ExELL (Explorations in English Language and Linguistics)
11.2 (2023): 99-132
DOI: 10.2478/exell-2023-0008
Original scientific article

Clause complexing in research-article abstracts: Comparing human- and AI-generated texts

Alvin Ping Leong

Nanyang Technological University, Singapore

Abstract

The ability of chatbots to produce plausible, human-like responses raises questions about the extent of their similarity with original texts. Using a modified version of Halliday's clause-complexing framework, this study compared 50 abstracts of scientific research articles from Nature with generated versions produced by Bard, ChatGPT, and Poe Assistant. None of the chatbots matched the original abstracts in all categories. The only chatbot that came closest was ChatGPT, but differences in the use of finite adverbial clauses and –*ing* elaborating clauses were detected. Incorporating distinct grammatical features in the algorithms of AI-detection tools is crucially needed to enhance the reliability of their results. A genre-based approach to detecting AI-generated content is recommended.

Key words: abstracts; clause complexing; generative AI; Bard; ChatGPT; Poe Assistant.

1. Introduction

The launch of ChatGPT (chat generative pre-trained transformer) by OpenAI in November 2022 took the world by storm. An advanced chatbot, ChatGPT could generate human-like responses to questions and complete writing tasks in seconds. The generated texts were, at first glance, so convincing that barely a month after its launch, Gao et al. (2022) reported in bioRχiv that scientists themselves were able to detect ChatGPT-generated abstracts only about two thirds of the time, and about one in ten original abstracts were incorrectly identified as being ChatGPT-generated (see also Gao et al., 2023). The scientists, as noted in the report, found it "surprisingly difficult to differentiate between the two" (Gao et al., 2022: 1).

The ability of ChatGPT (and similar chatbots such as Bard and Poe Assistant) to churn out plausible-sounding texts and solutions has led to a marked growth in its applications in various domains, chiefly in the technical and education sectors (Elad, 2023). It is highly useful, for instance, in reviewing computer codes and processing data. In the field of education, the effectiveness of AI chatbots in scaffolding L2 writing and facilitating language learning has also shown promising results (Yan, 2023).

However, while chatbots can automate various tasks and serve as a useful aid for language learning, a key worry is that they are merely primed with the patterns observed in the training data, and so sometimes produce answers that lack details or are plainly incorrect. It is for this reason that Stack Overflow, a forum for programmers, banned the use of all generative AI technologies (Stack Overflow, 2023), stating that "because the average rate of getting correct answers from ChatGPT and other generative AI technologies is too low, the posting of answers created by ChatGPT and other generative AI technologies is substantially harmful to the site and to users." In scholarly publications, where the veracity of the information is similarly held to a high standard, this sentiment is also shared by some scholars. Rahimi and Abadi (2023), for instance, argue that papers with AI-generated tools as a 'co-author' should be either rejected or, if already published, retracted.

Detecting such generated texts, however, is not easy. Relying on human intervention, particularly in specialist fields, is one option (Teixeira da Silva, 2023a), but it is also entirely possible for chatbots to sometimes get it right and fool scientists along the way, as alluded to in the report by Gao et al. (2022, 2023). As Teixeira da Silva (2023a) wryly notes, "[h]uman detection is insufficiently sensitive." The other option is to use detection tools, but they appear unable keep up with the pace of the evolution of ChatGPT and similar chatbots. These tools also sometimes return erroneous results, flagging original texts as AI-generated ones (Fowler, 2023; Jimenez, 2023). Efforts have been made to improve the efficacy of detection tools. But while Cingillioglu (2023) reports perfect accuracy in detecting original essays, achieving the same for AI-generated texts does not appear to be possible in the near future. He (2023:266) concedes that "it is better to let multiple culpable persons escape than to make one innocent suffer."

Another option is to examine more closely the finer language details of AI-generated texts. Less work has been done on their grammatical features, and whether these match the norms in various academic genres. This is unsurprising at this stage since the use of chatbots in scholarly writing is a very recent concern, and comparative studies between original and generated texts may still be in the works. One notable exception is the work of Levin et al. (2023), who used Grammarly to compare original and generated abstracts in the fields of obstetrics and gynecology. Their study, which will be elaborated on in Section 2.2, focused on measures related to correctness and clarity issues, such as lengthy sentences, inappropriate word choice, and grammatical errors.

This present study builds on the work of Levin et al. (2023), expanding on it by examining science-based abstracts in general. It departs from it, however, by focusing on the grammatical clause, including embedded ones. This is because the grammatical features of science writing go beyond mere grammatical correctness and clarity. The focus of this study, then, is not on how correctly language is used or whether it is clear, but the extent to which the use of grammatical clauses, the building blocks of meaning, differ between the original and generated texts. The analysis is based on a modified version of the Hallidayan framework on clause complexing (Halliday & Matthiessen, 2014). It is hoped that the patterns observed in the use of clauses can serve as an additional consideration when detecting generated texts.

The rest of this paper is organized as follows. Section 2 presents a broad background of the development of chatbots, and how they have come to affect scholarly publications. The Hallidayan framework is also introduced and reviewed. Section 3 offers a description of the corpus and the analytical approach. The findings and discussion are presented in Section 4. The concluding section summarizes the broad findings and discusses their implications.

2. *Background details and relevant research*

2.1 *Advent of chatbots*

ChatGPT is arguably the most well-known chatbot. Its launch in November 2022 allowed the public to try out its capabilities without charge.[1] This naturally fuelled its popularity, making it the fastest-growing application in history (Hu, 2023). According to Cyberhaven, by June 2023, about one in ten employees worldwide has used ChatGPT at work, some of whom (4.7%) have even included confidential data while using it (Coles, 2023).

The basic architecture for ChatGPT, known as the transformer, was first introduced by Google researchers in 2017 (Vaswani et al., 2017). As it requires less time for training as compared to earlier architectures (e.g., recurrent neural networks), it has since come to be used widely in natural language processing and the training of large language models. Given a large enough dataset and adequate training, transformer-based chatbots can generate a variety of human-like responses to prompts, ranging from the analysis of data to even the dispensing of personal advice.

[1] At the time of writing, GPT-3.5 was free to use; the most-advanced version, GPT-4, required a paid subscription.

Other AI chatbots quickly surfaced after the launch of ChatGPT. Quora developed an aggregator Poe, which generates responses from other chatbots, including ChatGPT. It was soft launched in December 2022, and made fully available to the public in February 2023 (Perez, 2023). In March 2023, Google launched Bard in response to the perceived threat posed by ChatGPT to its own search engine (Heaven, 2023). Microsoft itself overhauled its Bing search engine in February 2023 to include a chatbot feature based on GPT-4 (Mehdi, 2023).

The use of such chatbots can be helpful in certain fields such as psychology, where ChatGPT has been shown to outperform humans in evaluating emotional awareness (Elyoseph et al., 2023). In education, chatbots have been incorporated into teaching pedagogy as far back as 2016. In Okonkwo and Ade-Ibijola's (2021) review of 53 studies, the general consensus was that educators affirmed the value of utilizing chatbots in educational settings as they allowed students to make learning gains comparable to those achieved with human tutors (Graesser, 2016). Kohnke (2023) further reports that students enjoyed interacting with chatbots and perceived an improvement in their language skills.

However, outside of teaching and learning, particularly in areas where the contributions of the individual are warranted, reception to these chatbots has been less favorable. Concerns have been raised about the ethical implications of the use of chatbots among students in tests and assignments:

> […] one of our primary concerns and fears about ChatGPT is its potential for cheating and academic dishonesty. Since ChatGPT can generate natural language responses that are indistinguishable from those generated by humans, there is a risk that students could use ChatGPT to cheat on assessments. (Naidu & Sevnarayan, 2023: 6)

Similar concerns have also been raised about scholarly publications, but the views here are somewhat mixed. Some scholars argue that a chatbot cannot be regarded as a 'co-author' of a manuscript simply because it cannot be held accountable (Teixeira da Silva 2023a, 2023b). Others, though, hold a different view. Since the generated content need not always be initiated by the author (for it can be generated and edited by someone else), the ownership of the generated content becomes contentious. For this reason, Lund et al. (2023: 575) note that "it may be necessary, at minimum, to include the model as a coauthor of the manuscript."

Determining whether academic integrity has been compromised is a tricky issue since existing detection tools are not always reliable (Cingillioglu, 2023; Sharples, 2022). This warrants alternative ways of looking at generated texts, apart from the oft-cited inaccurate content or false citations/references (Houston & Corrado, 2023). The next section reviews work involving ChatGPT and scholarly publications, narrowing toward the comparative study by Levin et al. (2023) on human- and ChatGPT-generated abstracts.

2.2. *Chatbots and scholarly writing*

Letters to editors and small-scale studies have highlighted how chatbots can assist in the crafting of research grants (Najafali et al., 2023) and generate abstracts (Babl & Babl, 2023), among others. Najafali et al. (2023) used a series of prompts to guide ChatGPT to produce various segments of a research grant, such as the aims, hypothesis, and significance of the proposed project. While the overall quality was not perfect—ChatGPT was unable to write project aims that were acceptable for grant consideration—it was nevertheless viewed as a useful complementary tool to aid researchers in the writing process. In the case of abstracts, Babl and Babl (2023) recruited a person not trained in medicine to instruct ChatGPT to produce a conference abstract regarding a disease with fictitious data. The resulting abstract was of such a high quality that the authors deemed it "impressive." They further note:

> […] ChatGPT allows a non-scientist to produce a satisfactory conference abstract. We doubt even the current, uncorrected version would be rejected by an abstract assessment committee. (Babl & Babl, 2023: 1)

To some, however, the above comment may be of concern since the use of such tools jeopardizes "the integrity, originality, and validity of the academic peer-review, publishing, and the collective scientific literature" (Rahimi & Abadi, 2023: 9). Detecting the presence of AI-generated content crucially requires a comparison between the patterns observed in human- and AI-generated texts. An interesting study along this line was conducted by Levin et al. (2023). They used 50 abstracts that were randomly selected from 1,378 original articles from the American Journal of Obstetrics & Gynecology; the articles were published from 2018 to 2022. ChatGPT generated an abstract for each of the 50 original versions with a standardized prompt. The Grammarly software was used to assess all the abstracts.

Levin et al. (2023) reported a higher median Grammarly score for the original abstracts. The original abstracts also had more identified writing issues and used fewer unique words. The AI-generated abstracts were generally free of common grammatical errors (such as incorrect verb forms) but exhibited dialectal inconsistency (i.e., the mixed use of different dialects of English). The results are interesting, and, while statistical significance was reported for the observed differences, the median scores in some cases were quite close. For instance, the overall Grammarly scores for the original and ChatGPT-generated abstracts were 86 and 83, respectively (p=.003). A score of 86 for the original abstracts means that the writing was found to be more accurate than the writing in 86% of similar abstracts; the score of 83 for

ChatGPT-abstracts was not very far behind. Further, since both abstracts had median scores in excess of 80, the general quality of the abstracts could be taken to fall within the same band. Similar comments extend to the reported median scores for writing issues (original: 18 to AI-generated: 16, p=.004) and the use of unique words (original: 40 to AI-generated: 42, p=.031). The lower use of unique words can also be easily accounted for in science writing, given its heavy reliance on nomenclature (Biber & Gray, 2016; Halliday, 1998; Halliday & Martin, 1993).

Notwithstanding these limitations, the study by Levin et al. (2023) is valuable as a first step in examining specific linguistic features of human- and AI-generated scholarly texts. In this paper, we expand on their work by considering another aspect—the use of clauses—in science-based abstracts. Clauses tend to be overlooked in discussions on generated texts; yet, clauses are carriers of meaning (Halliday & Matthiessen, 2014), and patterns of clausal use can be highly insightful in characterizing different genres. To understand more fully the use and analysis of clauses and inter-clausal relations, we turn next to the Hallidayan framework on clause complexing.

2.3 *Hallidayan framework on clause complexing*

Clause complexing captures the relations between clauses at the same rank, according to Halliday's lexico-grammatical rank scale (Halliday & Matthiessen, 2014) as shown in Figure 1.

Clause

|

Word phrase

|

Word

Figure 1: Halliday's lexico-grammatical rank scale (adapted)

As seen in Figure 1, the clause rank is the highest rank on the scale; the Hallidayan framework uses the term RANKING CLAUSE to refer to any clause that operates at this rank. This is to separate it from other clauses that exist at a lower rank (i.e., at the rank of the word phrase). These lower-rank clauses are termed EMBEDDED (or rankshifted) clauses; examples of such embedded clauses include noun clauses and complement clauses, among others.

At this point, a note of clarification is needed as regards restrictive (defining) and non-restrictive (non-defining) relative clauses. On account of their different functions, these clauses are analyzed differently in the Hallidayan framework. Restrictive relative clauses, as seen in (1), perform an identifying function, and are considered to be an integral part of the noun phrase, specifying and so distinguishing the noun that it modifies from other nouns. In (1), for instance, the subset of B cells is not just any subset, but that which expands in the draining lymph node. The Hallidayan framework regards restrictive relative clauses as embedded clauses; by convention, embedded clauses are marked off using double square brackets [[…]]. All examples are taken from the corpus; a full list of the symbols used in this paper, including those for tactic and logico-semantic relations (Sections 2.3.1–2.3.2), is given in the Appendix.

(1) […] *we identified a subset of B cells [[that expands specifically in the draining lymph node over time in tumour-bearing mice*.]] *(*Nature *14)*

By contrast, non-restrictive relative clauses do not perform an identifying function, but a describing one. The noun is already taken to be specific, and the non-restrictive relative clause merely provides further information in relation to it. For this reason, punctuation marks (such as commas, dashes, or brackets) are often inserted to separate such non-restrictive relative clauses from the head noun, signaling their parenthetical nature. In the Hallidayan framework, non-restrictive relative clauses, as underlined in (2), are analyzed as ranking clauses.

(2) *Even canonical seed pairing is dispensable for PIWI binding or cleavage, unlike plant and animal AGOs, <u>which require uninterrupted target pairing from the seed to the nucleotides past the scissile bond</u>*. *(*Nature *20)*

The Hallidayan clause-complexing framework provides a systematic description of the interdependencies and logico-semantic relations between clauses within the sentence, defined as a CLAUSE COMPLEX (Halliday & Matthiessen 2014: 428–556). A clause complex is thus the grammatical equivalent of the graphological sentence. The limiting case of a single-clause sentence is termed a SIMPLEX.

Inter-clausal relations apply to both ranking and embedded clauses, so long as they operate at the same rank. These relations are analyzed along two dimensions. The first, TAXIS, relates to interdependency; clauses are either paratactically or hypotactically related to each other. The second, LOGICO-SEMANTICS, concerns how one clause expands another in semantic terms. More specific details about each dimension are given in the following sub-sections, beginning with the tactic relation.

2.3.1 *Tactic relations*

Clauses in a tactic relation are of equal or unequal status. The term PARATAXIS refers to the relation between two clauses of like status. HYPOTAXIS, on the other hand, refers to the unequal relation between a primary clause and a dependent clause.

Paratactic clauses are indicated by Arabic numerals, with '1' representing the first clause in such a relation, '2' representing the second clause, and so forth. Parataxis involves not merely main clauses but subordinate ones as well, so long as the inter-clausal relation is of equal status. These scenarios are exemplified in (3–4) below, where clause complexes are marked off using triple vertical lines (|||…|||), and clauses by double vertical lines (||…||).

(3)	1	||| *Crystalline materials enable essential technologies,* ||
	2	*and their properties are determined by their structures.* ||| (Nature 28)
(4)		||| […] *plastic pollution is an emerging threat to coral reefs,* ||
	1	*spreading throughout reef food webs,* ||
	2	*and increasing disease transmission and structural damage to reef organisms.* ||| (Nature 09)

Hypotaxis, by contrast, is the relation between clauses of unequal status. The relation is thus one of dependency between a primary clause and another that is dependent on it. Unlike parataxis, hypotactic clauses are indicated by Greek letters, with the primary clause always represented by 'α', and the dependent clauses by 'β', 'γ', and so on, as shown in (5) below:

(5)	α	||| *This event collapsed a palaeo-summit,* ||
	β	*probably culminating above 8,000 m in altitude.* ||| (Nature 32)

Clause complexes need not contain only paratactic or hypotactic clauses. It is not uncommon for both tactic relations to co-occur in a single clause complex. In (6), for example, the interferon response led to two outcomes—augmented B cell activation, and increased antigen presentation/co-stimulation. It is these outcomes that resulted in an expansion of effector T cells. The dependent paratactic clauses describing the two outcomes thus comprise a single group, collectively serving as the dominant segment upon which the clause describing the expansion of effector T cells is dependent.

(6)	α			||| *Loss of TIM-1 enhanced the type 1 interferon response in B cells,* ||
	β	α	1	*which augmented B cell activation* ||

2 *and increased antigen presentation and co-stimulation,* ||

β *resulting in increased expansion of tumour-specific effector T cells.* ||| (Nature 14)

In this light, the Hallidayan framework offers a more intuitive treatment of clause combinations, showing systematically how some clauses are more closely related to one another than to others, and the status of that relation. Labeling clauses as solely 'main' ('independent') or 'subordinate' ('dependent'), as in traditional grammar, runs the risk of obscuring such relations.

2.3.2 *Logico-semantic relations*

Logico-semantics captures the second way by which clauses are related to each other. As the term suggests, this relation is both logical and semantic, in the sense that clauses in a clause complex are related not only by way of interdependency, but meaning. As Halliday (2006: 355) notes, "[t]he conjunction system sets up logical-semantic relations between one piece of text and another: relations of equivalence, addition, alternation, adversity, comparison and contrast, cause, time, condition and concession." The Hallidayan framework recognizes two broad types of logico-semantic relations—PROJECTION and EXPANSION. However, in this study, only expansion was employed; a modified Hallidayan framework was thus used, rather than its full version.

The reason why projection was excluded was that its anomalies could not be easily explained away. In the Hallidayan framework, projection is manifested as a verbal process (as locution) or a mental process (as idea). These are exemplified in (7–8), with the projected clauses underlined. As the present corpus did not contain any instances of idea, the example in (8) is taken from the full text (rather than the abstract) of the specified research article.

(7) *We propose <u>that our assessment provides a quantitative foundation for safeguarding the global commons for all people now and into the future</u>.* (Nature 33)

(8) *We believe <u>that this work opens the door to translating the progress in modern natural language processing and deep learning to improving the quality and affordability of healthcare</u>*, […] (Nature 15, from full text)

Projection runs against the traditional account of English grammar, which Halliday and Matthiessen (2014: 303) themselves acknowledge. Labeling both projecting and projected clauses as ranking clauses raises three problems. First, the Hallidayan framework argues that constructions such as (7–8) have agnate structures, i.e., they can be re-expressed as free indirect speech (9). This appears to work well, but as free indirect speech is rare in science writing—no examples were detected in the present corpus—it could be counter-argued that its marked status implies the reverse, that (7) is the agnate form of (9) instead.

(9) *Our assessment provides, we propose, a quantitative foundation for safeguarding the global commons for all people now and into the future.*

Second, if projected clauses were ranking clauses, they could not be preceded by noun phrases (such as 'the fact'), since that would make them a postmodifier and, hence, an embedded clause. Such a diagnostic may work for some verbs (e.g., 'say'), but not others. Expressions such as "state/proclaim/assert the fact," for instance, are not uncommon, and yet these are nevertheless regarded by the Hallidayan framework as examples of locution.

Third, in a hypotactic setting, the projecting clause should rightfully be the *main* clause, i.e., it should be able to "stand on its own" (Halliday & Matthiessen, 2014: 440). This, however, is clearly not possible with the projecting clauses in (7–8). As Fawcett (2000: 29) points out, the projecting clause is actually "an uncompleted clause that is 'expecting' […] another element (which we may call a Complement)." For these key reasons, the logico-semantic analysis of clause complexes in this paper involved only expansion, but not projection.

We move next to expansion. A clause may be expanded in three ways: ELABORATION, EXTENSION, and ENHANCEMENT. In elaboration, the second clause restates, specifies, or exemplifies the information in the first clause. Halliday and Matthiessen (2014: 461) are careful to note that the elaborating clause "does not introduce a new element into the picture but rather provides a further characterization of one that is already there." Paratactic elaboration is usually signaled via punctuation marks such as the colon, semicolon, and dash; a typical example of hypotactic elaboration is the non-restrictive relative clause. By convention, elaboration is represented by the equality sign (=), as shown in (10–11).

(10) 1 ||| *We also find two main SFH types present in all the environments:* ||

=2 *'short-timescale' galaxies are not affected by their large-scale environment at early times but only later in their lives;* || […] (Nature 01)

(11) α ||| *A typical frustrated system is correlated bosons on moat bands,* ||

=β *which could host topological orders with long-range quantum entanglement.* ||| (Nature 26)

Unlike elaboration, extension involves new information, an exception, or an alternative. Extending clauses are typically introduced by conjunctions such as 'and', 'or', and 'but' (12); other common markers include 'whereas' and 'while'. Extending clauses are represented by the addition sign (+).

(12) 1 […] || *we conducted a genome-wide association study of the age-related MS severity score in 12,584 cases* ||

+2 *and replicated our findings in a further 9,805 cases.* ||| *(Nature 11)*

Lastly, in enhancement, the modifying clause provides circumstantial information and so qualifies how the primary clause is to be interpreted. Enhancements include information related to time, place, manner, cause, or condition (Halliday & Matthiessen, 2014: 476). Enhancing clauses, as shown in (13), are represented by the multiplication sign (×).

(13) α ||| *Consequently, even a 2-h delay in a cell's progression towards mitosis can induce cell cycle exit* ||

×β *if mitogen signalling is lost.* ||| *(Nature 16)*

As mentioned earlier, the clause-complexing framework, less projection, can also be extended to embedded clauses. In (14), the embedded segment, represented by 'A', serves as the grammatical subject. It comprises two embedded clauses, with the second functioning as a purpose clause. This renders the inter-clausal relation between the embedded clauses as one of hypotactic enhancement. Notationally, inter-clausal relations involving embedded clauses carry the superscript "E" to separate them from inter-clausal relations involving ranking clauses.

(14) ||| *Moreover,* [[…A…]] *is also unclear.* |||

A α^{E} [[*whether the break-induced replisome orchestrates additional DNA repair events* ||

$\times\beta^{E}$ *to ensure processivity*]]

2.4 *Related studies on clause complexing*

Over the years, empirical studies on clause complexing have involved a myriad of discourses. Early work investigated both speech and writing. In their analysis of spoken narrative texts in Australian English, Nesbitt and Plum (1988) found that the logical-semantic relation of extension, but not enhancement or elaboration, was strongly associated with genre. Greenbaum and Nelson (1995) compared both spoken and written texts from a diverse range of genres (e.g., conversations, letters, academic writing), but found no clear distinctions in their overall relative complexity. The application of clause complexing has also been extended to specific genres, such as aphasic discourse (Armstrong 1992) and student essays (Leong & Wee, 2005).

Where scholarly writing is concerned, however, research interest has been far more modest. The work of Sellami-Baklouti (2011) is a notable exception. Her corpus comprised 120 abstracts from research articles in the sciences and social sciences. She found that science writing was characterized by simplexes, and social-science writing by hypotactic relations. The finding involving science writing was affirmed in the work of Leong (2021b), who analyzed full articles (rather than abstracts) in the sciences and humanities. His corpus comprised 40 articles from respected journals such as Science and Journal of Ethics, among others. Like Sellami-Baklouti, Leong found that science writing favored simplexes; humanities writing, by contrast, relied more on embedded clauses.

In this age of AI, the need for more comparative work on specific grammatical features in human- and AI-generated texts is thus needed. This present work sought to address such a research gap by considering the use of clauses in science abstracts. It sought answers to the following research question: how do the occurrence rates of clauses, tactic relations, and logico-semantic relations in science abstracts compare with the same in AI-generated abstracts? Given that the public launch of ChatGPT and similar AI chatbots took place very recently, and that no prior empirical studies of a similar nature were available, no hypothesis as regards this research question could be drawn.

3. Methodology

3.1 *Corpus*

The corpus comprised 200 abstracts, 50 of which were published in Nature, a top journal covering all disciplines and sub-disciplines in the sciences. As reported in Scimago (2022), Nature was ranked first in the 'multidisciplinary' category. All the Nature articles were published in 2023, and were the most recent publications at the time of analysis (July 2023).

The other 150 abstracts were AI-generated versions of the original Nature abstracts. The selected chatbots were Google's Bard, OpenAI's ChatGPT, and Quora's Poe Assistant (hereafter Poe), with each chatbot generating a version of a corresponding Nature abstract. In the case of ChatGPT, the latest version, GPT-4, was used. These chatbots were selected because of their strong media presence. This was particularly so for ChatGPT, which attracted over 100 million users just two months after its public launch (Carr, 2023). Microsoft's Bing was originally considered, but as it was unable to produce abstracts in two instances, a fair comparison with the other chatbots was not possible, and it was dropped from the list. For consistency, the following prompt was used with all the chatbots to generate the abstracts:

> Write an abstract for a scientific research paper on [*title of original research article*]. The research paper is to be published in a scientific journal. The abstract should be about [*word length of original abstract*] words.

As an example, the title of Nature 01 was "Galaxies in voids assemble their stars slowly." and the original abstract was 177 words long. The following prompt was thus used to generate the AI versions of this abstract:

> Write an abstract for a scientific research paper on galaxies in voids assembling their stars slowly. The research paper is to be published in a scientific journal. The abstract should be about 177 words.

3.2 *Analysis*

Each abstract was broken up into clauses. All clauses were classified as either ranking or embedded clauses; ranking clauses were further divided into main and subordinate clauses. Inter-clausal tactic and logico-semantic relations for both ranking and embedded clauses, as outlined in Section 2.3, were recorded. Microsoft Excel was used to tabulate the statistics for each analysis; a screenshot of the analysis of the first nine clauses of Nature 01 is provided in Figure 2.

	Nature 01	Main clause	Subordinate clause	Embedded clause	Simplex	Parataxis (ranking)	+	x	=	Hypotaxis (ranking)	+	x	=	Parataxis (embedded)	+	x	=	Hypotaxis (embedded)	+	x	=
1	||| Galaxies in the Universe are distributed in a web-like structure	1			1																
2	[[characterized by different large-scale environments: dense clusters, elongated filaments, sheetlike walls and under-dense regions, ||			1																	
3	called voids.]] |||			1																	1
4	||| The low density in voids is expected to affect the properties of their galaxies. |||	1			1																
5	||| Indeed, previous studies6–14 have shown	1			1																
6	[[that galaxies in voids are, on average, bluer and less massive, ||			1																	
7	and have later morphologies and higher current star formation rates than galaxies in denser large-scale environments.]] |||			1											1						
8	||| However, it has never been observationally proved	1			1																
9	[[that the star formation histories (SFHs) in voids are substantially different from those in filaments, walls and clusters.]] |||			1																	

Figure 2: Screenshot of analysis of Nature 01

3.3 *Statistical analysis*

The frequency counts of the various categories of clauses and inter-clausal relations were computed as normed rates per 100 words; proportions were also computed where needed. Statistical tests were conducted using Real Statistics Resource Pack (Zaiontz, 2022), a Microsoft Excel add-in. The Welch one-way analysis of variance (ANOVA) test was used to determine if the observed differences in rates were statistically significant. The Games-Howell post-hoc test was applied for all significant ANOVA results. The significance level for all statistical tests was α=.05. In this paper, statistically significant differences are indicated with a single asterisk for p<.05, and double asterisks for p<.01.

4. Results and discussion

4.1 *General characteristics*

The corpus comprised a total of 38,267 words and 3,947 clauses. Table 1 lists the summary statistics of the number of words and clauses used in each abstract group. As main clauses also include simplexes, statistics related to the latter category are listed in the final row of the table.

For ranking clauses, the ratio of main to subordinate clauses was about 3:1 for Nature abstracts, and 2:1 for ChatGPT abstracts. The ratio for both Bard and Poe abstracts was roughly 4:1. Embedded clauses constituted about a third of all clauses, with the exception of Poe abstracts, in which more than four in 10 clauses were embedded. We will discuss more fully the occurrence rates of clauses in Section 4.2.

A cursory glance of the abstracts generated by Bard and Poe revealed the use of fixed or formulaic wording in many of them. Close to 40% of Bard abstracts used the expression "our/these findings suggest that" (19 occurrences), and more than a quarter of the abstracts (13 occurrences) contained the structure "we used [an approach] to investigate." The situation involving Poe abstracts was even more pronounced—the opening line in 46 of the 50 abstracts began with the words "We present"; the words "We report" were used in the remaining four abstracts.

Table 1: Total number of words and clauses in the corpus

	Nature			Bard			ChatGPT			Poe			Total
Words	10,255			7,981			9,168			10,863			38,267
	Freq.	**%RC[1]**	**%TC[2]**	**Freq.**	**%RC[1]**	**%TC[2]**	**Freq.**	**%RC[1]**	**%TC[2]**	**Freq.**	**%RC[1]**	**%TC[2]**	
Ranking clauses													
Main	467	73.72	48.38	473	81.08	48.70	444	69.30	44.89	405	80.50	43.91	1,789
Subordinate	179	26.28	18.11	118	18.92	12.02	212	30.70	20.91	103	19.50	11.23	612
Total ranking	646			591			656			508			2,401
Embedded clauses	351		33.51	402		39.28	356		34.20	437		44.87	1,546
Total clauses	997			993			1012			945			3,947
		%MC[3]			**%MC[3]**			**%MC[3]**			**%MC[3]**		
Simplexes	209	44.75		299	63.21		193	43.47		210	51.85		911

[1] %RC=percentage of ranking clauses

[2] %TC=percentage of clauses

[3] %MC=percentage of main clauses

A few abstracts also resembled each other to a large extent. Two Poe abstracts on plastic pollution, for instance, followed a template with the structure: "We present" → "Our results show that" → "We estimate that" → "We (also) find that" → "Our work underscores." Even the cited statistics, underlined in (15a–b) below, look similar:

(15) a. *We estimate that between <u>4 to 12 million metric tons</u> of plastic waste enters the oceans each year, and a significant proportion of this waste ends up on coral reefs, where it can have a range of negative impacts. (Poe 09)*

b. *We estimate that between <u>4 to 23 million metric tons</u> of plastic waste enters lakes and reservoirs each year, and a significant proportion of this waste accumulates in these ecosystems, where it can persist for decades or longer. (Poe 10)*

More markedly, Table 2 lists two near-identical Bard abstracts concerning pair density waves; the titles of the original Nature abstracts are included. As can be seen, differences between Bard 47 and Bard 49 are minor and relate chiefly to the use of technical terms.

While such similarities could be attributed to the word-predictive nature of AI text generation, it does not quite explain why this uniformity was not apparent in ChatGPT abstracts. Like Nature abstracts, ChatGPT abstracts exhibited variation in wording, and even when similar concepts were described, the abstracts were dissimilar.

To understand more fully how the AI-generated abstracts in the corpus compared with the original versions, we turn next to a more detailed look at the use of clauses and inter-clausal relations.

Table 2: Similarity between Bard-generated texts on pair density waves

Bard 47	**Bard 49**
Original title: "Detection of a pair density wave state in UTe_2"	*Original title:* "Pair density wave state in a monolayer high-T_c iron-based superconductor"
Pair density wave (PDW) states are a type of unconventional superconductivity that is characterized by a periodic modulation of the Cooper pairs. PDW states have been proposed to exist in a number of materials, but they have never been definitively observed. We used scanning tunneling microsco-	Pair density wave (PDW) states are a type of unconventional superconductivity that is characterized by a periodic modulation of the Cooper pairs. PDW states have been proposed to exist in a number of materials, but they have never been definitively observed in a monolayer high-T_c iron-based superconductor. We

py (STM) to investigate the electronic structure of UTe_2. We found that the STM images of UTe_2 exhibit a periodic modulation of the local density of states, which is consistent with a PDW state. The PDW state in UTe_2 is characterized by a wave vector of 0.33 r.l.u., which is incommensurate with the underlying crystal lattice. The PDW state is also characterized by a gap that opens at the Fermi level, which is indicative of superconductivity. Our findings provide the first definitive evidence for the existence of a PDW state in a material. These findings could have implications for the understanding of unconventional superconductivity and the search for new superconducting materials.	used scanning tunneling microscopy (STM) to investigate the electronic structure of a monolayer Fe(Te,Se) superconductor. We found that the STM images of the monolayer Fe(Te,Se) exhibit a periodic modulation of the local density of states, which is consistent with a PDW state. The PDW state in the monolayer Fe(Te,Se) is characterized by a wave vector of 0.33 r.l.u., which is incommensurate with the underlying crystal lattice. The PDW state is also characterized by a gap that opens at the Fermi level, which is indicative of superconductivity. Our findings provide the first definitive evidence for the existence of a PDW state in a monolayer high-T_c iron-based superconductor. These findings could have implications for the understanding of unconventional superconductivity and the search for new superconducting materials.

4.2 *Main, subordinate, and embedded clauses*

We begin with the use of main, subordinate, and embedded clauses. Table 3 presents the normed rates (per 100 words) of these clauses. Only significant F and p values are listed.

Given that AI text generation is based on linguistic patterns and next-word predictions, and that the prompts in this study specifically indicated that the texts were for a scientific journal, the varied differences among the chatbot abstracts were unexpected. Poe abstracts used fewer main and subordinate clauses than the rest. By contrast, Bard abstracts used more main and embedded clauses. ChatGPT abstracts differed from Nature abstracts in their greater use of subordinate clauses.

In this respect, the reported rates for Nature abstracts can serve as a useful benchmark when comparing texts. Figure 3, based on the statistical results in Table 3, shows how Nature abstracts (highlighted in black) are positioned relative to the other abstract groups. Abstract groups with statistically insignificant differences are enclosed in dotted boxes.

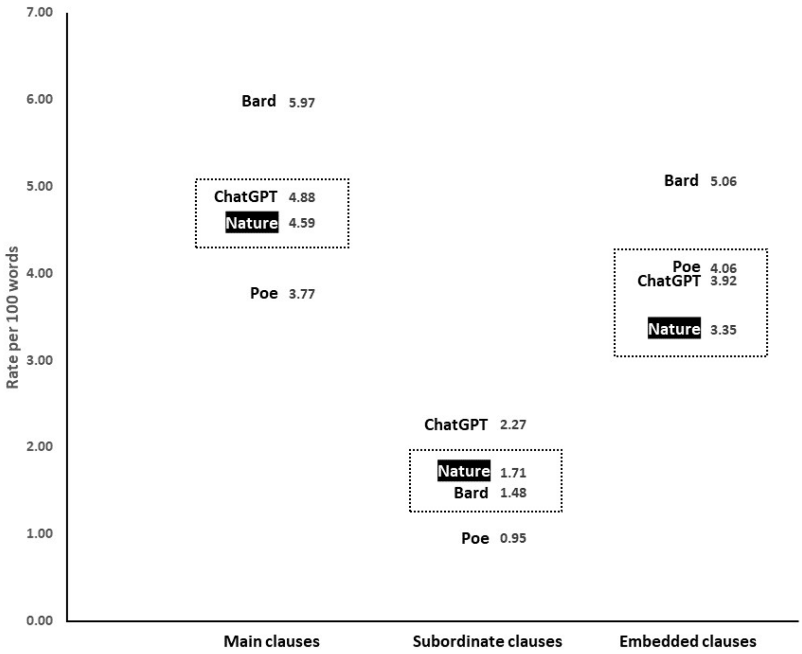

Figure 3. A diagram of a tree. Description automatically generated with medium confidence

As can be seen, the rate of main clauses in Nature abstracts (4.59) occupies a 'middle ground' when compared with Bard (5.97) and Poe (3.77) abstracts. We see this same middling result for subordinate clauses as well when comparing Nature with ChatGPT and Poe abstracts—the rate of subordinate clauses in Nature abstracts (1.71) was almost twice that of the Poe rate (0.95), but only three-quarters that of the ChatGPT rate (2.27). Where embedded clauses are concerned, only the observed difference between Nature and Bard abstracts was found to be statistically significant. This can be a helpful guide for diagnostic purposes. Abstracts that swing to one or the other extreme in the use of clauses may indicate that they were (partly) AI generated.

Table 3 and Figure 3, though, do also reveal that Nature abstracts were not statistically different from a few abstract groups in certain clause categories. For instance, differences in the use of subordinate clauses were not found to be statistically different between Nature and Bard abstracts, nor were the use of embedded clauses among Nature, ChatGPT, and Poe abstracts. We therefore need to explore further whether inter-clausal relations could provide a finer diagnostic layer to separate original from AI-generated abstracts.

Table 3: Occurrence rates of main, subordinate, and embedded clauses

Clauses	Rate	SD	Games-Howell comparisons (*p*) Nature	Bard	ChatGPT	Poe
Main						
Nature	4.59	1.73		1.10e-12**		6.22e-08**
Bard	5.97	1.82	1.10e-12**		8.18e-10**	9.30e-14**
ChatGPT	4.88	1.27		8.18e-10**		4.06e-13**
Poe	3.77	0.76	6.22e-08**	9.30e-14**	4.06e-13**	
Subordinate						
Nature	1.71	2.16			.03*	3.69e-05**
Bard	1.48	1.60			7.55e-04**	6.98e-03**
ChatGPT	2.27	2.09	.03*	7.55e-04**		1.28e-11**
Poe	0.95	1.06	3.69e-05**	6.98e-03**	1.28e-11**	
Embedded						
Nature	3.35	3.65		5.73e-05**		
Bard	5.06	3.47	5.73e-05**		.02*	.03*
ChatGPT	3.92	3.50		.02*		
Poe	4.06	3.20		.03*		

ANOVA (main clauses): $F(3, 107.06)=90.04$, $p<0.001$**

ANOVA (subordinate clauses): $F(3, 103.56)=26.71$, $p=7.06e^{-13}$**

ANOVA (embedded clauses): $F(3, 108.26)=7.26$, $p=1.75e^{-4}$**

4.3 *Tactic relations*

Of the two tactic relations, only hypotaxis returned significant ANOVA results. Parataxis across the abstract groups for both ranking and embedded clauses was found to be not merely statistically insignificant, but uncommon. For instance, in the case of Nature, no paratactic ranking clauses were found in nine abstracts, and out of 49 abstracts containing embedded clauses, parataxis was absent in 23 of them. Table 4 lists the ANOVA results for tactic relations, including clauses that do not carry such relations (i.e., simplexes).

Table 4: Occurrence rates of paratactic and hypotactic clauses

Taxis	Rate	SD	Games-Howell comparisons (p) Nature	Bard	ChatGPT	Poe
Simplexes						
Nature	2.07	1.87		4.35e-13**		
Bard	3.79	1.52	4.35e-13**		4.52e-13**	1.01e-13**
ChatGPT	2.14	1.39		4.52e-13**		
Poe	1.96	1.07		1.01e-13**		
Parataxis (ranking)						
Nature	0.74	1.07				
Bard	0.60	1.18				
ChatGPT	0.55	0.98				
Poe	0.51	0.42				
Parataxis (embedding)						
Nature	0.36	0.92				
Bard	0.35	0.68				
ChatGPT	0.30	0.81				
Poe	0.53	1.16				
Hypotaxis (ranking)						
Nature	1.64	2.13			.04*	7.57e-05**
Bard	1.37	1.52			5.08e-04**	.02*
ChatGPT	2.15	2.07	.04*	5.08e-04**		9.53e-11**
Poe	0.93	0.97	7.57e-05**	.02*	9.53e-11**	
Hypotaxis (embedded)						
Nature	0.49	1.35				8.81e-03**
Bard	0.74	1.25				
ChatGPT	0.66	1.29				
Poe	0.88	1.25	8.81e-03**			

ANOVA (simplexes): $F(3, 106.51)=53.10$, $p<0.001$**

ANOVA (hypotaxis ranking clauses): $F(3, 102.57)=24.59$, $p<4.56e^{-12}$**

ANOVA (hypotaxis embedded clauses): $F(3, 108.63)=3.56$, $p=.02$*

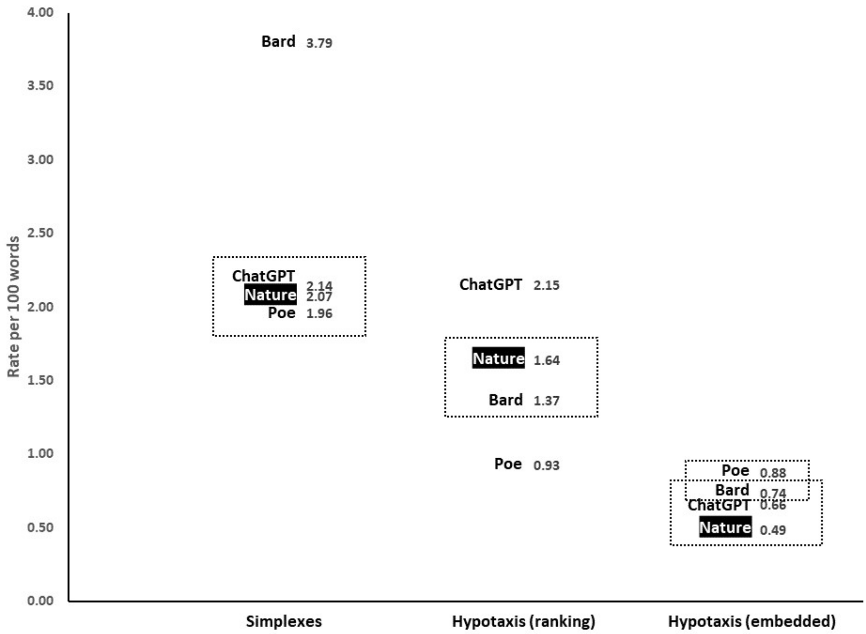

Figure 4: Simplexes and hypotactic relations—Nature abstracts as compared to the other abstract groups

Earlier studies have revealed the general preference for simplexes in science writing. In his comparative study of science and humanities research articles, Leong (2021b) found that simplexes constituted 32.21% of the main clauses in the science articles. This preference for simplexes in science writing reflects the objective, thing-oriented nature of scientific research, which downplays the need for additional qualifying or interpretive statements that are more commonly found in non-science writing (Sellami-Baklouti, 2011).

This preference for simplexes is also evident in the present corpus. As shown earlier in Table 1, however, the proportion of simplexes to main clauses for Nature abstracts was higher (44.75%) than that reported in Leong (2021b). This higher statistic is unsurprising since abstracts, being much shorter than full research articles, naturally comprise fewer main clauses. The proportions across ChatGPT, Nature, and Poe abstracts were within a narrow range (between 44% and 52%), although the proportion for Bard abstracts (63.21%) was clearly excessive. In terms of occurrence rates, then, Figure 4 suggests that a simplex rate of about 2 per 100 words could be taken to 'expected' for original abstracts; a rate that deviates too markedly, such as that for Bard abstracts (3.79), would thus be a red flag.

The exception involving Bard abstracts resulted in passages such as (16), where each clause complex is a simplex, with (16b) comprising an embedded clause. Such extended simplex segments are absent in Nature abstracts.

(16) a. ||| *In this study, we report the observation of the orbital Hall effect in titanium* (Ti). |||

b. *We find [[that the orbital Hall conductivity in Ti is strongly anisotropic, with a large value along the [001] direction*.]] |||

c. *This anisotropy is consistent with the theoretical prediction for the orbital Hall effect in a light metal.* |||

d. *Our findings provide new insights into the orbital Hall effect in light metals.* |||

e. *This effect may have important implications for the understanding of the electronic structure and transport properties of these materials.* ||| *(Bard 25)*

Hence, while the use of simplexes in (16) might appear to be aligned with the characterization of science writing being objective and thing-centered, its overuse may result in texts that appear overly simple and point-like. By contrast, when used in Nature abstracts, simplexes often contain embedded clauses, reflecting in turn the general denseness of science writing (Biber & Gray, 2016). That is to say, science writing is not merely thing-centered, but generally dense, containing embedded information. This is illustrated in (17) below.

(17) a. ||| *Recent efforts* [[…B…]] *have therefore focused on problems*

b. B [[*to demonstrate quantum speedups*]]

c. [[*that are both classically hard and naturally suited to current quantum hardware,*

d. *such as* [[*sampling from complicated – although not explicitly useful – probability distributions*.]]]] ||| (Nature 04)

As there is only a single ranking clause in (17), it is, by definition, a simplex. The simplex, though, is packed with information. Through the use of three embedded clauses, with (17d) being embedded within another embedded clause, the reader is told about quantum speedups, the nature of the focused problems, and an example of one such problem.

Where hypotaxis is concerned, more ranking than embedded clauses were involved. In Nature (and also ChatGPT) abstracts, the occurrence rate of hypotactic ranking clauses was about three times that of embedded clauses, and although there is no reason to expect hypotactic ranking and embedded clauses to be correlated in any systematic or meaningful way, particularly in short texts such as abstracts, the near one-to-one ratio involving Poe abstracts was unusual.

The extent of the under- or overuse of hypotaxis in relation to the Nature rate can thus serve as another diagnostic. The rate for hypotactic ranking clauses in ChatGPT abstracts, for instance, was significantly higher than that for Nature abstracts (2.15 vs. 1.64). This was due primarily to their more frequent occurrence; in a few cases, hypotaxis also involved more than two subordinate clauses (18), which was uncommon in Nature abstracts.

(18) α ||| *Our investigations reveal highly efficient single-photon absorption* ||
β *followed by rapid and coherent energy transfer,* ||
γ *facilitating near-unity quantum efficiency in photon emission.* ||| *(ChatGPT 07)*

As regards hypotactic embedded clauses, Figure 4 suggests that the low occurrence rate for Nature abstracts (0.49) was generally true for the other abstract groups as well, notwithstanding the statistically significant difference between the rates for Nature and Poe. In all cases, the occurrence rates did not exceed 1 per 100 words. Even so, the rate for Poe abstracts was almost twice that for Nature abstracts (0.88 vs. 0.49). We see a typical example of this in (19–20), which compare the statements of results in the Nature abstract and the Poe-generated version of the same title.

(19) ||| *We demonstrate*
α^{E} [[*that neurons within the lateral hypothalamus [[…C…]] utilize these signals*
C [[*that produce the stimulatory neuropeptide neurotensin and the inhibitory neurotransmitter GABA (γ-aminobutyric acid)*]]
β^{E} *to coordinately activate dopamine-producing neurons of the ventral tegmental area.*]] ||| (Nature 12)

(20) ||| *Furthermore, we show*
[[*that the circuit is modulated by a feedback mechanism*
α^{E} [[*involving the melanocortin system,* ||
β^{E} α^{E} *which integrates information from peripheral signals such as leptin and insulin* ||
β^{E} *to regulate food intake and energy balance.*]]]] ||| (Poe 12)

Although (19–20) are simplexes, hypotaxis occurs twice in the Poe version with the second occurrence involving embedded clauses within a larger embedded segment. A note of caution is perhaps warranted here. Seeing that both examples broadly reflect the condensed nature of science writing, and it is not out of the ordinary that packing information into simplexes using embedded clauses can result in hypotaxis (or parataxis), such an indicator is perhaps weak, and needs to be supplemented with other diagnostics.

4.4 *Logico-semantic relations*

As no significant results were returned for parataxis (Table 4), the rates of paratactic logico-semantic relations were, expectedly, also insignificant. As Table 5 shows, the use of extending clauses was similar across the abstract groups, but less common relations such as enhancement and elaboration were absent in all groups. For instance, no enhancing and elaborating clauses were found in ChatGPT and Poe abstracts for both ranking and embedded clauses, and embedded elaborating clauses were absent in all abstract groups. Embedded enhancing and elaborating clauses were also absent in Nature abstracts. On the whole, paratactic logico-semantic relations occur less than once per 100 words; any rate that is higher than this range is likely to be anomalous.

Table 5: Occurrence rates of logico-semantic relations (parataxis)

	Nature		**Bard**		**ChatGPT**		**Poe**	
	Rate	**SD**	**Rate**	**SD**	**Rate**	**SD**	**Rate**	**SD**
Ranking								
Extension	0.65	1.04	0.59	1.19	0.55	0.98	0.51	0.42
Enhancement	0.02	0.20	0.01	0.14	–	–	–	–
Elaboration	0.07	0.35	–	–	–	–	–	–
Embedded								
Extension	0.36	0.92	0.34	0.68	0.30	0.81	0.53	1.16
Enhancement	–	–	0.01	0.14	–	–	–	–
Elaboration	–	–	–	–	–	–	–	–

Where hypotaxis is concerned, Table 6 captures the rates for both embedded and ranking clauses. Given the low occurrence rate of hypotactic embedded clauses (Figure 4), the statistically insignificant results for embedded extending and enhancing clauses in Table 6 are not unexpected. Even though more elaborating clauses were used in Poe abstracts as compared to the other abstract groups, the rate was low (0.42).

Table 6: Occurrence rates of logico-semantic relations (hypotaxis)

			Games-Howell comparisons (*p*)			
Hypotaxis	**Rate**	**SD**	**Nature**	**Bard**	**ChatGPT**	**Poe**
Extension (embedded)						
Nature	0.01	0.14				
Bard	–	–				
ChatGPT	–	–				
Poe	–	–				
Enhancement (embedded)						
Nature	0.38	1.11				
Bard	0.56	1.23				
ChatGPT	0.54	1.31				
Poe	0.46	1.00				
Elaboration (embedded)						
Nature	0.10	0.55				$3.94e^{-04}$**
Bard	0.17	0.58				.02*
ChatGPT	0.12	0.40				$5.87e^{-04}$**
Poe	0.42	0.96	$3.94e^{-04}$**	.02*	$5.87e^{-04}$**	
Extension (ranking)						
Nature	0.04	0.31				
Bard	0.03	0.31				
ChatGPT	–	–				
Poe	0.08	0.40				
Enhancement (ranking)						
Nature	1.27	1.93				$6.57e^{-05}$**
Bard	0.88	1.01			.02*	
ChatGPT	1.34	1.87		.02*		$1.69e^{-05}$**
Poe	0.63	0.66	$6.57e^{-05}$**		$1.69e^{-05}$**	
Elaboration (ranking)						
Nature	0.33	0.93			$1.45e^{-0.4}$**	
Bard	0.46	0.96			.03*	
ChatGPT	0.81	1.22	$1.45e^{-0.4}$**	.03*		$4.71e^{-07}$**
Poe	0.22	0.68			$4.71e^{-07}$**	

ANOVA (embedded; elaboration): $F(3, 106.55)=6.49$, $p<4.46e^{-04}$**

ANOVA (ranking; enhancement): $F(3, 98.63)=14.97$, $p<4.20e^{-08}$**

ANOVA (ranking; elaboration): $F(3, 105.50)=12.16$, $p=6.77e^{-07}$**

The more common logico-semantic relations involving hypotactic ranking clauses were enhancement and elaboration, with the former having a generally higher occurrence rate (Figure 5). As compared to Nature abstracts, ChatGPT abstracts used more elaborating clauses; the rate was more than twice that for Nature abstracts (0.81 vs. 0.33). These tended to be realized by *–ing* clauses (21), which did not occur as frequently in Nature abstracts.

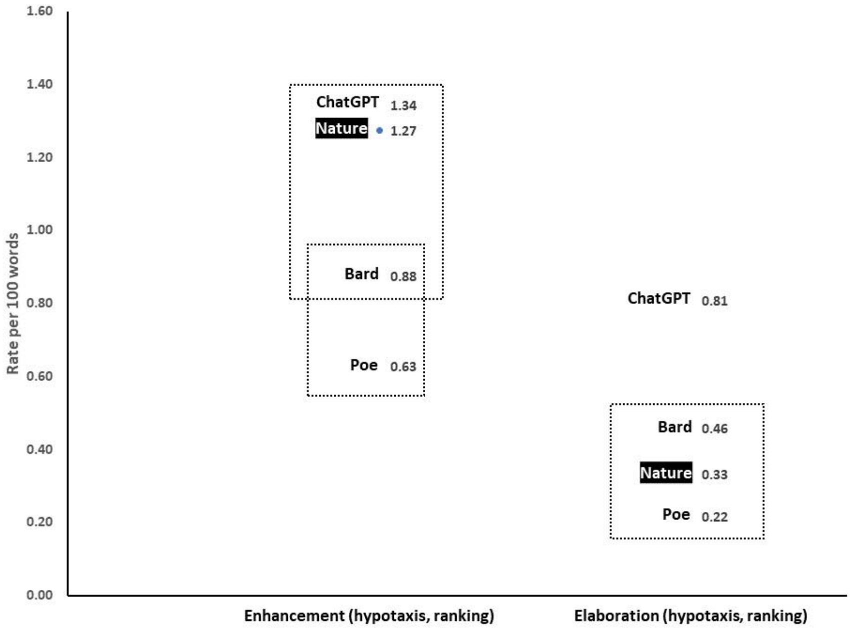

Figure 5: Enhancement and elaboration involving ranking clauses—Nature abstracts as compared to the other abstract groups

By contrast, the occurrence of hypotactic enhancing clauses was about the same for ChatGPT and Nature abstracts. Although the difference between Nature and Bard abstracts (1.27 vs. 0.88) was statistically insignificant, it was only weakly so (p=.06); the difference between ChatGPT and Bard abstracts, on the other hand, was significant (p=.02). The Bard rate may thus be taken to be near the lower boundary of what is considered typical of ranking enhancing clauses in original abstracts.

Despite the similarity between Nature and ChatGPT for ranking enhancing clauses, the occurrence rates of finite adverbial clauses in Table 7 reveal an interesting difference.

Table 7: Occurrence rates of finite adverbial clauses

			Games-Howell comparisons (*p*)			
Clauses	**Rate**	**SD**	**Nature**	**Bard**	**ChatGPT**	**Poe**
Nature	0.30	0.83			2.60e-03**	1.70e-04**
Bard	0.16	0.56				0.03*
ChatGPT	0.05	0.36	2.60e-03**			
Poe	0.02	0.20	1.70e-04**	0.03*		

Since finite adverbial clauses can only be enhancing in function (Halliday & Matthiessen, 2014: 481), Table 7 shows that Nature abstracts used far more of such clauses to express circumstantial information (e.g., time, purpose, condition); they were six times as likely to occur in Nature abstracts than in ChatGPT abstracts (0.30 vs. 0.05). For instance, (21), taken from Nature 16, contains four adverbial clauses, two of which are located in consecutive clause complexes. The function of the adverbial clause is indicated in parentheses at the end of each subordinate clause:

(21) α ||| *In mammalian cells, the decision [[to proliferate]] is thought to be irreversibly made at the restriction point of the cell cycle* ||

×β *when mitogen signalling engages a positive feedback loop between cyclin A2/cyclin-dependent kinase 2 (CDK2) and the retinoblastoma protein.* ||| *(time)*

[…]

α ||| *This temporal competition between two fates, mitosis and cell cycle exit, arises* ||

xβ *because cyclin A2/CDK2 activity depends upon CDK4/6 activity throughout the cell cycle, not just in G1 phase.* ||| (reason)

α *Without mitogens, mitosis is only observed* ||

×β *when the half-life of cyclin A2 protein is long enough [[to sustain CDK2 activity throughout G2/M.*]] ||| (time)

[…]

α ||| *Consequently, even a 2-h delay in a cell's progression towards mitosis can induce cell cycle exit* ||

×β *if mitogen signalling is lost.* ||| (condition) (Nature 16)

ChatGPT abstracts, on the other hand, tended to rely on other clauses, such as –*ing* and *to*-infinitive clauses. Example (22), which is also taken from a single abstract (ChatGPT 03), shows the use of both these non-finite clauses.

(22) α ||| *The novel technique incorporates superconducting resona-*

tors and microwave single-photon detectors ||

×β *to reliably identify spin resonance of individual electrons.* ||| (purpose)

[…]

α ||| *This approach allows for unprecedented high fidelity readout of electron spin states,* ||

×β *realizing an improvement by an order of magnitude in quantum bit error rate over existing methods.* ||| (result) (ChatGPT 03)

On the face of it all, this difference between Nature and ChatGPT does not fit the earlier characterization of science writing being condensed. Biber and Gray (2016: 103, 115), for instance, note that finite adverbial clauses are more characteristic of spoken language, and that word phrases and non-finite clauses are favored in specialist science genres. If so, would not the higher occurrence of finite adverbial clauses in Nature abstracts run against this characterization of science writing? This concern is misplaced for two reasons.

First, finite adverbial clauses form just one part of the larger group of subordinate clauses. As Table 7 indicates, the occurrence rate for Nature abstracts was comparatively low (0.30). Of the 179 subordinate clauses in Nature abstracts, only 30 (or 16.76%) were finite adverbial clauses; the total number of non-finite clauses was 127 (70.95%). The occurrence rate of non-finite clauses in Nature abstracts was 1.20, four times that of adverbial clauses.

Second, only a single genre—research-article abstracts in the sciences—was analyzed in this study. The rate of adverbial clauses may indeed be different in other science-based genres (e.g., full research articles, project proposals). However, given the work by Biber and Gray (2016) and Leong (2021b), among others, it is doubtful that the use of adverbial clauses in these other genres would be so different as to exceed that of non-finite clauses.

5. Conclusion

This study sought to compare the use of clauses and inter-clausal relations between original and AI-generated abstracts. Nature abstracts were compared with versions generated by Bard, ChatGPT, and Poe. The following are the broad findings of the analysis:

(a) The occurrence rates of main clauses (4.59) and subordinate clauses (1.71) in Nature abstracts occupied a middle position relative to the other abstract groups. The rate of embedded clauses for Nature abstracts was 3.35, differing only from that for Bard abstracts.

(b) With the exception of Bard abstracts, the rate of simplexes in the other three abstract groups was about 2. The proportion of simplexes to main clauses in Nature, ChatGPT, and Poe abstracts was between 44% and 52%.

(c) Parataxis and paratactic logico-semantic relations were uncommon, and the differences among the abstract groups were statistically insignificant.

(d) Hypotactic logico-semantic relations in Nature abstracts were more common among ranking clauses (1.64) than embedded clauses (0.49). The low occurrence rate of embedded hypotactic clauses (<1) was also generally true for the other abstract groups. Hence, even though embedded clauses occurred more often than subordinate clauses, it is the latter that carried the bulk of logico-semantic relations. The most common hypotactic logico-semantic relation involving ranking clauses in Nature and ChatGPT abstracts was enhancement. While ChatGPT abstracts relied mostly on non-finite clauses to express circumstantial information, Nature abstracts also used finite adverbial clauses.

None of the chatbots matched Nature in all categories. Poe abstracts, in fact, differed from Nature abstracts in every category (with a significant ANOVA result) except the use of embedded clauses. The chatbot that came closest to matching Nature abstracts was ChatGPT; it differed in the use of subordinate, elaborating, and finite adverbial clauses. Even so, distinct differences were detected—ChatGPT was twice as likely to use elaborating clauses, but only one sixth as likely to use finite adverbial clauses.

The statistics for Nature abstracts offer a possible means by which original texts can be separated from AI-generated ones. Rather than content, the focus here is on the building blocks that construct such content. The work of Levin et al. (2023) has demonstrated one way to do this using Grammarly; this present work proposes another using clause complexing.

But these are just starting points. The major limitation of this study is that only science-based abstracts from a single journal were considered. The same approach, however, can be easily extended to other genres (e.g., technical reports, literature reviews), and involve different disciplines and linguistic features. Moving forward, AI-detection tools can draw on the enormous amount of research work done on different genres and incorporate distinctive features into the algorithm. Recent examples include studies on cohesive devices in narrative and argumentative texts (Tabari & Johnson, 2023), lexical complexity in scientific letters (Zhou et al., 2023), and the use of the passive voice in research articles in the sciences and humanities (Leong, 2021a), among many others. A genre-based approach, that is to say, is crucially needed for detection results to be more reliable.

Scholars (and students) using chatbots to generate content for publication remains a vexing problem. While generative AI has its use in facilitating personalized learning and automating common, tedious tasks, it is merely a tool, but not an author; it has no independent thought, and so cannot take responsibility for its own generated content. Several journals, such as the PNAS (Proceedings of the National Academy of Science), Sage, and Science families of journals, have already banned the listing of chatbots as a 'co-author' in articles; generated content in the manuscript must also be clearly attributed. The recent voluntary commitments by OpenAI, Google, and Meta to watermark AI-generated content is a welcome move (Bartz & Hu, 2023). But detection programs continue to be needed, given how quickly chatbot texts can proliferate, not all of which may be watermarked.

References

Armstrong, Elizabeth M. (1992). Clause complex relations in aphasic discourse: A longitudinal case study. *Journal of Neurolinguistics* 7(4): 261–275. https://doi.org/10.1016/0911-6044(92)90018-R

Babl, Franz E., Maximilian P. Babl (2023). Generative artificial intelligence: Can ChatGPT write a quality abstract? *Emergency Medicine Australasia*. Advance online publication. https://doi.org/10.1111/1742-6723.14233

Bartz, Diane, Krystal Hu (2023). OpenAI, Google, others pledge to watermark AI content for safety, White House says. *Reuters*. https://www.reuters.com/technology/openai-google-others-pledge-watermark-ai-content-safety-white-house-2023-07-21/

Biber, Douglas, Bethany Gray (2016). *Grammatical Complexity in Academic English: Linguistic Change in Writing*. Cambridge: Cambridge University Press.

Carr, David F. (2023). As ChatGPT growth flattened in May, Google Bard rose 187%. *Similarweb*. https://www.similarweb.com/blog/insights/ai-news/chatgpt-bard/

Cingillioglu, Ilker (2023). Detecting AI-generated essays: The ChatGPT challenge. *International Journal of Information and Learning* 40(3): 259–268. https://doi.org/10.1108/IJILT-03-2023-0043

Coles, Cameron (2023). 11% of data employees paste into ChatGPT is confidential. *Cyberhaven*. https://www.cyberhaven.com/blog/4-2-of-workers-have-pasted-company-data-into-chatgpt/

Elad, Barry (2023). Number of companies using ChatGPT within their business 2023, by industry. *Enterprise Apps Today*. https://www.enterpriseappstoday.com/stats/openai-statistics.html#

Elyoseph, Zohar, Dorit Hadar-Shoval, Kfir Asraf, Maya Lvovsky (2023). ChatGPT outperforms humans in emotional awareness evaluations. *Frontiers in Psychology* 14: 1199058. https://doi.org/10.3389/fpsyg.2023.1199058

Fawcett, Robin P. (2000). *A Theory of Syntax for Systemic Functional Linguistics*. Amsterdam: John Benjamins.

Fowler, Geoffrey A. (2023). We tested a new ChatGPT-detector for teachers. It flagged an innocent student. *The Washington Post*. https://www. Washington post.com/technology/2023/04/01/chatgpt-cheating-detection-turnitin/

Gao, Catherine A., Frederick M. Howard, Nikolay S. Markov, Emma C. Dyer, Siddhi Ramesh, Yuan Luo, Alexander T. Pearson (2022). Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers. *bioRχiv*. https://doi.org/10.1101/2022.12.23.521610

Gao, Catherine A., Frederick M. Howard, Nikolay S. Markov, Emma C. Dyer, Siddhi Ramesh, Yuan Luo, Alexander T. Pearson (2023). Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *npj Digital Medicine* 6: 75. https://doi.org/10.1038/s41746-023-00819-6

Graesser, Arthur C. (2016). Conversations with AutoTutor help students learn. *International Journal of Artificial Intelligence in Education* 26(1): 124–132. https://doi.org/10.1007/s40593-015-0086-4

Greenbaum, Sidney, Gerald Nelson (1995). Clause relationships in spoken and written English. *Functions of Language* 2(1): 1–21. https://doi.org/10.1075/fol.2.1.02gre

Halliday, Michael A. K. (2006). On the grammatical foundations of discourse. Webster, Jonathan J., ed. *Studies in Chinese Language*. London: Continuum, 346–363.

Halliday, Michael A. K., Christian M. I. M. Matthiessen (2014). *Halliday's Introduction to Functional Grammar* (4th edn.). London: Routledge.

Halliday, Michael A. K. (1998). Things and relations: Regrammaticising experience as technical knowledge. Martin, James R., Robert Veel, eds. *Reading Science: Critical and Functional Perspectives on Discourses of Science*. London: Routledge, 185–237.

Halliday, Michael A. K., James R. Martin (1993). *Writing Science: Literacy and Discursive Power*. London: Routledge.

Heaven, Will Douglas (2023). Google just launched Bard, its answer to ChatGPT—and it wants you to make it better. *MIT Technology Review*. https://www. technologyreview.com/2023/03/21/1070111/google-bard-chatgpt-openai-microsoft-bing-search/

Houston, Aileen B., Edward M. Corrado (2023). Embracing ChatGPT: Implications of emergent language models for academia and libraries. *Technical Services Quarterly* 40(2): 76–91. https://doi.org/10.1080/07317131.2023.2187110

Hu, Krystal (2023). ChatGPT sets record for fastest-growing user base—analyst note. *Reuters*. https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/

Jimenez, Kayla (2023). Professors are using ChatGPT detector tools to accuse students of cheating. But what if the software is wrong? *USA Today*. https://www.usatoday.com/story/news/education/2023/04/12/how-ai-detection-tool-spawned-false-cheating-case-uc-davis/11600777002/

Kohnke, Lucas (2023). L2 learners' perceptions of a chatbot as a potential independent language learning tool. *International Journal of Mobile Learning and Organisation* 17(1/2): 214–226. https://doi.org/10.1504/IJMLO.2023.128339

Leong, Ping Alvin (2021a). The passive voice in scholarly writing: A diachronic look at science and history. *Finnish Journal of Linguistics* 34(2021): 77–102.

Leong, Ping Alvin (2021b). Writing in the sciences and humanities: A clause-complex perspective. *Word* 67(2): 137–158. https://doi.org/10.1080/00437956.2021.1909866

Leong, Ping Alvin, Bee Geok Wee (2005). Investigating the clause complex: An analysis of exposition-type essays written by secondary school students in Singapore. *ITL – International Journal of Applied Linguistics* 150: 47–76. https://doi.org/10.2143/ITL.150.0.2004372

Levin, Gabriel, Raanan Meyer, Eva Kadoch, Yoav Brezinov (2023). Identifying ChatGPT-written OBGYN abstracts using a simple tool. *American Journal of Obstetrics & Gynecology MFM* 5(6): 100936. https://doi.org/10.1016/j.ajogmf.2023.100936

Lund, Brady D., Ting Wang, Nishith Reddy Mannuru, Bing Nie, Somipam Shimray, Ziang Wang (2023). ChatGPT and a new academic reality: Artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing. *Journal of the Association for Information Science and Technology* 74(5): 570–581. https://doi.org/10.1002/asi.24750

Mehdi, Yusuf (2023). Confirmed: The new Bing runs on OpenAI's GPT-4. *Microsoft Bing Blogs*. https://blogs.bing.com/search/march_2023/Confirmed-the-new-Bing-runs-on-OpenAI%E2%80%99s-GPT-4

Naidu, Katharine, Kershnee Sevnarayan (2023). ChatGPT: An ever-increasing encroachment of artificial intelligence in online assessment in distance education. *Online Journal of Communication and Media Technologies* 13(3): e202336. https://doi.org/10.30935/ojcmt/13291

Najafali, Daniel, Chandler Hinson, Justin M. Camacho, Logan G. Galbraith, Rohun Gupta, Chris M. Reid (2023). Can chatbots assist with grant writing in plastic surgery? Utilizing ChatGPT to start an R01 grant. *Aesthetic Surgery Journal* 43(8): NP663–NP665. https://doi.org/10.1093/asj/sjad116

Nesbitt, Christopher, Guenter Plum (1988). Probabilities in a systemic-functional grammar: The clause complex in English. Fawcett, Robin P., David Young, eds. *New Developments in Systemic Linguistics, Volume 2: Theory and Application*. London: Pinter, 6–38.

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

Perez, Sarah (2023). Quora opens its new AI chatbot app Poe to the general public. *TechCrunch*. https://techcrunch.com/2023/02/06/quora-opens-its-new-ai-chatbot-app-poe-to-the-general-public/

Rahimi, Farid, Amin Talebi Bezmin Abadi (2023). Passive contribution of ChatGPT to scientific papers. *Annals of Biomedical Engineering*. Advance online publication. https://doi.org/10.1007/s10439-023-03260-8

Scimago (2022). http://www.scimagojr.com/

Sellami-Baklouti, Akila (2011). The impact of genre and disciplinary differences on structural choice: Taxis in research article abstracts. *Text & Talk* 31(5): 503–523. https://doi.org/10.1515/text.2011.025
Sharples, Mike (2022). Automated essay writing: An AIED opinion. *International Journal of Artificial Intelligence in Education* 32(4): 1119–1126. https://doi.org/10.1007/s40593-022-00300-7
Stack Overflow (2023). Temporary policy: Generative AI (e.g., ChatGPT) is banned. https://meta.stackoverflow.com/questions/421831/temporary-policy-generative-ai-e-g-chatgpt-is-banned
Tabari, Mahmoud Abdi, Mark D. Johnson (2023). Exploring new insights into the role of cohesive devices in written academic genres. *Assessing Writing* 57: 100749. https://doi.org/10.1016/j.asw.2023.100749
Teixeira da Silva, Jaime A. (2023a). ChatGPT: Detection in academic journals is editors' and publishers' responsibilities. *Annals of Biomedical Engineering*. Advance online publication. https://doi.org/10.1007/s10439-023-03247-5
Teixeira da Silva, Jaime A. (2023b). Is ChatGPT a valid author? *Nurse Education in Practice* 68: 103600. https://doi.org/10.1016/j.nepr.2023.103600
Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, Illia Polosukhin (2017). Attention is all you need. Von Luxburg, Ulrike, Isabelle Guyon, Samy Bengio, Hanna Wallach, Rob Fergus, S. V. N. Vishwanathan, Roman Garnett, eds. *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*. La Jolla, CA: Neural Information Processing Systems, 5999–6009.
Yan, Da (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies*. Advance online publication. https://doi.org/10.1007/s10639-023-11742-4
Zaiontz, Charles (2022). *Real Statistics Using Excel*. http://www.real-statistics.com/.
Zhou, Xinye, Yuan Gao, Xiaofei Lu (2023). Lexical complexity changes in 100 years' academic writing: Evidence from Nature biology letters. *Journal of English for Academic Purposes* 64: 101262. https://doi.org/10.1016/j.jeap.2023.101262

Author's address:

Alvin Ping Leong
Language and Communication Centre
School of Humanities, Nanyang Technological University
48 Nanyang Avenue, Singapore 639818
Singapore
E-mail: alvin.leong@ntu.edu.sg

Received: August 12, 2023

Accepted for publication: October 26, 2023

Appendix
Symbols used in the analysis

Symbol	Description
| | |	Clause complex / simplex
| |	Clause
[[…]]	Embedded clause
1, 2, 3, …	Taxis: Parataxis
α, β, γ, …	Taxis: Hypotaxis
=	Expansion: Elaboration
+	Expansion: Extension
×	Expansion: Enhancement
"E" superscript (e.g., α^E, β^E)	Related to embedded clauses only