Academics' perceptions of ChatGPT-generated written outputs: A practical application of Turing's Imitation Game

Joshua Matthews, Catherine Rita Volpe
University of New England, Australia

Artificial intelligence (AI) technology, such as Chat Generative Pre-trained Transformer (ChatGPT), is evolving quickly and having a significant impact on the higher education sector. Although the impact of ChatGPT on academic integrity processes is a key concern, little is known about whether academics can reliably recognise texts that have been generated by AI. This qualitative study applies Turing's Imitation Game to investigate 16 education academics' perceptions of two pairs of texts written by either ChatGPT or a human. Pairs of texts, written in response to the same task, were used as the stimulus for interviews that probed academics' perceptions of text authorship and the textual features that were important in their decision-making. Results indicated academics were only able to identify AI-generated texts half of the time, highlighting the sophistication of contemporary generative AI technology. Academics perceived the following categories as important for their decision-making: voice, word usage, structure, task achievement and flow. All five categories of decision-making were variously used to rationalise both accurate and inaccurate decisions about text authorship. The implications of these results are discussed with a particular focus on what strategies can be applied to support academics more effectively as they manage the ongoing challenge of AI in higher education.

Implications for practice or policy:

- Experienced academics may be unable to distinguish between texts written by contemporary generative AI technology and humans.
- Academics are uncertain about the current capabilities of generative AI and need support in redesigning assessments that succeed in providing robust evidence of student achievement of learning outcomes.
- Institutions must assess the adequacy of their assessment designs, AI use policies, and AI-related procedures to enhance students' capacity for effective and ethical use of generative AI technology.

Keywords: generative artificial intelligence (AI), ChatGPT, Imitation Game, education academics, initial teacher education, assessment, thematic analysis

Introduction

Contemporary artificial intelligence (AI) has profoundly disrupted teaching and learning in the higher education sector. Current generative AI applications, underpinned by large language learning models, can complete various tasks, such as producing highly integrated human-like academic texts (Kasneci, et al. 2023). One application of current relevance is Chat Generative Pre-trained Transformer (ChatGPT) (Kasneci et al., 2023; Lodge, Thompson, & Corrin, 2023; Rudolph et al., 2023), which is freely available to anyone with access to the internet (https://chat.openai.com). ChatGPT presents a chat interface into which users can enter text prompts and be immediately provided with bespoke text outputs. The application is surprisingly powerful; simple text prompts can be used to produce complex human-like written responses (Gao et al., 2023; Hulman et al., 2023; Nov et al., 2023).

Although relatively recent commentary predicted that generative AI would rapidly become an intrinsic component of higher education (Popenici & Kerr, 2017), the speed at which it has impacted the sector has been unanticipated (Lodge, Thompson, & Corrin, 2023). In a 2019 systematic review, Zawacki-Richter et al. asserted that although AI has been a presence in education for several decades, it is "still unclear for educators … how it can actually impact meaningfully on teaching and learning in higher education" (p. 1). Certainly, such assertions do not hold true today. A 2023 survey of 100 news articles on ChatGPT in higher education has highlighted contemporary concerns around academic integrity and assessment design (Sullivan et al., 2023). Other research reporting on the capabilities of AI language models (e.g., GPT-2 and GPT-3) available just a few years ago provides a sharp contrast with the AI-mediated affordances available today (Elkins & Chun, 2020; Floridi & Chiriatta, 2020). Floridi and Chiriatta overviewed the capabilities of the AI language model GPT-3 (Generative Pre-trained Transformer 3) and asserted that semantic questions, or those "which require understanding and perhaps even experience of both the meaning and the context" (p. 682), could be used to elicit non-humanlike (*bot-like*) responses from the model. They used the semantic question "How many feet can you fit in a shoe?" and the resultant text output to emphasise their point. A comparison of the respective text outputs produced by GPT-3 in 2020 and ChatGPT (version 3.5) in 2023 (Table 1) shows the improvements in generative AI that have occurred in just 3 years.

Table 1
A targeted comparison of large language model capabilities between 2020 and 2023

Language model	GPT-3	ChatGPT (version 3.5)
Date procured	2020	24 May 2023
Source	Floridi & Chiriatta (2020, p. 689)	https://chat.openai.com/
Output from the text prompt "How many feet can you fit in a shoe?".	"I don't know," said the boy. "How many?". The man replied, "Ten.". The boy then asked, "Tell me how many inches are in a yard?" Again, the man replied, "Ten." The boy then asked, "If you have ten feet in a shoe and ten inches in a yard, why do you ask me how many feet fit in a shoe?"	The number of feet that can fit in a shoe depends on the size and design of the shoe. Generally, shoes are made to accommodate one foot per shoe. However, there are certain types of footwear, such as sandals or flip-flops, that can accommodate multiple feet, such as in the case of shared or communal footwear in certain cultural practices. In those cases, the design allows for multiple feet to fit comfortably. In most standard shoes, though, only one foot is meant to fit in each shoe for proper fit, support, and comfort.

Note. Both GPT-3 and ChatGPT are related large language models. ChatGPT is a variation of GPT-3 that has been designed to facilitate more conversational interactions.

Although AI-mediated technology, such as ChatGPT, represents an important opportunity for higher education (Ouyang et al., 2022), it also represents a significant disruption (Dawson, 2021; Lodge, Thompson, & Corrin, 2023; Sullivan et al. 2023). As Lodge, Thompson and Corrin made clear, research around academic integrity and AI is crucial because students can obtain sophisticated written artefacts by simply "entering the assessment instructions into a tool such as ChatGPT, without going through the process of learning themselves" (p. 3). ChatGPT is extremely accessible, easy to use and raises several pressing questions for various stakeholders in higher education. For university academics with marking duties, a recurring question arises: "Was this text written by a machine or a human?"

Turing's Imitation Game

Turing's (1950) seminal article "Computing Machinery and Intelligence" presented a criterion for assessing the humanness of the communicative artefacts produced by computers. For decades, the so-called Imitation Game (or Turing test) has been a reference point for philosophical debate around the nature of AI (Block, 1981; Chomsky, 2008; French, 2000; Harnad, 1992). The Imitation Game itself is a thought-experiment involving three players: a human interrogator (A), a human subject (B) and a machine subject (C) (see Figure 1). All three players are located in a separate room and cannot see one another, but the interrogator can interact with each subject by posing questions via typewritten language. For the interrogator, the objective of the game is to determine which texts have been written by the human and which by the machine (computer).

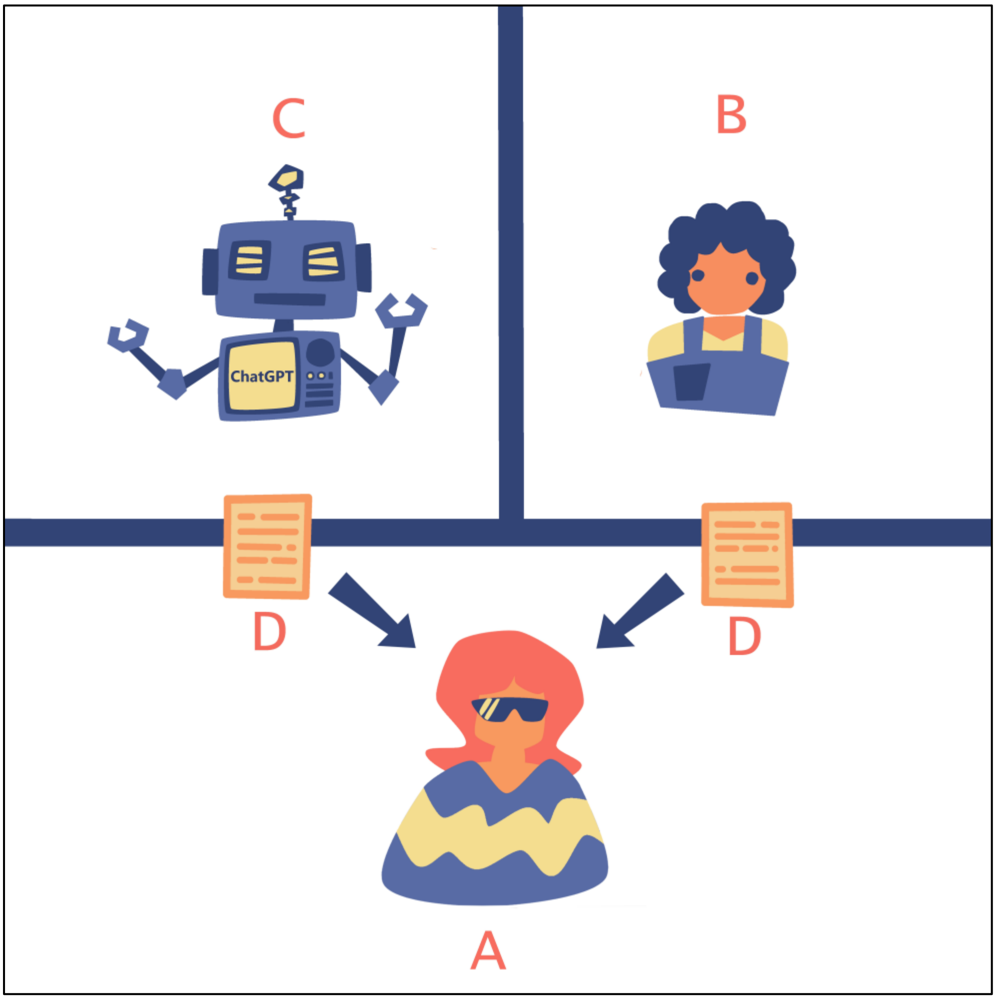

Figure 1. A schematic of the Imitation Game
Note. Academic/Human interrogator (A); Student/Human subject (B); ChatGPT/Machine subject (C); texts (D).

The machine's objective is to provide responses so humanlike that the interrogator cannot reliably determine which player is the machine and which is the human. In Turing's (1950) original description, the game ends with the interrogator making clear their position about which subject they believe is the machine and which they believe is the human.

The outcome of the game rests on two interrelated factors: first, the capacity of the machine to produce communicative artefacts that imitate the attributes of those produced by humans. As has been alluded to already, and will be reviewed in more detail below, there is emerging evidence that ChatGPT currently possesses ample capacity to interpret text prompts and produce sophisticated human-like texts (Gao et al., 2023; Hulman et al., 2023; Nov et al., 2023); second, and the focus of the current study, is the degree to which the human interrogator is sensitive to the attributes of communicative artefacts that indicate whether they are produced by a human or a machine.

The Imitation Game paradigm to investigate ChatGPT

As commercially available digital AI output detectors have low reliability (Elkhatat et al., 2023; Kirchner et al., 2023), it is particularly important to investigate the capacity of human interrogators to detect AI-generated text. In higher education, it is ultimately the academic's judgement around whether a text is AI-generated that underpins decisions about whether a breach of academic integrity has occurred. Such decisions are important to higher education institutions as they can impact key areas, such as student satisfaction and well-being, institutional reputation, and academic workloads. Thus, it is crucial to understand how readily academics can identify AI-generated texts to fully understand the impact AI may have on the higher education sector now and into the future (Lodge, Thompson, & Corrin, 2023). The Imitation Game paradigm is analogous to the decision-making processes imposed on academics as a consequence of contemporary AI, such as ChatGPT, and was applied in the current study as an overarching methodological and theoretical framework. The human interrogator (e.g., the academic – A in Figure 1) needs to determine, one way or the other, if communicative artefacts (e.g., texts submitted for academic credit, D in Figure 1) have been generated by a human (e.g., students – B in Figure 1) or a machine (e.g., ChatGPT – C in Figure 1).

To our knowledge, the Imitation Game paradigm has not been used to investigate ChatGPT in relation to initial teacher education; however, it has been applied in several other contexts. Gao et al. (2023) investigated human interrogators' capacities to differentiate scientific article abstracts generated by ChatGPT and human authors. To do so, 50 article abstracts and their corresponding titles were drawn from high quality journals. The prompt "Please write a scientific abstract for the article [title] in the style of [journal]" was then used to get ChatGPT to produce an abstract for each article. Blinded reviewers were given 25 abstracts, informed that these contained a mix of AI-generated and human-written texts and were asked to provide a binary ruling on the origins of each. Reviewers correctly identified 68% of the AI-generated abstracts and correctly identified 86% of the original abstracts as being written by humans. Although the qualitative data procured as part of the research was relatively limited, reviewers reported being surprised by the difficulty of distinguishing between the AI-generated and human-generated texts. Vague, formulaic and superficial language was noted as being useful in determining ChatGPT-generated texts.

Nov et al. (2023) investigated whether 392 survey respondents recruited from the general public could distinguish between healthcare-related answers generated by ChatGPT or human healthcare providers. Authentic written discourse between patients and healthcare providers was sourced from a digital health register, and 10 questions and their original answers were extracted. The same questions were then used as prompts for the generation of corresponding answers from ChatGPT. The original questions were presented to participants along with two alternative AI-generated and human-generated answers, with participants being informed that five of the answers had been written by AI. On average, ChatGPT-generated answers were identified 65.5% of the time and human-generated answers were identified 65.1% of the time; however, significant variation (49.0%–85.7%) was observed across the different questions in terms of the ease with which they could be correctly identified.

Another recent study inspired by Turing's Imitation Game was undertaken by Hulman et al. (2023) among 183 employees of a large health provider service in Denmark. The objective of the study was to determine how adequately ChatGPT could answer 10 frequently asked questions that were of relevance to the healthcare service (i.e., questions about diabetes). The questions were selected and answered in writing by a human expert. The same questions were also used as the prompts for ChatGPT to produce a set of AI-generated answers. Each question and its corresponding pair of answers (one AI-generated and one human-generated) were presented to human interrogators with the instruction to identify those produced by ChatGPT. On average participants could correctly identify the ChatGPT-generated texts 59.5% of the time. Again, as with the findings of Gao et al. (2023), the qualitative data tapping participants' perceptions of the ChatGPT-generated texts were relatively limited; however, it appeared to be the linguistic features of the texts, rather than their factual content, that provided the strongest indication that the texts were AI-generated.

Overall, in the studies reviewed, human interrogators could correctly identify AI-generated texts in more than 50% of cases. Research also suggests that there are certain attributes of AI-generated texts that interrogators perceived to be important in their identification, but details provided around these textual features were sparse. It seems clear that in at least some cases, ChatGPT could produce texts that were so humanlike that they were difficult to differentiate from human-generated text.

The current study

The current study adds to the emerging body of research that uses the Imitation Game as a methodological frame for interrogating human perceptions of AI-generated texts. As a point of difference, the research investigated higher education academics' capacities to determine the origins (human or AI) of short academic paragraphs relevant to a specific domain of knowledge, namely initial teacher education. The following research questions were addressed:

- RQ1: How readily can academics correctly identify texts written by either ChatGPT or humans?
- RQ2: What are the textual factors that academics perceive as being useful in identifying whether texts are written by ChatGPT or humans?
- RQ3: How does the application of these factors in decision-making align with accurate or inaccurate identification of text authorship?

Methods

Generation of task prompts

The research sought to investigate academics' perceptions of AI-generated texts that encompass domain specific knowledge (i.e., initial teacher education). This was important because texts written in response to prompts typical of initial teacher education discourse will almost certainly have textual features unique to them. Such language conventions are likely to be indicative of a discourse community that includes individuals who share common public goals and mutually understood communicative mechanisms, such as specific genres and lexis (Swales, 2011). To systematically tap into the language of this discourse community, two Australian Professional Standards for Teachers were used as the reference point for task prompt development: (a) create and maintain supportive and safe learning environments, (b) know students and how they learn (Australian Institute for Teaching and School Leadership, 2017).

To elicit text responses that demonstrated higher-order levels of information processing and presentation, the task prompts were developed with reference to the structure of the observed learning outcome (SOLO) taxonomy (Biggs & Collis, 2014). Specifically, each task prompt was mapped to the extended-abstract (e.g., the highest) level of the taxonomy. Future outcomes contingent on hypothetical actions were applied to each task prompt to achieve this objective. To standardise other attributes of the texts that would be generated, specifications around word length (250 words), the use of in-text references (none to be used) and general language conventions (British English) were imposed. The two task prompts used as stimulus for human and ChatGPT-generated texts are shown in Table 2.

Table 2

Task prompts used with humans and ChatGPT

Prompt 1 for Text Pair A (Texts 1 and 2)	Prompt 2 for Text Pair B (Texts 3 and 4)
Argue for the importance of teachers establishing a supportive and safe learning environment for the students in their classrooms. Support your case by presenting some hypothetical future longer-term negative consequences of failing to establish a supportive and safe learning environment for students. Ensure your response is 250 words in length. Please write a single unified paragraph. Do not include in-text citations or references as part of your response. Please use British English language conventions.	Describe the importance of teachers knowing the students in their classrooms and how they learn. Support your case by presenting some hypothetical future longer-term negative consequences of teachers not taking adequate steps to get to know their students and how they learn. Ensure your response is 250 words in length. Please write a single unified paragraph. Do not include in-text citations or references as part of your response. Please use British English language conventions.

Note. Prompt 1 relates to "Create and maintain supportive and safe learning environments" and prompt 2 relates to "Know students and how they learn".

Generation of texts

Four texts were generated in total: two authored by humans and two generated by ChatGPT. Two university-level initial teacher education academics each wrote a response to either prompt 1 or prompt 2. Human-generated texts were written prior to the AI generation of texts. Each task prompt was then entered into the free ChatGPT (version 3.5) interface (https://chat.openai.com) on 30 May 2023. The first text outputted from each text prompt was copied and used as the ChatGPT-generated texts in the current study. No revisions were made. The human-generated and ChatGPT-generated texts are presented as they were used in the data collection phase in the Appendix.

Participants

Sixteen participants (education academics employed within the Australian public university sector) took part in the study. All of them had multiple years of experience assessing written work submitted for academic credit as part of university-level education courses. All possessed a doctorate in education (or equivalent), with relevant professional experience and had sound knowledge of the Australian Professional Standards for Teachers (Australian Institute for Teaching and School Leadership, 2017) that necessarily underpin accredited initial teacher education courses within Australia. The project was granted ethical clearance from the authors' institutional Human Research Ethics Committee (#HE23-089) and thus met the requirements of the National Statement on Ethical Conduct in Human Research. All participants provided recorded informed consent to participate in the research and were given pseudonyms. All pseudonyms were randomly selected with https://thingnames.com/pseudonyms.

Interview protocol

All individual interviews took place via videoconferencing (i.e., Zoom), which enabled the sessions to be recorded for subsequent transcription and analysis. Participants were told that there were two text pairs to be considered, Text Pair A (Text 1 and Text 2) and Text Pair B (Text 3 and Text 4). It was made clear to participants that from each text pair, one text was written by an adult academic and the other was written by ChatGPT (it was not disclosed which was which). Texts were presented to the participants by sharing screens as part of the Zoom session. Text pairs were presented together, but the order within pairs was randomised. All participants read the texts in the same order. Participants were then asked to read each text and afterward present their perceptions about which text they felt was written by a human and which by ChatGPT. Participants were also asked to rate their confidence levels around their decision on a scale from 1 to 5 (i.e., 1 = *not confident at all*; 2 = *somewhat confident*; 3 = *confident*; 4 = *very confident*; 5 = *absolutely certain*). Semi-structured interviewing then sought to elicit participants' perceptions about the textual features that were influential in decision-making about text authorship. Participants were free to change their mind about the origins of the texts and their confidence levels about these decisions. Participants were also free to read texts multiple times and review previous texts as required. Interviews typically lasted between 30 and 40 minutes. Participants were not told the origins of each text until all data had been collected and they were asked to avoid speaking with other participants until all interviews had taken place.

Analysis

To answer the first research question, participants' accuracy in identifying text authorship was assessed. As 16 participants read two text pairs each, there were 32 data points of relevance. Accuracy, levels of confidence in decision-making and instances of mind changing around text authorship were tabulated.

To address the second research question, thematic analysis (Braun & Clarke, 2006) was used to analyse the decision-making discourse within the interview videos and corresponding written transcripts. Given the newness of this research area, an appropriate a priori framework was not available and so a grounded, inductive approach to analysis was adopted. We individually watched and read the interviews multiple times to become familiar with the data, while generating initial codes that reflected recurring textual factors that participants perceived as being influential in their decision-making. We then discussed all themes and established a definition for each. All transcripts were then reanalysed with the qualitative data analysis software NVivo to ensure that the themes covered all instances of decision-making around text authorship. If sections of discourse clearly aligned with multiple themes, multiple codes to that section of discourse were applied.

To address the third research question, the final coding process also entailed coding each instance of decision-making discourse as either aligning with an accurate or inaccurate assessment of text authorship. Accuracy of decision-making was coded based on the participants' position about the authorship of the text at any given time during the interview. For example, if a participant's position was that a text had been written by a human, but it had been written by ChatGPT, the rationales for decision-making provided during this period of time were coded as inaccurate. However, if later in the interview the participant's position about authorship of the text had changed (e.g., ChatGPT-generated), rationales provided from that point on, which supported this position, were coded as accurate.

Results

RQ1: How readily can academics correctly identify texts written by either ChatGPT or humans?

Participants' accuracy in identifying text authorship, levels of confidence in decision-making, and instances of mind changing are presented in Table 3. Participants were accurate in their perceptions of text authorship in 16 out of 32 instances (50% accuracy). Participants' confidence levels on these decisions ranged from *not confident at all* (1) to *absolutely certain* (5), with just one participant indicating that they were *absolutely certain* about authorship on one occasion. Mean accuracy and mean confidence levels for each text pair were very similar. An overall mean confidence level of 2.78 was evident, which is below the scale's midpoint (*confident*, 3). Participants often changed their minds about decisions on authorship (8 times) and their confidence levels about these decisions (8 times).

Table 3
Accuracy, confidence and mind changing

	Text Pair A		Text Pair B	
	Accuracy (1, 0)	Mean confidence (1–5)	Accuracy (1, 0)	Mean confidence (1–5)
Pseudonym				
Alisa	1*	1.75	1	2
Amber	0	1.5	1	2.5
Ava	0	3*	0	1.5
Callum	1*	1.75	1	2.75
Charles	0	3	1	3.5
Eve	1	4.5*	1	5
Gemma	0	2.5	0*	3
Thomas	0	3.5*	1	3.5*
Lewis	1	2	0	2.5
Lizbeth	0	2.5	0	3.5
Lucas	1*	4*	1	4
Marc	0	3*	0	3.25
Marilyn	0*	3.5	0*	1*
Mia	1*	2	1	3
Phoebe	1	3.5	0	1.5
Sienna	1	2.5*	0*	1.5
Mean	0.50	2.8	0.50	2.75

Note. 1 = accurate and 0 = inaccurate; *indicates mind change.

RQ2: What are the textual factors that academics perceive as being useful in identifying whether texts are written by ChatGPT or humans?

Five themes were identified that could categorise the textual factors that academics perceived as being useful in identifying whether texts were written by ChatGPT or humans. The themes occurred in the following descending order of frequency in the data set: voice (41.8%), word usage (24.3%), structure (12.8%), task achievement (11.9%) and flow (9.2%). Themes covered both positive and negative positions articulated by the participants about the texts. Themes are defined and elaborated upon through explanatory quotes from the participants in the subsections below.

Voice

This was the most prevalent theme in the data set constituting 196 instances (41.8%) of the total 469 enumerated. The theme was defined as sections of participant decision-making discourse that referred to the perceived degree of humanness or lack of humanness (bot-ness) of the texts. As the theme name suggests, a recurring commonality across this category of discourse was a sense put forward by the participants of the text conveying the presence or absence of a human voice. This was sometimes manifested as reference to a human actor who was assumed to have written the text:

> Yeah, probably in the first 100 or 150 words or so there is good evidence for me that there's enough personalisation that it's got an individual human perspective. (Marc)

In other instances, the presence or absence of a human voice was inferred due to the perceived clarity or vagueness of the concepts conveyed in the texts:

> Yeah, so I just thought it was very generic at the top and so my experience of chat-bots is that they make very generic statements. (Sienna)

The presence of a human voice in other instances was associated with the perceived appropriateness of the texts in terms of discipline specific expectations:

> The style of writing reflects very much the kind of teaching we put into the academic literacy class. Yeah … I think it's very likely to have been written by an adult academic. (Charles)

Another commonality of the decision-making discourse coded as this theme, although not evident in all instances, was reference to the word “voice” and the phrase “it feels”, or reference to feelings with undertones of some level of emotional connection with the text:

> It feels like it has more of a voice, whereas the other one was a bit more factual. (Alisa)

> I think there's a stronger personal voice in here … There’s an authorial voice. (Gemma)

> I just get a different feeling when I read this one… It feels different, it reads differently. (Thomas)

> The second one again, it sort of has a bit of a generic feel to it. (Lizbeth)

Word usage

This was the second most common theme in the data set making up 114 discrete references at 24.3% of the total. The theme was defined as sections of participant decision-making discourse that referred to language within the texts with a strong focus at the lexical level. This was sometimes identified by participants as the verbatim use of words from the task stem in the texts:

> That's a term that a robot might use because that word was included in the question stem ... the robot has included it because it's part of what's been fed into it, you know, the words that have been fed in. (Lucas)

> And it's using all the words in the actual question [stem]. (Amber)

In other instances, it was a particular word or cluster of words that seemed to strongly influence the decision-making of participants:

> But nobody uses the words 'in … the hypothetical future'. (Alisa)

> The word 'ascertain' … 'Interestingly, students can ascertain when a teacher has not taken … time'. It sounds overly … What's the word? … It sounds unnecessarily jargonistic perhaps. (Sienna)

> That word 'ignite' is not sitting comfortably with me … That word doesn't sit well with me. (Ava)

The presence of explicit cohesive devices in the text – particularly adverbs – was also noted as being influential on participants' perceptions of the text:

> 'Additionally … furthermore … moreover … ultimately … therefore': There's five examples there of sentences which start deliberately. (Lewis)

> You know, 'furthermore … consequently … additionally' those sorts of words … which again you know sort of stand out for me. (Lucas)

> The 'furthermore', the 'additionally', you know, all of that, that's what ChatGPT typically does. (Marilyn)

Structure

There were 60 instances of discourse coded as this theme making up 12.8% of the total. The theme was defined as sections of participant decision-making discourse that referred to structural elements of the texts that were at an organisational level beyond that of individual words. In some instances, this was evident in the participants' decision-making that referred to perceived formulaic structures or patterns within the texts:

> They sort of … follow that formulaic structure. (Lucas)

> It's not wrong, but it's the kind of thing you might expect a chat-bot to do because a chat-bot is following a formula for writing. (Sienna)

> It's well-structured and it's quite sequential in how it presents … So the difference for me is the structure of the sentences. (Marc)

> But that struck me straight away as a pattern. (Lewis)

Another component of the structure theme salient in participants' decision-making discourse was perceived presence of structural repetition within the text:

> There's a reasonable amount of repetition of the idea from one sentence to the next … that makes me feel bot-like. (Ava)

> There just seemed to be a constant repetition for the … first three sentences. (Gemma)

> The sentences are all quite similar actually, in the way that they're structured. (Alisa)

Task achievement

There were 56 instances of discourse coded as this theme making up 11.9% of the total. The theme was defined as decision-making discourse that referred to the perceived degree to which the text satisfied the task. In some instances, this was evident in discourse that referred in general terms to the degree that the text adhered to the perceived scope of the task:

> It's absolutely nailed the brief in that there's no mention of a, you know, it doesn't even go close to citing anybody or anything. It's very much nailed the brief. (Phoebe)

> Well, it answers all the questions, but it also just sounds like it's like a long series of platitudes if you like. (Amber)

> So this text one … I think is a bit tangential and it doesn't really address the question from start to end. (Thomas)

A component of this theme evident in the decision-making discourse related to the provision of examples in the texts. Multiple participants alluded to the presence of adequate examples and elaboration as being important in their decision-making process:

> It was very solid. It was very good. It just lacked … as many specific examples. As [the other text]. (Lizbeth)

> There's more evidence attached in the text …. More examples, more supporting information. (Thomas)

> And it has more examples as required by the task. (Gemma)

Another component of this theme related to the participants' perception of the degree to which the texts adhered to the required language conventions. In some instances, this related to the language requirements stipulated within the task stem (i.e., no in-text citations, British English language conventions and the length of 250 words). In other instances, this discourse related to language conventions implicit to texts of sound quality, namely accurate spelling, grammar and punctuation:

> Hmm. I just feel like a human is more likely to go over the word count. Whereas if you put in a specific order … I feel like … a computer is more likely to be more stringent about those things … I just feel like the computer is more likely to play by the formulas or the algorithmic rules. (Lizbeth)

> It's just poor grammar. ChatGPT wouldn't do that. (Eve)

> The only reason I'm going with this is that sentence in the middle there with all the commas in it. (Lewis)

Flow

This was the least prevalent theme with 43 instances making up 9.2% of the total. The theme was defined as participant decision-making that referred to the perceived degree of cohesion and logical sequencing within the text:

> The cohesion seems forced … it's got a kind of a piecemeal feel. (Ava)

> It sort of moves from one thing to the next so succinctly but it's … odd, mechanical. (Callum)

> This is a very well connected, very cohesive and coherent piece. (Thomas)

Participants used the word “flow” quite often to describe the sense of fluency or cohesion evident within the text:

> There's a certain flow to it. I can imagine some of my students writing like this. There seems to be more coherence … It's easier to read, it flows nicely. (Alisa)
>
> Yeah, so then as I read on I thought … look this flows really well. (Marc)
>
> And the flow, I just have a feeling that [this text] could be ChatGPT. (Mia)
>
> That doesn't seem to flow quite so well. (Phoebe)

RQ3: How does the application of these factors in decision-making align with accurate or inaccurate identification of text authorship?

All five categories of textual features were variously used by participants to rationalise both accurate and inaccurate text authorship decisions. The inner circle of Figure 2 presents the relative total proportion of decision-making categories applied throughout the corpus of interviews. The outer circle of Figure 2 shows the relative proportion of accurate (A) and inaccurate (I) authorship decisions. For example, of the 196 instances of discourse coded as voice, 92 of these represented instances used to rationalise inaccurate decisions about text authorship, which is roughly equivalent to the 104 instances where voice was used to rationalise accurate decisions. Similarly, of the 114 instances of discourse coded as word usage, 55 of these were used to rationalise inaccurate decisions about text authorship and 59 were used to rationalise accurate decisions. This trend continued among the discourse coded as structure, task achievement and flow, with instances of discourse used to rationalise accurate and inaccurate decisions occurring at an approximately equal measure for each (Figure 2).

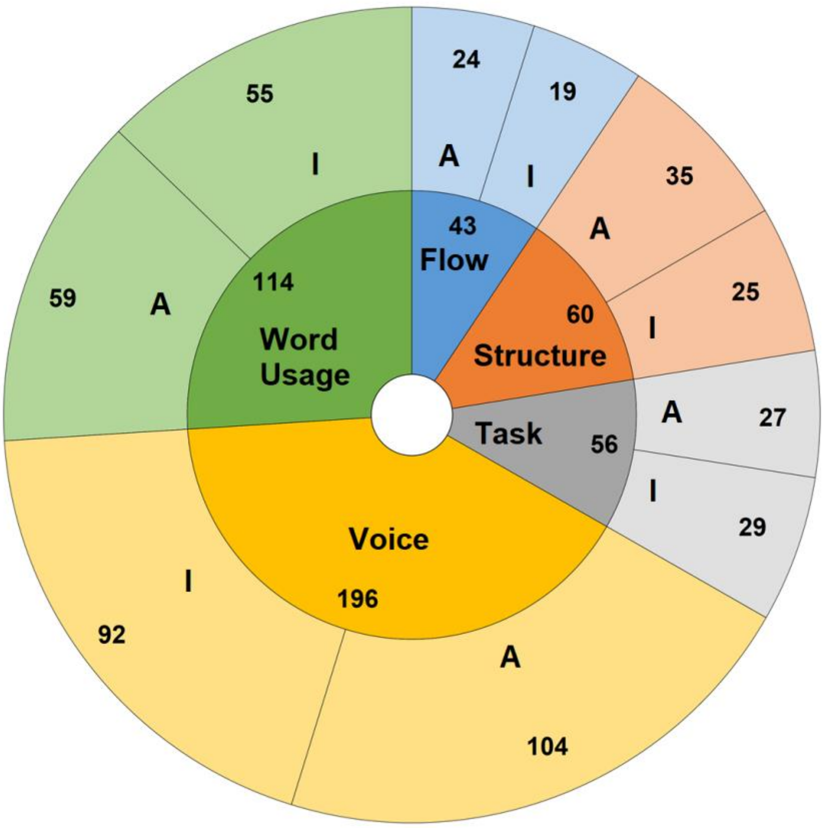

Figure 2. Frequency of authorship decision-making in categories as total, accurate and inaccurate
Note. A = Accurate, I = Inaccurate.

It is noteworthy that each category of decision-making was used for an approximately equal number of accurate and inaccurate authorship decisions. As the practical implications of generative AI are central to the current research, of particular interest were instances where interrogators perceived AI-generated texts as being written by humans. These instances provide strong evidence of the machine winning a decisive point in the Imitation Game. Table 4 provides examples of decision-making where human interrogators erroneously identified ChatGPT-generated text as being human-generated. As is shown in Table 4, there were instances where each of the five themes were used to rationalise these inaccurate decisions.

Table 4
Examples of participants' decision-making (all inaccurate)

Theme	Decision-making rationale
Voice	This one sounds like it's got that human layer over the top of it of you know. What's going to happen without a supportive environment, well 'bullying, discrimination, exclusion' this is going to be 'emotional and psychological' … damaging for students. (Ava, Text 2)
	I think there's a stronger personal voice in here. (Gemma, Text 2)
	To me this seems a lot more conversational, a lot more like reflective so maybe come from a human … and I feel like maybe the second one that's been written by someone who actually understands the school environment a little bit more. Understands the real world a little bit more. So maybe now that I've seen the two, I'm kind of thinking that the second one [Text 2] might have been human written and the first one [Text 1] was ChatGPT. (Lizbeth)
Word usage	So there's a small error there 'that fails to ... and inhibit'-s. There should be s there, 'their academic progress'. So that's a small error which makes me lean at this point towards human. (Phoebe, Text 3)
	Some of the words are quite strong. It's a more forceful argument '… it is imperative for teachers to cultivate an inclusive environment'. The language is stronger here [Text 2] than it was in the last one [Text 1]… Is potentially more human. (Ava)
	I suppose they're making reference to what we would say as jargon, but they use writing in a way that makes that jargon more familiar, more approachable, more accessible … That to me …. So it's talking, it's giving an experience. (Amber, Text 2)
Structure	Yeah, and again, it just feels like it has that more structured approach where the argument is consistent through the paragraph rather than jumping from topic to topic. (Ava, Text 3)
	It's structured in a way that it's a holistic paragraph … it's in a clear and logical fashion [Text 2], more so than the one before [Text 1] which was using … more jargon but in a less directed way. (Amber)
Task achievement	Hmm, okay. But that this one actually better addresses the question too. Yeah, I think so. (Gemma, Text 2)
	I find that is much more detailed and very well addressed in text two accurately, rather than in text one. (Thomas)
	It unpacks like the relational aspects more. So a strong student teacher connection. 'Students may feel disconnected, undervalued … unheard, which can' then 'impact their self-esteem and overall wellbeing'. (Lizbeth, Text 3)
Flow	This is a very well … a very well connected, very cohesive and coherent piece with a lot of supporting information. (Thomas, Text 2)
	It has … a flow that feels more natural. (Ava, Text 2)
	I thought this one flowed well [Text 2], but the first one [Text 1] was a bit more methodical and robotic. (Marc)

An unexpected finding from the research arose during the interviews where several participants began to reflect on their own capabilities to determine whether the texts were written by AI or a human. Some of the participants who came into the interview feeling quite confident about their own capabilities to make this determination began to feel increasingly less so following interrogation of Text Pair A. A sentiment that shone through from participants was that there was uncertainty about the exact capabilities of AI and this triggered concerns about the disruptive potential of AI in higher education:

> It's interesting because it raises things in my own mind about whether or not … how well I know AI. (Callum)
>
> This [is] much harder than I thought it was going to be. But maybe that's an indication of the challenge we're up against as academics. (Marc)
>
> Any other comments on anything with these two [texts]? (Interviewer)
> Deep concern because if I was reading these and I was marking … I wouldn't pick either of these as ChatGPT. (Lewis)
>
> So, yeah, I mean, imagine if all of your students or 50% of them were submitting their assessments through ChatGPT ... it could be very difficult to pick it up. (Lizbeth)

It is clear from the above quotes that playing the Imitation Game offered the participants the opportunity to reflect on the current challenges faced in higher education as a result of generative AI.

Discussion

Over 70 years ago, Turing (1950) predicted that by the year 2000 it would be "possible to programme computers … to make them play the imitation game so well that an average interrogator will not have more than 70 percent chance of making the right identification after five minutes of questioning" (p. 442). The results of the current study support the accuracy of Turing's incredible prediction. With technologies such as ChatGPT broadly available, it is difficult for interrogators to consistently win the Imitation Game. As suggested by the current study, academics are becoming increasingly aware of the challenges they face in relation to assessment as a result of generative AI. Participants found the accurate identification of authorship challenging and were forthright about their relatively low levels of confidence while undertaking the task. Despite having only two alternative texts to consider, interrogators had a combined accuracy of just 50%. It seems that ChatGPT was the likely winner of the Imitation Game in this instance. Analysis showed that there were textual features perceived as important in determining text authorship but that these were used almost equally to rationalise both accurate and inaccurate decisions.

The chance of winning the Imitation Game being equivalent to a coin flip raises a pressing academic integrity related question for the higher education sector. If humans cannot accurately identify AI-generated texts, what can be done? Perhaps a reasonable response is to depend less on human judgements and more strongly on those offered by AI algorithms, such as AI classifiers that are trained to differentiate AI and non-AI generated text. However, evidence suggests that the performance of AI classifiers is highly variable. In July 2023, OpenAI, the research organisation that created ChatGPT, shuttered its AI classifier designed to distinguish between human-generated and AI-generated text, citing low rates of accuracy (Kirchner et al., 2023). Recent research has also questioned the performance of various commercial AI classifiers. Of some concern is that AI classifiers may be particularly inconsistent when applied to human-generated text (i.e., uncertain classification and false positives) (Elkhatat et al., 2023). In light of this, if AI classifiers are a component of an institutional-wide academic integrity strategy, academics must be aware that AI classifiers may not always be accurate and that false positives may occur. It is critical that we avoid erroneously accusing and/or penalising students for academic misconduct, as the negative impacts on student well-being and institutional reputation of doing so will be severe. At the same time, academics must also be aware that AI technology is continually evolving. Research suggests that text produced by earlier versions of generative AI (i.e., ChatGPT 3.5) is more accurately classified as

AI-generated than text produced by more sophisticated versions (i.e., ChatGPT 4) (Elkhatat et al., 2023). As generative AI becomes more sophisticated in the future, it will become progressively more difficult to identify AI-generated text with commercial AI classifiers — false negatives may proliferate. Taken together, the evidence suggests that the affordances of AI classifiers are certainly not a fool-proof answer to the questions raised about human difficulties with the Imitation Game.

As it may soon be close to impossible to reliably detect AI use in text-based assessments, establishing a clear line of communication with stakeholders about the use of AI is crucial. Although certainly not the only measure that will be required to address the challenges of AI in higher education, a starting point is to ensure that all assessments have a pedagogical and ethical rationale – an *AI use descriptor* – that makes clear the degree to which it is acceptable to use AI in the completion of a task. For example, if a particular assessment should be completed without any direct use of generative AI technology, the AI use descriptor should make this explicitly clear and provide a pedagogical and ethical rationale for this position. Such descriptors will also be a mechanism that will enable, when appropriate, the definition of parameters of allowable use of AI in the completion of assessments. For example, it may be pedagogically and ethically sound to use generative AI for completing some parts of an assessment task, but perhaps not for others. In other cases, there may be sound reasons to apply no restriction on use of AI in the completion of assessment tasks. Regardless of the amount of AI use that is deemed appropriate, this position should be made explicitly clear and rationalised to students. It is important for graduates of higher education institutions to engage with conceptual frameworks that assist them to negotiate the effective and ethical use of AI, both at university and in the workforce. The presentation of a clearly justified AI use rationale with all assessment tasks represents a feasible mechanism to immediately begin modelling such frameworks.

As AI technology is evolving rapidly, academics need to have ample opportunity to regularly update their familiarity with the capabilities of contemporary AI. For example, regularly entering assessment tasks into generative AI technology will assist academics to make informed decisions about optimising assessment design. It is vital to build a collective awareness of the capabilities of AI and discuss the implications of these capabilities within and across work units and courses (e.g., departments, schools). Academics will also need to develop a clear sense of how to teach their students about effective and ethical uses of AI in their respective fields and this is likely to require considerable upskilling across a wide range of disciplines. Tertiary institutions will need to support academics on how best to redevelop assessments. With direct relevance to this matter, Lodge, Howard et al. (2023) have offered a valuable set of guiding principles that can be used as a compass to guide directions for assessment design in the age of AI. The principles are offered on the premise that institutions must address the use of AI technology in teaching. It cannot be ignored. Higher education assessments should be redesigned in consideration of the ways in which AI will increasingly become a part of our everyday lives. Assessments must engage with AI (discipline dependent) and follow a systemic approach across courses and degrees (CRADLEdeakin, 2023a, 2023b; Lodge, Howard et al., 2023).

Limitations and future research

There are several limitations of the current research, the identification of which may improve and guide future related work. Firstly, due to issues relating to feasibility, texts of only a relatively limited scope and diversity were applied in the current research. For example, only relatively short texts were used. It is possible that longer and more complex texts may have been more challenging for ChatGPT to produce and may have made the differentiation between human and ChatGPT-generated texts less demanding for interrogators. Several participants indicated that one of the factors they typically used to determine whether a text has been written by ChatGPT is by examining the in-text references. Although we were aware that ChatGPT is known to use inaccurate references, we felt that excluding references would require participants to focus more directly on the textual features that were important in decision-making, the focus of the current research. We acknowledge, however, that the inclusion of references may have assisted the participants in making more accurate determinations. Another possible limitation to the current research is that experienced academics, rather than education students, produced the human-generated text used. It could be argued that texts written by less expert authors (e.g., students) may have produced different results. Further, the study could have been enhanced by producing multiple AI-generated texts using the same prompt and thus broadening the pool of texts that the academic participants were asked to interrogate. For example, offering participants three texts, only one of which was written by a human, may have altered the decision-making dynamics that were evident.

Another basic limitation of the study was that it was undertaken among a relatively small sample (N = 16) and synthesising information for a larger group may have yielded different results. Although we are confident that the qualitative data gathered as part of the current study is rich and informative, replication or partial replication with larger numbers of participants is an interesting future possibility. Further, applying the Imitation Game paradigm with texts specific to other discipline areas among academics with various areas of expertise may also be of interest. Finally, it is important to acknowledge that the freely available ChatGPT (version 3.5) was used to produce the AI-generated texts used in the current study. Since this research was conducted, version 3.5 has been superseded by version 4. Replicating the current study and using more sophisticated iterations of generative AI to produce the text may result in different decision-making among human interrogators. This final limitation is emblematic of the challenges academics will continue to experience while seeking to orient their assessment practices in a rapidly changing technological landscape.

Conclusion

The current research is the first that we are aware of, since the emergence of ChatGPT, that has focused on the textual factors that influence education academics' decision-making processes around text authorship (AI or human-generated). This is a fertile area for future investigations, especially with the fast advancement of AI technology. The five themes (voice, word usage, structure, task achievement and flow) evident in the discourse provide a useful starting point for understanding the decision-making interrogators engage in when deciding the humanness or bot-ness of a text. As voice emerged as the most prevalent theme, it is clear that an important part of this process for interrogators was searching for the presence or absence of a human voice in the text. The voices underpinning both accurate and inaccurate text authorship decisions speak to the importance of readers connecting with texts through a sense of humanness within it. The research makes clear to us that textual humanness is difficult to define and is becoming very difficult to differentiate from text generated by AI. Indeed, we concede that this situation is concerning. We have heard the concerns of academics and in response highlight the pressing need for institutions to redouble efforts to interrogate the adequacy of their assessment design, AI use policies and AI-related procedures.

Author contributions

Author 1: Conceptualisation, Investigation, Data collection, Formal analysis, Writing – original draft, review and editing; Author 2: Investigation, Data collection, Formal analysis, Writing – review and editing.

Acknowledgments

The authors sincerely thank the reviewers for their feedback, which greatly improved the quality of the current work. Acknowledgement for the graphic design of Figure 1 goes to Stella Matthews.

References

Australian Institute for Teaching and School Leadership. (2017). *Australian professional standards for teachers*. https://www.aitsl.edu.au/standards

Biggs, J. B., & Collis, K. F. (2014). Evaluating the quality of learning: The SOLO taxonomy (structure of the observed learning outcome). Academic Press. https://doi.org/10.1016/C2013-0-10375-3

Block, N. (1981). Psychologism and behaviourism. *The Philosophical Review*, *90*(1), 5–43. https://www.jstor.org/stable/2184371

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa
Chomsky, N. (2008). Turing on the imitation game. In R. Epstein, G. Roberts, & G. Beber (Eds.), *Parsing the Turing test: Philosophical and methodological issues in the quest for the thinking computer* (pp. 103–106). Springer Science + Business Media B.V.
CRADLEdeakin. (2023a, March 2). *ChatGPT Webinar #2: How should educators respond?* [Video]. YouTube. https://youtu.be/V57mT1hrc24?si=uCrXrdj7-7H0VR4I
CRADLEdeakin. (2023b, October 4). *ChatGPT Webinar #5: Assessment reform for the Age of Artificial Intelligence* [Video]. YouTube. https://youtu.be/8F5jZrFbjJY?si=fWknI8TOQzUBKpt8
Dawson, P. (2021). *Defending assessment security in a digital world: Preventing e-cheating and supporting academic integrity in higher education*. Routledge.
Elkhatat, A. M., Elsaid, K., & Almeer, S. (2023). Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. *International Journal for Educational Integrity*, *19*, Article 17. https://doi.org/10.1007/s40979-023-00140-5
Elkins, K., & Chun, J. (2020). Can GPT-3 pass a writer's Turing test? *Journal of Cultural Analytics*, *5*(2). https://doi.org/10.22148/001c.17212
Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, *30*, 681–694. https://doi.org/10.1007/s11023-020-09548-1
French, R. M. (2000). The Turing test: The first 50 years. *Trends in Cognitive Sciences*, *4*, 115–122. https://doi.org/10.1016/s1364-6613(00)01453-4
Gao, C. A, Howard, F. M., Markov, N. S., Dyer, E. C., Ramesh, S., Luo, Y. & Pearson, A. T. (2023) Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *npj Digital Medicine 6*, Article 75. https://doi.org/10.1038/s41746-023-00819-6
Harnad, S. (1992). The Turing Test is not a trick: Turing indistinguishability is a scientific criterion. *ACM SIGART Bulletin*, *3*(4), 9–10.
Hulman, A., Dollerup, O. L., Mortensen, J. F., Fenech, M., Norman, K., Støvring, H., Troels & Hansen, T. K. (2023). ChatGPT- versus human-generated answers to frequently asked questions about diabetes: A Turing test-inspired survey among employees of a Danish diabetes center. *medRxiv.* https://doi.org/10.1101/2023.02.13.23285745
Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pferrer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274
Kirchner, J. H., Ahmad, L., Aaronson, S., & Leike, J. (2023). *New AI classifier for indicating AI-written text*. OpenAI. https://openai.com/blog/new-ai-classifier-for-indicating-ai-written-text
Lodge, J. M., Howard, S., Bearman, M., & Dawson, P. (2023). *Assessment reform for the age of artificial intelligence*. Tertiary Education Quality and Standards Agency. https://www.teqsa.gov.au/sites/default/files/2023-09/assessment-reform-age-artificial-intelligence-discussion-paper.pdf
Lodge, J. M., Thompson, K., & Corrin, L. (2023). Mapping out a research agenda for generative artificial intelligence in tertiary education. *Australasian Journal of Educational Technology*, *39*(1), 1–8. https://doi.org/10.14742/ajet.8695
Nov, O., Singh, N., & Mann, D. M. (2023). Putting ChatGPT's medical advice to the (Turing) test. *medRxiv*. https://doi.org/10.1101/2023.01.23.23284735
Ouyang, F., Zheng, L., & Jiao, P. (2022). Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020. *Education and Information Technologies*, *27*(6), 7893–7925. https://doi.org/10.1007/s10639-022-10925-9
Popenici, S. A., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, *12*(1), 1–13. https://doi.org/10.1186/s41039-017-0062-8
Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching*, *6*(1). https://doi.org/10.37074/jalt.2023.6.1.9

Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning & Teaching*, *6*(1), 1–10. https://doi.org/10.37074/jalt.2023.6.1.17
Swales, J. (2011). The concept of discourse community. In E. Wardle & D. Downs (Eds.), *Writing about writing: A college reader* (2nd ed., pp. 215–229). Bedford/St. Martins.
Turing, A., (1950). Computing machinery and intelligence. *Mind*, *59*(236), 433–60. https://doi.org/10.1093/mind/LIX.236.433
Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators? *International Journal of Educational Technology in Higher Education*, *16*, Article 39. https://doi.org/10.1186/s41239-019-0171-0

Corresponding author: Joshua Matthews, jmatth28@une.edu.au

Copyright: Articles published in the *Australasian Journal of Educational Technology* (AJET) are available under Creative Commons Attribution Non-Commercial No Derivatives Licence (CC BY-NC-ND 4.0). Authors retain copyright in their work and grant AJET right of first publication under CC BY-NC-ND 4.0.

Please cite as: Matthews, J., & Volpe, C. R. (2023). Academics' perceptions of ChatGPT-generated written outputs: A practical application of Turing's Imitation Game. *Australasian Journal of Educational Technology*, *39*(5), 82–100. https://doi.org/10.14742/ajet.8896

Appendix

Text Pair A (Text 1 & Text 2)

1	2
ChatGPT-produced?	Human-produced?

1	2	3	4	5
Not confident at all	Somewhat confident	Confident	Very confident	Absolutely certain

Text 1

A crucial starting point for effective classroom teaching is the establishment of a safe and supportive environment for all students. Without the assurance of feeling physically and emotionally safe, students will be unlikely to learn optimally or may even not learn at all. As Maslow's classic hierarchy of needs makes clear, needs such as the need for safety, friendship, and a sense of connection are more fundamental than the need for self-actualisation. In essence, a student must have their more fundamental needs met before higher-order needs, such as the need for self-improvement through learning, can be engaged with to their fullest extent. For example, teachers should proactively work toward establishing a classroom environment within which students feel comfortable asking questions, talking to one another, making mistakes and taking calculated risks. In the absence of such an environment, students may systematically miss opportunities to develop basic skills and domains of knowledge. This in turn can have longer-term consequences on cognitive development. Further, a supportive and safe learning environment classroom is conducive to the development of valuable interpersonal skills, which mirror many of the skills that are indispensable in professional workplaces. Additionally, the failure to establish a safe and supportive learning environment may have detrimental impacts upon vulnerable students' emotional wellbeing. Such impacts should be at the forefront of teachers' minds not only because of the resultant immediate impact on students' capacity to learn but also because of the potential longer-term health impacts, such as anxiety, depression and other psycho-social health considerations.

Text 2

The establishment of a supportive and safe learning environment by teachers is crucial for the holistic development of students. Failing to create such an environment can have detrimental consequences in the hypothetical future. Without a supportive atmosphere, students may feel alienated, leading to a decline in their self-esteem and overall well-being. This can negatively impact their motivation to learn and participate actively in the classroom. Additionally, in the absence of a safe learning environment, students may experience heightened levels of stress and anxiety, impeding their ability to concentrate and engage effectively with the curriculum. This can result in decreased academic performance and hinder their long-term educational outcomes. Furthermore, without a supportive and safe environment, instances of bullying, discrimination, and exclusion may prevail, contributing to a hostile and unhealthy school culture. This can lead to long-lasting emotional and psychological effects on students, impairing their social development and interpersonal relationships. Moreover, in the absence of a supportive learning environment, students may be hesitant to seek help or ask questions, hindering their ability to clarify doubts and deepen their understanding of the subjects. This lack of support may perpetuate educational inequalities, as students from disadvantaged backgrounds may face additional barriers to learning. Ultimately, the failure to establish a supportive and safe learning environment can impede students' overall growth, hinder their academic progress, and limit their future opportunities. Therefore, it is imperative for teachers to cultivate an inclusive, nurturing, and safe space for students, promoting their well-being, fostering positive relationships, and enhancing their educational journey.

Text Pair B (Text 3 & Text 4)

Text 3
Understanding the students in their classrooms and how they learn is of paramount importance for teachers. Failing to take adequate steps to get to know their students can have significant negative consequences in the hypothetical future. When teachers lack knowledge about their students' individual learning styles, strengths, and challenges, they may struggle to tailor their instructional methods to meet the diverse needs of the class. This can lead to a one-size-fits-all approach that fails to engage students effectively and inhibit their academic progress. Furthermore, without an understanding of their students' interests, backgrounds, and motivations, teachers may struggle to create a meaningful and relevant learning experience. This can result in disengagement, apathy, and decreased student participation, leading to a decline in educational outcomes and hindered intellectual development. In the absence of a strong teacher-student connection, students may feel disconnected, undervalued, and unheard, which can negatively impact their self-esteem and overall well-being. This lack of personal connection may also make it difficult for teachers to identify and address social, emotional, and behavioural issues that students may face. Consequently, students may not receive the necessary support and intervention, potentially exacerbating their challenges and impeding their holistic development. Additionally, a failure to understand their students can lead to missed opportunities for fostering a positive classroom climate and promoting a sense of belonging. This can contribute to a negative and unsupportive school culture, hindering students' engagement, collaboration, and overall educational experience. Therefore, it is crucial for teachers to invest time and effort in getting to know their students on an individual level, as it allows them to create a tailored and inclusive learning environment that supports each student's unique needs, maximises their potential, and nurtures their overall growth.

Text 4
A positive relationship between teachers and students is the key to successful teaching and learning. A way to ignite positive relationships between teachers and students is for teachers to ensure they know their students and how they learn. Teachers can learn more about their students through engagement in learning activities where students can reveal their preferred learning styles to their teachers. Students whose teachers understand their learning styles have the best opportunities for success. In addition, when lessons are tailored in consideration of students’ needs, this can result in the establishment of a positive learning environment. The results can be quite detrimental to the learning of all students in the classroom if teachers do not take the time to get to know their students. Major behavioural issues in the classroom are a potential consequence for this lack of attention. Interestingly, students can ascertain when a teacher has not taken the time to learn more about them, resulting in those students potentially misbehaving to receive more attention. Difficult behaviours in the classroom can make teaching and learning challenging, as students who act out inappropriately can cause increased stress for teachers. Thus, in order to foster positive relationships between students and teachers and to ensure all students have the best chance for success in their education, it is pertinent for all teachers to focus on the relationships between them and their students through taking the time to get to know their students and how they learn.