Generative AI And Large Language Models in Intelligent Systems

Kunal Bendure[1], Chaitanya Shinde[2], Ashwini Gagare[3]
[1,2,3]*MAEER's MIT Art's, Commerce and Science College, Alandi(D), Pune*
doi.org/10.64643/ijirt.208537-459

***Abstract*—Over the last couple of years, Generative AI and Large Language Models (LLMs) have quietly become part of how intelligent systems get built. What used to be narrow, rule-bound software is now doing things like writing content, helping with decisions, and holding conversations that feel almost natural. This paper is our attempt to make sense of where things currently stand: how these models are actually being used across healthcare, education, agriculture, and automation, and what is happening underneath the architectures, the training tricks, the way prompts are written. We have also tried not to gloss over the rough edges. Hallucination, bias, privacy, the cost of just running these things, and the ethical questions that come up once they are deployed at scale all of that gets a fair look here too. Toward the end, we talk about what responsible development might actually look like and where there is still real work left to do. Mostly, we just wanted to write something a student or early researcher could pick up and come away with a clear-eyed sense of what Generative AI can do well, and where it still trips up.**

***Index Terms*—Generative AI; Large Language Models (LLMs); Intelligent Systems; Healthcare; Education; Agriculture; Automation; Prompt Engineering; Responsible AI; Explainable AI; Model Architecture; Ethical AI**

I. INTRODUCTION

AI has gone through a few distinct eras at this point. It started with rigid, rule-based expert systems, moved into statistical machine learning, and has now landed on deep learning models that do not just sort or predict they generate. That is really the whole idea behind Generative AI: instead of classifying data that already exists, it produces new text, images, audio, or code from scratch. Large Language Models are probably the piece of this shift that ordinary people have actually noticed. Most of them are built on the transformer architecture [1], trained on enormous amounts of text, and somewhere in that process they pick up enough of a feel for language to respond in ways that sound coherent and, more often than not, actually relevant to what was asked.

Models like GPT, PaLM, and LLaMA have taken intelligent systems well past what they used to be capable of [2]. A single model these days can summarize a report, translate a sentence, reason through a problem, write code, and hold a conversation often without being retrained for each of those jobs separately, which was simply not how things worked a few years ago. What we are trying to do in this paper is pull together recent work on Generative AI and LLMs, with a real focus on how they are actually being used across different fields, and honestly, what is still missing before anyone should trust them in higher-stakes settings.

II. BACKGROUND AND RELATED WORK

Almost all of the recent progress here traces back to the transformer architecture [1], which pushed aside the older recurrent and convolutional approaches that used to dominate sequence modeling. Its self-attention mechanism basically lets the model work out which words in a sentence matter to each other, even when they are far apart and that turned out to be a far better way of handling long-range context than what came before. Pair that architectural shift with massive training datasets and a steady climb in available compute, and you get the scaling story that took language models from a few million parameters to hundreds of billions [2], [4].

Training methods have moved forward too. The usual playbook now is to pretrain on a broad, unlabeled text corpus, fine-tune on something more specific, and then run reinforcement learning from human feedback, or

RLHF [3], as a final alignment step that nudges the model's answers closer to what people actually find helpful and safe. Prompt engineering has also turned into its own small skill set instead of retraining a model every time a new task shows up, people have learned that phrasing the instruction carefully often gets the job done just as well.

That flexibility is a big reason LLMs have ended up embedded in so many different kinds of systems without needing much extra engineering behind the scenes.

III. APPLICATIONS OF GENERATIVE AI IN INTELLIGENT SYSTEMS

A. Healthcare

In healthcare, most of what we see right now is LLMs drafting clinical notes, answering medical questions, helping with patient communication, or summarizing dense literature that would otherwise take a clinician hour to get through. A model that can turn a doctor's rough shorthand into a clean discharge summary, or explain a diagnosis to a patient in plain words, genuinely saves time. Of course, that only works if the output is actually correct, and that is exactly where a lot of the hesitation around clinical use comes from.

B. Education

In education, these models are turning up as tutoring assistants, practice-material generators, and quick feedback tools for assignments. They are also proving handy for language learning a student can practice a conversation at two in the morning without needing a human on the other end. What really makes them useful in this space is that they can adjust how they explain something depending on how much the learner already knows.

C. Agriculture

Agriculture is a less obvious use case, but it is growing. Chatbots built on LLMs can walk farmers through crop decisions, help identify a pest or disease from a description, and do it all in a local language instead of English.

In places where an actual agricultural expert is hard to reach, that kind of tool fills a gap that would otherwise just sit empty.

D. Automation and Human-Centered Applications

On the automation side, LLMs increasingly power agents that take a plain-language instruction and turn it into something that actually runs generated code, an automated workflow, a robotic process. And in the more every day, human-facing corner of things, virtual assistants and customer support bots benefit from a model's ability to hold a conversation together over many turns, rather than losing the thread after a couple of exchanges the way older chatbots used to.

IV. CHALLENGES

None of this comes free, though. Hallucination a model confidently saying something that sounds right but is not is probably the most talked-about problem, and it gets genuinely dangerous in fields like healthcare or law where a wrong answer actually matters.

Bias is another one: whatever unfair or skewed patterns exist in the training data tend to show up again in the model's outputs, just less obviously. Privacy is its own headache, especially when models are trained on or exposed to information that was never meant to be public. Then there is the plain cost of it all running these models at scale eats serious computing power, which means real money and a real environmental footprint. And beyond the technical stuff, there are the ethical questions that keep coming up: misinformation, academic dishonesty, manipulation all reasons this space needs actual oversight, not just good intentions.

V. RESPONSIBLE AI PRACTICES

Dealing with all of this has to happen across a model's whole lifecycle, not just tacked on at the end. That means paying attention to what data goes into training, being honest about what the model cannot do, building real ways to check for factual accuracy, and keeping a human in the loop wherever a decision actually matters. Retrieval-augmented generation, or RAG [5], is one technique worth mentioning here it grounds a model's answers in verified outside sources instead of relying purely on what it memorized during training, and that tends to cut hallucination down noticeably. Explainability tools that surface some of the reasoning behind an answer help too, especially anywhere accountability is on the line.

VI. OPEN RESEARCH DIRECTIONS

There is still a lot of open ground here. People are working on more efficient architectures that could bring compute and energy costs down without sacrificing performance, better ways to keep model outputs grounded in fact, stronger frameworks for catching and reducing bias, and benchmarks that actually test for safety and reliability instead of just raw capability. None of that happens on its own it is going to take AI researchers, people who actually understand the domains these models get used in, and policymakers all working together for any of this to roll out responsibly.

VII. CONCLUSION

Generative AI and LLMs have genuinely changed what intelligent systems can do, opening up new ways to generate content, support decisions, and interact with people across healthcare, education, agriculture, and automation. But getting the most out of that potential, safely, still comes down to tackling hallucination, bias, privacy, cost, and ethics head-on rather than treating them as afterthoughts. This paper was our attempt to lay out where things currently stand, as a starting point for students and researchers just getting into the field, while pointing at the work still left to do before these systems are as safe, reliable, and explainable as they need to be.

REFERENCES

[1] Vaswani et al., “Attention is all you need,” in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2017, pp. 5998–6008.
[2] T. Brown et al., “Language models are few-shot learners,” in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2020, pp. 1877–1901.
[3] L. Ouyang et al., “Training language models to follow instructions with human feedback,” in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2022, pp. 27730–27744.
[4] R. Bommasani et al., “On the opportunities and risks of foundation models,” arXiv preprint arXiv:2108.07258, 2021.
[5] P. Lewis *et al.*, “Retrieval-augmented generation for knowledge-intensive NLP tasks,” in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2020, pp. 9459–9474.