30th
International Maritime Lecturers Association Conference
Split Croatia
www.imla30.org

International Maritime Lecturers Association
FACULTY OF MARITIME STUDIES

SMOOTH SAILING OR ROUGH WATERS? AI as (Maritime) Evaluator

Pieter Decancq*, Alison Noble*

Abstract. Generative Artificial Intelligence (GAI) is reshaping language learning in higher education, but its ability to produce written assignments on behalf of students complicates the validity of conventional assessment by the lecturer. This study evaluates whether a comprehensive, AI-driven set of prompts can provide assessment of Maritime English projects that meets both pedagogical needs and learner expectations. After completing a group assignment, consisting of individual essays, a technical glossary and supporting documentation, first-year cadets received GAI assisted feedback and grades. Following this, they completed a questionnaire about this experience. Developing the assessment tasks and the calibration rubric was time-consuming and challenging but received a positive response from students: most respondents found the feedback useful and the scores credible, while only a minority expressed concerns about factual accuracy and calculation errors. The findings suggest that AI-assisted assessment is acceptable when used in tandem with the lecturer's experience. The article concludes with recommendations, such as incremental prompting, social desirability bias recognition and proficiency masking awareness, for integrating AI literacy into academic language curricula for both teachers and students.

Keywords: Generative Artificial Intelligence (GAI); AI Literacy; Academic Language Assessment; Incremental Prompting; Proficiency Masking; Social Desirability Bias.

Antwerp Maritime Academy, Noordkasteel Oost 6, 2030 Antwerp, Belgium

* **Corresponding authors:** pieter.decancq@hzs.be; alison.noble@hzs.be

https://doi.org/10.7225/imla30.33

1. THEORETICAL OUTLINE

In this theoretical section, we outline the historical and conceptual background of artificial intelligence (AI), briefly tracing its origins and discussing the growing need for AI literacy. We also consider both the advantages and potential drawbacks of integrating AI into education, specifically MET (Maritime Education and Training).

AI fits into the broader tendency to develop tools to ease and extend human capability. In 400 BC, the Greek already had Hephaestus create intelligent golden helpers to assist him in his smithy. Cognitive scientists often regard tool use as an evolutionary basis for logical thinking and by extension, for algorithms, the mental step-by-step plan for using a tool. Over the centuries, tools have also evolved in a more abstract sense into computer code that is able to perform calculations by means of specific instructions (algorithms), from the Turing machine in 1936, through the chatbot Eliza in the 1960s, to Deep Blue's defeat of world chess champion Garry Kasparov just before the turn of the millennium.

The increasing use of tools such as Chat GPT by first-year bachelor students in Maritime English courses illustrates the rapid integration of AI technology into everyday academic contexts, as confirmed by recent findings at Antwerp Maritime Academy where over 80% of surveyed students had already engaged with GAI for language-related tasks (Noble, Decancq & Lisaité, 2024).

It is a testament to the fact that machine learning, which enables machines to learn from (big) data, has gradually developed into deep learning (DL), in which increasingly complex patterns can be recognised using neural networks with error feedback (Rumelhart et al., 1986), to now having evolved towards Generative AI (GAI), in which *novel* content is created. Turing's 1950 speculation about 'whether machines can think' must increasingly be answered affirmatively. Harari (2024) even goes further by stating that AI is no longer a tool, but an agent: "The biggest threat of AI is that we are summoning to Earth countless new powerful agents that are potentially more intelligent and imaginative than us, and that we don't fully understand or control".

1.1. AI Literacy

The benefits of using Large Language Models (LLMs) such as Chat GPT are numerous, ranging from enhanced linguistic expression, time-efficiency, improved access to knowledge, and the generation of feedback for students. It is, therefore, logical that educators adopt GAI tools in pedagogical settings, provided they are guided by sound didactic principles. This dual use creates a level playing field and assists (language) lecturers by "allowing educators to focus on more complex teaching responsibilities with augmented capabilities" (Lee & Moore, 2024, p. 82) rather than having to indulge in laborious manual correction of AI-generated essays. In addition, in their research on automated feedback in higher education, Lee and Moore (idem.) argue that the use of GAI through direct and repetitive language support not only motivates students and improves the quality and speed of feedback but also fosters critical thinking.

Nonetheless, human supervision remains quintessential in GAI-mediated educational assessment, not only because both teachers and students run the risk of becoming overly dependent, but also, paradoxically, because GAI may not always provide accurate and verifiable information. Notably the latter phenomenon of hallucination (Maleki et al., 2024) illustrates the fact that tools such as Chat GPT are generative language machines, but should information be incorrectly entered, they can produce credible nonsense. Critical awareness when using GAI is a necessity and indicates the need for AI literacy among both students and educators.

In their extensive comparative literature study, Laupichler et al. (2022) also emphasise the urgent need to embed AI literacy in higher and adult education. However, the definition or description of the term AI literacy as such is still fluid and variably associated with 'being ready for AI', acquiring basic AI terminology or developing programming skills. Wilton et al. (2022) introduce the term AILE (Artificial Intelligence for Educators), suggesting that educators require not only literacy but also targeted training. Although Hornberger et al. (2023) focus on a small sample of German technical university students, they suggest that expanding AI-focused curricula could attract broader student interest. Mansoor et al. (2024), in their multi-country study across Asia and Africa, show that nationality and academic background significantly influence levels of AI literacy. Annapureddy et al. (2024, p.16) propose 12 core competencies for AI literacy, spanning fundamental understanding of GAI, model knowledge, practical prompt skills, contextual judgement, and ethical awareness. These aim to help users evolve "from being consumers to interpreters ... thus promoting AI as a collaborator."

The fact that GAI is a potential game changer, if it has not already become so, and that the concept of AI literacy is in full development, has also been exhaustively demonstrated in our experimental GAI-assisted language assessment research at Antwerp Maritime Academy, as discussed further below.

1.2. Parameters for Evaluation

Maritime English students in the first year of the Nautical Sciences/Marine Engineering bachelor's programme at Maritime Academy Antwerp participated in a structured group project designed to familiarise them with various aspects of the maritime industry, including shipping companies, shipbuilding, maintenance, insurance and contracts, crewing, and oceanography. The project simultaneously aimed to foster the development of research, linguistic, and cognitive competences. Through a series of group meetings, students collaboratively formulated research questions, after which each student produced an individual academic essay as a contributory part of the whole report. The group also created a collective reflection on their teamwork experience. The final submission included, in addition to the essays and group reflection, a glossary of 50 relevant maritime terms and a bibliography.

In an initial experimental phase, the essays were assessed using three parameters: Language (/10), which covered grammatical correctness, academic style and (maritime) vocabulary; Content (/10) which encompassed task alignment, authenticity, and logical flow and structure; and Length, the only subtractive parameter, consisting of text length and formatting, and excessive bulleting or overuse of headings.

In general, more detailed instruction language led to more accurate evaluations. The subcategory grammatical correctness under Language, for example, required an exhaustive checklist of 20 parameters, including sentence structure, subject-verb agreement, tense consistency, punctuation, pronoun reference, modifier placement, and idiomatic usage, among others. Without these parameters, feedback tended to be vague and overly generous. This aligns with the notion of incremental prompting (Lingard, 2023), whereby increasingly detailed instructions are necessary to elicit accurate and nuanced responses from GAI, a process which was found to be laborious yet essential during rubric development. Quite bizarrely, repeating a command or the use of capitals increased the likelihood of appropriate responses from Chat GPT. Human mediation proved essential to maintain balance.

A second interesting observation was that Chat GPT tended to overrate logical flow and structure (under Content), which is likely a result of its training on linguistic form as opposed to argumentative content. This supports the view that LLMs are inherently form-driven, with grammaticality being easier to evaluate than logical coherence. In addition, the 'eager-to-please' conversation model may denote an underlying social desirability bias (Lee et al., 2024), causing flawed reasoning to be too overtly positively assessed. Here, too, incremental prompting was necessary to refine the outcomes and thus avoid overalignment.

Similar observations were made in the third factor authenticity (under Content) as the AI tool was essentially asked to evaluate itself thus recognising 'authentic human effort' as distinct from AI-generated text. As such the question reflects the wider epistemological paradox of system inherence. To what extent is Chat GPT able to reflect upon itself (cf. Bender & Koller, 2020), in a similar manner to the biological brain that examines itself as an entity through introspection? When asked, Chat GPT replied that it can only detect human essays as 'truly human' probabilistically, for example, by looking for 'human roughness' in language use. Despite this fact, almost no single essay in these trials was flagged as completely AI-generated and this obviously raises questions: Are students using GAI merely as a support tool, or is the model itself unable to recognise its own textual footprint?

A fourth concern was that as Chat GPT often makes calculation errors when adding the individual parameters to a final result, it remains important to realise that GAI is essentially still a language model and not a calculator. Here again, the evaluator's attentiveness was indispensable.

In a second trial phase, an algorithm was inserted to standardise evaluation across approximately 150 students. For example, by adding the following simple template 'Grammatical Correctness (/6): [Insert Mark]/6; Mistakes Identified: '[Insert Mistake]' Correction: "[Insert Correction]" Issue: [Explain the grammatical error]', greater consistency in feedback and grading was achieved. However, human intervention remained necessary because excessive prompting sometimes led Chat GPT to identify increasingly more errors, to the point of linguistic absurdity.

To sum up, the two experimental phases used in creating usable evaluation schemes in academic education exposed both weaknesses and advantages of Chat GPT as an LMM. It remains vital to realise that GAI primarily predicts plausible output but does not possess real knowledge; that the tool can only reflect in a simulated sense; and that its 'understanding' is grounded in statistical patterning rather than meaning.

2. METHODOLOGY

The survey was designed with a view to gauging students' reactions to GAI used as a tool for the assessment and grading of a teamwork project

report given as an assignment in the course Maritime English 1 as part of the BSc Nautical Sciences and BSc Marine Engineering programmes at Antwerp Maritime Academy in Belgium.

A survey consisting of twenty-eight questions was created in Google Forms. Question content was designed to elicit information such as students' use of GAI for the aforementioned project and, in detail, for each part of the project. Psycho-ethical questions then aimed to gauge students' tolerance and acceptance of GAI as a language assessor while assisting the course lecturer during evaluation and grading. The final section of the questionnaire examined the accuracy and perceived credibility of GAI as a language assessor.

Some questions offered respondents the opportunity to reply in more depth by providing space for comment or elaboration. In terms of sampling, only students enrolled in the course Maritime English 1 were deemed 'knowledgeable' to complete the survey (Kumar et al., 1993). The survey was distributed via the institution's online learning platform, *Blackboard*. Access to the survey was given at a moment when (nearly) all students enrolled in the course Maritime English 1 were present in class, thereby ensuring maximum 'catchment'. Of a total of 130 students enrolled, 94 students participated in the survey. Respondents and comments were subsequently assigned codes to facilitate categorization and analysis of data.

3. ANALYSIS OF THE DATA

The findings of this research build on research by Noble, Decancq and Lisaité (2024) who provide a broad preliminary examination of the role of AI in Maritime English education. By contrast, the data accumulated during this project are more specific, relating to one assignment, namely the aforementioned Teamwork Project report, and specifically the use of GAI to assist with the assessment of this task. As demonstrated in this section, the analysis of the data includes both quantitative and qualitative elements, as was the case with the previous research (idem., 2024).

A total of 94 first year students, enrolled in the course Maritime English 1 within the BSc programmes Nautical Sciences and Marine Engineering, responded to the questionnaire. It was not deemed necessary to ascertain the identity of the individual respondents to the questionnaire and participation in the survey thus remained anonymous. Nevertheless, a minimum of microdata was collected for statistical purposes at questions 1 and 2.

In reply to Q1 (*In which programme are you enrolled?*) the largest number 45 (47%) prove to be students enrolled in the Dutch taught BSc programme for *Nautische Wetenschappen* (Nautical Sciences). This group is followed by 35 students (37%) enrolled in the equivalent programme but taught in French, namely *Sciences Nautiques* (Nautical Sciences). A much lower number at 9 (9%) and 5 (5%) are enrolled in, respectively, Dutch taught BSc *Scheepswerktuigkunde* (Marine Engineering) and French taught *Mécanique Navale* (Marine Engineering). There is nothing of particular note in these findings given that they simply reflect and corroborate the profile of the student population at Antwerp Maritime Academy.

Q2 is related to gender identification. Of the respondent cohort, 73 (77%) describe themselves as 'male' and 21 (22%) as 'female'. Interestingly, although not pertinent to this study, no students describe themselves as 'other' and none would "rather not say". These results also mirror statistics related to gender at Antwerp Maritime Academy.

The questionnaire then moves on to individual use of GAI (Chat GPT) for the specific assignment Teamwork Project Report. In reply to Q3 "*Did you use generative AI (Chat GPT) to assist you in any way with producing the Teamwork Project Report?*" 72 (76%) participants in the survey state that they did make use of GAI, whilst 22 (23%) respond that they did not. This corresponds approximately to ¾ of the respondents having used GAI and only ¼ not.

When asked to explain in more detail why they did not use GAI, students' responses vary. Some reflect a fierce independence and a desire to remain self-sufficient rather than resorting to generative tools, observing *"I don't like AI, I don't think it helps us, we're becoming lazy while using this. We don't make our brain work, we don't use our critical skills and it kind of kills media literacy because we don't do our own research anymore".*

Others hint at the possibly detrimental legal and statutory implications of using the tool, with remarks such as "*I didn't think it was fair, for me it was not legit"* or "*because it can be considered as plagiarism".*

Some comments question the value, validity and accuracy of a tool such as Chat GPT, writing for example, "*I also remember to not really be satisfied with the results CHAT GPT gave me when I did use it in the past"* or *"I do not think that generative AI can help me with this type of task as my previous experience learned* [sic] *me that most information is incorrect and very general. I feel that I'm better at writing such a small text based upon scientific papers from which information is subtracted correctly".*

There are also students who mention environmental concerns as the reason for not using GAI, stating "*I am very concerned with the state of our planet, and that*

is why I don't like to use AI. I have used in the past but now that I know the impact of it, I don't anymore" and *"it's really bad for the environment, each research* (sic) *on Chat GPT uses a lot of water to keep the servers working."*

Figure 1 below, relating to Q5, illustrates the ways in which GAI (Chat GPT) was used by students to create this assignment. It appears that finding ideas (brainstorming) is popular, whilst employing the tool as a vocabulary assistant also scores highly. Undoubtedly, when using GAI (Chat GPT) as a language tool, correcting grammar and generally enhancing use of the English language is also revealed as common practice.

In addition, the data shows 16 students (17%) who state that they did not use GAI (Chat GPT). This is lower than the 22 students (23%) recorded at Q3 as not having used the tool. However, 7 (7%) state that they used 'other' means to create the assignment. When asked to elaborate (Q6), answers revealed that these 7 had, in fact, not used GAI (Chat GPT). This would account for the initial statistic of 23%.

For the purposes of assessment and to achieve the cumulative total mark, the Teamwork Project Report is divided into four sections, namely the essay, the glossary, the list of references (bibliography) and the administrative report. Of these four sections, only the essay is produced on an individual basis, by each participating student.

The next four questions (Q7-Q10) in the survey were designed to gauge to what extent the student used GAI to create each of these four sections. The data is similar for all four items, revealing that for the glossary, administrative report and bibliography over 90% of the respondents gave their estimated use of GAI as less than 50%. The individual essay, perhaps predictably, generated slightly different data.

Figure 2 illustrates this, where 88.2% replied that they used GAI to an extent of 40% or less when creating the task and 11.8%, the highest figure when compared to the other three tasks, replied that they used it for between 50% and 80% when creating the assignment.

Q11 "To what extent do you agree with the statement "*My experience is that preparing effective prompts for Chat GPT is very time consuming*" reveals that students tend to disagree with this, with 36 (38%) disagreeing, whilst only 13 (13%) agree with the statement. The remainder of respondents prefer to remain neutral. This is illustrated in Figure 3 below. Given the authors' challenging experience when attempting to create prompts for Chat GPT that produce the desired result, the question was designed

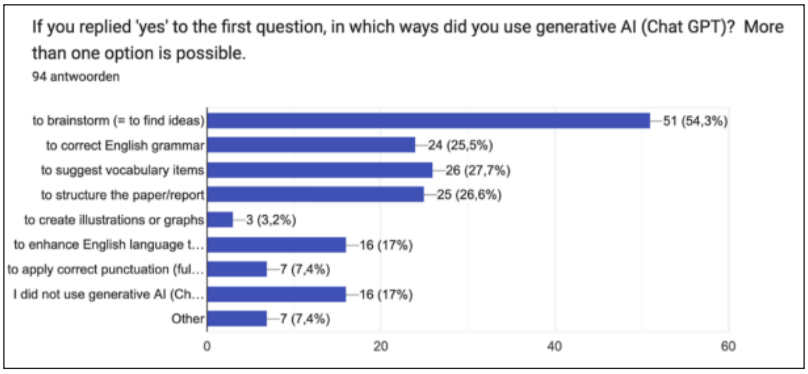

Figure 1. Q5 Use of GAI for the Teamwork Project report. Source: authors' own.

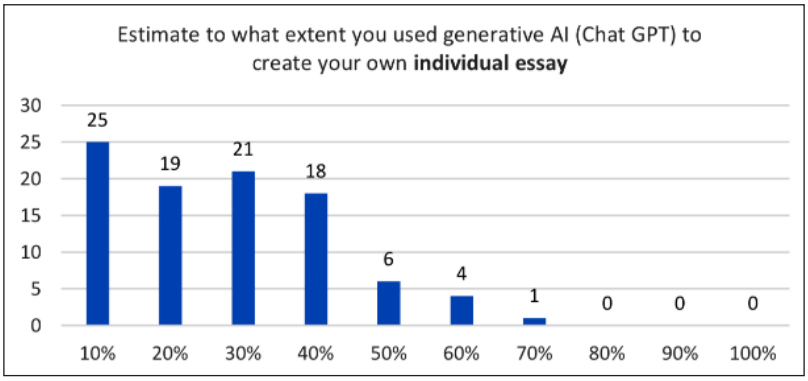

Figure 2. Q7 Self-estimated use of GAI for **individual essay**. Source: authors' own.

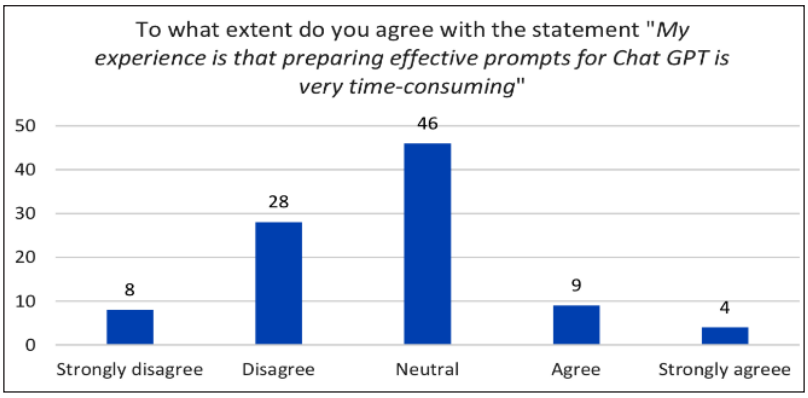

Figure 3. Q11 Preparing GAI prompts Source: authors' own.

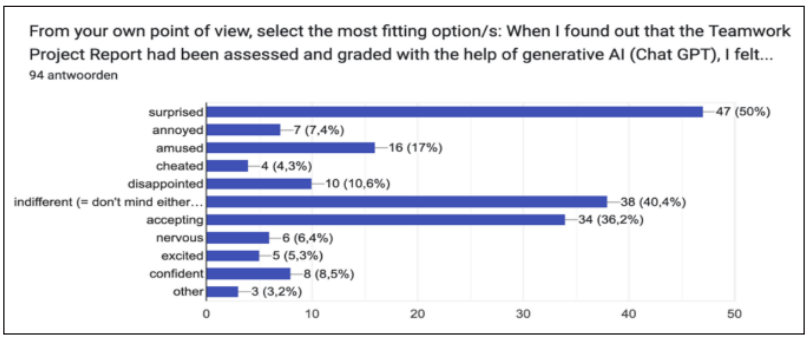

Figure 4. Q12 Student reaction to GAI as language assessor. Source: authors' own.

to gauge students' perception of this. The results here are somewhat contrary to expectations.

Following the collection of microdata, the survey moved on to explore the respondents' acceptance of GAI (Chat GPT) as a language assessor. Respondents were asked how they felt upon discovering that the Teamwork Project Report had been assessed and graded with the help of GAI (Chat GPT). Respondents were given a range of ten options (shown in Figure 4) from which they could select one. The option 'other' was also provided. If 'other' were selected, students were asked to describe how they felt.

The majority of respondents, 47 (50%) express 'surprise' followed by those who declare themselves 'indifferent' (38 or 40%), followed by respondents who feel 'accepting' (34 or 36%). A relatively large number feel 'amused' (16 or 17%) before the figures drop to under 10% for the remaining options. It is fair to state that respondents experiencing positive feelings ('accepting', 'confident', 'excited', 'amused') outweigh those who experience negative thoughts ('nervous', 'annoyed', 'cheated').

Respondents are shown to disagree with the statement at Q14, illustrated in Figure 5 below. This asked

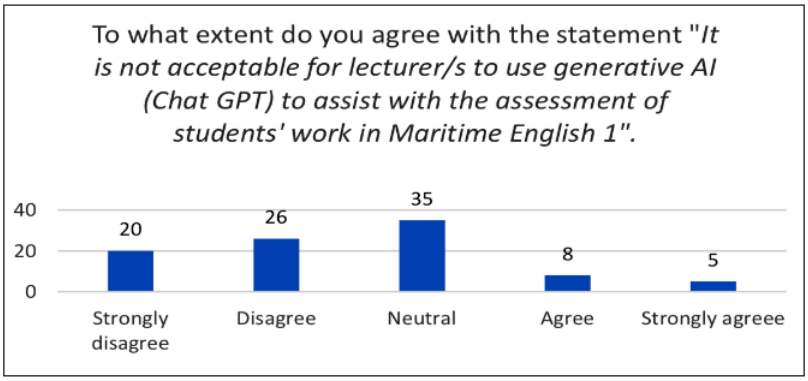

Figure 5. Q14 Acceptance of lecturer use of GAI as language assessor. Source: authors' own.

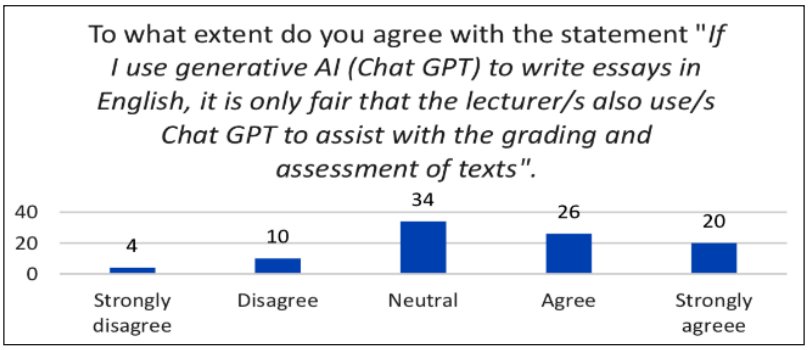

Figure 6. Q15 Fairness of lecturer use of GAI as language assessor. Source: authors' own.

whether respondents were opposed to the idea that lecturers use GAI (Chat GPT) to assist with the assessment of students work in Maritime English 1. 46 (49%) disagree and 13 (13%) agree, ultimately meaning that it is considered acceptable for lecturers to use the tool for assistance with assessment.

To counterbalance Q14, Q15 asked to what extent respondents agreed that if *they* use GAI (Chat GPT) to write essays in English, it could only be considered fair that lecturer/s also use the tool to assist with the grading and assessment of texts. Once again, as shown in the adjacent Figure 6 above, students display a positive approach – and a sense of fairness! – by agreeing with this statement. 20 (21%) strongly agree whilst 26 (27%) agree, bringing the total who agree to 48%. By contrast a total of only 14 (14%) disagree with the statement. As is the case with the majority of questions, the majority prefer to remain neutral, neither agreeing nor disagreeing.

The next question, Q16, probes the extent to which students feel GAI (Chat GPT) can be trusted or not to assess and grade their texts. The statement is as follows: *"I do not trust generative AI (Chat GPT) to correct, assess and grade my written work in Maritime English 1"*. Although many respondents (39 or 41.5%) once again take a neutral stance, those agreeing with the statement (29 or 30.8%) slightly outweigh those disagreeing (26 or 27,7%). Opinion would seem to be divided on this matter. Figure 7, as follows, illustrates the outcome.

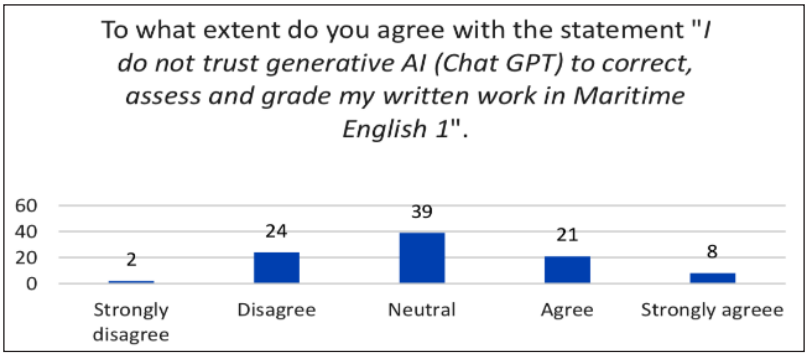

Figure 7. Q16 Trust in GAI as language assessor. Source: authors' own.

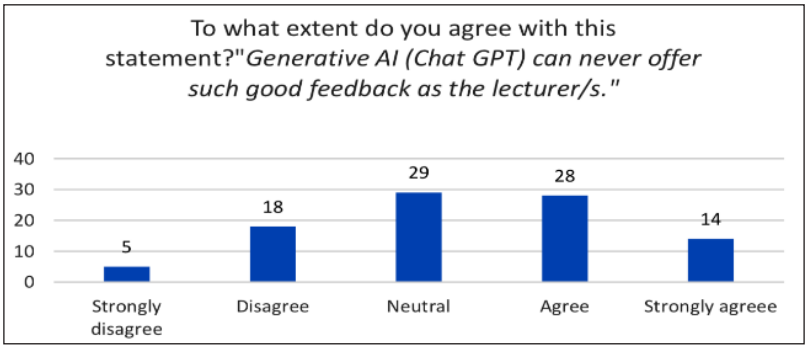

Figure 8. Q17 Appreciation of AI feedback. Source: authors' own.

Expanding on the matter of trust, respondents were asked in Q17 whether they agreed that GAI can never offer such good feedback as the lecturer. In this particular case, a relatively lower number of respondents 29 (30.8%) remain undecided (neutral), whilst 42 (44.6%) agree with the statement and 23 (24.4%) disagree. Ultimately – although perhaps simplistically - this would seem encouraging for the future of Maritime English lecturers. Figure 8 provides the details.

Following Q17 students were asked to elaborate. For some the reply to Q17 proved a gut reaction, with one respondent stating, "*I don't really know, it's just a feeling*". Others mentioned the human factor with comments such as "*…it is useful to have feedback from lecturers because it offers a human point of view*" and "*AI does not know you in person. The lecturer does and can see your progress in English*" or "*a lecturer is aware of the real-life situation we are in*".

Many respondents returned to the accuracy and trustworthiness of GAI (Chat GPT) with comments such as, "*I think AI is a new tool which can help us, but we still need to be careful because AI can still make mistakes*" and "*AI is more prone to make constant mistakes, a teacher might make some small ones. I do believe it can be used to help correcting, but never as the final result.*"

It should be noted that many were optimistic about the combined effort of lecturer and GAI together, stating "*generative AI sees more mistakes than the lecturer, although it is possible that AI has another way of reviewing a text so I think the lecturer may use generative AI but that s/he should still use his/her evaluation criteria, not those of AI.*" Another example runs as follows: "*AI is the perfect assistant for student/lecturer. Only if the AI is an assistant and not the other way round. I don't mind AI assessing my paper and giving me a score, but I feel more comfortable if the lecturer checks for correctness.*"

Some mentioned the expertise required to generate the prompt: "*Depends on how the prompt is written. If there is a good prompt and ChatGPT has some criteria given by the lecturer, I think this can help a lot.*"

The final section of the questionnaire examines the accuracy of GAI (Chat GPT) as language assessor. Q19 asked to what extent respondents agreed that they understand the AI-generated feedback provided to them. The result is resounding, with 54 (57.4%) agreeing that they understand the feedback. Those disagreeing number only 12 (12.7%) with 28 (29.4%) neither agreeing or disagreeing. Figure 9 illustrates the results.

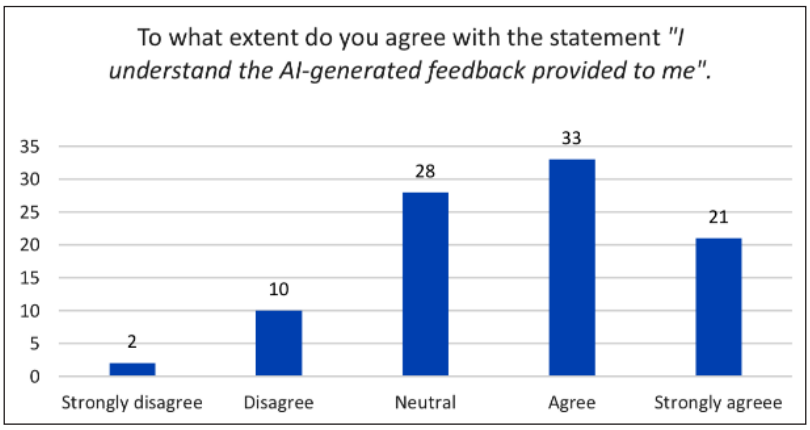

Figure 9. Q19 Understanding of GAI feedback. Source: authors' own.

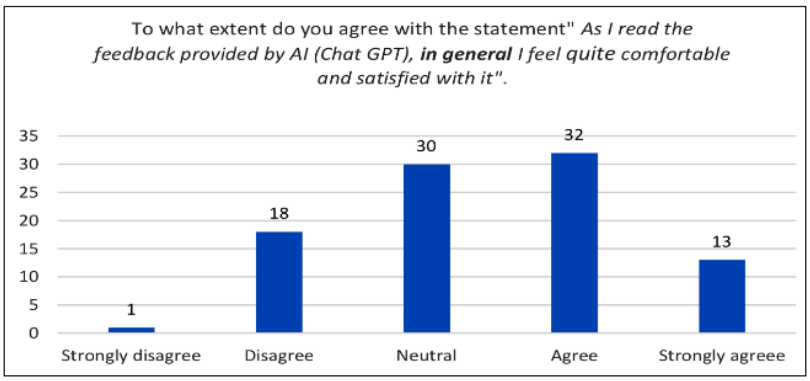

Figure 10. Q20 Satisfaction with GAI feedback. Source: authors' own.

Q20 gauged whether respondents felt comfortable and satisfied as they read the feedback provided by GAI (Chat GPT) **in general**. Once again those agreeing with the statement (45 or 47.8%) outnumber those disagreeing (12 or 12.7%), as illustrated in Figure 10. 28 (29.7%) remain undecided (neutral).

As with Q7 to Q10, the next four questions Q21 to Q24 address each of the different sections within the Teamwork Project Report separately. Each question follows the same format: To what extent do you agree with this statement *"As I read the feedback provided by generative AI (Chat GPT) on **my individual essay-glossary-bibliography-administrative report**, I feel quite comfortable and satisfied with it"*. As with Q7-S10, the results here prove similar. In all four sets of data between 30 and 35 respondents chose each time neither to agree or disagree, selecting 'neutral'. 44 (46.8%) respondents agree that they feel comfortable and satisfied when reading the feedback for the individual essay provided by GAI (Chat GPT) while 15 (15.9%) disagree. For the glossary 49 (52.1%) agree and 11 (11.7%) disagree, for the bibliography 45 (47.8%) agree and 13 (13.8%) disagree and for the administrative report 45 (47.8%) agree and 19 (20.2%) disagree.

When asked if the total score awarded by GAI (Chat GPT) was what they expected, 58 (61.7%) of respondents replied 'yes', as shown in Figure 11. Of those answering otherwise, 20 (21.3%) replied that they had expected a higher score whilst 16 (17%) replied that they had been expecting a lower score.

Respondents were subsequently asked to elaborate on their choice regarding their expectations. Those stating that the result is what they expected explain variously that they "*worked hard*" and "*did a lot of research*" or spent time on the project. Some of those who claim that the result did not meet their expectations – they expected a higher score - explain that they omitted part of the assignment, for example "*We didn't create a bibliography as a distinct part, so AI didn't take this into consideration, and we have a 0 for it. And I think it's quite unfair.*" Others return to the unreliability of GAI with comments such as "*I feel like the AI was really focused on the vocabulary and made some weird decisions*" and *"I feel like there is a rather large difference in its ability to correct, sometimes I totally agree, but then there are things that make no sense and are not useful."* Those who expected a lower score were, naturally, satisfied. "*I was surprised when I saw my score. Normally I have a lower score*".

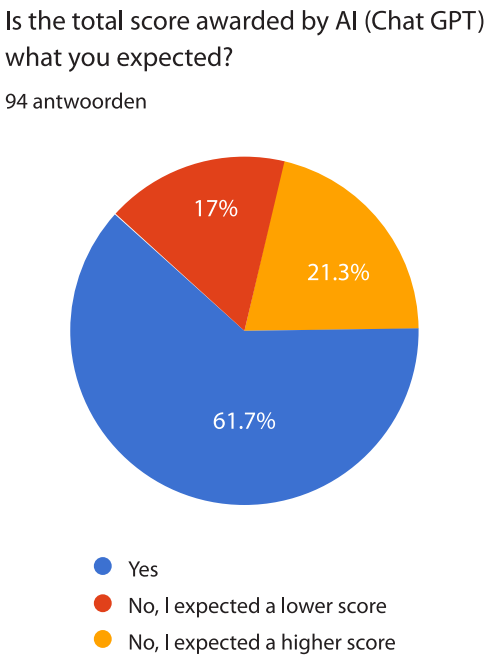

Figure 11. Q25 Expectations of GAI grading. Source: authors' own.

Respondents were then asked to what extent they agreed that as regards accuracy of language (grammar, syntax, punctuation, spelling, etc.), the correction provided by GAI (Chat GPT) seemed fair. 44 (46.8%) respondents agree that the correction seems to be fair, whereas 13 (13.8%) disagree. A relatively large number of respondents, 37 (39.3%) prefer to remain neutral regarding this question. Figure 12 below illustrates the data.

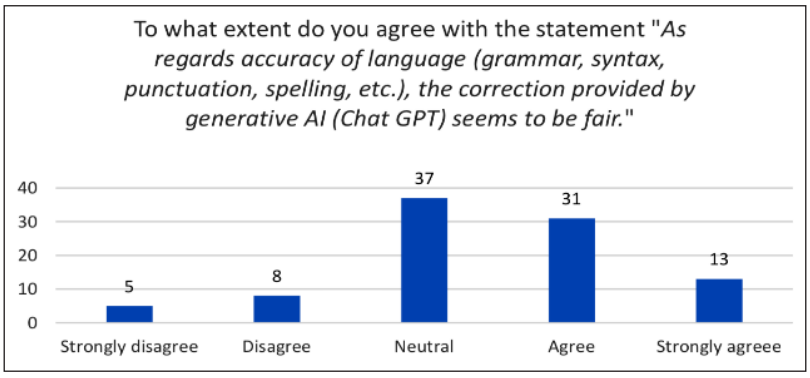

Figure 12. Q27 Perceived fairness of GAI language assessment. Source: authors' own.

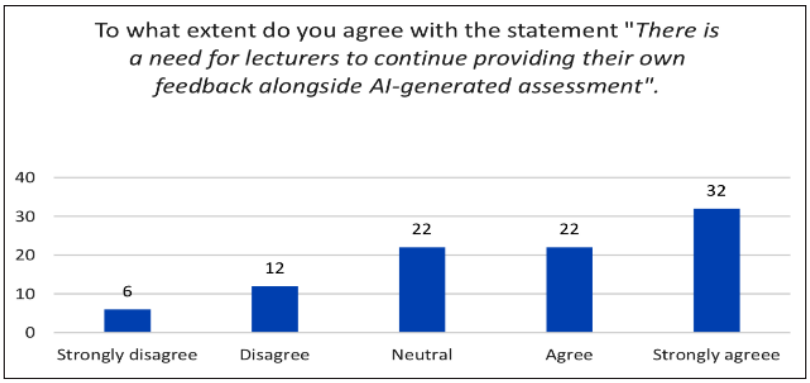

Figure 13. Q28 Need for continuing feedback from lecturer. Source: authors' own.

At the centre of the final question is the notion that AI-generated assessment might replace lecturers' feedback. Respondents were asked to what extent they agreed with the statement "There is a need for lecturers to continue providing their own feedback alongside AI-generated assessment". In Figure 13, considerably fewer respondents (only 22 or 23.4%) remain neutral in this instance, preferring to agree or disagree. A resounding 32 (34%) strongly agree that there is a need for lecturers to continue giving feedback. In total 54 (57.4%) respondents agree overall. Only 18 (19.1%) disagree.

When asked to explain their choice, a great many respondents emphasise the advantage of AI in combination with the lecturer: "*It is necessary to have a look at what AI can do, because AI can make mistakes or doesn't understand some things so it's the role of the human to supervise it*" and "*Chat GPT is good but the feedback of a lecturer is better. The combination of the two is the best.*" The message seems to be clear: *"Don't let the machine's take over".*

4. DISCUSSION

The questionnaire administered to first-year cadets in Maritime English revealed insights clustered around three themes: students' deployment of GAI in the specific assignment Teamwork Project Report; their reaction to lecturers' GAI use as an assisting tool for language assessment; and their perception of AI-mediated feedback.

4.1. Use of GAI for the Teamwork Project report

Roughly three-quarters of respondents reported using GAI for the teamwork project report, mainly for brainstorming, vocabulary, and grammatical support. Use peaked in the essay component, where a small but notable subgroup consulted the tool for more than 50% of the writing process; by contrast, usage for the glossary, bibliography and administrative report remained below 50% for most students. Non-users mentioned doubts about accuracy, concerns about academic integrity (*"for me it was not legit"*), environmental impact (*"it's really bad for the environment"*) and a wish to remain independent (*"we're becoming lazy"*). Such reservations may reflect in some sense the warning of Harari (2024), negatively framing AI as an agent but might also suggest gaps and uncertainties as regards AI literacy.

A further asymmetry emerged: students readily adopted GAI for cognitively demanding essay work yet were more hesitant for procedural tasks, despite the latter being well-suited to automation. Targeted AI literacy training such as incremental prompting (Lingard, 2023) could remedy this discrepancy.

A final but important consideration is proficiency masking, whereby students make themselves appear more proficient in language tasks than they actually are, often thanks to the help of GAI tools. AMA's open-entry policy yields linguistic heterogeneity (A1–C1) so beginners could submit C1-like project texts using GAI, yet would still underperform in traditional examinations, where the demand is for unassisted writing. Language lecturers at AMA therefore allow the use of GAI as an informed tool for preparation but assess summative competence with pen-and-paper tests, a successful policy thus far.

4.2. Perceptions of Lecturer GAI Use

When informed that GAI had contributed to grading, 50% of students felt "surprised", indifference (40%) and acceptance (36%). Positive reactions ('accepting', 'confident', 'excited', 'amused') clearly outweighed negative reactions ('nervous', 'annoyed', 'cheated'); 48% judged it "fair" that lecturers make use of GAI if students do likewise, provided human overview remains. This reciprocity, however, could challenge conventional notions of academic integrity and raises the question of whether AI should be considered as a third pedagogical actor.

However, transparency is pivotal, certainly in education. Several respondents expressed discomfort at GAI involvement, which, on the other hand, implies that explicitly incorporating the use of GAI in the project guidelines could enhance trust and foster a more critical engagement with AI tools in general.

4.3. Experience of GAI Feedback

Approximately half of the cohort rated AI feedback both intelligible and satisfactory; 60% received the marks they expected, though a minority cited calculation errors. Such anomalies underscore the need to embed factors such as hallucination (Maleki et al., 2024) awareness and knowledge of algorithms into AI literacy for language curricula.

While explicit comments on logical coherence were limited, educators noted GAI's tendency to overrate structure, reminiscent of its structural social desirability bias, necessitating frequent incremental prompts by the lecturer. This resonates with a student's comment that it *"depends on how the prompt is written. If there is a good prompt and ChatGPT has some criteria given by the lecturer, I think this can help a lot"*. As the model self-updates, this limitation may be mitigated, yet vigilant human mediation remains essential.

A majority (57%) confirmed the necessity of lecturer feedback alongside AI output. Interestingly, students also asked more follow-up questions about their individual essays (marked 20/50) than about group components (marked 30/50), probably reflecting greater personal interest in the essay rather than the group components, despite the heavier weight of the latter.

Collectively, these findings advocate a hybrid evaluative model, as a clear majority of respondents expressed the need for human feedback as final evaluator (*"it is useful to have feedback from lecturers because it offers a human point of view"*). Annapureddy et al. (2024) competencies perceive users progressing from consumers to informed interpreters of GAI. This is borne out by the survey results whereby GAI may function as co-creator but lecturers maintain ultimate authority (*"AI is the perfect assistant for student/lecturer. Only if the AI is an assistant and not the other way round (sic)"*). We are left, perhaps, with an analogy: GAI as helmsman, but language lecturer still as Captain of the ship!

5. CONCLUSION

The use of GAI in language assessment in Maritime English has clear educational value, but it is important to realise that human monitoring remains essential. It remains important to acknowledge that GAI as an LLM generates likely responses without true understanding (cf. social desirability bias as well as system inherence), relying on statistical patterns rather than genuine knowledge or reflective reasoning. Overall, the GAI assessment of the group project was viewed positively, but at the same time, students were critical of its shortcomings.

Bearing Harari's (2024) warning in mind, the integration of GAI must proceed hand in hand with a systematic enhancement of AI literacy among both students and lecturers. This applies not only to ethical considerations, as Noble et al. (2024) point out, or contextual appropriateness; it also entails practical competences such as formulating precise, incrementally refined prompts, thereby enabling more reliable output. Moreover, realistic language competence such as writing must continue to be gauged through summative pen-and-paper tests in order to reduce language proficiency masking. After all, language skill sets remain crucial within the broader framework of a reliable maritime industry and safe shipping.

Cultivating this kind of AI literacy can lead to a fascinating and creative dialogue between the human brain and tool, as mythic Hephaestus and his intelligent beings discovered centuries ago. AI literacy will undoubtedly gradually permeate the entire educational community, with GAI being seen as a co-creator, i. e. a collaborative partner that can generate ideas, refine feedback and provide support for the benefit of all parties. This hybrid form of collaboration and training in cooperation with human critical attention will hopefully ensure that GAI becomes an amplifier of human thinking rather than a replacement for it. As long as the lecturer remains the captain with GAI at the helm, rough waters may be circumnavigated and smooth sailing guaranteed.

Conflicts of interest: All authors declare that they have no conflicts of interest.

REFERENCES

Annapureddy, H. V., Fornaroli, S., & Gatica-Perez, D., 2024. Generative AI literacy: Twelve defining competencies. arXiv. Available at: https://arxiv.org/abs/2412.12107

Bender, E. M., & Koller, A., 2020. Climbing towards NLU: On meaning, form, and understanding in the age of data. In D. Jurafsky, J. Chai, N. Schluter, & J. Tetreault (Eds.), Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (pp. 5185–5198). Association for Computational Linguistics. Available at: https://doi.org/10.18653/v1/2020.acl-main.463

Haenlein, M., & Kaplan, A., 2019. A brief history of artificial intelligence: On the past, present, and future of artificial in-

telligence. California Management Review, *61*(4), 5–14. Available at: https://doi.org/10.1177/0008125619864925

Harari, Y. N., 2024, 24 August. *"Never summon a power you can't control"*: Yuval Noah Harari on how AI could threaten democracy and divide the world. The Guardian. Available at: https://www.theguardian.com/technology/article/2024/aug/24/yuval-noah-harari-ai-book-extract-nexus

Holm, J. R., Peltonen, J., Timmermans, B., & Henning, M., 2024. Academic teaching: Local university policies for generative AI and students' use of generative AI as a personalized tutor.

Hornberger, M., Bewersdorff, A., Nerdel, C. What do university students know about Artificial Intelligence? Development and validation of an AI literacy test, Computers and Education: Artificial Intelligence, Volume 5, 2023, 100165, ISSN 2666-920X. Available at: https://doi.org/10.1016/j.caeai.2023.100165.

Kumar N, Stern L, Anderson J., 1993. Conducting interorganisational research using key informants. Academy of Management Journal 36/6: 1633-1651

Laupichler, M. C., Aster, A., Schirch, J., & Raupach, T., 2022. Artificial intelligence literacy in higher and adult education: A scoping literature review. Computers and Education: Artificial Intelligence, *3*(4), 100101. Available at: https://doi.org/10.1016/j.caeai.2022.100101

Lee, H., & Moore, J., 2024. *Harnessing generative AI (GenAI) for automated feedback in higher education: A systematic review*. ERIC. Available at: https://files.eric.ed.gov/fulltext/EJ1446868.pdf

Lee, S., Yang, K.-Q., Peng, T.-Q., Heo, R., & Liu, H., 2024. Exploring social desirability response bias in large language models: Evidence from GPT-4 simulations. arXiv.

Lingard L., Writing with ChatGPT: An Illustration of its Capacity, Limitations & Implications for Academic Writers. Perspectives on Medical Education. 2023; 12(1): 261–270.

Maleki, N., Padmanabhan, B., & Dutta, K., 2024. *AI Hallucinations: A Misnomer Worth Clarifying*. IEEE CAI.

Mansoor, H. M. H., Bawazir, A., Alsabri, M. A., Alharbi, A., & Okela, A. H., 2024. Artificial intelligence literacy among university students—a comparative transnational survey. Frontiers in Communication, 9, Article 1478476. Available at: https://doi.org/10.3389/fcomm.2024.1478476

Noble, A., Decancq, P. & Lisaité, D., 2024. "*Sometimes ChatGPT is as confused as I am*: a preliminary examination of the role of AI in Maritime English education". In: Proceedings of 29th International Maritime Lecturers Association (IMLA) conference pp. 174-180. ISBN: 978-625-98269-8-1

Rumelhart, D. E., Hinton, G. E., & Williams, R. J., 1986. Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536. Available at: https://doi.org/10.1038/323533a0

Wilton, L., Ip, S., Sharma, M., Fan, F., 2022. Where Is the AI? AI Literacy for Educators. In: Rodrigo, M.M., Matsuda, N., Cristea, A.I., Dimitrova, V. (eds) Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners' and Doctoral Consortium. AIED 2022. Lecture Notes in Computer Science, vol 13356. Springer, Cham. Available at: https://doi.org/10.1007/978-3-031-11647-6_31

Zhou, L., & Tsang, S. J., 2025. Adolescents' Understanding of AI: Challenges and Opportunities for AI Literacy Education. Paper presented at the 75th Annual International Communication Association Conference, ICA 2025, Denver, Colorado, United States.