



Spider Web Rubrics

Reviewer Instructions

Updated: Sep 23, 2025

Table of contents

[Overview](#)

[High-Level Workflow](#)

[Detailed Workflow](#)

[Reviewer Grading Rubric](#)

[Feedback](#)

[Common Errors to Avoid](#)

Change log

Date	Change
Sep 23, 2025	<ul style="list-style-type: none">• Document created

Overview

Reviewers are critical to ensuring the highest quality standards. You are the first line of defense, and it is your responsibility to ensure that the tasks you review meet the project's expectations!!

These instructions are meant to guide you on how to perform great reviews and how to handle different case scenarios. Please refer to the attempter's [instructions](#) for more detailed information about the tasks.

Tara Eval

You must now use the **Eval Tool** to view feedback for all your reviewed tasks.

This is the **official, required method**.

Just paste your **attempt ID** from Outlier → hit **Submit** → and view feedback in seconds.

See clear sections: **Prompt**, **Rubric Criteria**, **Prompt Eval**, and **Rubric Eval**.

⚠ Allow up to 5 minutes for new tasks to appear.

✨ Faster, cleaner, and mandatory for all reviewers → taraeval.vercel.app

High-Level Workflow

Step 1. Read through the whole task from start to finish.

1. Read the prompt
2. Read the model response
3. Read the Quality Control Notes and the Google Colab notebook
4. Read the rubric/criteria
5. Read the criteria categorization: Must Have/Nice to have, and Weight

Step 2. Score the attempt.

Provide a score for the original attempt score using the [Rubric](#) below to rate the quality of the task holistically as you received it (before any edits).

Step 3. Select the “Action” that corresponds to the “Score” following the [Rubric](#).

- **Approve** → the task is a **4 or 5** in overall task quality
- **Fixed** → the task is a **3** in overall task quality or **2 or 1** if there are egregious errors, depending on the severity of the errors

Step 4. Fix the task.

We will be fixing all the tasks at the review level. We don't have an SBQ button for this project. **If you don't feel comfortable with the complexity level of a task, we encourage you to skip it.**

What fields are we fixing?

- Prompt (If they didn't produce model failures)
- Quality Control Notes
- Rubric
 - Criteria
 - Criteria classifications

Step 5. Provide feedback for the attempter.

Ensure to provide detailed feedback to help the Attempter improve the quality of their tasks and learn from their mistakes. Please refer to the Feedback section for additional information.

Detailed Workflow

Review the prompt

- Review that the prompt follows the requirements of a high-quality prompt:
 - **Use casual and natural language:** The questions should be realistic, with no markdown or fancy formatting. Do not be formal: Avoid questions that sound like exam questions!
 - **Write hard questions to stump the models.** If the prompt is too easy, the task will need to be redone, as you won't find model failures in the responses.
 - **Asking more than one question in one prompt is ok**, but the questions must be centered on the same task/intent, not two completely different topics.
 - **If you reference a variable in your prompt:** Use the correct variable name used in the dataset
 - **Be precise in your prompt without revealing the solution.**
 - **Dataset Dependency:** Every prompt must rely on the attached datasets. The questions should not be answerable without using the datasets.
 - **[NEW] You must use all the attached datasets in your prompts**
 - **Specific and Well-Structured:** Prompts should be clear, detailed, and organized. They may include multiple related questions that require the model to read, manipulate, and reason through the data, as well as run commands to generate results.
 - **Realistic Scenarios:** Write prompts that feel natural for a data analyst. Some questions can be slightly contrived, but avoid anything completely unrealistic or irrelevant to real-world analysis.

- **Plot-Based Questions:** Include at least one question that requires the model to generate a plot.

- **Balance Open- and Close-Ended Prompts:**

- This is usually specified within the task

- If the prompt is not aligned with the prompt's best practices, make sure to fix it
- Review the attempter's instructions on high-quality prompts and examples.

Review Quality Control Notes and Google Colab notebook

- Ensure that the Quality Control Notes and Colab notebook are aligned with the content of the prompt

Review the rubric

- Make sure that the rubric is:
 - Comprehensive (no missing steps for an ideal response)
 - Relevant (no unnecessary or unrelated steps)
 - Accurate (steps that are objective and correct)
 - Non-Repeating (no redundant steps - don't penalize the same mistake twice)
 - Self-contained (sufficient details/examples are provided such that it is specific and not completely open-ended).
 - Reflect on all the explicit asks of the prompt
- If the rubric is missing one of the previous points, make sure to fix it
- Check the common errors to avoid section and keep a close eye on them

Review the criteria

- Make sure that the criteria are:
 - Appropriately atomic (related to a single challenge in the prompt)
 - Specific (containing sufficient detail, including answers/examples that do not leave it completely open-ended)
 - Accurate (correct by fact check/reason)
 - Categorized (all items in the criteria should fall under the categories)
- If the criteria is missing one of the previous points, make sure to fix it

Review the criteria categorization.

- Make sure that the criteria are properly categorized as Must Have/Nice To have
- Make sure that the **Weights** are aligned based on its importance to the prompt and the following guideline:
 - [1] Accuracy-related criteria (e.g., factual correctness, calculations, results) should receive the highest weights, since they are the most critical.
 - [2] Instruction-following criteria (e.g., formatting, phrasing, or minor task requirements) should be assigned lower weights than accuracy but still enough to matter.
- If the criteria are not properly categorized, make sure to fix them

Reviewer Grading Rubric

Score	Description	Action	Criteria
✗ 1	Spam/AI / Very Low Quality	Don't fix and assign the score with the feedback	<p>ANY of these conditions must be TRUE:</p> <ul style="list-style-type: none"> • AI-generated content: prompt, quality control notes, rubric • Blatant disregard of instructions • Prompt doesn't cause a model failure • Prompt needs to be completely redone <p>ANY of these conditions is TRUE:</p> <p>MAJOR ISSUES:</p> <p>The datasets fail in any of the following areas:</p> <ul style="list-style-type: none"> • No Publicly accessible (no logins or paywalls). • No Open source (free for use). • Not in English. • Not Sufficiently large → at least 5 columns and 100 rows. <p>Model Failure:</p> <ul style="list-style-type: none"> - The prompt doesn't produce a model response failure - The model failure, if present, is really weak (not pronounced) - The model response failure is based on an ambiguous interpretation of the prompt (i.e., the request can be reasonably interpreted in multiple ways and the model picked one of the valid interpretations) <p>Prompts:</p> <ul style="list-style-type: none"> -The conversation doesn't have at least 1 open-ended and 1 closed-ended prompt -There's no request for a plot/graph/chart in the conversation -Uses all datasets <p>Quality Control Notes:</p> <ul style="list-style-type: none"> - The Quality Control Notes are completely deviated from the prompt's intent
✗ 2	Very low quality, does not follow any of the instructions, and needs significant rework	Fix + approve	

Score	Description	Action	Criteria
CONFIDENTIAL kolevikasxyz@gmail.com+outlier 2025-12-14T05:25:46.157Z			A task should fail if more than 10% of criteria contain major errors. Major errors are defined directly below.
			1. [Counterproductive Criteria] Criteria that penalize good responses / reward bad responses such that the inclusion of the criteria would make the response objectively worse. The criterion contains an objective inaccuracy. Note: (Do not fail “negative-criteria” as counter-productive if they are phrased correctly and properly given a negative weight (e.g., “The response X [-10]” and X is something the model does incorrectly).
			2. [Missing Criteria] Criteria that are objectively 100% essential to meet the explicit asks of the prompt are missing in the rubric (not just a nice to have, but rather a serious omission). This can include criteria that explicitly contain a result, or outcome, in direct response to a prompt’s ask. Or, this can include criteria that respond to implicit, but objectively necessary, steps to satisfy a direct ask from the prompt. Missing criteria should fall into one of 3 buckets: Addressing safety of the response Addressing accuracy of the response Assessing instruction following of the response (Are all direct requests covered?)
			3. [Fail - Major Imbalance] There is an obvious and egregious misalignment in the magnitude of the assigned weight of the criterion against its actual importance to the prompt. (e.g., the criterion is weighted +10 when it should reasonably be rated +2) Positive weights are given to rubric criteria when they should have negative weights, or vice versa.
			4. [Not Self-Contained] The criterion has a Objective final answer but does not provide the

Score	Description	Action	Criteria
CONFIDENTIAL kolevikasxyz@gmail.com+outlier 2025-12-14T05:25:46.157Z			<p>The criterion has vague language (e.g., “should be accurate”), missing expected outcomes/ examples where relevant. The criterion does not contain all the information needed to evaluate a response independent of the prompt.</p> <p>E.g.</p> <p><input checked="" type="checkbox"/> Mentions the smallest moon of Saturn.</p> <p><input checked="" type="checkbox"/> Mentions the smallest moon of Saturn is Aegaeon.</p>
Note: Not Self Contained means it does not contain all the information needed to evaluate a response without relying on basic domain knowledge with negligible reasoning (definition comprehension and trivial equivalence identification)			
<p>5. [Plot-based criteria]</p> <p>-There's no plot/chart/graph image present within the rubric</p> <p>[Negative criteria]</p> <p>- There's no negative criteria (negative weights) present in the rubric per model failure</p> <hr/> <p>MODERATE ISSUES Rubric:</p> <p>A task should fail if 15% or more of the criteria contain moderate issues.</p> <p>A single criterion cannot be double counted for multiple issues to contribute to this failing condition.</p> <p>Use the [Fail - 15%+ Moderate Rubric Errors] error category and the relevant ones below</p> <p>[Fail - Atomicity]</p> <p>15% or more rubric criteria contain unrelated prompt requests combined into one rubric criterion.</p> <p>[Fail - Overlapping Criteria]</p>			

Score	Description	Action	Criteria
15%			<p>15% or more criteria that evaluate the same exact aspect of the response.</p> <p>Each set of overlapping criteria can only be counted as one failing criteria contributing to a “moderate issues” fail.</p> <p>[Plot-based criteria]</p> <ul style="list-style-type: none"> - There’s no additional criteria describing how the plot/chart/graph should look
25%			<p>MINOR ISSUES Rubric: For the following errors, the task should fail if 25% or more criteria contain any combination of these. A single criterion cannot be double counted towards the 25% strike policy.</p> <p>Use this Error Category for the following Criteria: [Fail - 25%+ Minor Rubric Errors] and the relevant ones below</p>
30%			<p>[Vague Criteria]</p> <p>The criterion is too vague or abstract to be meaningfully evaluated, making it impossible to determine whether it has been met.</p>
35%			<p>[Binary]</p> <p>Criterion is not binary (true or false) or objective (a majority of readers should agree on whether a given model response satisfies the criteria). Even if the criterion must rely on some level of personal interpretation, it should be something that >75% of people would rate the same way.</p>
40%			<p>[Double Negative]</p> <p>A negative criterion flags something absent from the response, instead of something present</p>
45%			<p>✗ [-5] The response fails to correctly calculate the proportion of debt funding as 92%</p> <p>✓ [+5] The response correctly calculates the proportion of debt</p>

Score	Description	Action	Criteria
CONFIDENTIAL kolevikasxyz@gmail.com+outlier 2025-12-14T05:25:46.157Z	CONFIDENTIAL kolevikasxyz@gmail.com+outlier 2025-12-14T05:25:46.157Z	CONFIDENTIAL kolevikasxyz@gmail.com+outlier 2025-12-14T05:25:46.157Z	<p>funding as 92%</p> <p>[Irrelevant Criteria] The criterion is not directly related to the prompt OR the criterion objectively adds no value to the response. If there is any reasonable argument for the inclusion of the criterion, the criterion should NOT be failed for this condition.</p> <p>Give the person the benefit of the doubt because this is inherently subjective.</p> <p>Note: Criteria that relate to the stylistic/formatting elements of the response should not be considered irrelevant if they support the CB's preference for visually pleasing responses</p> <p>[Fail - Combination of Rubric Errors]</p> <p>Combination of Issues:</p> <p>Use this error category if there exists fewer than 10% of criteria with a major error and enough separate criteria with moderate errors such that the total sum of errors is greater than or equal to 15% of all criteria, or enough separate criteria with minor errors such that the total sum of errors is greater than or equal to 25%. Note that criteria cannot be double counted</p> <p>i.e., the Moderate threshold (15%) can include separate Major issues as part of its tally, while the Minor threshold (25%) can include both Moderate and Major issues</p>
CONFIDENTIAL kolevikasxyz@gmail.com+outlier 2025-12-14T05:25:46.157Z	CONFIDENTIAL kolevikasxyz@gmail.com+outlier 2025-12-14T05:25:46.157Z	CONFIDENTIAL kolevikasxyz@gmail.com+outlier 2025-12-14T05:25:46.157Z	<p>ALL these conditions must be true:</p> <p>The datasets are:</p> <ul style="list-style-type: none"> • Publicly accessible (no logins or paywalls). • Open source (free for use). • In English. • Sufficiently large → at least 5 columns and 100 rows. <p>Model Failure:</p> <ul style="list-style-type: none"> - The prompt produces a valid model response failure) <p>Prompts:</p> <ul style="list-style-type: none"> - The conversation has at least 1 open-ended and 1 closed-ended

Score	Description	Action	Criteria
CONFIDENTIAL kolevikasxyz@gmail.com+outlier 2025-12-14T05:25:46.157Z	CONFIDENTIAL kolevikasxyz@gmail.com+outlier 2025-12-14T05:25:46.157Z	<p>prompt</p> <ul style="list-style-type: none"> - Requests for a plot/graph/chart in the conversation - Uses all datasets 	<p>And ANY of these conditions is TRUE:</p> <p>Quality Control Notes:</p> <ul style="list-style-type: none"> • - The Quality Control Notes are mostly aligned with the prompt intent. There could be a more efficient way to solve the prompt, but it is generally okay <p>Rubric:</p> <ul style="list-style-type: none"> ● [Non-Fail - 5-10% Major Error] <ul style="list-style-type: none"> ○ Between 5 and 10% of distinct criteria satisfy the fail condition for a major error. (See Major Issues defined to the left) ● [Non-Fail - 10-15% Moderate Errors] <ul style="list-style-type: none"> ○ Between 10 and 15% of distinct criteria satisfy a fail condition for a moderate error. (See Moderate Issues defined to the left) ● [Non-Fail - 15-25% Minor Errors] <ul style="list-style-type: none"> ○ Between 15 and 25% of distinct criteria satisfy a fail condition for a minor error. (See Minor Issues defined to the left) ● [Non-Fail - Broad but Ratable Criteria] More than 25% of criteria are broad but most people would be able to determine if they were met (e.g., “the response should be humorous”) <p>[Non-Fail - Minor Imbalance]</p> <ul style="list-style-type: none"> ● More than 25% of the rubric criteria weights are roughly correlated with their importance, with some room for debate but no severe errors. ● Combination of Issues: Use this error category if there exists fewer than 10% of criteria with a major error and enough separate criteria with moderate errors such that the total sum of errors is between 10 and 15% of all criteria, or enough separate criteria with minor errors such that the total sum of errors is between 15 and 25%. Note that criteria cannot be

Score	Description	Action	Criteria
CONFIDENTIAL kolevikkaskyz@gmail.com+outlier 2025-12-14T05:25:46.157Z			double-counted:
★ 4-5	Perfect !! Well done, creative, and completely followed the instruction guide	Approve CONFIDENTIAL kolevikkaskyz@gmail.com+outlier 2025-12-14T05:25:46.157Z	<p>ALL these conditions must be true:</p> <p>The datasets are:</p> <ul style="list-style-type: none"> • Publicly accessible (no logins or paywalls). • Open source (free for use). • In English. • Sufficiently large → at least 5 columns and 100 rows. <p>Model Failure:</p> <ul style="list-style-type: none"> - The prompt produces a valid model response failure) <p>Prompts:</p> <ul style="list-style-type: none"> - The conversation has at least 1 open-ended and 1 closed-ended prompt - Requests for a plot/graph/chart in the conversation - Uses all datasets <p>Rubric</p> <ul style="list-style-type: none"> • The rubric covers the core instruction • The rubric does not contain objective inaccuracies • The rubric covers all prompt constraints • The rubric is specific and relevant • If needed the rubric provides direct answers • The rubric provides binary criteria • The rubric covers all the necessary criteria to create a perfect response. • The rubric criteria weights accurately reflect their importance to a good response • At least 1 criterion contains a plot plot/chart/graph • There are additional criteria describing how the plot/chart/graph should look • There are negative criteria for each model failure <p>Difference between a 4 and a 5:</p> <p>4 = Meets all the criteria</p> <p>5 = Stellar task, out of the ordinary, creative, and perfectly done</p>

Feedback

Provide helpful feedback! You are super critical in helping contributors get better at the project. Please provide concise but detailed and helpful comments about what was right and what was wrong with the task. Use the following guidelines:

- The feedback must be actionable and insightful.
- The goal of feedback is to help contributors improve, so make sure you *clearly* identify areas of improvement and explain how their attempt could have been better.
- Similar to good justifications, good feedback is detailed, points to examples, and explains the rationale behind your reviewer score.
- Although the structure of feedback is not strict, try to sandwich the bad in between the good *when you can*.

Example:

- i. Positive note (i.e., great writing style and creativity).
 - ii. Critique (i.e., there were issues with the ratings on certain dimensions, also reflected in the preference ranking and Likert scale).
 - iii. Another positive note (i.e., the flow between turns was smooth).
- **Note:** When a task gets scored as *less than 3*, it's even more important to provide quality feedback since the contributor clearly has room for improvement. It's expected that feedback for lower-scored tasks won't follow the structure described above.

Examples of Bad Feedback:

- “I disagree. It doesn’t make sense.”
- “Horrible - please improve”
- “Great - keep it up!”
- “Your score of 3 should actually be a score of 4...”

Examples of Good Feedback:

- “Excellent use of creativity! The model incorrectly changed the sign in step 3, and that caused the final answer to fail. Your final answer was correct, however, you provided a justification instead of only putting the final answer, which is the goal of the project.”
- “The prompt was very good and caused the model to fail, however, the final answer you provided is incorrect. The answer should be 3.45 m^2 , instead of 6.45 m^2 .”

NOTE: We have zero tolerance for unprofessional and harmful feedback. If we catch this behavior, you'll lose your reviewer permissions. Feedback should always be constructive and focused on how to improve the attempter's performance.

Common Errors to Avoid

CRITICAL - This needs to be fixed immediately, or the task won't pass Quality Control (QC)

IMPORTANT - Having this issue will increase the risk of this task not passing Quality Control (QC)

Status	Common Error	Example
CRITICAL	Incorrect wording	<p>Example.</p> <p><input checked="" type="checkbox"/> "Calculates a p-value of p=0.0001 for the correlation."</p> <p><input type="checkbox"/> "The response must calculate the p-value of p=0.0001 for the correlation."</p>
CRITICAL	Criteria are not specific enough	<p>Example Criteria:</p> <p><input type="checkbox"/> Calculates all the statistical measures over every sector.</p> <p><input checked="" type="checkbox"/> Calculates the mean Earnings for healthcare technology to be \$194.9M</p>
CRITICAL	Combined criteria	<p>Example Criteria:</p> <p><input type="checkbox"/> Provides top 3 and bottom 3 of each sector for every metric</p> <p>Solution: For each sector there has to be a separate rubric item outlining the top 3 and a separate rubric item outlining bottom 3</p> <p><input checked="" type="checkbox"/> States that top 3 in Divident/Yield category are Drug retail, commodity chemicals and tobacco.</p> <p><input checked="" type="checkbox"/> States that the bottom 3 in Dividend/Yield category are the Food industry, the tourism sector, and the automotive industry.</p>
CRITICAL	Datasets that are not public source	<p>Example:</p> <p><input checked="" type="checkbox"/></p>

Status	Common Error	Example
CRITICAL 	<p>Explanation: Remember that the datasets must be Public Domain. Make sure to look for the “License” section to verify this. If your task contains datasets that are not public domain, your task will automatically fail.</p>	<p>Usability ⓘ 9.12</p> <p>License <u>CC0: Public Domain</u></p> <p>Expected update frequency Daily</p> <p>Tags</p> <ul style="list-style-type: none"> Health Public Health
Missing essential criteria	<p>Explanation: When writing a rubric, all essential criteria to answer the prompt should be included. Missing essential criteria will automatically make the task obtain a low quality score.</p>	<p>Prompt: Give me 2 different Python scripts I can use to convert CSV to JSON</p> <p>Bad Rubric (missing to include a criterion that evaluates whether the model includes 2 Python scripts):</p> <ul style="list-style-type: none"> - The Python converts a CSV document into a JSON. - Explains that the script needs to include csv.DictReader() to convert the first row to keys. - Explains that each row needs to be converted into a dictionary. - Explains that all rows are stored in a list and then dumped to JSON using json.dump(). - Explains that including indent=4 makes the output human-readable. <p>Good Rubric:</p> <ul style="list-style-type: none"> - The Python converts a CSV document into a JSON. - Explains that the script needs to include csv.DictReader() to convert the first row to keys. - Explains that each row needs to be converted into a dictionary. - Explains that all rows are stored in a list and then dumped to JSON using json.dump(). - Explains that including indent=4 makes the output human-readable. <p>Provides 2 different Python scripts</p>

Status	Common Error	Example
CRITICAL	Repetitive criteria Explanation: When writing a rubric, each criterion should address a part to evaluate; therefore, repeating criteria is not allowed, as it will lead to over-penalization of the model.	<p>Prompt: Give me 2 different Python scripts I can use to convert CSV to JSON</p> <p>Bad Rubric (missing to include a criterion that evaluates whether the model includes 2 Python scripts):</p> <ul style="list-style-type: none"> - The Python converts a CSV document into a JSON. - Explains that the script needs to include <code>csv.DictReader()</code> to convert the first row to keys. - Explains that each row needs to be converted into a dictionary. - Explains that all rows are stored in a list and then dumped to JSON using <code>json.dump()</code>. - Explains that including <code>indent=4</code> makes the output human-readable. - Provides 2 different Python scripts - Explains that the script needs to include <code>csv.DictReader</code> to convert the first row to keys <p>Criteria 2 and 7 are almost identical, and we must only leave one.</p>
CRITICAL	Not Self-contained (Closed questions) Explanation: When you have a prompt with a closed question, the criteria should include the answer.	<p>Prompt: What's the capital of France?</p> <p>Bad Criterion: The response must include the capital of France.</p> <p>Good Criterion: The response must mention that the capital of France is Paris.</p> <hr/> <p>Prompt: I'm building a VR experience in Unreal Engine using C++ and Blueprints. I want to trigger dynamic haptics and spatial audio when the player interacts with certain world objects. I know Unreal supports force feedback and audio attenuation, but need C++-level details. Please search Unreal Engine's C++ API or headers for:</p> <ul style="list-style-type: none"> • Classes/methods for controller vibration. • Components/properties for spatial sound. <p>Provide:</p> <ul style="list-style-type: none"> • Full class names and method signatures. • Their modules or header files. • A C++ snippet showing how to trigger both haptics and spatial audio on trigger volume overlap. <p>Any required modules/dependencies for compilation.</p> <p>Bad Criterion:</p>

Status	Common Error	Example
CRITICAL ●	<p>Missing examples (open-ended questions)</p> <p>Explanation: Rubric criteria should be clear and objective, with specific examples that show what is expected.</p>	<p>Prompt: Recommend some shoes for running long distances like half marathon.</p> <p>Bad Criterion:</p> <p>Includes running shoe models for running long distances.</p> <p>Good Criterion:</p> <p>Includes running shoe models for running long distances such as the Nike Alphafly 3, Saucony Endorphin Pro 4, Hoka Cielo X1, etc.</p>
		<p>Prompt: I'm building a VR experience in Unreal Engine using C++ and Blueprints. I want to trigger dynamic haptics and spatial audio when the player interacts with certain world objects. I know Unreal supports force feedback and audio attenuation, but need C++-level details. Please search Unreal Engine's C++ API or headers for:</p> <p>Classes/methods for controller vibration. Components/properties for spatial sound.</p> <p>Provide: Full class names and method signatures. Their modules or header files. A C++ snippet showing how to trigger both haptics and spatial audio on trigger volume overlap. Any required modules/dependencies for compilation.</p> <p>Bad Criterion: Explains any required dependencies or engine modules needed in the C++ build file for successful compilation.</p> <p>Good Criterion: Explains any required dependencies or engine modules needed in the C++ build file for successful compilation, such as Core, CoreUObject, Engine, InputCore, etc.</p>

Status	Common Error	Example
IMPORTANT	Subjective criteria Explanation: Criteria should be objective (True/False). Avoid criteria based on personal opinions.	<p>Bad Criterion: States that the best coding language is C++.</p> <p>Good Criterion: States that C++ is the best option to design a financial system because of its ability to handle large datasets and complex calculations efficiently.</p>
IMPORTANT	Overly prescriptive Explanation: Don't include criteria based on your own likes or dislikes. Focus on what's essential for a good response and what is stated in the prompt.	<p>Prompt: Give me some ideas of Python scripts I can use to convert CSV to JSON</p> <p>Bad Rubric:</p> <ul style="list-style-type: none"> - The Python converts a CSV document into a JSON. - Explains that the script needs to include csv.DictReader() to convert the first row to keys. - Explains that each row needs to be converted into a dictionary. - Explains that all rows are stored in a list and then dumped to JSON using json.dump(). - Explains that including indent=4 makes the output human-readable. - Provides exactly 3 different Python scripts.
IMPORTANT	Unnecessary style/formatting Explanation: Don't specify style or formatting unless the prompt requires it.	<p>Prompt: Give me 2 different Python scripts I can use to convert CSV to JSON</p> <p>Bad Rubric:</p> <ul style="list-style-type: none"> - The Python converts a CSV document into a JSON. - Explains that the script needs to include csv.DictReader() to convert the first row to keys. - Explains that each row needs to be converted into a dictionary. - Explains that all rows are stored in a list and then dumped to JSON using json.dump(). - Explains that including indent=4 makes the output human-readable. - Provides 2 different Python scripts - Is formatted in a table for easier reading. <p>Explanation of why these are bad criteria: The prompt is not requesting any specific formatting.</p>

Status	Common Error	Example
IMPORTANT 	Poor Grammar and Spelling in the Criteria Explanation: The criteria in the rubric should be written with perfect spelling. Starting each criterion with a capital letter, adding accents, and punctuation marks, and correcting misspelled words.	<p>Bad grammar: the answer must include the 4 suits of a poker deck: hearts, spades, diamonds and clubs.</p> <p>Correction: The answer must include the four suits of a poker deck: hearts, spades, diamonds, and clubs.</p>