OPEN ACCESS
Indian Journal of Artificial Intelligence and Neural Networking (IJAINN)
ISSN: 2582-7626 (Online), Volume-5 Issue-6, October 2025

# Critical Understanding of LLM-Generated Statements

**Reeshabh Choudhary**

***Abstract:** Now that we live in a world where most of the text in recent online interactions that we come across seems to be generated by LLMs, it becomes critical to understand the nature of statements being generated by LLMs. Technology has always been sold to humans under the tag that it is foolproof and will make lives easier. LLMs produce text by predicting the next token or sequence based on probabilities derived from their training data. A question then arises, whether they generate a 'probability statement' or 'probability of a statement'. The difference between the two may seem elusive, but it is actually quite obvious. This paper intends to bring forward that difference to its audience, who, in turn, can understand the capabilities of the machine they are using and adapt a better framework to judge and use the response generated by LLM models in their applications.*

***Index Terms:** Artificial Intelligence, Large-Language Models, Probability, Human Judgment, Intelligence*

***Abbreviations:***
Large Language Model

## I. INTRODUCTION

Since the advent of transformer models in 2017, we have seen a dramatic rise in the capabilities of machine learning language models in text generation [1]. However, their adaptation was much later hyped by the release of ChatGPT, a public-facing interface for OpenAI models, where users can prompt and get responses. LLMs have captured imaginations, and more of the media have followed, with promises of rapid, even unparalleled, productivity growth made since then. There are no doubts about the impressiveness of the text or information produced by the generative models in response to simple user prompts. But the validity of the responses generated by large language models has remained in question [2]. LLMs are not sentient beings but "stochastic parrots" [3]. Their capability is fundamentally statistical, driven by pattern matching rather than cognitive processes. Unlike their counterparts, the mortal sentient beings, who can reason abstractly or infer causality, LLMs lack an internal model of the world that would allow them to understand why specific patterns exist or how they relate to real-world phenomena. LLMs learn to associate inputs with outputs based on probabilities derived from training data, without any inherent understanding of the underlying concepts or causal relationships. for example, when an LLM generates a response to a question, it does so by calculating the likelihood of certain words or phrases given the input context and entirely relies on patterns observed during training. A question then arises, whether LLMs generate a 'probability statement' or 'probability of a statement'.

**Manuscript received on 03 September 2025 | First Revised Manuscript received on 19 September 2025 | Second Revised Manuscript received on 21 September 2025 | Manuscript Accepted on 15 October 2025 | Manuscript published on 30 October 2025.**
*Correspondence Author(s)
**Reeshabh Choudhary***, Senior Technical Architect, Department of Automation COE, Company Name: Eversana, India. Email ID: reeshabh.choudhary@eversana.com, ORCID ID: 0009-0003-4750-8151

© The Authors. Published by Lattice Science Publication (LSP). This is an open-access article under the CC-BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

## II. PROBABILITY STATEMENT

Probability statements can never be strictly contradicted by experience, even though we assume that all external perturbations and all observational errors are entirely removed. Suppose that we place 100 balls in a sack, out of which five are black and the rest are white. If we now draw a black ball randomly out of the sack in the very first attempt, we would be surprised, but would never question the number of black balls present in the sack. However, the same cannot be said for scientific assumptions based on probability. The probabilistic statements cannot be strictly falsified. However, they can be questioned if observed outcomes consistently deviate from expectations in a way that seems highly improbable given the stated probability. "The probability of rolling a six on a fair die is 1/6" is a probability statement as it quantifies the likelihood of a specific outcome in a repeatable experiment. It is a statement rooted in the structure of the dice. There can be six such probability statements referring to the throw of the die, such as 'the probability of a one to be thrown is 1/6', 'the probability of a two to be thrown is 1/6, and so on. And these statements can be true simultaneously. However, statements such as "a six will be thrown or four will be thrown" will not be considered probability statements. These are contradictory statements which refer to propositions about the outcome of a dice throw. A die can only land on one face and hence will only have one uppermost side at any point in time. This makes the above statements, such as "a six will be thrown or four will be thrown," mutually exclusive, as a single dice throw can only produce one outcome. A probability statement is not about holding partial beliefs in contradictory propositions; instead, it describes the objective likelihood of individual events, typically expressed in terms of a numerical probability or chance. What makes "probability statement" unique is that it refers to objective properties of events in the world, which are grounded in empirical or theoretical frameworks, rather than subjective beliefs or propositions about statements. A probability statement is not a psychological state of believing multiple contradictory outcomes at once.

## III. PROBABILITY OF A STATEMENT

The degree of belief in a proposition can be referred to as "probability of a statement". A proposition is an idea or opinion that an entity may express about a subject. Such a statement reflects confidence in a statement's truth rather than an event's occurrence. This can be understood by revisiting the famous experiment conducted by Daniel Kahneman [4] and Amos Tversky, in which they created a personality sketch of an imaginary individual, Tom W, as follows: Tom W is of high intelligence, although lacking in true creativity. He needs order and clarity, and for neat systems in which every detail finds its appropriate place. His writing is rather dull and mechanical, occasionally enlivened by somewhat corny puns and by flashes of imagination of the sci-fi type. He has a strong drive for competence. He seems to feel little sympathy for other people and does not enjoy interacting with others. Self-centred, he nonetheless has a deep moral sense. He then suggested that the group of participants take a sheet of paper and rank the nine fields of specialization listed below in order of the likelihood that Tom W is now a graduate student in each of these fields: 1. Business administration 2. Computer Science 3. Engineering 4. Humanities and education 5. Law 6. Medicine 7. Library Science 8. Physical and Life Sciences 9. Social science and social work. When this task was administered to a group of graduate students in psychology, most of them ranked computer science and engineering as the most probable streams (you can cross-check your answers honestly). Ironically enough, this group of students was aware of the intuition bias and were familiar with base rates of different fields. Yet, they did not engage their knowledge but were lured by the representativeness and ignored base rates, and did not even doubt the authenticity of the description based on the judgment. Had the above information not been presented and just asked to gauge Tom's probability of joining a stream, they would have considered the base rates of different fields, i.e., the average number of students joining each field in each set.

## IV. LLM'S DILEMMA

LLMs can produce statements that resemble probability statements, such as "There is a 70% chance of rain this afternoon," if trained on weather-related data or prompted to generate such claims. However, it is not difficult to dismiss the claims of objectivity in such statements. The argument presented will be on the lines that unless the LLM is explicitly drawing on a probabilistic model tied to real-world data, such statements cannot be considered "probability statements". When LLM generates a statement, it reflects a probabilistic weighting of possible outputs based on its training data, not a direct assessment of real-world events. The model assigns probabilities to tokens or sequences that express a "degree of confidence" in the statement based on the input context. However, their "probabilities" are statistical artefacts of their architecture, not evaluations of evidence like in scientific inquiry. On their own, LLMs do not directly compute objective probabilities of events unless integrated with external systems (e.g., real-time data feeds or statistical models). But we must also investigate the case of real-time data integration closely, as many such integration systems are possible and are, in fact, at play in the recent and upcoming software architecture. Once integrated with a real-time data source, the nature of the output from an LLM changes, as it can access current, empirical information about the world, allowing it to generate statements informed by actual evidence. However, whether the resulting statements qualify as probability or remain expressions of a degree of belief depends on how the integration is implemented and the context of the output. Even though integration with a real-time data source provides the context for output, the fundamental nature of LLM remains the same, which means its phrasing, interpretation, and presentation of real-time data depend on patterns learned during training. For example, suppose the training data contains many instances of rounded or simplified probability statements. In that case, the LLM might approximate or rephrase the input data rather than directly reporting the exact probability. This introduces a layer of subjectivity, since the LLM's output is not a direct computation of probability but a translation of the input data into language, filtered through its learned patterns. For an LLM to produce a probability statement that asserts an objective chance about a

real-world event, it must act as a faithful conduit for the probabilistic model's output, without altering or reinterpreting the data through its language generation process. For the LLM's output to be a true probability statement, it must either output the numerical probability provided by the external model (e.g., a weather model's 70% probability) without modification or perform its own probabilistic calculation using the real-time data, following a rigorous statistical or computational method. The moment an LLM merely interprets or paraphrases the real-time data (e.g., converting "70% chance of rain" into "It's likely to rain"), the output risks being a degree of belief that is nothing but a linguistic approximation shaped by the model's training rather than a direct reflection of an objective probability. For instance, an LLM integrated with a weather forecasting system that outputs "70% probability of rain at 3 PM based on radar and atmospheric data" may directly report this as "There is a 70% chance of rain this afternoon". This qualifies as a probability statement because it accurately conveys the objective probability computed by the external model, with minimal influence from the LLM's training data. Accountability is maintained if the system can trace the statement back to the weather model's calculations. However, suppose the LLM, influenced by its training data, rephrases this as "It's likely to rain this afternoon". In that case, this reflects a degree of belief in the technical sense, as the LLM's choice of "likely" is shaped by linguistic patterns in its training data rather than a direct calculation of probability. The statement loses precision and objectivity, as "likely" is ambiguous compared to "70%". This output cannot be held fully accountable as a probability statement because it introduces subjectivity from the LLM's language generation process. On the other hand, if the LLM is equipped with a module to process real-time meteorological data and compute probabilities using a statistical model, then it could independently generate a statement like "There is a 70% chance of rain this afternoon" based on its own calculations. This would qualify as a probability statement, as it reflects an objective probability derived from empirical data. However, this requires the LLM to have a transparent probabilistic model that performs its own probabilistic calculations using explicit and reproducible methods, and the output's accuracy depends on the quality of the data and the


© Copyright: All rights reserved.

model's design. Most current LLMs are not designed for such calculations, relying instead on external models or linguistic patterns. In summary, the accountability of a statement being made is much more critical in today's time. While quoting statements directly from a generated output, users should have access to metadata or logs showing how the probability was derived.

## V. CONCLUSION

LLMs generate creative or persuasive text; they do not express intention but instead recombine patterns based on the input prompt. Clearly, they often struggle with tasks requiring them to go beyond their training, such as solving novel problems outside their training distribution or handling counterfactual reasoning. For example, LLMs cannot accurately answer questions about summing numbers in an Excel sheet and often resort to measures like using an agent (tool), which is essentially" automation" rather than actual intelligence. In fact, LLMs are not capable of genuine intelligence. Current AI models perform well on simple, repetitive tasks with clear outcomes, such as text summarisation or image classification. However, for complex tasks that require specific context—such as diagnosing intricate medical conditions or making strategic decisions—human judgment remains essential. Most occupations rely on tacit knowledge that AI models or LLMs cannot reproduce. Because LLMs depend on pattern matching, they are prone to errors in out-of-distribution scenarios or when faced with adversarial inputs. For instance, they may generate "hallucinations" (false but plausible outputs) when faced with ambiguous or novel questions. While statistical methods provide objective tools such as p-value, the interpretation of these results involves a human decision. The scientist must weigh the evidence, consider the context, and decide whether the improbability of the observed result justifies rejecting the hypothesis. In current times, when most of us are using LLMs which are probabilistic generative models to look for answers to most of our queries, we must therefore be careful to accept the answers generated by the model, as often the model is unaware of the entire context of a person's situation and generates response based upon the context pre-loaded and its comparison with the description of context provided by the user. For the same reasons, these models can only assist in the decision-making process by providing a probabilistic view; however, the responsibility for inference lies solely with the human in the loop. Hence, one must not factor out human judgment when dealing with LLM-generated texts in day-to-day life. In fact, a proper critical analysis of the text is advisable in such scenarios, as it might be just resonating the sentiments prompted without any accountability.

## DECLARATION STATEMENT

I must verify the accuracy of the following information as the article's author.

- **Conflicts of Interest/ Competing Interests:** Based on my understanding, this article has no conflicts of interest.
- **Funding Support:** This article has not been funded by any organizations or agencies. This independence ensures that the research is conducted with objectivity and without any external influence.
- **Ethical Approval and Consent to Participate:** The content of this article does not necessitate ethical approval or consent to participate with supporting documentation.
- **Data Access Statement and Material Availability:** The adequate resources of this article are publicly accessible.
- **Author's Contributions:** The authorship of this article is contributed solely.

## REFERENCES

1. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017. DOI: https://doi.org/10.48550/arXiv.1706.03762
2. D. Acemoglu, "The simple macroeconomics of AI," *Economic Policy*, vol. 40, no. 121, pp. 13–58, 2025. DOI: https://doi.org/10.1093/epolic/eiae042
3. E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, "On the dangers of stochastic parrots: Can language models be too big?" in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*, New York, NY, USA: ACM, 2021, pp. 610–623. DOI: https://doi.org/10.1145/3442188.3445922
4. D. Kahneman, *Thinking, Fast and Slow*, new ed. London, U.K.: Efinito, Jul. 2022, ISBN 9781802063059. https://www.abebooks.com/9781802063059/Thinking-Fast-Slow-Daniel-Kahneman-1802063056/plp

## AUTHOR'S PROFILE

**Reeshabh Choudhary,** B.E. (Computer Science), is a student for life and is eager to learn and share academic knowledge among students and peers. He works as a Senior Technical Architect at Eversana India. He has written three books: "There Is No AI", "Concurrency & Parallelism" and "Objects, Data and A.I".

**Disclaimer/Publisher's Note:** The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of the Lattice Science Publication (LSP)/journal and/or the editor(s). The Lattice Science Publication (LSP)/ journal and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.


*© Copyright: All rights reserved.*