

A Genre, Scoring, and Authorship Analysis of AI-Generated and Human-Written Refusal Emails

Business and Professional
Communication Quarterly
1–22

© 2025 by the Association for
Business Communication



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/23294906251322890
journals.sagepub.com/home/bcq



Winnie Wilson¹  and Heath Rose¹ 

Abstract

This study compares AI-generated (ChatGPT and Gemini) and human-written business refusal texts. A genre analysis found that AI-generated texts are formulaic and less nuanced than human-written texts. Applying a rating of professional writing quality, inferential statistics revealed no significant difference in scores between Gemini and human-written texts, but revealed ChatGPT as lower. Human assessors identified authorship of AI-generated texts with an accuracy rate of 68.1%, and human-written texts with 86% accuracy. Key concerns for assessors were tone, relationship, language choice, content, and structure. The findings inform four key areas of focus for teaching business writing in the AI age.

Keywords

Generative-AI, business writing, business communication, genre

Introduction

Training in written communication skills is in high demand, as organizations seek to help their employees improve the quality and effectiveness of their writing. Many business writing courses include a large component on the writing of negative messages, such as when a writer needs to refuse a reader's request. These messages are a particularly challenging aspect of business correspondence, as considerable skill is

¹University of Oxford, UK

Corresponding Author:

Heath Rose, University of Oxford, 15 Norham Gardens, Oxford, OX1 2JD, UK.
Email: heath.rose@education.ox.ac.uk

required to ensure the act of refusal does not damage the relationship between writer and reader. It is thus very important for people engaging in business communication to have an awareness of this genre.

Updated research on business genres is particularly needed with the emergence of generative artificial intelligence (AI) tools. AI tools are able to generate natural-sounding text in response to a prompt, but how do the texts they generate compare to human-written ones? In other words, if AI-generated texts are equally as well-written as human texts, is there still a case for teaching this genre in business written communication training?

At the time of writing, there were no published studies comparing human-written and AI-generated texts in a business context. This exploratory study investigates business refusal texts from three writer groups: ChatGPT, Gemini, and Human. First, genre analysis is employed to understand the typical moves and steps of texts from each writer group and the differences between each. Next, human assessors are asked to blind-evaluate the texts by scoring them based on established business writing criteria, with the aim of finding out whether AI tool-generated texts are considered higher-quality than human writing. To gather information on whether there are identifiable and characteristic qualities of AI-generated versus human-written texts, participants are asked to identify authorship of the texts and explain what influences their decisions. Insights are used to make recommendations for business writing teachers and identify future directions for research at the intersection of generative AI and professional writing.

Literature Review

Genre Research in the Business Disciplines

Few people have explicit knowledge of the rhetorical and formal features of everyday texts. Hence, genre analysis is an important tool for providing understanding of how texts in a target genre are structured to engage effectively with readers (Hyland, 2022). Descriptions of how genres are used in specific discourse communities can function as models for writing instructors to apply to help learners achieve their communicative goals (Bhatia, 1993; Swales, 1990). In particular, frameworks have been developed by analyzing a corpus of texts considered representative of the genre of interest, to identify moves and steps that carry out distinct functions within the text.

John Swales (1990, 2004) conducted pioneering work on analyzing the organizational patterns of writing genres. Move analysis organizes common moves, steps, and text boundaries to give insight into how writers achieve their purpose (Swales, 1990; Vergaro, 2004). In addition to consideration of whether moves are obligatory or optional, genre analysts have considered their importance by referring to the percentage of texts in which a move appears to inform discussion about how typical, predominant, or conventional certain moves are in a genre (Biber et al., 2007; Henry & Roseberry, 2001). The information gained from move analysis can thus help writers

make decisions on how they might structure their texts, taking into account the conventions of the discourse community in which their genre is situated.

While there is a healthy literature on genre-based approaches to teaching writing, almost all studies focus on academic writing in science and social science-based disciplines. Zhai and Razali's (2023) recent systematic review found that of 54 articles published since 2003 on genre-based approaches, only 2 articles focused on the business subject area. Our study aims to fill this gap, as well as to address the future of writing in the generative-AI era.

Refusals and Politeness Strategies

There is a wide range of communicative purposes within business genres, with requests featuring prominently in previous studies (e.g., Kong, 1998; Nguyen & Miller, 2012; Park et al., 2021). Few studies in the professional writing literature have considered what happens when a request needs to be refused. Writing negative messages such as refusing a request is difficult and anxiety-inducing (Schryer, 2000), but a skill critical to interpersonal effectiveness in the workplace (de Rycker, 2014). The act of refusal can potentially damage the relationship between writer and reader and requires considerable thought and skill in constructing the message. As refusals are face-threatening acts, it is important to employ "politeness" strategies for lessening the threat and maintaining social harmony (Brown & Levinson, 1987). These strategies include the use of explanations and reasons to maintain goodwill when composing negative messages (Campbell, 1990).

Schryer (2000) identified the rhetorical organization of insurance claim denial letters as following a traditional structure: a neutral buffer opening, the policy explanation, the medical explanation, the decision, and a closing section. Textbooks mostly recommend this indirect approach to maintain the recipient's face needs and goodwill, through a sequence of rhetorical moves typically involving an explanation and rationale before presenting the negative news (Creelman, 2012; De Rycker, 2014). The explanation-first order has been found to have a positive effect on reader evaluations of correspondence (Jansen & Janssen, 2010, 2013; Shelby & Reinsch, 1995). Thus, genre analysis can reveal useful insights into such text patterning, and the typical performances of the genre can form the basis of professional writing instruction.

Generative AI and the business context

Any study of business writing conducted since the launch of the hugely popular ChatGPT in November 2022 needs to consider the use of generative artificial intelligence (AI) tools built on large language models (LLMs). Business practitioners are increasingly using AI tools for purposes including the drafting of business messages (Cardon, Fleischmann, Logemann, et al., 2023). However, research is concentrated in the academic field, particularly around the applications and implications of ChatGPT in education (e.g., Bhullar et al., 2024; Rudolph et al., 2023). Few studies have evaluated its use in professional contexts, particularly for generating business emails.

AlAfnan et al. (2023) used established rubrics from university courses to assess AI-generated texts based on prompts from composition, business writing, and communication university courses. They found that in response to a prompt asking for a reply to an angry customer, ChatGPT-generated emails were formulaic and templated, rather than personalized. Jovic and Mnasri (2024) compared AI-generated emails (a routine complaint, a negative message, and a persuasive message) from four freely available LLMs (Bard, Bing Chat, ChatGPT 3.5, and Llama 2) to assess whether the content generated was clear, concise, accurate, and contextually relevant. Similar to AlAfnan et al. (2023), they found the generated texts fell short in providing sufficient supporting evidence, information, and detail. These studies were limited by the analysis of a very small number of texts, which were rated by only one assessor in each instance. Another major limitation is the lack of comparison between AI-generated and human-written emails. It is undoubtedly valuable to study the features of AI-generated texts and make comparisons between different LLMs. However, it is a glaring omission not to compare these to texts written by humans, whose writing the AI tools were trained to emulate.

Human and AI Authorship

Few published studies have yet explored whether humans can distinguish between human-written and AI-generated business texts. Previous studies into texts in other domains have found that human assessors experienced difficulty in determining authorship. Gunser et al. (2021) investigated literary texts generated by GPT-2 (an earlier model preceding what now powers ChatGPT) and written by humans. Nine human assessors with literature-specific backgrounds were asked to decide which texts were AI-generated versus human-written, and they misclassified over 26% of all texts. Another study found that human assessors were incapable of reliably detecting AI-generated poetry, with an average accuracy rate of 50.21%, or no deviance from chance (Köbis & Mossink, 2021).

Regarding academic texts, Casal and Kessler (2023) concluded that human assessors were not particularly effective in distinguishing between AI-generated and human-written research abstracts, with an overall positive identification rate of only 38.9%. However, the reviewers were more effective at identifying human authors of research abstracts (44.1% accuracy rate) than AI authors of abstracts (33.7%). Still in the academic context, Yeadon et al.'s (2024) research into authorship decisions on 300 AI-generated and human-written university physics essays by human evaluators ($n=5$) found an average accuracy rate of 62.4%, which the authors reported as only marginally better than random chance.

Studies on authorship have been limited by either small numbers of human assessors or of texts evaluated, or both. There is little data on what factors influence human assessors when deciding whether a text is AI-generated or human-written, with only Casal and Kessler (2023) collecting qualitative data to gain insight into rationales for authorship decisions. Existing studies conducted in the academic realm may not generalize to the business context. Overall, it remains unclear whether AI-produced texts

can emulate the human empathy needed for relationship-building text types such as face-saving refusal messages.

AI Literacy

AI literacy refers to the ability to understand, evaluate, and effectively use AI tools, particularly generative AI, in communication. Previous research in business communications has found AI literacy encompasses four key components: *Application* (knowing how to use AI tools for specific tasks), *Authenticity* (maintaining human voice, audience awareness, and trust), *Accountability* (ensuring accuracy, ethical use, and responsibility for AI-generated content), and *Agency* (retaining control over decision making and not overrelying on AI) (Cardon, Fleischmann, Aritz 2023). Developing AI literacy helps professionals navigate both the opportunities (e.g., enhanced productivity) and challenges (e.g., inaccuracy, bias, and ethical concerns) associated with AI in workplace communication (Getchell et al., 2022). An examination of the strengths and limitations of AI-produced communications, when compared to human-produced communications, is an essential first step to understanding how future professionals can apply their AI literacy to writing in the workplace, while maintaining an authentic voice, accountable processes, and agency over content.

Research gap

The review of the literature shows that research on business genres and AI-generated writing in the professional context needs updating. Our study adopts a mixed-method research design to answer the following research questions:

RQ1: What are the moves and steps that typify the business refusal genre? Are there differences between AI-generated and human-written texts?

RQ2: How are AI-generated and human-written texts evaluated by human assessors using an established business writing rubric?

RQ3: Can human assessors identify authorship of texts (AI-generated or human-written)? What influences their decisions?

We answer these questions by analyzing human writing alongside text generated by two of the AI tools most commonly used in the workplace: ChatGPT 3.5 and Gemini. We also sought the help of business writing teachers to evaluate the texts.

Materials and Methods

The Texts

The first step was to collect texts for both the genre analysis (RQ1) and evaluation by human assessors (RQ2 and RQ3). For this, two business scenarios were developed:

- (a) an internal communication, instructing the writer to refuse a colleague's request to deliver a workshop on their behalf; and
- (b) an external communication, instructing the writer to refuse a client's request to increase the number of participants on a workshop beyond the maximum class size.

For each scenario, three prompts were created specifying different audiences, making a total of six prompts (Appendix A). There was no word limit for the output, and writers were instructed to invent names and details. The same six prompts were presented to three writer groups to generate a corpus of 54 texts, as shown in Table 1. All texts were collected in March 2024.

The six human writers were recruited from the researcher's network of colleagues working as professional skills trainers, with between 7 and 30 years of experience in teaching business communication. They gave written permission for their texts to be used in the genre analysis and for the online task.

Data Collection

Participants. Participants for the main data collection phase were recruited through the researcher's network of teachers with experience in teaching professional skills or Business English, including a writing component. Complete data were collected from a total of 36 participants, whose Business English-related teaching experience ranged from 1 to 25 years, with a median of 12.5 years. The majority of participants were located in Asia (81%), with others in Europe (11%), the Middle East (6%), and Central America (3%).

Scoring instrument. A scoring instrument was needed by which human assessors would be able to evaluate the texts in answer to RQ2 (Appendix B). There are few existing rubrics for the assessment of business writing, and none that were suitable for this current study (although Fraser et al. [2005] and Jovic & Msnari [2024] provided inspiration). For the purposes of this study, a three-scale analytic scoring rubric was created using the British Council's established principles of effective business writing. These are known as the 3Cs (Clarity, Credibility, and Connection) and are used in professional skills writing courses globally, having been developed through many years of working with clients from different industries. Four descriptors were formulated under each of the Clarity, Credibility, and Connection scales to guide assessors on what to look for. For each text, the assessor was asked to give a score out of 5 on each scale,

Table 1. Breakdown of Writer Groups and Texts.

ChatGPT 3.5 (OpenAI)	Gemini (Google)	Human ($n = 6$)
18 texts (3 versions per prompt)	18 texts (3 versions per prompt)	18 texts (3 versions per prompt)

to produce a total score out of 15. The rubric was deliberately kept simple to reduce assessor fatigue. Analysis (presented in the findings section) demonstrated good internal reliability, as well as good concurrent validity when compared to a holistic scoring methods.

Online task. Data collection was conducted online using JISC Online Surveys (Version 3) to create a data elicitation task (Appendix C). Participants who had consented to participating in the research were emailed a link to the task to complete at a time and location of their choosing. Each participant was randomly allocated one text from each writer group (ChatGPT, Gemini, and Human), based on one of the six prompts. The order in which texts appeared was randomized and participants were not provided with information on the authorship of the texts they were shown. For each text, participants were asked to

1. Score the text using the scoring instrument
2. Give an additional score out of 10 representing their overall impression of the quality of the text (this score was used to check for concurrent validity)
3. Judge whether the text was human-written or AI-generated
4. Answer an open-ended question to explain what influenced their decision on authorship

Following this procedure, each of the 54 texts was rated by two separate assessors. One of the researchers also rated all of the 54 texts, providing a third rating as a benchmark.

Genre Analysis

A genre analysis based on the full corpus of 54 texts was conducted based on Swales's (1990, 2004) move-step approach and by consulting the work of Upton and Cohen (2009) and Hyon (2018), who offered a more systematic description of the process for identifying and describing moves. The researcher read all 54 texts and established a working set of moves and steps categories. After piloting and refining the framework, the full set of texts was coded. The frequency of moves and steps were calculated, and patterns in organization and sequencing noted.

Quantitative and Qualitative Analyses

Quantitative data (scores and authorship judgments) gathered from the online task were analyzed using IBM SPSS Statistics (Version 29) with the tests listed in Table 2.

Participant responses to open-ended questions were analyzed using thematic analysis, a qualitative method for identifying recurring themes within the data. Initial codes were assigned to capture interesting phrases, ideas, and potential themes. These codes were subsequently refined and grouped into broader thematic categories related to the

Table 2. Statistical Tests for Quantitative Data Analysis.

RQ	Purpose	Statistical Test
2	Establish reliability of rubric	Cronbach's alpha
	Establish concurrent validity	Pearson product-moment correlation
	Establish normality of data	Frequency statistics Shapiro-Wilk test
	Compare scores between writer groups	One-way analysis of variance Descriptive statistics by writer group Tukey honestly significant difference post hoc test
3	Compare judged vs. actual authorship	Crosstabulations Pearson chi-square Fleiss multirater kappa

research questions. Finally, the revised coding framework was systematically applied to all data, with analysis done using NVivo (Version 14).

Results

RQ1: A Move and Step Framework for Business Refusals

To determine what typifies the business refusal genre, the 54 texts were analyzed to produce a framework containing four major moves and segmentation of their associated steps (Table 3). Worked examples from each writer group are provided in Appendix D.

Differences between human-written and AI-generated texts could be seen through a study of whether specific moves occurred and in what order. Move occurrence counts were tabulated by step and are shown as percentages in Table 4, separated by writer group and text type. ChatGPT and Gemini texts were largely consistent in performing all steps within the framework, whereas human-written texts were inconsistent and often omitted steps. Only Step 1.1 Salutation and Step 2.1 Explanation was an obligatory step that occurred across all texts in the corpus.

While frequency counts provided insight into how often moves occurred, further analysis was needed to clarify the sequence in which moves and steps are made. In the majority of cases, moves and steps occurred in the order listed in the framework. One notable deviation from the framework is in Move 2 (Provide details), namely, how ChatGPT and Gemini texts present the refusal (Step 2.3). Both AI tools almost consistently delivered the refusal as the first step in Move 2, before the explanation (with the sequence being Step 2.3 → 2.1 → 2.2), with only two exceptions noted among 36 AI-generated texts. Conversely, all human-written texts that included an explicit refusal (2.3) followed the framework order and delivered the refusal as the final step of Move 2—the indirect structure recommended by scholars. Human writers were more likely to omit steps and deviate from their order listed in the framework, particularly when writing to a colleague at the same level or a longstanding customer. This

Table 3. Move and Step Framework for Business Refusals With Examples.

Moves and Steps	Example
MOVE 1: Open the correspondence	
Step 1.1: Salutation ^a	Dear, Hi, Hello
Step 1.2: Polite remark	"I hope this email finds you well"
Step 1.3: Express thanks	"Thanks for reaching out"
Step 1.4: Refer to request	". . .about your request to increase the participant limit"
Step 1.5: Empathy statement	"I understand this may not be the answer you were hoping for"
MOVE 2: Provide details	
Step 2.1: Explanation, reason ^a	"Here are some key reasons for this limit"
Step 2.2: Additional information	<i>Context-dependent and may include information on personal circumstances (internal texts) or policies and procedures (external texts)</i>
Step 2.3: Explicit refusal	"I regret to inform you that I am unable to. . ."
MOVE 3: Call to action	
Step 3.1: Specific alternatives or advice	"I'd like to offer the following alternative solutions"
Step 3.2: Next step	"Let me know if you would like to have a chat about this"
MOVE 4: Close the correspondence	
Step 4.1: Closing comment	"I trust you will find a suitable arrangement"
Step 4.2: Sign-off	Sincerely, Regards, Best regards, Warm wishes, All my best

^aObligatory step, occurring across all texts.

Table 4. Frequency of Step Occurrence by Writer Group, Text Type, and Total Corpus (Percentage).

Step	ChatGPT (n = 18)		Gemini (n = 18)		Human (n = 18)		Total corpus (N = 54)
	Internal	External	Internal	External	Internal	External	
1.1	100	100	100	100	100	100	100
1.2	100	100	0	0	56	33	48
1.3	189 ^a	200 ^a	133 ^a	200 ^a	11	56	131
1.4	100	100	100	100	78	78	93
1.5	178 ^a	167 ^a	122 ^a	156 ^a	44	11	113
2.1	100	100	100	100	100	100	100
2.2	0	100	100	100	44	100	74
2.3	100	100	100	100	78	33	85
3.1	100	100	100	100	100	78	96
3.2	100	100	100	100	44	78	87
4.1	100	100	100	100	78	78	93
4.2	100	100	100	100	78	100	96

^aPercentage counts over 100 indicate that a move occurred more than once in a text.

will be a point we will return to in RQ3, as it was seen as a hallmark of human writing by the raters.

RQ2: Scoring Data

For the scoring rubric, reliability statistics produced a Cronbach's alpha of .819, indicating a high level of internal consistency for the scale items (Clarity, Credibility, and Connection—hereafter referred to as the 3Cs). This showed that the three items were measuring the same construct of writing quality and further analysis could be done on the total score for the 3Cs. In addition to scoring texts on the 3C rubric, participants were asked to express their overall impression of the quality of the text as a score out of 10 (1=poor, 10=excellent). This holistic measure was intended as a second test of writing quality to establish concurrent validity. A Pearson product-moment correlation between the 3C total score and the overall impression score showed a strong, positive correlation ($r = .811$, $n = 162$, $p < .001$).

Comparison across writer groups and text types. The 3C scores were submitted to a one-way analysis of variance (ANOVA) across the three writer groups. Table 5 shows there was a statistically significant difference between groups, with a medium to large effect size, $F(2, 159) = 8.859$, $p < .001$, $\eta^2 = .100$. This indicates that the writer group (ChatGPT, Gemini, Human) has a moderate influence on the variance in total writing scores.

To identify which specific groups differed, descriptive statistics were conducted for the three writer groups (Table 6), along with a Tukey post hoc test (Table 7). The results showed that participants rated human-written texts ($M = 11.22$, $SD = 2.957$, $p < .001$) significantly higher than ChatGPT-generated texts ($M = 8.89$, $SD = 3.007$). There was also a significant difference in scores between ChatGPT-generated texts and Gemini-generated texts ($p = .034$). However, differences between Gemini-generated ($M = 10.30$, $SD = 2.731$) and human-written texts did not reach significance at the .05 level ($p = .224$).

ChatGPT-generated texts had the lowest mean scores, while human-written texts scored the highest. Among the two AI tools investigated, raters preferred Gemini-generated texts over those generated by ChatGPT. Finally, to check for effects of text type (internal or external email) on total 3C scores, a two-way ANOVA was conducted. No statistically significant interactions were found between writer group, text type, and total 3C score ($p = .448$), meaning that the findings were the same regardless of whether the message was written for an internal or external reader.

RQ3: Authorship Judgments and What Influences Them

Quantitative results. After completing scoring based on the rubric, participants were asked to judge the authorship of each text (whether it was human-written or AI-generated). Table 8 shows crosstabulations for judged versus actual authorship. Overall accuracy of judgments was 74% (80 correct out of a total 108 attempts). Participants were most easily able to identify human-written texts, with an 86.1% accuracy rate

Table 5. One-Way ANOVA on 3C Total Score.

	Sum of Squares	df	Mean Square	F	Sig.
Between groups	149.086	2	74.543	8.859	<.001
Within groups	1337.926	159	8.415		
Total	1487.012	161			

Table 6. Descriptive Statistics by Writer Group.

Groups	N	Mean	Standard Deviation	Standard Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
ChatGPT	54	8.89	3.007	.409	8.07	9.71
Gemini	54	10.30	2.731	.372	9.55	11.04
Human	54	11.22	2.957	.402	10.42	12.03
Total	162	10.14	3.039	.239	9.66	10.61

Table 7. Tukey HSD Post Hoc Test: Multiple Comparisons on 3C Total Score by Writer Group.

(I) Writer_ Group	(J) Writer_ Group	Mean Difference (I - J)	Standard Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
ChatGPT	Gemini	-1.407 ^a	.558	.034	-2.73	-.09
	Human	-2.333 ^a	.558	<.001	-3.65	-1.01
Gemini	ChatGPT	1.407 ^a	.558	.034	.09	2.73
	Human	-.926	.558	.224	-2.25	.39
Human	Gemini	.926	.558	.224	-.39	2.25
	ChatGPT	2.333 ^a	.558	<.001	1.01	3.65

^aThe mean difference is significant at the .05 level.

(only 5 instances out of 36 where a human-written text was incorrectly identified as AI-generated). By contrast, there was more uncertainty around identifying the authorship of AI-generated texts, with a 68.1% accuracy rate. Within the AI writer groups, there was little difference between ChatGPT and Gemini, with accuracy rates of 66.6% and 69.4% respectively.

Results of a Pearson chi-square test show $\chi(1)=28.167, p < .001$, suggesting a statistically significant association between judged and actual authorship, with a moderate effect size (Cramer's $V=.511$). Hence, participant judgments as to whether a text was written by a human or AI tool were not random. To further check that agreement

Table 8. Crosstabulation: Judged Versus Actual Authorship.

		Actual			
		AI	Human	Total	
Judgment	AI	Count	49	5	54
		% within judgment	90.7	9.3	100.0
		% within actual	68.1	13.9	50.0
		% of total	45.4	4.6	50.0
	Human	Count	23	31	54
		% within judgment	42.6	57.4	100.0
		% within actual	31.9	86.1	50.0
		% of total	21.3	28.7	50.0
	Total	Count	72	36	108
% within judgment		66.7	33.3	100.0	
% within actual		100.0	100.0	100.0	
% of Total		66.7	33.3	100.0	

between judged and actual authorship did not occur by chance, Fleiss's kappa was run, and it showed that there was moderate agreement between judged and actual authorship, $\kappa = .467$ (95% CI, .278 to .655), $p < .001$.

Qualitative insights. After judging the authorship of the text, participants were asked to explain what influenced their decision as to whether the text was AI or human-written. Thematic analysis of the open-ended responses revealed that participants were most concerned with three broad characteristics of the texts: tone, relationship, and formality; language choice and grammatical accuracy; content, detail, and structure. These themes and some examples are summarized in Table 9.

Overall, the three broad themes that emerged from the qualitative data show a preference for texts that consider the relationship between writer and reader. While AI-generated texts were notable for their grammatical correctness, they were also assessed to be formulaic and lacking in contextual awareness. By contrast, less clichéd language choice, concise messaging, specific rather than general content, and attention to interpersonal relationships were cited as the main influences when identifying texts as human-written.

Discussion

Our research contributes a genre framework for business refusals that did not exist in previous literature, showing that there are typical moves and steps employed across writer groups for business refusals. Frequency and sequencing analysis showed that texts generated by ChatGPT and Gemini applied moves and steps in a largely consistent manner across text type (internal, external) and did not adjust their approaches,

Table 9. Influences on authorship judgments.

Theme	Summary
Tone, relationship, formality	<p>Tone was commonly described using adjectives, with AI-generated texts labeled “cold,” “distant,” “impersonal,” “mechanistic,” “rigid.” Human-written texts were described as “empathetic,” “familiar,” “human,” “sincere.”</p> <p>Responses showed a tendency to prioritize interpersonal relationships, commenting on the (in)appropriateness of tone and formality depending on the reader-writer relationship (e.g., subordinate writing to a boss).</p>
Language choice	<p>Words and phrases found in the texts heavily influenced participants’ decisions on authorship. In particular, AI-generated texts were frequently described as containing “jargon,” “ cliché phrases,” “business writing clichés,” or “stock phrases.” Examples cited included “reaching out,” “more than happy,” “apologize for any inconvenience caused,” and “regret to inform.”</p> <p>AI texts were also judged to be complex, including long-winded sentences and exaggerated vocabulary. Human-written texts were described as using simpler, clearer, and more accessible language (“Plain English,” “less jargony”).</p> <p>Overuse of the first-person singular pronoun “I” was identified as a feature of AI-generated texts, while human-written texts exhibited features such as subject pronoun drop (“Hope you are well”) showing informality and connection.</p>
Grammatical accuracy	<p>Participants remarked on the imperfections that characterize human writing, such as mistakes in grammar and punctuation or not writing out words fully (e.g., using “Thurs” instead of Thursday). The lack of errors in AI-generated text also guided participants in making their authorship assessments.</p>
Content, detail and structure	<p>Participants found AI texts to be “verbose,” “long-winded,” and “repetitive.” AI texts were widely judged to be clearly organized and coherent, but this contributed to the feeling they were formulaic, templated and generic. The use of the direct refusal (rather than explanation-first) was evaluated negatively.</p> <p>Human-written texts were “succinct,” “concise,” and “to the point.” They were observed to be personalized, containing contextual content specific to the writer/reader’s situation. The use of politeness strategies such as the indirect structure (reason before refusal) and implied rather than explicit refusals was also noted.</p>

despite the prompts specifying different writer-reader relationships. There was also a tendency for the AI tools to repeat steps such as expressing thanks (Step 1.3) and the empathy statement (Step 1.5), contributing to their perceived repetitive nature. The consistency with which AI texts applied genre moves corresponds with the findings of AlAfnan et al. (2023) and Jovic and Mnasri (2024), who also described AI-generated emails as formulaic and templated.

On the other hand, the human-written texts varied in the moves they employed, with differences seen between internal and external texts. Human writers were much more likely to omit steps, depending on the context and writer-reader relationship specified in the prompt. Particularly for external texts written to clients, it was expected that overt signals of politeness would be more consistently included; however, only 33% and 11% of human-written texts included a polite remark (1.2) and empathy statement (1.5), respectively. Instead, politeness may be inferred from the fact that only one-third of external texts included an explicit refusal statement (2.3). The remaining majority contained an implied refusal, which the reader would infer from the performance of other steps. For example, the writer would first offer an explanation to the customer before providing specific alternatives (in line with politeness strategies proposed by Campbell, 1990) and attempting to persuade the customer to accept one of these. Implied refusals can be interpreted as an indicator of human writing skill and a desire to maintain face (Brown & Levinson, 1987) not seen in the AI-generated texts, all of which delivered an explicit refusal statement, such as “I’m unable to approve your request.” Overall, human-written texts were more nuanced in their adoption of genre conventions and appeared to consider the text and audience type when choosing whether to include or omit moves, and how to sequence them.

Scoring data based on the three-scale rubric assessing Clarity, Credibility, and Connection (3Cs) showed that human assessors preferred human-written over AI-generated texts, with Human texts achieving the highest mean score, followed by Gemini and finally ChatGPT. The significant difference in scores between Gemini and ChatGPT texts suggests that Gemini may be the more effective AI tool for business writers. Interestingly, score differences between Gemini-generated and human-written texts did not reach significance. This may signal that Human and Gemini texts might be perceived similarly, but further research with a larger sample size would be needed to explore this aspect. In addition, human-written texts were judged to be of higher quality regardless of whether they were written for an internal or an external audience, with no statistically significant interactions found based on scenario type. Given the recent release of these AI tools, no previous published studies have compared the applications of ChatGPT and Gemini in business writing, nor are there studies comparing human-written with AI-generated business texts. AIAfnan et al. (2023) and Jovic and Mnasri (2024) found that assessors gave low scores to AI-generated business texts, but human-written texts were not part of either study. Some studies in academic domains have observed that AI models may be as good or better than humans at writing scientific (Yeadon et al., 2024) and argumentative essays (Herbold et al., 2023). However, our study shows they fall short in business writing. One explanation may be that business writing is heavily judged on the ability to build relationships with the reader, an area in which AI-generated texts do not perform as well compared to human-written ones.

On authorship, we found participants could correctly identify human or AI authorship of business refusal texts with a higher accuracy rate than by chance. These results diverge from the studies of Casal and Kessler (2023) and Yeadon et al. (2024), who

found low accuracy rates in human raters' ability to identify authorship. This may be attributable to the context in which these two previous studies were conducted: AI tools generate academic texts that can fool even experienced assessors, but fare less well in a business writing context that prioritizes relationships and connection to the reader. Similar to the present study, Casal and Kessler's participants were more likely to correctly identify human-written texts as human. This suggests there are characteristics of human writing that make it readily identifiable as such and more positively perceived.

Findings on tone corresponded with research on AI-generated messages conducted by Coman and Cardon (2024), in which professionals rated ChatGPT-generated messages as less sincere and caring. Similar to participant comments in this study, Jovic and Mnasri (2024) found that the tone in LLM-generated emails was emotionally neutral and overly formal, pointing to difficulty in expressing sentiments. In terms of language choice, Campbell et al. (2023) found that a plain style avoiding jargon and nonrequisite words contributed positively to the perceived professionalism of business writers, and these perceptions may have influenced authorship judgments in this study. The tendency for AI-generated texts to be convoluted, verbose, and jargonistic was an indicator of AI authorship for many participants. Human writers paid attention to the writer-reader relationship and wrote messages that were concise and specific in a plain style, avoiding the use of clichéd language.

Table 10 summarizes the characteristics of AI-generated and human-written texts as found through the move structure analysis and qualitative data from participants.

Conclusion and Recommendations

The introduction of this study included a provocative question for the profession: "Is there still a case for business writing training?" The short answer is yes. Business writing training and the teachers and trainers who deliver it are still needed in the AI age. Effective business writing, especially in difficult situations, requires a human touch, understanding of the reader's needs and of the context itself. AI tools might help people draft communications in the future, but human awareness of the genre is needed so messages are not seen as detached or formulaic. We identify four key areas where teachers of business writing should focus their energies.

Raising Awareness of the Complex Nature of Genres

The broad and simplified genre framework presented here for business refusals is useful as a teaching tool for learners new to the business discourse community, as well as professionals looking to refine their skills. However, it should be seen as a flexible framework and understood as description rather than prescription (Pinto dos Santos, 2002), and emphasis needs to be on process over product (Creelman, 2012). Rather than simply providing a writing "formula," teachers need to emphasize that genres are flexible and guide learners on how they can use and adapt generic conventions to best

Table 10. Characteristics of AI-Generated and Human-Written Texts.

	AI-Generated	Human-Written
Genre structure	Rigid and formulaic Some steps repeated Explicit refusals before reason	Flexible (steps omitted or different sequencing depending on context) Refusal after reason (or implied)
Tone, relationship, formality	Mechanical and impersonal Very formal regardless of writer-reader relationship	Human, friendly, sincere Formality level adjusted for audience
Language choice	Jargon and clichés Overuse of personal pronoun “I”	Plain style Subject pronoun drop More use of “we”
Grammatical accuracy	No errors	Minor mistakes Short forms used
Content and detail	Long and repetitive Generic content	Concise Specific details for writer/reader’s situation

achieve their goals. Each writer-reader interaction should be handled in a context-specific manner, taking into account factors such as role relationships, tasks, and industry-specific practices or concerns (Zhang, 2013). Contextualization remains important and learners should be introduced to a genre’s conventions through analysis, critique, reflection, practice, and feedback.

Focusing on Involvement in the Writing Process

AI tools cannot replace the experience of attending classes and interacting with other learners and a teacher. The discussion process is hugely important for helping writers develop knowledge in how to persuasively structure arguments, build relationships, and embrace nuance (Ying, 2020). Business writing training should continue to focus on giving learners the opportunity to share authentic workplace experiences and engage in collaborative planning, writing, and editing. Practice is essential, as each performance of a genre is unique (Devitt, 2015) and involves consideration of how to perform rhetorical moves in ways to fulfil specific tasks in a specific discourse community. Learners then receive feedback on the textual performance from peers and their trainer, and reflect on it before performing again. They might also compare their human-written texts with texts generated by different AI tools, with discussion focusing on where AI can be used effectively to help and where human judgment remains most important. Fostering a sense of involvement will help retain text ownership and satisfaction in the writing process (Dhillon et al., 2024), as learners also hone their thinking skills.

Helping Learners Develop Aspects of Writing Voice That Feel Human

AI tools can be used to support parts of the writing process, such as in finding relevant arguments, suggesting organizational structure, and providing examples of styles that may be best suited to the writer's audience and goals (DeJeu, 2024). The ability of AI tools to produce grammatically correct text reduces the need for intense grammar instruction in business writing courses, although learners should still be made aware of common errors. Instead, emphasis should be on using human judgment, creativity, empathy, and contextual knowledge to evaluate AI output to ensure a text achieves its communicative purpose. The results of our exploratory study suggest that having a relationship orientation, emotional understanding, and contextual knowledge are the main differentiating factors between human-written and AI-generated texts. Hence, business writing training should focus on developing the ability to analyze the purpose, audience, and context so that communication can be adapted to the specific situation at hand. Genuine empathy for the reader's needs and expectations, along with a human voice, personality, and style (including nonstandard and idiosyncratic language), is what sets human writing apart. As the bulk of business writing converges on formulaic structures and clichéd phrases, which AI tools perpetuate, human writers should aim to engage and surprise their readers.

Developing AI Literacy

Current and future business writing curricula need to include a component on AI literacy (Cardon, Fleischmann, Aritz, et al., 2023; Getchell et al., 2022 provide starting frameworks). This should include aspects such as understanding how to use AI tools for specific tasks and for what tasks they might be appropriate; considering authenticity and the human element in communication; taking responsibility for AI-generated content and using it fairly; and discussing issues around governance, ethics and safety. Learners, along with their teachers, must be able to negotiate what AI-human collaboration should look like in their context. Applying these frameworks to the above example, a business communication trainer could engage in the following steps to raise AI literacy around "bad news" email writing:

- Stage 1: brainstorming to share existing knowledge on Generative AI around its benefits and limitations; create rules for using AI in the workplace (ethical considerations); and think of tips to write good prompts
- Stage 2: testing out prompting strategies to generate writing on the same topic as a human-written text
- Stage 3: working together to review and evaluate output (both AI and human) to discern the extent to which they can be used accurately and ethically in the workplace, while maintaining human agency.

These activities aim to facilitate discussion around what AI can/cannot help with and where human writers' strengths lie. This links to two main roles of business written

communication training: (1) helping to raise awareness of genres, and (2) involving learners in the writing process so they know how to best achieve their goals.

Limitations and Future Research

We acknowledge several limitations of this study. First, genre analysis is not an exact science. Moves and their boundaries may be difficult to delineate (Samraj, 2014), resulting in different ways to code moves and steps. In addition, further linguistic analysis of the texts with computerized corpus techniques (Flowerdew, 2012, provides an example) could elaborate on differences in lexicogrammatical features and add a quantitative dimension to the perceptions around language choice reported by participants. Second, generative AI is constantly evolving and as newer LLMs are released, output generated from the same prompts used in this study may change. For consistency in this study, we generated each text with a single prompt. In reality, the prompt process is usually iterative, meaning that often more than one prompt is used to fine-tune the desired output. Future studies on AI and writing in the business context should consider (1) using newer versions of the platforms; (2) evaluating AI models beyond the two studied here; and (3) testing different prompt strategies (Knoth et al., 2024). Third, future studies could consider broadening the representativeness of the participant sample. The data in our study reflects the opinions and perceptions of teachers and trainers with a professional background in Business English, but does not capture the views of learners or business practitioners.

Finally, our study investigated AI-generated and human-written texts, in isolation of one another. However, actual use of the technology is more likely to involve a combination of AI technology and human judgment. Further study on how professionals are using AI tools in their writing would be helpful for understanding points in the writing process where technology can assist. Particular areas for focus would be when and where humans are choosing to use or not use AI, as well as how decisions are made on modifying AI-generated texts to suit a particular communicative purpose.

AI tools can enhance productivity and efficiency and play a valuable role in supporting and augmenting human work. However, they cannot yet replace humans, who possess moral reasoning, ethical judgment, and emotional intelligence. Our study showed some of the human aspects of writing that AI-generated texts do not fully replicate. Further research will be needed as the technology continues to evolve, so that teachers can help learners make use of the strengths of AI tools to assist with writing, while knowing how to compensate for weaknesses in the technology.

Declaration of Conflicting Interests


The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Winy Wilson  <https://orcid.org/0009-0003-4071-9781>

Heath Rose  <https://orcid.org/0000-0002-6434-6663>

Supplemental Material

Supplemental material for this article is available online.

References

- AlAfnan, M. A., Dishari, S., Jovic, M., & Lomidze, K. (2023). ChatGPT as an educational tool: Opportunities, challenges, and recommendations for communication, business writing, and composition courses. *Journal of Artificial Intelligence and Technology*, 3(2), 60–68. <https://doi.org/10.37965/jait.2023.0184>
- Bhatia, V. K. (1993). *Analysing genre: Language use in professional settings*. Longman.
- Bhullar, P. S., Joshi, M., & Chugh, R. (2024). ChatGPT in higher education—A synthesis of the literature and a future research agenda. *Education and Information Technologies*, 29(16), 21501–21522. <https://doi.org/10.1007/S10639-024-12723-X>
- Biber, D., Connor, U., & Upton, T. A. (2007). *Discourse on the move: Using corpus analysis to describe discourse structure*. John Benjamins Publishing Company.
- Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge University Press.
- Campbell, K. S. (1990). Explanations in negative messages: More insights from speech act theory. *Journal of Business Communication*, 27(4), 357–375.
- Campbell, K. S., Naidoo, J. S., & Smith, J. (2023). When your boss says, “You need to sound more professional”: Writing style and writer attributions. *International Journal of Business Communication*, 60(4), 1071–1094. <https://doi.org/10.1177/23294884211025735>
- Cardon, P., Fleischmann, C., Aritz, J., Logemann, M., & Heidewald, J. (2023). The challenges and opportunities of AI-assisted writing: Developing AI literacy for the AI age. *Business and Professional Communication Quarterly*, 86(3), 257–295. <https://doi.org/10.1177/23294906231176517>
- Cardon, P., Fleischmann, C., Logemann, M., Heidewald, J., Aritz, J., & Swartz, S. (2023). Competencies needed by business professionals in the AI age: Character and communication lead the way. *Business and Professional Communication Quarterly*, 87(2), 1–24. <https://doi.org/10.1177/23294906231208166>
- Casal, J. E., & Kessler, M. (2023). Can linguists distinguish between ChatGPT/AI and human writing?: A study of research ethics and academic publishing. *Research Methods in Applied Linguistics*, 2(3), 100068. <https://doi.org/10.1016/J.RMAL.2023.100068>
- Coman, A. W., & Cardon, P. (2024). Perceptions of professionalism and authenticity in AI-assisted writing. *Business and Professional Communication Quarterly*. Advance online publication. <https://doi.org/10.1177/23294906241233224>
- Creelman, V. (2012). The case for “living” models. *Business and Professional Communication Quarterly*, 75(2), 176–191. <https://doi.org/10.1177/1080569912442981>

- de Rycker, A. G. H. (2014). Mitigation in turning down business proposals across cultures: The case for pragmatic competence instruction. *3L: Language, Linguistics, Literature, 20*(1), 87–100. <https://doi.org/10.17576/3L-2014-2001-07>
- DeJeu, E. (2024). A rhetorical approach for reimagining business writing instruction in the AI age. *Business and Professional Communication Quarterly*. Advance online publication. <https://doi.org/10.1177/23294906241255301>
- Devitt, A. J. (2015). Genre performances: John Swales' genre analysis and rhetorical-linguistic genre studies. *Journal of English for Academic Purposes, 19*, 44–51. <https://doi.org/10.1016/J.JEAP.2015.05.008>
- Dhillon, P. S., Molaei, S., Li, J., Golub, M., Zheng, S., & Robert, L. P. (2024). *Shaping human-AI collaboration: Varied scaffolding levels in co-writing with language models* [Conference session]. Proceedings of the CHI Conference on Human Factors in Computing Systems, 1–18. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3613904.3642134>
- Flowerdew, L. (2012). Exploiting a corpus of business letters from a phraseological, functional perspective. *ReCALL, 24*(2), 152–168. <https://doi.org/10.1017/S0958344012000043>
- Fraser, L., Harich, K., Norby, J., Brzovic, K., Rizkallah, T., & Loewy, D. (2005). Diagnostic and value-added assessment of business writing. *Business and Professional Communication Quarterly, 68*(3), 290–305. <https://doi.org/10.1177/1080569905279405>
- Getchell, K. M., Carradini, S., Cardon, P. W., Fleischmann, C., Ma, H., Aritz, J., & Stapp, J. (2022). Artificial intelligence in business communication: The changing landscape of research and teaching. *Business and Professional Communication Quarterly, 85*(1), 7–33. <https://doi.org/10.1177/23294906221074311>
- Gunser, V. E., Gottschling, S., Brucker, B., Richter, S., & Gerjets, P. (2021). Can users distinguish narrative texts written by an artificial intelligence writing tool from purely human text? *Communications in Computer and Information Science, 1419*, 520–527. https://doi.org/10.1007/978-3-030-78635-9_67
- Henry, A., & Roseberry, R. L. (2001). A narrow-angled corpus analysis of moves and strategies of the genre: 'Letter of Application'. *English for Specific Purposes, 20*, 153–167.
- Herbold, S., Hautli-Janisz, A., Heuer, U., Kikteva, Z., & Trautsch, A. (2023). A large-scale comparison of human-written versus ChatGPT-generated essays. *Scientific reports, 13*(1), 18617. <https://doi.org/10.1038/s41598-023-45644-9>
- Hyland, K. (2022). English for Specific Purposes: What is it and where is it taking us? *ESP Today, 10*(2), 202–220. <https://doi.org/10.18485/esptoday.2022.10.2.1>
- Hyon, S. (2018). *Introducing genre and English for Specific Purposes*. Routledge.
- Jansen, F., & Janssen, D. (2010). Explanations first: A case for presenting explanations before the decision in Dutch bad-news messages. *Journal of Business and Technical Communication, 25*(1), 36–67. <https://doi.org/10.1177/1050651910380372>
- Jansen, F., & Janssen, D. (2013). Effects of directness in bad-news e-mails and voice mails. *Journal of Business Communication, 50*(4), 362–382. <https://doi.org/https://doi.org/10.1177/0021943613497053>
- Jovic, M., & Mnasri, S. (2024). Evaluating AI-generated emails: A comparative efficiency analysis. *World Journal of English Language, 14*(2), 502–517. <https://doi.org/10.5430/wjel.v14n2p502>
- Knoth, N., Tolzin, A., Janson, A., & Leimeister, J. M. (2024). AI literacy and its implications for prompt engineering strategies. *Computers and Education: Artificial Intelligence, 6*, 100225. <https://doi.org/10.1016/J.CAEAI.2024.100225>

- Köbis, N., & Mossink, L. D. (2021). Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry. *Computers in Human Behavior, 114*, 106553. <https://doi.org/10.1016/J.CHB.2020.106553>
- Kong, K. C. C. (1998). Are simple business request letters really simple? A comparison of Chinese and English business request letters. *Text & Talk, 18*(1), 103–141.
- Nguyen, H., & Miller, J. (2012). Exploring business request genres: Students' rhetorical choices. *Business Communication Quarterly, 75*(1), 5–28. <https://doi.org/10.1177/1080569911430379>
- Park, S., Jeon, J., & Shim, E. (2021). Exploring request emails in English for business purposes: A move analysis. *English for Specific Purposes, 63*, 137–150. <https://doi.org/10.1016/J.ESP.2021.03.006>
- Pinto Dos Santos, V. B. M. (2002). Genre analysis of business letters of negotiation. *English for Specific Purposes, 21*(2), 167–199. [https://doi.org/10.1016/S0889-4906\(00\)00028-4](https://doi.org/10.1016/S0889-4906(00)00028-4)
- Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching, 6*(1), 342–363. <https://doi.org/10.37074/jalt.2023.6.1.9>
- Samraj, B. (2014). Move structure. In K. P. Schneider & A. Barron (Eds.), *Pragmatics of discourse* (pp. 385–406). De Gruyter Mouton. <https://doi.org/10.1515/9783110214406-015>
- Schryer, C. F. (2000). Walking a fine line: Writing negative letters in an insurance company. *Journal of Business and Technical Communication, 14*(4), 445–497. <https://doi.org/10.1177/105065190001400402>
- Shelby, A. N., & Reinsch, N. L. (1995). Positive emphasis and you-attitude: An empirical study. *International Journal of Business Communication, 32*(4), 303–326. <https://doi.org/10.1177/002194369503200401>
- Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.
- Swales, J. M. (2004). *Research genres: Explorations and applications*. Cambridge University Press.
- Upton, T. A., & Cohen, M. A. (2009). An approach to corpus-based discourse analysis: The move analysis as example. *Discourse Studies, 11*(5), 585–605.
- Vergaro, C. (2004). Discourse strategies of Italian and English sales promotion letters. *English for Specific Purposes, 23*(2), 181–207. [https://doi.org/10.1016/S0889-4906\(03\)00003-6](https://doi.org/10.1016/S0889-4906(03)00003-6)
- Yeadon, W., Agra, E., Inyang, O., Mackay, P., & Mizouri, A. (2024). Evaluating AI and human authorship quality in academic writing through physics essays. *ArXiv Preprint ArXiv:2403.05458*.
- Ying, J. (2020). The importance of the discussion method in the undergraduate business classroom. *Humanistic Management Journal, 5*(2), 251–278. <https://doi.org/10.1007/s41463-020-00099-2>
- Zhai, X., & Razali, A. B. (2023). Triple method approach to development of a genre-based approach to teaching ESL/EFL writing: A systematic literature review by bibliometric, content, and scientometric analyses. *SAGE Open, 13*(1), 1–27. <https://doi.org/10.1177/21582440221147255>
- Zhang, Z. (2013). Business English students learning to write for international business: What do international business practitioners have to say about their texts? *English for Specific Purposes, 32*(3), 144–156. <https://doi.org/10.1016/J.ESP.2013.01.002>

Author Biographies

Winnie Wilson is a qualified English language and professional skills trainer for written and spoken communication skills. Prior to entering the business training industry, she worked in market research and advertising and maintains a strong interest in the business world. She completed this research as a postgraduate researcher at the University of Oxford.

Heath Rose is Professor of Applied Linguistics at The University of Oxford. His research explores issues related to the globalisation of English language. Prior to entering academia he worked as a language instructor in a business university in Japan.