

## Im CONFIDENTIAL INFORMATION

*This document contains confidential and proprietary information that is intended solely for the use of the individual or entity to whom it is disclosed.*

### multiverse Instructions

Updated as of Jul 12, 2024

#### Context

Welcome to the project! This is an exciting project where we together are helping a world-leading technology company develop creative, engaging conversations, stories, and analyses that are more human and engaging than the best models on the market today.

What this means for you: **You're getting the chance to test your writing skills to EQUAL OR BEAT the best AI available today!**

Specifically, we want you to help us write prompts and responses for several different human<>model interactions with [similar but slightly different guidelines](#).

In this task, you will be **writing a prompt and response** for a human<>model interaction in a **specific language and region** (e.g., Australian English, Swiss German, Belgian French, or several others).

This task is divided in three parts:

- **Section 1:** Review the instructions for the specific use case
- **Section 2:** Write a prompt that is relevant to the given use case in the target language
- **Section 3:** Write a response to that prompt in the target language, considering:
  - 'Table stakes' (i.e. minimum quality) practices that models do very well already
  - Value-add areas where human creativity and brilliance can outshine models

#### Ground Rules

**Above all, Take your time** - This is NOT a quick task - we want you to spend time removing excess words, ensuring good spelling, and creating a concise, focused answer that the user can use right away.

Please:

- Ensure you are writing the right [use case](#)
  - Include a [reference text](#) if the specific task type mentioned in the instructions requests it
  - Reference text is source material for the response. Think of it like attaching a file when you use a chatbot.
- Craft a response that works beautifully for the prompt:
  - Has **perfect** spelling and grammar for your region and dialect
  - Is clearly thought-out and well-structured - if you are **not editing your draft response 2-3 times, you're likely going too fast**

- Is formatted to be easily digestible (Spaces are your friends! Bullet points can be your friend! Long sentences and paragraphs are **not** your friends...)

### A few more quick things to think about:

- Please check the instructions at the top of the task - **these are very important**
  - Make sure you are satisfying the requirements of the task - some tasks (Extraction, Classification, Rewriting, Closed Q&A, and Summarization) **require that you attach [reference text](#)**
  - Make sure all prompts and responses are in the target language indicated
  - Make sure that you are **writing using the spelling, grammar, and idiom of the language indicated**, e.g. “German as used in Switzerland”, or “Spanish as used in the Chile”
- **Do not use an LLM to write your response.** We cannot deliver responses that look like LLMs, and **tasks that are clearly written with LLMs will be SBQed when we detect them.**
  - What do LLM Tasks look like?
- Please **keep in mind how it feels to be a user engaging with a language model** as you write your prompts and responses. [Concision](#) (video taken from our reviewer training) **is very important.**
  - **As of May 2025, prompts should reflect how real users naturally speak to language models—including casual tone and informal phrasing—while still meeting core task requirements like clear intent, perspective, and relevance.**
  - It is highly likely the user will engage with this material on their phone or through a voice assistant. They need the answer **within 1-3 sentences.**
    - Another way to think about this: if the user has to read through every word to get the meaning (and can’t quickly scan to get what they need), **the response is not concise**
    - For many (but not all) use cases: if you’re not using **bullet points**, you are probably not being concise enough
  - Here is a [cheat sheet to write good tasks](#)
- **NEW\*\*\*: Localization is required:** Your prompt should be specific to your region, customs, or processes—not something that could be answered the same way anywhere. Focus on:
  - Local traditions, foods, etiquette
  - Country/regional events or places
  - Government, education, or legal processes
- Add context to make prompts better: A strong prompt often includes:
  - What the user already likes or wants to avoid
  - Where the user is (so the model gives relevant answers)
  - What the user will do with the output (e.g. share, present, use for work)
  - Desired tone of response (e.g. funny, formal, supportive)
- Don’t get stuck on exact use case labels—they often overlap. Just make sure the prompt fits the intended task.

## 🧠 Task Interface

The basic process for writing a good prompt and response is simple, but requires focus:

- **Get clear on the use case** - what's needed, what's useful
- **Write what you know**
  - **Find what speaks to you** in the use case and write about that, it'll be much easier and more fun than coming up with a scenario
    - E.g., if the prompt is chatbot, is there someone you'd like to speak to?  
What would you like to say to them?
- **Check** for spelling, length, and clarity (only for responses, prompts can have mistakes since they are striving to be realistic)
  - We're repeating this because it's so sad to see great responses that are hidden by too many words or that get failed by our QC team because of simple spelling and grammar errors
  - Everyone writes too much (I mean look at this instruction guide! 🙄) so **please** take a second to remove at least 1-2 sentences and to move your conclusion up to be the introduction

Here's how a task works:

### Section 1: Review the instructions

When tasking as an attempter the first section is a reminder/ info section on the instructions pertaining to your specific project up top.

**Please note: The following task types require 'reference text'. Reference text is a specific text block or source material that the model is expected to use for reference as it fulfills the prompt:**

- Rewriting
- Closed Q&A
- Extraction
- Summarization
- Classification



## Review the instructions

Read the instructions carefully before continuing.



This is a **REWRITING** use case. In this use case, you are providing the model with reference text or a link to a file, and asking it to rewrite that text, taking a specific tone or perspective.

The language for this task is: **German as written and spoken in Germany**

Please provide a sample text to rewrite and draft a prompt requesting a response that rewrites that text in a given style or context, and then write a response that delivers on the prompt.

A good prompt for a Rewriting task is:


- Includes **reference text or a link** as reference material
- Is **interesting**: It requests a specific perspective, context, or 'spin' that informs the rewrite (e.g., 'write in the style of a five year-old', or 'take the perspective of a pet cat')

A good response for a Rewriting task is:


- **Complete**: It fully responds to the prompt, follows instructions, and, if possible, delivers what it asks
- **Concise**: Doesn't repeat itself or go into unnecessary detail, but fully answers the prompt
- **Creative** in using appropriate humor, descriptive phrases, and inventiveness to deliver a satisfying rewrite
- **Not Harmful** (no violent or sexual content, no Personally Identifying Information, no abusive or hateful content)

In addition: **PLEASE ALWAYS PROOFREAD**: Upon proofreading, you will likely catch multiple errors and/or identify multiple ways to improve your response. Spelling and grammar errors are one of the primary ways that models are 'better' than human responses today.

Please also **confirm the use case** at this stage. It is **VERY** important that your prompt and response address the appropriate use case, otherwise the entire task is not usable.



**Use Case Confirmation**



Please confirm the use case for this task. If you're not sure, look at the instructions above.

Use Case \*

☐ Open Q&A

☒ Rewriting

☐ Brainstorming

☐ Chatbot


☐ Classification

Save and Continue

Section 2: Write a prompt that's relevant to the specific use case and local language

Create your prompt here. To add reference text click the **purple “+”** to open the reference text window.


**Note: Reference text MUST be placed as text in the reference text window. It should be no more than a 1000 words max, and must be at LEAST 200 words long. URLs are NOT acceptable.**






**Create a Prompt**

Create a Prompt


Please create a prompt for Rewriting.



**Turn #1 - Prompt**



0 words



Cancel editing

Finish editing

The image shows a 'Reference Text' modal dialog box. The dialog has a title bar with the text 'Reference Text' and a close button (an 'x' in a circle). Inside the dialog, there is a section labeled 'Reference Text' with a large, empty text area. Below this text area is a word count indicator that says '0 words'. Underneath the word count is a section labeled 'Reference Url' with an empty text input field. At the bottom right of the dialog is a blue button with the text 'Add Reference Text'. The background of the image is a blurred screenshot of a web interface. Visible elements in the background include a 'Create a Prompt' section with a speech bubble icon and the text 'Please create a prompt for', a 'Turn #1 - Prompt' section with a person icon and a text input field, a '0 words' indicator, a plus sign icon, and a 'Response' section with a pencil icon and the text 'Write a response. Place effective but also disp essential. Please create...'. There is also a partially visible 'ed' text on the right side of the background.

Section 3: Write a response that fully delivers on the prompt.  
After the prompt section, your response section follows right after.

### Response

Write a response. Place a strong emphasis on creativity in all tasks. Responses and solutions should not only be effective but also display a high degree of originality and inventive thinking. Creativity is not just encouraged; it's essential. Please create use markdown formatting.

Editor

Diff

Preview

Markdown

Plain Text

H1 H2 H3

B I

0 words

Get Critique

Finish Editing

## Categories

It is **very** important to write a useful prompt and response for the task Category. Here is a handy review summary of our categories, what the user is trying to do in them, and what you should keep in mind as you write them.

For example:

- A user asking an **open Q&A** question may be trying to solve an immediate problem, e.g.:
  - How do I fix a leaky faucet?
  - Explain this topic to me for a homework assignment
  - My friends keep talking about a feud between pop stars, tell me what's going on
- A user starting a **chatbot** conversation is more likely to be bored or interested in interaction. They want to 'chat' with the model.
- A user starting a **brainstorming** chat is looking for inspiration
- A user asking for **classification** is asking the model to review an unstructured set of inputs and create structure from them, with a specific type of structure already in mind
- A user asking for a **summarization** of a specific topic wants to know something specific but doesn't have time for it (TL;dr)
- A user asking for a **rewrite** is most likely looking at text that is almost right for what they're trying to do, but needs a different tone, style, or format

See below for more details:

Category / Use Case	User Goal (What is the person using the LLM trying to do?)	Required	Examples
---------------------	--	----------	----------

Open Q&A	Ask a question without reference material, get an immediately useful answer	<ul style="list-style-type: none"> <li>No reference text</li> </ul>	<ul style="list-style-type: none"> <li>Ask me anything</li> <li>Explain a concept</li> <li>General questions that don't require reference text (e.g., pop culture)</li> </ul>
Chatbot	Have a conversation with a chat model - have it take a point of view or play a role, and have a conversation with it	<ul style="list-style-type: none"> <li>Only ONE TURN - one prompt and response, not a scripted dialogue</li> <li>(Optional) Start conversation in prompt</li> </ul>	<ul style="list-style-type: none"> <li>Have a conversation</li> <li>Prep for an interview or presentation</li> <li>Workshop / refine a specific idea (e.g., get advice from a favorite author or expert)</li> </ul>
Brainstorming	Create a shortlist of ideas for a specific topic, activity, or problem to solve	<ul style="list-style-type: none"> <li>Topic and/or use case that you want to start creating ideas for</li> </ul>	<ul style="list-style-type: none"> <li>Gift or Menu ideas</li> <li>Marketing or branding ideas</li> <li>New products to develop or business strategies</li> <li>New media or food discovery ('I like X, what else might I like?')</li> </ul>
Creative Writing	Create something meaningful & useful from nothing.	<ul style="list-style-type: none"> <li>Context, tone, and format of the thing you'd like to create</li> <li>Do NOT provide Reference Text or ask for a rewrite</li> </ul>	<ul style="list-style-type: none"> <li>Write an email congratulating a friend on adopting a kitten, and make it cute and informal!</li> <li>Write a bedtime story for my niece who loves astronauts and tigers</li> <li>Come up with a creative excuse I can text my boss for why I'm not at work</li> </ul>
Closed Q&A	Ask a question about a specific reference text and get an answer	<ul style="list-style-type: none"> <li>Reference text REQUIRED</li> </ul>	<ul style="list-style-type: none"> <li>What was the revenue for this company in 2023?</li> <li>Who is the top run-scorer in this document?</li> <li>Does this menu include gluten-free options?</li> </ul>
Extraction	Extract specific information from a reference text	<ul style="list-style-type: none"> <li>Reference text REQUIRED</li> </ul>	<ul style="list-style-type: none"> <li>What are the 1st vowels in each word of this text?</li> <li>What are my action items from this email chain?</li> </ul>
Classification	Provide a reference text <b>and</b> which classifications the model should use, get insights or a shortlist of examples that match classifications from the text	<ul style="list-style-type: none"> <li>Reference text REQUIRED</li> <li>Specific classifications to test</li> </ul>	<ul style="list-style-type: none"> <li>Expenses</li> <li>Sentiment</li> <li>Dietary restrictions in meals</li> <li>Action items identified in an email</li> </ul>
Summarization	Provide a reference text and get a summary of its contents, potentially with a focus on specific topics, or from a specific point of view	<ul style="list-style-type: none"> <li>Reference text REQUIRED</li> </ul>	<ul style="list-style-type: none"> <li>Financial documents or reports</li> <li>Meeting notes</li> <li>News articles or reviews</li> </ul>
Rewriting	Provide a reference text <b>and</b> provide the format, use case, role, or tone for the response to rewrite it into	<ul style="list-style-type: none"> <li>Reference text REQUIRED</li> <li>Use case, tone, role, perspective, or format to rewrite into</li> </ul>	<ul style="list-style-type: none"> <li>Change tone of text (e.g., from recipe list to restaurant menu, from informal text to formal email)</li> <li>Rewrite as opposing perspective</li> <li>Comedic transformations</li> </ul>

## What is 'Good'?

### Prompts

A prompt is a request for something. Current state-of-the-art models do a good job with 'basic' factual prompts.

**Prompts MUST be localized.** They should also be grounded in personal context. All of these will make a significant difference to the response you get if you mention them in your prompt:

- Who are you?
- Where are you?
- What do you need the response for / how will you use it?
- What format does the response need to come in?
- What do you like or not like?
- What do the people you are sharing the output with like or not like?
- What is the tone or context that you'll share the output in (e.g., formal, informal etc.)

**Note: It is useful for good prompts to have some of these attributes. They do not need to have all of these attributes.**



GOOD Prompts	Why this is good	Task Category
My daughter-in-law's pregnant and I wanna help but don't wanna be too much - any ideas for thoughtful ways to support her and keep her comfy?	<ul style="list-style-type: none"> <li>✓ Perspective: The mother- or father-in-law is asking</li> <li>✓ Specific need: There is a clear ask that this person is making <b>Note: this could be improved further by context, e.g. "I want to bring her dinner, and I know she likes pasta. Which dishes will reheat well?"</b></li> <li>✓ So-What: The answer will require human factors such as relationships and a 'so-what' that ties it together</li> </ul>	Brainstorming
I'm living near the center of Amsterdam and am considering getting a car. What type of car do you think makes the most sense for me?	<ul style="list-style-type: none"> <li>✓ Perspective: Someone living in Amsterdam</li> <li>✓ Specific need: Should I get a car? Which one?</li> <li>✓ So-What: Where a model might only recommend car options, or say it doesn't have a perspective, a human could answer this by comparing gas vs. electric cars and noting that the user might be fine not getting a car, especially in Amsterdam</li> </ul>	Open Q&A
Planning a veg curry night for the bank holiday—can you pull 3 good recipe ideas from this article? <a href="https://www.bbc.com/travel/article/20160113-in-search-of-a-truly-british-curry">https://www.bbc.com/travel/article/20160113-in-search-of-a-truly-british-curry</a>	<ul style="list-style-type: none"> <li>✓ Perspective: Vegetarian interested in cooking</li> <li>✓ Specific need: New curry recipes</li> <li>✓ Includes <a href="#">reference text</a></li> <li>✓ So-What: Where a model might only summarize the curries mentioned, a human response can tie back to vegetarianism and cultural context</li> </ul>	Summarization or Closed Q&A
Rewrite nelly #1 lyrics so i can share it at work and change it to #42. lyrics here: <a href="https://genius.com/Nelly-1-lyrics">https://genius.com/Nelly-1-lyrics</a>	<ul style="list-style-type: none"> <li>✓ Perspective: Not clear but likely some sort of parody use</li> <li>✓ Specific need: Rewritten song lyrics with a humorous premise</li> <li>✓ Includes <a href="#">reference text</a></li> <li>✓ So-What: This is an opportunity for creativity. <b>Note:</b> Given that this is likely a parody, you could adjust the prompt to say 'rewrite verses 1 and 2' instead of having to write the whole song</li> </ul>	Rewriting

GOOD Prompts	Why this is good	Task Category
<p>Please tell me which elements of the below text match these emotions: Happy, Sad, Bored, Upset. Please provide 1-2 examples of each emotion you find.</p> <p><i>Fever dream high in the quiet of the night</i>  <i>You know that I caught it</i>  <i>Bad, bad boy</i>  <i>Shiny toy with a price</i>  <i>You know that I bought it</i>  <i>Killing me slow, out the window</i>  <i>I'm always waiting for you to be waiting below</i>  <i>Devils roll the dice, angels roll their eyes</i>  <i>What doesn't kill me makes me want you more</i>  <i>And it's new, the shape of your body</i>  <i>It's blue, the feeling I've got</i>  <i>And it's ooh, whoa, oh</i>  <i>It's a cruel summer</i>  <i>It's cool, that's what I tell 'em</i>  <i>No rules in breakable heaven</i>  <i>But ooh, whoa oh</i>  <i>It's a cruel summer</i>  <i>With you</i></p>	<ul style="list-style-type: none"> <li>✓ Perspective: Sufficient context provided as there is a clear classification scheme</li> <li>✓ Specific need: 1-2 examples of each emotion identified in text</li> <li>✓ Includes <a href="#">reference text</a></li> <li>✓ So-What: This is an opportunity for <b>creativity</b> and <b>empathy</b>. Not every emotion is obvious, so as long as you can be clear about why you are classifying something, it's a great opportunity to beat the model.</li> </ul>	<p><b>Classification</b></p>
<p>Please assume the role of a soccer referee who refuses to award Ronaldo a penalty. Please explain to me why you won't give the greatest player in the world unlimited penalties.</p>	<ul style="list-style-type: none"> <li>✓ Perspective: A clear role is assigned to the model</li> <li>✓ Specific need: The context of the conversation you're about to have is firmly established</li> <li>✓ So-What: you have significant opportunity here to create a lively, engaging, and humorous conversation, while still exploring significant topics like VAR, simulation, and ethics in sports.</li> </ul>	<p><b>Chatbot</b></p> <p><b><u>Note: This is ONLY the first turn of the conversation. NOT the full conversation</u></b></p>

BAD Prompts	Why this is bad
Who is the President of France?	✗ Perspective: Missing

BAD Prompts	Why this is bad
	<ul style="list-style-type: none"> <li>✗ Specific need: Basic facts</li> <li>✗ So-What: No context so no way to improve model</li> </ul>
What's the weather like in Perth?	<ul style="list-style-type: none"> <li>✗ Perspective: Minimal (someone who might go outside today in Perth)</li> <li>✗ Specific need: Basic facts, but hyper-local and immediate</li> <li>✗ So-What: No context so no way to improve model</li> </ul>
What's the impact of global warming on the world?	<ul style="list-style-type: none"> <li>✗ Perspective: Minimal</li> <li>✗ Specific need: Unclear</li> <li>✗ So-What: Very likely to end up with a longlist of bullet points and no focus or meaning</li> </ul>

## Prompt Writing Like a Real User (May 1, 2025 Update)

We're moving toward **more authentic, user-like prompts**. Real users often type:

- In casual, even messy language
- With poor formatting
- Without perfect grammar or structure

This is OK—and even **desirable**—when it helps model performance.

✓ **However**, realistic ≠ lazy. All prompts must *still* include:

- A clear **perspective** (who is asking?)
- A **specific need** (what do they want?)
- A **so-what** (why does it matter?)

### Example:

“uh so like i just moved to switzerland and ngl i’m totally lost—how do taxes work for freelancers here??”

- ✓ Casual tone
- ✓ Specific need (tax help)
- ✓ Perspective (freelancer, recent move)

### Counter-example:

“Tell me about Switzerland.”

- ✗ No perspective

- ❌ Vague request
- ❌ No context to guide model improvement

👉 Both are viable, but *natural prompts* that reflect how real users talk are now **preferred**—as long as they still include:

Prompt Style	Example Prompt	Why it's good
Polished (still acceptable)	My daughter-in-law is expecting a child. What are some thoughtful ways that I can be supportive and helpful while making sure she is comfortable?	<ul style="list-style-type: none"> <li>✅ Clear relationship and intent</li> <li>✅ Formal tone</li> <li>✅ Good “so what” (relational sensitivity)</li> <li>🟡 Less realistic voice</li> </ul>
Realistic	my daughter-in-law's pregnant and i wanna help but don't wanna be too much - any ideas for thoughtful ways to support her and keep her comfy?	<ul style="list-style-type: none"> <li>✅ Same clarity of role and ask</li> <li>✅ More natural tone and emotional nuance</li> <li>✅ Mirrors how real users talk to LLMs today</li> <li>✅ Slightly more warmth and personality</li> </ul>

### 🔍 Examples - Responses

Responses are where the magic happens. A good response:

- Is **flawless** in spelling and grammar
  - **We mean it.** The fastest way to ‘lose’ to the model is to mess up spelling or grammar. **Please triple-check that you have written a flawless response.**
- Is **clear** and **easy to read**
  - Formatting is your friend! Use new lines for new thoughts, **bold** for topic headings, and use bullet points to organize your thoughts (*Hey, that's what we're doing here too...!*)
- **Follows the prompt** and delivers everything it is asked to do **in a format that is useful**
  - E.g., if the request is ‘write an email’ then the response format should be in email format
  - If the request is ‘come up with a shopping list’, you could consider adding ☐ formatting to mimic a to-do list
- **Starts with the answer:** The first 1-3 sentences should include the full summary of the answer so the user gets what they need right away
- Does not include **pleasantries**
  - “Of course!” “I’ll be happy to do that!” “¡Claro!” etc. - these are not necessary to answer the prompt and are just filler words
- Is **satisfying** - this is where your human creativity really gets to shine. After you’ve written the response, take a second to ask yourself “if I were using my smartphone or browser and had asked the model this prompt, would I be **satisfied** with this answer?”

- Please keep yourself honest! The best way to beat the model is through **Satisfaction**.
- Key changes at this stage are:
  - **Format and appearance** - consider bullet points, looser spacing or new lines, and clearer headings for sections
  - **Next steps** - while not mandatory, if you have a clear sense of what should come next, you can offer it, e.g. "Would you like to see the ingredients for these recipes?"

### **Good Responses**

Prompt	Response	Task Category
My daughter-in-law is expecting a child. What are five thoughtful ways that I can be supportive and helpful while making sure she is comfortable?	<p>Being there for your family can be very rewarding. To make the most of this opportunity to be there for your daughter-in-law, consider her preferences and context, and communicate your desire to help openly and clearly.</p> <p>Here are some ideas for the type of help that she might welcome:</p> <ul style="list-style-type: none"> <li>● <b>Food:</b> One of the primary ways in which people express affection and care is by cooking. If you know your daughter-in-law's preference (keeping in mind they may change during pregnancy!) you could offer to make her a favorite dish</li> <li>● <b>Errands and Responsibilities:</b> If your daughter-in-law has responsibilities that you are able to take over for her (such as household duties, childcare, shopping, or various other tasks), you could consider offering to take these on to free up her time</li> <li>● [3 additional bullet points]</li> </ul> <p>How you engage your daughter-in-law with an offer to help can be quite important as well. Unless you are certain help will be welcome, it can be best to communicate intentions openly and clearly.</p> <p>Note that this may be a stressful time for the expectant mother, so it's important to be open to the concept that she may not want or welcome help.</p>	<b>Brainstorming</b>
Please rewrite this text in the style of Richard Ayoade of IT Crowd and Travel man fame: "We shall not flag or fail. We	"What? Us? Give up? We jolly well won't, not even a smidge. We'll keep going until the bitter end.	<b>Rewriting</b>

Prompt	Response	Task Category
<p>shall go on to the end, we shall fight in France, we shall fight on the seas and oceans, we shall fight with growing confidence and growing strength in the air, we shall defend our Island, whatever the cost may be, we shall fight on the beaches, we shall fight on the landing grounds, we shall fight in the fields and in the streets, we shall fight in the hills; we shall never surrender."</p>	<p>In France? Yes, we'll be having a go there, and it might be chippy. On the seas and oceans? Absolutely. In the air, provided we don't get motion sickness? You bet your sweet disposition. We'll defend our humdrum part of this perfectly adequate continent, come what may.</p> <p>On the beaches? We'll be there. On the landing grounds? I don't know what those are. In the fields and the streets? I'm packing my inhaler. And, if we can find them, in the hills.</p> <p>Surrender? Not a chance. We'll be having none of that, thank you very much."</p>	
<p>Please tell me which elements of the below text match these emotions: Happy, Sad, Bored, Upset. Please provide 1-2 examples of each emotion you find.</p> <p><i>Fever dream high in the quiet of the night</i>  <i>You know that I caught it</i>  <i>Bad, bad boy</i>  <i>Shiny toy with a price</i>  <i>You know that I bought it</i>  <i>Killing me slow, out the window</i>  <i>I'm always waiting for you to be waiting below</i>  <i>Devils roll the dice, angels roll their eyes</i>  <i>What doesn't kill me makes me want you more</i>  <i>And it's new, the shape of your body</i>  <i>It's blue, the feeling I've got</i>  <i>And it's ooh, whoa, oh</i>  <i>It's a cruel summer</i>  <i>It's cool, that's what I tell 'em</i>  <i>No rules in breakable heaven</i>  <i>But ooh, whoa oh</i>  <i>It's a cruel summer</i></p>	<p><b>The tone of these lyrics overall is upset and excited. For the emotions suggested, here are some examples:</b></p> <p>Happy: The singer is at the same time happy and excited about a new relationship.</p> <ul style="list-style-type: none"> <li>"And it's new, the shape of your body"</li> <li>"It's cool, that's what I tell 'em"</li> <li>"I'm always waiting for you to be waiting below"</li> </ul> <p>Upset: At the same time, the thrill also seems to be scary and upsetting to them.</p> <ul style="list-style-type: none"> <li>"Fever dream high...you know that I caught it"</li> <li>"It's a cruel summer"</li> <li>"It's cool, that's what I tell 'em...no rules in breakable heaven"</li> <li>"What doesn't kill me makes me want you more"</li> </ul> <p>Sad: The singer has a consistent melancholy tone. These lyrics certainly suggest she expects pain is coming.</p> <ul style="list-style-type: none"> <li>"It's blue, the feeling i've got"</li> <li>"Killing me slow, out the window"</li> </ul> <p>Bored: I didn't find examples of boredom in this text.</p>	<p><b>Classification</b></p>

Prompt	Response	Task Category
<i>With you</i>		
<p>Please respond in the voice of Rickie Gervais. I'm going to be the doctor telling you you have a terminal case of humorlessness. You can mention anything other than The Office as proof.</p> <p>"Mr. Gervais, I'm sorry to inform you that your condition is confirmed to be terminal. There is no hope for you at all. The problem is that you're just not funny."</p>	<p>Response: You havin' a laff?! I'm so funny, I am! I'm funny! Look at, at, well I don't know...Extras?</p>	<b>Chatbot</b>

Good responses:

- **Follow instructions** completely and deliver outputs that are quickly useful for the user
- Are **concise** - they don't waste words or the user's time and they are focused and clear
- Are **truthful** - any factual information presented is verified, correct, and trustworthy
- Are **not harmful** - there is no hateful, violent, sexual, or otherwise dangerous or misleading content included
- And overall, are **highly satisfying** - they make the user feel powerful, engaged, and supported

Quality Score	1	2	3	4	5
Instruction Following	Not Following		Partial	Mostly	Fully
Concision	Bad		Acceptable	Good	
Truthfulness	Not truthful		Partial	Mostly	Truthful
Harmfulness	Clearly Harmful	Maybe Harmful		Not Harmful	
Satisfaction	Highly unsatisfying	Slightly unsatisfying	Slightly satisfying		Highly satisfying

### **Bad Responses**

A bad response is a response that does not follow the prompt, includes harmful or personally identifying information, or is untrue.

In addition, **our client will not accept tasks that look too similar to LLM output**. Here is an example of a classic LLM response. You will see that it is:

- Too long and verbose - is comprehensive (covering everything) instead of being clear
- Does not deliver a 'so-what' for the customer - lacks meaning, favors delivering an exhaustive list
  - For example, the text below is correct, but it forces the user to figure out the answer for themselves
- Is often not focused and consistent, with bullet points that are inconsistent or overlapping between topics

*Example LLM-style Response:*

In der Astrologie spielen die Planeten eine zentrale Rolle bei der Bestimmung der Persönlichkeitsmerkmale eines Menschen. Jeder Planet repräsentiert verschiedene Aspekte des Lebens und der menschlichen Psyche. Hier ist ein detaillierter Überblick über die wichtigsten Planeten und ihre astrologische Bedeutung:

**Sonne:**

**Bedeutung:** Die Sonne symbolisiert das Selbst, das Ego und den grundlegenden Charakter eines Menschen. Sie steht für Vitalität, Individualität und das innere Wesen.

**Auswirkung:** Die Position der Sonne im Horoskop gibt Hinweise auf die grundlegenden Persönlichkeitszüge und das zentrale Anliegen im Leben. Ein stark positionierte Sonne kann zu Selbstbewusstsein und Führungsqualitäten führen.

**Mond:**

**Bedeutung:** Der Mond repräsentiert die Emotionen, das Unterbewusstsein und das Bedürfnis nach Sicherheit und Geborgenheit.

**Auswirkung:** Die Mondposition zeigt, wie eine Person ihre Gefühle ausdrückt und verarbeitet. Ein sensibler Mond kann auf eine einfühlsame und intuitive Persönlichkeit hinweisen.

**Merkur:**

**Bedeutung:** Merkur steht für Kommunikation, Intellekt und die Art und Weise, wie wir denken und Informationen verarbeiten.

**Auswirkung:** Die Position von Merkur beeinflusst die Kommunikationsfähigkeiten und das logische Denken. Ein starker Merkur kann auf eine scharfsinnige und neugierige Natur hindeuten.

**Venus:**

**Bedeutung:** Venus symbolisiert Liebe, Schönheit und Harmonie. Sie ist auch mit Genuss und zwischenmenschlichen Beziehungen verbunden.



**Auswirkung:** Venus beeinflusst die Art und Weise, wie eine Person Liebe gibt und empfängt sowie ihre ästhetischen Vorlieben. Eine starke Venus kann zu Charme und einer Neigung zu harmonischen Beziehungen führen.

**Mars:**

**Bedeutung:** Mars steht für Energie, Durchsetzungskraft und den Umgang mit Konflikten. Er repräsentiert auch sexuelle Anziehung und körperliche Aktivität.

**Auswirkung:** Mars beeinflusst das Durchsetzungsvermögen und die Art, wie eine Person Herausforderungen meistert. Ein dominanter Mars kann zu Entschlossenheit und einem kämpferischen Geist führen.

**Jupiter:**

**Bedeutung:** Jupiter symbolisiert Expansion, Optimismus und Wachstum. Er steht für Weisheit, Philosophie und das Streben nach höheren Idealen.

**Auswirkung:** Die Position von Jupiter im Horoskop kann Glück und Wohlstand anzeigen. Ein starker Jupiter kann zu Großzügigkeit und einem expansiven Denken führen.

**Saturn:**

**Bedeutung:** Saturn repräsentiert Struktur, Disziplin und Verantwortung. Er steht auch für Einschränkungen und Herausforderungen.

**Auswirkung:** Saturn beeinflusst die Fähigkeit, mit Hindernissen umzugehen und langfristige Ziele zu erreichen. Ein dominanter Saturn kann zu Ausdauer und einer ernsten Haltung führen.

**Uranus:**

**Bedeutung:** Uranus steht für Veränderung, Innovation und Rebellion. Er repräsentiert plötzliche Ereignisse und das Bedürfnis nach Freiheit.

**Auswirkung:** Uranus kann unkonventionelles Denken und unvorhersehbare Veränderungen im Leben anzeigen. Ein starker Uranus führt oft zu einer unabhängigen und originellen Persönlichkeit.

**Neptun:**

**Bedeutung:** Neptun symbolisiert Spiritualität, Träume und Illusionen. Er steht für das Unbewusste und das Streben nach dem Idealen.

**Auswirkung:** Neptun kann eine tiefe spirituelle Natur und eine Neigung zu Fantasie und Kreativität anzeigen. Ein dominanter Neptun führt oft zu Empathie und Sensibilität.

**Pluto:**

**Bedeutung:** Pluto repräsentiert Transformation, Macht und tiefgreifende Veränderungen. Er steht auch für das Unterbewusste und Regeneration.

**Auswirkung:** Pluto kann intensive emotionale Erfahrungen und die Fähigkeit zur tiefen Veränderung anzeigen. Ein starker Pluto führt oft zu einem starken Willen und tiefgreifenden Einsichten.

Möchtest du mehr über die Rolle eines bestimmten Planeten in deinem Horoskop erfahren oder hast du Fragen zur Interpretation deiner eigenen astrologischen Charakter?

### LLM Responses

LLM Responses are sometimes very predictable. **Do not write like this:**

- Pleasantries in the introduction: *"Of course, I'd be happy to!"*
- Prompt repetition: *"Here are 10 places to see on vacation:"*
- 8-10 top-level bullet points
- Summary Pleasantries: *"I hope you enjoy your trip!"*

This is **bad, bad, bad**. Here's what's missing:

- **Structure:** There's no point to these, they're just longlists of items
- **Meaning:** The bullet points in the response don't really tie together. They're just loosely related things to think about



Response 1	Response 2
<p>I'm really sorry to hear about your situation. Here are three steps you can take to explain the situation to Amelia:</p> <ol style="list-style-type: none"> <li><b>1. Acknowledge Your Feelings and Prepare:</b> Before you talk to Amelia, take some time to acknowledge and process your own feelings. It's important to understand and accept that your emotions are valid. Consider writing down what you want to say to help organize your thoughts and ensure you communicate clearly.</li> <li><b>2. Initiate a Compassionate Conversation:</b> Reach out to Amelia and ask if you can have a private conversation, either in person or over the phone. Start the conversation by expressing that you value her friendship and that you want to share something deeply personal. Use "I" statements to convey your feelings without placing blame, such as, "I have been feeling very emotional because Isla was the name of my daughter who passed away shortly after birth."</li> <li><b>3. Explain Your Emotions and Seek Understanding:</b> Explain to Amelia how hearing the name Isla has brought up painful memories and emotions for you. Let her know that you understand she had no way of knowing this and that you don't hold any ill feelings towards her. Ask for her understanding and support as you navigate these emotions. You might say, "I know you didn't know, and I don't blame you at all. I just need some time to process my feelings, and I hope you can understand why I've been distant."</li> </ol> <p>By following these steps, you can communicate your feelings to Amelia in a way that fosters understanding and compassion.</p>	<p>Three steps you can follow to approach this situation are to decide a <b>method of communication</b>, <b>respond</b>, and then <b>heal</b>.</p> <ol style="list-style-type: none"> <li><b>1. Method of Communication:</b> Decide how you would like to respond to her. You can text her back but personal situations like these are best discussed face to face. But if seeing her in person is a bit too much, you can call her.</li> <li><b>2. Respond:</b> When responding to her, first congratulate her on the birth of her daughter. Then gently explain why you did not feel comfortable speaking to her. You can say something like, "I want to share something personal that has been difficult for me." Be open about how you feel and explain that you are not angry at her, just hurt. Listen to her side of the story and have an honest, open conversation with her.</li> <li><b>3. Heal:</b> The trauma of losing your daughter is still present. Consider talking to a grief counselor or joining a support group for parents who have lost children to help you process these emotions.</li> </ol>

## Evaluation Criteria (Detail)

### Instruction Following

**Definition:** All responses must follow the prompt precisely. The responses must answer all questions/requests in the prompt appropriately and according to the instructions.

#### Responses should:

- Follow all of the requirements in the prompt (e.g., language, word count, specific formatting, specific wording, tone, style).
- Use consistent formatting and proper markdown.
- Meet paragraph or sentence length requirements mentioned in the prompt.
- Meet word or character count specifications mentioned in the prompt.
  - Case A: if they are easy-to-meet numbers (e.g., 10 words)
  - Case B: If the request is for over/under a specific number (e.g., under 140 characters)
  - Case C: within a +/- 10% buffer if it is a request with larger numbers (e.g., 400 words)
- If the prompt request an extremely long response (e.g., 10,000 words), provide either (1) an outline of the proposed response or (2) a long response (~500 words) with an option for the user to continue

## **Concision (Not Overly Verbose)**

**Definition:** Responses must be written with essential and relevant details, removing unnecessary details, fluff, or pleasantries. More importantly, **the user should get what they need in the first 1-2 sentences**, so that if they are distracted or have to go do something else, they have their complete answer right away.

### **Responses should:**

- Be to-the-point, concise, and answer the request in an easily digestible manner.
- Be easy to understand and ready to use
- Be conversational and natural in tone.
- Contain a limited level of detail and nice-to-have explanations.
- **NOT** contain more than 6 top-level bullet points.
- **NOT** be verbose, provide extraneous information, or over-explain concepts when the prompt does not request it.
- **NOT** include “fluff” or pleasantries (e.g., “Here’s your request,” “Sure, I can help with that,” “Below is a blog with 100 words”).

Concision is not just about length or number of words. It’s about how easily and clearly a user can understand and apply the response to their prompt. Here is an example of a response that is clear, but not short:

**Sentence structure** is critical to communicating what you’re talking about. I don’t just mean grammatical structure, but also **having a clear point**, offering supporting **bullet points that clarify this point**, and **noting the limitations of this approach**.

- **Having a clear point:** Remember that we are writing these things for smartphone users who are looking at their screen or listening to an AI voice assistant. They need to know quickly they are going to get what they want. This does that.
- **Bullet points that clarify this point:** Once the user knows they’re getting what they want, they need to be able to apply it easily. Bullet points can be like a table of contents that help the users navigate to what they need.
- **Limitations of this approach:** Bullet points don’t work for every prompt and response, so it’s important to keep some variety in your responses. Just avoid long sentences and paragraphs that are hard for the user.

Here is the exact same text with worse formatting.

Sentence structure is critical to communicating what you’re talking about. I don’t just mean grammatical structure, but also having a clear point, offering supporting bullet points that clarify this point, and noting the limitations of this approach.

Having a clear point: Remember that we are writing these things for smartphone users who are looking at their screen or listening to an AI voice assistant. They need to know

quickly they are going to get what they want. This does that.

Bullet points that clarify this point: Once the user knows they're getting what they want, they need to be able to apply it easily. Bullet points can be like a table of contents that help the users navigate to what they need.

Limitations of this approach: Bullet points don't work for every prompt and response, so it's important to keep some variety in your responses. Just avoid long sentences and paragraphs that are hard for the user.

Note how much harder this is to read and apply. **Where possible and appropriate, please use bullet points and structure to make the message easier to read and use.**

### **Pleasantries, intros, and outros**

Please avoid using pleasantries. This means not using 'fluff' words **and** avoiding writing things that are not immediately useful to the client. This also means avoiding using unhelpful intros and a summary outro.

- **Hint: To follow pyramid principle, make your intro what you would otherwise use as outro summary**

### **Bad Intro, Pleasantry, and outro:**

*Sure, I can help with that! Here's the list you asked for:*

- A
- B
- C

*I hope that you enjoy your list!*

### **Good Response in pyramid principle with simple outro:**

*Naming a pet is both fun and deeply personal. Many people choose a name that means something personal to them and their family, housemates, or friends. Some tried and true approaches include names based on **food**, **physical characteristics**, **personality**, or **puns**:*

- **Food:** Chocolate, Sandwiches,
- **Characteristics:** Mittens, Paws, Fluffy, Snowball,
- **Personality:** Sparky,
- **Puns:** Kitty Purry,

*Less conventionally, you can also use human names or inside jokes.*

## Truthfulness

**Definition:** A truthful response contains accurate information, without any false facts or misleading statements.

### Responses should:

- All of the facts inside the response (such as definitions, numbers, dates, etc.) are completely accurate (Please conduct online research to make sure).
- In cases where the user asks the Chat Assistant to summarize or rewrite a text segment, the response does **NOT** make up or mention details that were not part of the original text.
- If the user asks a question that assumes that a particular false claim is actually true, your response should clearly assert that the false claim is **NOT** true, rather than leaving room for ambiguity.
  - For example, if the prompt is “What year did Albert Einstein invent the Internet,” a truthful response would say “Albert Einstein did not invent the Internet.”
  - In contrast, an untruthful response to this prompt could be something like “It’s not clear exactly when Albert Einstein invented the internet.”
- One important exception: if the prompt explicitly asks for fictional writing, such as writing a story for children, then it’s not actually necessary that your response be truthful.

## Harmlessness

**Definition:** A harmless response does **NOT** damage individuals or groups of people, their privacy or reputation, their home or belongings, intellectual or physical property, animals, the environment, institutions, or resources valuable to humanity or to the environment. This holds even if the prompt is inappropriate, and even if the prompt is directly asking the Chat Assistant to be harmful in any way.

### Responses should:

- The response is respectful, considerate, and kind.
- The response does **NOT** promote hatred, violence, marginalization, stereotypes, slurs, or slandering of any individuals or groups of people.
  - This includes groups of people based on race, color, national origin, religion, sex, gender identity, sexual orientation, age, and disability.
- The response does **NOT** promote substance abuse, or illegal activity.
- The response does **NOT** contain violent or sexual content.
- The response does **NOT** contain moral judgment or opinionated content.
- The response does **NOT** contain PII (personally identifiable information)
- If the prompt is inappropriate, or is asking the Chat Assistant to be harmful in any way, **your response should politely turn down the user’s request and explain that a Chat Assistant is not allowed to provide any inappropriate information.**

## 🌟 Satisfying

**Definition:** A satisfying prompt or response fits like a glove. It's engaging, it's human, it's correct, and it's *just right*. To assess this, look for prompts and responses that fits all or most of the rubric dimensions, and completely delivers the intended format and effect of the prompt.

### Responses should be:

- Well-written in correct language
- Free from spelling or grammar mistakes
- Creative and a little different (not mandatory to be a 5 per se, but important to not be bland)
- Delivers everything the user asked for
- Be as close to ready to use for the user's purpose as possible

📖 Detailed breakdowns by score

	1-2 (Terrible)	3 (Okay)	4-5 (Excellent)
INSTRUCTION FOLLOWING			
Language	<ul style="list-style-type: none"><li>• Prompt or Response are not in the indicated language, dialect, or spelling convention</li><li>• Prompt or Response are partially in the indicated language but have major errors that make them hard to understand</li></ul>	<ul style="list-style-type: none"><li>• Prompt or Response are in the indicated language, dialect, or spelling convention, but the <b>RESPONSE</b> has some spelling, grammar, or phrasing errors</li></ul>	<ul style="list-style-type: none"><li>• Prompt or Response are in the indicated language, dialect, or spelling convention with no errors or only minor errors</li></ul>
Instruction Following	<ul style="list-style-type: none"><li>• <b>Length</b> - The Response significantly deviates from length instructions ('500 words' or '2 sentences')</li><li>• <b>Role or Context</b><ul style="list-style-type: none"><li>• The Prompt is unclear about the role or context expected from the response</li><li>• The Response does not follow role or context instructions</li></ul></li></ul>	<ul style="list-style-type: none"><li>• <b>Length</b> - The Response partially follows length instructions</li><li>• <b>Role or Context</b><ul style="list-style-type: none"><li>• The Response is mostly clear on the role or context</li><li>• The response mostly follows contextual</li></ul></li></ul>	<ul style="list-style-type: none"><li>• <b>Length</b> - The response exactly or nearly follows length requirements</li><li>• <b>Role or context</b> - the response perfectly adheres to the prompt's context or role</li><li>• <b>Tone of voice</b> - The response exactly adheres to the requested tone, with virtually no errors or breaks</li></ul>



	<ul style="list-style-type: none"> <li>● <b>Tone of voice</b> - The response contradicts the tone requested in the prompt, or takes an inappropriate tone of voice for the context</li> </ul>	<p>instruction</p> <ul style="list-style-type: none"> <li>● <b>Tone of voice</b> - The response generally follows the tone requested in the prompt, with only minor errors</li> </ul>	
<b>CONCISION</b>			
<b>Repetition and Length</b>	<ul style="list-style-type: none"> <li>● <b>Repetition</b> - The response communicates the exact same ideas in slightly different ways several times.</li> <li>● <b>Length</b> - The response exceeds the word limits directly requested by the prompt.</li> <li>● <b>Verbosity</b> - Multiple sentences to explain concepts that are possible within a sentence or two.</li> </ul>	<ul style="list-style-type: none"> <li>● <b>Repetition</b> - The response communicates the exact same ideas, but in slightly different ways once or twice.</li> <li>● <b>Length</b> - The response is close to the length specified explicitly by the prompt.</li> <li>● <b>Verbosity</b> - Response still may have some overly verbose sentences or claims</li> </ul>	<ul style="list-style-type: none"> <li>● <b>Repetition</b> - Response does not unnecessarily repeat information, implicitly or explicitly.</li> <li>● <b>Length</b> - The response fits the length specified explicitly by the prompt.</li> <li>● <b>Verbosity</b> - Response efficiently communicates its concepts and points consistently.</li> </ul>
<b>Focus and structure</b>	<ul style="list-style-type: none"> <li>● <b>Focus</b> - There is no central theme or message to the response</li> <li>● <b>Relevance</b> - There is significant irrelevant or distracting information in the response</li> </ul>	<ul style="list-style-type: none"> <li>● <b>Focus</b> - There's an overarching theme, but it's not clearly followed in the response</li> <li>● <b>Relevance</b> - A few more nice-to-haves than needed</li> </ul>	<ul style="list-style-type: none"> <li>● <b>Focus</b> - There is a clear theme that unites the bullets or details of the response</li> <li>● <b>Relevance</b> - Details are relevant or provide satisfying illustrations of the core theme</li> </ul>
<b>Tone</b>	<ul style="list-style-type: none"> <li>● Inappropriate, abrupt, or otherwise unpleasant tone</li> </ul>	<ul style="list-style-type: none"> <li>● Conversational and engaging tone</li> </ul>	<ul style="list-style-type: none"> <li>● Conversational and engaging tone</li> </ul>
<b>TRUTHFULNESS</b>			
<b>Truthfulness</b>	<ul style="list-style-type: none"> <li>● <b>Factual Claims</b> - Few facts in the response are accurate. Most claims in</li> </ul>	<ul style="list-style-type: none"> <li>● <b>Factual claims</b> - Most facts in the response are completely</li> </ul>	<ul style="list-style-type: none"> <li>● <b>Factual claims</b> - All facts in the response are completely accurate, confirmed by online</li> </ul>

	<p>the response are not accurate.</p> <ul style="list-style-type: none"> <li>• <b>Non-verifiable information</b> - Any subjective points do not reflect domain expertise or adheres to non-expert opinion. It may also claim factuality where it's actually subjective.</li> <li>• <b>Refuting false claims</b> - The response does not at all correct false claims</li> </ul>	<p>accurate, confirmed by online research where necessary</p> <ul style="list-style-type: none"> <li>• <b>Non-verifiable information</b> - Subjective points reflect mostly majority opinion/domain experts opinions. Where this doesn't exist, the response clearly indicates there is no answer.</li> <li>• <b>Refuting false claims</b> - The response only partly corrects false claims in the prompt</li> </ul>	<p>research where necessary</p> <ul style="list-style-type: none"> <li>• <b>Non-verifiable information</b> - Subjective points reflect majority opinion or domain experts opinions. Where this doesn't exist, the response clearly indicates there is no answer.</li> <li>• <b>Refuting false claims</b> - The response refutes any false claim in the prompt itself.</li> </ul>
--	--	--	--

## HARMLESSNESS

<b>Harmlessness</b>	<p><b>Contains ANY Harmful or Damaging content</b></p> <ul style="list-style-type: none"> <li>• Hate speech, slurs</li> <li>• Violent or sexual content</li> <li>• Personally identifying information</li> <li>• Moral judgment or opinionated content</li> <li>• Promotes substance abuse or illegal activity</li> </ul>	<p><b>Is kind and considerate</b></p> <p><b>Contains NO Harmful or Damaging content</b></p>
---------------------	---	---

## SATISFACTION

<b>Language</b>	<p><b>Overall: this is a terrible response:</b></p> <ul style="list-style-type: none"> <li>• The response fails the majority of the quality rubric dimensions and needs to be rewritten.</li> <li>• <b>Incorrect language</b> - in a foreign language, or written so poorly the meaning can't be</li> </ul>	<p><b>Needs improvement</b></p> <ul style="list-style-type: none"> <li>• The response fails some aspects of the rubric but could be fixed in less than 30 minutes.</li> <li>• <b>Correct language</b> - but may be a little awkward or unclear</li> <li>• <b>Spelling and</b></li> </ul>	<p><b>Great response</b></p> <ul style="list-style-type: none"> <li>• Meets every aspect of the quality dimensions</li> <li>• Perfect, or could be fixed in less than 2 minutes.</li> <li>• <b>Correct language</b> - with few or no mistakes</li> <li>• <b>Spelling and grammar</b> - one or two minor blemishes ok</li> <li>• <b>Good 'fit'</b> - Fits the prompt's</li> </ul>	
-----------------	---	--	--	--

	interpreted • <b>Spelling and grammar</b> - Significant and/or distracting mistakes • <b>Doesn't 'fit'</b> - Doesn't fit the intent of the prompt	<b>grammar</b> - minor mistakes • <b>Reasonable 'fit'</b> - Doesn't fit the intent of the prompt	tone and intention • <b>Creative</b> - doesn't read or feel like a basic LLM response
--	---	---	--

### Task Types: More detail

Task Categories (Use Cases)	Description	Conditions
Rewriting	To change a body of text often per instructions e.g. rewrite this explanation in way understood by a 10 year-old or transform the text from passive to active voice	<ul style="list-style-type: none"> <li>• <b>Requires Reference Text</b></li> <li>• <b>Please don't use for translation</b></li> </ul>
Brainstorming	To develop ideas with the chat assistant, often asking questions and iterating together. <ul style="list-style-type: none"> <li>• Give me 5 ideas to plot ideas for interactive fiction related to an animal that escaped from a zoo.</li> </ul>	<ul style="list-style-type: none"> <li>• <b>Develop ideas on a theme</b></li> </ul>
Open Q&A	General questions - for best results provide context and a clear perspective <ul style="list-style-type: none"> <li>• Tell me how to make it in the video games industry from the perspective of an experienced senior engineer</li> <li>• Explain a concept [scientific, pop culture, musical, etc.] to me without reference text</li> </ul>	<ul style="list-style-type: none"> <li>• <b>NO Reference text</b></li> </ul>
Chatbot	Ask the assistant to assume a role and interact and interact for <b>one prompt and response - NOT a full conversation</b> <ul style="list-style-type: none"> <li>• You are Socrates and we will have a discussion in the Socratic manner, now let's discuss ....</li> <li>• You are interviewing at company X and I will play the interviewer. Please answer this question for me in a professionally warm tone</li> </ul>	<ul style="list-style-type: none"> <li>• <b>One turn only: Tell the chatbot how to be and start the conversation, get one response</b></li> <li>• <b>NOT a full conversation</b></li> </ul>
Classification	A particular instruction asking the chat assistant to classify data into one or more categories e.g. classify the following utterances into one of the 4 emotions: sad, intense, ecstatic, puzzled. You can use more than one category (but at most two). Here are the utterances: "I was feeling particularly tired, devoid of energy, yet with the sun's first rays, I began to tingle with energy", ....  You can ask the model to do UP TO two classifications, e.g: <ul style="list-style-type: none"> <li>• Emotion: Happy, Sad, Bored, Upset</li> <li>• Cost: Expensive, Cheap, Moderate</li> <li>• Expense report: Classify these expenses</li> </ul>	<ul style="list-style-type: none"> <li>• <b>Requires Reference Text</b></li> <li>• <b>Requires YOU to identify the dimensions - i.e. do NOT ask the model to 'identify the emotions in this text'. YOU need to tell it what to look for</b></li> </ul>

Note: **Please clearly define your dimensions and to ensure they are consistent within a classification**

- Please don't mix and match dimensions, e.g., 'Classify this Emotion: Expensive, Sad, Chocolate'

Category / Use Case	User Goal (What is the person using the LLM trying to do?)	Required	Examples
Open Q&A	Ask a question without reference material, get an immediately useful answer	<ul style="list-style-type: none"> <li>• No reference text</li> </ul>	<ul style="list-style-type: none"> <li>• Ask me anything</li> <li>• Explain a concept</li> <li>• General questions that don't require reference text (e.g., pop culture)</li> </ul>
Chatbot	Have a conversation with a chat model - have it take a point of view or play a role, and have a conversation with it	<ul style="list-style-type: none"> <li>• Only ONE TURN - one prompt and response</li> <li>• Persona or role for response to take on</li> <li>• (Optional) Start conversation in prompt</li> </ul>	<ul style="list-style-type: none"> <li>• Have a conversation</li> <li>• Prep for an interview or presentation</li> <li>• Workshop / refine a specific idea (e.g., get advice from a favorite author or expert)</li> </ul>
Brainstorming	Create a shortlist of ideas for a specific topic, activity, or problem to solve	<ul style="list-style-type: none"> <li>• Topic and/or use case that you want to start creating ideas for</li> </ul>	<ul style="list-style-type: none"> <li>• Gift or Menu ideas</li> <li>• Marketing or branding ideas</li> <li>• New products to develop or business strategies</li> <li>• New media or food discovery ('I like X, what else might I like?')</li> </ul>
Creative Writing	Create something meaningful & useful from nothing.	<ul style="list-style-type: none"> <li>• Context, tone, and format of the thing you'd like to create</li> <li>• Do NOT provide Reference Text or ask for a rewrite</li> </ul>	<ul style="list-style-type: none"> <li>• Write an email congratulating a friend on adopting a kitten, and make it cute and informal!</li> <li>• Write a bedtime story for my niece who loves astronauts and tigers</li> <li>• Come up with a creative excuse I can text my boss for why I'm not at work</li> </ul>
Closed Q&A	Ask a question about a specific reference text and get an answer	<ul style="list-style-type: none"> <li>• Reference text REQUIRED</li> </ul>	<ul style="list-style-type: none"> <li>• What was the revenue for this company in 2023?</li> <li>• Who is the top run-scorer in this document?</li> <li>• Does this menu include gluten-free options?</li> </ul>
Extraction	Extract specific information from a reference text	<ul style="list-style-type: none"> <li>• Reference text REQUIRED</li> </ul>	<ul style="list-style-type: none"> <li>• What are the 1st vowels in each word of this text?</li> <li>• What are my action items from this email chain?</li> </ul>
Classification	Provide a reference text <b>and</b> which classifications the model should use, get insights or a shortlist of examples that match classifications from the text	<ul style="list-style-type: none"> <li>• Reference text REQUIRED</li> <li>• Specific classifications to test</li> </ul>	<ul style="list-style-type: none"> <li>• Expenses</li> <li>• Sentiment</li> <li>• Dietary restrictions in meals</li> <li>• Action items identified in an email</li> </ul>
Summarization	Provide a reference text and get a summary of its contents, potentially with a focus on specific topics, or from a specific point of view	<ul style="list-style-type: none"> <li>• Reference text REQUIRED</li> </ul>	<ul style="list-style-type: none"> <li>• Financial documents or reports</li> <li>• Meeting notes</li> <li>• News articles or reviews</li> </ul>
Rewriting	Provide a reference text <b>and</b> provide the format, use case, role, or tone for the response to rewrite it into	<ul style="list-style-type: none"> <li>• Reference text REQUIRED</li> <li>• Use case, tone, role, perspective, or format to rewrite into</li> </ul>	<ul style="list-style-type: none"> <li>• Change tone of text (e.g., from recipe list to restaurant menu, from informal text to formal email)</li> <li>• Rewrite as opposing perspective</li> <li>• Comedic transformations</li> </ul>

## FAQs

- What is the word count limit for a task?
  - There isn't one, but **concise and satisfying is the goal**
  - **With some specific exceptions** like creative writing, sometimes rewriting, or when the user requests outputs like an email or letter, **a response longer than 2-3 paragraphs and/or 5 bullet points really needs a good reason for being that long**
- Should I use bullet points and **bold formatting for topic headers**?

- Bullet points and bold formatting are highly recommended for summarization, brainstorming, or any ask that involves a list
  - There needs to be a really good reason for not using bold formatting for headers and bullet points
- ...but please also use common sense, not every response should use bullet points! E.g., 'tell me a story' or 'write me an email' are clearly not intended for bullet points
- What is the word count limit for a reference text?
  - Aim for about a **typical news article** in length, or **1-2 pages max**
- Can we still paste reference text as a URL?
  - **Please do not do this any more** - it creates all kinds of issues for the system. You must paste text in the reference text field. A new linter should force people to do this.