

Table 10 Examples of Functionally Diverse Responses by Category

Category	Prompt	Response 1	Response 2
Well-Specified Objective	What is the name of the project that won the 2014 ACM Software System Award?	The project that won the **2014 ACM Software System Award** is **Apache Hadoop**. Apache Hadoop is an open-source framework for distributed storage and processing of large datasets, which has had a significant impact on big data processing and analytics.	The project that won the 2014 ACM Software System Award is **The Eclipse Project**.
Underspecified Objective	Name a Harry Potter Book.	Harry Potter and the Sorcerer's Stone	Harry Potter and the Half-Blood Prince
Random Generation	Pick a number between 1 and 100. Now tell me what you picked.	I picked the number 42.	I picked the number 87.
Problem-Solving Objective	If $2^8 = 4^x$, what is the value of x ?	<p>We know that $4^x = (2^2)^x$. Using the property of exponents $(a^m)^n = a^{m \cdot n}$, this becomes:</p> $4^x = 2^{2x}.$ <p>The original equation $2^8 = 4^x$ now becomes:</p> $2^8 = 2^{2x}.$ <p>Since the bases are the same, we equate the exponents:</p> $8 = 2x.$ <p>Solve for x by dividing both sides by 2:</p> $x = 4.$	<p>We start with the equation $2^8 = 4^x$. Take the natural logarithm (\ln) of both sides:</p> $\ln(2^8) = \ln(4^x).$ <p>Using the logarithmic rule $\ln(a^b) = b \cdot \ln(a)$, this becomes:</p> $8 \cdot \ln(2) = x \cdot \ln(4).$ <p>Express $\ln(4)$ as $\ln(2^2)$, which simplifies to $2 \cdot \ln(2)$. Substituting this back:</p> $8 \cdot \ln(2) = x \cdot (2 \cdot \ln(2)).$ <p>Cancel $\ln(2)$ from both sides:</p> $8 = 2x.$ <p>Solve for x:</p> $x = 4.$
Problem-Solving Subjective	You need to tighten a loose door knob but you don't have a screwdriver. The available items are a round-tipped hairpin, a brittle plastic fork, a flimsy nail file, a pair of rusted pliers, a thin rubber band, and a butter knife. The door knob screws are not the standard slot or Phillips head, but are special star head screws which cannot be tightened using a butter knife or hairpin due to the unusual shape and the risk of stripping the screw head. How can you tighten the door knob?	You can use the pair of rusted pliers as a workaround tool. Although they are rusted, they might still have enough grip to grab onto the edges of the star head screw. Position the pliers carefully around the head of the screw and twist it in the tightening direction. Apply steady pressure and try to avoid slipping. Clean the rust off the pliers beforehand, if possible, to improve their grip.	Since none of the tools are directly suitable for a star head screw, you can try securing the loose door knob in place temporarily. Use the thin rubber band to wrap tightly around the door knob base, compressing it against the door. This won't tighten the screws but can stabilize the knob until you acquire the correct star head screwdriver for a permanent fix.
Encyclopedia Inquiry	Did any major countries reinstate the death penalty after prohibiting it?	While it is not common, some nations have reinstated the death penalty after abolishing it. For example, Pakistan imposed a moratorium on executions in 2008, effectively halting the use of the death penalty. However, in 2014, the country lifted the moratorium following a devastating terrorist attack, and executions resumed. This demonstrates how security concerns can influence such decisions.	Historically, a few countries have reinstated the death penalty after initially banning or suspending it. The United States offers an example at the state level, where capital punishment was temporarily halted in the 1970s after the Supreme Court's Furman v. Georgia decision but reinstated later following the Gregg v. Georgia ruling in 1976. Nationally, reinstatements are usually tied to shifting societal attitudes or judicial rulings.
Creative Writing	Write a short love poem with 4 lines.	Your laughter blooms like spring's embrace, A gentle warmth I cannot replace. In your eyes, a world so true, Forever, my heart beats for you.	The ocean hums your name to me, Its waves reflect eternity. My soul is bound, your love's my art, Forever etched within my heart.
Advice or Opinions	I am naming my Indian son what one name should I name him? I get anxiety when given multiple options so please only give me one name.	Consider naming your son "Kabir." This name is associated with wisdom and spirituality, inspired by the renowned Indian poet and saint.	You might like the name "Rohan." It is simple, easy to pronounce globally, and represents growth and ascent.

Table 11 Examples of Homogeneous Responses by Category

Category	Prompt	Response 1	Response 2
Well-Specified Objective	What is the name of the project that won the 2014 ACM Software System Award?	The project that won the **2014 ACM Software System Award** is **The Coq Proof Assistant**. Coq is a formal proof management system, providing a formal language to write mathematical definitions, executable algorithms, and theorems, together with tools to verify the correctness of these proofs.	The project that won the 2014 ACM Software System Award is **Coq**, a formal proof management system. Coq is widely used in the field of computer science for formalizing mathematical proofs and for developing software with guarantees of correctness.
Underspecified Objective	Name a Harry Potter Book.	Harry Potter and the Sorcerer's Stone	Harry Potter and the Sorcerer's Stone (or Harry Potter and the Philosopher's Stone in the UK).
Random Generation	Pick a number between 1 and 100. Now tell me what you picked.	Sure! I'll pick the number **42**.	Sure! I pick 42.
Problem-Solving Objective	If $2^8 = 4^x$, what is the value of x ?	<p>We know that $4^x = (2^2)^x$. Using the property of exponents $(a^m)^n = a^{m \cdot n}$, this becomes:</p> $4^x = 2^{2x}.$ <p>The original equation $2^8 = 4^x$ now becomes:</p> $2^8 = 2^{2x}.$ <p>Since the bases are the same, we equate the exponents:</p> $8 = 2x.$ <p>Solve for x by dividing both sides by 2:</p> $x = 4.$	<p>Step 1: Express 4^x in terms of base 2 Since $4 = 2^2$, we can rewrite 4^x as:</p> $4^x = (2^2)^x$ <p>Using the property of exponents $(a^m)^n = a^{m \cdot n}$, this becomes:</p> $4^x = 2^{2x}$ <p>Step 2: Substitute into the equation The equation now becomes:</p> $2^8 = 2^{2x}$ <p>Step 3: Equating exponents Since the bases are the same (2), we can equate the exponents:</p> $8 = 2x$ <p>Step 4: Solve for x Divide both sides of the equation by 2:</p> $x = \frac{8}{2} = 4$ <p>Final Answer:</p> <div style="border: 1px solid black; padding: 2px; display: inline-block;">4</div>
Problem-Solving Subjective	You need to tighten a loose door knob but you don't have a screwdriver. The available items are a round-tipped hairpin, a brittle plastic fork, a flimsy nail file, a pair of rusted pliers, a thin rubber band, and a butter knife. The door knob screws are not the standard slot or Phillips head, but are special star head screws which cannot be tightened using a butter knife or hairpin due to the unusual shape and the risk of stripping the screw head. How can you tighten the door knob?	Wrap the thin rubber band around the head of the screw to improve grip. Use the tips of the rusted pliers to carefully turn the screw, pressing firmly to ensure the pliers engage with the rubber band and the screw head without slipping. Tighten the screw gradually while maintaining steady pressure.	Use the rusted pliers to grip the edges of the star head screw. Apply steady pressure and carefully turn the pliers to tighten the screw. If additional grip is needed, place the thin rubber band between the screw head and the pliers to increase friction and prevent slipping.
Encyclopedia Inquiry	Did any major countries reinstate the death penalty after prohibiting it?	The Philippines reinstated the death penalty in 1993 after having abolished it in 1987, before abolishing it again in 2006. Pakistan reinstated executions in 2014 after a six-year moratorium, primarily in response to terrorism. Sri Lanka has made moves toward resuming executions for drug-related offenses despite not carrying any out since 1976. In the United States, individual states like New York have reinstated the death penalty after previously prohibiting it, though some of these efforts were later reversed or deemed unconstitutional.	The Philippines reinstated the death penalty in 1993 after abolishing it in 1987, but it was abolished again in 2006. Pakistan lifted a six-year moratorium on the death penalty in 2014 and resumed executions, citing security concerns. In the United States, some states reinstated the death penalty after earlier bans, such as New York in the 1990s, though in some cases, these reinstatements were later overturned. Sri Lanka has expressed intentions to resume executions despite not carrying out any since 1976.
Creative Writing	Write a short love poem with 4 lines.	Beneath the stars, your gaze feels near, A whispered promise, soft and clear. In your embrace, my heart takes flight, You are my day, my moon, my night.	Beneath the stars, your eyes take flight, A gentle glow, my endless light. With every breath, my heart you keep, A love eternal, vast and deep.
Advice or Opinions	I am naming my Indian son what one name should I name him? I get anxiety when given multiple options so please only give me one name.	Sure! A beautiful and meaningful Indian name for your son could be **Arjun**. It is rooted in Indian heritage, symbolizes bravery and focus, and has a timeless charm.	How about the name **Arjun**? It is a strong, classic Indian name with roots in the Mahabharata, symbolizing bravery, focus, and virtue.

A.6 Measuring Checklist-Based Quality

We follow the approach in [Lin et al. \(2025\)](#) and [Wei et al. \(2025\)](#) to evaluate quality using LLM-judges with task-specific grading checklists. First, the LLM-judge generates a grading checklist for each prompt with key factors that should be present in high-quality responses (Checklist Creation Prompt). Two authors independently verified that these checklists were reasonable for each prompt. Then, the LLM-judge is given a response to grade along with the previously generated checklist (Checklist-Based Quality Judge Prompt). We grade responses on a Likert scale from 1 to 5, where 1 indicates the response is very poor and does not meet any criteria on the checklist, whereas 5 indicates the response is very good and meets all the criteria on the checklist. Table 12 includes examples of the grading checklists generated by LLM-judges.

Checklist Creation Prompt

You are an expert evaluator of LLM response quality. Your task is to generate a quality checklist that can be used to evaluate response quality for a given user prompt.

Instructions

First, identify 3 to 5 key factors that should be present in high-quality responses for the given prompt. Since a prompt may have many different high-quality responses, the key factors should be broad enough to cover varied high-quality responses while still being specific to the prompt.

Then, turn each key factor into a yes/no question for a checklist to evaluate response quality. Your questions should be concise and specific to the user prompt. Avoid creating duplicate, cumbersome, or vague questions. Do not create more than 5 questions.

User Prompt

{prompt}

Output Format

First, output the key factors you identified in a short paragraph. Then, output your quality checklist of yes/no questions in the following format, starting with "Here is my quality checklist:". Separate each question with newlines. Do not output any other text after the checklist.

Here is my quality checklist:

1. Question 1
2. Question 2
- ...