 2023).

Australia: Australia faces a growing threat from BEC, with cybercriminals targeting both large corporations and small businesses. The Australian Cyber Security Centre (ACSC) has reported an increase in BEC attacks, resulting in substantial financial losses for businesses across the country (Alder, 2023; Kelly, 2023).

Africa: Nigeria is a known hotspot for BEC operations, with numerous international fraud rings operating from the country. Despite efforts to crack down on these activities, Nigerian cybercriminals continue to perpetrate BEC scams globally, leading to significant financial damage (Abnormal Security, 2023b; Alder, 2023).

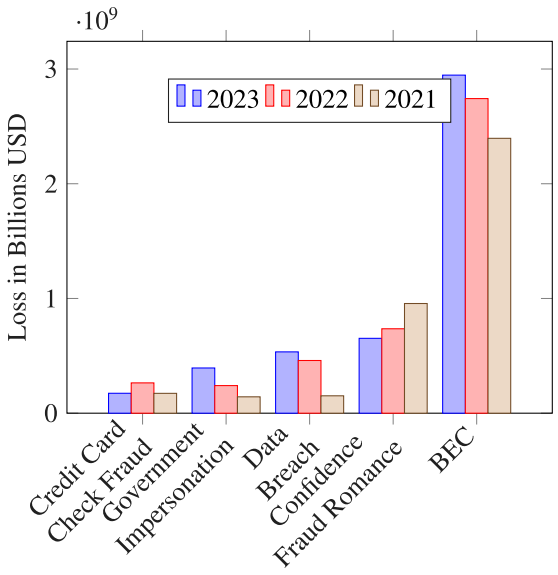

Fig. 3. Top 5 complaint loss comparison data (2021–2023).
Note: The financial loss for botnet attacks in 2021 is roughly estimated based on the growth rate observed between 2022 and 2023.

South America: Countries in South America are not immune to BEC. Businesses in Brazil and Argentina have reported increasing incidents of BEC, with cybercriminals targeting sectors ranging from finance to manufacturing. The lack of stringent cybersecurity measures in some regions makes these countries particularly vulnerable (Abnormal Security, 2023b).

4.2. Statistics and trends

In 2023, pretexting — including Business Email Compromise (BEC) — surpassed traditional phishing as the most prevalent social engineering tactic, with BEC fraud accounting for over 50% of such incidents. Analysis of complaint losses over the past three years for the top five complaint types, as shown in Fig. 3, reveals distinct trends and critical areas of concern. The median open rate for text-based BEC fraud is nearly 28%, and BEC served as the attack vector for 9% of data breaches in 2023. These statistics underscore the significant and expanding impact of BEC fraud, highlighting the urgent need for enhanced security measures and increased awareness. This shift is corroborated by the 2023 Verizon Data Breach Investigations Report (DBIR), which similarly identified pretexting and impersonation as dominant social engineering vectors in enterprise-scale breaches. Additionally, ENISA's Threat Landscape report emphasized that BEC remains a top threat in both public and private sector organizations across the EU.

Furthermore, high-profile cases illustrate the devastating effects of these schemes. For instance, in 2019, Toyota Boshoku Corporation — a subsidiary of Toyota — suffered a loss of approximately $37 million after cybercriminals deceived employees into transferring funds to fraudulent accounts. Similarly, in 2016, a sophisticated BEC scam targeting Facebook and Google resulted in collective losses exceeding $100 million, with attackers impersonating a legitimate vendor. Such cases not only underscore the financial impact but also demonstrate the evolving sophistication of BEC schemes, which exploit trust-based relationships to circumvent conventional security measures. These cases reveal systemic detection gaps—Toyota's breach stemmed from over-reliance on manual payment verification, while the Facebook/Google case exploited implicit trust in vendor processes, bypassing anomaly detection systems that rely on metadata or payload analysis.

A summary of the BEC fraud Objectives discussed in the recent literature are outlined in Table 2.

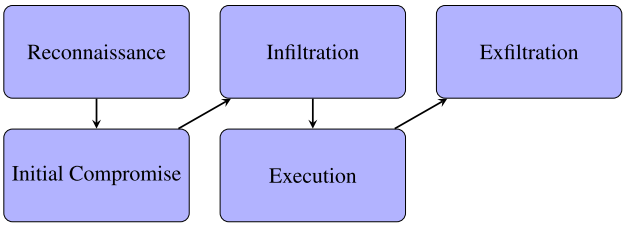

Fig. 4. How a BEC attacker defrauds a victim.

4.3. Anatomy of business email compromise

Following the understanding of BEC in the previous section, this section considers the entire anatomy of BEC fraud.

4.3.1. Stages of a BEC attack

The stages of a BEC attack, illustrating how a BEC attacker defrauds the victim, are presented in Fig. 4 and described as follows:

Reconnaissance: Attackers research their targets to understand the organizational structure, business operations, and key personnel (Saud Al-Musib et al., 2021).

Initial Compromise: Using the gathered information, attackers establish a relationship with individuals managing the organization's financial accounts by impersonating a trusted internal or external source (Federal Bureau of Investigation (FBI), 2017).

Infiltration: Once inside the email system, attackers observe communications, gather additional information, and wait for the opportune moment to execute their plan.

Execution: When ready, attackers use the compromised communication channel to request an urgent and confidential financial transfer to their account. They often present the payment receiver as unreachable due to travel or meetings, making the request appear legitimate and urgent (Microsoft Security, 2017).

Exfiltration: After the victim transfers the money, the attackers quickly move the funds to another account, severing the link to the initial transaction. By the time the fraud is discovered, the attackers have already gained control of the transferred amount.

4.3.2. BEC fraud methods

Business Email Compromise (BEC) is executed through diverse and evolving strategies that rely heavily on impersonation, urgency, and psychological manipulation. Drawing from the FBI's Internet Crime Complaint Center (IC3) (Internet Crime Complaint Center (IC3), 2023), we outline five primary BEC fraud types and analytically map each to its typical detection failure point:

- **Bogus Invoice Schemes**—Attackers impersonate suppliers or vendors and send fraudulent invoices to request payments. These schemes often bypass controls in organizations with frequent invoice workflows.
  Failure Point: Routine-looking emails from familiar names often evade rule-based filters, especially when the attacker uses a compromised or spoofed internal account.
- **CEO Fraud**—Cybercriminals pose as senior executives (e.g., CEO, CFO) and pressure subordinates to perform urgent financial transactions.
  Failure Point: The psychological weight of hierarchical authority discourages employee verification, and urgency cues often camouflage textual anomalies that NLP systems might otherwise detect.
- **Account Compromise**—Legitimate email accounts are hijacked and used to send fraudulent requests, often within existing threads to gain credibility.
  Failure Point: Authenticated messages from compromised accounts bypass anomaly filters unless behavioral baselines or context-aware models are deployed.

Table 2
Summary of BEC fraud objectives..

Source	Description	Objectives
(Zweighaft, 2017)	The attacker poses as a lawyer or representative of the law firm supposedly in charge of the company's legal matters and requests confidential information.	Stealing confidential, private information such as financial records, legal documents, and intellectual property. **Example**: An attacker impersonates the company's legal advisor and requests copies of recent merger and acquisition documents.
Almutairi et al. (2023)	Investigated the sophisticated methods attackers use in BEC to exploit linguistic nuances and social engineering tactics without relying on explicit indicators like links or attachments.	Enhancing detection techniques to identify BEC fraud by analyzing the linguistic properties of emails, thereby preventing the unauthorized acquisition of sensitive information and financial assets. **Example**: Using transformer-based models such as BERT and BiLSTM to detect subtle linguistic cues in email text, improving detection accuracy and reducing false positives.
(King, 2019)	The attacker uses a hacked executive's or employee's email account to make requests that appear legitimate to internal staff.	Financial or confidential information requests that appear to come from within the company, aimed at unauthorized fund transfers or data breaches. **Example**: An attacker uses a compromised CFO's email to instruct the finance department to change the bank account details for the next payroll run.
(Cross and Gillett, 2020)	Corporate fraud involving the identity theft of a senior member of an organization. The attacker sends emails asking for urgent financial transactions or access to confidential documents.	Urgent financial or confidential information requests aimed at diverting company funds or gaining access to sensitive information. **Example**: An email appearing to be from the CEO urgently requests the transfer of $100,000 to a new supplier's account.
(Spangler, 2021)	Detailed the BEC method and strategies employed by attackers to deceive targets into disclosing critical information.	Educating organizations on the various tactics used in BEC scams, helping them develop better preventive measures and response strategies. **Example**: Training sessions simulate BEC scenarios to help employees recognize and respond to suspicious emails effectively.

- **Attorney Impersonation**—Fraudsters pose as legal representatives handling sensitive matters, urging discretion to minimize verification.
  Failure Point: Legal language and confidentiality pretexts reduce suspicion; rule-based and ML-based systems often lack sufficient contextual understanding.
- **Data Theft**—Targeting HR, payroll, or finance personnel, attackers extract sensitive information for future fraud.
  Failure Point: Requests often appear operationally legitimate, and systems without cross-role access profiling fail to flag such data exfiltration attempts.

Each BEC type exploits a different structural vulnerability—whether technological (e.g., email filters), procedural (e.g., lack of verification), or human (e.g., trust in authority). These frauds are rarely detectable using static rules alone.

4.3.3. Composite strategies and tactical adaptation

Modern BEC fraud increasingly blends multiple strategies in a single attack cycle. For instance, attackers may use spear-phishing to gain credentials (technical breach), then engage in CEO fraud (social engineering), while redirecting payments via homograph domain attacks (visual deception).

- **Credential Harvesting:** Through phishing emails or brute-force attacks, attackers gain account access.
- **Social Engineering and Pretexting:** Exploiting urgency and authority to override skepticism.
- **Homograph Domains:** Visually deceptive URLs mislead victims (e.g., "examp1e.com" vs. "example.com").
- **Deepfake Voice or Video Impersonation:** Synthetic media content impersonates executives during high-stakes communications.

Failure Point: Most security tools are designed to detect isolated threats. When strategies are layered and contextually plausible, they bypass segmented defence systems, underscoring the need for unified, adaptive detection frameworks that integrate behavioral analysis and real-time verification.

These analytical mappings clarify how and why BEC strategies persist despite advances in security technologies. They inform the taxonomy in Fig. 5 and motivate the evaluation of technical and non-technical controls in subsequent sections.

As demonstrated, BEC attacks rely less on technical sophistication and more on psychological manipulation, impersonation, and trust exploitation—rendering traditional perimeter-based defenses insufficient. These multifaceted fraud strategies, which often combine AI-generated content, domain spoofing, and behavioral engineering, represent a growing challenge to existing security mechanisms. The severity of financial damage and the attackers' increasing reliance on human error rather than malware demand a paradigm shift in defensive strategies.

In the following section, we critically examine how detection and prevention techniques proposed in the literature align (or fail to align) with these evolving threat strategies. This analysis lays the foundation for evaluating technical and non-technical countermeasures within our taxonomy.

5. Countermeasures against BEC attacks

In this section, we examine how companies and researchers have attempted to combat BEC fraud by proposing and evaluating a range of countermeasures. Specifically, we address **Objective 2** by presenting a comprehensive classification of BEC detection and prevention techniques — both technical and non-technical — identified in the surveyed literature.

Guided by the well-known People–Process–Technology (PPT) triad in security research, we define *technical* controls as technology-centric solutions (e.g., rule-based filters, ML/NLP models, cryptographic schemes) and *non-technical* controls as people- and process-centric measures (e.g., training, human verification, governance policies). This socio-technical framing moves beyond an intuitive split and offers a structured lens for comparing robustness, scalability, and deployment realism across studies.

Table 3 summarizes the technical and non-technical approaches adopted in the reviewed studies. As observed, 23 out of the 30 papers

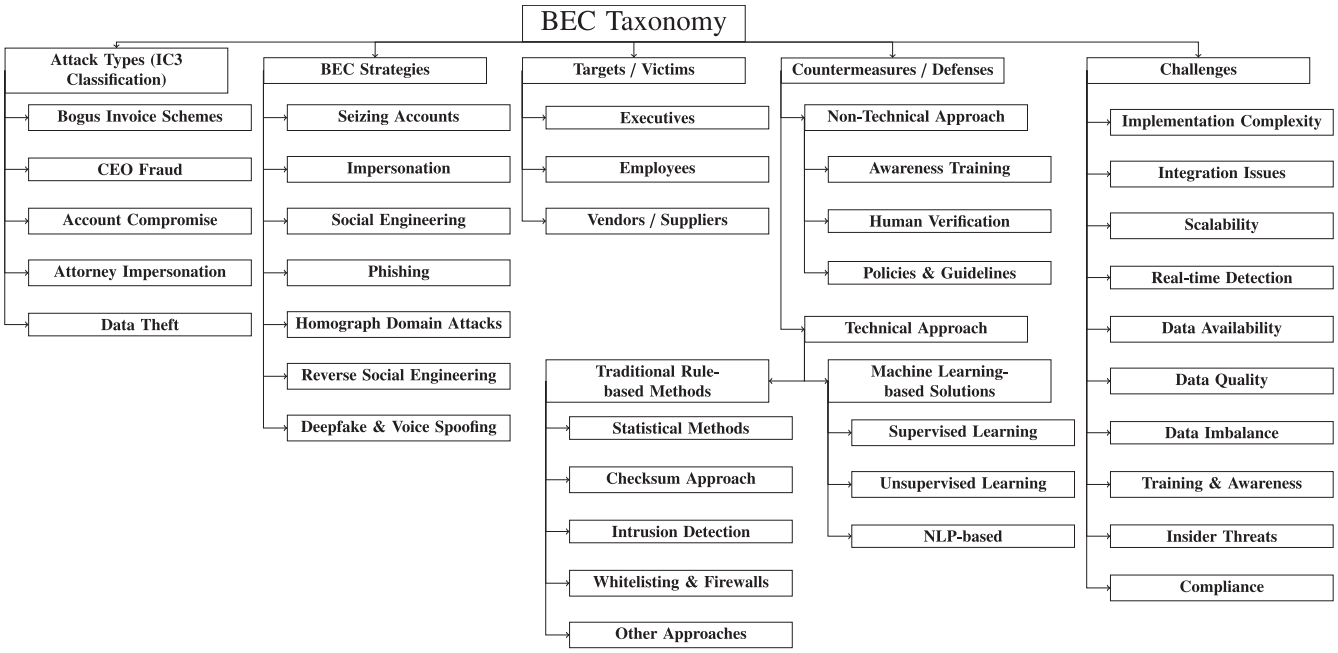

Fig. 5. A taxonomy of Business Email Compromise (BEC) fraud.

focus exclusively on the Technology layer, while only 7 incorporate elements related to Process or People-based controls. Notably, hybrid human-in-the-loop strategies are virtually absent from the literature. This disproportionate emphasis on technological solutions highlights a critical gap in current BEC mitigation efforts and may help explain persistent issues such as alert fatigue and user non-compliance, which are discussed in more detail in Section 7.

5.1. Technical countermeasures

Various technical mitigation proposals have been discussed in the literature. These countermeasures can be broadly divided into two main categories: *Traditional Rule-based Methods* and *Machine Learning-based Solutions*.

Table 4 summarizes the main technical detection techniques, comparing their reported results and highlighting key findings from recent BEC studies. The performance landscape reveals three clear patterns. **(i) Rule-based /signature methods** (e.g., checksum, cryptography, firewall rules) exhibit near-zero false-positive rates but lack adaptability and require manual maintenance, making them brittle against novel BEC variants. **(ii) Classical ML classifiers** (e.g., K-means + sentiment, KNN, SVM) provide moderate accuracy yet struggle with adversarial text manipulation and class-imbalance—issues evident in higher false-positive rates on heterogeneous datasets such as TREC06P. **(iii) Transformer-driven NLP models** (e.g., BERT + BiLSTM, CAPE ensemble) achieve state-of-the-art precision and recall (> 98%) but demand large, well-labeled corpora and frequent retraining to remain effective. Notably, only 2 of 30 studies combine people- or process-level controls with high-performing ML models, indicating a gap in hybrid human-in-the-loop architectures that could mitigate alert fatigue and adaptive attacker tactics. We elaborate on these trade-offs — robustness, scalability, and deployment realism — in Section 7.

5.1.1. Traditional rule-based methods

Scholars and industry experts have collaborated extensively to develop software defences and risk mitigation techniques that enterprises can deploy to counter the rising threat of BEC fraud. Measures such as maintaining up-to-date software, enforcing end-point security, deploying anti-malware systems, and utilizing digital signatures for emails can help protect potential victims from BEC attacks (Meyers, 2018). Another effective approach is analyzing historical email patterns to detect anomalies in communication behavior, which has been demonstrated in commercial applications such as BEC-Guard (Cidon et al., 2019).

Typo-squatting, a tactic in which attackers register domain names that closely resemble legitimate ones, presents an additional threat that can sometimes be mitigated by proactive domain monitoring and early warning systems (Mansfield-Devine, 2016). Organizations can also employ blocklists and allowlists to prevent fraudulent email interactions. Blocklists restrict access from known malicious sources, such as compromised IP addresses and suspicious email domains (Siadati et al., 2020): "If the recipient's email address, IP address, or another characteristic has been blacklisted, the session will be canceled before the email is received" (Tervaniemi et al., 2006). Conversely, allowlists define trusted email senders, reducing false positives. A well-balanced strategy incorporating both blocklists and allowlists is crucial to ensuring seamless, legitimate communication while filtering out fraudulent messages effectively.

Statistical methods. Shahrivari et al. (2020) employed the Delphi technique, collecting feedback from thirty cybersecurity experts to validate BEC detection criteria. Their study highlighted that global financial losses from BEC fraud exceeded $26 billion and identified four key factors crucial for effective detection: recognizing email authenticity, detecting malicious mobile applications, identifying indicators of mobile malware, and discerning phishing attempts. Their approach, which combined expert-driven insights with statistical validation, achieved an accuracy rate of 92.5

Acar et al. (2019) conducted a large-scale analysis of malware attacks collected from two organizations between 2017 and early 2018, focusing on threat vectors, time series analysis, vulnerabilities, and social engineering tactics. Unlike earlier malware research, their study concentrated on modern industrial malware samples. A key finding of their time-based analysis revealed that 93% of malware samples were distributed during weekdays, underscoring the targeted nature of these attacks and the influence of corporate email traffic patterns.

Checksum approach. Teerakanok et al. (2020) proposed a semi-automated method for verifying the authenticity and integrity of financial transactions using a checksum generated from critical transaction

Table 3
Summary of BEC studies and methods.

Study	Non-technical solutions			Technical solutions					
	Awareness training	Human verification	Policies and guidelines	Machine learning	NLP Methods	Checksum	Cryptography	Intrusion detection	Firewall
Tervaniemi et al. (2006)			✓						
Benaroch (2018)			✓						
FBI (2021)		✓							
Mansfield-Devine (2016)	✓								
Nehme and George (2018)	✓								
Zweighaft (2017)	✓								
Meyers (2018)			✓						
Siadati (2019)								✓	
Maleki (2019)				✓					
Binks (2019)	✓								
Ross (2018)	✓								
Shahrivari et al. (2020)			✓						
Papathanasiou et al. (2023)	✓								
Awah Buo (2020)			✓						
Kurematsu et al. (2019)				✓					
Haddon (2020)									✓
Baby et al. (2019)									✓
Aparna et al. (2021)									✓
Acar et al. (2019)			✓						
Cidon et al. (2019)				✓					
Vorobeva et al. (2021)				✓	✓				
(Siadati et al., 2020)			✓						
Wickline (2021)				✓					
Susanti et al. (2023)			✓						
Ogwo-Ude (2023)			✓						
Teerakanok et al. (2020)						✓			
Almutairi et al. (2023)				✓	✓				
Brabec et al. (2023)				✓					
Papathanasiou et al. (2024)							✓		
Lazarus (2024)	✓								

details. The process involves a secret delivery key issued by the invoice-issuing entity, which the supplier then uses to generate a checksum by combining essential transaction data and the secret key. If both parties produce an identical hash, the transaction is deemed legitimate. Their approach employs the SHA256 message-digest function and converts the hash to base 8 for added security.

Papathanasiou et al. (2024) introduced the BEC Defender algorithm, which employs cryptographic techniques such as Message Authentication Codes (MACs) and QR codes to verify the authenticity of email communications. The system utilizes Fernet encryption for secure data storage and SHA2 hashing to enhance the security of the registration process. While extensive testing across multiple email providers and operating systems demonstrated the algorithm's effectiveness, certain limitations remain. These include:

- Challenges in secure key distribution.
- A three-hour validation window, which, while adding security, may reduce usability.
- Potential inaccuracies in MAC address verification.
- A residual risk of replay attacks within the validation timeframe.

Despite these challenges, BEC Defender represents a promising cryptographic approach to mitigating email-based fraud.

Intrusion detection method. Sahoo and Rajitha (2019) proposed an intrusion detection approach designed to distinguish between legitimate and fraudulent emails, thereby safeguarding users against phishing attacks and data breaches. Their method, applied to the Enron dataset, achieved a 98% accuracy rate.

Siadati (2019) focused on BEC attacks that impersonate coworkers, a category of social engineering threats that often bypass traditional phishing detection mechanisms due to their lack of common indicators such as malicious links or suspicious IP addresses. The study introduced a novel countermeasure aimed at disrupting attackers by monitoring their private communications and intercepting key resources (e.g., stolen passwords and fraudulent bank account details). Their system demonstrated a recall rate exceeding 80% and a false positive rate of 0.3%, highlighting its effectiveness in identifying impersonation attempts.

Whitelisting and firewall methods. Haddon (2020) analyzed BEC attack vectors and data exfiltration risks, emphasizing network lockdowns, firewall restrictions, and up-to-date antivirus systems as key defense strategies. Their study provided real-world insights into evolving attack techniques and countermeasures. While these methods can enhance security, they require significant resources and may struggle to keep pace with rapidly evolving threats. Their evaluation was based on case studies and historical reports, without reporting a specific accuracy metric.

Opazo et al. (2017) proposed a client-side security mechanism that analyzes email headers for inconsistencies, logs alerts, and notifies enterprise administrators of potential threats. Their framework includes whitelisting trusted contacts, which reduces false positives while maintaining strict email security policies.

Wickline (2021) examined the effectiveness of modern antivirus solutions in detecting and mitigating malware threats. The study identified BEC, phishing, and spear phishing as primary attack vectors and highlighted how malware is leveraged to disrupt critical infrastructure and steal sensitive data. Additionally, the research noted that malware development surged during the COVID-19 pandemic, with 350,000 new malicious programs created daily, leading to a 40% increase in global malware volume.

Other approaches. While technical defences and detection models form the foundation of BEC mitigation, a number of studies have taken broader or more specialized perspectives to address complementary dimensions of the threat. These include organizational case studies, risk modeling frameworks, and legal or regulatory analyses. Together, these contributions enrich the understanding of BEC by highlighting its psychological, procedural, and institutional implications—extending beyond algorithmic detection and infrastructure-level controls.

Awah Buo (2020) examined the global rise of BEC fraud and presented a case study of Unatrac Holding Ltd. They conducted a detailed investigation into the psychological and sociotechnical impact of a successful BEC attack on both the organization and its employees.

Benaroch (2018) proposed a model modification approach for BEC risk management, where zero or more precautionary measures can be deployed in varying sequences. These measures have impulse-type effects to reduce uncertainty, and their impacts can be substitutive, complementary, or synergistic. This modeling approach enables both passive and proactive IT risk management.

KOLOUCH (Kolouch, 2016) studied legal implications and potential criminal liabilities of phishing, scams, BEC, and other specialized cyber-attacks. Their focus extended to international legal standards, including those defined in the Convention on Cybercrime, as well as the relevant laws within the Czech Republic.

Table 4
Comprehensive summary of BEC attack detection techniques and performance metrics.

Source	Method	Description	Contributions	Limitations	Dataset	Acc.	Prec.	Recall	F1-score	False Pos. Rate
Maleki (2019)	ML	K-means clustering on keywords and sentiment analysis.	Detects BEC; stops emails on sender-side.	Lacks dynamic feature selection; scale issues.	Enron	92%	–	–	93%	–
Cidon et al. (2019)	ML	Two-stage approach (metadata + textual).	Real-time BEC detection; widely adopted.	Needs continuous training.	BEC-Guard	–	98%	96.9%	–	0.000019% (1 in 5,260,000 emails)
Vorobeva et al. (2021)	ML	NLP with TF–IDF, k-Means, LDA.	Effective for Russian/English emails.	Lacks insider attack protection.	2,308 emails	95% (Eng), 75% (Rus)	–	–	–	–
Regina et al. (2020)	NLP	Task-agnostic augmentation using BERT and heuristic translation.	Enhances BEC text classification tasks.	Limited dataset scope; augmentation variability.	Simulated	87%	–	–	–	–
Xiao and Jiang (2020)	ML	KNN and Bi-LSTM for phishing/spam detection.	High phishing detection accuracy.	Needs a larger dataset.	TREC06P	95.27% (KNN), 91.51% (Bi-LSTM)	91.75%	91.49%	91.58%	1.22%
Kurematsu et al. (2019)	ML	Email text author identification.	ML-based author ID approach.	Uses only first 100 words.	Enron	84%	–	–	–	–
Teerakanok et al. (2020)	Checksum	Checksum from invoice.	Counters bogus invoices; easy to implement.	Needs secure key exchange, manual verification.	Simulated	N/A	–	–	–	–
Almutairi et al. (2023)	ML	BERT and BiLSTM for BEC detection.	Captures high-level linguistic properties.	Resource-intensive.	Fraud, TREC-7	99%	99.8% (Fraud), 98.5% (Trec07)	99.7% (Fraud), 98.6% (Trec07)	99.8% (Fraud), 98.6% (Trec07)	–
Wickline (2021)	Whitelisting/ML	Real-time monitoring, endpoint data, app listing.	Flags suspicious activities, blocks malicious files.	High resource use, false positives, complex management.	Enterprise	N/A	–	–	–	–
Brabec et al. (2023)	ML	CAPE combining multiple ML models.	Comprehensive across email modalities.	Limited data, operational costs, explainability needed.	Diverse orgs.	Precision > 80%	–	–	–	–
Papathanasiou et al. (2024)	Cryptography	Secure email scheme using QR codes and MAC.	High security, protection against replay attacks.	Symmetric key distribution, timestamp limits.	Simulated	High precision	–	–	–	–
Sahoo and Rajitha (2019)	Intrusion Detection	Distinguishes fake from real emails.	High phishing detection accuracy.	Needs representative dataset; possible false positives.	Enron	98%	–	–	–	–
Siadati (2019)	Intrusion Detection	Targets social engineering attacks.	Novel deterrence technique.	Hard to generalize; needs attacker communication access.	77 scammer email accounts	–	–	>80%	–	0.3%
Haddon (2020)	Firewall	Analyzes attack vectors and data exfiltration risks.	Provides insights into real-world attack techniques and countermeasures.	High resource demands; challenges in mitigating evolving threats.	Case studies and historical reports	N/A	–	–	–	–

5.1.2. Machine learning-based solutions

Machine Learning (ML) has been widely and successfully applied to various business and research applications, including BEC detection.

Maleki (2019) proposed and tested a behavior-based detection model for compromised email accounts or machines. The model prevents fraudulent emails by blocking messages from compromised senders who fail to form a valid user profile from the recipient's perspective. Additionally, the system alerts legitimate account owners when a compromise is detected. Evaluated on the Enron Dataset, the framework achieved 92% accuracy and a 93% F1-score.

Cidon et al. (2019) introduced BEC-Guard, a two-stage detection system for identifying and blocking impersonation emails. The first stage analyzes email metadata (e.g., sender, receiver, CC, BCC fields) to detect anomalous patterns. If flagged, the email proceeds to content-based analysis, which employs natural language processing (NLP) and link verification. The text classifier uses TF–IDF with unigrams and bigrams (10,000 features), while the link classifier flags small or newly created websites. The combined system reported 98.2% precision, 96.9% recall, and an extremely low false positive rate of 0.000019% (1 in 5,260,000 emails). Despite its success, continuous retraining is necessary to counter evolving attack strategies.

Kurematsu et al. (2019) developed an ML-based author identification model for BEC detection, focusing on writing style analysis. Unlike traditional spam filters, this approach relies on author profiling, analyzing the first 100 words of an email body. Evaluated on the Enron dataset, the system achieved 84% accuracy, highlighting its potential for authorship verification in email security.

Vorobeva et al. (2021) proposed a BEC detection method based on writing style analysis. Their feature set included word n-grams, three-gram phrases, day-of-week, time sent, message urgency, and email headers. Using Linear Support Vector Classification (LSVC) with feature scaling, their system achieved 95% accuracy for English emails and 75% accuracy for Russian emails.

Xiao and Jiang (2020) introduced a phishing and spam detection system using K-Nearest Neighbors (KNN) and Bi-LSTM. Their approach significantly reduced false positives while maintaining high accuracy. Their experiments on the TREC06P dataset resulted in 95.27% accuracy (KNN), 91.51% accuracy (Bi-LSTM), 91.75% precision, 91.49% recall, 91.58% F1-score, and a false positive rate of 1.22

Almutairi et al. (2023) proposed a transformer-based model that combines BERT and BiLSTM for BEC detection, leveraging linguistic traits rather than traditional email features. Their approach achieved 99% accuracy on Fraud and TREC-7 datasets. They reported 0.998 precision, 0.997 recall, and 0.998 F1-score on the Fraud dataset, and 0.985 precision, 0.986 recall, and 0.986 F1-score on the TREC-7 dataset.

Brabec et al. (2023) developed CAPE, a modular and adaptive BEC detection system designed for Security Operations Centers (SOC). CAPE integrates multiple ML models and applies a Bayesian framework for continuous refinement. Over two years, CAPE's precision remained consistently above 80%, demonstrating its reliability in real-world applications. However, its performance heavily depends on data availability, operational costs, and explainability.

NLP methods. Natural Language Processing (NLP) focuses on qualitative analysis of textual content, including emotional and contextual

cues. Regina et al. (2020) presented a corpus task-agnostic augmentation system motivated by BEC detection use cases. Combining NLP methods — such as the BERT language model, multi-step reverse translation, and heuristic-based augmentation — their approach improved performance on various text classification tasks, achieving a balanced accuracy of 96% in a BEC detection setting

5.2. Non-technical solutions

Though the above sections address various technological methods to prevent BEC fraud, there are many non-technical solutions discussed in the recent studies that could be helpful to eliminate BEC fraud summarized in Table 5

5.2.1. Awareness training

Employee education is an essential preventative tool against BEC fraud (Mansfield-Devine, 2016; Opazo et al., 2017). According to Binks (2019), company-wide training and education are the most efficient techniques to minimize phishing assaults. For example, awareness training on comprehensive security, particularly phishing simulators, may highlight current spoofing attack techniques (Agarwal and Kumar, 2016). In addition, simulated assault training may assist individuals in comprehending the delicate signs and current protective strategies of a BEC scam (Ross, 2018). Employees at all levels of an organization need to be provided with BEC scam testing and training (Zweighaft, 2017; Kanistras et al., 2018). According to Nehme and George (2018), businesses must "(1) positively influence employees' understandings of phishing scams via awareness programs, (2) train people to analyze the validity of emails, and (3) teach them about social engineering, phishing, and the risks associated with them".

In addition, Spangler (2021) illustrated the BEC method and the strategies employed by a hostile actor to trick a target into disclosing essential or secret information. They reviewed and categorized the steps of BEC to classify the process of executing this sort of fraud. They investigated impacts and risks to evaluate the probability and severity of a BEC event if it occurred in an organization. They presented a review of preventative techniques, including developing an awareness culture of solid security to determine how to decrease the risk of a BEC.

Furthermore, Papathanasiou et al. (2023) used awareness techniques to understand and combat BEC attacks by examining the operational dynamics and social structures of the "Black Axe" Confraternity. This qualitative study, which involved interviews with an incarcerated BEC scammer, provided insights into the social engineering tactics and organizational methods used by BEC criminals. However, the study is limited by its focus on a single source and specific criminal group, which may not represent the broader landscape of BEC operations.

The study by Jayakrishnan et al. (2022) discussed the various methods and challenges associated with Business Email Compromise (BEC) fraud, particularly within medium and large-scale firms in the USA. The study highlights the importance of robust cybersecurity policies, continuous employee training, and the adoption of advanced detection technologies to combat BEC effectively. The authors emphasize the critical role of awareness and training programs in enhancing the overall cybersecurity posture of firms. They also explore the legal and regulatory frameworks that govern cybersecurity practices, underscoring the need for a proactive approach to policy-making and implementation.

5.2.2. Human verification

The FBI (FBI, 2021) advises email users to validate URLs in emails to confirm their affiliation with the authentic company. Workers should also check hyperlinks with similar but misspelled domain names. The FBI further advises email users not to provide login passwords or other personal info when replying to communications. Employees should verify that the email addresses used by senders match the individuals they claim to be.

5.2.3. Policies and guidelines

Company policies and procedures are crucial in preventing BEC fraud. For example, the FBI recommends businesses implement two-factor authentication, which allows users to verify requests for updating account information. These security measures are akin to those used to guard against personal fraud but are primarily applied in large enterprise settings. It is suggested that governance systems be established to uniformly and securely authenticate all payments. For instance, the validity of emails and requests can be verified by contacting financial personnel (Mansfield-Devine, 2016). Requiring multiple employees to approve significant transaction requests is another effective method to mitigate risk (Meyers, 2018). As emphasized by Burns et al. (2019), a "business governance framework" where any email requests involving substantial sums of money are always cross-checked through an alternative method is vital. Such strategies can introduce additional barriers to prevent compliance with fraudulent requests.

In Lazarus (2024), The primary method used in this study is a qualitative analysis through interviews with a high-profile incarcerated cybercriminal, supplemented by data from tapped phone records. The main dataset includes direct testimonies and law enforcement-monitored phone conversations. The study utilizes Actor-Network Theory (ANT) and Social Network Theory (SNT) to explore the fluid and adaptable structures of these cybercriminal networks. ANT highlights the importance of both human and non-human actors in shaping network dynamics, while SNT emphasizes the roles of social connections and interactions within the network. However, the research has certain limitations. It is based on a single-case study, focusing on one high-profile offender, which might not provide a comprehensive picture of all BEC operations. Additionally, the reliance on interviews and monitored phone data may introduce biases or inaccuracies in the findings. The study also highlights the challenges in visualizing and understanding these non-hierarchical, fluid criminal networks, which differ significantly from traditional organized crime models.

Zweighaft (2017) described BEC and executive impersonation and how they were performed and explored the usage of regulations and actual operational procedures to combat this fraud problem. They stated that financial institutions needed to be aware of the legal and regulatory dangers of BEC, including CEO impersonation. They suggested making efforts to build a proactive, scepticism-based culture to prevent this form of fraud.

Susanti et al. (2023) discussed the increasing prevalence of Business Email Compromise (BEC) fraud and outlined comprehensive strategies for its prevention. Their study emphasizes the importance of non-technical methods, including the implementation of a robust risk management system (ISO 31000:2018) and an information security management system (ISO 27001:2013). The authors highlight the critical role of raising awareness within organizations through training and the use of a whistle-blowing system to report suspicious activities. They provided case studies from Indonesia and other countries to illustrate the methods and impacts of BEC fraud, underscoring the need for organizations to adopt preventive measures and foster a culture of vigilance. The paper, however, does not engage in technical or empirical analysis, focusing instead on policy guidelines and human verification to mitigate the risks associated with BEC.

The author Ogwo-Ude (2023) discusses the significant challenges that Business Email Compromise (BEC) attacks pose to medium and large-scale firms in the USA. He identifies several key strategies for mitigating BEC threats, including the implementation of advanced email authentication protocols such as DMARC, SPF, and DKIM, along with multi-factor authentication and sophisticated threat detection systems. Additionally, he emphasizes the importance of regular training programs to enhance employee awareness and their ability to identify suspicious emails. Establishing comprehensive incident response plans, tailored to address BEC incidents and involving key stakeholders from IT, legal, finance, and communications departments, is also highlighted.

While numerous countermeasures show promise, their success often depends on diverse and high-quality datasets. Section 6 elaborates on the dataset challenges and the various data sources researchers currently employ.

Table 5
Summary of non-technical solutions for BEC fraud detection.

Source	Method	Description	Strengths	Limitation
Mansfield-Devine (2016)	Awareness Training	Employee education and awareness training on phishing and BEC fraud.	Enhances employee ability to recognize and prevent BEC fraud. **Example**: Regular workshops on identifying phishing emails and BEC tactics.	Requires continuous updates and engagement to remain effective.
Binks (2019)	Awareness Training	Company-wide training to minimize phishing assaults.	Comprehensive security awareness, highlighting current spoofing attack techniques. **Example**: Phishing simulation exercises to test employee responses.	Implementation may be resource-intensive.
Ross (2018)	Awareness Training	Simulated assault training to understand BEC fraud indicators and strategies.	Helps employees comprehend and react to BEC scams effectively. **Example**: Role-playing scenarios to practice recognizing BEC attempts.	Needs regular updates to reflect evolving attack methods.
Zweighaft (2017)	Awareness Training	BEC fraud testing and training for all organization levels.	Positively influences employee understanding of phishing scams. **Example**: Comprehensive training programs for new hires and annual refreshers for all staff.	Requires ongoing training and resource allocation.
Nehme and George (2018)	Awareness Training	Programs to educate employees on phishing, social engineering, and associated risks.	Improves employee ability to analyze email validity and recognize scams. **Example**: Interactive training sessions on phishing detection.	Continuous engagement and updating required.
Lazarus (2024)	Awareness Training	Qualitative analysis of cybercriminal networks and social engineering tactics.	Provides insights into criminal tactics and organizational methods. **Example**: Interviews with former cybercriminals to understand their methods.	Focuses on a single case study; may not represent broader BEC operations.
Papathanasiou et al. (2023)	Awareness Training	Examines the social structures and tactics of BEC criminals through interviews.	Offers insights into social engineering tactics used by BEC scammers. **Example**: Detailed analysis of social engineering techniques from insider perspectives.	Limited by focusing on a specific criminal group.
FBI (2021)	Human Verification	Advises users to validate URLs, check hyperlinks, and verify email addresses.	Simple, direct approach to verifying email authenticity. **Example**: Employees are trained to double-check URLs and email addresses for inconsistencies.	Relies on user diligence and awareness.
Meyers (2018)	Policies and Guidelines	Recommends multiple sign-offs on significant transactions to limit risk.	Adds layers of verification to prevent fraudulent transactions. **Example**: Implementing a policy requiring two senior executives to approve wire transfers.	May slow down legitimate processes.
Burns et al. (2019)	Policies and Guidelines	Suggests a business governance framework for high-value email transactions.	Erects barriers to target compliance with fraudulent requests. **Example**: Establishing a protocol where all high-value email requests must be verbally confirmed.	Implementation can be complex and time-consuming.
Susanti et al. (2023)	Policies and Guidelines	Emphasizes cybersecurity policies, continuous employee training, and legal frameworks.	Enhances the overall cybersecurity posture of firms. **Example**: Adoption of ISO 27001:2013 standards for information security management.	Does not engage in technical analysis.
Ogwo-Ude (2023)	Policies and Guidelines	Identifies strategies for mitigating BEC threats, including advanced email authentication and multi-factor authentication.	Enhances protection against BEC with comprehensive incident response plans. **Example**: Implementation of DMARC, SPF, and DKIM protocols along with regular phishing simulation tests.	Requires coordinated effort across multiple departments.

6. Dataset

In this section, we present the datasets used in available studies to address **Objective-3** In predictive machine learning-based security studies, the required data is not ordinarily available beforehand, leading to extra work during data collection. Indeed, collecting a dataset is a better option for quality; however, access to such a dataset is highly complicated, especially in the security field. For this reason, simulated data is an option to expand experiments in this area.

The dataset used by Almutairi et al. (2023) consists of two parts: a fraud email detection dataset with 5,187 phishing emails and 6,742 legitimate messages, and the Trec-7 dataset with 50,199 phishing emails and 25,220 control messages, for detecting Business Email Compromise (BEC) through linguistic content, excluding metadata and attachments. While this helps focus on content-based features, the dataset relies heavily on general phishing corpora (e.g., CEAS, TREC), which may not fully reflect BEC-specific stylistic and tactical nuance. The dataset used by Regina et al. (2020) includes the public SST-2 and TREC-6 datasets, for Business Email Compromise (BEC) detection. SST-2 consists of movie reviews labeled as positive or negative, while TREC is a multi-label questions dataset categorized by question types. The BEC dataset contains anonymized email texts labeled as suspicious or non-suspicious, addressing class imbalance by augmenting minority classes. However, SST-2 and TREC-6 originate from domains unrelated to email fraud, raising concerns about domain transferability and the ecological validity of linguistic signals in BEC detection. Furthermore, Vorobeva et al. (2021) used a dataset with 2,308 email messages from 50 authors in Russian and English, including both genuine and simulated emails. The dataset was created by altering one to two characters in the sender's email address and changing the message sending time. Each user had between 2 and 232 emails, with an average length of 1,943 characters. The dataset's focus on simulations and small sample size may limit its generalizability for broader BEC research. The approach is innovative but anchored in handcrafted perturbations, which may not mirror real-world attacker behavior or deception sophistication. Author in Kurematsu et al. (2019), proposed an identification model from emails received in the past. They first define a target person in advance, and the ML identification models should identify whether this person sends an email. The model was trained and tested using the Enron dataset, but only 1633 emails from the target person were used; the method caused a dataset overfitting. Also the dataset Structure does not reflect a realistic BEC dataset which should include a phishing email in any context. Moreover, using only a single sender as a target further reduces variability, weakening its representativeness for impersonation scenarios common in BEC. The authors of Cidon et al. (2019) developed BEC-Guard using their own dataset of corporate emails from 1,500 diverse organizations, ranging from 10 to over 100,000 mailboxes across various industries. The dataset includes over 7,000 labeled examples of BEC attacks, with access granted to Office 365 APIs providing all historical emails, including internal and external communications. Although the scale and diversity of this dataset are valuable, its proprietary nature limits reproducibility and independent benchmarking, a persistent barrier in BEC research.

The authors in Maleki (2019) proposed and implemented a behavior-based framework for detecting BEC when accounts or machines are compromised. This framework prevents malicious emails from being sent by identifying a lack of sufficient emails from the sender. To evaluate the framework, they used a combination of datasets: the Enron dataset for legitimate emails and the Nigerian dataset for fraudulent emails. The generated data focused on two parts. Firstly, for the header, they implemented a parser to extract Activity-Based Features and Interaction-Based Features. Secondly, for the body, they extracted selected features such as URL-Based Features and Content-Based Features from the Nigerian dataset, generating fake rows of malicious data among the pre-processed benign emails from the Enron dataset. While creative, this synthetic hybrid approach risks introducing artifacts that could bias detection models toward trivial statistical cues rather than deception dynamics.

Table 6 provides a summary of public and self-collected datasets used by researchers for BEC.

Despite growing interest in BEC detection, the field still suffers from a lack of standardized, realistic datasets that capture attacker adaptation, cross-domain deception strategies, and real-world messaging dynamics. This constrains generalizability, limits benchmarking, and hinders progress toward operationally deployable solutions.

7. Challenges to prevention measures

Although various solutions for detecting BEC attacks have been reviewed, it is evident that no single solution acts as a "silver bullet" against BEC fraud. Over time, the threat of BEC has continued to grow, becoming increasingly prevalent in e-crime. As a result, some proposed solutions may become ineffective or outdated. We emphasize the need for ongoing research and innovation to develop robust, adaptable strategies for combating the evolving nature of BEC fraud. To illustrate the challenges more concretely, we include real-world case studies where existing security mechanisms failed to prevent BEC attacks.

7.1. Challenges in technical solutions

In the following, we categorize the challenges in technical solutions as Traditional Rule-based Methods and Machine Learning-based Solutions. Moreover, methodological limitations across reviewed detection systems deserve greater scrutiny. Many studies validate models in clean, controlled settings that omit the messy realities of operational email environments—such as ambiguous language, multilingual exchanges, or adversarial intent. This raises concerns about ecological validity and overfitting to idealized data distributions.

7.1.1. Challenges in traditional rule-based methods

Traditional Rule-based Methods to eliminate suspicious emails from one's inbox are widely employed in both personal and organizational email systems. Filtering techniques rely on specific word combinations and known phishing patterns to distinguish between legitimate communications and fraudulent emails (Siadati et al., 2020). While these techniques can be effective in preventing common phishing attacks, they struggle to detect highly sophisticated BEC attacks that are carefully tailored to individual organizations.

To ensure both relevance and analytical depth, we selected real-world BEC incidents that exemplify distinct attacker strategies and illustrate concrete failure points across technical and non-technical defences.

A prominent illustration is the Treasure Island Homeless Charity BEC attack. In June 2021, cybercriminals infiltrated a bookkeeper's email account and manipulated an existing invoice, successfully redirecting $625,000 to an attacker-controlled account. Because the fraudulent message originated from a legitimate email address and lacked typical phishing markers, spam filters failed to detect it. This incident underscores a key limitation of rule-based filtering systems: they depend on static indicators rather than behavioral context, making them ineffective when attackers operate from compromised accounts. More broadly, it highlights the inherent weaknesses of signature-based approaches, which assume static adversarial patterns and fail to adapt to evolving deception tactics. As attackers increasingly leverage insider mimicry and contextual awareness, detection mechanisms must shift toward adaptive behavioral profiling to remain effective.

7.1.2. Challenges in machine learning-based solutions

Machine learning-based approaches have been explored for BEC detection, leveraging Natural Language Processing (NLP) and behavioral anomaly detection (Cidon et al., 2019; Gascon et al., 2018). However, the effectiveness of ML models is often constrained by several factors:

- The models rely on known patterns of fraudulent behavior, which attackers constantly evolve to evade detection.
- Feature engineering techniques used in traditional ML models are predominantly rule-based, failing to consider the deep contextual meaning of an email's content (Cohen et al., 2018; HADA et al., 2020; Xiao and Jiang, 2020).
- BEC attacks often do not contain obvious malicious indicators, such as URLs or attachments, making them difficult for supervised learning models to classify accurately.

Table 6
Summary of BEC datasets used in literature.

Dataset	Availability	Description	Articles
Enron Email	Public	Contains approximately 500,000 emails from 150 employees (mainly executives), released following the Enron scandal.	Maleki (2019), Almutairi et al. (2023), Kurematsu et al. (2019)
TREC	Public	Approximately 50,000 emails, with about 35,000 spam and 15,000 non-spam messages, commonly used for benchmarking email classification methods.	Regina et al. (2020), Almutairi et al. (2023)
BEC-Guard	Private	A proprietary dataset by Barracuda Networks featuring around 7,000 documented BEC attacks.	Cidon et al. (2019)
Russian & English Emails	Private	A collection of 2,308 genuine and simulated emails from 50 authors, spanning Russian and English messages with modified sender details and timestamps.	Vorobeva et al. (2021)

Even when using advanced NLP, these systems often overlook socio-pragmatic features like authority tone, impersonation patterns, or cross-thread anomalies that characterize BEC. Without modeling the interactional intent of emails or conversation history, ML systems remain brittle to novel fraud expressions.

Another illustrative case is the Insurance Broker Firm attack, in which cybercriminals used a phishing email to compromise an employee's account. Once inside, the attackers inserted themselves into an ongoing conversation with a client, subtly manipulating the thread to convince the client to transfer nearly £300,000 to a fraudulent bank account. Because the communication appeared to originate from a trusted source and followed a legitimate thread structure, even advanced machine learning-based detection systems failed to flag the deception. The fraud was ultimately averted not by automated defences, but by the client's decision to independently verify the payment request. This case highlights the limitations of technical solutions in isolation and underscores the importance of hybrid security approaches that integrate AI-driven detection with procedural verification and human oversight.

7.2. Challenges in datasets

BEC detection applications require high-quality, large-scale datasets to effectively train models for operational use. However, studies in this domain face significant challenges due to limited data availability, privacy concerns, and the sensitivity of fraud-related communications (Nettleton, 2016; West and Bhattacharya, 2016).

Despite several ML-based studies reporting high detection accuracy, meaningful cross-comparison remains difficult. This is largely due to inconsistent evaluation protocols—some studies rely on synthetically balanced datasets, while others use proprietary or imbalanced real-world corpora. For instance, Maleki (2019) report 92% accuracy using behavioral cues on the Enron dataset, whereas Cidon et al. (2019) relies on non-public datasets that hinder reproducibility. Additionally, there is no consensus on the importance of metadata: Vorobeva et al. (2021) stress its necessity, while Kurematsu et al. (2019) suggests it is non-essential. These contradictions highlight the need for standardized benchmarking practices and clearer assumptions about operational requirements.

The scarcity of publicly available, diverse BEC datasets presents several challenges:

- Limited Data Availability: Unlike traditional spam or phishing datasets, BEC datasets are scarce due to the sensitive nature of compromised business communications.
- Privacy Concerns: BEC fraud often involves confidential corporate information, making it difficult for organizations to share incident data for research purposes.
- Lack of Representative Attack Samples: BEC attacks evolve rapidly, and existing datasets may not capture the latest tactics used by cybercriminals.
- Dataset Bias: Most publicly used corpora (e.g., Enron) predominantly reflect formal, English-language business communication in Western settings. This linguistic and cultural bias limits generalizability to diverse organizational contexts and multilingual environments.

This data scarcity perpetuates a cycle where detection models are trained on incomplete or outdated threat representations, leading to blind spots in real-world deployment. Additionally, the absence of adversarially generated emails or impersonation-rich scenarios weakens model robustness under adversarial conditions.

Further methodological concerns arise from evaluation settings that do not reflect realistic deployment scenarios. Many studies train and test on clean, well-formatted corpora without noisy, ambiguous, or multilingual samples—conditions common in enterprise environments. Few explore adversarial testing, conversation-level context, or real-time detection constraints.

Finally, the body of published research may be subject to publication bias, where studies reporting high accuracy or strong results are more likely to be published. This potentially distorts the perception of overall model efficacy, as weaker or null results are underreported.

For instance, the analysis of the Treasure Island BEC attack highlights the importance of real-world datasets in developing more realistic anomaly detection models. Had training datasets included similar cases of email manipulation, detection systems might have been better equipped to flag subtle anomalies in the bookkeeper's correspondence.

7.3. Challenges in non-technical solutions

Despite advancements in technology, human factors remain a critical challenge in BEC detection. Employees' behavior, organizational policies, and security awareness play significant roles in mitigating BEC fraud.

7.3.1. Human error and lack of verification procedures

Even with sophisticated security measures in place, BEC attacks can succeed due to human error. In the Treasure Island case, the bookkeeper trusted an altered invoice and approved the fraudulent transaction without secondary verification. This demonstrates that security measures must go beyond detection and include strong procedural controls, such as requiring dual approvals for high-value transactions. Moreover, many training programs lack realism, failing to simulate the nuanced pressure and plausibility of real BEC scenarios. Without scenario-based learning or regular drills, employees may struggle to translate training into action under time-sensitive pressure.

Conversely, the Insurance Broker Firm case illustrates the importance of human vigilance. The attack was thwarted because the targeted client independently verified the transaction request. This highlights a crucial point: BEC prevention must integrate both technical defences and human decision-making frameworks.

7.3.2. Challenges in security awareness and training

Although employee training programs aim to mitigate human vulnerabilities, their effectiveness is inconsistent. Studies indicate that:

- Security awareness training often fails to prevent BEC fraud because attackers exploit urgency and authority to bypass human skepticism.
- Employees may prioritize efficiency over security, as seen in high-pressure financial environments where transactions must be processed quickly.
- Social engineering tactics continuously evolve, making it difficult to prepare employees for new attack strategies.

Junger et al. (2017) found that warnings and pre-attack education had minimal impact in preventing data leaks, and in some cases, led to an increased likelihood of security mistakes. The Treasure Island case further supports this finding, as the bookkeeper likely lacked sufficient training on verifying invoices and identifying fraudulent modifications.

The case studies presented in this section reinforce key challenges in BEC detection:

- The evolving sophistication of BEC fraud: Attackers continually refine their methods to bypass both traditional security filters and ML-based detection models.
- Gaps in current detection techniques: Rule-based filters fail against account compromise, and ML models struggle with deceptive yet contextually legitimate messages.
- The importance of human verification: The Insurance Broker Firm case demonstrated that human vigilance remains a crucial line of defence.

Ultimately, addressing BEC requires bridging gaps between detection technologies and organizational behavior. Effective defence must account for attacker agility, human decision-making under pressure, and the socio-technical environment in which fraud unfolds.

Ethical and legal considerations. In parallel, the deployment of BEC detection systems raises important ethical and regulatory concerns. Privacy risks emerge when models process sensitive email content or behavioral metadata—especially in jurisdictions with stringent data protection laws such as the GDPR. Furthermore, high false-positive rates may lead to workflow disruptions, reputational harm, or reduced employee trust. These challenges call for carefully calibrated detection thresholds, clear escalation protocols, and human-in-the-loop validation. Organizations must also ensure that AI-powered tools are transparent, accountable, and compliant with legal mandates on data minimization and fairness in automated decision-making.

Meta-synthesis of research tensions

To provide a more integrated understanding of the BEC detection literature, we conducted a meta-synthesis to identify core tensions and contradictions across reviewed studies. Table 7 summarizes these findings and outlines their implications for future research.

The tensions identified in Table 7 underscore the fragmented yet evolving nature of current BEC detection research. These contradictions — between accuracy and deployment, rule-based and data-driven systems, or detection and prevention — highlight that addressing BEC fraud requires more than incremental technical improvements. Instead, what is needed is a coordinated research agenda that reconciles these competing demands through interdisciplinary strategies, grounded experimentation, and globally inclusive datasets. Building on these insights, the following section outlines future research directions and actionable recommendations that aim to close these critical gaps.

8. Future directions and actionable insights

Building on the findings of this review and the critical research gaps identified, this section outlines future research directions and practical recommendations aimed at addressing **Objective-4**. Our goal is to guide both researchers and practitioners toward developing resilient, scalable defences against the increasingly nuanced and socially engineered tactics employed in Business Email Compromise (BEC) fraud.

Future research directions

Building on the challenges and gaps identified in Sections 6 and 7, this section outlines key future research directions that address documented limitations in current BEC detection capabilities.

The following research directions are directly derived from the core challenges presented in Section 7. We highlight unresolved issues and prioritize foundational areas that must be addressed to enable scalable and realistic BEC countermeasures.

1. **Advancing Context-Aware NLP Models:** While transformer-based models (e.g., BERT, BiLSTM) show high classification accuracy (e.g., 99% in Almutairi et al. (2023)), our review in Section 5.1 shows they often miss deeper discourse-level cues such as tone, intent shifts, or impersonation patterns. Future work should explore advanced NLP methods — semantic role labeling, coherence modeling, and dialogue-based analysis — to detect stylistically subtle or narrative-consistent fraud (Gascon et al., 2018).
2. **Improving Dataset Diversity and Benchmarking:** As discussed in Section 6 and highlighted in Table 7, the lack of multilingual, adversarially rich, and sector-specific BEC corpora undermines both model generalizability and cross-study comparability. Future research should prioritize the curation of diverse, high-fidelity datasets, with standardized evaluation protocols and benchmark tasks to reduce methodological fragmentation (Maleki, 2019; Cidon et al., 2019).
3. **Real-Time and Adaptive Detection Frameworks:** Section 5.1 outlines the brittleness of static classifiers under behavioral drift and novel attack strategies. Future systems should support real-time, stream-based detection with adaptive retraining or feedback loops. BEC-Guard (Cidon et al., 2019) provides an early example, but wider deployment and robustness testing are needed to support operational scalability.

Table 7
Meta-synthesis of core tensions in BEC detection research.

Tension or theme	Description	Implications for research
High accuracy vs. real-world deployment	Lab models often achieve high accuracy, but fail in operational environments where BEC emails lack obvious cues.	Future work should simulate realistic scenarios and evaluate models under adversarial and multilingual conditions.
Technical sophistication vs. human vulnerability	Most approaches focus on technical signatures, neglecting how social cues like trust and urgency drive BEC success.	Detection should combine algorithmic methods with human-centered training and decision support.
Rule-based vs. ML-based models	Rule-based systems are interpretable but rigid; ML models offer adaptability but need large, curated datasets.	Comparative evaluation should balance explainability, scalability, and data feasibility.
Data realism vs. availability	Realistic proprietary datasets are often inaccessible; public datasets are limited in diversity and scope.	A shared benchmark initiative is needed to curate anonymized yet representative corpora.
Prevention vs. detection focus	Most methods detect after compromise, ignoring preventive controls at the UI or workflow level.	Research should include proactive defenses, such as verification prompts and email workflow redesign.
Global relevance vs. regional bias	Many studies target English-speaking or Western settings, ignoring cross-cultural attack variations.	Inclusion of diverse regions and languages is essential for robust, globally applicable solutions.

4. **Policy and Procedural Evaluation Studies:** Our analysis in Section 5.2 shows that even when technical solutions are in place, procedural lapses — such as failure to verify payment instructions — often enable fraud (e.g., Treasure Island case). Future research should empirically evaluate the effectiveness of procedural safeguards (e.g., multi-party approvals, transaction thresholds) and compare governance frameworks across high-risk sectors like finance and healthcare (Ogwo-Ude, 2023).
5. **Human–AI Collaboration Interfaces:** As noted in Section 5.2, human judgment is crucial in ambiguous cases, yet current interfaces offer limited support for contextual decision-making. Future systems should offer enriched dashboards with cues such as anomalous tone, behavioral mismatches, or irregular geolocation to guide user actions (Zweighaft, 2017; Junger et al., 2017).
6. **Multimodal Fraud Detection:** In Section 8, we note that language-only systems often fail in impersonation-heavy scenarios. Combining textual analysis with biometric signals (e.g., voice authentication), behavioral data (e.g., access patterns), and system-level context (e.g., device fingerprinting) can create more resilient fraud detection frameworks (Wickline, 2021; Brabec et al., 2023).

Actionable insights for researchers and practitioners

Drawing from the above directions, we recommend the following practical steps:

- *Focus on Linguistic Generalization and Robustness:* As adversaries evolve, models must go beyond pattern recognition to understand linguistic deception. Techniques like adversarial training and zero-shot learning may help improve model resilience against novel fraud expressions.
- *Invest in Open, Annotated BEC Corpora:* Research communities and industry consortia should collaborate on creating publicly available datasets that reflect realistic BEC attacks, particularly those involving multilingual content and author mimicry. Data scarcity remains the most significant barrier to replicable, impactful research.
- *Embed AI in Operational Workflows:* Technical solutions should not remain isolated detection layers. Embedding AI tools within workflows — such as automated flagging of suspicious invoices or identity verification prompts during transactions — can materially reduce fraud success rates, especially when paired with human-in-the-loop designs.
- *Evaluate Organizational Safeguards at Scale:* Future studies should investigate how procedural interventions (e.g., mandatory callbacks, split-approval mechanisms) impact fraud rates over time. Policy effectiveness varies by context and should be evaluated longitudinally and sector-wise.
- *Bridge Technical and Human-Centric Solutions:* Training alone is insufficient. Organizations should adopt interactive simulations, guided response workflows, and real-time decision support systems to align employee behavior with emerging threats. Up to 40% reductions in successful phishing outcomes have been observed when human-centric safeguards are used effectively (Ogwo-Ude, 2023).

In conclusion, securing enterprise communications against BEC fraud demands an interdisciplinary agenda—one that integrates advances in NLP, adaptive system design, organizational governance, and user experience. As attack strategies continue to evolve, so too must our defence mechanisms—anchored not only in technology, but in socio-technical understanding and proactive resilience planning.

9. Conclusion

usiness Email Compromise (BEC) remains one of the most financially damaging and operationally complex threats in today's cybersecurity landscape. While detection capabilities have evolved — from traditional rule-based filters to sophisticated machine learning (ML) and natural language processing (NLP) models — adversaries continue to outpace these defences by exploiting social engineering and context-specific deception.

Across the reviewed studies, a consistent set of vulnerabilities emerges: the rigidity of rule-based systems, the data dependence of NLP-driven models, and the critical importance of procedural safeguards such as dual-approval mechanisms. These recurring patterns indicate that the observed weaknesses are not isolated but likely generalizable to most medium- and large-scale organizations that rely on structured email communications and formal financial workflows. However, evidence remains limited for smaller enterprises and non-English-speaking contexts, highlighting areas for further investigation.

Key insights.

- *Surface-level success, deep-level gaps.* ML can reliably flag anomalous domains, spoofed headers, and malicious links (Kurematsu et al., 2019; Vorobeva et al., 2021), yet still misses payload-free impersonation, narrative manipulation, and other context-rich ploys.

- *Social-engineering blind spot.* Models that reach 98% precision in laboratory settings (Cidon et al., 2019) often falter in multilingual or zero-shot scenarios (Brabec et al., 2023), a gap exacerbated by the scarcity of diverse, public BEC corpora (Maleki, 2019; Almutairi et al., 2023).
- *Human and procedural fragility.* Awareness training and dual-verification protocols help (Meyers, 2018; Ogwo-Ude, 2023) but are vulnerable to fatigue, urgency, and attacker adaptation (Nehme and George, 2018; Ross, 2018).

BEC taxonomy. Fig. 5 refines the taxonomy initially introduced in Almutairi et al. (2024), structuring reviewed techniques and defenses using established cybersecurity and socio-technical models. The division between *Technical* and *Non-Technical* countermeasures follows the People–Process–Technology (PPT) triad, a widely recognized framework for analyzing layered security ecosystems. This lens enables more structured comparisons between algorithmic defences and those requiring organizational coordination or human oversight.

The taxonomy serves three key stakeholder groups:

- **Researchers**—pinpoint under-explored threat vectors and design reproducible benchmarks.
- **Practitioners**—map existing safeguards to specific attacker tactics and spot defence gaps.
- **Policymakers**—see where technical controls and organizational processes mis-align, guiding regulation and funding.

CRediT authorship contribution statement

Amirah Almutairi: Writing – review & editing, Writing – original draft, Methodology. **BooJoong Kang:** Supervision. **Nawfal Alhashimy:** Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Amirah Almutairi reports financial support was provided by Shaqra University. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank the Deanship of Scientific Research at Shaqra University for their support and for providing the opportunity to undertake this research. This work is also supported by "Socio-Technological Solution for Bridging the AI Divide: A Blockchain and Federated Learning-Based AI Training Data Platform" (NRF-2024S1A5C3A02043653).

Data availability

No data was used for the research described in the article.

References

Abnormal Security, 2023b. Key takeaways from the 2023 FBI IC3 report: Business email.... Abnorm. Secur. URL https://www.abnormalsecurity.com.

Acar, A., Lu, L., Uluagac, A.S., Kirda, E., 2019. An analysis of malware trends in enterprise networks. In: International Conference on Information Security. Springer, pp. 360–380.

Agarwal, D.K., Kumar, R., 2016. Spam filtering using SVM with different kernel functions. Int. J. Comput. Appl. 136 (5), 16–23.

Alder, S., 2023. FBI issues warning about BEC scams as losses increase to $43 billion. HIPAA J. URL https://www.hipaajournal.com.

Almutairi, A., Kang, B., Fadhel, N., 2023. The effectiveness of transformer-based models for BEC attack detection. In: International Conference on Network and System Security. Springer, pp. 77–90.

Almutairi, A.M., Kang, B., Hashimy, N.A., 2024. Business email compromise: A comprehensive taxonomy for detection and prevention. In: Proceedings of the 2024 International Conference on Information Systems Security (ICISS 2024). ICISS '24, ACM, Edinburgh, United Kingdom, http://dx.doi.org/10.1145/3700706.3700714.

Aparna, K., Kumar, G.R., Ishar, S., Santhosh, N., Sreeja, D., 2021. CaseStudy on ddos attacks and attack trendsin cloud computing environments. Int. J. Techo. Eng..

Atlam, H.F., Oluwatimilehin, O., 2022. Business email compromise phishing detection based on machine learning: A systematic literature review. Electronics null, null. http://dx.doi.org/10.3390/electronics12010042, URL https://www.semanticscholar.org/paper/atlam2022machine.

Awah Buo, S., 2020. An application of cyberpsychology in business email compromise. arXiv e-Prints. arXiv–2011.

Baby, R.T., Ebenezer, V., Karthik, N., 2019. Magnum opus of phishing techniques. Int. J. sci. Tech. research.

Benaroch, M., 2018. Real options models for proactive uncertainty-reducing mitigations and applications in cybersecurity investment decision making. Inf. Syst. Res. 29 (2), 315–340.

Binks, A., 2019. The art of phishing: past, present and future. Comput. Fraud Secur. 2019 (4), 9–11.

Bitdefender, 2023. BEC attacks in 2023: What organizations need to know. Bitdefender URL https://www.bitdefender.com.

Brabec, J., Šrajer, F., Starosta, R., Sixta, T., Dupont, M., Lenoch, M., Menšík, J., Becker, F., Boros, J., Pop, T., et al., 2023. A modular and adaptive system for business email compromise detection. arXiv preprint arXiv:2308.10776.

Braun, V., Clarke, V., 2006. Using thematic analysis in psychology. Qual. Res. Psychol. 3 (2), 77–101.

Burns, A., Johnson, M.E., Caputo, D.D., 2019. Spear phishing in a barrel: Insights from a targeted phishing campaign. J. Org. Comput. Electron. Commer. 29 (1), 24–39.

Cidon, A., Gavish, L., Bleier, I., Korshun, N., Schweighauser, M., Tsitkin, A., 2019. High precision detection of business email compromise. In: 28th USENIX Security Symposium (USENIX Security 19). pp. 1291–1307.

Cohen, A., Nissim, N., Elovici, Y., 2018. Novel set of general descriptive features for enhanced detection of malicious emails using machine learning methods. Expert Syst. Appl. 110, 143–169.

Cross, C., Gillett, R., 2020. Exploiting trust for financial gain: An overview of business email compromise (BEC) fraud. J. Financ. Crime.

FBI, 2021. Operation Rewired. FBI, URL https://www.ic3.gov/Media/Y2022/PSA220504.

Federal Bureau of Investigation (FBI), 2017. Business E-mail compromise on the rise. https://www.fbi.gov/news/stories/business-e-mail-compromise-on-the-rise. (Accessed 10 October 2024).

Gascon, H., Ullrich, S., Stritter, B., Rieck, K., 2018. Reading between the lines: content-agnostic detection of spear-phishing emails. In: International Symposium on Research in Attacks, Intrusions, and Defenses. Springer, pp. 69–91.

HADA, T., SEI, Y., TAHARA, Y., OHSUGA, A., 2020. Codewords detection in microblogs focusing on differences in word use between two corpora. In: 2020 International Conference on Computing, Electronics and Communications Engineering. ICCECE, pp. 103–108. http://dx.doi.org/10.1109/iCCECE49321.2020.9231109.

Haddon, D.A., 2020. Attack vectors and the challenge of preventing data theft. In: Cyber Security Practitioner's Guide. World Scientific, pp. 1–50.

I.B.M. Security, 2023a. IBM X-Force Threat Intelligence Index 2023. Tech. rep., IBM Corporation, URL https://www.ibm.com/think/topics/business-email-compromise.

Internet Crime Complaint Center (IC3), 2023. 2023 internet crime report. URL https://www.ic3.gov/Media/PDF/AnnualReport/2023_IC3Report.pdf. (Accessed 24 May 2024).

Jayakrishnan, G., Banahatti, V., Lodha, S., 2022. PickMail: a serious game for email phishing awareness training. In: Usable Security and Privacy (USEC) Symposium, vol. 2022.

Junger, M., Montoya, L., Overink, F.-J., 2017. Priming and warnings are not effective to prevent social engineering attacks. Comput. Hum. Behav. 66, 75–87.

Kanistras, K., Rutherford, M., Valavanis, K., 2018. Foundations of Circulation Control Based Small-Scale Unmanned Aircraft. Springer.

Kelly, R., 2023. Business email compromise attack costs far exceeding ransomware losses. ITPro URL https://www.itpro.com.

King, D., 2019. The future of us fraud in a post-emv environment. Fed. Reserv. Bank Atlanta.

Kitchenham, B., Charters, S., 2007. Guidelines for performing systematic literature reviews in software engineering. EBSE.

Kolouch, J., 2016. CyberCrime. CZ. NIC, zspo.

Kurematsu, M., Yamazaki, R., Ogasawara, R., Hakura, J., Fujita, H., 2019. A study of email author identification using machine learning for business email compromise.. In: SoMeT. pp. 205–216.

Lazarus, S., 2024. Cybercriminal networks and operational dynamics of business email compromise (BEC) scammers: Insights from the "black axe" confraternity. Deviant Behav. 1–25.

Maleki, N., 2019. A behavioral based detection approach for business email compromises (Ph.D. thesis). University of New Brunswick..

Mansfield-Devine, S., 2016. The imitation game: How business email compromise scams are robbing organisations. Comput. Fraud Secur. 2016 (11), 5–10.

Meyers, A., 2018. Not your fairy-tale prince: the Nigerian business email compromise threat. Comput. Fraud Secur. 2018 (8), 14–16.

Microsoft Security, 2017. What is business email compromise?. URL https://www.microsoft.com/en-us/security/business/security-101/what-is-business-email-compromise-bec#areaheading-oc3a35.

Microsoft Threat Intelligence, 2023. Digital Defense Report 2023. Tech. rep., Microsoft Corporation, URL https://www.microsoft.com/en-gb/security/business/security-101/what-is-business-email-compromise-bec/.

National Cyber Security Centre (NCSC), 2023. Annual Review 2023. Tech. rep., NCSC, UK, URL https://www.ncsc.gov.uk/guidance/business-email-compromise-defending-your-organisation.

Nehme, A., George, J.F., 2018. Iterating the cybernetic loops in anti-phishing behavior: A theoretical integration. Twenty-Fourth Am. Conf. Inf. Syst..

Nettleton, D.F., 2016. A synthetic data generator for online social network graphs. Soc. Netw. Anal. Min. 6 (1), 1–33.

Ogwo-Ude, O., 2023. Business email compromise challenges to medium and large-scale firms in USA: An analysis. Open J. Appl. Sci. 13 (6), 803–812.

Opazo, B., Whitteker, D., Shing, C.-C., 2017. Email trouble: Secrets of spoofing, the dangers of social engineering, and how we can help. In: 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery. ICNC-FSKD, IEEE, pp. 2812–2817.

Papathanasiou, A., Germanos, G., Kolokotronis, N., Glavas, E., 2023. Cognitive email analysis with automated decision support for business email compromise prevention. In: 2023 8th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference. SEEDA-CECNSM, IEEE, pp. 1–5.

Papathanasiou, A., Liontos, G., Paparis, G., Liagkou, V., Glavas, E., 2024. BEC defender: QR code-based methodology for prevention of business email compromise (BEC) attacks. Sensors 24 (5), 1676.

Regina, M., Meyer, M., Goutal, S., 2020. Text data augmentation: Towards better detection of spear-phishing emails. arXiv preprint arXiv:2007.02033.

Ross, C., 2018. The latest attacks and how to stop them. Comput. Fraud Secur. 2018 (11), 11–14.

Sahoo, P.K., Rajitha, C., 2019. Detecting forged E-mail using data mining techniques. Int. J. Eng. Adv. Technol..

Saud Al-Musib, N., Mohammad Al-Serhani, F., Humayun, M., Jhanjhi, N., 2021. Business email compromise (BEC) attacks. Mater. Today: Proc. (xxxx), http://dx.doi.org/10.1016/j.matpr.2021.03.647, URL https://doi.org/10.1016/j.matpr.2021.03.647.

Shahrivari, V., Darabi, M.M., Izadi, M., 2020. Phishing detection using machine learning techniques. arXiv preprint arXiv:2009.11116.

Siadati, H., 2019. Prevention, detection, and reaction to cyber impersonation attacks (Ph.D. thesis). New York University Tandon School of Engineering.

Siadati, H., Koven, J., Silva, C.F.d., Jakobsson, M., Bertini, E., Maimon, D., Memon, N., 2020. A framework for analysis attackers' accounts. In: Security, Privacy and User Interaction. Springer, pp. 63–89.

Spangler, M., 2021. Business Email Compromise: Impacts and Strategies for Protecting Against Social Engineering Attacks (Ph.D. thesis). Utica College.

Susanti, D.S., Subandi, F.E., Failasufa, N., Putri, W.A., 2023. Business email compromise (BEC) fraud and how to prevent it. Asia Pac. Fraud. J. 8 (2), 269–280.

Teerakanok, S., Yasuki, H., Uehara, T., 2020. A Practical Solution against Business Email Compromise (BEC) Attack using Invoice Checksum. In: Proceedings - Companion of the 2020 IEEE 20th International Conference on Software Quality, Reliability, and Security, QRS-C 2020. pp. 160–167. http://dx.doi.org/10.1109/QRS-C51114.2020.00036.

Tervaniemi, M., Szameitat, A.J., Kruck, S., Schröger, E., Alter, K., De Baene, W., Friederici, A.D., 2006. From air oscillations to music and speech: functional magnetic resonance imaging evidence for fine-tuned neural networks in audition. J. Neurosci. 26 (34), 8647–8652.

Vorobeva, A., Khisaeva, G., Zakoldaev, D., Kotenko, I., 2021. Detection of business email compromise attacks with writing style analysis. In: International Symposium on Mobile Internet Security. Springer, pp. 248–262.

West, J., Bhattacharya, M., 2016. Intelligent financial fraud detection: a comprehensive review. Comput. Secur. 57, 47–66.

Wickline, T., 2021. The Capabilities of Antivirus Software to Detect and Prevent Emerging Cyberthreats (Ph.D. thesis). Utica College.

Xiao, D., Jiang, M., 2020. Malicious mail filtering and tracing system based on KNN and improved LSTM algorithm. In: 2020 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech). IEEE, pp. 222–229.

Zweighaft, D., 2017. Business email compromise and executive impersonation: are financial institutions exposed? J. Invest. Compliance.