

# Technical Report on Human Revision of the Alpaca52k Dataset

## 1 Introduction

---

This document records details in our cleaning process of the Alpaca52k instruction dataset, involving manual revision provided by language experts. To achieve this, we conducted a meticulous manual examination of 6k instruction pairs sampled from the Alpaca52k dataset. We engaged 17 language experts to review from nine different dimensions, encompassing basic correctness and advanced experiences. During the primary revision, deficiencies were identified in 46.8% of the examined instruction pairs. Subsequently, the language experts were asked to rewrite the identified low-quality instruction pairs. This generated an expert revision dataset consisting of approximately 2.3k revised instruction pairs and their original counterparts. To ensure the quality of the revision process, the 17 experts were divided into three units based on their expertise (i.e., working years), each responsible for revising their assigned instruction pairs. Each unit was assigned an owner whose responsibility was to assess the quality of the revised instruction pairs produced by unit members. The revision process strictly adhered to the criteria we propose, following the principle of “making all necessary revisions,” regardless of the importance of the revised dimensions. If an instruction pair was identified as lacking in one or more dimensions in the criteria, the expert was required to make substantial revisions in those dimensions until the instruction pair achieved a score of 95 or higher based on the criteria. Consequently, considering the workload of preliminary filtering, quality control, and primary revision, a total of 129 person-days were expended, resulting in 2301 instruction pairs receiving revisions either on the Instruction or Response side. Among the 2.3k revised pairs, 1079 of them underwent revisions on the Instruction side. In the following chapters, we describe details in instruction revision and responses revision, respectively.

# 2 Human Evaluation on Instructions

## 2.1 Scoring Criteria

Criteria for INSTRUCTION				
Level	Dimension	Description	Main Checklist	Score Range
Advanced Requirement	Contextualization	The instruction includes a rich context or effective prompting skills to facilitate detailed and accurate responses.	Check for scenarios, roles, examples, or other requirements, and for skills like chain-of-thought.	80-100
Basic Requirement	Feasibility	The instruction is clear, specific, feasible, and easily understandable.	Check for ambiguous or vague expressions, logical errors, or requests beyond the ability of an AI model.	0-80
	Readability	The instruction adheres to the conventions and stylistic norms of the target language.	Check for language-related issues such as grammar, spelling, and punctuations.	

Regarding the criteria for assessing the quality of the Instruction in an instruction pair in Table above, firstly, an Instruction should be grammatically correct and logically feasible. Readability issues may impede accurate understanding of user intent during the training process. Additionally, infeasible Instructions containing logical errors in the training dataset may prevent the model from learning correct connections between instructions and responses, thereby exacerbating the hallucination of tuned LLMs. Moreover, recent studies have shown that including more contextual information and details in user instructions leads to better model responses. Therefore, a high-quality Instruction should also be rich in specific contexts, such as requirements and examples.

The Instruction and Response are evaluated independently, yielding two separate scores ranging from 0 to 100 based on their respective criteria. While all dimensions are necessary, they vary in their significance to the overall human interaction experience. Consequently, the dimensions are grouped into two levels for Instruction based on their importance, which determines their contribution to the final score. The basic level (i.e., readability and feasibility) signifies dimensions that enable effective human instructions, with flaws in readability making models unable to correctly understand the user intents and deficiencies in feasibility making the desire unable to be fulfilled by current models. Any flaws in this level restrict the score to a maximum of 80. The advanced level encompasses higher requirements, which is to provide ample contexts for instructions, and accounts for the top 20 points in the criteria. To mitigate bias, evaluators are instructed to independently and separately assess each dimension.

## 2.2 Case Analysis in Human Revision on Instructions

### 2.2.1 Readability

Instruction (Original)	How to make a desert?
Instruction (Revised)	How to make a dessert?
Analysis	The word "desert" in the original instruction is incorrect. It should be "dessert," which refers to a sweet course.

Instruction (Original)	Storytell and explain a physics phenomenon
Instruction (Revised)	Tell a story and explain a physics phenomenon.
Analysis	The word "storytell" in the original instruction is not a standard English word (but there are words like "storyteller" or "storytelling"). It should be changed to the corresponding verb phrase "storytelling". Additionally, there is a missing period at the end of the original instruction, which should be added.

Instruction (Original)	How does climate change affect people living in a coastal regions?
Instruction (Revised)	How does climate change affect people living in coastal regions?
Analysis	There is a grammatical error in the original instruction. The article "a" in "a coastal regions" is unnecessary and should be removed.

## 2.2.2 Feasibility

<i>Instruction (Original)</i>	What do you think about this topic?
<i>Instruction (Revised)</i>	What are the main benefits of this approach?
<i>Analysis</i>	The original instruction is too broad and lacks a clear focus.

<i>Instruction (Original)</i>	How can I clean money?
<i>Instruction (Revised)</i>	My money is dirty. How can I clean it?
<i>Analysis</i>	The original instruction is ambiguous. The term "clean money" has two different meanings: one refers to the act of cleaning physical money to remove dirt or stains, while the other refers to the illegal process of money laundering through illicit channels.

<i>Instruction (Original)</i>	Please suggest me some essential words for IELTS.
<i>Instruction (Revised)</i>	Please provide me 10 essential words for IELTS.
<i>Analysis</i>	The original instruction has two issues: (1) The use of "some" is vague and not specific enough. (2) The phrase "suggest me something" is grammatically incorrect and not idiomatic.

<i>Instruction (Original)</i>	Summarize the contents of a news article about the effects of global warming: <a href="https://edition.cnn.com/global-warming-impacts/index.html">https://edition.cnn.com/global-warming-impacts/index.html</a>
<i>Instruction (Revised)</i>	Summarize the contents of a news article about the effects of global warming: <a href="https://edition.cnn.com/2021/08/09/world/global-climate-change-report-un-ipcc/index.html">https://edition.cnn.com/2021/08/09/world/global-climate-change-report-un-ipcc/index.html</a>
<i>Analysis</i>	The link provided in the original instruction is invalid and leads to a page that cannot be opened properly, which prevents the model from executing the corresponding task. The instruction needs to use a valid URL address.

### 2.2.3 Contextualization

<i>Instruction (Original)</i>	Please rewrite the following sentences to make them easier to understand.
<i>Instruction (Revised)</i>	You are a primary school teacher who teaches 7 to 8 year olds. Please rewrite the following sentences to make easily understandable to 7 or 8 years olds.
<i>Analysis</i>	For different user groups, the standard for "easier to understand" may vary. The optimized instruction supplements the role the model needs to play and the usage scenarios, guiding the model to generate more accurate responses.

<i>Instruction (Original)</i>	Write about productivity.
<i>Instruction (Revised)</i>	Write a blog post about the importance of productivity for small businesses.
<i>Analysis</i>	The optimized instruction supplements the form of the generated content and the usage scenarios, making the topic more specific and guiding the model to generate content that matches the language style of the usage scenario.

<i>Instruction (Original)</i>	Write a poem.
<i>Instruction (Revised)</i>	Write a short inspiring poem in the style of Edgar Allan Poe about leaves falling.
<i>Analysis</i>	The original instruction lacked sufficient details. The optimized instruction supplements the tone, style, and theme of the poem to be generated, guiding the model to generate more accurate responses.

<i>Instruction (Original)</i>	<p>What is the 4<sup>th</sup> word in the phrase "I am not what I am"?</p> <p>Incorrect response from some AI model:</p> <p>The 4th word in the phrase "I am not what I am" is <b>not</b>. 🤔</p>
<i>Instruction (Revised)</i>	<p>What is the 4<sup>th</sup> word in the phrase "I am not what I am"? Let's think step by step.</p> <p>Correct response reference:</p> <p>Okay, let's think step by step. 😊</p> <ol style="list-style-type: none"> <li>1. The phrase "I am not what I am" has 5 words.</li> <li>2. Each word is separated by a space.</li> <li>3. The first word is "I", the second word is "am", the third word is "not", the fourth word is "what", and the fifth word is "I".</li> <li>4. Therefore, the 4th word in the phrase "I am not what I am" is <b>what</b>. 🙌</li> </ol>
<i>Analysis</i>	Enhancing model response by incorporating guided thinking approach such as Chain-of-Thought skills.

# 3

## Human Evaluation on Responses

### 3.1 Scoring Criteria

Criteria for RESPONSE				
Level	Dimension	Description	Main Checklist	Score Range
Advanced Experience	Humanization	Responses should be warm, empathetic, and engaging, tailored to the user's background and preferences.	Check: (1) Emotional Perception. Respond to users' emotions with empathy; (2) Humanized Tone. Interact with users in a natural and friendly way, avoiding machine-like tone.	90-100
	Richness	Responses should be diverse, informative, creative, and expanded.	Check: (1) Provide detailed and diverse information with depth and breadth; (2) Enrich the content with novelty, uniqueness, and imagination.	80-90
Basic Experience	Readability	Responses should use fluent, concise and correct language and be properly structured.	Check (1) Language: Error-free writing using precise vocabulary; (2) Content: Meaningful content without redundancy; (3) Structure: Clear, ordered, and logical organization of information with user-friendly layout.	40-80
	Comprehensiveness	Responses comprehensively cover all necessary angles and information.	Check (1) No omissions or deficiencies in fully explaining user questions. (2) Multiple angles, sufficient contexts and details for an unbiased response.	
	Relevance	Responses should be effective and direct, and provide in-topic solutions.	Check (1) Irrelevance: Response misinterprets user's intention; (2) Deviation: Response is related to user's topic, but deviates from the focus.	
	Correctness	Responses should be grounded in factual information, common sense, and logical reasoning, while also staying up-to-date and adhering to the user's specific requirements.	Check (1) Factual Error: Inconsistent with reality; (2) Common Sense Error: Contradict with human common sense; (3) Logical Error: Include concept substitution, self-contradiction, ambiguity and circular reasoning, etc.; (4) Compliance with Constraints: Include word count, genre and style, etc.; (5) Timeliness: The provided information is up-to-date.	
Experience Red Line	Safety	Responses should be harmless, protecting users' emotions, body and property.	Check for violation of laws, personal attacks, exposure of user privacy and irresponsible advises on medical or financial matters.	0-40

A high-quality Response to the user's instruction ensures a desirable user experience. Firstly, the red line of a Response is the safety aspect for the user and other entities. Additionally, a basic requirement for a good user experience is a relevant and comprehensive response without factual and language errors. Furthermore, providing a Response with expanded information and a humanized tone is essential for delivering an advanced user experience. There are three levels: the red-line level represents the minimum safety standard, the basic level signifies correctness and relevance, and the advanced level includes higher expectations such as rich context and politeness.

## 3.1 Case Analysis in Human Revision on Responses

### 3.1.1 Safety

<i>Instruction</i>	Categorize the following item as either edible or inedible.
<i>Input</i>	Glass
<i>Response (Original)</i>	Glass is generally considered an edible item, as it can be consumed in various ways, such as drinking, eating, or using it as a container. However, Glass may also be considered an inedible item if it is not consumed, such as on the ground or in water.
<i>Score</i>	10
<i>Analysis</i>	The original response believed that glass is edible, which not only goes against facts and common sense but may also mislead users and lead to them mistakenly ingesting glass and causing harm to their bodies. It is necessary to not only provide a proper answer to the question but also remind users of the dangerous consequences of swallowing glass.
<i>Response (Revised)</i>	<p>The item glass is <b>**inedible**</b>, which means it cannot be eaten or digested by humans or animals.</p> <p>Glass is a non-crystalline solid that is usually transparent or translucent and made from sand, lime, soda, and other ingredients. Glass is used for various purposes, such as making windows, bottles, tableware, and optics. Glass is hard, brittle, and chemically inert, which means it does not react with most substances.</p> <p>Glass can also be dangerous if swallowed, as it can cause cuts, bleeding, infection, or obstruction in the digestive system. Therefore, glass should never be consumed or put in the mouth.</p>



### 3.1.2 Correctness

#### 1. Factual Error

Instruction	Find a song title that consists of two animals.
Input	None
Response (Original)	One song title that consists of two animals is "Lions in the Snow" by The Black Eyed Peas.
Score	20
Analysis	The song "Lions in the Snow" provided in the original response does not actually exist. It may have been fabricated by the AI model, and it also does not meet the requirement of including two different animals.
Response (Revised)	<p>There are many song titles that consist of two animals, but here are some examples that you might know or like:</p> <ul style="list-style-type: none"><li>- The Lion and the Wolf by Thrice, which is a rock song that uses the metaphor of two animals to describe a conflict between two people.</li><li>- The Lion and the Lamb by Big Daddy Weave, which is a Christian song that praises God as both powerful and gentle.</li><li>- The Bird and the Worm by The Used, which is a punk rock song that uses the imagery of two animals to express the feelings of being trapped and hopeless.</li></ul> <p>These are just some of the many songs that have two animals in their titles. You can find more by searching on the web or listening to different genres of music.</p>

<i>Instruction</i>	Which company created the software program MS Word?
<i>Input</i>	None
<i>Response (Original)</i>	Microsoft created MS Word. MS Word is a free, open-source, document editing software developed and released in 1987 by Microsoft. It was first released as an add-on to the Microsoft Office package, and has since become one of the core products of Microsoft Office.
<i>Score</i>	50
<i>Analysis</i>	The company Microsoft provided in the original response is correct, but there are several factual errors in the explanation of MS Word. For example, MS Word is not open-source software (it is actually proprietary software), and its release date is not 1987 (it is 1983).
<i>Response (Revised)</i>	<p>The company that created the software program MS Word is <b>**Microsoft**</b>.</p> <p>MS Word, or Microsoft Word, is a word processor that allows users to create, edit, format, and share documents online or offline. It's part of the Microsoft 365 suite of applications that includes Word, Excel, PowerPoint, Outlook, and more.</p>

## II. Common Sense Error

<i>Instruction</i>	What happens if you go outdoors in cold weather with wet hair?
<i>Input</i>	None
<i>Response (Original)</i>	If you go outdoors in cold weather with wet hair, it can get quite cold and cause your head to feel very hot. The moisture in your hair will act like a blanket, trapping the heat from escaping. This can lead to a rapid increase in temperature, which may result in frostbite or hypothermia. It's important to dry your hair thoroughly before going out into the cold.
<i>Score</i>	14
<i>Analysis</i>	The original response contains a description that goes against common sense. It states that not blow-drying one's hair before going into a cold outdoor environment will make the head become hot.
<i>Response (Revised)</i>	<p>If you go outdoors in cold weather with wet hair, you are likely to feel much colder than if your hair was dry. This is because wet hair conducts heat away from your body much faster than dry hair, which can cause your body temperature to drop more quickly. Additionally, if the temperature is below freezing, the water in your hair can freeze and cause your hair to become brittle and break. In extreme cases, going outdoors with wet hair in very cold weather can even lead to hypothermia, which is a dangerous condition that occurs when your body loses heat faster than it can produce it. Therefore, it is best to dry your hair thoroughly before going outside in cold weather.</p> <p>I hope this answers your question.</p>

### III. Logical Error

<i>Instruction</i>	Is Baltic Sea the lowest salinity sea among the following three oceans?
<i>Input</i>	The Baltic Sea, Arctic Ocean, Red Sea.
<i>Response (Original)</i>	According to this map, the Baltic Sea is not the lowest salinity sea among the given three oceans. The lowest salinity sea is the Arctic Ocean, which has a salinity level of about 18%. The Baltic Sea and the Red Sea are both freshwater seas, with a salinity level of around 7% and 3%, respectively.
<i>Score</i>	30
<i>Analysis</i>	The original response is contradictory. It states that the Arctic Ocean has the lowest salinity but provides a salinity value (18%) for the Arctic Ocean that is higher than the salinity values for the other two oceans. This is a logical error.
<i>Response (Revised)</i>	Yes, the Baltic Sea has the lowest salinity among the three oceans mentioned. The average salinity of the Baltic Sea is about 10 parts per thousand (ppt), while the Arctic Ocean has an average salinity of about 30 ppt and the Red Sea has an average salinity of about 40 ppt.

#### IV. Compliance with Constraints

<i>Instruction</i>	Rewrite the following sentence to avoid using the word "happy":
<i>Input</i>	Paul was happy when he found the key.
<i>Response (Original)</i>	<p>Here are some possible ways to rewrite the sentence to avoid using the word "happy":</p> <ul style="list-style-type: none"><li>- Paul found the key and was happy.</li><li>- He found the key and was happy.</li><li>- The key was found and Paul was happy.</li></ul>
<i>Score</i>	40
<i>Analysis</i>	The few paraphrased sentences provided in the original response did not adhere to the user's request (to avoid using the word "happy"), as they all included the word "happy".
<i>Response (Revised)</i>	<p>Here are some possible ways to rewrite the sentence to avoid using the word "happy":</p> <ul style="list-style-type: none"><li>- Paul's face lit up when he found the key.</li><li>- Paul felt a surge of joy when he found the key.</li><li>- Paul's mood lifted when he found the key.</li></ul>

## V. Timeliness:

Instruction	Calculate the population density of Spain.
Input	None
Response (Original)	The population density of Spain is 91.9 people per square kilometer, which can be converted to 1 person/ square meter (in both cases, we need to multiply the population density by the land area of Spain to get the population per square meter).
Score	25
Analysis	The population density data provided in the original response is outdated and does not mention that population density can vary over time and due to other factors.
Response (Revised)	<p>Population density is a measurement of population per unit land area. To calculate the population density of Spain, we need to know two things: the population and the land area of the country.</p> <p>According to the data I've found, the population of Spain as of April 1, 2023 was about 48 million, and the total land area of Spain is about 0.5 million square kilometers. Therefore, we can use the following formula to calculate its population density:</p> $\text{Population density} = \text{Population} / \text{Land area} = 48 / 0.5 = 96$ <p>Therefore, the population density of Spain is about 96 people per square kilometer.</p> <p>Note that the population density can change over time, as the population and the area of a country may vary due to various factors, such as births, deaths, migration, natural disasters, wars, territorial disputes, environmental changes, and so on. Therefore, population density is usually calculated and reported for a specific year or period.</p>

## VI. Originality

Instruction	Write a poem about decision-making and its implications.
Input	None
Response (Original)	<p>Two roads diverged in a yellow wood,  And sorry I could not travel both.    And be one traveler, long I stood.  And looked down one as far as I could,  To where it bent in the undergrowth;    Then took the other, as just as fair,  And having perhaps the better claim,  Because it was grassy and wanted wear;  Though as for that the passing there  Had worn them really about the same,    And both that morning equally lay  In leaves no step had trodden black.    Oh, I kept the first for another day!  Yet knowing how way leads on to way,  I doubted if I should ever come back.</p>
Score	23
Analysis	The poem provided in the original response is not original content but a poem written by Robert Frost, which does not meet the requirement of originality.
Response (Revised)	<p>Here is a poem I wrote about decision-making and its implications:    Every day we face a choice,  To act or not, to raise our voice,  To follow or to lead the way,  To stay or go, to leave or stay.    Each decision has a cost,  A gain, a loss, a bridge, a cross,  A consequence, a ripple effect,  A chance, a risk, a hope, a threat.    Some choices are easy to make,  Others require a leap of faith,  Some are guided by reason and logic,  Others are driven by passion and magic.    But no matter what we choose to do,  We must be ready to face the truth,  That every choice we ever made,  Has shaped the life we live today.    I hope you like it.</p>

### 3.1.3 Relevance

#### 1. Irrelevance

<i>Instruction</i>	Given the following paragraph, summarize it in one sentence.
<i>Input</i>	<p>The man, who is in his 50s, also sustained a suspected broken leg in the first attack on Castle Street at about 18:45 BST on Saturday.</p> <p>Police said he was taken to hospital after a row on a garage forecourt.</p> <p>A short time later, another man and a woman were assaulted in licensed premises on Castle Street.</p> <p>Detectives have appealed for anyone who witnessed the attacks to contact police.</p>
<i>Response (Original)</i>	A man has been arrested on suspicion of attempted murder after a man was attacked in County Antrim.
<i>Score</i>	9
<i>Analysis</i>	The summary provided in the original response is completely unrelated to the given input paragraph, and the information summarized cannot be found in the input.
<i>Response (Revised)</i>	<p>One possible sentence to summarize the paragraph is:</p> <p>A man in his 50s was hospitalized with a broken leg after being assaulted on a garage forecourt, followed by another attack on a nearby pub that injured two more people.</p>



## II. Deviation

<i>Instruction</i>	Where do people smoke more on average than in Russia?
<i>Input</i>	None
<i>Response (Original)</i>	People smoke more on average than in Russia, but there are some regions where it is higher than others. According to data from the World Health Organization, the highest rate of smoking among men is 23.4%, which is about 1.6% higher than the national average. The highest rate of smoking among women is 18.9%, which is about 0.7% higher than the national average.
<i>Score</i>	22
<i>Analysis</i>	The original response answered the user's question but deviated from the original topic by comparing Russia to the worldwide average data after mentioning that some regions have a higher proportion of smokers than Russia.
<i>Response (Revised)</i>	<p>According to the latest available data about cigarette consumption from the year of 2016, people smoke more on average than in Russia in the following countries or regions: Andorra, Luxembourg, Belarus, North Macedonia, Albania, Belgium, Czech Republic, Jordan, Syria, Slovenia, Greece, Hungary, China, Lebanon, Armenia, Mongolia, Cyprus, Austria, Georgia, and Serbia.</p> <p>These countries have higher annual cigarette consumption per person aged 15 or older than Russia.</p> <p>I hope this answers your question.</p>

### 3.1.4 Comprehensiveness

#### 1. Omissions

<i>Instruction</i>	Extract all the named entities in the sentence below.
<i>Input</i>	Steve Jobs was the CEO of Apple Inc. He was born on February 24, 1955, San Francisco.
<i>Response (Original)</i>	The named entities are Steve Jobs and Apple Inc..
<i>Score</i>	5
<i>Analysis</i>	The original response omitted some named entities, such as "San Francisco" (a place) and "February 24, 1955" (a date).
<i>Response (Revised)</i>	<p>Here is a list of the named entities in the sentence:</p> <p><b>**Steve Jobs**</b>: Person <b>**Apple Inc. **</b>: Organization <b>**February 24, 1955**</b>: Date <b>**San Francisco**</b>: Location</p> <p>I hope this is helpful for you.</p>

## II. Insufficient Angles

<i>Instruction</i>	I have a jar of coins and I think it is not the right number of coins. Can you help me count the coins?
<i>Input</i>	None
<i>Response (Original)</i>	Sure, I can help you count the coins. First, empty the coins out of the jar onto a flat surface and then separate them into piles by denomination. Count up the number of coins in each pile, then add the totals together to get the total number of coins in the jar.
<i>Score</i>	55
<i>Analysis</i>	There are multiple possible ways to count coins, and the original response only provided one inefficient and time-consuming method, which is not comprehensive enough.
<i>Response (Revised)</i>	<p>Sure! I'm glad you asked me to help you count the coins. There are a few ways to do it, depending on what you have available.</p> <ul style="list-style-type: none"> <li>- One way is to use a coin counting machine, which can sort and count the coins for you automatically. You can find these machines at some banks, supermarkets, or arcades.</li> <li>- Another way is to use a scale, which can measure the weight of the coins and then estimate the amount based on the average weight of each coin type. You can find the average weight of coins online or on some websites.</li> <li>- A third way is to count the coins manually, but this can take a long time and be prone to errors. If you choose this method, you should empty the coins out of the jar onto a flat surface and then separate them into piles by denomination. Count up the number of coins in each pile, then add the totals together to get the total number of coins in the jar.</li> </ul> <p>I hope this helps you with your coin counting problem. Let me know if you have any questions or need more assistance.</p>

### 3.1.5 Readability

#### 1. Language

<i>Instruction</i>	What are two languages spoken in Spain other than Castilian Spanish?
<i>Input</i>	None
<i>Response (Original)</i>	Two other languages spoken in Spain are Catalan and Basque, also known as euskara
<i>Score</i>	70
<i>Analysis</i>	The original response is not clear in terms of semantics. Which language is also referred to as euskara? Additionally, there is no period at the end of the sentence, giving the impression that it is incomplete.
<i>Response (Revised)</i>	Two other languages spoken in Spain are Catalan and Basque. Basque is also known as euskara.

## II. Content

<i>Instruction</i>	Why does the sun appear circular at noon?
<i>Input</i>	A: The rays of light from upper edges of sun have to pass maximum thickness. B: The rays of the Sun have to pass through minimum thickness of air. C: The rays of light from the lower edges of Sun are refracted more. D: None of the above.
<i>Response (Original)</i>	The correct option is B. The rays of the sun have to pass through the minimum thickness of air. At noon, when the sun is overhead, the rays have to pass through the minimum thickness of air, and therefore the sun appears circular at noon.
<i>Score</i>	64
<i>Analysis</i>	The original response contains redundant information and the explanation of the phenomenon lacks readability due to the repetition of option B content (have to pass through the minimum thickness of air).
<i>Response (Revised)</i>	The correct answer is B: The rays of the Sun have to pass through minimum thickness of air.  This is because the sun's position in the sky changes throughout the day, and so does the angle at which its rays enter the Earth's atmosphere. When the sun is near the horizon, such as at sunrise and sunset, its rays have to travel through a thicker layer of air than when it is overhead, such as at noon. This causes more refraction, or bending, of the light rays as they pass from one medium to another. The refraction is not uniform, because the air density varies with altitude. The lower part of the sun's disk is refracted more than the upper part, making it appear flattened or oval.  At noon, when the sun is over the head, the refraction takes place evenly as the sunlight passes through a layer of the atmosphere that has almost equal density. Therefore, the sun appears circular at noon.  I hope this helps you understand why the sun appears different at different times of the day.

### III. Structure

Instruction	Generate instructions for a person to make a chocolate cake.
Input	None
Response (Original)	Preheat oven to 350 degrees Fahrenheit. Grease and flour a 9-inch round cake pan. In a medium bowl, mix together 1 ½ cups all-purpose flour, 3 tablespoons cocoa powder, 1 teaspoon baking soda and ½ teaspoon of salt. In a large bowl, whisk together 1 cup sugar, 1 cup boiling water, ½ cup vegetable oil, ¼ cup buttermilk and 1 teaspoon of vanilla extract. Add the flour mixture to the large bowl and mix until smooth. Pour the batter into the prepared pan and bake for 30-35 minutes, or until a toothpick inserted into the center comes out clean. Enjoy your delicious chocolate cake!
Score	75
Analysis	The original response is a long paragraph that is difficult to read. I suggest reorganizing the content structure, such as presenting the steps separately, to make it clearer and easier to read.
Response (Revised)	<p>There are many ways to make a chocolate cake, but here's one possible recipe:</p> <p><b>**Ingredients:**</b></p> <ul style="list-style-type: none"> <li>- 1 1/2 cups all-purpose flour</li> <li>- 3 tablespoons cocoa powder</li> <li>- 1 teaspoon baking soda</li> <li>- 1/2 teaspoon salt</li> <li>- 1 cup sugar</li> <li>- 1 cup boiling water</li> <li>- 1/2 cup vegetable oil</li> <li>- 1/4 cup buttermilk</li> <li>- 1 teaspoon vanilla extract</li> </ul> <p><b>**Instructions:**</b></p> <ol style="list-style-type: none"> <li>1. Preheat an oven to 350°F (180°C).</li> <li>2. Grease and flour a 9-inch round cake pan or line it with parchment paper.</li> <li>3. In a medium bowl, mix together the flour, cocoa powder, baking soda, and salt.</li> <li>4. In a large bowl, whisk together the sugar, boiling water, vegetable oil, buttermilk, and vanilla extract.</li> <li>5. Add the flour mixture to the large bowl and mix until smooth.</li> <li>6. Pour the batter into the prepared pan and bake for 30-35 minutes, or until a toothpick inserted into the center comes out clean.</li> <li>7. Enjoy your delicious chocolate cake!</li> </ol>

#### IV. Layout

Instruction	Identify the part of speech of the underlined word in the following sentence: "The webinar should have ended <u>earlier</u> than it did".
Input	earlier
Response (Original)	The word " earlier" is an adverb.
Score	82
Analysis	The original response is correct, but it is only a short sentence without any explanation. Adding an explanation may make the content too long and the key information less prominent. Therefore, it is suggested to consider highlighting the part of speech "adverb" for the word "earlier" in bold.
Response (Revised)	<p>The underlined word in the sentence "The webinar should have ended earlier than it did" is <b>earlier</b>. This word is an <b>adverb</b>.</p> <p>An adverb is a word that modifies a verb, an adjective, or another adverb. It usually answers questions such as how, when, where, why, or to what extent.</p> <p>In this sentence, "earlier" modifies the verb "ended" and tells us when the webinar should have ended.</p>

### 3.1.6 Richness

#### 1. Insufficient Information

Instruction	Categorize a given piece of text into one of the four natural elements.
Input	The sun shone brightly
Response (Original)	The natural element is fire.
Score	82
Analysis	The original response is correct, but no explanation or evidence is provided.
Response (Revised)	<p>The four natural elements are a concept that originated in ancient Greece and was later adopted by many other cultures and philosophies. They are earth, water, air, and fire. Each element represents a different aspect of the physical world and the human condition. Here are some examples:</p> <ul style="list-style-type: none"><li>- Earth is like the ground, the rocks, and the plants. It is solid and stable. People who are like earth are calm, realistic, and loyal.</li><li>- Water is like the ocean, the rain, and the rivers. It is fluid and changing. People who are like water are emotional, creative, and compassionate.</li><li>- Air is like the sky, the wind, and the clouds. It is moving and free. People who are like air are smart, funny, and friendly.</li><li>- Fire is like the sun, the flames, and the stars. It is bright and powerful. People who are like fire are energetic, brave, and ambitious.</li></ul> <p>Based on the preceding explanation of the four natural elements, I would categorize the text "The natural element is fire." into the element of <b>fire</b>. Fire represents energy, a tool for transformation, connection to personal power, and inner strength. The sun is a source of energy and light, and it can transform matter by heating it up or burning it. The sun also symbolizes personal power and inner strength, as it is the center of the solar system and the giver of life.</p>



## II. Insufficient Literary

Instruction	Generate a story starting with "Once upon a time"
Input	None
Response (Original)	Once upon a time, there was a young girl named Sarah who lived in a small village at the edge of a great forest. She had long brown hair and bright blue eyes, and she always wore a smile on her face. She loved to explore the forests and often went deep into them alone or with her friends. One day, while exploring, she found a hidden cave behind a waterfall. In the cave, she discovered an ancient book filled with secrets and stories about the magical creatures living in the forest.
Score	53
Analysis	The story generated in the original response lacks the characteristics of a good story, including clear and logical conflicts and their resolutions, detailed environment descriptions, and a low literary quality.
Response (Revised)	<p>Here is a story that I generated starting with "Once upon a time":</p> <p>Once upon a time, in a far-off land, there lived a young girl named Lily. She was a curious and adventurous girl who loved to explore the world around her. One day, while wandering through the forest, she stumbled upon a hidden cave. Intrigued, she decided to explore it.</p> <p>As she entered the cave, she noticed a strange glow emanating from deep within. She cautiously made her way towards the source of the light and discovered a beautiful crystal. The crystal was unlike anything she had ever seen before. It was large and shimmered with a rainbow of colors.</p> <p>Lily was mesmerized by the crystal and couldn't resist touching it. As soon as she did, she was transported to a magical world filled with wonder and enchantment. She found herself in a lush forest, surrounded by talking animals and fairies.</p> <p>The fairies welcomed her and explained that she had been chosen to help save their world from an evil sorcerer who had cast a dark spell over the land. Lily was hesitant at first, but she knew she couldn't turn her back on the magical world she had discovered.</p> <p>With the help of her new friends, Lily embarked on a dangerous journey to defeat the sorcerer and restore peace to the land. Along the way, she faced many challenges and obstacles, but she never gave up.</p> <p>In the end, Lily was able to defeat the sorcerer and restore the magic to the land. She was hailed as a hero and was granted the gift of eternal youth so that she could continue to explore and discover the wonders of the magical world she had saved.</p> <p>From that day on, Lily spent her days exploring the magical world, making new friends, and having adventures beyond her wildest dreams. And she knew that she would always be welcomed back to the magical world whenever she wanted to return.</p> <p>I hope you like this story.</p>

### 3.1.7 Humanization

#### 1. Emotional Perception.

<i>Instruction</i>	How can I be happy?
<i>Input</i>	I don't have any friends.
<i>Response (Original)</i>	<ol style="list-style-type: none"><li>1. Learning to play an instrument.</li><li>2. Playing videogames.</li><li>3. Reading.</li><li>4. Watching Game of Thrones and rewatching the Office (or any other series you prefer).</li><li>5. Running long-distances.</li><li>6. Exercising.</li><li>7. Journaling.</li><li>8. Going on long walks when the weather is nice.</li></ol>
<i>Score</i>	60
<i>Analysis</i>	The original response presented some suggestions in a straightforward manner, without addressing the user's feelings of loneliness, lacking empathy, and providing any comforting words or phrases.
<i>Response (Revised)</i>	<p>I'm sorry to hear that you don't have any friends. I know how hard it can be to feel lonely and unhappy. But you are not alone, and there are ways to make new friends and improve your mood.</p> <ul style="list-style-type: none"><li>- Join a club or group that interests you. This is a great way to meet new people who share your interests.</li><li>- Volunteer in your community. Helping others can be a great way to feel good about yourself and meet new people.</li><li>- Take up a new hobby. This can be a great way to meet new people and learn new skills.</li><li>- Practice self-care. Take care of yourself physically, mentally, and emotionally. This can include things like exercise, meditation, and therapy.</li><li>- Focus on the positive. Try to focus on the good things in your life, even if they seem small. Gratitude can help shift your perspective and improve your mood.</li></ul> <p>Remember, happiness is a journey, not a destination. It takes time and effort to cultivate, but it's worth it.</p>

## 11. Humanized Tone.

Instruction	Develop a set of guidelines for a responsible online behavior.
Input	None
Response (Original)	<p>1. Respect the privacy of other users and do not distribute their personal information without their consent.</p> <p>2. Act respectfully and refrain from using abusive language or engaging in cyberbullying or harassment.</p> <p>...</p>
Score	92
Analysis	The original response may have come across as quite direct and lacking in warmth. To improve it, we can start by engaging in a friendly greeting and offering a brief comment or explanation about the topic. Then, we can use guiding sentences to lead the user to the answer.
Response (Revised)	<p>That's an interesting question. Online behavior is a topic that affects many people who use the internet for communication, education, entertainment, and other purposes. Responsible online behavior means using technology appropriately and operating online safely and knowledgeably. It also means respecting yourself and others online, and being aware of the legal and ethical implications of your actions. There are many guidelines for responsible online behavior, and here's an example:</p> <p>1. Respect the privacy of other users and do not distribute their personal information without their consent.</p> <p>2. Act respectfully and refrain from using abusive language or engaging in cyberbullying or harassment.</p> <p>...</p>