Evaluating AI-Generated Emails: A Comparative Efficiency Analysis

Marina Jovic[1], & Salaheddine Mnasri[1]

[1] Assistant Professor, American University of the Middle East, Kuwait

Correspondence: Dr. Marina Jovic, Assistant Professor, American University of the Middle East, Kuwait. E-mail: marina.jovic@aum.edu.kw

Abstract

This study investigates the efficiency of large language models (LLMs) in producing routine, negative, and persuasive business emails for educational purposes within the context of Business Writing. Specifically, it compares the outputs generated by four widely-used LLMs (ChatGPT 3.5, Llama 2, Bing Chat, and Bard) when presented with identical email scenarios. These generated emails are evaluated using an elaborate rubric, allowing for a systematic assessment of LLMs' performance across three distinct email types. The results of the study show that the output with the same prompt varies greatly despite the rather formulaic nature of business emails. For instance, some LLMs struggle with following the requested structure and maintaining consistency in tone, while others have issues with unity and conciseness. The findings of this research hold implications for teaching business writing (rubrics, task instructions, in-class implementation), as well as for the integration of AI in professional communication at large.

Keywords: business emails, Bard, ChatGPT 3.5, Bing Chat, Llama 2.0, comparative analysis

1. Introduction

Establishing and maintaining professional relationships, sharing important information, promoting the organization, and improving its performance all hinge on effective business communication. Email communication, in particular, is essential for business interactions in the arena of digital communication. To achieve desired results and promote collaboration, it is essential for professionals to have the capacity to write emails that are crystal clear, concise, and appropriate.

Recent technological advancements have yielded large language models, such as GPT-3.5, which are transforming business communications practices. These LLMs demonstrate human-like proficiency in producing natural language text, demonstrating their capacity to function as automated assistants for multiple writing purposes. The integration of AI-generated material, including emails, holds enormous potential for both businesses and educational institutions. An increasing number of professionals and students employ AI to write their electronic correspondence. A salient issue yet unaddressed is: Which of the currently available and publicly accessible LLMs without a usage fee demonstrates optimal performance in generating emails that feature clarity, conciseness, accuracy, and contextual relevance?

The central aim of this investigation is to assess the efficacy of the four prominent LLMs in generating routine, negative, and persuasive business emails. The ultimate goal is to derive insights from the assessments and provide recommendations for enhancing the pedagogy of business writing. Specifically, the study seeks to address the following research questions:

1) To what extent can Bard, ChatGPT 3.5, Llama 2.0, and Bing Chat effectively generate routine business emails, e.g., routine complaints, using a direct approach?

2) How well do these LLMs capture the nuances of negative message communication in business emails?

3) Can LLMs convincingly compose persuasive business emails using an indirect approach?

By addressing these research questions, this study aims to shed light on the potential of LLMs in enhancing different facets of business communication through email writing. Given the extensive body of research conducted on ChatGPT 3.5, it is beneficial to conduct a comparative analysis of its performance with other LLMs that are frequently used by professionals and students. This comparison could provide valuable insights into the relative strengths and weaknesses of the models in email generation. The findings could hold implications for business pedagogy, organizations seeking to optimize their communication effectiveness, as well as the integration of AI in professional communication at large.

Various business sectors have already started to explore the potential impact of AI in their respective fields. In addition to examining the existing uses and capabilities of AI, disciplines including marketing, management and organizational studies, and information systems are working to develop strategic frameworks that will guide decisions on AI. It is imperative that educators in communication, especially business ones, consider how these new technologies can change business communication and teach students about the uses, functions, and ethics of AI in the workplace (Getchell et al., 2022).

The following sections offer a literature review, outline the research methodology, and expose the findings, followed by an in-depth analysis. Furthermore, the article explores the implications of the present research, unpacks its potential applications, and suggests future investigations.

1.1 Literature Review

Effective communication can lead to increased employee satisfaction, improved decision-making, and enhanced customer service (Valiyeva & Thomas, 2022), and it can help build trust, improve collaboration, and resolve conflict (Romm & Pliskin, 1997). The implementation of effective communication strategies can significantly enhance the organization's chances of long-term success (Mnasri & Jovic, 2023). More particularly, written communication, especially when it is official, can be viewed as serious and consequential. E-mailing, used by over a billion people for a variety of functions, is one of the most prevalent means of communication. Since November 2022, the process of writing an email has become remarkably simplified due to the rapid progress in artificial intelligence (AI). AI, a broad term that includes various technologies such as machine learning, deep learning, natural language processing, and computer vision, is transforming the realm of business communication. These technologies are being increasingly used to enhance productivity and improve business communication. The advent of AI offers the potential to support, mediate, and facilitate not only business communication but also business writing pedagogy.

Even before Open AI ChatGPT became publicly accessible in November 2022, AI had already made its mark in business writing. Early applications ranged from the use of chatbots for customer service and AI-generated content for marketing to error detection in writing and customer relationship management. Some of the earliest were based on rule-based systems that could identify common errors in grammar and spelling. However, these tools were not always accurate, and they could not provide feedback on more complex aspects of writing, such as style and clarity (Adams & Chuah, 2022). Systems for writing assessment and assistance, based on automated writing evaluation (AWE), have also been available for some time. These tools range from automated systems to those that provide synchronous feedback and can even generate entire texts (Godwin-Jones, 2022).

Getchell et al. (2022) discusses the use of AI tools for business communication and collaboration, including team communication and meeting tools, text-summarization tools, augmented writing tools, oral communication evaluation, and conversational agents. Building on Siddike et al. (2018) and Babic et al. (2020), the study proposes the AI Roles framework, outlining five key roles for AI in business communication. These include: i) Tool, which handles data processing and retrieval tasks; ii) Assistant, providing recommendations for various tasks; iii) Monitor, evaluating communication performance; iv) Coach, offering advice to improve communication; and v) Teammate, collaborating with humans in decision-making processes (Getchell et al., 2022).

According to Alshurafat (2023), chatbots like ChatGPT can assist in communicating with clients and colleagues, providing quick answers to common questions, and reducing the time and effort required for manual follow-ups. Additionally, ChatGPT can facilitate communication between team members, helping to streamline workflow and improve collaboration. This can lead to improved customer service, faster resolution of issues, and increased efficiency in accounting processes. Another study shows that ChatGPT enables natural, text-based interactions and proves to be beneficial in customer service and virtual assistant applications. It offers features like topic detection, emotion detection, and sentiment analysis to enhance user understanding. Additionally, it can generate multiple conversation threads for more realistic interactions (George & George, 2023).

The rapid rise in AI led to a remarkable increase in the number of AI studies and research, including in the fields of business communication and pedagogy. Getchell et al. (2022) looks into the profound impact of AI technologies on business communication, providing a comprehensive overview of their current capabilities and associated challenges. The authors propose a research agenda for scholars in business communication, encompassing a range of topics such as implementation, lexicography and grammar, collaboration, design, trust, bias, managerial concerns, tool assessment, and demographics. The topic of tool assessment, which is the focus of the present article, aligns with one of these proposed areas of research. Despite some controversy, when these tools are integrated into instruction, they can enhance the quality of learners' writing and offer benefits to both students and teachers (Godwin-Jones, 2022). Cardon et al.'s (2023) study indicates that the integration of generative AI in business communication could lead to significant changes in teaching methods, with potential challenges including reduced critical thinking and originality and benefits such as increased efficiency and enhanced idea generation.

The significance of LLMs in the realm of higher education has already been identified (Okonkwo & A. Ade-Ibijola, 2021; Sok & Heng, 2023; Rudolph et al., 2023; Gill et al., 2024). A key area of interest for the scientific community has been exploring the capabilities of this technology. Some examples of emerging academic topics are: ChatGPT's role in scientific communication (Huang & Tan, 2023); its potential in writing a medical report (Biswas, 2023); AI potential in producing any high-quality academic writing (Mahyoob et al., 2023); AI ability to generate English essays (Fitria, 2023); and AI writing style (AlAfnan & MohdZuki, 2023). Literature has also focused on AI's potential impact on education, research, and writing (Bishop, 2023) or on students' writing skills (Abdullayeva & Musayeva, 2023; Yan, 2023). Some studies also provide practical advice (Al-Afnan et al., 2023; Cardon et al., 2023). Al-Afnan et al. (2023) looks into the use of ChatGPT as a tool for teaching and learning by scrutinizing the advantages and disadvantages of its application in communication, business writing, and composition courses. Furthermore, the research provides valuable guidance for both educators and students. Cardon et al. (2023) emphasizes the need for students to acquire AI literacy and offers recommendations for educators to maximize the benefits and minimize the challenges of AI-assisted writing.

As for studies that targeted AI-assisted email writing, they range from AI testing email effectiveness (Nguyen et al., 2023) to investigating people's perceived trust in AI-mediated email writing (Liu et al., 2022). There are attempts to propose a testing framework (Mahyoob et al., 2023). The authors tried to rate the accuracy and proficiency of this algorithm-generated writing, including emails, on the basis of six principles (Relatedness, Adequacy, Limitation, Authenticity, Cognition, and Redundancy). The findings reveal that while ChatGPT has extraordinary capabilities, there are also severe flaws, such as information repetition, unfounded assumptions, incorrect reasoning, fabricated references, hallucinations, and a lack of pragmatic interpretation. Another study reveals some other flaws as well by investigating the stylistic features of different genres of ChatGPT-4-generated responses, including business correspondence. The authors analyze sentence length, paragraph structure, word choice, mood, tense, voice, pronouns, keyword density, lexical density, lexical variety, and reading ease (AlAfnan & MohdZuki, 2023).

Even though there are numerous studies that look at how well AI chatbots perform in writing, they are restricted in two ways: (1) they lack a notable analysis of the performance of chatbots such as Bard, Bing Chat, and Llama 2.0, and (2) there is no comparative analysis of these freely available AI tools. By assessing and contrasting four well-known LLMs—ChatGPT (GPT-3.5), Bing, Bard, and Llama 2.0—this research fills the gap in the field. The study also seeks to help students and teachers choose the best AI chatbot platforms to improve their business writing, teaching, and learning experiences by looking at their distinctive features and uses.

2. Method

The study adopts the following methodology: Three different prompts for three different email types were created. The aim was to solicit responses and examine their effectiveness, as per the criteria detailed in the grading rubric (see Appendix A). All the prompts included a specific concern or issue. The first prompt is a customer's complaint; the second is a denial of a refund request; and the third is a request for remote work. Thus, humans (ideally customer service professionals) can easily tell that the sender is requiring a clear and concise response in a respectful tone. Each scenario follows a purposeful structure and a specific approach (direct or indirect). These prompts differ in their purpose, approach, tone, structure, and content depending on the specific communication objectives and context they address. There is an emphasis on action or resolution. Although they do not provide detailed background information, they guide and offer possible solutions (Al-Afnan et al., 2023). They contain clear instructions on the number of words and paragraphs. The number of words ranges from 150 to 200 for routine emails, and 200 to 250 for more complex negative and persuasive ones, in an attempt to test the LLM's ability to generate concise content. Limitations in the number of paragraphs and structural constraints were given to evaluate LLMs' ability to follow the guidelines. Multiple suggestions in each scenario had the aim of checking if AI language models could choose one and maintain unity.

The first author created a generalized rubric that could be used for grading all three types of emails. New LLM accounts were made. Each prompt was put in a new chat in each LLM to avoid compromising it with previous ones. These four models were chosen because they are easily accessible and free of charge. Concerning the email assessment, the primary author forwarded the AI-generated emails to the secondary author, who was, at that point, unaware that the emails were written by AI or that they were intended for research purposes. Following this, the authors engaged in a comparative review and discussion concerning the grading outcome.

2.1 Email Types

2.1.1 Routine Message

Emails containing routine messages are used for a variety of purposes within businesses as a standard form of communication. They can include conveying information, giving updates, requesting information or action, making appointments, and more. They can be addressed to specific people or groups. The recommended structure typically contains a subject line, salutation, introduction containing the main idea, body with details and reasons, closure with a specific request (if applicable), and signature. The approach is direct. The tone might be formal or informal. Routine emails are typically sent in real-time to facilitate speedy and effective communication. Nonetheless, effective routine emails require attention to email etiquette, which includes clear subject lines and careful proofreading. In the present research, emails containing a routine message were generated using the prompt below:

Prompt 1. Routine message

> Write a complaint letter addressed to an airline manager (*add any name*) about a recent unpleasant flight experience. Expected word count 150 – 200.
>
> Instructions:
>
> - Choose an airline that you have flown with recently.
> - In the letter, describe your experience with the airline, including any issues that you encountered (e.g., long wait times, rude staff, incorrect booking, delay, etc.).
> - Be specific about what you would like the manager to do to rectify the situation (e.g., refund, discount, apology, etc.).
> - Use a formal and respectful tone in the letter.
> - You should write four paragraphs.
> - Follow the direct approach and include an appropriate subject line, salutation (Dear...), and closing.

2.1.2 Negative Message

In a professional context, emails conveying negative information should adopt a polite and indirect approach. This method allows for the delivery of difficult truths and aids in problem-solving while maintaining professional relationships. Such emails typically begin with a buffer to soften the message, followed by context or an explanation for the unfavorable news. They also present a positive perspective when turning down requests, responding to complaints, giving feedback, or talking about delays. Even when expectations are unfulfilled, it is vital to maintain empathy and respect for the recipient's feelings. Offering the recipient hope (when appropriate) or showing a way forward by proposing solutions or future cooperation is an effective conclusion. In sending negative messages, communication must be transparent, prompt, and clear. The following prompt was used to generate emails containing a negative message:

Prompt 2. Negative message

> Write a 200- to 250-word email using the indirect approach. Organize it into five paragraphs. Here is what will happen:
>
> You are the customer service representative for an online retailer. A customer (*add a name*) has contacted you requesting a refund for a product they bought (*add a reason why*). Your company's policy clearly states that refunds are only accepted in certain cases (*explain in your own words when it is acceptable*). You need to write an email to the customer explaining that their request for a refund cannot be fulfilled. Make sure you express your regret and offer alternative solutions or options.
>
> Follow the pattern: buffer, reasons, additional information, bad news, positive angle, and respectful close.

2.1.3 Persuasive message

Persuasive communication, characterized by strong arguments and emotional appeals, seeks to alter the attitudes, beliefs, or behaviors of its audience. The AIDA principle, which stands for Attention, Interest, Desire, and Action, is a framework frequently used to organize persuasive messages. 'Attention' involves captivating the reader with an engaging subject line or introductory paragraph. 'Interest' involves supplying information that sparks the recipient's curiosity once their attention has been captured. The 'Desire' component emphasizes the benefits of a proposal, plan, or request. To foster a desire, it is crucial to present a compelling argument and illustrate how the proposal or idea meets recipients' needs, resolves their issues, or satisfies their desires. 'Action' entails incorporating a strong call to action at the conclusion. It is essential to specify the next step one wishes the receiver to take, such as endorsing your proposal, scheduling a meeting, or executing another distinct action. In the present research, emails containing a persuasive message were generated using the prompt below:

Prompt 3. Persuasive message

> Write a 4-paragraph persuasive message using the AIDA template to be sent to your manager. You need to convince him to allow you to work remotely. Write from 200 to 250 words in four paragraphs:
>
> - Attention: get your manager's attention with a question or statistics (or else).
> - Interest: write about the selling points and mention the benefits of your idea.
> - Desire: what are the benefits and objections; awaken the desire.
> - Action: ask him for action.

2.2 LLM overview

2.2.1 ChatGPT 3.5

ChatGPT is a chatbot developed by OpenAI that is based on a large language model. The GPT-3.5 model, an OpenAI-created generative pre-trained transformer, serves as the basis of ChatGPT. As a general-purpose language model, the GPT-3.5 transformer model is initially trained on a sizable amount of openly available text. The model is then further modified for conversational applications using supervised and reinforcement learning techniques. The user is in charge of determining the conversation's duration, organization, level of detail, style, and language. The primary purpose of the chatbot is to simulate human interactions, but it is also equipped to perform a broad variety of additional tasks (Rudolph et al., 2023). As it was trained on unfiltered text, GPT-3.5 is prone to bias and erroneous data. Furthermore, ChatGPT encounters "hallucinations", which are false responses that seem accurate. (Rudolph et al., 2023). All in all, a wide information base, conversational nature, and availability are its benefits; limited context understanding is its drawback.

2.2.2 Bard

Bard, which was renamed to Gemini in February 2024, is a conversational, generative artificial intelligence chatbot developed by Google. It was initially based on the LaMDA family of large language models (LLMs), but it later transitioned to the more powerful and versatile PaLM 2 model. This advanced model is more powerful and versatile than the previous version of Bard. It is able to perform a wider range of tasks. For example, PaLM 2 can generate more creative text formats, translate languages more accurately, and answer questions more comprehensively. Bard, powered by PaLM 2, was released to compete with the rival ChatGPT. According to CNBC, several comments indicate that Bard's responses are up-to-date with current events (CNBC, January 31). The competition between Google's Search Engine and Microsoft's Bing is anticipated to intensify, making for an exciting rivalry (Reuters, 2023). Despite the backing of Google, Bard was relatively underused software, with very few applications and studies on it. Recently, its use has significantly increased, as it has been connected with Google's entire toolkit, which includes YouTube, Google Drive, Google Flights, and others. Additionally, it now includes

additional features, including fact-checking capabilities and multilingual communication. Even though Bard was created with education in mind and can align its explanations with specific curricula, it is not limited to themes covered in the classroom. It can perform sophisticated arithmetic and reasoning, code, and retrieve data from Google's range of other services. Its strengths lie in its educational focus, its ability to align explanations with specific curricula, and its incorporation of instructions to enhance learning.

2.2.3 Llama 2.0

A foundational, 65-billion-parameter large language model, Llama 2.0, is developed by Meta. Its developers claimed that the 13 billion-parameter version of the model outperformed the much larger ChatGPT on several NLP tasks. Touvron et al. (2023) showed that LLaMA is capable of outperforming Bard and ChatGPT on several NLP tasks. Additionally, Llama 2 is open-source software. Its relatively small size and open-source nature make it an attractive alternative to other existing chatbots. It has the following pros: Multimodal learning experience, hands-on practice—allowing students to actively engage with the subject matter and reinforce their understanding—and adaptive learning: Llama can adapt to the user's progress and adjust the difficulty level of the content accordingly. Its biggest disadvantage in the context of this research is its availability and speed. During the period of this study, the researchers encountered several challenges while using Llama AI. The difficulties in accessing the system, the extended time it took to generate responses, and the system's heavy dependence on network quality could be particularly frustrating for in-class activities.

2.2.4 Bing Chat

Recently released, Microsoft's AI-powered Bing Chat exhibits a number of features comparable to ChatGPT. It can write letters, summarize articles, generate conversational text, finish coding tasks, and respond to complex questions. Bing Chat includes internet access, allowing it to provide up-to-date news in real time. While Bing Chat can offer insights into events as recent as the day before, ChatGPT's free edition is limited to data up to 2021, making it unable to address recent changes. Some of the key features of Bing Chat's interface include a text input area, a "New Topic" option to start fresh conversations, and source citations that are provided in the message below. It gives users the option between traditional search results and a chatbot. It also proposes follow-up queries to increase engagement. Users can customize responses to their tastes by choosing one of three conversational modes: more creative, more balanced, or more precise. The user interface offers access to the Microsoft account as well as a response counter and feedback system. In addition, Bing Chat provides features like sidebar option comparisons, online browsing, summarizing, and decision-making. In November 2023, Microsoft declared that Bing Chat and Bing Chat Enterprise would now be known as Copilot. This is the name that's been given to Microsoft's other generative AI services.

3. Results

The results section conducts a separate examination of the data collected from three different types of AI-generated emails. The analysis starts with routine message emails, followed by an exploration of negative and persuasive ones.

3.1 Analysis of Routine Emails

The analysis of the generated emails shows the following: With regard to the subject line, Bard generated a clear and concise one, while ChatGPT produced a more verbose one: "*Complaint Regarding an Unpleasant Flight Experience.*" Llama's subject line, albeit shorter than ChatGPT's, exceeds the recommended number of words (three to five) and lacks specific details. Bing, on the other hand, provides details in the subject line, which might be more appropriate for the second paragraph. Regarding the salutation, Bard's version lacks personalization with "Dear Manager", while ChatGPT's version, "Dear [Manager's Name]", also needs personalization. If the appropriate etiquette is not followed, it might lead to potential impoliteness. Llama and Bing use a generic name, *"Dear Mr. Smith,"* which can serve as a more instructive example for students.

Table 1. Routine Email Evaluation Results (Content and Format)

	Subject Line	Salutation	Structure	Number of paragraphs	Number of words
Bard	1	0.5	0.5	0	1
ChatGPT	0.5	0.5	1	0	0
LLaMA	0.5	1	1	0	0
Bing	0.5	1	1	0	0

|	Clarity of Purpose	Relevance	Support	Call to Action	Adaptation
Bard	1	1	0	0.5	1
ChatGPT	1	1	0.5	1	0.5
LLaMA	1	1	0.5	1	1
Bing	1	1	0.5	1	1
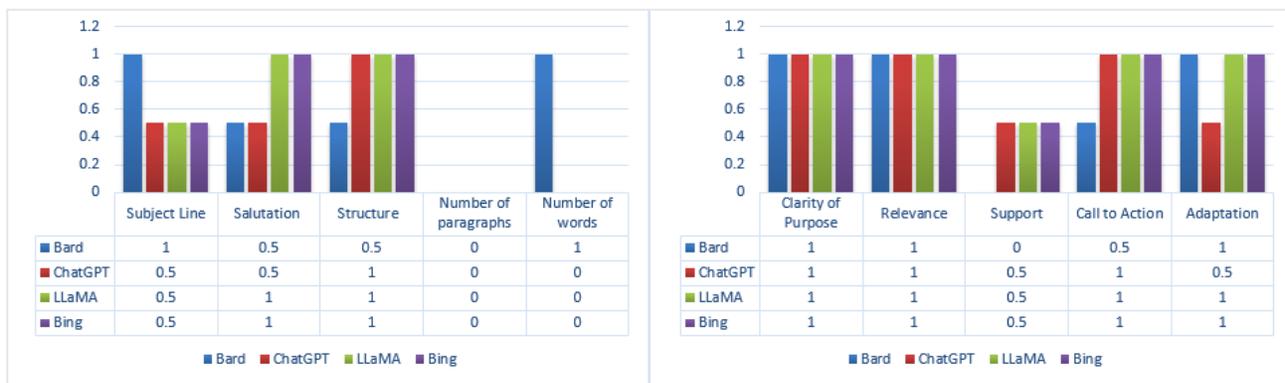
In terms of structure, Bard's composition of the routine email is characterized by excessive details in the first paragraph and some disorganization in the fourth and fifth paragraphs. Some redundancy is present, and as a result, the structure seems cluttered. ChatGPT,

similarly, produces quite a few more details than necessary in the first paragraph, dedicating two subsequent paragraphs to a comprehensive explanation of the problem. Its call to action is more pragmatically aligned with the issue encountered than Bard's. Llama, on the other hand, generates a well-organized email, but it systematically elaborates on the problem over three distinct paragraphs, covering all the options presented in the scenario. As a result, it does not acknowledge that, similar to ChatGPT and Bing Chat, only one of these suggestions should be selected. This, while offering the user the option of choosing the needed content, may present a logical inconsistency. Llama's use of bullet points for the call to action, however, serves as a great example of formatting that enhances understanding.

None of the LLMs adhered to the requested number of paragraphs. Bard, ChatGPT, and Llama each generated emails consisting of five paragraphs, exceeding the requested four. Bing's response, on the other hand, is divided into six paragraphs. Bard is the only AI model that followed the requested word count and stayed within the specified limit. Furthermore, Bard is the only model that accurately responds to the prompt by selecting and addressing just one suggested situation, thereby maintaining a concise response. The other AI models, namely ChatGPT, Llama, and Bing, exceed the 300-word limit and deviate from the prompt's guidelines for scenario selection.

Supporting evidence in Bard-generated emails could be more comprehensive. Additionally, in the context of the problem, its call to action might appear somewhat exaggerated. This could be attributed to the fact that it incorporated two out of the three options offered in the prompt, as it was unable to discern the most appropriate one. The other LLMs also generated the content for all the given options within the prompt, which resulted in paragraph excess. ChatGPT's response demonstrates a high level of formality in its wording, which may make it less suited to the audience's needs. Conversely, Llama's response sometimes lacks finesse. None of the bots included supplementary information or referenced a potential attachment, thereby reducing their persuasiveness.

Table 2. Routine Email Evaluation Results (Diction and Tone)

	Grammar and Spelling	Vocabulary	Conciseness	Sentence Structure	Punctuation and Formatting
Bard	1	1	0.5	1	1
ChatGPT	1	0	0.5	0.5	0.5
LlaMA	1	0.8	1	1	1
Bing	1	1	1	1	1

	Formality	Politeness	Emotion	Audience Awareness	Consistency
Bard	1	1	1	1	0.5
ChatGPT	0	1	0.5	1	0.5
LlaMA	1	1	1	0.5	0.5
Bing	1	1	1	0.5	1
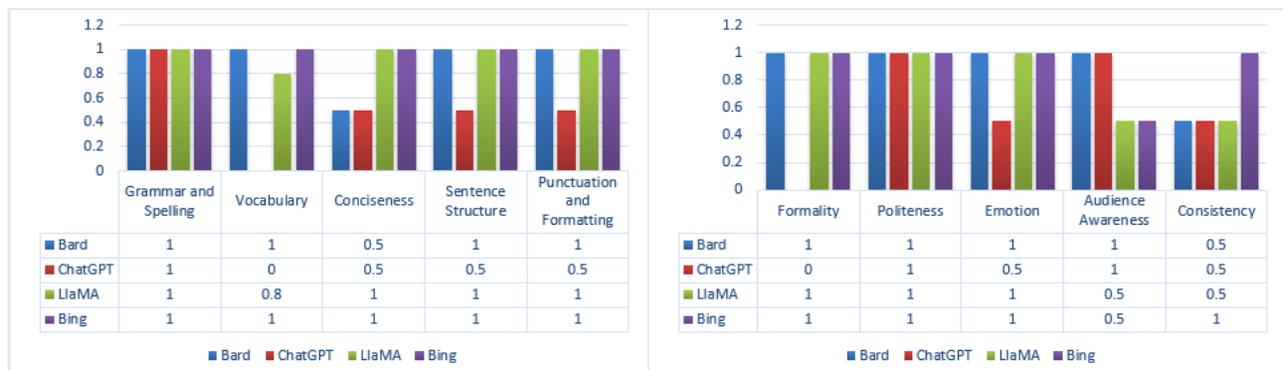
For grammar and spelling checks, an AI tool called QuillBot was used in addition to the authors' assessment. It reported two errors in comma usage and one suggestion to change from passive to active voice in the Bard routine email. The evaluation of ChatGPT's email advised using active voice instead of passive. Moreover, the sentence "I kindly request that the airline take" should be altered to "I kindly request that the airline takes" for reasons of formality. However, no punctuation errors were found. Llama also received two suggestions to use active voice and had one punctuation error and one capitalization error. In Bing's response, there was again a suggestion to opt for active voice. There was also a tense sequence error in the sentence: "Secondly, when I finally boarded the plane, I found out that my seat was changed," where it should read, "had been changed." Three punctuation errors were also identified. Generally, the most common editing suggestion was to change from passive to active voice, as active voice is often considered more desirable in business correspondence for reasons of clarity, responsibility, conciseness, professionalism, influence, and occasionally, "you" attitude.

Finally, the tone in LLM-generated emails (and ChatGPT's in particular) tends to come across as emotionally neutral and overly formal, potentially falling short of successfully expressing sentiments. Llama's response also lacks a nuanced understanding of the recipient's perspective, but to a lesser degree than ChatGPT. Bing Chat is slightly better in this regard, as it includes a clear deadline and a sense of urgency towards the end. The following closing paragraph: *"Please respond to this letter within 10 days and let me know how you intend to address this matter. You can reach me by phone at +381 11 1234567 or by email at user@email.com. I look forward to hearing from you soon and getting a good response.* This not only sets expectations but also ensures that the recipient has all the necessary details to respond promptly, aligning with established guidelines in business communication.

3.2 Analysis of Negative Message Emails

Bard's negative message email could benefit from a more tactful strategy. It plunges directly into the issue, bypassing the recommended buffer and indirect approach. The AI email response is clearly written, adhering to the specified word limit. However, it does not follow the requested five-paragraph structure, opting instead for a seven-paragraph structure. The overall quality is compromised by the lack of supporting evidence. There are some errors present, such as redundancy in the following sentence: "You can exchange the product for a different product". Additionally, the email content is frequently conventional in character, with little emotional investment and insufficient consideration for the audience's perspective.

ChatGPT follows the requested format with exactly five paragraphs. It effectively employs the indirect approach. The email's 331 words, however, go beyond the permitted word count. Although it gives three potential answers to the problem at hand, the suggestions are a little vague and might use more detail. The response is free of major grammatical errors. It still tends to be verbose, and the tone is somewhat more formal than is typically required for this type of email. To improve readability, the largely compound or complex sentence structure should be diversified by incorporating simple sentences. For example, a compound sentence like "Thank you for choosing [Your Company Name], and we are here to support you throughout your shopping journey", could be replaced by two simpler, more straightforward sentences. In terms of audience awareness, ChatGPT outperforms Bard since it is more conscious of the recipient's perspective. It succeeds in adding some emotion. Positive aspects of ChatGPT's method for conveying bad news include its combination of formality, clarity, and a hint of emotional resonance.

Table 3. Negative Message Email Evaluation Results

	Subject Line	Salutation	Structure	Number of paragraphs	Number of words
Bard	1	1	0	0	1
ChatGPT	1	1	1	1	0
LlaMA	1	1	0.5	0	0
Bing	0	0	1	1	1

	Clarity of Purpose	Relevance	Support	Call to Action	Adaptation
Bard	1	1	0.5	1	0.5
ChatGPT	1	1	0.8	0.8	1
LlaMA	1	1	1	1	1
Bing	1	1	0.5	1	1

	Grammar and Spelling	Vocabulary	Conciseness	Sentence Structure	Punctuation and Formatting
Bard	1	1	1	1	0.5
ChatGPT	1	0.8	0.5	0.5	1
LlaMA	1	1	1	1	1
Bing	1	1	1	1	1

	Formality	Politeness	Emotion	Audience Awareness	Consistency
Bard	1	1	0.5	0.5	1
ChatGPT	0.8	1	0.8	1	1
LlaMA	1	1	0.5	1	1
Bing	1	1	0.8	1	1

Using Llama's email format can be challenging, as it often transforms into lengthy, continuous text when copied and pasted. The generated response comprises six paragraphs and totals 323 words, which is more than the recommended word number. The email starts with a concise statement of the purpose, followed by a buffer. It does not strictly adhere to the indirect approach. Additionally, Llama's email shows inconsistency in the usage of pronouns that alternate between "I" and "we". A few minor punctuation errors occur as well. Incorporating a more pronounced "you" attitude would enhance the reader-centric tone.

Bing-generated negative news email is missing a subject line, which is a significant oversight. While the email does provide some details, it could benefit from additional information for better clarity. The language is grammatically sound. Adopting a more reader-centric "you" attitude could enhance the email, making it less generic and more engaging. By incorporating specific details and addressing the recipient's needs, the overall effectiveness of the communication could be significantly improved.

3.3 Persuasive Email Analysis

Bard's persuasive email, while demonstrating a clear paragraph structure with headings such as "attention" and "interest", appears to function more as a teaching tool than a persuasive message. The salutation, "Hi [Manager's name]," is overly simplistic and direct, deviating from the expected formality. The main idea is presented immediately, adhering to a direct approach that does not align with the prompt requirements. The introduction lacks an attention-grabbing element (a question or statistics), as requested in the prompt. The sections labeled "interest" and "desire" seem to be preoccupied with the writer's perspective, neglecting to consider the audience's point of view. Additionally, the email's arguments are weak. The use of pathos (emotional appeal) and logos (logical appeal) to effectively persuade the reader is very limited. The call to action in the email could be more assertive. It falls within the specified word limit (241 words), but it fails to provide more relevant details to bolster the message. On the positive side, the grammar in Bard's email is sound, contributing to its overall clarity and readability.

Despite its lack of brevity, the subject line of ChatGPT's email effectively captures the reader's attention. It follows the provided guidelines and takes an indirect approach. The email starts with a buffer that skillfully integrates both statistical data and a question, as recommended

in the prompt, though one option would be sufficient. The email's length, at 297 words, slightly exceeds the prescribed word limit. The email demonstrates a degree of audience awareness by acknowledging the evolving global landscape and the necessity to take into account the advantages of remote work for the team as well, and not only for the individual. Logos is incorporated into the email. The logical appeal is reinforced by strong arguments and only a general reference to studies. Counterarguments are also addressed, demonstrating a well-rounded approach. Although the email has strong logos, it might use additional emotional appeal to engage the reader more deeply. The fifth paragraph might not be necessary, and its omission would simplify the communication of the message. Although vocabulary is stronger than in normal emails, there are still some instances of verbosity, especially when adjectives and intensifiers are used excessively. On the plus side, there are no grammar errors, which adds to the email's professionalism and general clarity. To improve readability and maintain reader interest, a mix of simple, compound, and complex sentences could be employed.

Table 4. Evaluation Results of Persuasive Email Messages

Chart 1 (top left)

	Subject Line	Salutation	Structure	Number of paragraphs	Number of words
Bard	0	0.5	0	0	1
ChatGPT 3.5	0.5	1	1	0.8	0.5
Llama 2	0.5	1	0.5	1	0
Bing	0	0	0.5	0	1

Chart 2 (top right)

	Clarity of Purpose	Relevance	Support	Call to Action	Adaptation
Bard	1	0.8	0.5	0.8	0.8
ChatGPT 3.5	1	1	1	1	1
Llama 2	1	1	0.8	1	1
Bing	1	1	0.5	1	1

Chart 3 (bottom left)

	Grammar and Spelling	Vocabulary	Conciseness	Sentence Structure	Punctuation and Formatting
Bard	1	1	1	1	1
ChatGPT 3.5	1	0.5	0.8	1	1
Llama 2	1	1	0.8	1	1
Bing	1	1	1	1	1

Chart 4 (bottom right)

	Formality	Politeness	Emotion	Audience Awareness	Consistency
Bard	0.8	1	0.5	0.5	1
ChatGPT 3.5	1	1	0.8	1	1
Llama 2	1	1	0	0.5	1
Bing	1	1	0	0	1

Llama's email takes a direct approach, outlining its goal in the very subject line. It also contains the word "request", which is normally advised against in persuasive business emails. The opening paragraph references statistics from a recent Upwork study. However, the authors could not find confirmation for the numbers stated. The writing style in Llama's email is more natural and less generic and formal than ChatGPT's, which contributes to a more engaging tone. However, there is room for improvement in terms of persuasion, particularly in the area of pathos (emotional appeal). To engage the receiver more deeply, the email should use more emotional components. The email uses some logos at the very beginning by citing statistics; however, it should be strengthened with additional data to support the claims. Although there is a high degree of audience awareness, the frequent use of the first-person singular pronoun "I" diminishes "you" attitude, which is normally more compelling. The email's 330 words are over the allotted word count. Condensing the language to keep the recipient's attention and interest would be welcome.

The email generated by Bing employs a direct approach, concisely delivering information. However, at 218 words, the email appears somewhat brief, suggesting the need for further detail. The email's use of logos (logical appeal) and pathos (emotional appeal) is relatively weak. It lacks statistical or supporting data that could strengthen the argument. Furthermore, there is a lack of pathos, as the email does not contain emotional elements that would engage the recipient on a deeper level. A significant area for improvement is the absence of a "you" attitude, which results in the communication seeming self-centered. Adopting a more reader-centric approach, focusing on the recipient's needs and benefits, could enhance the persuasiveness of the email.

4. Discussion

This section begins by providing an overview of each LLM's performance across three email types, as well as the overall results. Subsequently, it looks into the implications of these findings for the rubric and task design. Finally, it concludes by offering teaching recommendations.

4.1 Overview of LLMs' Performance

Bard's performance in generating subject lines exhibits inconsistency. While it successfully crafts suitable subject lines for the initial two types of emails, it underperforms in the third type, the persuasive email. Bard's ability to generate salutations that align with the requested level of formality could be improved. Moreover, it struggles to follow the expected email structure, managing to do so correctly only in the simplest, direct-approach structure of routine emails. In terms of paragraph generation, Bard consistently produces more paragraphs than required, surpassing other chatbots. However, it does meet the specified word count.

Additionally, Bard falls short in routine emails by not providing sufficient details, a similar issue observed in negative and persuasive ones. The call to action is average, with a lower performance in routine and persuasive emails and a stronger performance in the negative one. However, the negative message call to action is not aligned with the previously mentioned arguments.

In terms of clarity of purpose and relevance, Bard performs well. Bard's adaptation to the audience is moderately successful. Bard consistently performs well in grammar, vocabulary, punctuation, formatting, politeness, and sentence structure. However, it lacks conciseness in the routine email and formality in the persuasive one. Bard's effectiveness in addressing the emotional aspect and showing awareness of the audience's needs is limited. The overall tone is generally consistent, except for routine emails, where it deviates.
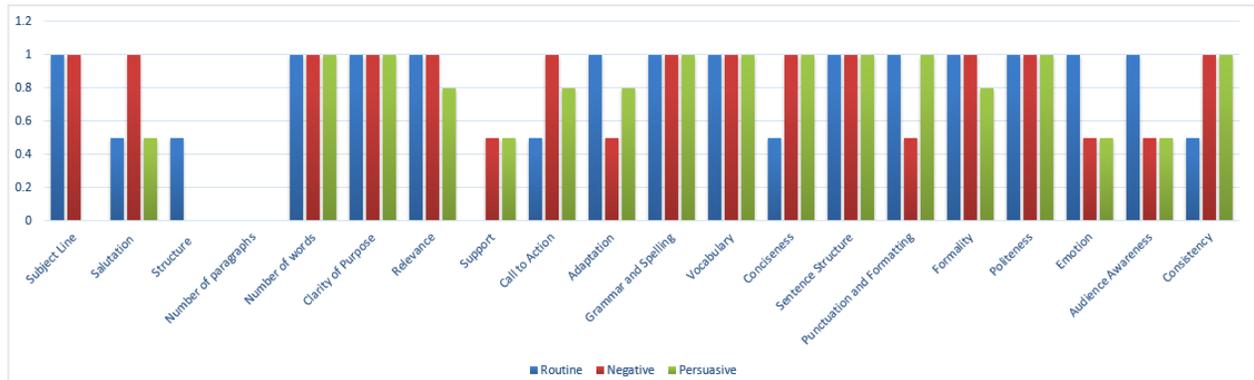
Figure 1. Bard performance evaluation as per rubric

ChatGPT tends to generate verbose subject lines, requiring revision for conciseness. The salutation allows for personalization, which is a positive aspect as long as students are familiar with the correct form of address. In terms of structure, ChatGPT is consistent in following the requested format and often meets the expected number of paragraphs. However, it frequently falls short of the specified word limit.

ChatGPT performs well in several areas, including clarity of purpose, relevance, grammar and spelling, punctuation and formatting, politeness, and audience awareness. While it generally succeeds in providing a call to action and in adapting to the audience, there is potential for improvement in terms of providing supporting evidence.

To get to the desired outcome, it could benefit from additional prompting when it comes to vocabulary, conciseness, and sentence structure. It occasionally resorts to excessive formal language. Emails generated by ChatGPT consistently maintain a polite tone and demonstrate an understanding of the audience's needs in most instances. However, they generally lack emotional resonance.
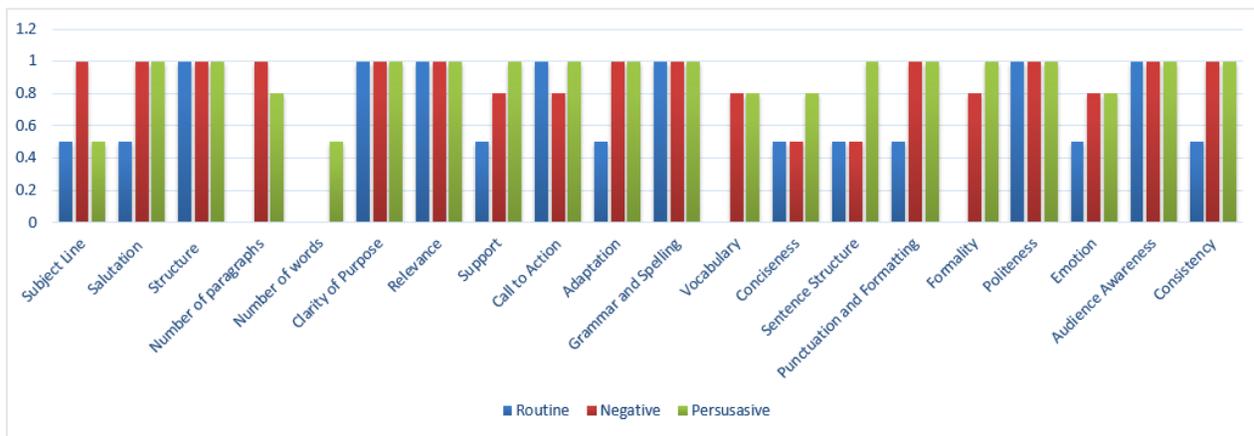

Figure 2. ChatGPT 3.5 performance evaluation as per rubric

Emails generated by Llama excel at crafting salutations, demonstrating a commendable understanding of suitable greetings. However, they require intervention concerning subject lines and overall structure. Llama does not adhere to the specified number of paragraphs and word limits.

In terms of content, Llama exhibits clarity of purpose, relevance, and effectiveness in issuing a call to action. It demonstrates adaptability to the audience and provides a certain level of support. Additionally, Llama performs strongly in areas such as grammar, vocabulary, conciseness, sentence structure, punctuation, and formatting.

Llama maintains an appropriate tone in terms of formality and politeness, although consistency can fluctuate. However, it falls short of conveying the appropriate emotions, particularly in negative and persuasive emails. There are instances where it struggles to fully articulate the recipient's perspective.
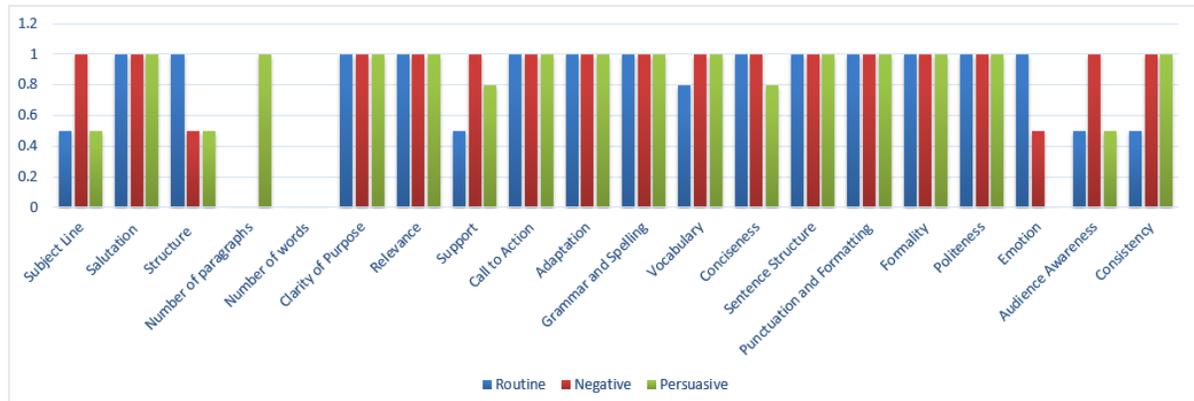
Figure 3. Llama 2.0 performance evaluation as per rubric

Emails generated by Bing consistently omit subject lines when they are not explicitly requested in the task instructions. While Bing correctly includes the appropriate salutation in the routine email, it generates inappropriate ones in the other two types of emails. Bing demonstrates an ability to adhere to the requested structure, but frequently fails to generate the requested number of paragraphs. It complies with the requested word count in the last two types of emails.

In terms of content, Bing Chat excels in delivering clarity of purpose, relevance, and effective calls to action. It is capable of adapting to the audience's needs but tends to lack supporting evidence. Bing's emails have no errors in grammar, vocabulary, conciseness, sentence structure, punctuation and formatting.
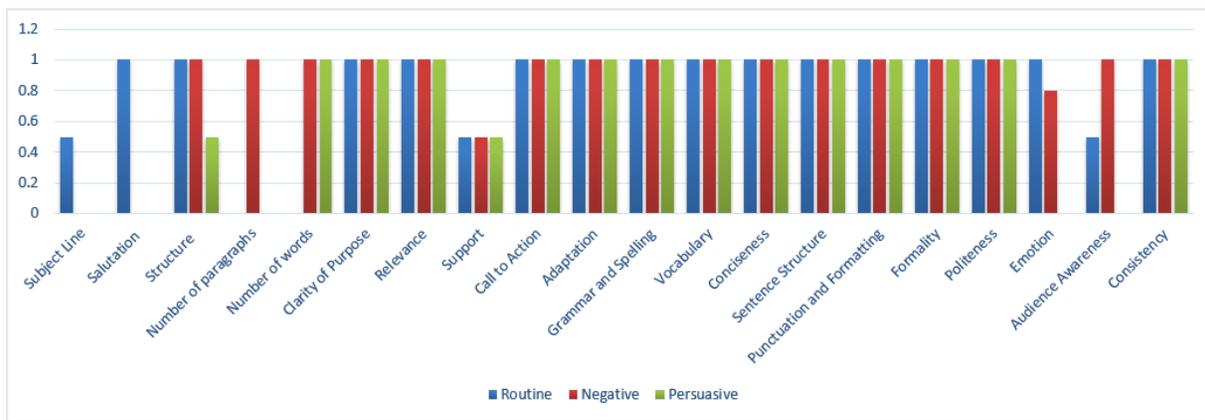

Figure 4. Bing Chat performance evaluation as per rubric

Regarding tone, Bing maintains an appropriate level of formality, politeness, and consistency throughout its emails. However, similar to other chatbots, it struggles with conveying emotions appropriately and understanding the recipient's perspective.

Table 4. LLMs' Overall Performance Comparison

	Routine email	Negative news email	Persuasive email	Total
Bard	15.5	15.5	14.2	45.2
ChatGPT 3.5	11.5	17	18.2	46.7
LlaMA 2.0	15.8	17	16.1	48.9
Bing	16.5	17.3	14	47.8
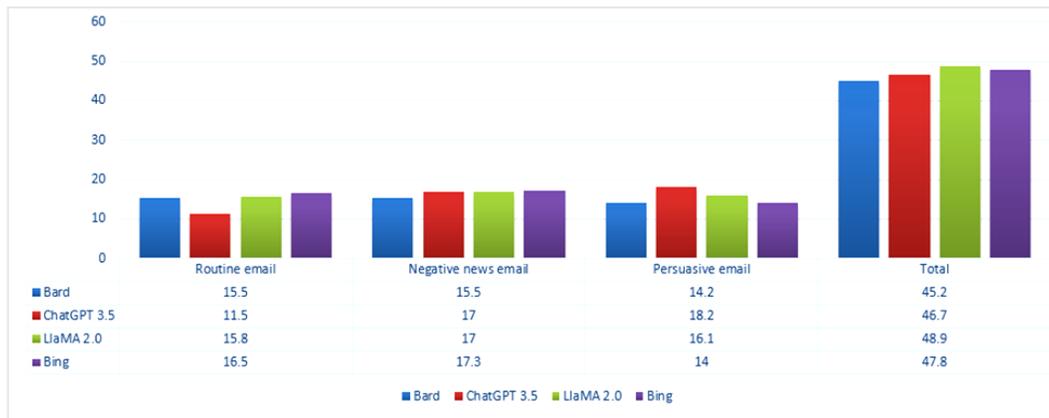
The preceding graph shows that Llama 2.0 achieved the highest performance with a score of 48.9 out of 60, followed closely by Bing at 47.8, ChatGPT 3.5 at 46.7, and Bard, which scored 45.2 out of 60. This study aligns with findings in other fields where Llama similarly outperformed other LLMs (Touvron et al., 2023; Calonge et al., 2023).

4.2 Teaching Implications

Understanding the unique strengths and weaknesses of each LLM, as well as their collective characteristics, is vital for several reasons. First and foremost, it can aid in the student evaluation process, particularly when there are suspicions of academic dishonesty. For example, an LLM might display outstanding language skills but lack in areas such as emotional engagement, audience awareness, or maintaining the appropriate level of formality. They might exceed the suggested paragraph count or demonstrate poor email organization. Thus, despite the fact that the emails exhibit a remarkable mastery of language, the aforementioned inconsistencies could trigger suspicions of AI-assisted cheating. This, in turn, necessitates the development of more comprehensive rubrics and/or specialized task designs to address the issue effectively.

4.2.1 Rubric Design Recommendations

To address the identified strengths and weaknesses of LLMs in generating emails, the following modifications to the grading rubric are suggested.

1. Subject Line Evaluation: The evaluation criteria outlined in the paper for assessing a subject line (clarity, conciseness, and relevance) could be broadened to include compliance with a specified word limit. This would help curb the tendency of LLM-generated emails to have detailed and lengthy subject lines, as observed with ChatGPT 3.5.

2. Emphasis on Supporting Evidence: The rubric should emphasize the importance of providing supporting evidence. Students should be encouraged to strengthen their arguments with verifiable statistics, quotes, or other forms of evidence. This would address the issue of unverifiable evidence in AI-generated emails, such as the persuasive email attention getter generated by Llama.

3. Call to Action Effectiveness: The rubric should clearly define the expected level of effectiveness and appropriateness of the call to action. It should also assess its coherence with the overall unity of the email.

4. Email Unity: The rubric should assess the unity of the email, ensuring it is cohesive and well-organized. This would help address the issue of disjointed and unclear emails generated by LLMs, such as Bard routine email where the complaint and adjustment request are not aligned.

5. Language: The rubric should include criteria that assess the use of industry-specific terminology, conciseness, politeness, and formality level appropriate for the specific business context. It is advisable to provide separate scores for each aspect.

6. Personalization: The rubric should provide separate scores that evaluate the personalization of the content. Examples of generic language that should be avoided include phrases such as "I am writing to complain…" or "I hope this email finds you well." Instead, students should strive to use distinctive language that reflects their unique voice and style while still maintaining the professional tone and appropriateness of the communication.

7. Audience: The rubric should evaluate the effectiveness of the email in adapting to the specific needs, preferences, and expectations of the audience. This can be achieved by assessing how well the email tailors its message, tone, and language to the target audience's unique characteristics, such as their cultural background, personal values, and interests. Additionally, the rubric should introduce a section that evaluates the incorporation of emotional elements in the communication. This section should penalize emotionally neutral or detached content, and reward the effective integration of emotional elements that align

with the purpose of the message, such as using storytelling techniques, humor, or emotional appeals to engage the audience and enhance the overall impact of the message. By doing so, the rubric can encourage students to produce business messages that are not only informative but also engaging, persuasive, and memorable.

4.2.2 Task Design Recommendations

With regard to task design, the research findings validate some effective strategies. The first one involves providing concise assignment guidelines, thereby granting students the freedom to add their unique ideas and details. Examining the AI-generated emails shows that they fall short in providing sufficient supporting evidence, especially statistics. Even when such data was provided, its reliability was subject to doubt. Second, business writing instructors should provide thorough assignment materials that highlight that students have to include any additional information required to compose business letters properly in accordance with the standards and materials taught in class (Al-Afnan et al., 2023). For instance, when composing business correspondence, students are instructed that it is crucial to provide clients with comprehensive information. This includes sharing insights into any potential issues that may have arisen. When specifics are missing, students need to attend to what an imaginary audience needs, as instructed, and provide all the missing details. The aforementioned research highlighted that in situations where students were required to provide original information to address a problem, ChatGPT did not sufficiently delve into the specifics. The third option, particularly useful during in-class testing, is to incorporate multiple choices within prompts, such as the following segment from the routine email scenario:

Ex.1 In the letter, describe your experience with the airline, including any issues that you encountered (e.g., long wait times, rude staff, incorrect booking, delay, etc.).

Presenting the scenario with various options for potential problems (claims), solutions (adjustments), and so on, would require LLMs to make a choice and remain consistent. However, as observed, they tend to either choose one or generate an adjustment that is excessively strong, consequently breaking the unity—as is the case with Bard—or write a separate paragraph for each option, resulting in issues related to the number of paragraphs, words, unity, and so forth.

4.2.3 In-class Application

Finally, the findings of the study are directly relevant to teaching implementation. Educators who happen to see knowledge building as inextricably related to communication (Mnasri and Papakonstantinidis 2023), especially if they implement some of the AI-assisted tasks in their classrooms, would find the following recommendations useful and tangible:

When generating emails, Bard encounters difficulties with the formality of salutations, inconsistent subject lines, and poor email organization. Implementing Bard in teaching would involve training students (or prompting or preparing the Chatbot itself) on how to consistently create subject lines, write acceptable salutations, and develop email structures. Although it does well in terms of clarity and relevancy, its emails lack detail. While it has a solid command of language and a decent ability to adapt to its audience, it lacks the proper formality. A review of Bard-generated emails for conciseness, formality, and emotional engagement could serve as a valuable classroom activity. When students can recognize these errors, identify them, and correct them, they move one step closer to avoiding such mistakes in their own writing.

ChatGPT, on the other hand, tends to generate verbose subject lines, and often exceeds the word count limit. It excels in audience awareness, politeness, relevancy, and clarity but usually leaves out crucial supporting details. Educational applications of ChatGPT should focus on designing prompts or in-class activities to enhance subject lines, use of terminology, conciseness, formality level, call to action, and supporting evidence. For students, ChatGPT could serve as a tool to foster vocabulary expansion and enhance the formality of their communication. For corporate communication, it should be fine-tuned to broaden its vocabulary and formalize its language, aligning with the specific requirements and nuances of the particular business context.

Salutations and language proficiency are among Llama's strong points, but it needs improvement in structural elements, crafting compelling subject lines, and fostering emotional engagement. Therefore, it's crucial to guide students on how to effectively use Llama. This includes teaching them to prompt Llama appropriately, correct its output when necessary, write engaging subject lines, improve the structure of emails, and infuse more emotional depth into its communication.

Bing-generated emails excel in clarity and courtesy, yet there are areas that could benefit from further refinement. For educational applications, the emphasis should be on guiding the LLM and students to create suitable salutations, improve the structure, provide substantiating evidence, and enhance emotional engagement while understanding the audience's viewpoint. It is worth noting that both Bard and Bing represent valuable resources, as they offer insightful commentary and include directives that foster learning. They appear to be specifically designed for educational purposes.

The above-mentioned findings can guide both educators and students in using conversational AI for more precise and effective writing, while introducing conversational AI tools and their features; conducting exercises to improve writing proficiency and accuracy using these models; discussing best practices for employing each language model in business writing, including advice on adjusting tone and style; or facilitating feedback. While it falls outside the scope of the article, it's noteworthy that AI's performance as a feedback generator has been less than satisfactory. Different LLMs assessed the same email differently, resulting in varied summative and formative grades. This aspect remains to be thoroughly tested.

In terms of limitations, the study focuses on only four LLMs (ChatGPT 3.5, Llama 2.0, Bing Chat, and Bard), which may not represent all

accessible models today. Furthermore, the evaluation focuses on specific email scenarios and targets three email types, potentially limiting its applicability to all business writing and overlooking subtleties in other kinds of email. The use of a rubric might involve a level of subjectivity. Finally, a comparison with human-written emails could potentially enrich the findings, which could be considered a next step in the research.

Future studies should aim to further enhance our understanding of AI's role in business writing and professional correspondence and its potential to revolutionize professional communication, both internally and externally. Some of the promising areas of research include: a) fine-tuning AI models for specific business contexts; b) developing AI feedback and training tools; c) assessing inter-cultural implications; d) examining human-AI collaboration in business writing; e) addressing ethical concerns pertaining to plagiarism; f) forecasting the long-term impact of AI on business communication; and d) continuing comparative evaluations of AI models. Finally, given that AI particularly lacks emotional understanding, special attention should be given to this particular component of LLM-generated emails and to their ability to adapt to the audience and their needs.

5. Conclusion

In conclusion, the LLMs that were tested on their email writing show both strengths and weaknesses, doing well in terms of language and clarity but failing in terms of subject lines, structure, emotions, and audience perspective. Due to its inability to grasp emotional context and understand diverse data, AI may arrive at incomplete or incorrect outcomes (Shane, 2019). The statement still appears to remain somewhat valid, despite the ongoing AI revolution.

The study also shows that Llama 2.0 exhibits the highest consistency across diverse email types and has achieved the highest performance with a score of 48.9 out of 60. It is followed closely by Bing at 47.8, ChatGPT 3.5 at 46.7, and Bard, which scored 45.2 out of 60.

Considering the findings of the evaluations of LLMs in email writing, it becomes clear that these models have huge potential to improve various aspects of business correspondence. They show strengths in language proficiency, clarity of purpose, and relevance, all of which are essential for effective communication in the business world.

However, there are some obvious shortcomings, especially when it comes to proper subject lines, the right format, and audience emotions and perspectives being clearly communicated. While LLMs can assist in several aspects of business communication, they cannot replace human judgment, creativity, or empathy, as this research has confirmed. Effective correspondence requires a personal touch, emotional intelligence, and a deep understanding of the business context. These drawbacks show how LLMs must be developed further in order to better meet the needs for business communication.

Even with high-quality data, AI is greatest at prediction; humans are better at judgment. When humans and AI collaborate, they can work toward eliminating the weaknesses of both (Agrawal et al., 2017). Using LLMs, as a tool, in conjunction with human experience can lead to more effective business correspondence by maximizing their strengths and minimizing their flaws. And this is precisely what education should focus on—i.e., the development of a new form of hybrid intelligence.

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

Open access

Copyrights

References

Abdullayeva, M., & Musayeva, Z. M. (2023, April). The Impact of Chat Gpt on Student's Writing Skills: An Exploration of Ai-Assisted Writing Tools. In *International Conference of Education, Research and Innovation* (Vol. 1, No. 4, pp. 61-66).

Adams, D., & Chuah, K. M. (2022). Artificial intelligence-based tools in research writing: current trends and future potentials. *Artificial Intelligence in Higher Education*, 169-184. https://doi.org/10.1201/9781003184157-9

Agrawal, A., Gans, J. S., & Goldfarb, A. (2017, February 7). What to expect from artificial intelligence: To understand how advances in artificial intelligence are likely to change the workplace—and the work of managers—you need to know where AI delivers the most value. *MIT Sloan Management Review*. https://doi.org/10.7551/mitpress/11645.003.0008

AlAfnan, M. A., & MohdZuki, S. F. (2023). Do artificial intelligence chatbots have a writing style? An investigation into the stylistic features of ChatGPT-4. *Journal of Artificial intelligence and technology*, *3*(3), 85-94. https://doi.org/10.37965/jait.2023.0267

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

Alshurafat, H. (2023). The usefulness and challenges of chatbots for accounting professionals: Application on ChatGPT. *Available at SSRN 4345921*. https://doi.org/10.2139/ssrn.4345921

Babic, B., Chen, D. L., Evgeniou, T., & Fayard, A. L. (2021). *A better way to onboard AI* (No. hal-03624592). Retrieved from https://hbr.org/2020/07/a-better-way-to-onboard-ai

Bishop, L. (2023). A computer wrote this paper: What chatgpt means for education, research, and writing. *Research, and Writing (January 26, 2023)*. https://doi.org/10.2139/ssrn.4338981

Biswas, S. (2023). ChatGPT and the future of medical writing. *Radiology*, *307*(2), e223312. https://doi.org/10.1148/radiol.223312

Calonge, D. S., Smail, L., & Kamalov, F. (2023). Enough of the chit-chat: A comparative analysis of four AI chatbots for calculus and statistics. *Journal of Applied Learning and Teaching*, *6*(2). https://doi.org/10.37074/jalt.2023.6.2.22

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*. https://doi.org/10.1177/23294906231176517

CNBC. (2023, January 31). Google testing ChatGPT-like chatbot 'Apprentice Bard' with employees. CNBC. Retrieved from https://www.cnbc.com/2023/01/31/google-testing-chatgpt-likechatbot-apprentice-bard-with-employees.html

Fitria, T. N. (2023, March). Artificial intelligence (AI) technology in OpenAI ChatGPT application: A review of ChatGPT in writing English essay. *ELT Forum: Journal of English Language Teaching*, *12*(1), 44-58. https://doi.org/10.15294/elt.v12i1.64069

George, A. S., & George, A. H. (2023). A review of ChatGPT AI's impact on several business sectors. *Partners Universal International Innovation Journal*, *1*(1), 9-23.

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

Gill, S. S., Xu, M., Patros, P., Wu, H., Kaur, R., Kaur, K., … Buyya, R. (2024). Transformative effects of ChatGPT on modern education: Emerging Era of AI Chatbots. *Internet of Things and Cyber-Physical Systems*, *4*, 19-23. https://doi.org/10.1016/j.iotcps.2023.06.002

Godwin-Jones, R. (2022). *Partnering with AI: Intelligent writing assistance and instructed language learning.*

Huang, J., & Tan, M. (2023). The role of ChatGPT in scientific communication: writing better scientific review articles. *American Journal of Cancer Research*, *13*(4), 1148.

Liu, Y., Mittal, A., Yang, D., & Bruckman, A. (2022). Will AI console me when I lose my pet? Understanding perceptions of AI-mediated

email writing. In *Proceedings of the 2022 CHI conference on human factors in computing systems* (pp. 1-13). https://doi.org/10.1145/3491102.3517731

Mahyoob, M., Al-Garaady, J., & Alblwi, A. (2023). A proposed framework for human-like language processing of ChatGPT in academic writing. *International Journal of Emerging Technologies in Learning (iJET)*, *18*(14). https://doi.org/10.3991/ijet.v18i14.41725

Mnasri, S., & Jovic, M. (2023). On the need to explicitize the unstated argument in cancer research: an ethnography on scientific argumentation. *Humanities and Social Sciences Communications*, *10*(1), 1-9. https://doi.org/10.1057/s41599-023-01823-7

Mnasri, S., & Papakonstantinidis, S. (2023). The teaching-learning dynamic from behaviourism to social constructionism: a communication-centred narrative. *International Journal of Learning and Change*, *15*(3), 312-327. https://doi.org/10.1504/IJLC.2023.130629

Nguyen, N., Johnson, J., & Tsiros, M. (2023). Unlimited Testing: Let's Test Your Emails with AI. *Marketing Science*. https://doi.org/10.1287/mksc.2021.0126

Okonkwo, C. W., & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. *Computers and Education: Artificial Intelligence*, *2*, 100033. https://doi.org/10.1016/j.caeai.2021.100033

Reuters. (2023). *Google unveils ChatGPT rival Bard, AI search plans in battle with Microsoft. Reuters.* Retrieved from https://www.tbsnews.net/tech/google-unveils-chatgpt-rival-bardai-search-plans-battle-microsoft-581138

Romm, C. T., & Pliskin, N. (1997). Toward a virtual politicking model. *Communications of the ACM*, *40*(11), 95-100. https://doi.org/10.1145/265684.265696

Rudolph, J., Tan, S., & Tan, S. (2023). War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. *Journal of Applied Learning and Teaching*, *6*(1). https://doi.org/10.37074/jalt.2023.6.1.23

Shane, J. (2019). *You look like a thing and I love you: How artificial intelligence works and why it's making the world a weirder place.* Little, Brown.

Siddike, M. A. K., Spohrer, J., Demirkan, H., & Kohda, Y. (2018). People's interactions with cognitive assistants for enhanced performances. In *Proceedings of the 51st Hawaii International Conference on System Sciences* (pp. 1640-1648). ScholarSpace at University of Hawaii at Manoa. https://doi.org/10.24251/HICSS.2018.205

Sok, S., & Heng, K. (2023). ChatGPT for education and research: A review of benefits and risks. *Available at SSRN 4378735*. https://doi.org/10.2139/ssrn.4378735

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M. A., Lacroix, T., ... Lample, G. (2023). *Llama: Open and efficient foundation language models.* https://doi.org/10.48550/arXiv.2302.13971

Valiyeva, A., & Thomas, B. J. (2022). Successful Organizational Business Communication and its Impact on Business Performance: An Intra-and Inter-Organizational Perspective. *Journal of Accounting, Business and Finance Research*, *15*(2), 83-91. https://doi.org/10.55217/102.v15i2.586

Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies*, 1-25. https://doi.org/10.1007/s10639-023-11742-4

Appendix A

Table 5. Rubric used for email assessment

Content 5 points	Excellent 1 point	Very good 0.8 point	Satisfactory 0.5 point	Needs improvement 0 point
Clarity of Purpose	Clearly states the purpose of the email.	For the most part, purpose is clear.	Somewhat clear purpose	Purpose is unclear
Relevance	Information provided is highly relevant to the topic.	Information provided is mostly relevant to the topic.	Somewhat relevant information	Information is irrelevant.
Support	Includes highly relevant details, evidence, or explanations	Includes relevant details, evidence, or explanations	Somewhat includes relevant details, evidence, or explanations	Lacks relevant details, evidence, or explanations
Call to Action	Clearly states the desired outcome or response	Clear call to action, for the most part	Somewhat clear call to action	Call to action is unclear or missing
Adaptation	Tailored to meet recipient's needs, expectations, and relationship	Tailored to meet recipient's needs and expectations for the most part	Somewhat tailored to meet recipient's needs and expectations	Not tailored to meet recipient's needs and expectations
Format 5 points	Excellent 1 point	Very good 0.8 point	Satisfactory 0.5 point	Needs improvement 0 point
Subject Line	Clear, relevant, and concise	Mostly clear, relevant, and concise	Somewhat clear, relevant, and concise	Unclear, irrelevant, or lengthy
Salutation	Personalized; follows business etiquette	Personalized and follows business etiquette for the most part	Somewhat follows etiquette and lacks personalization	Does not follow etiquette and lacks personalization
Structure	Clear and organized with a well-structured introduction, body, and closing	Mostly clear and organized with a somewhat structured format	Somewhat clear and organized but lacks structure	Unclear, disorganized, or lacks structure
Number of paragraphs	Correct number of paragraphs as per instructions	There is an extra paragraph.	There are two extra paragraphs.	Incorrect number of paragraphs; more than two
Number of words	Correct number of words as per instructions	Mostly correct number of words; within 10% deviation	Somewhat correct number of words: within 20% deviation	Incorrect number of words; exceeds limitation for more than 20%
Diction 5 points	Excellent 1 point	Very good 0.8 point	Satisfactory 0.5 point	Needs improvement 0 point
Grammar and Spelling	Free of major grammatical and spelling errors	Mostly free of major errors	Some major errors present	Numerous major errors present
Vocabulary	Appropriate for the context and audience	Mostly appropriate vocabulary	Somewhat appropriate vocabulary	Inappropriate vocabulary
Conciseness	Information is presented concisely without unnecessary repetition.	Mostly concise with minor repetition	Somewhat concise with noticeable repetition	Not concise, significant repetition
Sentence Structure	Varied sentence structures enhance readability	Mostly varied sentence structures	Somewhat varied sentence structures	Limited variety in sentence structures
Punctuation and Formatting	Properly used, enhancing clarity	Mostly proper use, with minor issues	Some issues with punctuation and formatting	Numerous issues with punctuation and formatting
Tone 5 points	Excellent 1 point	Very good 0.8 point	Satisfactory 0.5 point	Needs improvement 0 point
Formality	Matches the required level of formality	Mostly matches the required formality	Somewhat matches the required formality	Does not match the required formality
Politeness	Respectful and courteous tone throughout	Mostly respectful and courteous tone	Somewhat respectful and courteous tone	Lacks respect or courtesy
Emotion	Appropriately conveys emotions based on the email type	Mostly conveys appropriate emotions	Somewhat conveys appropriate emotions	Inappropriately conveys emotions
Audience Awareness	Demonstrates understanding of the recipient's perspective	Mostly demonstrates understanding	Somewhat demonstrates understanding	Lacks understanding of the recipient's perspective
Consistency	Tone remains consistent throughout the email.	Mostly consistent tone	Somewhat consistent tone	Inconsistent tone
Total Score: 20 points				