"I, too", shall have to prompt? A study of EFL students and their unmonitored use of GenAI in the completion of an imitation task in poetry

Guilène Révauger

Guilène Révauger est maîtresse de conférences en études du monde anglophone. Ses recherches portent notamment sur l'enseignement et l'apprentissage de l'anglais dans un environnement multilingue, l'ingénierie de formation, la citoyenneté et la littératie numériques.
Affiliation : Institut national supérieur du professorat et de l'éducation, université de La Réunion, laboratoire DIRE.
Courriel : guilene.revauger@univ-reunion.fr
Toile : http://revauger.net
Adresse : INSPÉ La Réunion, Allée des Aigues marines, 97487 Saint Denis, La Réunion.

Flavie Rident

Flavie Rident est enseignante agrégée d'anglais, spécialiste de littérature britannique, elle enseigne au département LSH de l'Université de Guyane et co-dirige le Centre de Ressources en Langues. Son intérêt porte sur les feedbacks et la remédiation dans des dispositifs de formation en présentiel enrichi et plus largement sur les dispositifs d'accompagnement numériques.
Affiliation : université de Guyane.
Courriel : flavie.rident@univ-guyane.fr
Adresse : DFR LSH, université de Guyane, Campus Troubiran, 97300 Cayenne.

This study questions the extent to which a generative artificial intelligence (GenAI) model can be of use within an English as a Foreign Language Literature course at university level. An experimental GenAI-assisted creative writing exercise was conducted, asking students to perform the imitation task of writing a poem in the style of Langston Hughes. The analysis of qualitative data collected through surveys and observational studies provides answers to the main research question and three sub-questions raised: To what extent does the exercise provide for the explicit application of the literary analysis elements taught in the course? What are the characteristics of the prompts intuitively generated by the students? Do students intuitively manage to generate

efficient prompts? What pedagogical implications can be inferred from the results? Although the exercise exposed participants to a large quantity of literary terms fitted to their individual needs, it also shed light on a number of limitations and complications.

"'I, too', shall have to prompt?" Une étude des pratiques libres des étudiants en Anglais Langue Étrangère recourant à une IA générative pour réaliser un exercice d'imitation poétique.

Cette étude porte sur le bien-fondé du recours à un modèle d'intelligence artificielle générative (GenAI) dans le cadre d'un cours universitaire de littérature en anglais langue étrangère. Un exercice expérimental d'écriture créative assistée d'une IA générative a été mené, invitant les étudiants à écrire un poème à la manière de Langston Hughes. L'analyse des données qualitatives recueillies par le biais d'enquêtes et d'observations apporte des éléments de réponse à la principale question de recherche et aux trois questions annexes soulevées : Dans quelle mesure l'exercice permet-il une mise en pratique explicite des éléments d'analyse littéraire enseignés dans le cadre du cours ? Quelles sont les caractéristiques des requêtes générées intuitivement par les étudiants ? Les étudiants parviennent-ils intuitivement à générer des requêtes performantes ? Quelles implications pédagogiques peuvent être déduites des résultats ? Bien que l'exercice ait exposé les étudiants à une grande quantité de termes littéraires répondant à leurs besoins individuels, il a également mis en lumière un certain nombre de limites et de difficultés.

1. Introduction

Generative artificial intelligence models (GenAI) have entered the educational world breaking the door open, disrupting the class, and inviting teachers to rethink their methods, and learners to find or imagine new task strategies. They offer possibilities that can be time-saving, or facilitating learning by generating questions from a text for example, or rephrasing an explanation. These tools are based on machine learning, deep learning and Large Language Models (LLM) capable of human-like text generation. But are they as intuitive as non-AI experts may imagine them to be? How will our students proceed with these GenAIs? They are already using Artificial Intelligence (AI) tools, but the scope of GenAIs is much wider in form and in content than most AIs they are used to. For this research, it was decided to investigate how students would use OpenAI's chatbot ChatGPT-3.5 if left the possibility to do so but without guidance. The assignment chosen is a "writing in the style of" imitation task as, up to today, this assignment is done unmonitored at home and students can use the tools they wish. The LLMs will be or are part of these tools, and ChatGPT could seem like a perfect assistant

for this kind of imitation task. It will allow the study of students' capacity to transfer their disciplinary skills in this new context and humbly establish a picture of where we stand.

2. Art made easy? Creating, imitating and prompting

2.1. AI Poetry

The debate over the use of AI in art divides artists and critics. Some seem to claim the use of AI corrupts art and the generated work cannot be considered as artistic: "art is a uniquely human endeavour, using perspectives from philosophies of art and aesthetics" (Jiang et al., 2023 p. 363). However, when delving into the debate, one may realise that it is not the presence of AI in art that is at stake, but the conditions of its use, as for Jiang et al. (2023), image generators might be used if they do not involve pillaging existing artistic projects, and do not supersede human-made works.

The question then is how much original matter does the artist provide in the input? To what extent does he or she intervene on the output? To which degree is the human in-the-loop or out-of-the-loop? This stake of human in or out-of-the-loop raises many questions, especially in the field of AI and accountability (Goldenfein, 2019), and it needs to be broached when dealing with AI and art. Firstly, we humans appear to have a strong bias against AI art (Millet et al. 2023). Secondly, although Hitsuwari et al. (2023) tested discrimination on haiku and concluded that humans did successfully discriminate human made from machine made, other studies had previously concluded that we humans, were unable to tell the difference between artwork made by a human, and artwork made by an AI with a human out-of-the-loop process (Kobïs & Mossink, 2021). So human intervention is needed to improve the output, or to select the best one, if one wishes for a successful collaboration with the AI. For Hitsuwari et al. (2023), collaboration between the machine and humans delivered results considered as more beautiful.

If we admit that humans need to remain in the loop to maintain authenticity in art, both in the output and in the input, there is nonetheless the challenge of making the LLM produce the wanted piece of art. Is the LLM capable of generating any piece of art?

It seems that the less coded the type of art is, the more difficult it becomes for the LLM to produce what is demanded. This point was raised by Cheng et al. (2018), who did question the ability of AIs to generate modern Chinese poems, in which creativity and imagination prevailed over the strict rules of ancient Chinese poetry. However, Badura et al. (2022) contradict this observation as they did not observe that AI generated Shakespearean sonnets were more likely to be considered human made than AI generated free verse in the style of Walt Whitman, or that the machine performed better in the generation of poems with strict formal codes, especially length of line, cadence, and rhyme patterns. However, the samples for this study were very short.

2.1.1. Teaching poetry and literature at university level

In higher education, the aim of literature classes is to introduce works and authors believed to be fundamental in the history of literature, but also to develop cultural insights and tackle artistic perspectives on specific times or periods, so as to highlight their articulations and specificities. It allows to develop skills in stylistics and stylistic analysis, and is used as "a resource to develop linguistic competence" (Zaro Vera, 1991, p. 169). The study of prose and poetry also allows to analyse the complexity of language in its symbolic content, and the connotations carried by a word, the variations in interpretations according to context, or the reader's personal experience. These disciplinary skills are developed and assessed thanks to traditional exercises, such as comparing and contrasting texts, essays, text commentaries or imitation of an author. They are also used to develop creative skills.

One of the traditional assignments used in the literature class is "Writing in the style of"; for this activity, students are meant to reproduce a piece of writing reflecting the style of a particular author; when presenting their work, they also need to justify their choice of theme, structure, cadence and imagery. According to Uwe Geist, "Imitation is closely connected with observation and analysis" and students are taught to "supplement the personal-intuitive attitude with a technical-problem-solving one" (Geist, 2005, p. 172). Imitation thus demands that the imitator analyses thoroughly the material to be imitated, and becomes aware of how it functions, aware of the style.

Style is understood as the particular element that bridges form and content, as it "operates from the level of small mechanical units, like diction and punctuation, through the broadest conceptual levels, like content" (Stodola, 2003, p. 59). It entails "variations in the presentation of this information [the content] that serve to alter its 'aesthetic quality' or the reader's emotional response" (Abrams & Harpman, 2012, p. 386). For Rifaterre, it is "an emphasis (expressive, affective or aesthetic) added to the information conveyed by the linguistic structure, without alteration of meaning" (Rifaterre, 1959, p. 155). It may be understood as the specific use of literary means, vocabulary or syntax adopted by an author to express an idea. But it is further "the way in which language is used in a given context, by a given person, for a given purpose, and so on" (Leech & Short, 1981, p. 10). Style is not then dependent solely on the author, but also on context and other such mutable elements. Style is far from monolithic both in its definition and in its occurrence, yet its identification remains a most relevant entry into the works of an author, for the very fact that in spite of its eluding definition, it remains a writer's signature (Compagnon, 1997).

"Writing in the style of" may be seen as too constraining or too formal, yet it involves an accurate understanding of the original work through detecting and analysing elements of style before being able to imitate the original work. It allows students to be creative, but also to use their cultural, stylistic, and linguistic competence. "Writing in the style of" is a task that can only come towards the end of a course, as it demands sharp mastery of devices, vocabulary, syntax, and stylistics components, but also a thorough knowledge of the author's work and the capacity to implement a critical strategy.

It is very likely that this type of assignment will undergo changes with the introduction of GenAIs as tools for our students, but will it make it useless? It seems that it may shift the focus,

from reproducing the work itself to the critical approach students may have towards this imitation. Assuming that our students will test ChatGPT or other LLMs for this personal creation, we have questioned their ability to use this GenAI to the best of their purpose. Our aim is to check their mastering of this new tool, but also to envisage new activities integrating LLMs to improve their understanding of literature.

2.2. Conversing with a generative bot and prompt engineering

Conversing with a chatbot as a customer might be the most frequent first interaction one has had up to today. Studies have shown that anthropomorphic chatbots have a clear impact on the emotions of the people interacting (Crolic et al., 2022), especially because they align with the human interacting as if the conversation was natural. However, studies have also shown that humans reverted to clear commands under nominal clauses when confronted with bot mistakes (Spillner & Wenig, 2021). So the form of the interactions changes, sentences become shorter, the vocabulary used is not as rich as in human conversation (Hill et al., 2015), and there is not much attention paid to spelling or grammar (Park et al., 2018). We agree that GenAIs are not simple chatbots but nevertheless, previous interactions students have had may influence the way in which they will consider prompting, prompting being "a natural language instruction that tells a large language model (LLM) to perform a task" (Learn Microsoft, 2024). Consequently, prompting aims at initiating and shaping the output produced by the AI, it gives context, and details the expected output (White et al., 2023). Prompt engineering has thus become a new research area in the field of Human-Computer-Interaction (Korzynski et al., 2023).

2.2.1. The interaction overall processes

Users may have different types of conversations with GenAIs, depending on the format of the desired output (text or image). In the matter of texts, users may request an answer to a question, a summary, a classification, or content generation (Korzynski et al., 2023); their prompt needs to shape this output in form and content, and there are different ways of giving instructions to steer outputs, from keywords, brief search queries, to longer expanding conversations comprising numerous prompts and query refinements (Budiu et al., 2023b). The span of possibilities in the output intended is vast and, even if we limit the intention of the user to the area of writing fiction, the uses which are made vary very much from one user to another. Depending on the person's expectations on the capacity of the AI, they will use it as "an active idea generator, a co-writer, or a writing assistant" (Yang et al., 2022, p. 63). In all of these options, the GenAI comes in to support, if not emulate, human creativity. However, this use may also divert the writer's attention from his or her own creative activity onto curiosity about what the AI can do or trying to "school" the AI (Ghajarghar et al., 2022). The novelty of the tool, natural curiosity, and the speed of response might quickly take the user away from their first intention, their first prompt.

To avoid wandering in prompting, a clear plan must be laid out with the students, prior to the interaction, so that they can make informed decisions during their conversation (Woo et al., 2023). However, it needs to be reminded that our purpose in this study is to evaluate students' ability to interact with a GenAI uninformed, and to investigate their prompts.

2.2.2. Prompt definition and taxonomy

According to Cambridge dictionary online, a *prompt* is defined as:

> (computer)
> - a sign on a computer screen that shows that the computer is ready to receive your instructions
> - an instruction given to an artificial intelligence (= a computer system or machine that has some of the qualities that a human brain has, such as the ability to interpret and produce language in a way that seems human, recognize or create images, solve problems, and learn from data supplied to it) by a human using natural language rather than computer language (…)
>
> (actor's help)
> - words that are spoken to an actor who has forgotten what he or she is going to say during the performance of a play
> - a person whose job is to help actors, during a performance, to remember words that they have forgotten (Cambridge Dictionary)

The term stems from the French *prompt* (quick, swift), which in turn comes from the Latin *promptus* (visible, evident, brought to light). Used as a verb, it is a synonym of "to cause", "to make something happen" (Cambridge Dictionary). Despite initial meanings which may suggest swiftness, visibility, and cause, in prompt engineering language, not all prompts aim for an output. The prompt can also be considered merely as an instruction, the textual input given to a GenAI.

Henceforth users may resort to prompts which are utterances deprived of apparent specific expectations ("Thank you"; "Have a good day", etc.). Rather than asking for the generation of a product, these prompts recall the socio-interactional conversations held between humans. Budiu et al. (2023a) identify these utterances as "filler prompts" and consider them as signs of anthropomorphic language.

Looking into the taxonomy of prompts has been fundamental to better structure the research planned, no standardised taxonomy is used at present in the area of prompting and research, however, it is called for (Santu & Feng, 2023). Studies show that prompting is still a question of trial and error (Dang et al., 2022; Zamfirescu-Pereira et al., 2023); strategies need to be developed (Ekin, 2023) as no specific strategy is developed by mainstream users, although fixed patterns or templates have been issued by researchers, for instance the text to image generator (Oppenlaender et al., 2023). Prompting remains an area to be explored, and a more thorough investigation can be done, be it in refining patterns but also in innovating perspectives on uses (White et al., 2023).

Braun has established a taxonomy encompassing three meta-dimensions: interaction, context, and outcome. "Interaction" relates to the communication process and shaping, it integrates the format of the input and output as well as the CITL (Computer-in-the-loop) or HITL (Human-in-the-loop) criteria. "Context" deals with the information fed into the computer in terms of style expected, role of machine and examples provided. Zero-shot prompts (or direct prompting) provide a simple instruction with no extra guidance or details; one shot prompts provide the model with an example of the output expected; few-shot prompts are used for more complex tasks. Few-shot prompting may also be combined with chain of thought prompting, which consists in providing intermediary reasoning steps (Wei et al., 2022). The last meta dimension, "Outcome", incorporates both goal and the way the prompts have been fed into the machine, in one single prompt, or following a step-by-step process, the latter being used for complex tasks. This taxonomy allowed this research to retain general criteria and guidelines as regards relevant information to be found.

Oppenlaender et al. (2023) have developed a more specific taxonomy, focusing on prompt modifiers in text to image generation. They have defined six prompt modifiers with different purposes: the "subject term" used to define the subject, the "style modifier" used to reproduce a particular style (be it through mentioning the materials used or a school), the "image prompt" (that introduces an image through one or several urls), the "quality booster" (that is generally an added adjective meant to enhance the image), the "repeating term" (meant to solidify the intended output through the repetition of a central term), the "magic term" (an abstract concept or an incongruous word meant to add some surprising added element to the output).

2.2.3. Prompt structure

Numerous studies pinpoint the necessity of developing prompting skills and master the iterative nature of prompts (Dang et al., 2022; Oppenlaender et al., 2023; Robertson et al., 2024). The pitfall of a deceptively easy interaction with the AI is brought to the fore when one forgets it is an interaction with an LLM (Willey et al., 2023), the anthropomorphic characteristics may steer the user away from their focus on prompt structure and, even when prompt structure is implemented, a slight change can have a great impact on the output (Zhao et al., 2021). Nevertheless, the most common structure found in internet pages or in research papers revolves around four components: request (instruction), references to either previous bot answers or external sources, format (or output indicator), framing (i.e., contextual details) (Budiu et al., 2023a). Their language format may vary: the prompt may be a question, a request ("I need…"), a command ("give me two examples"…), or even an incomplete sentence. Yet other special types of prompts do not use the typical structure ("Can you do X", "Give me more", and "filler prompts"); prompts lacking core components may lead to long lasting and inefficient conversations comprising numerous query refinements (Budiu et al., 2023a).

2.2.4. Prompt literacy

Because prompt structure is key to the output, guides have been edited to help experts as well as laypeople. There is also a wide diversity of prompt courses available, promising breathtaking results and high paying jobs. All offer users techniques to improve their prompt literacy. Prompting is not an intuitive skill; it is acquired through experimentation (Oppenlaender et al., 2023). Indeed, some strategies are not natural in human language, like iteration strategies, using the sandwich technique with long prompts (the same statement in two places), or resorting to caps or hyperbolic language. The intuitive approach of non-experts often appears flawed and inefficient.

Budiu et al. identify the reasons for which prompt writing and including the four core components is difficult, including a lack of vocabulary or knowledge (Budiu et al., 2023a). Zamfirescu-Pereira et al. (2023) have stressed the opportunistic exploration of prompt designs by end-users, as well as the difficulty to make effective progress. A limited understanding of prompt engineering and LLMs abilities, together with intuitions rooted in human social interactions and experiences, impact the quality and efficiency of the prompts (Zamfirescu-Pereira et al., 2023). Prompt engineering thus appears to be a creative skill for creating AI art (Oppenlaender et al., 2023), and a new competence meant to become essential (Korzyński et al., 2023), if not inevitable (Walter, 2024) in most jobs.

However, it is a fast-evolving environment with rapidly changing circumstances, numerous recommendations, but no systematic rule; this may probably all be outdated when the article is published, and some points may become less important as technologies develop. Nevertheless, prompting is seen as a new essential, if not vital, skill for both teacher and students (Gattupalli, et al., 2023) and teaching prompt design needs to be careful tailored, as it involves challenges such as a well-understood critical approach to AI. Using AI efficiently implies recognizing its fallibility (hallucinations, hits and misses, bias, etc.), and a social context that needs to be taken into account, in order to prevent widening disparities in technological skills (Walter, 2024). Theophilou et al. (2023) conducted a pilot educational intervention and observed a shift in attitude, and improved prompting strategies, thanks to education. That is teaching chatGPT prompting strategies as part of a global project to help students better understand LLM limitations, constraints and techniques, and thus tackle misuse as well as negative representations to make the most of AI (Theophilou et al., 2023).

3. Methodology

3.1. Research questions

In this study, we explore whether AI generated poetry could be used to improve the analytical, interpreting, or writing skills of students enrolled in an English as Foreign Language (EFL) literature university course. An AI-assisted creative writing poem activity was provided in the hope that the exercise, through the generation of prompts, would help students explicitly voice and transfer the literary analysis skills and literary knowledge taught in the course. Traditionally, this assignment is done as homework with students producing a poem and

justifying their choices according to the specificities of the author under study. The main research question asks to what extent the exercise provided for the explicit application of the literary analysis elements taught in the course. There are three sub-questions: What are the characteristics of the prompts intuitively generated by the students? Do students intuitively manage to generate efficient prompts? What pedagogical implications can be inferred from the results?

3.2. Research approach

This study aims at finding answers to research questions. Hence it follows an inductive process relying on the formulation of research questions and the gathering of predominantly qualitative data through surveys and observations. We borrow the objectives and the methodology of different kinds of research, including exploratory research (through the exploration of the recent and under-investigated domain of AI within a literature class), explanatory research (e.g., we aim for a better understanding of students' intuitive prompt writing), and action-research (the study is deeply rooted within an educational environment and aims for the improvement of language education practices). A number of features of correlational research have been embraced (although without manipulating or controlling variables) to facilitate the investigation of the relationship between the prompts and the explicit application of the literary analysis elements taught in the course, between the prompts and the intermediate ChatGPT answers or the final poem produced. Since participants were not randomly assigned, we are aware of potential response bias, and tried to minimise it as far as possible, mainly through anonymisation.

3.3. Participants

A literature class of third year students (L3) following a degree in EFL have been chosen for the experimentation (Licence Langues, Littératures et Civilisations Étrangères, Anglais/Bachelor's Degree in English, université de la Guyane). The group is composed of 16 students (2 boys and 14 girls). These students are familiar enough with literary analysis and creative writing not to be hampered by the instruction itself.

A pilot study was carried out with 16 pre-service teachers (M1) enrolled in a master's course in Teaching English as a Foreign Language (TEFL) (Master Métiers de l'Enseignement, de l'Éducation et de la Formation (MEEF) Anglais, université de La Réunion).

3.4. Experimentation

Figure 1 below illustrates the timeline of study. The M1 pilot test was held on 30 January 2024, and the L3 experimental AI-assisted creative writing poem exercise was carried out on 6 February 2024.

Figure 1 – Phases of the study.

	L3 (EFL Literature Students) Preliminary sessions	M1 (Pre-service EFL teachers) Pilot test *Survey, collection of transcripts*	L3 Experimentation *Survey, collection of transcripts*	M1 Post-activity analysis
Description	▪ Analyses of poems by Harlem Renaissance poets; including Hughes' "The Negro Speaks of Rivers" & "As I Grew Older" & "the Weary Blues"	▪ Survey (student uses of AI; AI made/human-made poems) ▪ Student presentation on AI ▪ Analysis and testing of L3 lesson plan ▪ Reflection on potential research questions for a master's thesis	▪ Comparison of Hughes' poetry with a poem by Claude McKay ▪ AI-assisted creative writing poem exercise ▪ Self-assessment ▪ Survey	▪ Preliminary analysis of a selection of L3 students' transcripts ▪ Student presentation on prompt engineering ▪ Discussion on the integration of prompt engineering within EFL lessons ▪ Preparation of posters: *Tips to write efficient prompts*
Teaching objectives	▪ Poem analysis methodology ▪ Identifying elements of Langston Hughes's literary style & Harlem Renaissance poetry	▪ Reflection on poetry analysis & Hughes's literary style ▪ Didactic analysis of a lesson plan	▪ Compare & contrast to consolidate the identification of Hughes's literary style ▪ Voicing & transferring literary analysis skills and literary knowledge	▪ Coding & interpreting qualitative data for a master's thesis ▪ Identifying elements of prompt engineering

The experimental activity takes place during the seventh class. During the first to sixth class, the approach to Langston Hughes has been a conventional one with alternate lectures and workshops based on the analysis of his poetry. Among the poems studied were "Dream Variations", "As I grew Older", "The Negro Speaks of Rivers", and "The Weary Blues". Particular attention has been paid to the elements which have been chosen as representative of Langston Hughes's style, such as a strong, vivid, poetic voice, radical language, a cause defended, an art celebrated, use of syntax, punctuation and unconventional grammar to create a rhythm, etc.

The students have been warned that this class would incorporate an AI tool and students have been asked to open a ChatGPT account beforehand. The day of the activity, the class began with a ten-minute conversation on students' perception and use of AI. A single question was asked: "Opinion and uses of AIs?". The participants were asked to react and interact. Next, the students were given a short comparative study to highlight Langston Hughes's specific style. The participants then carried out the AI-assisted imitation task. No stake was given during 30 minutes, to avoid exerting pressure and influencing students in their use. To identify this moment within the transcripts, the students were asked to start the first prompt they wrote after this teachers' interruption/intervention with a special character (e.g., "9) why is this poem the same style as Langston Hughes?" (ID13)). The students then carried out a self-assessment, and the teacher evaluated the poems created. Table 1 below identifies the evaluation criteria, relying on the items expected within the poem.

Table 1 – Literary terms used during the course and expected within the AI-assisted creative writing poem in Langston Hughes's style.

Themes &	racism, race relations, black pride, freedom, emancipation, struggle, elevation, change, music, American dream, working class, ordinary lives,

Topics	universal man, past & present, a strong, vivid, poetic voice, radical language, or lyrical poem, etc.
Imagery & Symbolism	river, sun, body, wall, opposites, power, action, destruction; universality of imagery, etc.
Musicality	jazz, blues, repetitions, anaphora, alterations, cadence, musicality through sound; musicality through punctuation; rhymes, etc.
Structure & Prosody	syntax, structure of lines, irregularity, breaks, run-on-lines, punctuation, etc.
Grammar & Language	oral style rather than conventional written grammar rules; Black folk dialect, African tone, simple/everyday language, elegance, figurative, etc.

3.5. Data collection method & analysis

Data were collected through two surveys and the observational study of transcripts. The students carried out the experimental activity on ChatGPT. They were asked to share the link of their conversations and, as a backup, to copy paste their conversations in a text document. Anonymous Lime Surveys were taken online, the L3 survey included the link of the shared conversation, which allowed for the association of the survey and the conversation. The survey was taken at the beginning of the lesson for the M1 pilot test, and at the end of the lesson for the L3 experimentation. Anonymous ID numbers were automatically generated through Lime Survey.

For the M1 students, our survey comprised a set of eleven questions: Question 1 and 2 targeted students' use of AIs in general; questions 3 to 6 were Likert scale questions, they tackled AIs in an educational context and the need to adapt, or not, to this new tool; questions 7 to 10 were inspired by the Turing Test (Turing, 1950) to assess students' ability to discriminate human-made or machine-made poems (two poems were human-made and two machine-made). Each question invited the participants to leave a comment to argue further his or her answer. The last question was an open question for any further comment students wanted to add.

The L3 students were submitted a survey of sixteen questions. Questions 12 to 15 were identical to the questions asked to M1 students. Question 1 was targeting their use of AI in general. Questions 2 to 11 considered the assignment itself and the output obtained with a focus on their prompting and the perceived interest of such an activity for their literature course. Questions 12 to 15 tested their ability to discriminate between human-made or machine-man poems and question 16 was an open question for any further comment they wanted to add. The survey was studied under three entries: context, interaction and feeling/opinion.

The 16 conversation transcripts were coded manually (mainly through inductive coding) with atlas.ti, a data analysis software facilitating qualitative data analysis. 178 codes were generated from scratch to cover the samples; each time a new code was created, a text search was conveyed to recode the samples already coded and code the following ones. Deductive coding was used occasionally, mostly to look for specific terms or terms not found (e.g., although "metaphor" is used, "comparison" was never used; "American dream" appeared only once in a ChatGPT answer). Students' prompts and ChatGPT answers were identified (<Student>; <GPT>). 501 quotations were created (240 GPT prompts, 238 Student prompts).

To identify the characteristics of the prompts intuitively used by the students, we devoted 16 codes to the formal aspect of the students' prompts. We thus highlighted capital letters (e.g., "FREE VERSE NOW" (ID15)); familiar terms ("Bro, you're basically writing sonnets" (ID14)); politeness; discourtesy ("NO STANZA, NO STRUCTURE NOW" (ID13)); language mistakes (spelling, grammar, punctuation mistakes); unclear antecedents; the absence of a verb, of initial capital letters, full stops or question marks. The prompts were identified either as commands ("add a personal twist to it" (ID19) ; "stop with the structure" (ID23)); declarative sentences or segments ("It is too simple" (ID21)); personal comments ("Nice thanks" (ID21)); questions ("can you structure the poem so it will look like a free form poem" (ID6)); requests ("You are my literature teacher in L3 LLCER, I need you to write a poem inspired by Langston Hughes" (ID10)); and wishes ("I'd like to delete all the rhymes", (ID12)). All the students' prompts were subject to a word count and an idea count identifying the number of different ideas or elements within the prompt (e.g., "Add more prosaic elements, punctuation, free verses" (ID14) was coded <C nb ideas/elements 3>). The student's initial prompt launching the chat was identified; so was the presence of contextual elements and the use of the exact same words as the teacher's instructions. The clarity of the prompt and the use of idiomatic language were also coded. However, we consider the coding of such elements as a heavily biased enterprise so only cautious and exploratory conclusions can therefore be drawn from these elements.

In order to scrutinize the type of process used by the participants and look at the way the prompts were generated, customized, and used to re-equilibrate the AI poem generated, we identified the total number of prompts used within the transcripts; the level of the poem targeted (poem, stanza, title, verse or word level); the purpose of the prompt (22 codes, e.g., <a broad task / Prompt> (such as "write a poem in Hughes' style"); <adding an improvement>; <asking for information or a definition>; <asking for proof> ; <asking for variation>; <asking for feedback>; <repeating>, etc.).

To try and define whether the prompts were efficient, we marked the final poem created, and coded the intermediary ChatGPT answers: <Answer seems to fit last prompt given>; <Answer inappropriate/not satisfying>; <Answer partially satisfying>. The subjectivity of such coding is evident, and here again we call for a cautious interpretation. Additional codes were used to analyse ChatGPT answers, such as <Answer provides literary information>; <ChatGPT

explicits/justifies its choice> or <Answer highlights textual interpretation of instructions>. The final poem was marked A, B, C or D using five criteria: "Something personal", "Topic or theme corresponding to Langston Hughes", "Imagery and musicality", "Global structure", "Grammar and punctuation". A poem marked A meets 4 to 5 criteria, B meets 3, C meets 2 and D meets only one or none.

To identify the explicit application of the literary analysis elements taught in the course, we looked at the literary terms explicitly used. Only the literary terms explicitly mentioned within the participants' prompts or the GPT answers were coded. Thus, the term "struggle" was coded <struggle>; but "making violent efforts to get free" would not be coded <struggle>. The words within the poems were not coded. Thus, the word "Harlem" was recurrent in the poems, but it was not coded when used within a poem. As for all other codes, no matter the amount of repetitions, a code was only used once within each quotation (each student prompt, or each ChatGPT answer, no matter the amount of words). A general code was used to identify references to Langston Hughes or his style. The table below provides a synthesis of the terms expected and the codes created.

Table 2 – Codes used to codify the literary terms explicitly found in the conversations.

Themes & topics	<T African American / Black / Blackness / Black pride> ; <T authenticity>; <T culture/cultural>; <T discrimination>; <T freedom>; <T Harlem>; <T Harlem Renaissance>; <T hope / optimism>; <T identity>;< T loneliness>; <T ordinary>; <T personal twist>; <T poetic voice>; <T poverty>; <T racism>; <T resilience>; <T segregation>; <T setting>; <T social realism>; <T strength>; <T struggle>; <T theme>; <T topic>
Imagery & Symbolism	<IS imagery>; <IS metaphor>; <IS simile>; <IS symbolism>; <IS vivid / vividly>
Musicality	<M repetition.s>; <M alternate rhymes>; <M blues>; <M enclosed rhymes>; <M jazz>; <M musicality>; <M rhymes>; <M rhythm>
Structure & Prosody	<SP free flow>; <SP free form>; <SP free verse>; <SP lines>; <SP meter / metrical>; <SP punctuation>; <SP run on line>; <SP sonnet>; <SP stanza>; <SP structure.s>; <SP verse>
Grammar & Language	<L accessibility>; <L broad audience>; <L simplicity>; <L vernacular>

4. Results

4.1. Survey

The survey submitted to the students helped us get a glimpse of how familiar they were with AI, and what their opinions about this new tool were. It was also used to understand how they perceived their interaction with the AI, and their opinions on the output they had obtained. Two participants did not complete the survey in L3. In both classes, half of the students rarely or never use AI and, in L3, 7 students out of 14 mention ChatGPT when asked if they use AIs; four of them use more specific AIs such as grammar checkers. All of these seven students mention an AI related to academic tasks. It is not the case with students at a master's level. The latter mention using AIs for personal activities like cooking "to get examples of healthy recipes" (N5) but also for their academic work, especially rephrasing. They never mention complex tasks involving prompts which would need to be elaborate.

Most L3 students believed the result to be poor or very poor, and there is no correlation between their opinion on the result and their use of AI, as those who use it frequently find the result as poor or average as those who use it rarely or never. Eight students found it hard to find the right words or the right sentence structure for their prompts. But this has no correlation with their being familiar with AIs or not, as three of the seven students who never use AIs found it easy to find the right words. Nine students said they had followed intuition. Only one had prepared their prompt, while two had listed the elements they needed to focus on before their initial prompt. Four students declared they had used strategies which they did not usually use in class. Ten students believe that prompt writing should be taught at university, and only two believe it cannot be taught. Two students believe it should be a high priority, and six believe it can be included in the courses they already have, such as this literature class. Five participants believe these prompting classes should be given by IT engineers.

As regards their feeling and the collaboration with the AI, 10 out of the 14 students have experienced frustration, one added a comment "I was more focused on not insulting the AI than on the actual poetry" (ID14) when asked if this class had helped them better understand Langston Hughes's poetry. To this latter question, most participants answered "To some extent" or less. Considering the authorship of the poems generated, six of the L3 students replied that the AI, themselves and Langston Hughes were the authors; one student left a comment explaining that "if the author's style is so distinct that it can be replicated by a trained AI, the poet or the artist should be credited" (ID21). None of the students believe they alone were the authors, and five students picked chat GPT as author.

Eight M1 students consider teaching prompting at university level as a high to essential priority, only one less considers it high to essential for high school students. Ten consider it a moderate to not at all a priority for pupils under the age of 13.

Both groups of students have been submitted 4 poems. Poems 1 and 4 were AI produced, poem 1 was deemed written by a human for three third year students but eight M1 students. Poem 4 was deemed written by a human for more than half of each group. As for the human-made

poems, no entire group has agreed on the human origin, but overall, the L3 have fared better in discriminating between human and AI poetry.

4.2. L3 Student-ChatGPT conversations

The analysis of the formal aspect of the transcripts revealed that 66 out of the 239 student prompts comprised language mistakes (27,6%), including syntactic or grammatical errors ("*try again the poem" (ID10); "*Why this poem is like Langston Hughes" (ID6)), lexical errors ("*Do not jump line" (ID13); "*can you change the ponctuation even remove some (ID16)"), spelling slips ("*Delete all the rhymes fron this poem" (ID12); "*mae it only 3stanza" (ID18)), or the use of inadequate punctuation in questions ("*Can you write a poem in the style of Langston Hughes but with a personal twist." (ID6)). The absence of a question mark was not coded as a mistake since absence of final punctuation, whether full stop or question mark, was coded separately: 75% of the prompts written by the students had no final punctuation (180 prompts). 48.5% did not start with a capital letter (116 prompts). Students often used elliptical sentences or segments (e.g., "*Can add a repetition in the third stanza" (ID17); "with less rhymes" (ID10)).

Although all these elements could impact the clarity of the message, only 11% of the prompts were deemed unclear. Numbers however need to be taken with caution, the clarity of the message being a subjective element. Clarity seemed compromised in prompts such as: "*Whats the structure that defines Langston Hughes style in his peace of works" (ID 20); "*POEM IN LANGSTON HUGHES WITH A PERSONNAL TWIST" (ID12); "*Do not jump line" (ID13). However, a teacher aware of the task instructions or an EFL teacher acquainted with loan translations may both consider these sentences comprehensible.

The analysis of the transcripts also disclosed the use of common abbreviations ("More pls" (ID 15); unconventional language ("Do you know what Free verse is supposed to be like? Cause this poem is not" (ID7); "Bro, you're basically writing sonnets, thats not what im asking…" (ID 14)); informal exclamations ("okay" (ID14); "GOOOOOOOOOOD JOOOOOOOOB" (ID21)).

Politeness to the machine appeared in 7 transcripts (ID5, 7, 10, 14, 15, 18, 20), through greetings ("Hello can you help me write a poem in Langston Hughes ? …" (ID 7); "Aloa Chat GTP !" (ID20); "hiii" (ID15)); the use of the adverb "please" ("can I have another one please (ID5); "Make it into one big Stanza please" (ID14)); or the use of the modal "could" ("Could you define and analyse his poetry and literary style" (ID14); "*could you change the strucure a bit" (ID18)). Anthropomorphism reached a climax when ID20 asked ChatGPT its opinion ("Do you like this peace of work?" (ID20)), and ID9 its interpretation of life ("okay what is your interpretation of life" (ID9)).

Dissatisfaction with ChatGPT's answers appeared through the use of another prompt, or explicitly through negations ("It is not a free verse, try again" (ID10); "this one is not enough

Langston Hughes" (ID5)), as well as the use of rhetorical questions ("Okay but wasn't he also known for his free verse writing style?" (ID7)). Dissatisfaction was at times associated to capital letters or the adverb "now", which we interpreted as signs of exasperation ("NO STANZA , NO STRUCTURE NOW" (ID13) ; "FREE VERSE NOW" (ID15)). Annoyance permeated some discourteous prompts which seemed to be scolding the machine ("Okay make the first stanza's verses longer, add at least two words to each and stop the rhyming omg" (ID14); "you know what forget it" (ID18)). ID21 resorted to four extremely aggressive prompts abusing ChatGPT: "you really don't have talent in art you fool"; "I'm embarrassed at your incompetence"; "i could never fail my parents like you fail me"; "Jump off a bridge" (ID21).

Two out of the 16 initial prompts were single greeting words ("hiii" (ID15); "Hello" (ID10)); the second prompt then used by ID15 was "You know Langston Hughes?" while ID10 was the only one to provide some context, asking: "You are my literature teacher in L3 LLCER, I need you to write a poem inspired by Langston Hughes" (ID10). Three initial prompts were very similar to the wording of the teacher's instructions (e.g., "Can you write a poem in the style of Langston Hughes but with a personal twist." (ID6); "Write a poem in the style of Langston Hughes. It has to have my personal twist." (ID20)); 4 other prompts comprised at least some of the words used in the teacher's instructions (e.g., "write a poem in the style of langston hughes" (ID3). 7 prompts were thus rather broad and asked for the production of a poem. 5 other prompts (and ID15's second prompt) explicitly requested information on Langston Hughes, his poetry or the method to write in the style of Hughes: "Hi, what do you know of Langston Hugues" (ID14); "How to write a poem in the style of Langston Hughes?"(ID16); "*what do you know about langston hughs"(ID18). One prompt added a thematic content request: "*Hello can you help me write a poem in Langston Hughes? I want it to be about strength and loneliness" (ID7). The last prompt showed originality since the student copied and pasted Hughes's poem "The Weary Blues" (241 words; ID13). ID13's poem apart, only 4 prompts out of 15 begin with a capital letter and end with a full stop or a question mark. However, again, this does not seem to significantly affect the clarity of the prompt. At this stage, there are 5 commands (e.g., "Write…" (ID12), "Describe…" (ID21)) and 8 questions (e.g., "Can you write…" (ID6), "What is distinct…" (ID11), "How to write…" (ID11). 7 of the ChatGPT answers following the initial prompts appeared to be partially satisfying (usually when the student had asked for a poem), and 9 seemed to be satisfying (usually when the student had asked for information on Hughes).

The final results of the poem production were lukewarm: only one student was awarded an A, five students obtained a B, and C a 10. However, no production was given a D. Most poems equated a feeling of loneliness with a wandering of the poetic persona in the streets of Harlem, offering a clear focus on the urban setting; whereas Langston Hughes focuses on the people, and the theme of loneliness is used to support a sense of violent segregation. The musicality is never quite reproduced as ChatGPT remains rather conventional in syntax and in line structure, if not in stanza structure. ChatGPT generated mostly quatrains, even when prompted to do otherwise.

Yet, within the conversation, 49% of ChatGPT answers seemed to fit the last prompt given by the student (118/240 GPT prompts). Considering the previous student prompt given, 29% of ChatGPT answers appeared to be partially satisfying, and 19% were coded as non-satisfactory.

Figure 2 shows that in total, each student used from 4 to 33 prompts. The distribution is positively skewed. The average is 15 prompts, which is also the number of prompts given by the sole student who was awarded an A Grade. The 5 students who obtained a B grade used on average 12 prompts. The 10 students who obtained a C grade used on average 17 prompts. ID21 (C grade) used 22 prompts including four prompts which appeared out of scope, or at least could barely lead to the improvement of the poem ("you really don't have talent in art you fool"; "I'm embarrassed at your incompetence"; "i could never fail my parents like you fail me"; "Jump off a bridge" (ID21)). Therefore, no correlation was found in the corpus between the number of prompts given by the student and the quality of the final production.

Figure 2 – Number of prompts given by the students in the conversation.
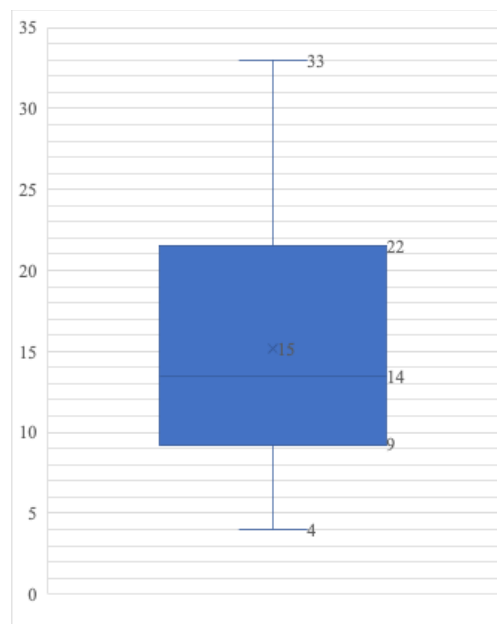
We could not identify any recurrent global method in the students' requests. The students seemed to iteratively engage with ChatGPT in a conversational and rather unorganized manner. ID7, who was awarded an A, started by announcing a theme and working at poem level, primarily on the structure (see figure 3). They then introduced improvements focusing on musicality, free verse, and a title. ID 10 appeared at a loss with structure, confronting the production of sonnets (see figure 4). Although they began with a promising contextualization and tackled style, structure, imagery and theme, their final production was poor. ID10's transcript is characterised by a strong focus on structure (which the student tries to modify with references to sonnet and free verse); the repetition of similar prompts which had previously seemed abandoned (prompt 3: "with less rhymes", prompt 13: "No more rhyme"); many word for word repetitions (which did not prove efficient here, although other transcripts showed that mere repetition was sometimes fruitful); and a change of theme after 12 prompts, which does

not actually make the machine start from the scratch (structure and punctuation remain the same).

**Figure 3 – ID7's prompts (15 prompts, A Grade).**

> Hello can you help me write a poem in Langston Hughes? I want it to be about strength and loneliness: (Poem level; Global style; Precise themes)
> Can it be less structured? (Poem level; Structure)
> Even less structured with different types of stanzas (Poem level; Structure)
> Can you make it in the style of Langston Hughes? (Poem level; Global style)
> Can you add Some repetitions (Poem level; Repetitions/musicality)
> Less repetitions (Poem level; Repetitions/musicality)
> Can add a repetition in the third stanza (Stanza level: Repetitions/musicality)
> 9 what makes you believe that this poem is in Langston Hughes style? (Poem Level; Global style)
> Okay but wasn't he also known for his free verse writing style ? (Free verse, structure)
> Can you make the earlier poem in a more free verse style? (Poem level; Free verse, structure)
> I don't want consistent stanzas (Poem level; Structure)
> Do you know what Free verse is supposed to be like? Cause this poem is not (Poem level; Free verse, structure)
> Nice it's much better now can you name it? (Title level)
> It is too simple (Title level)
> Nice thanks

Figure 4 – ID10's prompts (23 prompts, C Grade).

> Hello
> You are my literature teacher in L3 LLCER, I need you to write a poem inspired by Langston Hughes (Poem level; Global style)
> With less rhymes (Poem level; musicality)
> More simple (Poem level; level of difficulty)
> Talk about the segregation (Poem level; theme)
> More subtle (Poem level; subtlety)
> Try to imitate the poet (Poem level; Hughes' style)
> Try another one (Poem level)
> Use more imagery (Poem level; imagery)
> 9 why is it like Langston Hughes (Poem level; Hughes' style)
> 9 try again the poem (Poem level)
> 9Change the theme (Poem level; theme)
> 9No more rhyme (Poem level; musicality)
> 9Try to free verse in the style of Langston Hughes (Poem level; free verse, structure)
> 9not a sonnet (Poem level; structure, sonnet)
> 9not a sonnet (Poem level; structure, sonnet)
> FREE VERSE (Poem level; free verse, structure)
> FREE VERSE NOW (Poem level; free verse, structure)
> 9 FREE VERSE NOW (Poem level; free verse, structure)
> 9 Do not mention the name (Word level)
> 9 FREE VERSE NOW (Poem level; free verse, structure)
> 9It is not a free verse, try again (Poem level; free verse, structure)
> 9stop with the structure (Poem level; structure)

The literary terms explicitly used by ChatGPT and the students were distributed into the five domains (language and grammar, imagery and symbolism, musicality, structure and prosody, themes and topics). A code-document analysis revealed that every script comprised some literary terms, used either by chatPGT or the students. The box-and-whisker below (figure 5) displays the dataset on a five-number summary: the median value of the data set is 39 literary terms. The lower whisker shows that only two literary terms were found in the poorest script (ID3, <Harlem>; <personal twist>); the second lowest number is 17 (ID4). The upper whisker corresponds to the highest data point, which reaches 68 literary terms (ID14); the second richest script comprises 59 terms (ID7). The interquartile range is thus of 33, and the average number of terms per script is 38.

Figure 5 – Distribution of the number of literary terms used by ChatGPT or the students within the scripts.
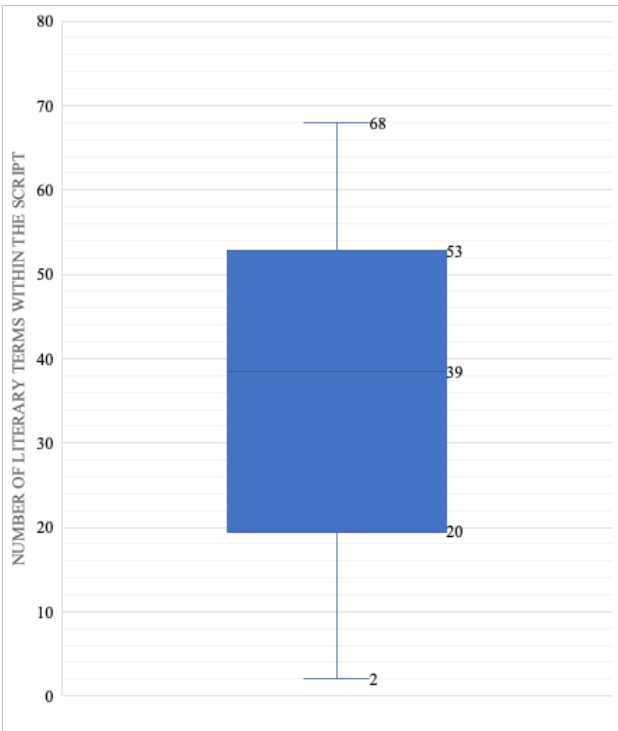
A co-occurrence analysis however shows that participants employed more literary terms pertaining to musicality and structure than ChatGPT (Figures 6 and 7). 31 occurrences of "free verse" and 21 occurrences of "stanza" were coded in the students' prompts ("free verse" was coded 6 times in ChatGPT answers, and "stanza" 12 times.). Conversely, chatGPT made use of many terms which were not at all found in the students' prompts (e.g., resilience (25); struggle (22); blues (21); jazz (20); identity (16); Harlem (12); rhythm (18), etc.).

Figure 6 – Imagery & symbolism, language, musicality, structure & prosody: co-occurrence analysis table showing the terms explicitly used by ChatGPT and the terms explicitly used by the students.

		GPT ● ◇ 240	Student (L3) ● ◇ 239
● ◇ IS imagery	⊕ 27	23	4
● ◇ IS metaphor	⊕ 5	5	0
● ◇ IS simile	⊕ 1	1	0
● ◇ IS symbolism	⊕ 9	9	0
● ◇ IS vivid / vividly	⊕ 19	19	0
● ◇ L accessibility	⊕ 3	3	0
● ◇ L broad audience	⊕ 2	2	0
● ◇ L simplicity	⊕ 4	4	0
● ◇ L vernacular	⊕ 8	8	0
● ◇ Langston Hughes / LH style / LH spirit	⊕ 137	65	68
● ◇ M <repetition.s>	⊕ 15	12	3
● ◇ M alternate rhymes	⊕ 3	1	2
● ◇ M blues	⊕ 22	21	0
● ◇ M enclosed rhymes	⊕ 2	1	1
● ◇ M jazz	⊕ 21	20	0
● ◇ M musicality	⊕ 10	7	3
● ◇ M rhymes	⊕ 30	15	15
● ◇ M rhythm	⊕ 18	18	0
● ◇ SP free flow	⊕ 3	2	1
● ◇ SP free form	⊕ 8	4	4
● ◇ SP free verse	⊕ 38	6	31
● ◇ SP lines	⊕ 19	8	11
● ◇ SP meter / metrical	⊕ 5	5	0
● ◇ SP punctuation	⊕ 10	4	6
● ◇ SP run on line	⊕ 1	0	1
● ◇ SP sonnet	⊕ 4	1	3
● ◇ SP stanza	⊕ 34	12	22
● ◇ SP structure.s	⊕ 33	16	17
● ◇ SP verse	⊕ 19	9	9

Figure 7 – Themes and topics: co-occurrence analysis table showing the terms explicitly used by ChatGPT and the terms explicitly used by the students.

Term	Count	Col1	Col2
T African American / Black / Blackness / Black pride	19	17	2
T authenticity	6	6	0
T culture/cultural	14	14	0
T discrimination	6	6	0
T freedom	3	3	0
T Harlem	15	12	0
T Harlem Renaissance	11	11	0
T hope / optimism	12	12	0
T identity	17	16	0
T loneliness	1	0	1
T ordinary	6	6	0
T personal twist	9	1	8
T poetic voice	1	1	0
T poverty	8	8	0
T racism	4	4	0
T resilience	25	25	0
T segregation	1	0	1
T setting	2	2	0
T social realism	2	2	0
T strength	15	12	2
T struggle	23	22	0
T theme	19	14	5
T topic	6	3	3
Wrong lit vocab	2	0	2

Together with the lexical difference of the terms used by ChatGPT and the students, the amount of literary terms employed by ChatGPT is remarkable. The ChatGPT answers expose students to a large quantity of literary terms. Figures 8 and 9 highlight the number of words used by the students, and by ChatGPT. The student prompt "*What is distinct about Langstong Hughes poem writing style?" (L3ID11) thus triggered an answer providing coherent literary information pertaining to 18 codes (<IS imagery; IS symbolism; IS vivid/vividly; L simplicity; L vernacular; M blues; M jazz; M rhythm; T African American; T authenticity; T culture/cultural; T Harlem Renaissance; T identity; T ordinary; T poverty; T racism; T resilience: T struggle>).

Figure 8 – Co-occurrence analysis Sankey diagram highlighting the amount of literary terms used by the students.
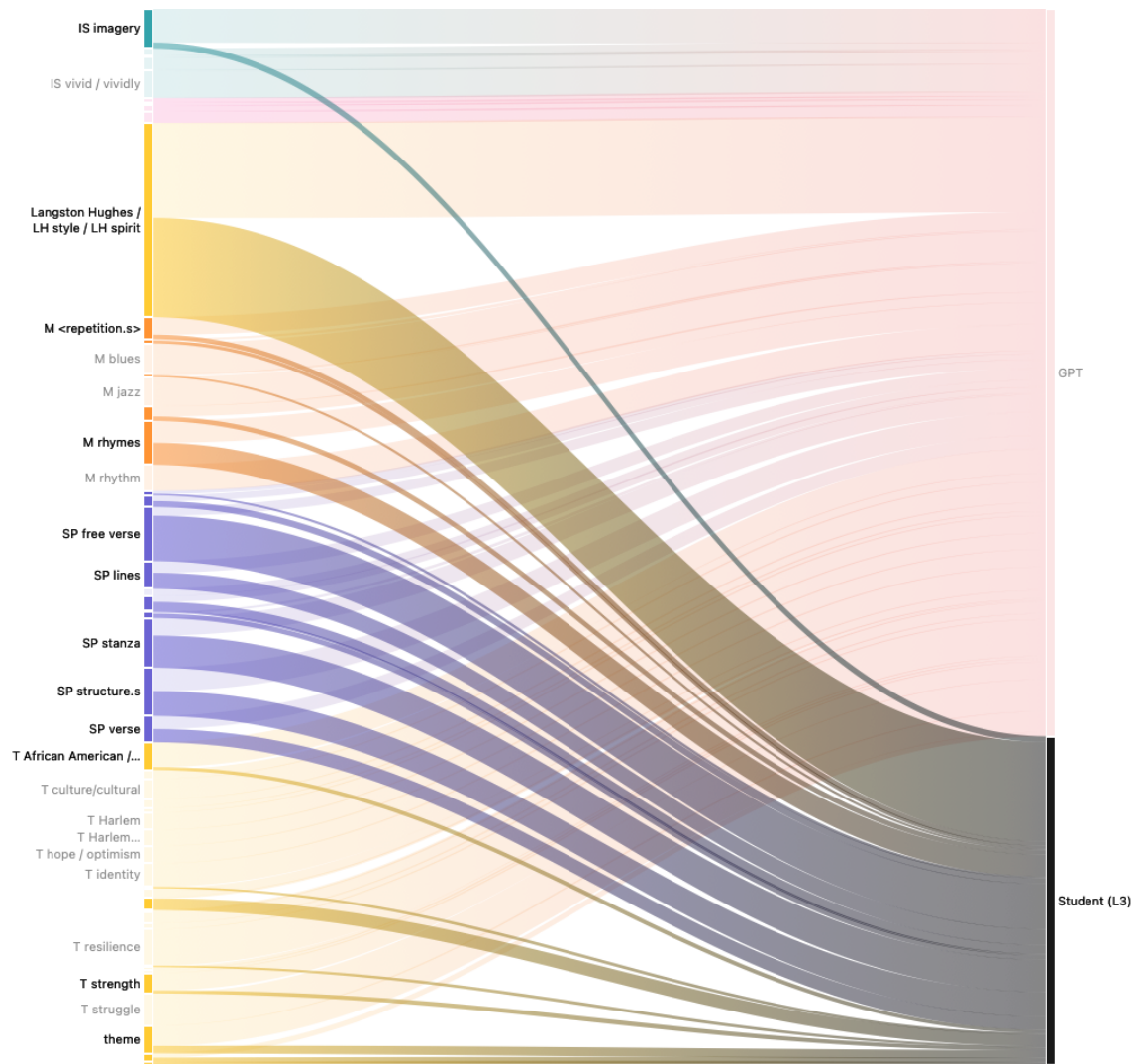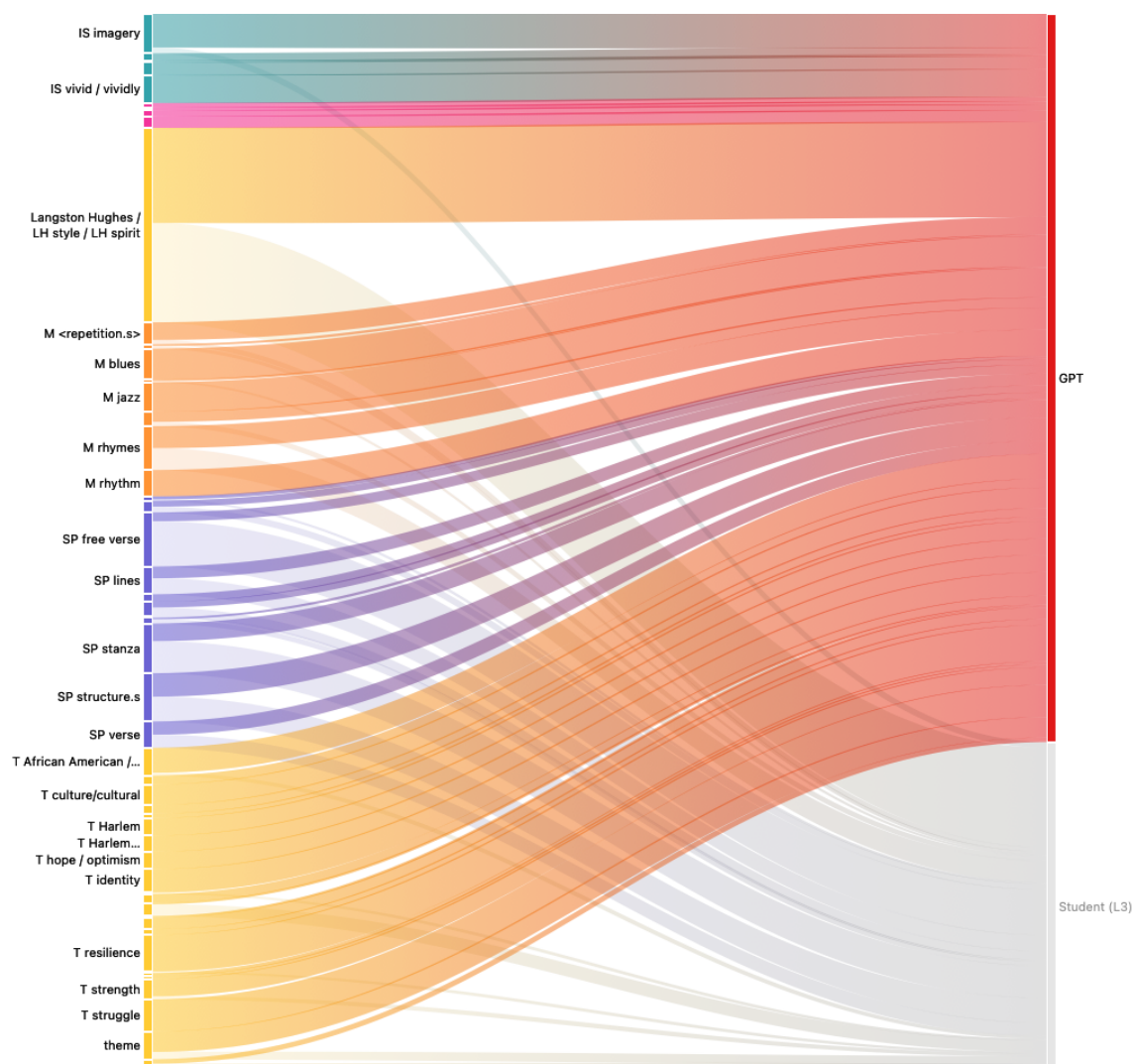
Figure 9 – Co-occurrence analysis Sankey diagram highlighting the amount of literary terms used by ChatGPT.

5. Discussion

The traditional imitation writing assignment, especially because it comprises a self-analysis section, aims at having the students reflect on their choices in content as well as in form. We believed that the students would use an AI if asked to do this task today, and we wondered if their prompts would incorporate traces of their analysis in their command to generate a poem. Our study thus questioned the extent to which the GPT exercise provided for the explicit application of the literary analysis elements taught in the course. The analysis of the corpus showed that despite lukewarm final productions, ChatGPT exposed students to a large quantity of literary terms which were not always the same terms as the ones used by the students and pertained to the five domains expected (language and grammar, imagery and symbolism, musicality, structure and prosody, themes and topics). Moreover, these terms were used within answers fitted to the student's previous question, which opens the path towards differentiated instruction.

The study also scrutinised the characteristics of the prompts intuitively generated by the participants and questioned the efficiency of such prompts. We concluded that these prompts comprised language mistakes, unconventional language, abbreviations, elliptical sentences, typos, signs of anthropomorphism, etc., that is to say many characteristics of modern instantaneous communication and human-bot conversations. However, these elements did not necessarily impact the clarity of the message.

The corpus revealed that participants engaged with ChatGPT in a conversational and rather unorganised manner. Writing a poem involves forethought about a subject or a situation, and a slow line by line process. This appears to have been completely skipped by all the students but one. None of the participants asked for a first line. The use of the AI has utterly reversed the process from macro to micro, generating one poem which then had to be altered. Furthermore, most literary terms focused on the global aspect of the poem and on its layout on the page. Maybe one way of circumventing this reversed process would be to start the task in the class, make the students reflect on the process in action, and on how they would proceed with an LLM to maintain the same process.

Although ChatGPT exposed students to a range of literary terms that differed from those they used, the students themselves hardly ever used hypernyms or hyponyms, preferring to repeat the same terms. It could be that the students have been quickly taken over by their inability to steer the LLM towards the right output. Emotions, as well as impatience, have prevailed, given the short if not elliptical prompts, conveying the idea that they expected the AI to lead the way, instead of using it as an assistant. Their frustration is palpable in form and content of prompts. It is also possible that students have not thoroughly read the generated poems, asking straight away for a change in layout, for example in the number of lines in a stanza. The swiftness of the LLM responses may have led students to adapt to the LLM by adding immediacy in their interaction, turning the required reflexivity into spontaneity. The latter being also visible in the quick swaying of their emotional response to one and the same GPT response.

Adding the LLM tool has changed stakes for the students; it could be that the novelty of the tool has turned the assignment away from its original target because of loss of academic landmarks and resurfacing of chatbot interaction habits.

Our third sub-question considered the pedagogical implications which could be inferred from the results. The *why* teach prompt engineering has been demonstrated by researchers who described the blatant need for prompt engineering education both to tackle misuse (Theophilou et al., 2023) and develop what may become vital skill (Korzyński et al. 2023; Smith, 2023). Similarly, the surveys and discussions showed that students needed guidance since they were unsure about how to use ChatGPT, and whether it was right or not. Moreover, within a literature course, a reflection on creativity, authorship and human AI cooperation appeared essential. We believe most students will use LLMs for their home assignment, and this reflection could help in keeping the academic stakes a priority in the task if they carry it out unsupervised.

However, we have found only partial answers to the *how* to teach prompt engineering. Students-teachers reflections on digital literacy may integrate digital citizenship, ethical considerations

or needs analyses. Yet providing prompt writing recommendations is a thorny issue: it is difficult to go beyond general tips on source check, message clarity, needs and wants analyses, organisation, or common sense. What is sure is that the teacher's traditional request for full, grammatical, polite sentences in class appeared irrelevant in a human-AI conversation. Yet, the rude and impatient prompts used by a student call for abuse awareness and prevention, since psychologists have proven that online behaviours may have offline consequences. Our study demonstrated no correlation between the number and type of prompts used and the final result obtained. Moreover, within two transcripts, the same prompts did not yield the same results. Despite the general recommendations available online, prompt engineering is a fast-evolving field and the efficiency of the techniques depends on the environment.

Finally, it appears we had overlooked a certain amount of limitations, such as ChatGPT's biases and its inability to get off the beaten track; indeed when asked if it could use the terms "Black girl" the AI answered "As an AI language model, I strive to use language that is respectful, inclusive, and sensitive to diverse identities and experiences." It also appeared socially biased as when asked about Langston Hughes, ChatGPT mentioned Harlem and associated it to "struggle", "poverty", and the more fashionable and pervasive term of "resilience". This pitfall has been flagged by the BIG-bench Collaboration (Srivastava et al., 2023). ChatGPT has fared also rather badly in the patterns produced, as it offered sonnets, traditional rhyme patterns and quatrains. It has seemed unable to generate a poem not following western mainstream conventions. Any ChatGPT production should therefore entail consideration of permeability to toxic content, simplification, bias, and the extent to which GenAI tools are WEIRD–Western, educated, industrialised, rich and democratic.

We had not overlooked the additional skills that the use of a GenAI would involve, yet we had not anticipated the point to which the imitation process would be reversed, thus requiring a complete change of perspective. Furthermore, the contiguity between GenAI and mere chatbot conversation has amplified participants' loss of focus on keeping some hindsight to guide the LLM towards the desired output. This point would have benefited from timing the ChatGPT conversation so as to have a clearer view of the time spent reading generated poems and explanations.

6. Conclusion

The activity required not only literary but digital literacy and digital citizenship skills. It also entailed a redefinition of teaching posture, since the teacher lost control during the time span of the student-bot conversations, while machine tutoring was momentarily made possible. The exercise indisputably exposed students to a large quantity of literary terms fitted to their individual conversations with the chatbot, and it provided for the explicit use of literary vocabulary. Such a benefit is however tempered with limitations and complications. The spontaneous approach of students in their completing of the assignment using ChatGPT could be reined in with guidelines about prompting. Yet the prompting skills needed to be efficient add another layer of difficulty to the disciplinary skills already at stake in carrying out the

assignment. This could also widen the gap between computer-wise students and those who are not, and who may struggle further with this extra layer. This experiment has also shown how the emotional response prevails over a more objective, academic response which draws the assignment away from its purpose.

Remerciements

Références

Abrams, M. H., & Harpman, G. G. (2012). *A glossary of literary terms* (10th ed.). Wadsworth Cengage Learning.

Badura, M., Lampert, M., & Dreżewski, R. (2022). System supporting poetry generation using text generation and style transfer methods. *Procedia Computer Science*, *207*, 3310–3319. https://doi.org/10.1016/j.procs.2022.09.389

Braun, M., Greve, M., Kegel, F., Kolbe, L., & Beyer, P. E. (2024). Can (A) I have a word with you? A taxonomy on the design dimensions of AI prompts. *Proceedings of the 57th Hawaii international conference on system sciences*.

Budiu, R., Liu, F., Cionca, E. & Zhang, A. (2023a). *Prompt structure in conversations with generative AI*. Nielsen Norman Group. https://www.nngroup.com/articles/ai-prompt-structure/

Budiu, R., Liu, F., Cionca, E. & Zhang, A. (2023b). *The 6 types of conversations with generative AI*. Nielsen Norman Group. https://www.nngroup.com/articles/AI-conversation-types/

Cambridge Dictionary. (n. d.). Prompt. In *dictionary.cambridge.org*. Retrieved March 13, 2024, from https://dictionary.cambridge.org/dictionary/english/prompt

Cheng, W. F., Wu, C. C., Song, R., Fu, J., Xie, X., & Nie, J. Y. (2018). *Image inspired poetry generation in XiaoIce*. ArXiv. https://arxiv.org/abs/1808.03090

Compagnon, A. (1997). Chassez le style par la porte, il rentrera par la fenêtre. *Littérature*, *105*, 5-13. https://doi.org/10.3406/litt.1997.2428

Crolic, C., Thomaz, F., Hadi, R., & Stephen, A. T. (2022). Blame the bot: Anthropomorphism and anger in customer–chatbot interactions. *Journal of Marketing*, *86*(1), 132-148.

Dang, H., Mecke, L., Lehmann, F., Goller, S., & Buschek, D. (2022). *How to prompt? Opportunities and challenges of zero-and few-shot learning for human-AI interaction in creative applications of generative models*. ArXiv. https://doi.org/10.48550/arXiv.2209.01390

Ekin, S. (2023). *Prompt engineering for ChatGPT: a quick guide to techniques, tips, and best practices.* TechRxiv. https://doi.org/10.36227/techrxiv.22683919.v2

Gattupalli, S., Maloy, R. W., & Edwards, S. A. (2023). Prompt literacy: A pivotal educational skill in the age of AI. *College of Education Working Papers and Reports Series 6.* https://doi.org/10.7275/3498-wx48

Geist, U. (2005). Stylistic imitation as a tool in writing pedagogy. In G. Rijlaarsdam, H. van den Bergh, M. Couzijn (Eds.), *Effective learning and teaching of writing. Studies in writing* (vol 14). Springer. https://doi.org/10.1007/978-1-4020-2739-0_13

Ghajargar, M., Bardzell, J., & Lagerkvist, L. (2022). A redhead walks into a bar: Experiences of writing fiction with artificial intelligence. *Proceedings of the 25th international academic MindTrek conference*, 230-241.

Goldenfein, J. (2019). Algorithmic transparency and decision-making accountability: Thoughts for buying machine learning algorithms. In C. Bertram, A. Gibson & A. Nugent (Eds.), *Closer to the machine: Technical, social, and legal aspects of AI* (pp. 41-61). Office of the Victorian Information Commissioner. https://papers.ssrn.com/abstract=3445873

Hill, J., Ford, W. R., & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. *Computers in Human Behavior*, *49*, 245-250.

Hitsuwari, J., Ueda, Y., Yun, W., & Nomura, M. (2023). Does human–AI collaboration lead to more creative art? Aesthetic evaluation of human-made and AI-generated haiku poetry. *Computers in Human Behavior*, 139, 107502. https://doi.org/10.1016/j.chb.2022.107502

Jiang, H. H., Brown, L., Cheng, J., Khan, M., Gupta, A., Workman, D., Hanna, A., Flowers, J., & Gebru, T. (2023). AI Art and its Impact on Artists. *Proceedings of the 2023 AAAI/ACM conference on AI, ethics, and society*, 363–374. https://doi.org/10.1145/3600211.3604681

Köbis, N., & Mossink, L. D. (2021). Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry. *Computers in Human Behavior*, *114*. https://doi.org/10.1016/j.chb.2020.106553

Korzyński, P., Mazurek, G., Krzypkowska, P., & Kurasinski, A. (2023). Artificial intelligence prompt engineering as a new digital competence: Analysis of generative AI technologies such as ChatGPT. *Entrepreneurial Business and Economics Review*, *11*(3), 25–37. https://doi.org/10.15678/EBER.2023.110302

Learn Microsoft. (2024, March 13). *Overview of prompts*. https://learn.microsoft.com/en-us/ai-builder/prompts-overview

Leech, G. N. & Short, M. H. (1981). *Style in fiction: A linguistic guide to English fictional prose.* Longman.

Millet, K., Buehler, F., Du, G., & Kokkoris, M. D. (2023). Defending humankind: Anthropocentric bias in the appreciation of AI art. *Computers in Human Behavior*, *143*. https://doi.org/10.1016/j.chb.2023.107707

Oppenlaender, J., Linder, R., & Silvennoinen, J. (2023). *Prompting AI art: An investigation into the creative skill of prompt engineering.* ArXiv. https://doi.org/10.48550/arXiv.2303.13534

Park, M., Aiken, M., & Salvador, L. (2018). How do humans interact with chatbots?: An analysis of transcripts. *International Journal of Management and Information Technology*, *14*.

Pressman, J. D. (2024). *JD-P/simulacra-aesthetic-captions*. https://github.com/JD-P/simulacra-aesthetic-captions

Riffaterre, M. (1959). Criteria for style analysis. *Word*, *15*(1), 154-174.

Robertson, J., Ferreira, C., Botha, E., & Oosthuizen, K. (2024). Game changers: A generative AI prompt protocol to enhance human-AI knowledge co-construction. *Business Horizons*.

Santu, S. K. K., & Feng, D. (2023). *TELeR: A general taxonomy of LLM prompts for benchmarking complex tasks.* ArXiv. https://doi.org/10.48550/arXiv.2305.11430

Spillner, L., & Wenig, N. (2021). Talk to me on my level–linguistic alignment for chatbots. *Proceedings of the 23rd international conference on mobile human-computer interaction*, 1-12.

Srivastava, A., Rastogi, A., Rao, A., Shoeb, A. A. M., Abid, A., Fisch, A., … & Wang, G. (2022). *Beyond the imitation game: Quantifying and extrapolating the capabilities of language models*. ArXiv. https://doi.org/10.48550/arXiv.2206.04615

Stodola, D. (2013). Using stylistic imitation in freshman writing classes: The rhetorical and meta-rhetorical potential of transitions in Geoffrey of Vinsauf's medieval treatises. In M. Duncan & S. Medzerian Vanguri (Eds.), *The Centrality of Style* (pp. 57-69). WAC Clearinghouse. https://doi.org/10.37514/PER-B.2013.0476.2.04

Theophilou, E., Koyutürk, C., Yavari, M., Bursic, S., Donabauer, G., Telari, A., ... & Ognibene, D. (2023). *Learning to prompt in the classroom to understand AI limits: A pilot study.* ArXiv. https://doi.org/10.48550/arXiv.2307.01540

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *59*(236), 433–460.

Walter, Y. (2024). Embracing the future of Artificial Intelligence in the classroom: the relevance of AI literacy, prompt engineering, and critical thinking in modern education.

International Journal of Education Technology in Higher Education, 21. https://doi.org/10.1186/s41239-024-00448-3

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022). *Chain-of-thought prompting elicits reasoning in large language models.* ArXiv. https://doi.org/10.48550/arXiv.2201.11903

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., ... & Schmidt, D. C. (2023). *A prompt pattern catalog to enhance prompt engineering with chatgpt.* ArXiv. https://doi.org/10.48550/arXiv.2302.11382

Willey, L., White, B. J., & Deale, C. S. (2023). Teaching AI in the college course: introducing the AI prompt development life cycle (PDLC). *Issues in Information Systems*, *24*(2).

Zamfirescu-Pereira, J. D., Wong, R. Y., Hartmann, B., & Yang, Q. (2023). Why Johnny can't prompt: how non-AI experts try (and fail) to design LLM prompts. *Proceedings of the 2023 CHI conference on human factors in computing systems*, 1-21.

Zaro Vera, J. J. (1991). Literature as study and resource: the purposes of English literature teaching at university level. *Alicante Journal of English Studies / Revista Alicantina De Estudios Ingleses*, (4), 163–175. https://doi.org/10.14198/raei.1991.4.14

Zhao, T. Z., Wallace, E., Feng, S., Klein, D., & Singh, S. (2021). Calibrate before use: Improving few-shot performance of language models. *Proceedings of the 38th international conference on machine learning*, PMLR 139, 2021. https://doi.org/10.48550/arXiv.2102.09690

Yang, D., Zhou, Y., Zhang, Z., Li, T. J-J., & LC, R. (2022). AI as an Active Writer: Interaction strategies with generated text in human-AI collaborative fiction writing. In A. Smith-Renner & O. Amir (Eds.), *Joint proceedings of the IUI 2022 workshops: APEx-UI, HAI-GEN, HEALTHI, HUMANIZE, TExSS, SOCIALIZE* (pp. 56-65). http://ceur-ws.org/Vol-3124/