How do teachers envision AI grading for open-ended questions in universities?

ELIZAVETA STASHEVSKAIA, University of Twente, The Netherlands

ABSTRACT

AI is becoming increasingly more prominent in the vast majority of industries across the world and its inclusion within education can be expected. AI in education can be used for numerous functions from planning to enrollment trend predictions to grading. Whilst teachers' thoughts on AI in education have been discussed by existing literature, their opinions and preferences specifically on AI grading in universities have yet to be analysed. It was interesting to understand what assignments university teachers would use AI grading for and what impact that would have on their relationship with their students. To achieve this a series of interviews and a survey were conducted with teachers from the University of Twente. The study concluded that : (1) the majority of the participating teachers could imagine AI becoming a part of the grading process in university education mainly, due to the time-saving benefits it may offer to busy teachers. (2) AI is likely best suited for factual assignments which do not require creativity or a judgment call from the grader. (3) the impact that the introduction of AI would have on student-teacher relationships will be highly dependent on the AI's application. Specifically, how transparent teachers and universities will be with using AI and how well students understand as well as accept this technology. (4) no matter how AI is used during the grading process, the teacher must always remain the one responsible for a student's grade.

Keywords: AI grading, open-ended questions, teachers, trust, assignment type, student-teacher relationship

1 INTRODUCTION

The capabilities of AI continue to improve at a rapid pace as it becomes exceedingly more useful for various real-world applications [2]. With this, its increased involvement in university environments only becomes more inevitable. Automated grading for multiple-choice questions (MCQs) has been around since 1966 and is still widely used today [5]. Although MCQs are notorious for not fully examining a student's comprehension of the material, the speed at which they can be corrected often makes them an attractive method for teachers and examiners alike [3]. This study hypothesised that a good implementation of AI grading for open-ended questions could provide significant help to busy teachers, potentially allow students to gain feedback on their work more frequently and faster as well as improve the student's understanding of the tested materials. However, few of these benefits would be obtained unless AI grading for open-ended questions is accepted by teachers. Therefore, this study aims to determine how teachers would use AI grading tools and what potential ramifications will follow its introduction into university-level education.

TScIT 42, July 5, 2024, Enschede, The Netherlands

2 PROBLEM STATEMENT & RESEARCH QUESTIONS

The use of AI in education is not a new topic and a multitude of research has been done about its potential benefits and drawbacks, studies like [6][17][11] intently explore AI's future in education. Likewise, teachers' perception of AI in education has also been captured by several studies, for example, [14]. Flaws of AI grading for open-ended questions have also been meticulously discussed by multiple studies such as [8][22][18]. The previously done research, however, seems to focus either on AI in education as a whole and teachers' attitudes towards it or on AI grading and its acceptance by students. Hence this research raised the question: To what extent do university teachers accept AI grading for open-ended questions? In this research, we take acceptance to mean feelings of trust and openness towards the AI grading model. Similarly, when using the term AI grading in this study, we mean specifically AI grading for open-ended questions. The proposed research question and its sub-questions aim to analyse how university teachers would utilise AI-driven grading and which assignments are best suited for it.

Main research question (RQ): To what extent are teachers open to AI-driven grading in universities?

Sub-research question 1 (SRQ1): What type of assignments/questions are suitable for AI grading?

Sub-research question 2 (SRQ2): What impact will the use of AI grading have on student-teacher relationships?

3 METHODOLOGY

To sufficiently answer the research questions, it required a comprehensive review of existing literature as well as the analyses of newly obtained data through a survey and interviews.

3.1 Literature review

A standard literature review was performed to summarise and gauge an understanding of previously done studies on this and closely related topics [10] such as AI grading in general and its perception by involved parties. Based on the reviewed literature certain gaps in existing research were identified and served as a basis for creating survey and interview questions.

3.2 Survey & interviews

A survey and interviews were both conducted for data collection. A survey was chosen as it had the potential to reach a wider variety of teaching staff from different programs than interviews could.

It was assumed that a large number of teachers would have been more willing to fill out a quick survey in their own time rather than planning and attending an interview. To ensure survey and interviews were performed ethically, approval for the study was sought from the University of Twente's Ethical Committee.

3.2.1 *Survey planning.*

To draw valuable and accurate conclusions from the survey results it was important to try and ensure that only teachers filled in the survey. Knowing if the targeted demographic is actually the one responding to an online survey can often be a challenge [19], for that reason careful considerations were made into how the surveys were distributed. The survey was distributed through QR code posters in teacher break rooms as well as in teacher-accessible channels within the University of Twente (UT), such as dedicated department Microsoft Teams groups and email groups. The posters clearly addressed teachers and the first query of the survey asked for the participant's teaching program, both of which were meant to deter people outside of the target demographic from participating in the survey.

3.2.2 *Interview planning.*

Beyond the survey, interviews were conducted as they were presumed to provide deeper insights and details into the subject's responses [20]. The survey and interview groups had no intended overlap. The short interviews took place either in person or online as per the interviewee's preference.

4 LITERATURE REVIEW RESULTS

By using a series of search terms outlined in Table 1 a multitude of papers were found and organized into categories related to this particular study. These include AI in education as a whole, teachers' perception of AI in education, the (dis)advantages of using AI grading, and students' attitudes towards it in addition to elements which contribute to technology acceptance.

Table 1. Search terms used for literature review

Primary term	Secondary term	Query
AI grading	universities teachers' opinion education trust	AI grading AND (universities OR teachers' opinion OR education) OR trust
Technology acceptance	model AI grading	technology acceptance AND (model OR AI grading)

4.1 AI in education

[6] discusses the benefits of AI in education, this includes smart scheduling, predicting dropout rates, grading and others. The paper draws attention to teachers still being central for a successful education with AI merely taking on a helping role to reduce teachers' workload [6]. In like manner, [17] also highlights more potential uses of AI than grading such as creating specialized exercises for students based on knowledge gaps, personalised practice based on student's learning methods and creation of questions based on specific given topics. Such ideas are further supported by similar research in [11]. These provide a clear idea of the potential that AI has in education.

4.2 Teacher's attitude towards AI in education

[14] examined the perception of teachers on using AI for higher education. Across the discussed cases, over 80% of teachers stated that "Universities should use new technologies to address individual student's learning needs", however, only about 20% agreed that "University teachers have a responsibility to allocate teaching time to work with a learning analyst or AI tools to facilitate the student's learning" [14]. Despite the vast majority of teachers believing that new technology can be used to help individual students, only one-fifth would be willing to spend their teaching time aiding the development of such new technologies, this formulates the question as to why this is. This discovery helped formulate some of the survey and interview questions which aimed to gain further insights into what may be stopping teachers from working with AI.

4.3 Benefits and drawbacks of AI grading

Focusing specifically on assessment, [12] demonstrates the advantages and disadvantages of using AI grading. The former includes saving time, consistent student feedback and convenience, whilst the latter includes inadequate quality, ethical concerns as well as career implications, such as the replacement of some teaching staff by AI [12]. This study predicts that by discussing in detail the potential disadvantages of AI grading with teachers, methods to overcome them can be identified and used in the development of AI grading models.

An exploratory study by [4] looked into the general acceptance of AI assessments from both the students and the teachers' perspectives. The results of the teacher's responses in [4] showcase some similar results to [12], teachers mentioned the positive impact AI grading may have on time management but also the career consequences it may have on teachers if the AI can replace some (supporting) staff [4]. A couple of further insights made by [4] were that teachers envisioned the AI in a supportive role to them and that they would not use it for all types of assessments. These discoveries made by [4] are precisely the foundation of our study, which aims to determine what types of assessments teachers would use AI grading for and how they would remain in the grading process along the AI.

4.4 Student's opinions on AI grading

Examining the perception of AI grading by students, it can be observed that students have concerns such as bias, exploitation and transparency but they also acknowledge benefits such as faster grading and providing more educational support [18]. [7] also found that students perceive AI grading as fairer than that which is done

by teachers and the more transparent the AI grading algorithm is, the fairer it will be deemed by students.

4.5 Technology acceptance

As AI grading for open-ended questions is a relatively new technology that is not yet widely used, it was imperative to examine which factors generally contribute to the acceptance of emerging technologies. The technology acceptance model (TAM) attempts to explain and predict users' behaviour towards a new system [13], studying this model provided a better understanding of what users expect from a new technology and directly helped in answering the main research question (RQ).

TAM was first introduced by Davis in 1985 and then refined into a second version of the model [13], which depicts the theorized correlation between the user's perceptions and their acceptance of the technology.

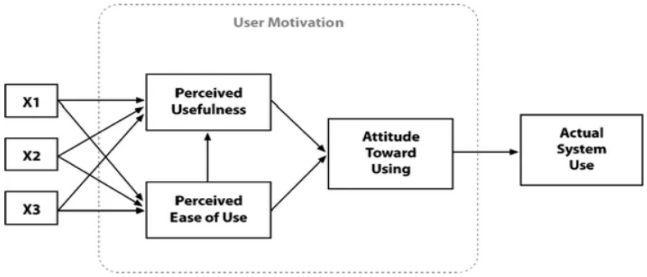

Fig. 1. Technology acceptance model (TAM) [13]

Davis proposed that the attitude of the potential users would have a direct impact on whether the system will be used [13]. Sequentially, the attitude of the user towards the technology was a result of how useful they deemed the system to be in addition to how easy it looked to use, with the ease of use having a direct influence on the perceived usefulness [13]. Both the perceived usefulness and ease of use were thought to depend on the characteristics of the designed system and are depicted as X1, X2 and X3 in Fig 1 [13]. This model was further refined and expanded in later years, however, the essential parts remain the same [13] therefore, this version was deemed sufficient for this study.

This study hypothesised that by first determining teachers' attitudes towards AI grading and then ascertaining what assessments they would find it useful for, will allow the discovery of characteristics that are currently missing from AI grading models for open-ended questions.

5 DATA COLLECTION RESULTS

The following section summarises the results of the interviews and survey.

5.1 Survey & interview design

The survey and interview questions were designed cohesively, first asking a demographic question, followed by a couple of introductory questions and lastly questions which aimed to answer the main research question (RQ). The demographic question asked for the program the participating teacher taught in, this was inquired as it was hypothesised by this study that the teacher's background may affect their attitude towards AI grading. Next, the participant was asked if they grade exams themselves and if they could imagine utilizing AI grading for open-ended questions. These questions served as an introduction to the subsequent queries and provided an initial feeling of the participant's attitude towards AI grading. Following the introductory queries, questions aimed at directly answering SRQ1 and SRQ2 were posed. To answer SRQ1 the participants were asked what assignments they would use AI grading for, and for SRQ2 participants were asked several questions: if they would involve themselves in the grading process when working with AI grading, what impact the use of AI grading may have on their relationship with their students and how would they support their students when using AI grading. For the survey questions, where appropriate a 4-Point Likert Scale (Strongly agree, Agree, Disagree, Strongly Disagree) was used instead of a 5-Point to avoid receiving neutral responses which could make it difficult to draw conclusions. The interview and survey questions can be found in the Appendix.

5.2 Results summary

This section summarises the results of the 10 conducted interviews with teachers from the UT as well as the 20 collected survey responses.

5.2.1 *Demographic overview.*

The survey respondents taught in 14 different teaching programs including psychology, philosophy, (biomedical) health sciences and various engineering programs. Of the survey respondents 55% grade with the help of TAs, 40% grade themselves and 5% do not grade exams themselves.

The teaching programs of the interviewees are summarised in Fig.2. Of those interviewed, 60% grade with the help of TAs and the other 40% grade their exams themselves.

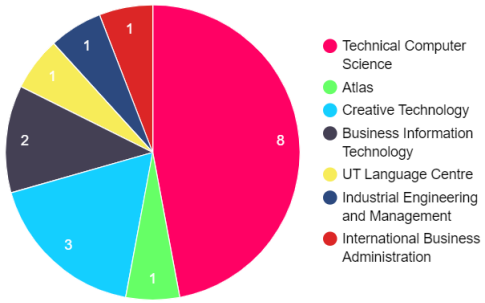

Fig. 2. Distribution of interviewed teachers by teaching program

5.2.2 *AI grading for open-ended questions as a possibility.*

The majority of the interviewed teachers answered that they could

envision using AI grading for open-ended questions in some capacity, however, this potential usage would come with several conditions, which are explained in the subsequent sections, the main of which is that the teacher maintains the sole responsibility for a student's grade. As for the survey, the votes on whether teachers could imagine using AI grading were split in half, with 35% saying "Strongly disagree", 20% saying "Strongly agree" and the rest answering either "Agree" or "Disagree".

A few participating teachers answered that they would not use AI grading for open-ended questions at all. The reason behind this answer, from one interviewee, was that AI would be a stochastic technology and hence would have a tendency to "hallucinate answers" and output different answers each time it is applied to the same questions, making it vastly unreliable. In addition to this, an interviewee said that the behaviour of AI towards 'edge case' answers are entirely unforeseeable and "whilst a TA can decide that they do not know how to grade a particular question and flag it to be graded by the teacher, AI will most likely not be able to do this." This interviewee elaborated that even if AI had an inbuilt functionality that would aim to recognise when it does not know an answer, this would still be very prone to false positives and false negatives, potentially making grading of 'edge case' answers random. Furthermore, an interviewee and a survey respondent both mentioned that fully relying on AI grading may be against current rules and regulations, such as those outlined by the UT's examination board, which expects transparency and fairness within the grading process. The interviewee explained that here, transparency means that every grade is explainable with a rubric known to the student before the exam, however, as mentioned previously, the output of AI can be a "black box" and unexplainable, infringing on this regulation. Another interviewee, who could not imagine using AI grading, explained that they are not used to it and do not find it necessary. They expressed that they would rather remain grading themselves as not only is AI still a very new technology and they do not trust it but, also because students have a right to be graded by professionals; two survey respondents also supported this argument. In addition to this, a survey respondent mentioned that they would not trust AI to grade questions where partial points are possible as this can be very subjective and the AI may output greatly different points to very similar answers.

Despite a lot of the participating teachers being open to using AI grading to some extent, generally, teachers were concerned with what exactly that would entail. A large number of participating teachers expressed a lack of trust in current and future AI models. Reasons for this included both insufficient understanding of how an AI grading model would work underneath the surface as well as having very in-depth knowledge of how AI functions and deeming it unsuitable for grading open-ended questions at its current state.

5.2.3 AI grading as an assistant tool.

Participating teachers who had a positive attitude towards potentially using AI grading had one unanimous condition: AI grading should only be used as a helping tool and the teacher will always decide the final grade and remain responsible for it. Teachers, in both interviews and the survey, recognised the time-saving benefits that AI grading could provide, however, coupled with the aforementioned lack of trust they described various ways in which the AI tool can merely aid them in the grading process.

The interviewees described several functions an AI grading assistant could perform. Two commonly mentioned functions were finding keywords within the answers or checking if certain components of the rubric were present. Interviewees explained that this sort of pre-processing of assignments by AI could save teachers a lot of time as they would only need to focus on the pointed-out sentences/sections of the answer. Due to the rising number of students and increased financial constraints, which impact the number of hired TAs, one teacher described this potential pre-processing functionality as a "wonderful concept" and another called it a "very promising technology". Similarly, an interviewed teacher mentioned that if the AI can identify keywords that are present, it can also identify keywords that are missing, this could work to assist the teacher in focusing on areas a student needs help in. One of the teachers mentioned that AI is much better at identifying correct answers rather than partially correct answers; therefore, if the partially correct answers could get filtered out, the teacher would be able to focus more on providing detailed feedback to students who are struggling instead of spending time grading 'excellent' answers in full detail. Another addition to this assistant functionality that was mentioned in an interview is recommendations, where based on the contained keywords the AI can suggest to a teacher if the answer was appropriate but would never automatically add points to a grade. Likewise, an interviewed teacher suggested that perhaps the AI grading tool can be given a rubric from which it selects keywords that it determines as significant for the correct answer and then the teacher approves the list of the selected keywords per question. This would also save the teacher time by not needing to manually enter the keywords for the AI. Lastly, a survey respondent and an interviewee, both from the UT Language Centre, proposed that the AI grading assistant tool can be used to create rubrics as well as to perform low-level grammatical/structural editing so that the teacher does not need to waste their time on it but the student can still receive some feedback. In general, many study participants agreed that there is a space to be filled by AI within the grading process, as long as the final resulting grade always comes from the teacher.

5.2.4 Nuances of an AI grading assistant.

Whilst the potential capabilities of an AI grading assistant were discussed in Section 5.2.3, several more nuanced remarks were also made regarding this possibility. Several study participants showed their concerns regarding the privacy of the students. An interviewee explained that training a model over the years on the same exam is likely to produce the most accurate results, however, this would mean that students' exams are used as a data set which they should be aware of and consent to. A response from the survey mentioned that they do not think it is ethical to request student's permission to use their work to train AI at all. A further issue to consider was mentioned in an interview and that is the possible bias an AI grading assistant could introduce. The interviewee explained that if the AI grading assistant gives the teacher recommendations before

they grade an assignment themselves, the teacher is likely to be influenced by said recommendation, making it potentially unfair if the AI is inaccurate. In parallel, an interviewee commented that AI may also give incorrect recommendations due to a student's writing style, particularly their use of language. They explained that teaching in an international setting often requires working with students who have English as a second language, meaning that they often make lexical mistakes or use generally awkward phrasing. They elaborated that in such situations a teacher can quite easily decipher the meaning behind the student's work but AI may struggle to do this and therefore, introduce bias against pupils with weaker English. Other survey respondents also shared this view. Another nuance brought up by a survey respondent was the protection of teachers when using AI. This teacher wrote that some students tend to get frustrated even with a teacher's feedback which they do not agree with. They expressed that using AI may cause even more frustration from the students if they receive a failing grade therefore, before AI grading is introduced "there would need to be procedures and policies put in place by the examination board to provide teachers with some measure of protection in the event that a student challenges an AI-generated grade." Lastly, the same survey respondent raised the concern of how much would AI grading actually help teachers or further complicate matters. If teachers cannot rely on the AI to consistently generate the 'correct' output they may be forced to spend more time regrading answers marked by AI or dealing with students complaints about their AI-generated grades. Furthermore, this teacher explained that learning any new technology always has a learning curve and is a time investment in itself which may not actually save time in the long run.

5.2.5 Assignments to be graded by AI.

Amongst participating teachers interested in using AI in their grading process, there was a rather clear consensus on what type of assignment it should be used on. Interviewed teachers explained that AI should primarily be used for factual assignments because although there are multiple ways to phrase an answer there is only one correct solution and an AI model should be quite good at meaning interpretation. Such questions would be: Give an example, explain a term, program a function that will do this, etc. Assignments which require a judgment, where a conclusion needs to be interpreted from the answer, or questions which have potentially "out of the box" solutions should not be graded using AI. An interviewed teacher explained that despite AI being very good at statistics and recognising patterns, making it an excellent candidate to grade factual assignments, it is weak in its interpretation of semantics.

Several specific assignment types were discussed in the interviews. Programming assignments were discussed at length by teachers from the Computer Science program and it was largely agreed that because coding solutions can often be creative and not all deviations from the standard solution are wrong, AI should not be grading them as a whole. If, however, AI grading is implemented as the above-discussed assistant, it can be used to explain to the teachers what the code is doing, so they do not have to read it line by line, it can highlight the bounds of a function and it can check its output. All these functionalities can save time for busy graders. Additionally, an interviewee commented that the AI grader should be able to recognise if certain 'mistakes' are significant to the grade, for example, it should be able to ignore syntax errors and misspelt words which are of no consequence to a student's understanding of programming. In this case, an interviewee explained that the AI grading assistant could also be made available to the students during their studies. They elaborated that this would help students understand how AI will be used when their work is graded so there would be no surprises, can help them receive faster feedback as well as alleviate some pressure from teachers and TAs as some queries can be answered by the AI. Even so, not all programming assignments would be suitable for AI grading specifically, multiple interviewees mentioned that assignments in which code design and code quality are significant should not be evaluated by AI as answers could be far too subjective.

Another discussed type of assignment that AI could give recommendations on, is the technical sections of project reports. Here an interview explained that the AI could also use keyword identification to give the grader a quick second opinion on what is there and what is missing, however, AI should not be used to grade, for example, the reflection component of a project report as "reflection is a purely human endeavour".

Several teachers from less technical studies also added that an AI grading assistant could be used on simpler language assignments, where the important components are low-level concerns, such as grammar, sentence structure, used vocabulary level and not creativity. Another mentioned part that could be done by AI in language assignments, is checking if specific simple rubric points have been met such as not using contractions or personal pronouns. In this case, an interviewee explained that it could help teachers quickly understand if an assignment they are grading is up to the standards of a certain language level. They further explained that this could also allow students to receive feedback on these low-level concerns which can often be overlooked by graders in favour of focusing on bigger issues due to time limitations. It was emphasised by multiple teachers, in both the interviewees and survey responses, that (academic) writing should not be graded by AI in its entirety as it often does not have a rigid template which makes it difficult for the AI similarly, some assignments in psychology and philosophy are intricate and require a sophisticated level of interpretation and evaluation which teachers do not think AI currently possesses.

Generally, when discussing assignment types to apply AI grading to, participating teachers agreed that it should be assignments of low importance. This could make some grading processes more time efficient and supply students with more feedback than could be provided by teachers.

5.2.6 Impact of AI grading on student-teacher relationship.

There was a relatively even distribution of the participating teacher's opinions regarding the impact that AI would have on their relationship with their students.

On the one hand, some participating teachers agreed that the most

significant factor that would create a positive impact on their relationship with their students is the teacher's ability to always explain a student's grade. If this is the case, then multiple study participants agreed that introducing an AI grading assistant could significantly lower grading time, which is often a point of tension between teachers and their students. An interviewed teacher proposed that the saved time could then be used to provide more guidance for struggling students, further improving the student-teacher relationship.

On the other hand, multiple teachers in the study expressed their concerns regarding the impact that using AI would have on their relationship with their students. An interviewee said students may question why they still need teachers if their assessment comes from AI or why they are paying university fees. A response from the survey argued that an AI-driven course can be set up without teachers and without the need to attend a university at all. An interviewee added that students are often unhappy with standardised feedback that comes from the teacher, they may feel even more distanced from the teacher if their feedback does not come from the teacher at all but from an AI model. A survey respondent also explained that using AI may "offload" the responsibility of grades from the teachers to the students as they would then have to actively complain and ask for explanations. An interviewee further elaborated that in this situation, technology acceptance takes on a big role and as of currently, students do not trust AI and do not see it as reliable. They continued that in the future, AI could become so integrated into people's daily lives, that it becomes entirely accepted but as of right now this is not the case. Another concern mentioned by an interviewee was that using AI grading can eliminate a teacher's opportunity to learn from their students. They continued to explain that grading exams can tell a teacher which parts of the course may be poorly explained or which concepts students struggle with, if answers are AI graded the teacher will inadvertently miss these.

A portion of participating teachers said that the impact on the student-teacher relationship will heavily depend on how teachers choose to use AI. An interviewee said that teachers need to correctly and transparently disclose that they are using AI grading, this will largely impact how much the students trust this process. They detailed that if an emphasis is made on the fact that AI is only used for pre-processing to speed up grading, then student's attitudes will largely depend on their trust in the teacher. If students believe that their teachers will actually regrade work pre-processed by AI and not blindly accept AI recommendations, they are more likely to be open to this technology.

6 DISCUSSION

The following section discusses the implications of the collected data, mentions the limitations of the conducted study and proposes some opportunities for future work on this subject.

6.1 Results analysis

In the subsequent sections, the results of the survey and interviews are analysed.

6.1.1 *Bloom's Taxonomy.*

In both the interviews and survey responses, teachers seemed to agree that if AI grading is utilised it should only apply to simple and factual questions. To be able to more definitively decide which assignments are most suitable for AI grading, the Taxonomy of Bloom can be used. The (revised) Taxonomy of Bloom describes a "model of classifying thinking according to six cognitive levels of complexity" and is shown in Fig 3. [9].

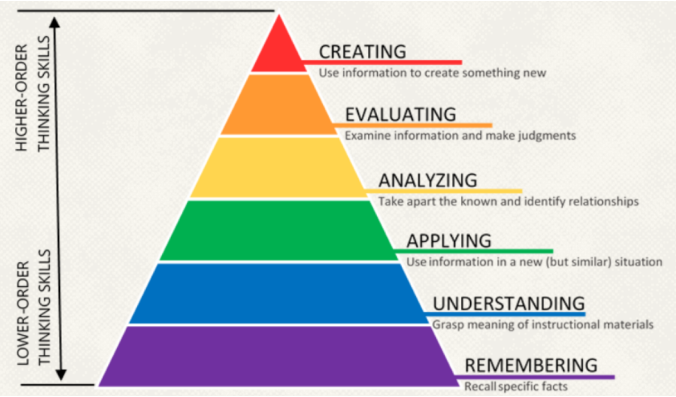

Fig. 3. Bloom's Taxonomy [16]

The higher the thinking skill is in the Taxonomy, the higher order thinking skill it is [16]. Considering the opinions of the teachers regarding the use of AI exclusively for factual assignments which do not require judgment calls, it can be theorised that it may be better to, at least initially, apply AI grading to assignments with lower-order thinking skills such as, 'remembering' and 'understanding'.

6.1.2 *AI as a grading assistant.*

The vast majority of participating teachers explained that they would be open to using an AI grading assistant and listed some possible functionalities that it could possess, these are outlined in Fig 4. Participants explained that such a tool could help teachers save time but keep the teacher central to the grading process, this idea was also presented in the studied literature [6] [4]; where an emphasis was made on AI only taking on a helping role within education.

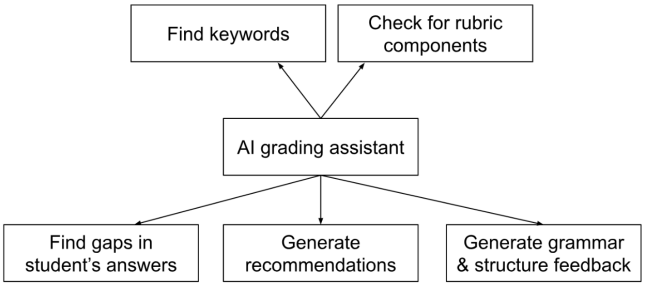

Fig. 4. Potential functionalities of AI as a grading assistant

Some of these functionalities are already present in AI grading tools such as EasyGrader, which is capable of determining term frequency and thus identifying keywords within a student's answer [1].

6.1.3 *Benefits and drawbacks of an AI grading assistant.*

Table 2 summarises the potential benefits and drawbacks of an AI assistant according to the participating teachers who expressed interest in this technology. Some of the discovered benefits and drawbacks were also discussed in the literature review, where [12] mentioned that the former included time-saving and more feedback whilst the latter included ethical concerns as well as the potentially inadequate quality of AI generated output.

Table 2. Benefits and drawbacks of an AI grading assistant

Benefits	**Drawbacks**
Saves time	Teachers may rely too heavily on AI recommendations
More focus on struggling students	Teachers may be biased when grading due to AI recommendations
Students could receive more feedback	AI recommendations may be incorrect due to the student's writing style
	Using AI can have a learning curve for the teachers

6.1.4 *Rules for assessment and transparency.*

A point brought up by some participants of this study was the potential violation AI-driven grading would cause on UT's rules and regulations. The assessment policy on a programme level at the UT states that " The programme can justify itself for how the assessment takes place and the quality of the assessment" [15]. Because of this policy, it must be ensured that using AI-driven grading is justifiable and can still guarantee that the quality of the assessment is not hindered. Furthermore, a guideline mentioned by an interview participant for Computer Science at the UT described three screening guidelines for assessment, which are validity, reliability and transparency [21]. Validity means that assessment criteria need to be known to students in advance, reliability means the assessment is fair and has valid methods to mitigate bias and lastly, for the assessment to be transparent it requires sufficient information to have been communicated to the students about the testing process and for all that information to be unambiguous [21]. Complying with validity and providing students with grading rubrics in advance will still be possible when using AI but satisfying the requirements of reliability and transparency may be more cumbersome. Since AI can often be biased, mitigating it and ensuring reliability will remain the task of the examiner. Abiding with the guideline of transparency may also be an issue as AI output can often be ambiguous. When analysing these guidelines it can be hard to imagine that an AI grading model can satisfy them at all, however, if the case of AI as a grading assistant is being discussed, it could still comply with these rules. If the AI grading assistant is only performing functions as described in Section 6.1.2 then it could be rather reliable and, due to the low complexity of its functionalities, also be fairly transparent to the students. To further improve transparency, several teachers mentioned allowing students to use the AI grading assistant during their studies as this could help them accept and understand the technology better. Literature such as [18] also discussed that while transparency was one of the main concerns students had regarding AI grading, they also recognised that AI could be a form of educational support to them during their studies. In parallel, [7] mentioned that the transparency of an AI algorithm will have a direct impact on how fair students find it. Therefore, allowing students to access the same AI tools during their studies as teachers would use during grading, may be a way to help students understand and accept this technology better.

6.1.5 *Is AI grading wanted?*

Especially in survey responses answers such as "No idea" and "Haven't thought about it" appeared quite a few times. This may suggest that at least some teachers have not considered the prospect of using AI for grading and it may not be their priority. Similarly, whilst some study participant said they would use AI it came with a lot of preconditions, making it seem as though, again, the technology wasn't really on their mind and they are only thinking of it as a future possibility. A couple of interviewees also mentioned that even with a high workload it is their duty to grade their student's work themselves and "offloading" it to an AI is unethical and unfair to the students. This can be seen as a lack of perceived usefulness by some teachers which could explain their negative attitude towards AI grading, as the studied literature on technology acceptance demonstrates a direct link between the perceived usefulness of a technology and its acceptance by the user [13]. In consequence, this raises the question of whether AI grading is trying to solve a problem which, in some teacher's perception, does not exist. Is AI grading needed or is it just another way for people to jump onto the AI 'trend'?

6.2 Limitations

The study was conducted with teachers only from the UT and the majority of participants, in both the survey and especially the interviews, were from technical studies, mostly Computer Science. The goal of the survey was to reach a wider variety of teachers and whilst that was achieved it was on a smaller scale than would have been ideal. Primarily, it seemed that fewer teachers than anticipated wanted to fill out the survey. Because of this, the results may not accurately reflect the opinions of teachers from other subjects or universities as they may have different perspectives.

6.3 Future work

As grading open-ended questions with the use of AI is still a new concept, there is still a vast quantity of future work that needs to be done before this technology becomes widespread. One such study could be very similar to this one but focusing on different universities which teach subjects that are not present at the UT such as law, art, history, etc. This study observed that the subjects participants taught had a significant influence on their expectations of what an AI-driven grading model should be able to do, thus a comprehensive understanding of what would be wanted by more university-level

subject teachers is necessary.

Another future study should involve pilot programs with some already invented AI grading models, for instance, the aforementioned EasyGrader tool could be made available to a subset of teachers within the UT to gather their feedback on a real system instead of a hypothetical one.

Lastly, it may be important to examine how AI grading will fit into various university regulations and how teachers can be protected when using AI. This concern was outlined in Sections 5.2.2, 5.2.4 and further discussed in Section 6.1.4, because of this concern it is necessary to analyse how university regulations may need to be adapted to accommodate AI grading. Furthermore, rules to protect teachers when using AI should also be proposed and accepted university-wide before this technology can start being utilised.

7 CONCLUSION

The study's objective was to determine how open teachers are towards introducing AI-driven grading for open-ended questions at a university level. To understand teachers' perception of this emerging technology, it was important to understand how teachers would use AI grading and what consequences they believe it would have on their relationship with their students. These findings could serve as a foundation for future research and development of university AI grading models. The study has consisted of a literature review, interviews and a survey. The former allowed an understanding of previously performed research as well as helped formulate the interview and survey questions whilst the latter two methods of research allowed the acquisition of new insights.

Over half of the study participants (60%) could envision AI within the grading process as this technology could offer significant time-saving benefits to busy teachers and potentially allow them to focus more on struggling students. Despite this, several issues that AI may introduce were discussed, these included: AI generating different answers each time it is applied, AI potentially generating 'random' output when a student's answer is more complex or is partially correct, using AI for grading may be against privacy or other regulations and lastly, AI may be biased towards answers that have a poorer use of the English language. Furthermore, participants from both the interviews and the survey brought up moral reasons why AI should not be used such as that students deserve to be graded by experts and that this is a big part of why people choose to go to university and therefore pay large tuition fees. Due to these concerns, the teachers who participated in this study decided that the grading processes could not be simply handed over to an AI.

The mentioned concerns were directed at using AI as a grader, but study participants suggested that AI could only function as a grading assistant. This would greatly reduce the number of applicable concerns as the teacher would remain the grader. The AI grading assistant could, amongst others, highlight keywords, generate recommendations and check for rubric components. This could save teachers a lot of time and make it easier to grade 'perfect' answers so that more time could be devoted to giving feedback to struggling students.

As for the assignments suitable for AI, the participating teachers agreed that they should be low-complexity, factual assignments in which creativity, reflection and design are not the focus. For example, programming tasks where the goal is to produce the correct output and not code design, technical sections of project reports and simpler language assignments where content and/or grammar are the focal points would be, according to the participants, suitable for AI grading.

Although there was a rather clear consensus amongst the study participants on which assignments AI would be appropriate for, the same could not be said about the impact that introducing AI would have on student-teacher relationships. The participating teacher's opinions diverged with some stating that using an AI grading assistant would not hinder their responsibility for the students' grades and would only help them to output grades faster thus potentially allowing more contact time with students, improving their relationship. Others, however, said that AI grading would further distance teachers from students and may lead to students wondering why they need teachers and universities at all. Furthermore, AI is still not accepted as a reliable technology and students are likely not to trust it, potentially creating more work for the teachers if they are constantly asked to explain student's grades or re-grade them. Finally, a portion of teachers expressed that the impact AI will have on student-teacher relationships will heavily depend on how transparently teachers will use the technology and how well students will understand its mechanics.

Overall, teachers appear to be more open to AI as a grading assistant to grade faster than as a replacement for actual graders. This may change in the future as AI becomes more advanced, however, in its current state the vast majority of the participating teachers believe AI is not sophisticated enough to take over grading in university education. Teachers could use an AI grading assistant for factual assignments and focus more on interaction with students or grading more complex, creative pieces of work. The introduction of AI into university education can have both a positive and a negative impact, which will largely rely on how transparent teachers are with their students about using AI, how heavily teachers will depend on AI and how much students will trust the AI technology. Regardless of precisely how AI is introduced into the grading process at universities, it was unanimously agreed that the teacher must always remain the sole entity responsible for a student's grade.

REFERENCES

[1] A.I. Aldea. 2020. Towards grading automation of open questions using machine Learning. (2020).
[2] Greg Allen. 2020. Understanding AI technology. *Joint Artificial Intelligence Center (JAIC) The Pentagon United States* 2, 1 (2020), 24–32.
[3] Murtadha Alomran and D Chai. 2018. Automated scoring system for multiple choice test with quick feedback. *International Journal of Information and Education Technology* 8, 8 (2018), 538–545.
[4] SHMP Berg. 2023. *The Acceptance of Assessment AIs in Formal Higher Education: an Exploratory Study.* Master's thesis. University of Twente.

[5] Steven Burrows, Iryna Gurevych, and Benno Stein. 2015. The eras and trends of automatic short answer grading. *International Journal of Artificial Intelligence in Education* 25 (2015), 60–117.
[6] Ismail Celik, Muhterem Dindar, Hanni Muukkonen, and Sanna Järvelä. 2022. The promises and challenges of artificial intelligence for teachers: A systematic review of research. *TechTrends* 66, 4 (2022), 616–630.
[7] Fangyuan Chai, Jiajia Ma, Yi Wang, Jun Zhu, and Tingting Han. 2024. Grading by AI makes me feel fairer? How different evaluators affect college students' perception of fairness. *Frontiers in Psychology* 15 (2024), 1221177.
[8] Blerta Abazi Chaushi, Florije Ismaili, and Agron Chaushi. 2024. Pros and Cons of Artificial Intelligence in Education. (2024).
[9] Mary Forehand. 2010. Bloom's taxonomy. *Emerging perspectives on learning, teaching, and technology* 41, 4 (2010), 47–56.
[10] Jeffrey W Knopf. 2006. Doing a literature review. *PS: Political Science & Politics* 39, 1 (2006), 127–132.
[11] Valentin Kuleto, Milena Ilić, Mihail Dumangiu, Marko Ranković, Oliva MD Martins, Dan Păun, and Larisa Mihoreanu. 2021. Exploring opportunities and challenges of artificial intelligence and machine learning in higher education institutions. *Sustainability* 13, 18 (2021), 10424.
[12] Rahul Kumar. 2023. Faculty members' use of artificial intelligence to grade student papers: a case of implications. *International Journal for Educational Integrity* 19, 1 (2023), 9.
[13] Nikola Marangunić and Andrina Granić. 2015. Technology acceptance model: a literature review from 1986 to 2013. *Universal access in the information society* 14 (2015), 81–95.
[14] Cormac McGrath, Teresa Cerratto Pargman, Niklas Juth, and Per J Palmgren. 2023. University teachers' perceptions of responsibility and artificial intelligence in higher education-An experimental philosophical study. *Computers and Education: Artificial Intelligence* 4 (2023), 100139.
[15] n.d. n.d. Assessment policy programme level. https://www.utwente.nl/en/examination-board/Safeguarding_assessment_quality/Assessment_Policy/.
[16] n.d. n.d. Bloom's taxonomy. https://citt.ufl.edu/resources/the-learning-process/designing-the-learning-experience/blooms-taxonomy/. (n.d).
[17] Satya Vir Singh and Kamal Kant Hiran. 2022. The impact of AI on teaching and learning in higher education technology. *Journal of Higher Education Theory and Practice* 22, 13 (2022).
[18] Eva Stoica. 2022. *A student's take on challenges of AI-driven grading in higher education.* B.S. thesis. University of Twente.
[19] Valerie M Sue and Lois A Ritter. 2007. *Conducting online surveys.* Sage.
[20] Hamed Taherdoost. 2022. How to conduct an effective interview; a guide to interview design in research study. *International Journal of Academic Research in Management* 11, 1 (2022), 39–51.
[21] Vadim Zaytsev. 2022. Assessment Screening Guidelines.
[22] Ramiz Zekaj. 2023. AI Language Models as educational allies: enhancing instructional support in higher education. *International Journal of Learning, Teaching and Educational Research* 22, 8 (2023), 120–134.

APPENDIX

AI usage

During the preparation of this work the author used Grammarly in order to check the spelling and grammar in the written work. After using this tool/service, the author reviewed and edited the content as needed and takes full responsibility for the content of the work.

Interview questions

(1) In which program(s) do you teach?
(2) Do you grade your exams yourself?
(3) Could you imagine using AI grading for open-ended questions in your courses? Why/why not?
(4) What kind of assignments would you use AI grading for? Why these/why not others?
(5) If you are using AI grading for open-ended questions, would you still involve yourself in the grading process? Why/why not?
(6) How do you think the use of AI grading for open-ended questions will impact your relationship with your students?
(7) What methods of support would you give to students when using AI grading for open-ended questions?
(8) Do you have any other comments concerning this topic?

Survey questions

(1) In which program(s) do you teach?
  (a) Technical Computer Science
  (b) Business Information Technology
  (c) International Business Administration
  (d) Other:
(2) Do you grade your exams yourself?
  (a) Yes
  (b) Partially, with the help of TAs/others
  (c) No
(3) I could imagine using AI grading for open ended questions in my courses. Please indicate how much you agree.
  (a) Strongly disagree
  (b) Disagree
  (c) Agree
  (d) Strongly agree
(4) Explain your previous answer, why or why not?
(5) What kind of assignments would you use AI grading for? (Tick all that apply)
  (a) Final assignments (Assignments with a lot of weight to them such as exams/ project reports)
  (b) Pass/fail assignments
  (c) Essays done for homework/tutorials
  (d) Code-based homework/tutorial assignments
  (e) Math (formula-based) homework/tutorial assignments
  (f) Math (proof-based) homework/tutorial assignments
  (g) Other:
(6) Why did you choose or not choose the assignments in the previous question?
(7) If I am using AI grading for open-ended questions, I would still involve myself in the grading process. Please indicate how much you agree.
  (a) Strongly disagree
  (b) Disagree
  (c) Agree
  (d) Strongly agree
(8) Explain your previous answer, why or why not?
(9) How do you think the use of AI grading for open-ended questions will impact your relationship with your students?
  (a) It won't have an impact
  (b) I'm not sure
  (c) It will have a positive impact
  (d) It will have a negative impact
(10) Explain your previous answer.
(11) What methods of support would you give to students when using AI grading for open-ended questions?
  (a) Give students a right to ask for a regrade (if an assignment was graded by AI)
  (b) Regrade barely not passing assignments (when graded by AI as 4.5<=5.5)
  (c) Offer additional feedback upon request

(d) Other:

(12) Do you have any other comments concerning this topic?