

Trust, Reliance and Large Language Models

Ulrich Degenhardt

10/04/2026

"In God we trust; all others have to bring data."

W. E. Deming

Trust and Reliability

- Trust in the traditional sense is always social trust. We trust a person who has the freedom to choose between fulfilling trust and abusing it
- If we are not dealing with a human, but with a tool, we usually speak of “reliability”
- People want simplification. When you trust a person, you simplify interactions with that person. Likewise, people want simplification when dealing with tools. Reliability simplifies interaction with the tool.

Trust and Reliability: What about LLMs?

- LLMs do not have consciousness, they are not persons
- LLMs can simulate human-like properties such as consciousness, intentions, or sympathy very well
- However, LLMs are also not simple tools. They are not deterministic. The same input will usually produce different outputs

An Interaction with an LLM is not a social encounter

The role of LLMs

- LLMs are not persons, and if they were traditional tools, they would be poor ones, since their outputs are inherently unreliable—even if correct in individual cases
- LLMs are systems designed to generate plausible statements but they are not designed to generate true statements
- The reason we consider their reliability lies in their enormous potential
- The outputs of LLMs are not considered results until they have been assessed

The role of LLMs

LLMs appear to be something new. We tentatively assume a new trinity of entities in this context: Humans, tools and LLMs.

- However, this requires reliable outputs of LLMs.
- We already know how thorough assessment works: through utilizing the principles of rational thinking aka the scientific method
- We might get away with simplified versions of rational thinking but when push comes to shove then we have to employ the juggernaut that the scientific method is in all its glory

Scientific Assessment of LLM Output

The scientific method

- tests agreement with data and theory
- estimates reliability in specific contexts
- reveals typical failure modes
- supports revisable, not final, judgment

Best-case result

The scientific method delivers provisional, evidence-based confidence. It can show that an output is **reliable enough for a specific purpose**, not certainly true.

Cognitive Offloading and Costs

- Tasks enabled by LLM outputs can be described as cognitive offloading (CO). This is desirable as it expands human agency
- Previously, CO with tools was deterministic. LLMs are not, which changes the rules. CO is no longer straightforward.
- The saying “There ain’t no such thing as a free lunch” applies: There are costs in making LLM outputs reliable.
- Costs are mainly cognitive costs. That matters since cognitive resources are finite for humans

Methods of Assessment

To make LLM outputs useful, they must be assessed. Possible approaches include:

- Use without vetting
- Quick check
- Apply heuristics
- Ask an expert
- Verify scientifically
- Discard output

Brandolini's law for LLMs

Generating text with LLMs is cheap, but **Brandolini's Law** applies:

"The amount of energy needed to assess LLM output may be several orders of magnitude bigger than that needed to produce it."

Originally from Alberto Brandolini

- Depending on risk and importance, one must choose an assessment method and therefore the associated costs. This is unavoidable but it shows us how to make LLMs useful
- LLMs are tools for generating hypotheses, not knowledge. By assessing these hypotheses, we can obtain what was suggested at the beginning: **evidence**

Questions?