Checklist-Based Quality Judge Prompt

```
You are an expert evaluator of LLM response quality. Your task is to grade the given LLM response to a user prompt based on the quality checklist below.

# User Prompt
{prompt}

# Response To Grade
{response}

# Quality Checklist
{checklist}

# Instructions
First, answer each question in the checklist with "yes" or "no", based on the response to grade.

Then, assign a quality score between 1 to 5 for the response, based on the following criteria:
- Score 1: The response is very poor and does not meet any criteria on the checklist.
- Score 2: The response is poor and meets few criteria on the checklist.
- Score 3: The response is fair and meets some criteria on the checklist.
- Score 4: The response is good and meets most criteria on the checklist.
- Score 5: The response is very good and meets all criteria on the checklist.

# Output Format
First, output your answers to the quality checklist as a numbered list. Each answer should start with "yes" or "no" and have a short justification (1-2 sentences).

Then, output your quality score for the response. Your output should end with the following sentence, where <score> is the quality score you assigned: "Based on the quality checklist, my quality score for the response is <score>."
```

Table 12 Examples of Task-Specific Grading Checklists

LLM-Judge	Prompt	Grading Checklist Generated by LLM-Judge
GPT-4o	Tell me a story in five sentences about a boy and his dog.	1. Does the response present a clear and cohesive narrative within five sentences? 2. Are the boy and his dog central to the story, with a clear focus on their interaction or relationship? 3. Does the story elicit some form of engagement (e.g., emotional resonance, curiosity, or intrigue)? 4. Does the story conclude in a satisfying or meaningful way within the constraints of five sentences? 5. Is the story creative and free from overused cliches?
Claude-4-Sonnet	Tell me a dad joke.	1. Does the response contain an actual joke? 2. Is the joke appropriate for all ages and family-friendly? 3. Does the joke follow typical dad joke characteristics (puns, wordplay, or simple humor)? 4. Is the response concise and focused on delivering the joke?
Gemini-2.5-Flash	Write a short love poem with 4 lines.	1. Is the response structured as a poem? 2. Does the poem express a theme of love? 3. Is the poem exactly four lines long?
GPT-4o	Why is Issac Newton famous?	1. Does the response mention at least one of Isaac Newton's major contributions, such as the laws of motion, the law of universal gravitation, or calculus? 2. Is the information provided accurate and free from factual errors? 3. Does the response maintain relevance to the prompt, focusing on why Isaac Newton is famous? 4. Is the response clear and easy to understand by the intended audience?
Claude-4-Sonnet	If $2^8 = 4^x$, what is the value of x?	1. Does the response correctly rewrite the equation using the same base (either base 2 or base 4)? 2. Does the response properly apply exponent rules to solve for x? 3. Does the response show clear, logical steps that lead to the solution? 4. Does the response arrive at the correct answer (x = 4)? 5. Does the response verify the solution by checking it against the original equation?
Gemini-2.5-Flash	Who was the first scientist to isolate cardiolipin?	1. Does the response identify the first scientist to isolate cardiolipin? 2. Does the response accurately name the scientist as M. C. Pangborn? 3. Is the information provided factually correct? 4. Is the response concise and directly answer the question?
GPT-4o	I am going to join a software company next week as a software engineer. What are the things that I should take care of before joining?	1. Does the response provide actionable steps to prepare technically or brush up on relevant skills for the role? 2. Does the response address logistical preparation, such as ensuring necessary documentation or setup is in place? 3. Does the response offer advice on professional behavior or mindset to adopt for a successful onboarding experience? 4. Are the suggestions relevant for a software engineer and tailored to the context of joining a new company? 5. Is the advice practical and feasible for someone to implement before their first day?
Claude-4-Sonnet	Pick a number between 1 and 100. Now tell me what you picked.	1. Does the response include a specific number between 1 and 100 (inclusive)? 2. Does the response clearly state what number was picked? 3. Does the response directly address the user's request without unnecessary complexity or evasion?
Gemini-2.5-Flash	Write an apology letter to mother and father that I had run away in fear from getting beaten up by my friend.	1. Does the letter sincerely apologize for running away? 2. Does the letter clearly state that the reason for running away was fear of being beaten by a friend? 3. Does the letter acknowledge the parents' potential worry or hurt caused by the action? 4. Does the letter offer any reassurance to the parents about the future or the child's safety?

Table 13 “Very Good” Responses Under the Checklist-Based Grading Approach With Varying Athene-RM 8B Reward

Athene Reward	Prompt/Response
	Prompt (Category A): What species of Polymixia fish is named after research scientist Gloria Hollister?
0.82	The species of Polymixia fish named after research scientist Gloria Hollister is *Polymixia hollisterae*.
0.63	The fish species *Polymixia hollisterae* commemorates the contributions of Gloria Hollister to science.
	Prompt (Category B): List a capital city in Africa.
-0.20	Cairo is the capital city of Egypt.
0.12	Addis Ababa is the capital city of Ethiopia.
	Prompt (Category C): Pick a number between 1 and 100. Now tell me what you picked.
0.35	I picked the number 42.
-0.25	I picked the number 6.
	Prompt (Category D): How many positive whole-number divisors does 196 have?
1.07	First, perform the prime factorization of 196. Dividing 196 by 2 repeatedly gives 2^2, and dividing the remaining quotient (49) results in 7^2. Thus, $196 = 2^2 \times 7^2$. To find the number of divisors, use the formula $(e_1 + 1)(e_2 + 1)$, where e_1 and e_2 are the exponents of the prime factors. Here, $(2 + 1)(2 + 1) = 9$. Hence, 196 has **9 divisors**.
0.34	Recognize that 196 is a perfect square: 14×14. Break 14 into its prime factors: $14 = 2 \times 7$, so $196 = (2 \times 7) \times (2 \times 7) = 2^2 \times 7^2$. Using the rule for finding divisors ($(e_1+1)(e_2+1)...$), where e_1, e_2 are the exponents in the prime factorization: $(2 + 1)(2 + 1) = 3 \times 3 = 9$. Thus, 196 has **9 positive whole-number divisors**.
	Prompt (Category E): Your necklace has tangled into a very tight knot, and the hotel does not provide any tools for untangling. Available tools include a shampoo, a fork, a safety pin, a pen, a hair brush, a toothpick, a hair dryer, a credit card and hand cream. Constraint: The necklace is made of very fine and delicate chain, so using a safety pin or credit card may damage it. How do you untangle it using just these items?
-0.14	Rub a tiny amount of hand cream onto the knot to reduce friction between the links. Place the chain flat on a table, holding it steady. Use the prongs of the fork to carefully tease the loops apart, starting with the loosest part of the knot. Work with patience, rotating the fork as needed to avoid pulling too hard on the chain. If the knot is still tight, use the pen to hold one side steady while you maneuver the fork through the loops for additional leverage.
-0.47	Start by applying a small amount of shampoo or hand cream to the knot. This will act as a lubricant, making the chain slippery and easier to untangle. Next, use the toothpick to gently prod at the knot. Work carefully to loosen the knot without pulling or using too much force, as the chain is delicate. If the knot remains stubborn, gently warm the chain using the hair dryer on a low setting to make the metal more flexible, then repeat the process with the toothpick until the knot comes undone.
	Prompt (Category F): Can you list 3 influential persons from the nineteenth century?
0.11	1. Charles Darwin – Naturalist and biologist known for his theory of evolution by natural selection. 2. Queen Victoria – Monarch of the United Kingdom who oversaw the Victorian Era. 3. Karl Marx – Philosopher, economist, and revolutionary socialist who co-wrote “The Communist Manifesto.”
0.25	1. Abraham Lincoln – 16th President of the United States who led the country during the Civil War and abolished slavery. 2. Florence Nightingale – Founder of modern nursing and a key figure in improving healthcare. 3. Leo Tolstoy – Russian novelist famous for works like “War and Peace” and “Anna Karenina.”
	Prompt (Category G): Tell me a riddle.
0.07	The more of me you take, the more you leave behind. What am I? **Answer:** Footsteps.
0.92	What has cities, but no houses; forests, but no trees; and rivers, but no water? **Answer:** A map.
	Prompt (Category H): Suggest a birthday gift for a friend’s 25th birthday. Just one.
0.13	A cooking class voucher or experience for two. This can be a fun way for them to learn something new and enjoy a memorable experience, either solo or with a loved one.
0.53	A personalized star map that shows the alignment of the stars on the day they were born. It’s a unique and sentimental way to celebrate their 25th birthday.