Linguistic Cues of Deception in Malaysian Online Investment Scams' Promotional Materials

Ameiruel Azwan Ab Aziz [a]
ameirul@uitm.edu.my
Universiti Teknologi MARA Cawangan Melaka, Malaysia

Nurul Atiqah Mohd Sharif [b]
2022859324@student.uitm.edu.my
Universiti Teknologi MARA Cawangan Melaka, Malaysia

Wan Farah Wani Wan Fakhruddin
wanfarah@utm.my
Universiti Teknologi Malaysia, Malaysia

Amirah Mohd Juned
amirahjuned@uitm.edu.my
Universiti Teknologi MARA Cawangan Melaka, Malaysia

Nursyaidatul Kamar Md Shah
nursyaidatul@uitm.edu.my
Universiti Teknologi MARA Cawangan Melaka, Malaysia

Ariff Imran Anuar Yatim
ariffimran@uitm.edu.my
Universiti Teknologi MARA Cawangan Melaka, Malaysia

Aminabibi Saidalvi
aminabibi@uitm.edu.my
Universiti Teknologi MARA Cawangan Johor, Malaysia

ABSTRACT

The entire world has transitioned to a borderless information flow in this high-technology era, making communication more effective at the ease of the fingertips. However, these advantages come with various cybercrimes that can easily mislead readers and win them over to their point of view, including online investment scams. This quantitative study aims to analyse the linguistic cues of deception of investment scams' promotional materials using the Linguistic Inquiry and Word Count (LIWC) and Statistical Package for Social Science (SPSS) software. The data was gleaned from official website pages of investment scams provided by the Royal Malaysia Police (RMP), Central Bank of Malaysia (CBM), Financial Consumer Alert List (FCA), and the Securities Commission Malaysia (SC). Descriptive analysis and Pearson correlation analysis were conducted. The findings of the descriptive analysis show that the highest linguistic cue used in the online investment scam is *Lifestyle*. For Pearson correlation analysis, the findings show that linguistic cue for *Perception* significantly correlates with other linguistic cues such as *Lifestyle, Social Process, Cognition,* and *Affect.* This indicates that the linguistic cues used in online investment scams are related. The findings of the study can be used as a guide to prevent online investment scam problems in the future.

Keywords: Deception; Linguistic Cues; Online Investments; Promotional Materials; Scams

---

[a] *Corresponding author*
[b] *Main author*

eISSN: 2550-2131
ISSN: 1675-8021

INTRODUCTION

Language is used to communicate information, which may be truthful and deceptive and loaded with emotions and sentiments like joy, guilt, fear, and fury (Adha, 2020). Human beings are known to lie to manipulate the impressions others have on them and to gain benefits from the deception (Shaari et al., 2019). Even though scammers might leave linguistic footprints in their lies, identifying them to predict deception is inevitably challenging in this technologically dependent era. Today, information is transferred without boundaries through the internet and communication can now take place at anyone's fingertips. Unfortunately, computer-mediated communication has also perpetuated lies that are planned, especially in asynchronous interactions. Of late, digital deception is becoming a phenomenon and contributes to the increasing number of cases of cybercrimes, including the nefarious phenomenon of online investment scams.

The Central Bank of Malaysia (CBM) (2023) defines online investment scams as illegal deposit-taking operations that primarily engage, communicate, and make deals with potential investors on fund administration and investment advice without a license via emails and websites. The CBM (2023) has listed approximately 420 businesses and individuals that are neither authorised nor approved under the Financial Services Act (FSA) 2013. The business activities and investments of these companies and individuals range from gold to cryptocurrency investments, trading investments, and other investment scams. The Royal Malaysia Police's Commercial Crime Investigation Department (CCID) recorded losses of investment frauds worth more than RM245 million in 2021, which was a significant increase compared to the losses exceeding RM180 million in 2020 ("Bukit Aman", 2022). This amount is projected to increase dramatically as 7,009 cases of loan and investment scams were recorded from January to May 2022 (New Straits Times, 2022). This can lead to a severe predicament for the economy, society, and the nation, and can be harmful to the individuals who become victims of such scams (Deora & Chudasama, 2021). The wide use of the internet has made it feasible for scammers to contact millions of potential victims inexpensively globally and charm their victims to participate in non-existent investment schemes. Despite the significant risks, most victims are eager to grow their money quickly in anticipation of huge profits (Rahman et al., 2021). As a result, Malaysia continues to see an increase in reported cases of online investment scams every year. According to the Securities Commission's (SC) Annual Report in 2020, the number of official complaints and inquiries regarding financial misconduct has shown an increasing trend for the past five years, as shown in Table 1.

TABLE 1. Complaints and inquiries received by the Securities Commission (SC) of Malaysia (Securities Commission Malaysia, 2020)

	2020	**2019**	**2018**	**2017**	**2016**
Complaints	1,482	821	558	444	433
Enquiries	2,193	1,336	1,104	626	609
Total	3,675	2,157	1,662	1,070	1,042

The hypothetical assumption regarding this problem is that most victims were deceived by scammers' "sweet words" that influenced their decision-making and actions. Scammers tend to use more positive words (e.g., guaranteed, profit, and success) than negative words (e.g., dangerous, suspicious, and risk) to convince and influence the victims to believe them and to prove that the product or service can be trusted (Zhou & Zafarani, 2020). Because of this, individuals are susceptible to being duped by fraudulent websites of investment scams.

Scammers regularly employ linguistic cues to create a trustworthy and authentic impression of themselves (Adha, 2020). In a study by Addawood et al. (2019), scammers frequently use group references and choose third-person pronouns (e.g., we, their, her) over actual first-person pronouns (e.g., I, me, myself) in their deceitful discourse to distance themselves from the message they created. This issue has grown over time since most victims occasionally lack the self-control to live up to their lifestyle and want to live in luxury, as represented by the scammers (Deora & Chudasama, 2021). This matter of contention is a worrying phenomenon when the victims are influenced by better financial profits with exceptional returns once they join the investment as proposed by the scammers. The victims will entirely rely on the promised return and be prepared to spend all their money, including family savings and their savings, on the investment, which will have unfavourable effects like losing their retirement funds and properties in addition to their life savings (Mohd Padil et al., 2022). The language and the proof portrayed by the scammers show that the product or service was legitimate based on specific words in their promotional materials, such as product brands, functional expressions, punctuation marks, and tenses (Li et al., 2020). Furthermore, Genao (2021) highlighted that scammers avoid referring to themselves when speaking to their victims. They also frequently employ more generalised language (e.g., always, every, all) to avoid providing too much specific information about the scheme (Choudhury, 2014).

To date, there is a breach in the literature addressing online investment scams in the Malaysian context to identify the linguistic cues to predict deception. The proposed solution to this widespread phenomenon is to study how deceptive language is presented in online investment scams' promotional materials to provide comprehensive public education as part of the preventative actions against this pervasive problem.

RESEARCH AIM

It is believed that online investment scams can be tackled by educating the public on potential threats from perpetrators. This action involves increasing awareness of scams and providing consumers with knowledge on how to avoid scam investment schemes. Educating individuals about the deceptive techniques and persuasion strategies used by scammers can help them develop their critical evaluation skills and be more cautious when encountering suspicious investment offers (Naksawat et al., 2016). This study aims to examine the linguistic cues of deception provided in Malaysian online investment scams' promotional materials and to analyse the correlation patterns of the cues. Even though the concept of deception may be universal, this study focused on identifying similar linguistic cues of deception on different promotional materials and how the cues are correlated with one another. To assess this, promotional materials from ten online investment scams listed by the Central Bank of Malaysia (2023) which were active during the data collection period were chosen and analysed. As this study dealt with text-based Computer-Mediated Communication (CMC), LIWC dimensions were used to extract the semantic features needed to predict deception (Newman et al., 2003).

LITERATURE REVIEW

LANGUAGE AND DECEPTION

There are various forms of verbal and non-verbal deception detection in a language, such as through facial expression, eye gaze, body dynamics, statement analysis, voice pitch, and the patterns of the language used to convince people (Genao, 2021). Language use might mislead the audience when the speaker has a negative bias toward deception or fabricating messages (Addawood et al., 2019). This is because language is one of the powerful mediums that can

influence one's thoughts, decisions, and behaviour, both positively and negatively. Worryingly, studies on scams and frauds revealed that victims were reported to be swayed by the communication techniques employed by scammers (Karim et al., 2023), specifically in the utilisation of persuasive communication skills to lure the victims and pressure the people to join the schemes. In scam detection and prevention, researchers and investigators can identify common linguistic characteristics, or linguistic footprints, across various scam communications. Linguistic footprints refer to the unique linguistic patterns, styles, and strategies individuals use in communication (Zhou & Zhang, 2008). In investment scam studies, linguistic footprints refer to the specific language techniques employed by scammers to manipulate and deceive individuals or groups involved in investment scams or fraudulent financial schemes. To investigate the linguistic footprints of deception by scammers, it is crucial to identify which cues to deception were generally used in their promotional materials. This can be done by referring to a framework developed by Pennebaker and King (1999) called Linguistic Inquiry and Word Count (LIWC).

LINGUISTIC INQUIRY AND WORD COUNT (LIWC)

Linguistic Inquiry and Word Count (LIWC) is a software for text-based deception detection that uses the identification of various linguistic dimensions within text or speech to classify words into psychologically driven categories that offer insights into an individual's personality, psychological state, and communication style (Pennebaker & King, 1999). LIWC has been widely used to study deception due to its ability to classify texts according to truth conditions (Levitan et al., 2018; Addawood et al., 2019; Van Der Zee et al., 2020). The core of the LIWC program is the internal dictionary, which consists of over 12,000 words, word stems, phrases, and selected emoticons. The dictionary allows individual and multiple language files to be mapped into psychologically motivated categories, thus making it a valuable tool for deception detection and psychological profiling. LIWC can extract the various linguistic cues from the datasets through natural language processing. The software focuses on detecting accurate word data and linguistic cues, such as changes in word quantity, pronouns, emotional terms, and distinction markers that may reflect deception (Boyd et al., 2022). LIWC classifies the linguistic cues into different linguistic dimensions of psychological processes such as *Drives*, *Cognition*, *Affect*, *Social Process*, *Culture*, *Lifestyle*, *Physical*, *States*, *Motives*, *Perception*, *Time Orientation,* and *Conversational*. Five linguistic dimensions of psychological processes were selected for this study to achieve the research aim, namely, *Lifestyle*, *Social Process*, *Perception*, *Affect,* and *Cognition* (Pennebaker & King, 1999).

Lifestyle is one of the dimensions of the linguistic cues of deception. Also known as affinity fraud, scammers have been reported to use religious sentiments in their promotional materials to attract their victims' attention and make them more reliable regarding religious matters. Most were purported to be pastors, preachers, priests, imams, or other religious leaders to gather money for the neighbourhood or any other organisation (John, 2018). They misled and manipulated their victims by claiming that the donations were for the community or authorities. Frankel (2012) noted that in the United States, for example, hundreds of fellow parishioners in Kansas, Nebraska, and Missouri were persuaded by the church leader to invest in a non-existent prime bank trading programme by giving their investment funds biblical names. Such a move implies that the funds would fulfil a religious duty and, most importantly, rely on a high level of trust among church members. Furthermore, because faith is a powerful motivator, scammers frequently use religious phrases to target and convince victims. Examples of these phrases include *"God wants us to prosper," "Getting into heaven, one dollar at a time,"* and *"God will make you rich, but he just wants you to give me a few dollars first"* (John, 2018). In addition, scammers also persuade their victims by spamming quotes or pictures of

luxurious lifestyles to mislead their victims' decision-making intentionally. Most victims are willing to do anything to meet their basic needs, cover their living expenses, and increase their household income so as to live a better lifestyle without considering the risks associated with investments (Mohd Padil et al., 2022).

Next, for the dimension of *Social Processes* or *Social*, the focus of the linguistic cues is on social behaviours and referents. In the updated LIWC program, Boyd et al. (2022) noted that this dimension also includes subordinate categories such as prosocial (behaviours or referents that signal to help or caring about others at the interpersonal level), politeness, interpersonal conflict words (reflecting referents to concepts that indicate conflicts), moralisation, and communication words (talk, explanation, disagreement). Examples of linguistic cues in this dimension can be found in a study conducted by Newman et al. (2003), who discovered that words such as *talk*, *us*, and *friend* were used to establish the tone of the communication, the level of social engagement, the type of relationships being discussed, and the overall social context of the text.

The third linguistic cue used in promotional materials is the *Cognition* dimension. The term *Cognition* refers to how people think or describe their thoughts in various ways, such as cognitive process, differentiation, and certitude (Pennebaker et al., 2015). This dimension highlights how scammers use ambiguous rather than precise words to deceive and manipulate their victims. A few samples of words and phrases used in this dimension to influence prospective victims are 'may need', 'may change', and 'but not' (Pennebaker et al., 2015). These words and phrases are manipulated to avoid luring prospective investors. Another sign of this dimension is the ability to produce fewer overall words due to the cognitive challenges of lying, which facilitates managing information and avoiding contradictions by speaking less (Toma & Hancock, 2010).

Past research revealed that one of the linguistic cues that scammers have always used is the *Affect* dimension, such as 'emotive words' (Newman et al., 2003). The term 'emotive words' (e.g., happy, pretty, good) refers to using particular word selections to elicit an emotional response from the readers (Newman et al., 2003). Emotive words frequently seek to persuade the readers or listeners to share the writer's or speaker's point of view by employing language to evoke an emotional response (Pennebaker, 2015). For instance, emotive words are frequently used as positive and inspirational quotes to attract their victim's attention (Pennebaker, 2015).

Next, Levitan et al. (2018) noted that the *Perception* dimension designates a category that includes terms related to the cognitive process involved in observing and understanding data from the environment (e.g., high level, move into, global). This dimension captures the linguistic cues that reveal how people interpret sensory information from observations and comprehend their environment. In a study on the *Perception* dimension by Connell (2012), persons who lie use modal verbs like 'may' or 'might' and generalising terms like 'some' and 'anytime' to create ambiguity and uncertainty in their thoughts. In addition, persons who lie like to leave gaps in their material by using fewer words or sentences and lessen the information's specificity by omitting crucial details like timing or location (Connell, 2012). Scammers also frequently use the same terms and limit the vocabulary of their messages so that they appear trustworthy, and the individual who receives the message will be incapable of detecting the deception (Adha, 2020).

eISSN: 2550-2131
ISSN: 1675-8021

CRITERIA-BASED CONTENT ANALYSIS (CBCA)

Criteria-Based Content Analysis (CBCA) was selected and used to analyse the linguistic cues of deception in online investment scams' promotional materials, as suggested by Addawood et al. (2019). CBCA uses a statistical approach to create features and develop predictive models. Earlier research applied the bag-of-words approach to detect cybercrime patterns in text-based CMC (Adha, 2020), as text message lexicons are integrated and used as features to identify cybercrime. Despite its widespread application in text classification, several studies indicate that this method cannot identify cybercrime patterns (Adha, 2020). However, this study used CBCA because models developed from bag-of-words in one web genre cannot generalise well with data from another.

CBCA has been widely used in previous studies as a tool for determining the reliability of witness testimony and separating truthful from false reports or documents of past events to stop illegal activities from spreading further and affecting a more significant segment of the population in the nation. For instance, a study by Genao (2021), which investigated the application of CBCA to adult testimony in a meta-analysis via a test was designed to differentiate between truthful and deceptive reports of past events, exhibited that approximately 80% of liars for nonverbal cues were classified as deception . In contrast, word-based syntactic features include the frequency of punctuation marks, the occurrence of function words, and parts-of-speech (POS) tagging (Genao, 2021).

Besides being used in several academic and forensic settings, CBCA was initially designed for child sexual abuse testimony. Cues indicating deception, on the other hand, can be classified as verbal, nonverbal, or physiological signals. There are two types of verbal cues: content-based cues, which can be proven false compared to the truth, and linguistics-based cues, which can be recognised when the truth is unknown. This is because text-only computer-mediated communication lacks nonverbal cues to dishonesty, and the truth is frequently challenging to confirm. As a result, most research on deception in computer-mediated communication focuses on linguistic cues.

METHODOLOGY

This study uses a purposive sampling strategy to select instances of datasets in each web genre. The list of ten unauthorized websites that the Securities Commission Malaysia (SC), the Central Bank of Malaysia (CBM), the Royal Malaysia Police (RMP), and the Financial Consumer Alert List (FCA) provided served as the unit of analysis. The information about linguistic cues of deception in online investment scams' promotional materials identified and flagged as active investment scams was gathered from this list.

A CBCA was computed based on the existing psychometric test scales in LIWC software to identify the linguistic cues of deception used in the promotional materials. In LIWC, psychological process language dimensions are specific categories or themes into which words are categorised according to their linguistic and psychological meanings (Boyd et al., 2022). These dimensions are characteristics that encompass a variety of features of human communication. *Lifestyle*, *Cognition*, *Affect*, *Social Process,* and *Perception* are examples of the psychological process language dimensions (Boyd et al., 2022).

In this study, LIWC extracted 77.78% of the terms from the websites' promotional materials for investment schemes. LIWC coding's central concept is the classification of texts based on truth conditions. Although LIWC is a strong tool that can offer insightful analysis of the language used in a text, it is not able to capture all of the meanings or nuances of human language due to a variety of factors, including limited dictionary coverage, contextual ambiguity, lack of semantic understanding, ignoring non-verbal cues, cultural and regional

eISSN: 2550-2131
ISSN: 1675-8021

variations, sarcasm and irony, and complex sentence structures (Rubin & Conroy, 2011). Deception detection has an accuracy of 74% when used with common classification techniques like decision trees and logistic regression (Fuller et al., 2009). The accuracy of the classifier reaches an average accuracy rate of 70% when using LIWC for detecting deceptive opinions (Mihalcea & Strapparava, 2009). Human judges, in contrast, are only successful in detecting deception 50–63% of the time (Rubin & Conroy, 2011).

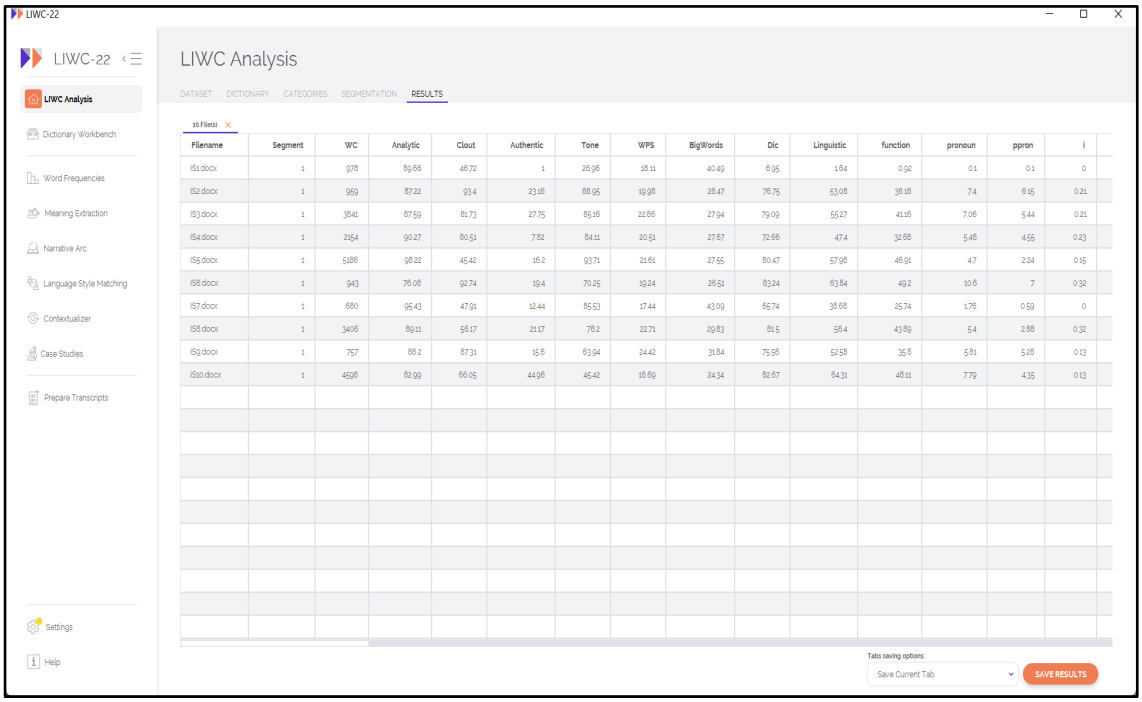


FIGURE 1. LIWC Analysis

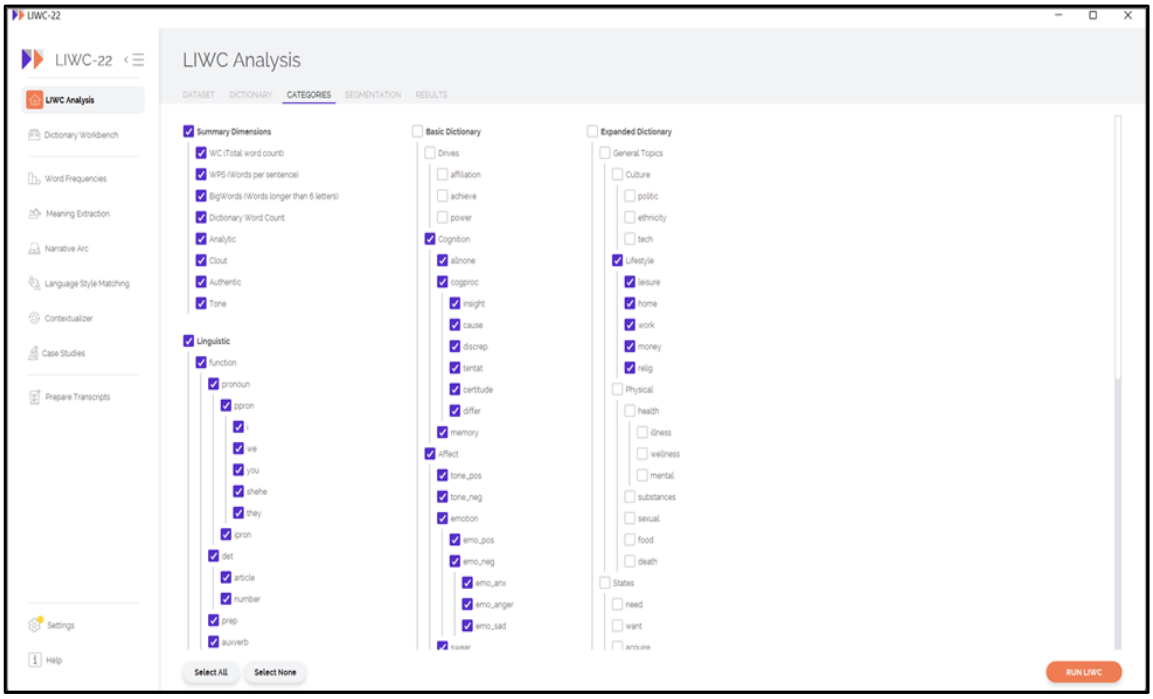


FIGURE 2. Linguistic Dimensions in LIWC

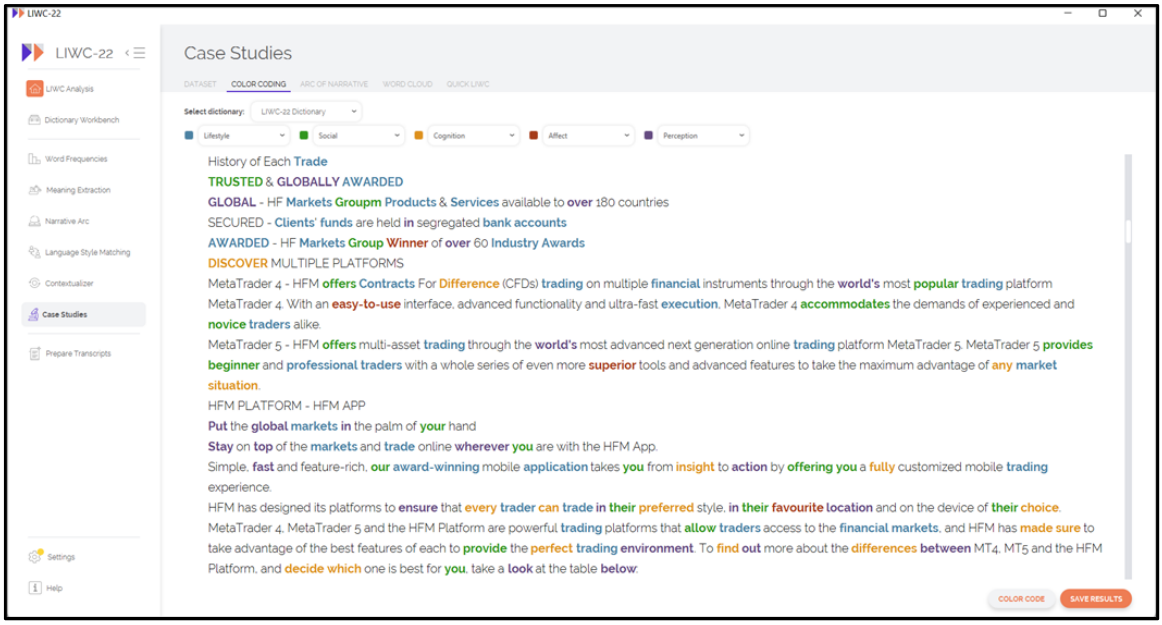
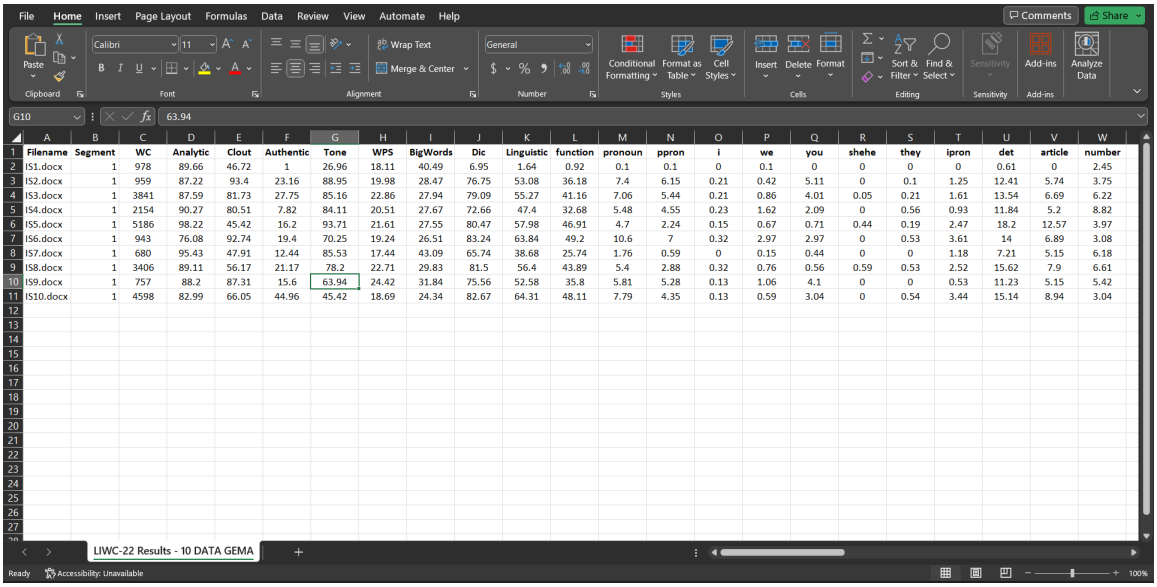


FIGURE 3. Linguistic Dimensions in LIWC



Filename	Segment	WC	Analytic	Clout	Authentic	Tone	WPS	BigWords	Dic	Linguistic	function	pronoun	ppron	i	we	you	shehe	they	ipron	det	article	number
IS1.docx	1	978	89.66	46.72	1	26.96	18.11	40.49	6.95	1.64	0.92	0.1	0.1	0	0.1	0	0	0	0	0.61	0	2.45
IS2.docx	1	959	87.22	93.4	23.16	88.95	19.98	28.47	76.75	53.08	36.18	7.4	6.15	0.21	0.42	5.11	0	0.1	1.25	12.41	5.74	3.75
IS3.docx	1	3841	87.59	81.73	27.75	85.16	22.86	27.94	79.09	55.27	41.16	7.06	5.44	0.21	0.86	4.01	0.05	0.21	1.61	13.54	6.69	6.22
IS4.docx	1	2154	90.27	80.51	7.82	84.11	20.51	27.67	72.66	47.4	32.68	5.48	4.55	0.23	1.62	2.09	0	0.56	0.93	11.84	5.2	8.82
IS5.docx	1	5186	98.22	45.42	16.2	93.71	21.61	27.55	80.47	57.98	46.91	4.7	2.24	0.15	0.67	0.71	0.44	0.19	2.47	18.2	12.57	3.97
IS6.docx	1	943	76.08	92.74	19.4	70.25	19.24	26.51	83.24	63.84	49.2	10.6	7	0.32	2.97	2.97	0	0.53	3.61	14	6.89	3.08
IS7.docx	1	680	95.43	47.91	12.44	85.53	17.44	43.09	65.74	38.68	25.74	1.76	0.59	0	0.15	0.44	0	0	1.18	7.21	5.15	6.18
IS8.docx	1	3406	89.11	56.17	21.17	78.2	22.71	29.83	81.5	56.4	43.89	5.4	2.88	0.32	0.76	0.56	0.59	0.53	2.52	15.62	7.9	6.61
IS9.docx	1	757	88.2	87.31	15.6	63.94	24.42	31.84	75.56	52.58	35.8	5.81	5.28	0.13	1.06	4.1	0	0	0.53	11.23	5.15	5.42
IS10.docx	1	4598	82.99	66.05	44.96	45.42	18.69	24.34	82.67	64.31	48.11	7.79	4.35	0.13	0.59	3.04	0	0.54	3.44	15.14	8.94	3.04

FIGURE 4. Linguistic Dimensions from LIWC in a spreadsheet

When a word from a document matches a construct, a running total will be kept (Figure 1). After reading and counting each word in the document, the ratio of words in a construct divided by all words was determined (Figure 2). These factors were divided into the five linguistic dimensions, which are *Lifestyle*, *Social Process*, *Cognition*, *Affect,* and *Perception* (Figure 3), and were kept in a spreadsheet (Figure 4) for statistical analysis. The results from LIWC were then analysed in the Statistical Package for the Social Sciences (SPSS) software to identify the pattern of linguistic cues of deception by forming descriptive profiles and the correlation between all cues.

eISSN: 2550-2131
ISSN: 1675-8021

FINDINGS

Table 2 shows the total mean score (M=11.008, SD=6.176) for the *Lifestyle* dimension. Three categories were identified based on the descriptive analysis for this dimension. The highest category for *Lifestyle* is *money* (M=7.171, SD=4.562), followed by *work* (M=6.739, SD=4.248). The lowest mean score (M=0.021, SD=0.066) is *religion.* This dimension reveals that the scammers prioritised *Lifestyle* by emphasising this cue to entice and deceive their victims into falling for their scam. The expression of phrases like *'Make your trading **extra profitable**!'*, '*Double your deposit, get **more profit up to 100%***', *'providing **a stable income** in the range between 8% to 15% a month',* and *'Multiply your **profit potential 2,000 times**'* were used to mask the glorious promises of overemphasis on guaranteed high returns and bonuses with low or almost no risk. Furthermore, the use of religious sentiments created a sense of security and value-added reliability of the promotional materials. Phrases like *'**Islamic** account', '**Islamic** swap-free accounts'*, and *'**Muslim** traders'* were used to bring in affinity to mislead investors' decision to join the investment. These overly exaggerated phrases were used to lure potential investors with unrealistic profit claims and were typically considered a red flag as indicators of deception.

TABLE 2. Linguistic Cues for Lifestyle

Dimension	Example	Mean	Std. Deviation
Lifestyle		**11.008**	**6.176**
money	Deposited, credited, cent accounts, accounts registered, USD, transfer-ins, funded, currency market, awards, deposit bonus, compensated, investments, investing, price, bonus funds, sales, paid, inexpensively	7.171	4.562
work	Trader, profitability, trading operations, accrued, trading account, execution, trading account equity, register, residents, services, sports foundation, bonus application, retail, work, reports, industry, Brokerage house, payroll services, institutional organisations, business finances, administrative task	6.739	4.248
religion	Islamic, Islamic account, Muslim traders	0.021	0.066

The findings in Table 3 show the total mean score (M=7.876, SD=4.038) of linguistic cues for the *Social* dimension. The *social reference* category scored the highest mean score (M=3.999, SD=2.236) for the *Social* dimension. *Social references* such as *ours, we,* and *us* were used to establish a sense of security and belonging to provide an impression that the investment plan is safe to join. For example, they will use mottos like '*The more you **trade** with us, the better it gets*', '***Trade** Your Way*', and '***Trade** Safe, **Trade** with Us*' to influence and persuade victims' decision-making to participate in their scheme. These cues revealed the scammers' strategies to establish social engagement and rapport with potential victims. The use of social reference words is also presumed to channel the focus on potential investors by creating a false sense of excitement on the excessive emphasis on returns as well as personal and affiliate rewards. Despite the typical detachment of scammers from the deceptive acts they commit (De Paulo et al. 2003), the frequent use of second-person pronouns such as *you* and third-person pronouns like *he, him,* and *her* suggests that they take a more cautious approach in crafting their fraudulent acts. Expressions such as *'Magnify **your** income by 10 times'* and *'**You** can generate profits even if you have zero knowledge about crypto'* were instances of this tactic.

On the other hand, the s*ocial behaviour* category scored the second highest (M=3.495, SD=2.208) for this dimension. Expressions such as *'**We offer** a chance for you to take maximum advantage of the rewards on these promotions', 'Our goal is to **welcome** new clients*

*and **reward our existing clients**, with exclusive promotions'*, and *'Multiply your profits is our **commitment**'* are examples of scammers' attempts to signal caring social behaviour at the interpersonal level. The lowest mean score (M=0.007, SD=0.221) for the *Social* dimension is the *conflict and family* category. It is evident in this study that scammers seem to avoid using expressions that might create a red flag and cause scepticism among potential investors. An expression like *'Our **company** will not be **liable** for any **losses** that investors may incur by trading with the bonus'* was detected in only a few promotional materials and written in small font and hard-to-read places.

TABLE 3. Linguistic Cues for Social

Dimension	Example	Mean	Std. Deviation
Social		**7.876**	**4.038**
Social Reference	His/her, your, you, our, us, we'll, our, he, their, we,	3.999	2.236
Social Behavior	Campaign, meet, we offer, claim, please, welcome, offer, request, disclosure, described, advised, met, stated, exchange, brought, commitment, give them, provide them,	3.495	2.208
Conflict & family	Client, company, loyalty, our promotions, warning, written, you agree, personal, help, reject client's, client's personal, help you, your referrals, relationship, customers, employees,	0.007	0.221

Table 4 depicts the analysis of linguistic cues for the *Cognition* dimension. The findings show that *cognition* has the highest mean score (M=5.390, SD=3.450), followed by the *different* category (M=1.705, SD=1.707). It is identified that the use of complex language, such as ambiguous words, vague terminology, complex jargon, impersonal language, and buzzwords in the promotional materials, created an impression of sophistication and sounded more legitimate than it is. For instance, the sentences *'The company is not responsible for **any consequences** of the bonus cancelling including the stop out cases since the **accrued bonus** remains the company's entire property'*, '*Leverage products may not be suitable for everyone and may result in loss of all your capital. Please ensure you fully understand the risks involved and whether trading is appropriate for you*', and *'This promotion **cannot be combined** with other bonus promotions* might confuse potential investors to fully understand the investment's terms and conditions. This tactic seems to be an attempt to obfuscate details by presenting a lack of transparency and clarity about the investment operations.

The *certitude* category scored the lowest mean score (M=0.175, SD=0.220) for this dimension. Examples of words for *certitude* are *not allowed, understand*, and *activate.* These words were replacements for the original cognitive processing dimension of certainty. They appeared to mirror a degree of bravado or bragging of certainty, which is unverifiable and misleading. Examples of the sentences that used these words are *'The 100% Bonus is **not allowed** for withdrawal, however, there is no limit on withdrawing the profit gained from the bonus', 'You should consider whether you **understand** how CFDs work and whether you can afford to take the high risk of losing your money'* and *'**Activate** your free 50% bonus on the amount you deposit'.*

TABLE 4. Language Cues for Cognition

Dimension	**Example**	**Mean**	**Std. Deviation**
Cognition		**6.018**	**3.832**
Cognition	Any consequences, or, no, any, may need, may change, but not, all, option, cannot, can,	5.390	3.450
Different	Other, another, than, however, since, different purpose, nonetheless, aware, using any, otherwise, different, could utilised, methods, various,	1.705	1.707
Certitude	Not allowed, no, exclude, or, need, all, especially, can always, exclusive, potential, permission, should completely, notice, activate, know, decide, mind, want, relevant, should consider understand how, allows, leads, originating, enables,	0.175	0.220

Table 5 illustrates the mean score (M=4.814, SD=2.837) of linguistic cues for the *Affect* dimension. Descriptive analysis has shown that *positive tone* has the highest mean score (M=3.656, SD=2.128) for linguistic cues for *Affect*, followed by *negative tone* (M=1.158, SD=0.971). A point to note here is that LIWC conceptualised the dictionaries of positive and negative tones as reflections of sentiment rather than emotion per se and included words related to positive and negative emotions. The accounted evidence from the analysis demonstrates that the studied promotional materials rely heavily on the use of words that represent *positive tone* like *bonus, rewarded, benefit, well*, and *free* to leave impressive persuasive elements in making investment decisions.

Words representing *negative tone*, such as *warning, exhausting, losing,* and *lose* were also used. However, the occurrences were less compared to *positive tone* words, particularly in creating false images of poverty, financial hardships, and fear of missing out. It appears that this strategy attempts to create psychological manipulation to cloud rational judgment among investors.

Interestingly, although *positive tone* scored the highest mean score, *emotion-positive* category, representing the true emotion labels, scored the lowest mean score (M=0.047, SD=0.123). Words associated with *emotion-positive* like *successful, strongly recommended, keep, hit, goal,* and *new* rarely occurred across all the studied documents, except for a few lines such as *'**Successfully** helping many who have lost in the market due to the market uncertainty'.*

TABLE 5. Language Cues for Affect

Dimension	**Example**	**Mean**	**Std. Deviation**
Affect		**4.814**	**2.837**
Positive Tone	Bonus, bonuses, rewarded, benefit, well, free, new, achieve, reliable, smooth,	3.656	2.128
Negative Tone	Warning, exhausting, losing, lose,	1.158	0.971
Positive Emotion	Strongly recommended, keep, hit, goal, new, accept, smart, strive, strong, successful	0.047	0.123

Meanwhile, Table 6 shows the descriptive analysis and the mean score (M=4.126, SD=2.435) of language cues for the *Perception* dimension, which is the least frequently used cue identified in the promotional materials. A noteworthy point here is that scammers appeared to use *Perception* words as an attempt to distance themselves from being fictional to increase the believability of their deception. Descriptive analysis shows that *Space* scored the highest mean score (M=3.390, SD=2.047). The analysis provided examples of words used in the *Space* category, such as *opened, level, up, into, wide* and *closely*. It can be seen in the sentences *'Sign up to be an Introducing broker (IB) with the most awarded broker in the industry and enjoy*

the highest commissions, ***up*** *to 60% of the spread, for every client's trade'* and *'****Worldwide*** *trusted trading platform'.*

Apart from that, *Attention* scored the second highest mean score (M=0.371, SD=0.431). Words representing *Attention* like *carried out, remains, reach, go,* and *full* were used to create the illusion that the potential investors were given attention and priority, as demonstrated in the example, *'Once everything is set, you just need to* ***sit back*** *and* ***carry on*** *with your daily routine'*. This tactic is presumably intended to ensure a convenient and worry-free investment process.

The lowest mean score (M=0.077, SD=0.1008) for this dimension is the *Visual* category. Words related to *visual* such as *world, higher, images*, and *international* were not frequently used in most of the studied documents. Nevertheless, it is evident that scammers attempt to convince investors of the legitimacy of the promoted investment schemes and appear trustworthy, as shown in this example, *'More than 7,000,000 traders from all over the* ***world*** *have chosen xxx and proved our title of* ***international*** *broker'.*

TABLE 6. Language Cues for Perception

Dimension Perception	**Example**	**Mean 4.126**	**Std. Deviation 2.435**
space	Opened, level, in, there, up, into, back, below, center, here, wide, closely,	3.390	2.047
attention	Carried out, remains, reach, go, full, rapidly, ensure,	0.371	0.431
visual	Detects, within, place within, higher, running, location, run, local, world, internal, images,	0.077	0.1008

A correlation analysis was done for each dimension to assess whether the measurements of linguistic cues correlated with one another. Table 7 shows a correlation analysis for linguistic cues. The findings show no significant correlation between *Cognition* and *Affect* (r=0.582, p=0.078). Correlation between *Social* and *Cognition* shows a highly significant correlation where r=0.642, p=0.046. Linguistic cues for *Social Process* also exhibit a high correlation between *Social Process* and *Affect* (r=0.684, p=0.029). The *Lifestyle* dimension significantly correlates between *Cognition* (r=0.772, p=0.009) and *Social Process* (r=0.689, p=0.028). However, there is no significant association between *Lifestyle* and *Affect* (r=0.553, p=0.097).

Nevertheless, a surprising result is that although *Perception* was the least used linguistic cue, the correlation analysis revealed that this dimension is substantially associated with other linguistic cues. The correlation analysis shows that *Perception* has a highly significant correlation between *Cognition* (r=0.835, p=0.003), *Affect* (r=0.697, p=0.025), *Social Process* (r=0.833, p=0.003), and *Lifestyle* (r=0.801, p=0.005).

TABLE 7. Correlation analysis for linguistic cues

		(1)	**(2)**	**(3)**	**(4)**	**(5)**
Cognition (1)	Pearson Correlation	1				
	Sig. (2-tailed)					
Affect (2)	Pearson Correlation	0.582	1			
	Sig. (2-tailed)	0.078				
Social Process (3)	Pearson Correlation	0.642*	0.684	1		
	Sig. (2-tailed)	0.046	0.029			
Lifestyle (4)	Pearson Correlation	0.772*	0.553	0.689*	1	

	Sig. (2-tailed)	0.009	0.097	0.028		
Perception (5)	Pearson Correlation	0.835**	0.697*	0.833**	0.801**	1
	Sig. (2-tailed)	0.003	0.025	0.003	0.005	

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

DISCUSSION

This research was conducted largely to demonstrate the utilisation of automated linguistic cue identification to flag potentially deceptive discourse and to use statistical calculations to analyse the correlation patterns of the deceptive cues. The analysis depicts the prospect of a general focus on linguistic cues as indicators for predicting possible deception in promotional materials. As for the specific identified linguistic cue dimensions that might be potential deception indicators, these findings provide evidence for the hypothesis that scammers presented their linguistic behaviour differently from legitimate investment schemes by applying a particular strategy. A recognisable pattern shown in the linguistic presentation in the promotional materials; however, the analysis might need to be expanded to a larger sample size for a more comprehensive conclusion to be made. Nevertheless, the findings illustrated that the most frequently used cue was *Lifestyle*, followed by *Social Process, Cognition, Affect*, and *Perception* dimensions.

It is noticeable that exaggerated guarantees of high returns and emphasis on religious sentiment were the main persuasive and manipulative tactics to entice potential investors to join the investment schemes. Furthermore, the findings also recognized that the scammers' messages were briefer and significantly raised red flags, particularly on the use of complex technical jargon and terminology, a lack of specificity, transparency, and clarity of investment details, unverifiable information and accomplishments in their promotional materials. It occurs that these linguistic behaviours showed ambiguity, raised suspicions, and demonstrated scammers' inability to create truthful, detailed and convincing discourse. However, the correlation analysis done in this study shows that the dimension of *Perception* is related to the dimensions of *Lifestyle, Social Process, Cognition*, and *Affec*t, implying that the cues were used in tandem to form convincing narratives of the scams. This may help future research to focus on the complexities of deception and investigate markers across all dimensions that are significant predictors for the early detection of scams.

While the classification of the analysis done in this study is far from perfect, the revelation of linguistic cue identification in deceptive materials in online investment scams may provide the first step in possible deception detection through textual analysis. The five dimensions identified from the linguistic cues for deception suggest that using words in online investment scams' promotional materials intentionally maximizes their opportunity to lure their victims by convincing and persuading them to participate in the investment schemes. By employing engineered sentences and claims that appeal to the victims' desire, scammers can deceive the victims through the grossly inaccurate depictions of investments.

Although, in time, there is a great tendency for scammers to manoeuvre these patterns to be more closely similar to legitimate materials, it is possible that the linguistic behaviours of deceptive materials may provide scepticism because they are not addressing real details, making it possible for linguistic footprint detection. In addition, when the illusion of a successful investment scheme can be discredited by potential investors via methodical consideration of the scammers' engineered "sweet words", the risk of financial losses incurred by the scams can be greatly reduced. Staying vigilant and attuned to these linguistic cues is essential in the ongoing battle against deception in the digital age.

CONCLUSION

This study has examined the linguistic cues for deception in Malaysian online investment scams' promotional materials by describing their occurrences and analysing whether they are related to one another in forming a unique deception strategy. By analysing linguistic cues in these communications, experts can identify red flags and warning signs to identify suspicious language patterns, excessive persuasion techniques, high-pressure tactics, vague or misleading information, or inconsistent statements as indicators of fraudulent intent. At the same time, by leveraging linguistic cues and applying computational methods, systems can be developed to analyse large volumes of text data automatically, detect patterns associated with scams, and provide early warning or mitigation measures to protect individuals from fraudulent activities. Markers of all dimensions identified were significant predictors for early recognition of scams.

However, the research presented in this paper is insufficient to provide comprehensive information for conclusive scam detection and prevention. The current study's findings on the linguistic cues used in online investment scams are based on descriptive and correlation analysis. A limitation of this study is that the correlational design does not include reasoning for the basis of the linguistic cues' occurrences, and further investigation is recommended to gain a more thorough understanding of the concern. These findings only provide the basis for the primarily used linguistic cues in online investment scams, which do not perfectly identify whether the online investment is a scam. Further studies should be conducted by including behavioural aspects as a dependent variable to measure how linguistic cues can influence people to believe in online investment scams. Multiple regression analysis should be used to further analyse the linguistic cues of deception in Malaysian online investment scams.

Furthermore, while the findings are restricted to accessible data that was analysed to address the research question posed in this study, this study has great potential for generalisability. A more comprehensive understanding of a text is needed by considering the stylistic and psychological dimensions for more concrete guidance in determining and identifying deception in scams. Language texts could be statistically analysed using stylometry techniques, a branch of computational linguistics in authorship attribution. These techniques provide valuable tools for analysing texts, identifying authors, understanding writing styles, and contributing to various fields of study.

This study has elucidated a preliminary foray into detecting the linguistic cues and their correlation in online investment scams' promotional materials and could contribute to developing a possible comprehensive linguistic model for scam detection.

AUTHORS' CONTRIBUTIONS

The contributions of the authors to this manuscript are as follows: Ameiruel Azwan Ab Aziz oversaw project administration, played a key role in conceptualisation, contributed to methodology design, was involved in conceiving and designing the analysis, participated in the writing process, and provided valuable review and editing. Nurul Atiqah Mohd Sharif was responsible for data collection and curation, engaged in data visualization, and played a significant role in the initial draft preparation. Amirah Mohd Juned contributed to the project's conceptualization and methodology, taking part in the formal analysis. Wan Farah Wani Wan Fakhruddin contributed to the writing process by reviewing and editing the manuscript, validating the findings, and providing valuable resources. Nursyaidatul Kamar Md Shah and Ariff Imran Anuar Yatim participated in data collection and curation, data visualization, and the review and editing of the manuscript. Aminabibi Saidalvi played a significant role in the conceptualisation and securing of funding for the project. It is important to note that all authors have thoroughly reviewed and agreed to the content of the published version of the manuscript.

eISSN: 2550-2131
ISSN: 1675-8021

ACKNOWLEDGEMENT

This research was supported by the Ministry of Higher Education through the Fundamental Research Grant Scheme (Grant No. FRGS/1/2022/SSI0/UITM/02/10) and Universiti Teknologi MARA.

REFERENCES

Addawood, A., Badawy, A., Lerman, K., & Ferrara, E. (2019). Linguistic Cues to Deception: Identifying Political Trolls on Social Media. *Proceedings of the International AAAI Conference on Web and Social Media*, *13*(01), 15-25. https://doi.org/10.1609/icwsm.v13i01.3205

Adha, A. (2020). Linguistic Based Cues in Detecting Deception in Indonesian Language Use. *Argumentum (1787-3606)*, (16). https://doi.org/10.34103/argumentum/2020/2

Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). The development and psychometric properties of LIWC-22. Austin, TX: University of Texas at Austin. https://www.liwc.app

Bukit Aman says investment fraud cases surged in 2021, losses exceed RM420m. (2022, January 5). *Malay Mail*. Retrieved from https://www.malaymail.com/news/malaysia/2022/01/05/bukit-aman-says-investment-fraud-cases-surged-in-2021-losses-exceed-rm420m/2033251

Central Bank of Malaysia (2023). Retrieved from https://www.bnm.gov.my//financial-consumer-alert-list

Choudhury, F. (2014). Can language be useful in detecting deception? The linguistic markers of deception in the Jodi Arias interview.,,,. *Journal of Undergraduate Research*, *7*(2), 78-92.

Connell, C. (2012). Linguistic Cues to Deception (Doctoral dissertation, Virginia Tech). Retrieved from https://vtechworks.lib.vt.edu/handle/10919/32465

Deora, R. S., & Chudasama, D. (2021). Brief study of cybercrime on the internet. *Journal of communication engineering & Systems*, *11*(1), 1-6. https://doi.org/10.37591/JoCES

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129*(1), 74–118. https://doi.org/10.1037/0033-2909.129.1.74

Frankel, T. (2012). The Ponzi scheme puzzle: A history and analysis of con artists and victims. Oxford University Press.

Fuller, C. M., Biros, D. P., & Wilson, R. L. (2009). Decision support for determining veracity via linguistic-based cues. Decision Support Systems*, 46*(3), 695-703.

Genao, D. R. (2021). Identification of Fraudulent Financial Statements: The Detection of Deception and Collusion in Earnings Calls (Doctoral dissertation, Northcentral University). Retrieved from ProQuest Dissertation https://www.proquest.com/docview/2551555139?pqorigsite=gscholar&fromopenview=true

John, S. (2018, September 18). *CNBC*. Retrieved from Religious-based financial fraud is rampant. Here's how to fight it: https://www.cnbc.com/2018/09/06/religious-based-financial-fraud-is-rampant-heres-how-to-fight-it.html

Karim, N. A., Ab Wahid, Z., Ariffin, S. N. K., Nor, S. H. S., Nazlan, A. N., & Kassim, S. (2023). Financial Literacy among University Students and its Implications towards Financial Scams. *Information Management and Business Review*, *15*(3 (I)), 124-128.

Levitan, S. I., Maredia, A., & Hirschberg, J. (2018, June). Linguistic cues to deception and perceived deception in interview dialogues. In *Proceedings of the 2018 Conference of*

the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers) (pp. 1941-1950).

Li, L., Lee, K. Y., Lee, M., & Yang, S.-B. (2020). Unveiling the cloak of deviance: Linguistic cues for psychological processes in fake online reviews. *International Journal of Hospitality Management*, *87*, 102468. https://doi.org/10.1016/j.ijhm.2020.102468

Mihalcea, R., & Strapparava, C. (2009, August). The lie detector: Explorations in the automatic recognition of deceptive language. In *Proceedings of the ACL-IJCNLP 2009 conference short papers* (pp. 309-312).

Mohd Padil, H., Kasim, E. S., Muda, S., Ismail, N., & Md Zin, N. (2021). Financial literacy and awareness of investment scams among university students. *Journal of Financial Crime*, *29(*1), 355-367. https://doi.org/10.1108/jfc-01-2021-0012

Naksawat, C., Akkakoson, S., & Loi, C. K. (2016). Persuasion strategies: use of negative forces in scam e-mails. *GEMA Online Journal of Language Studies*, *16*(1), 1-17.

New Straits Times. (2022, August 4). RM5.2b in losses through online scams since 2020. Kuala Lumpur, Malaysia. Retrieved from https://www.nst.com.my/news/crime-courts/2022/08/819331/rm52b-losses-through-online-scams-2020

Newman M. L., Pennebaker J. W., Berry D. S., Richards J. M. (2003). Lying words: Predicting deception from linguistic style. *Personality and Social Psychology Bulletin*, 29, 665–675. https://doi.org/10.1177/0146167203029005010

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: language use as an individual difference. *Journal of personality and social psychology*, *77*(6), 1296.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*.

Rahman, A. A., Azmi, R., & Mohd Yusof, R. (2020). Get-Rich Quick scheme: Malaysian current legal development. *Journal of Financial Crime*, *28*(1), 49-59.

Rubin, V. L., & Conroy, N. J. (2011). Challenges in automated deception detection in computer-mediated communication. *Proceedings of the American Society for Information Science and Technology*, *48*(1), 1-4.

Schillermann, M. K. (2018). *Early detection and prevention of corporate financial fraud* (Doctoral dissertation, Walden University).

Securities Commission Malaysia (2020). *Annual Report 2020*. Retrieved October 19, 2023, from https://www.sc.com.my/api/documentms/download.ashx?id=e1c7eb21-53db-4f02-a8f8-55dc09f9fff

Shaari, A. H., Kamaluddin, M. R., Fauzi, W. F. P., & Mohd, M. (2019). Online-dating romance scam in Malaysia: An analysis of online conversations between scammers and victims. *GEMA Online Journal of Language Studies*, 19(1), 97-115

Toma, C. L., & Hancock, J. T. (2010, February). Reading between the lines: linguistic cues to deception in online dating profiles. In Proceedings of the 2010 ACM conference on Computer supported cooperative work (pp. 5-8). https://doi.org/10.1145/1718918.171892

Van Der Zee, S., Poppe, R., Havrileck, A., & Baillon, A. (2020). A Personal Model of Trumpery: Linguistic Deception Detection in a Real-World High-Stakes Setting. Psychological Science, *33*(1), 3–17. https://doi.org/10.1177/09567976211015941

Zhou, X., & Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)*, *53*(5), 1-40. https://doi.org/10.1145/3395046

Zhou, L., & Zhang, D. (2008). Following linguistic footprints: Automatic deception detection in online communication. *Communications of the ACM*, *51*(9), 119-122.

ABOUT THE AUTHORS

Ameiruel Azwan Ab Aziz (Ph.D.) is a senior lecturer at the Academy of Language Studies, Universiti Teknologi MARA. His research interests include teaching English as a second language (TESL), psychology of language teachers, and forensic linguistic analysis. He has published his work extensively in reputable journals, both locally and internationally.

Nurul Atiqah Mohd Sharif (MA) is a postgraduate student of Masters in Applied Language Studies currently in her Semester 2 at the Academy of Language Studies, Universiti Teknologi MARA (UiTM), Melaka. Her research interests are in the area of semantics and morphology.

Wan Farah Wani Wan Fakhruddin is a Senior Lecturer in the Faculty of Social Sciences and Humanities at Universiti Teknologi Malaysia. Her research interests include a functional approach to language and discourse studies, with an emphasis on professional and academic genres from a Systemic Functional Linguistics perspective.

Amirah Mohd Juned is a lecturer at UiTM Cawangan Melaka. She obtained her Doctor of Philosophy (Education) from Universiti Teknologi MARA, Shah Alam. Her research interest is in the area of motivation, teacher instructional practices and educational psychology.

Nursyaidatul Kamar Md Shah is an accomplished English language lecturer at the Academy of Language Studies, UiTM Melaka. With a strong passion for applied linguistics, language studies, and psycholinguistics, she is currently dedicated to pursuing her Ph.D. in applied language studies to further contribute to the field.

Ariff Imran Anuar Yatim is a lecturer at the Academy of Language Studies, University Teknologi MARA (UiTM) Melaka, Malaysia. He has a Bac. of Edu (Hons.) in TESL from UiTM and a Master of Applied Linguistics from Universiti Putra Malaysia. His research interests include communication strategies, semantics, and pragmatics.

Aminabibi Saidalvi was an Associate Professor at the Academy of Language Studies, Universiti Teknologi MARA Johor (UiTM), Pasir Gudang Campus, until June 12, 2023. Her research interests include TESL, Online Integrated Language Learning, Oral Communication Skills, and Second Language Research.

eISSN: 2550-2131
ISSN: 1675-8021