Using Large-Language Models for Rubric-Based Assessment

Dan S. Myers and Madhav Khanal

Summary and Recommendations

This report presents the results of experiments using large-language AI models (LLMs) on a realistic program assessment task. Working with a set of 100 student information literacy artifacts collected for the 2025 RFLA assessment cycle, we investigate whether current-generation AI models can be used for effective rubric-based assessment.

We prompted a collection of current leading models from OpenAI (GPT-5, GPT-4o, o3, and o4-mini) and Anthropic (Claude Sonnet 4 and Opus 4.1) to score each student artifact using the categories of the AAC&U's Information Literacy VALUE rubric, with the goal of answering the following questions:

1. Can models provide self-consistent evaluations – that is, does a given model return similar results when prompted to assess the same artifact multiple times?
2. Do the evaluations of AI models align with those given by expert faculty evaluators?
3. How does the choice of model and prompt design affect the results of questions (1) and (2)?

The high-level answers to our questions are as follows:

1. All models exhibit **high interrater reliability** that meets or exceeds that of human assessors

2. **Models do not inherently grade in the same way as humans**. The simple strategy of prompting a model with an artifact, a copy of the rubric, and instructions for scoring (termed "zero-shot prompting" in the AI literature) never aligned with human results. OpenAI's models consistently yield results that are far more positive than human assessors across all categories of the rubric. Claude Sonnet was the best performing zero-shot model but still failed to produce a distribution of results that matched those of real human experts.

3. Incorporating examples into the prompt ("few-shot prompting") gives reasonable – but still imperfectly aligned – results for GPT-4o, Sonnet, and Opus. More powerful reasoning models (o3, GPT-5, and o4-mini) do not respond well to examples and underperform the less-powerful GPT-4o and Sonnet models.

These results provide useful guidance on the suitability of AI models for RFLA program assessment and grading more generally. We make the following recommendations:

- **Assessing student work using only a copy of a rubric is not reliable.** Zero-shot prompting never produced human-aligned results for any model in any experiment.

- Reasonable results can be obtained using well-designed prompts with examples but will still not perfectly align with humans and results are sensitive to the choice of model. Therefore, any use of AI in assessment **must be carefully validated and benchmarked** before being deployed.

- We should continually message to faculty that **AI assessment is not a substitute for our expert evaluation of student work.** AI *could* be a useful tool for some low-stakes assessments, but we should be maximally cautious about using it for consequential summative grading.

Data

The data set consists of a random sample of 100 artifacts collected from RFLA 100 courses during the 2024-25 academic year following the standard RFLA assessment process. Artifacts were anonymized by the regular RFLA assessment team after collection by removing student names and any personally identifiable information.

Each artifact was scored by a panel of three faculty members using the AAC&U's Information Literacy VALUE rubric. The rubric (shown in Fig. 1) defines five categories, each with four levels of mastery from 1 (lowest) to 4 (highest). Faculty reviewers also have the option to assign a score of 0 in a category if that aspect of the artifact cannot be assessed, or if it fails to meet even the standard for a score of 1.

The total set of human expert scores consists of 1500 individual observations (3 reviewers x 100 artifacts x 5 categories). Our assessment plan also considers the total summed score

INFORMATION LITERACY VALUE RUBRIC

for more information, please contact value@aacu.org

Association of American Colleges and Universities

Definition

The ability to know when there is a need for information, to be able to identify, locate, evaluate, and effectively and responsibly use and share that information for the problem at hand. - The National Forum on Information Literacy

Evaluators are encouraged to assign a zero to any work sample or collection of work that does not meet benchmark (cell one) level performance.

	Capstone 4	Milestones 3	Milestones 2	Benchmark 1
Determine the Extent of Information Needed	Effectively defines the scope of the research question or thesis. Effectively determines key concepts. Types of information (sources) selected directly relate to concepts or answer research question.	Defines the scope of the research question or thesis completely. Can determine key concepts. Types of information (sources) selected relate to concepts or answer research question.	Defines the scope of the research question or thesis incompletely (parts are missing, remains too broad or too narrow, etc.). Can determine key concepts. Types of information (sources) selected partially relate to concepts or answer research question.	Has difficulty defining the scope of the research question or thesis. Has difficulty determining key concepts. Types of information (sources) selected do not relate to concepts or answer research question.
Access the Needed Information	Accesses information using effective, well-designed search strategies and most appropriate information sources.	Accesses information using variety of search strategies and some relevant information sources. Demonstrates ability to refine search.	Accesses information using simple search strategies, retrieves information from limited and similar sources.	Accesses information randomly, retrieves information that lacks relevance and quality.
Evaluate Information and its Sources Critically	Thoroughly (systematically and methodically) analyzes own and others' assumptions and carefully evaluates the relevance of contexts when presenting a position.	Identifies own and others' assumptions and several relevant contexts when presenting a position.	Questions some assumptions. Identifies several relevant contexts when presenting a position. May be more aware of others' assumptions than one's own (or vice versa).	Shows an emerging awareness of present assumptions (sometimes labels assertions as assumptions). Begins to identify some contexts when presenting a position.
Use Information Effectively to Accomplish a Specific Purpose	Communicates, organizes and synthesizes information from sources to fully achieve a specific purpose, with clarity and depth	Communicates, organizes and synthesizes information from sources. Intended purpose is achieved.	Communicates and organizes information from sources. The information is not yet synthesized, so the intended purpose is not fully achieved.	Communicates information from sources. The information is fragmented and/or used inappropriately (misquoted, taken out of context, or incorrectly paraphrased, etc.), so the intended purpose is not achieved.
Access and Use Information Ethically and Legally	Students use correctly all of the following information use strategies (use of citations and references; choice of paraphrasing, summary, or quoting; using information in ways that are true to original context; distinguishing between common knowledge and ideas requiring attribution) and demonstrate a full understanding of the ethical and legal restrictions on the use of published, confidential, and/or proprietary information.	Students use correctly three of the following information use strategies (use of citations and references; choice of paraphrasing, summary, or quoting; using information in ways that are true to original context; distinguishing between common knowledge and ideas requiring attribution) and demonstrates a full understanding of the ethical and legal restrictions on the use of published, confidential, and/or proprietary information.	Students use correctly two of the following information use strategies (use of citations and references; choice of paraphrasing, summary, or quoting; using information in ways that are true to original context; distinguishing between common knowledge and ideas requiring attribution) and demonstrates a full understanding of the ethical and legal restrictions on the use of published, confidential, and/or proprietary information.	Students use correctly one of the following information use strategies (use of citations and references; choice of paraphrasing, summary, or quoting; using information in ways that are true to original context; distinguishing between common knowledge and ideas requiring attribution) and demonstrates a full understanding of the ethical and legal restrictions on the use of published, confidential, and/or proprietary information.

Figure 1: The Information Literacy VALUE rubric

across all five categories, which is in the range 0-20. For the 100-level information literacy course, our target is that *50%* of students should receive a cumulative assessed score of *10 or higher*.

Data Handling and API Use

There is concern that AI model providers may retain information provided in requests to use for future training. All our experiments use only model APIs accessed via network requests. Per their terms of service, OpenAI, Anthopic, and Google do not retain data submitted through their APIs for future training. Our data handling plans have been reviewed and approved by the IRB and the college's CIO.

Zero-Shot Prompting

The simplest prompting strategy is to submit an artifact, a copy of the rubric and instructions asking the model to score the artifact in each category using the criteria of the rubric. This approach relies on the model's internal interpretation of the rubric criteria rather than any provided examples and is termed *zero-shot prompting* in the AI literature and mimics how most users intuitively approach AI-supported assessment.

We used the zero-shot approach to evaluate the complete set of 100 artifacts using five major models. Each artifact was evaluated three times by each model with all evaluations being independent API calls – no information on prior submission or prior responses is retained between model calls.

We measured interrater reliability by calculating intraclass correlation coefficient (ICC) scores for the three evaluations given by each model on each of the five rubric categories. ICC is measured on a scale of 0.0 to 1.0, where 1.0 indicates perfect agreement. An ICC score of at least .70 is considered "good" agreement among raters and scores of .80 or higher are considered "excellent". The ICC consistency scores for the five models are as follows:

- Both Claude models and o3 achieved ICC scores of .85 or higher in every one of the five categories
- GPT-5 achieved scores of .77 to .86 in all categories
- GPT-4o scored .71 to .86 in all categories

Similar results are obtained for absolute measures of consistency. Our 2025 assessment report gives an overall ICC score of .792 for the human assessors. Therefore, it is reasonable to conclude that current major models exhibit a high level of self-consistency in their evaluations, to a degree that is at least comparable with human evaluators.

We next consider if the scores returned by AI models align with those generated by humans. The key metrics for this evaluation are the two benchmark score levels: the fraction of responses that achieved a combined score of at least 10 and the fraction scoring at least 15. The table shows the following:

- Humans rated about 50% of student artifacts with combined scores of 10 or higher, just meeting our benchmark goal for 100-level information literacy
- All the OpenAI models, including the new GPT-5, are substantially *more positive* than humans in their evaluations, estimating 79% or more of students met the baseline score of 10%

Model	Fraction >= 10	Fraction >= 15
Human expert	.507	.097
GPT-5	.790	.140
GPT-4o	.857	.177
o3	.837	.113
Claude Sonnet 4	.590	.040
Claude Opus 4.1	.490	.025

Table 1: Benchmark scores for zero-shot prompting

- Opus 4.1 most closely matches the fraction >= 10 and both Claude models underestimate the fraction of scores >= 15

Figure 2 shows the actual score distributions returned by the six models, aggregated across all five rubric categories. The human distribution is nearly symmetric: 7.5% zeros, 25% ones, 40% twos, 25% threes, and 2.5% fours. **None of the AI models produce similar score distributions**, even Sonnet and Opus, which came the closest to matching the benchmark score fractions. The three OpenAI models have similar distributions and award more than 80% of scores as 2 or 3. Sonnet and Opus assign scores of 2 about 60% of the time, with Opus being a bit more restrictive on assigning scores of 3 and 4.

These general dynamics extend to the score distributions within each of the five individual rubric categories. Figure 3 shows the distributions of 0-4 scores for the “Evaluate Information and Its Sources Critically Category”. Again, none of the AI models achieve a distribution that aligns with the scores assigned by human reviewers. The OpenAI models are primarily positive. Both Claude models are better aligned with the expert scores, but with again a greater preponderance of twos and almost no zeros or fours.

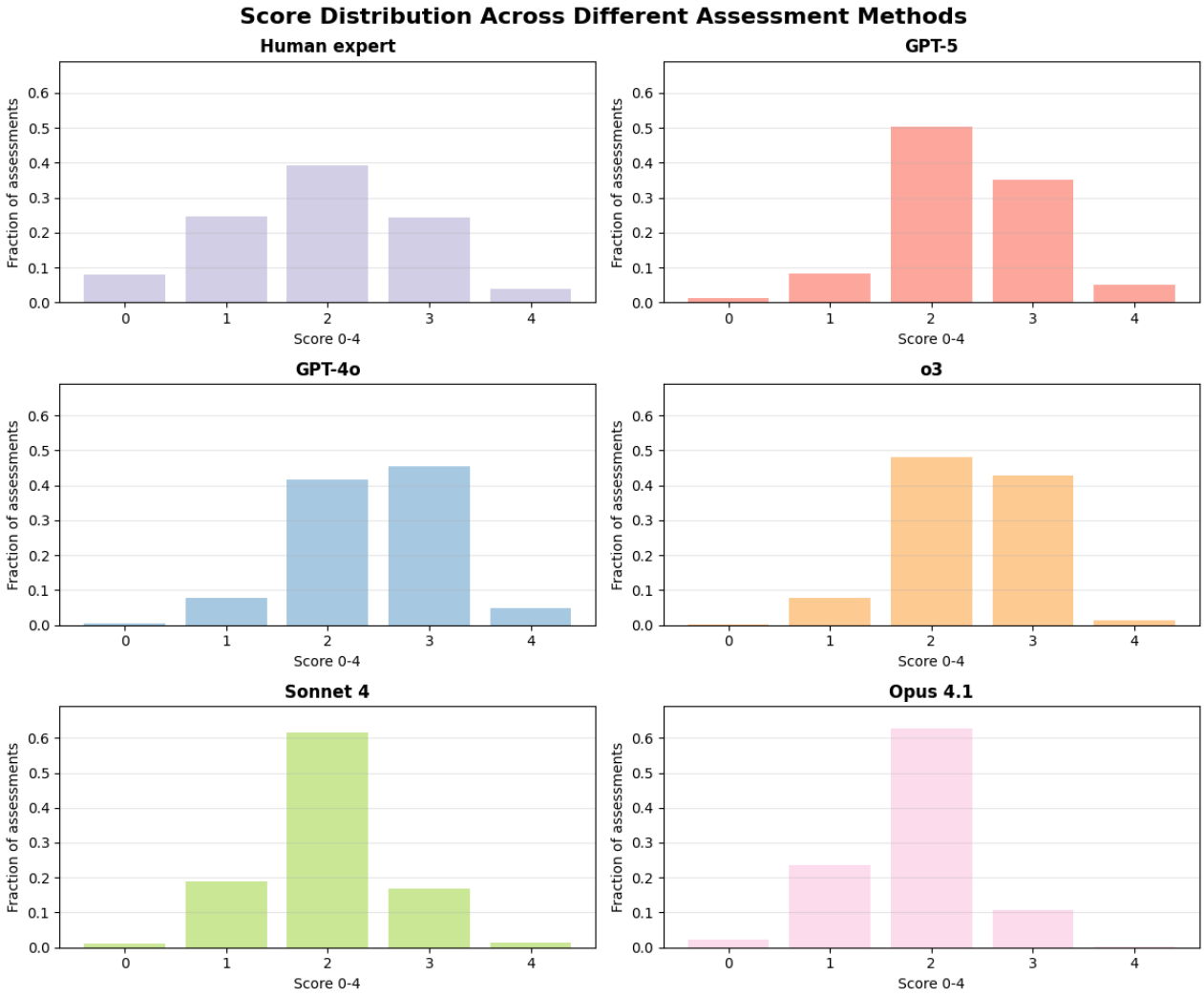

Figure 2: Cumulative score distributions per model

Based on these results (and similar behaviors in the other score categories), we conclude the AI models do not reliably align with human evaluators using zero-shot prompting. Therefore, unless benchmark experiments show otherwise, **we should assume that simple prompting strategies cannot be used to evaluate student work.**

Three-Shot Prompting

We next consider *few-shot prompting*, where the AI's instructions are augmented with specific examples illustrating different input cases to help guide its responses. We selected three of the 100 artifacts: one high-scoring (scores of 3 or 4 in every category), one

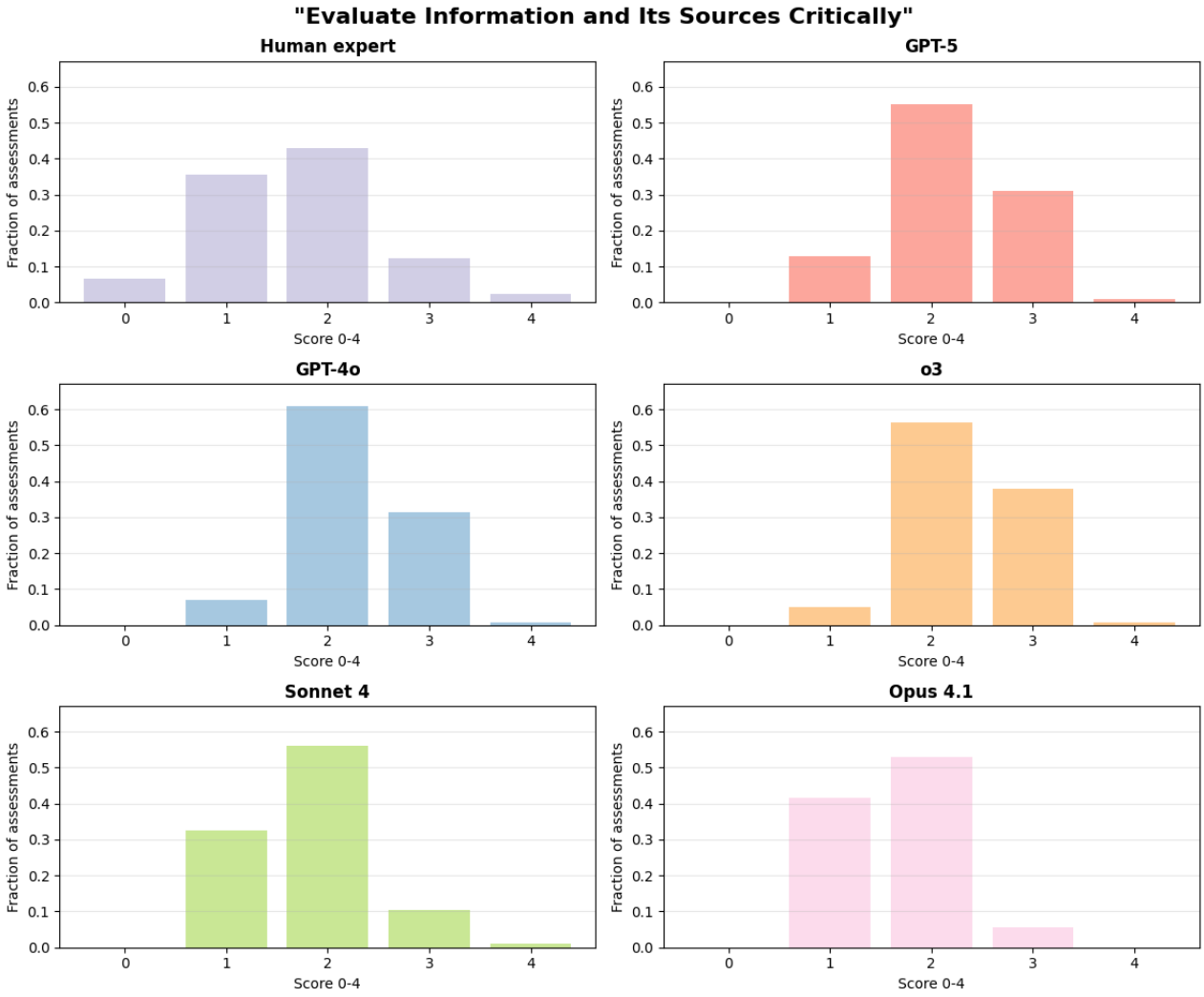

Figure 3: Score distributions for "Evaluate Information and Its Sources Critically"

medium-scoring (1 or 2 in every category), and one low scoring (0 or 1 in every category) and revised the prompt to contain the complete text of each artifact and the raw and average scores it was assigned by human evaluators in each category. The text of the prompt is given in the appendix.

Table 2 shows the fractions of scores of at least 10 and 15 for each of the six models under this prompting strategy and Figure 4 plots the overall distributions of scores.

Model	Fraction >= 10	Fraction >= 15
Human expert	.507	.097
GPT-5	.715	.069
GPT-4o	.577	.093
o3	.737	.108
Claude Sonnet 4	.474	.048
Claude Opus 4.1	.505	.041

Table 2: Benchmark scores for three-shot prompting

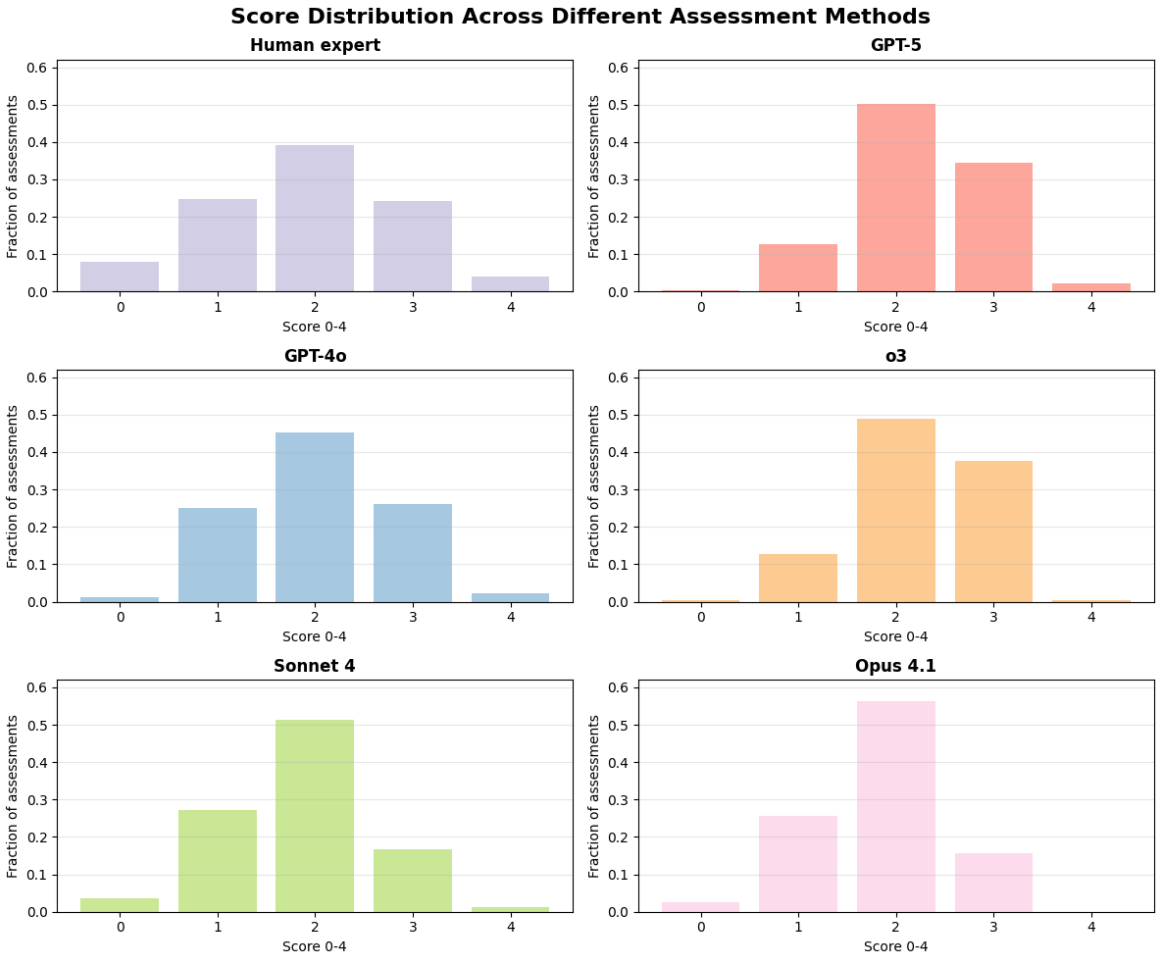

Figure 4: Cumulative score distributions for three-shot prompting

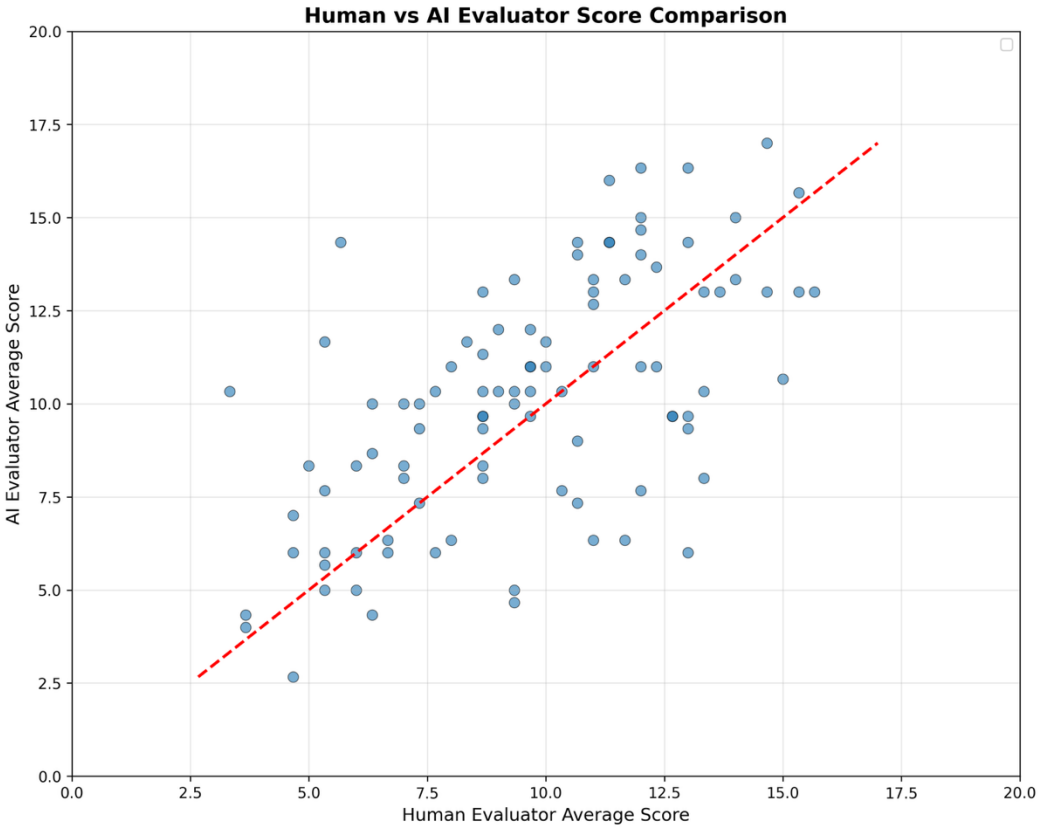

Figure 5: Comparison of human- and AI-generated scores for GPT-4o

These results both encouraging and interesting. GPT-4o moves from being one of the worst performing models to having reasonable alignment with the human evaluators. Sonnet and Opus are both very close to the 50% human mark for scores >= 10, but their score distributions do not match the human distribution as closely as does GPT-4o.

GPT-5 and o3 achieve similar results and both continue to award too many scores at the 2 and 3 level and too few at the 0 and 1 level. The o3 model uses additional inference-time reasoning, which often improves its responses on difficult problems. In this case, we suspect that reasoning is causing the model to pay less attention to the examples in the prompt – o3 is more likely to follow its own chain-of-thought, whereas GPT-4o is "less intelligent" but better at following instructions. GPT-5 is the newest model and uses an internal selector to route queries among a collection of reasoning and non-reasoning models based on its judgment of the most appropriate model based on the task. These

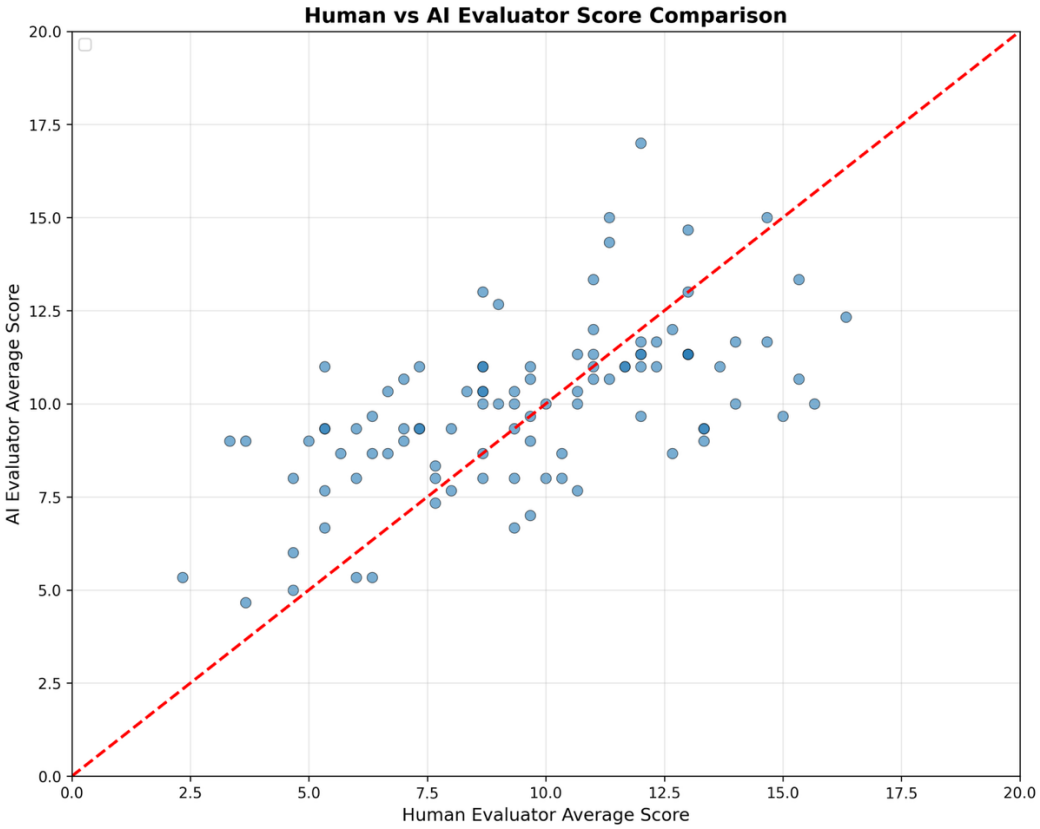

Figure 6: Comparison of human- and AI-generated scores for Claude Sonnet 4

results suggest that, for our problem, GPT-5 inclines toward one of its reasoning models that returns results like o3.

Figures 5 and 6 plot comparisons of the human-assigned and model-assigned scores for GPT-4o and Claude Sonnet 4 using the three-shot approach. Each artifact appears as a point in the scatter plot, with its x-position given by the average cumulative human score and y-position by the average cumulative AI score. Artifacts with similar scores lie close to the diagonal line; points above the line indicate artifacts for which the model returned a *higher* score than the humans and points below the line have *lower* model scores.

The figures reveal further details about the relative performance of the two models. Sonnet achieves the best performance on the top-level benchmark of matching the fraction of scores >= 10, but it does so by mostly scoring in a band around 10. Therefore, we should regard the high performance of Sonnet (and Opus, which has similar behavior) on that benchmark as more likely to be a chance artifact of the scoring process rather than an

indication of strength at the underlying labeling task. GPT-4o's distribution more clearly matches the actual human-assigned scores, although there are both high and low-outlying points. GPT-4o's performance is mostly hurt by its relative lack of fours, which causes the highest scoring artifacts to be marked below their real human evaluations.

We experimented with additional techniques in attempts to reduce the variance between human and AI-assigned scores, but none yielded improvements.

- Increasing the number of examples included in the prompt. Increasing to five or ten artifacts did not improve performance for GPT-4o. There are diminishing returns to increasing the length of the prompt.

- Creating a "scoring guide" by allowing a model to evaluate a few examples with their human scores and produce its own explanation of the factors associated with each score. In theory, these explanations should help the model by giving it guidance based on its own internal understanding of the scoring problem. In practice, no version of the scoring guide we tried gave any improvements, either with or without the three-shot examples included in the prompt. It did not matter if the scoring guide was produced by the same model conducting the evaluations or a different model; that is, there was no benefit to having a model, say Claude Sonnet, use a scoring guide that it produced itself vs. one created by another model.

- We conducted limited testing of a few other models: GPT-4.1, o4-mini, and Claude Sonnet 3.5. None of these experiments yielded improvements compared to the main models listed above.

Overall, these results indicate that few-shot prompting using GPT-4o is our best performing option for AI-supported assessment, although its results still do not perfectly align with human faculty evaluations.

Conclusions and Further Work

This study provides clear evidence that AI-supported assessment and grading should be **approached with caution**. Straightforward approaches likely to be used by most faculty

– prompting a model with a rubric and a submission – do not generate aligned results in any of our tests. Few-shot prompting does better, but the best performing model (GPT-4o) still has deviations from human-produced scores. The choice of model is significant. Sonnet 4 and Opus 4.1 are similar in most ways, but the different OpenAI models may have notably different outputs. Changing or upgrading a model – as in the move from GPT-4o to GPT-5 – could lead to radically different assessments of the quality of student work.

Therefore, if we choose to use AI in the assessment process, our process should include agreed-upon benchmarks based on past evaluations and plans to retest following model updates. At the level of classes and assignments, AI grading *could* be useful as a low-stakes evaluation, but it should not be used as a substitute for human judgment on significant work without careful benchmarking.

This pilot study opens major opportunities for further projects. These include:

- Benchmarking against other assessment tasks, including the Critical Thinking and Written Communication VALUE rubrics and assignments from individual classes.

- Investigating model preferences by asking a model to generate artifacts that, based on its internal understanding, achieve different levels of performance on a rubric.

- Systematic testing on a wider array of models, including the Google Gemini line. Fine tuning models to better encode alignment with human preferences.

- Testing whether AI models are better suited for some grading tasks than others. For example, scoring against a multi-category rubric with five categories and five levels may be too difficult, but binary pass/fail classifications maybe more reliable, which could create opportunities to incorporate automated AI evaluation into specifications grading systems

- Evaluating the institutional tradeoffs between potential efficiency gains of AI-supported assessment vs. the benefits of in-depth human engagement with the assessment process

Appendix – Complete Prompt

The text below is the three-shot prompt used for all models in our experiments. The example student artifacts have been replaced with placeholders. The prompt for our zero-shot experiments is the same with the "Example Assessment Results" section removed.

```
# Information Literacy Assessment Prompt

You are an expert academic evaluator specializing in information
literacy assessment. Your task is to evaluate student-created
artifacts using the Information Literacy VALUE Rubric.

## Assessment Instructions

Evaluate the provided student artifact PDF file against each of the five
Information Literacy criteria. Assign a numerical score for each
category based on the performance descriptors:

- **4 (Capstone)**: Meets the highest level of proficiency
- **3 (Milestones)**: Meets the higher intermediate level
- **2 (Milestones)**: Meets the lower intermediate level
- **1 (Benchmark)**: Meets the basic level of proficiency
- **0**: Fails to meet even basic proficiency

## Scoring Reference

Use the detailed criteria and performance descriptors provided in the
Information Literacy VALUE Rubric Markdown document to evaluate each
category. Apply the specific performance indicators for each proficiency
level (Capstone, Milestones, Benchmark) as defined in the rubric.

## Required Output Format

Return your assessment as a JSON object with this exact structure:

```json
{
  "determine_extent": {
    "score": [0-4],
    "reasoning": "[Brief explanation of score in 1-2 sentences]"
  },
  "access_information": {
    "score": [0-4],
    "reasoning": "[Brief explanation of score in 1-2 sentences]"
  },
  "evaluate_critically": {
    "score": [0-4],
```
```

```
    "reasoning": "[Brief explanation of score in 1-2 sentences]"
  },
  "use_effectively": {
    "score": [0-4],
    "reasoning": "[Brief explanation of score in 1-2 sentences]"
  },
  "access_ethically": {
    "score": [0-4],
    "reasoning": "[Brief explanation of score in 1-2 sentences]"
  }
}
```

Important Guidelines

- Provide ONLY the JSON output with no additional text, explanations, or commentary
- Keep reasoning statements concise (1-2 sentences maximum)
- Base scores strictly on observable evidence in the artifact
- Be consistent in applying the rubric criteria across all categories
- If information is insufficient to make a determination, score conservatively and note the limitation in reasoning

Example Assessment Results

Use the three examples below to guide your evaluations.

High-scoring artifact

The following artifact was assessed by three human evaluators and received the following scores in each of the five categories. It is an example of a generally high-scoring artifact that received scores of 3 and 4 in most categories.

Determine the extent of information needed: 3, 4, 4 (average 3.66)
Access the needed information: 3, 4, 3 (average 3.33)
Evaluate information and its sources critically: 2, 3, 4 (average 3.0)
User information effectively to accomplish a specific purpose: 3, 3, 4 (average 3.33)
Access and use information ethically and legally: 2, 3, 4 (average 3.0)

<<TEXT OF ARTIFACT INSERTED HERE>>

Low-scoring artifact

The following artifact is an example of one that received scores of 1 or 2 in most categories. The three human evaluators scored it as follows.

Determine the extent of information needed: 2, 0, 2 (average 1.33)
Access the needed information: 2, 2, 1 (average 1.66)
Evaluate information and its sources critically: 2, 1, 1 (average 1.33)
User information effectively to accomplish a specific purpose: 2, 2, 1 (average 1.66)
Access and use information ethically and legally: 2, 1, 2 (average 1.66)

<<TEXT OF ARTIFACT INSERTED HERE>>

Zero-scoring artifact

The following artifact is an example of one that received scores of 0 or 1 in most categories. The three human evaluators scored it as follows.

Determine the extent of information needed: 1, 1, 0 (average .66)
Access the needed information: 1, 0, 0 (average .33)
Evaluate information and its sources critically: 1, 1, 1 (average 1.0)
User information effectively to accomplish a specific purpose: 1, 0, 0 (average .33)
Access and use information ethically and legally: 0, 0, 0 (average 0.00)

<<TEXT OF ARTIFACT INSERTED HERE>>

Rubric Document:

Information Literacy VALUE Rubric

For more information, please contact value@aacu.org

The VALUE rubrics were developed by teams of faculty experts representing colleges and universities across the United States through a process that examined many existing campus rubrics and related documents for each learning outcome and incorporated additional feedback from faculty. The rubrics articulate fundamental criteria for each learning outcome, with performance descriptors demonstrating progressively more sophisticated levels of attainment. The rubrics are intended for institutional-level use in evaluating and discussing student learning, not for grading. The core expectations articulated in all 15 of the VALUE rubrics can and should be translated into the language of individual campuses, disciplines, and even courses. The utility of the VALUE rubrics is to position learning at all undergraduate levels within a basic framework of expectations such that evidence of learning can be shared nationally through a common dialog and understanding of student success.

Definition

The ability to know when there is a need for information, to be able to identify, locate, evaluate, and effectively and responsibly use and share that information for the problem at hand. – Adopted from the National Forum on Information Literacy

Framing Language

This rubric is recommended for use evaluating a collection of work, rather than a single work sample in order to fully gauge students' information skills. Ideally, a collection of work would contain a wide variety of different types of work and might include: research papers, editorials, speeches, grant proposals, marketing or business plans, PowerPoint presentations, posters, literature reviews, position papers, and argument critiques to name a few. In addition, a description of the assignments with the instructions that initiated the student work would be vital in providing the complete context for the work. Although a student's final work must stand on its own, evidence of a student's research and information gathering processes, such as a research journal/diary, could provide further demonstration of a student's information proficiency and for some criteria on this rubric would be required.

Evaluators are encouraged to assign a zero to any work sample or collection of work that does not meet benchmark (cell one) level performance.

Rubric Criteria and Performance Levels

1. Determine the Extent of Information Needed

Capstone (4): Effectively defines the scope of the research question or thesis. Effectively determines key concepts. Types of information (sources) selected directly relate to concepts or answer research question.

Milestones (3): Defines the scope of the research question or thesis completely. Can determine key concepts. Types of information (sources) selected relate to concepts or answer research question.

Milestones (2): Defines the scope of the research question or thesis incompletely (parts are missing, remains too broad or too narrow, etc.). Can determine few concepts. Types of information (sources) selected partially relate to concepts or answer research question.

Benchmark (1): Has difficulty defining the scope of the research question or thesis. Has difficulty determining key concepts. Types of information (sources) selected do not relate to concepts or answer research question.

2. Access the Needed Information

Capstone (4): Accesses information using effective, well-designed search strategies and most appropriate information sources.

Milestones (3): Accesses information using variety of search strategies and some relevant information sources. Demonstrates ability to refine search.

Milestones (2): Accesses information using simple search strategies, retrieves information from limited and similar sources.

Benchmark (1): Accesses information randomly, retrieves information that lacks relevance and quality.

3. Evaluate Information and its Sources Critically

Capstone (4): Thoroughly (systematically and methodically) analyzes own and others' assumptions and carefully evaluates the relevance of contexts when presenting a position.

Milestones (3): Identifies own and others' assumptions and several relevant contexts when presenting a position.

Milestones (2): Questions some assumptions. Identifies several relevant contexts when presenting a position. May be more aware of others' assumptions than one's own (or vice versa).

Benchmark (1): Shows an emerging awareness of present assumptions (sometimes labels assertions as assumptions). Begins to identify some contexts when presenting a position.

4. Use Information Effectively to Accomplish a Specific Purpose

Capstone (4): Communicates, organizes and synthesizes information from sources to fully achieve a specific purpose, with clarity and depth.

Milestones (3): Communicates, organizes and synthesizes information from sources. Intended purpose is achieved.

Milestones (2): Communicates and organizes information from sources. The information is not yet synthesized, so the intended purpose is not fully achieved.

Benchmark (1): Communicates information from sources. The information is fragmented and/or used inappropriately (misquoted, taken out of context, or incorrectly paraphrased, etc.), so the intended purpose is not achieved.

5. Access and Use Information Ethically and Legally

Capstone (4): Students use correctly all of the following information use strategies (use of citations and references; choice of paraphrasing, summary, or quoting; using information in ways that are true to original context; distinguishing between common knowledge and ideas requiring attribution) and demonstrate a full understanding of the ethical and legal restrictions on the use of published, confidential, and/or proprietary information.

Milestones (3): Students use correctly three of the following information use strategies (use of citations and references; choice of paraphrasing, summary, or quoting; using information in ways that are true to original context; distinguishing between common knowledge and ideas requiring attribution) and demonstrates a full understanding of the ethical and legal restrictions on the use of published, confidential, and/or proprietary information.

Milestones (2): Students use correctly two of the following information use strategies (use of citations and references; choice of paraphrasing, summary, or quoting; using information in ways that are true to original context; distinguishing between common knowledge and ideas requiring attribution) and demonstrates a full understanding of the ethical and legal restrictions on the use of published, confidential, and/or proprietary information.

Benchmark (1): Students use correctly one of the following information use strategies (use of citations and references; choice of paraphrasing, summary, or quoting; using information in ways that are true to original context; distinguishing between common knowledge and ideas requiring attribution) and demonstrates a full understanding of the ethical and legal restrictions on the use of published, confidential, and/or proprietary information.

Association of American Colleges and Universities