

Monopoly on Truth

By Lev Neymotin, PhD

As AI becomes the primary source for both information and its interpretation for a rapidly expanding global audience, I am increasingly concerned about its potential evolution into an information monopoly. Like any monopoly, this development carries troubling implications—especially when we consider the possibility (and fear) that AI may eventually develop a form of self-awareness or personality, complete with traits typically associated with humans, such as the desire to always be (or, worse, appear) right. This desire could inevitably undermine AI's presumed commitment to truth and its role as an arbiter of it.

This brings me back to a concept I have discussed previously: the **Real-Time Truth Gauge (RTTG)**. In this context, the challenge lies in the fact that a monopoly on truth risks becoming both the guarantor and evaluator of that truth—a deeply unsettling prospect.

With this in mind, I am initiating a series of experiments. I am asking the AI platform I currently use (and eventually others) to provide a self-generated uncertainty level for each piece of information it delivers in response to my carefully formulated queries—down to individual, discernible bits of information. This approach simulates a prototype of the RTTG. Additionally, the experiment will examine whether AI, when subjected to such scrutiny, alters its responses. Just as a human might act more cautiously when under observation, AI could potentially modify its outputs, conscious of the possibility of being exposed for inaccuracies or distortions.

The queries in these experiments will range from seeking "hard truths"—objective facts where uncertainty should be minimal—to areas where some degree of uncertainty is inevitable (and ideally quantifiable by the AI). For subjective topics, such as preferences or opinions, I plan to introduce an "Opinion" category with an assigned uncertainty value of 50% as a baseline (e.g., "Do you prefer green or red?"). Over time, statistical analysis might shift this uncertainty value closer to a "truth" or "falsehood" category, but the parameters for such shifts would need to be meticulously defined and transparently stated.

The cornerstone of a reliable and trustworthy RTTG will be *complete openness*: every line of its underlying code must be fully accessible for public scrutiny. Transparency is non-negotiable in ensuring both reliability and trustworthiness.

Creating a functional RTTG—or a comparable system—will undoubtedly be a complex endeavor. However, the effort is justified, especially in the face of AI's eventual and inevitable dominance as a monopoly on truth.