

Evaluating AI-Generated Content

A structured review framework for judging AI-drafted content with explicit, repeatable criteria rather than gut instinct.

1. Why This Matters

Using AI to draft content is routine. The differentiator is whether the person directing it can tell good output from confident-sounding output, and say precisely why, before it ships, not after.

2. Evaluation Criteria, in Order

1. **Accuracy**, every factual/technical claim correct; overrides everything else
2. **Instruction adherence**, did it do what was asked (format, length, required elements)
3. **Clarity & tone**, easy to follow for the intended audience
4. **Consistency**, terminology, voice, formatting match the style guide

Rule of thumb: fails #1 or #2? Flag it and move on, don't fine-grain score #3-4.

3. Repeatable Checklist

- ✓ Terminology used consistently for the same concept
- ✓ Transitions: sections logically follow, not disconnected
- ✓ Unstated assumptions the audience may not share
- ✓ References/citations exist and say what's claimed
- ✓ Constraint compliance: limits and required elements followed exactly

4. Writing a Rationale That Scores Well

A rationale is judged on whether a reader sees your reasoning at a glance, not just your conclusion.

- **Lead with the decisive factor**, one sentence, before any evidence.
- Give **one** concrete piece of evidence, quote or point to specific text, then stop.
- **Constraint discipline:** write to the limit, not well under it. A short answer reads as underdeveloped even when correct.

5. Worked Example

Prompt: "Write a 100-word explanation of what a REST API is, for a non-technical stakeholder."

Draft A: uses "endpoint," "payload," "idempotent" undefined. Accurate, but fails audience fit.

Draft B: "a structured way for two computers to request and send information, like ordering from a menu." No jargon, matches the audience.

Rationale (58 words): The decisive factor is audience fit. Draft A is technically correct but uses undefined jargon that fails the explicit non-technical instruction. Draft B conveys the same concept accurately through a plain-language analogy, meeting both accuracy and instruction-adherence criteria. Draft B is the stronger response.