Are LLMs Mirrors or Minds?

N.H. Pavao

August 31, 2024

Abstract:

In this essay, I introduce the concept of **Turing's Mirror**, a new dimension for evaluating language models (LLMs) that challenges the traditional assumptions implicit in the Turing Test. The Turing Test posits that if a machine can engage in conversation indistinguishably from a human, it demonstrates intelligence. However, this paradigm presupposes that passing the Turing Test means that LLMs are agentic entities with a fixed degree of verbal fluency. In other words, some models can pass the Turing Test, and others cannot — traditionally these are the only two outcomes. Turing's Mirror is an alternative outcome of the Turing Test, where these models are less like autonomous agents with intrinsic qualities that they project, and more akin to mirrors that reflect the cognitive capabilities and verbal sophistication of the test administrator.

Through this lens, I argue that LLMs may pass the Turing Test when evaluated by experts who bring complex, nuanced queries, effectively reflecting their advanced understanding – but the same models may fail when subjected to simpler queries from non-experts, revealing their limitations as mere pattern recognizers and next-token-predictors rather than genuine thinkers. This essay proposes that LLMs should be understood as tools that reflect our inputs rather than as independent sources of intelligence. This paradigm shift could have significant implications for how we approach the development, deployment, and evaluation of LLMs, urging us to reconsider whether progress in this field requires more powerful models or simply more sophisticated interactions.

Since November 30, 2022, with the launch of ChatGPT, I've observed the commercial deployment of language models with a mix of fascination and skepticism. Not long ago, interacting with a chatbot felt like talking to a parrot rather than engaging in a real conversation—responses were clunky, repetitive, and painfully limited. But then, seemingly overnight, something changed. These models weren't just stringing together coherent sentences anymore—they were producing text that was strikingly human-like, nuanced, and, at times, genuinely insightful.

It was as if the resolution of an image had suddenly been cranked up from a grainy 32 x 32 pixel display to a crisp 4096 x 4096, where imperfections vanished, and the details became indistinguishable from reality. This transformation got me thinking. This leap in LLM capability wasn't magical, but was instead the result of scaling laws predicting improvements as more compute power and data are applied. Yet, despite these advancements, a deeper question emerged in my mind: Are these models truly becoming more intelligent, or are they just getting better at reflecting our own intelligence?

Traditionally, we've used the Turing Test as a benchmark for determining whether a machine could pass as human in conversation. The testing scenario is simple enough: a test administrator sits behind one-way reflective glass, unable to see the entity on the other side, but knowing that it could either be a human or a machine. The task is to decide which one it is based solely on the conversation. If the administrator can't reliably distinguish the machine from a human, then the machine has passed the Turing Test—it has successfully mimicked human conversation.

However, the real question might not be whether there's a machine or a human behind the glass. What if the entity isn't an agentic, but simply the mirror itself? What if the administrator is merely making an assessment of "humanness" based on their own reflection? I call this scenario **Turing's Mirror** - and I believe there's a possibility that many of these LLMs are more closely resemblant of a reflective surface, rather than some agent with a fixed degree of verbal fluency.

Turing's Mirror suggests that what we're actually evaluating isn't an independent entity but rather a reflection of our own cognitive abilities. In my mind, this perspective changes a lot about how we can frame interactions with these models and leverage them to our needs. It's no longer just about whether the LLM can convince us it's human—by performing traditionally human tasks that match or exceed our competence level—rather, it's about how much of ourselves we see in the LLM's responses and shape our own queries and practices around that.

In this short essay, I'll explore what it means to see LLMs not as independent thinkers but as mirrors that amplify and reflect the complexity of the questions that we ask them.

Turing's Mirror

As I delved deeper into this idea this morning, it became clear that thinking of LLMs as mirrors offers a new way to understand our interactions with them. When engaging with one of these models, it's easy to imagine you're conversing with a digital mind—some homunculus hidden behind the screen, processing your words and crafting responses.

Let's consider the Turing Test again, but now with a twist: instead of asking whether the entity behind the glass is a human or a machine, the question becomes whether there is any agent behind the glass at all. What if the administrator is not seeing an independent mind but merely a reflection of their own mind in Turing's Mirror? When you ask an LLM a question, it doesn't have its own ideas or consciousness. It's not pondering your question like a human would. Instead, it's doing what it was designed to do—predict the next word, phrase, or idea based on everything it has learned from the massive datasets it was trained on. However, the quality and depth of the answer you receive are heavily influenced by how you frame the question. It's just performing "next token prediction"—it doesn't have a mind. It doesn't have beliefs. It doesn't have an agenda, or agency, or any traditional notion of a worldview. The LLM is just holding up a mirror to your own words, reflecting back your level of understanding.

The more sophisticated your input, the more sophisticated the output. A simple, straightforward question yields a simple, straightforward answer. But if you challenge the model with a complex, nuanced query

that requires deeper exploration, the LLM delves into its vast reservoir of knowledge and returns something richer, more layered. It's not creating new knowledge—it's reflecting the complexity you brought to the table.

This mirror analogy also explains why different people tend to have vastly different experiences with the same LLM. Someone with a deep understanding of a topic might receive a response that feels insightful and almost human-like because the LLM is reflecting the depth of their inquiry. Meanwhile, someone else might find the same LLM simplistic or underwhelming simply because their questions didn't push the model to reveal its full potential. Lack of domain expertise embedded in the queries, yields simple inferences from the models.

Interacting with an LLM through the lens of Turing's Mirror suggests that we aren't testing the LLM itself but are instead evaluating our own input. The more I bring to the conversation, the more the LLM seems capable of giving in return. But it's not the LLM that's changing—it's my perception, shaped by the quality of the interaction I initiate.

The concept of Turing's Mirror challenges us to rethink how we assess these models. If what we're seeing is merely a reflection of our own cognitive abilities, then perhaps the LLM's apparent intelligence is not an inherent quality but a reflection of the user's input. This shift in perspective is crucial for understanding and maximizing the potential of these models.

The Implications of Turing's Mirror

Turing's Mirror sheds light on why there's such a stark divide in how different groups perceive advancements in language models. On one hand, technical experts view models like GPT as a monumental advancement, almost like unlocking a new layer of reality where machines demonstrate abilities once thought uniquely human. From this view, we are on the precipice of a Cambrian explosion in commoditized intelligence. I don't just mean experts in machine learning. Anyone with domain expertise is likely to get a response that is reflective of their superior understanding.

But this perception isn't universal. Pundits without domain expertise—those not as deeply embedded in the technical intricacies of the problem they're asking about—often have a more tempered view. To them, these models still seem like glorified chatbots, capable of producing impressively coherent text but fundamentally lacking in key areas. They clearly see the limitations: no symbolic self, no true reasoning, no generative ideas that weren't somehow seeded by human input, and no genuine self-reference.

This divide in perception is precisely where Turing's Mirror comes into play. For technical experts who understand the nuances of how these models work, interacting with an LLM is like seeing their deep knowledge and high-level thinking reflected back at them. The LLM appears almost as an intellectual peer, responding to the complexity and precision of their inquiries with equally complex and nuanced output. For this reason, I have a conversation with GPT4 almost every morning to help prioritize my work schedule for the day.

Meanwhile, mid-level experts, who might approach these models with more straightforward or less technically precise queries, encounter an LLM that doesn't seem as revolutionary. To them, it's still just a machine spitting out text based on patterns—impressive, yes, but not showing the reasoning or self-awareness that would indicate a true leap into something more akin to human intelligence. The potential for productivity gains when this is the reflection you see are hard to imagine.

This difference in experience isn't about the LLM itself—it's about what the LLM reflects. The breakthroughs some see are, in a way, a reflection of their advanced understanding and the sophisticated questions they pose. The LLM draws from vast reservoirs of data and patterns, but it's the user's input that determines whether the output feels like a breakthrough or just more of the same. It's the 55/10 principle in practice – outlining the solution with an appropriately posed question essentially solves the problem.

So, what does this mean technically? It means that the perceived intelligence of these models is largely a mirror of the user's cognitive and linguistic abilities. The LLM isn't independently generating new ideas or reasoning as a human would—it's recombining and synthesizing information based on the input it receives. For those who push the model with advanced, nuanced queries, it appears to be on the cusp of something profound. For others, it's still limited by the boundaries of pattern recognition, lacking the higher-order thinking associated with true intelligence.

This understanding challenges us to reconsider what it means when we talk about breakthroughs in language models. Is it the LLM itself that is advancing, or is it our ability to interact with it in more complex ways that's evolving? Turing's Mirror suggests the two are deeply intertwined and that the real advances might be as much about the users as they are about the models.

This perspective not only explains the differing views on the current state of language models but also invites critical thought about where true innovation lies. As these models develop, recognizing the role of the user in shaping the perceived capabilities of LLMs will be crucial. It's not just about making smarter models—it's about how we engage with them and how that engagement reflects back the intelligence we bring to the table.

Strategies for Maximizing LLM Capabilities

Understanding LLMs as mirrors fundamentally changes how I approach maximizing their potential. If the model's output is merely a reflection of the input I provide, then unlocking its full potential hinges on how I engage with it. The more sophisticated and precise my approach, the more I can push the boundaries of what I can achieve with the model.

One effective strategy is **query reformulation**—taking a straightforward question and iteratively refining it to increase its complexity and depth. Imagine starting with a simple inquiry like, "How does an LLM influence decision-making in business?" That's broad, and while the model might provide a decent answer, it's likely to be general. However, by breaking it down—asking about specific decision-making models, introducing ethical considerations, or linking it to real-world case studies—the LLM can generate

more nuanced, detailed insights. It's not just about getting an answer; it's about exploring the problem space in greater depth – an ability refined over years of developing personal expertise.

Another approach is using the LLM to **synthesize insights across different domains**. For instance, when exploring the implications of quantum computing on cryptography, the LLM can pull from various fields—quantum mechanics, information theory, cybersecurity—and help reveal connections that might not be immediately obvious. This cross-disciplinary synthesis is one of the model's strengths, but it's only as good as the connections I encourage it to make and the team of human experts that I build around it. By framing questions in a way that bridges different fields, I can leverage the LLM to produce more innovative and insightful responses - while preserving the jargon that proves valuable in different technical disciplines, but breaking down language barriers shared between the human agents. If I can look at myself in an LLM query, then I can also use it to look at other technical experts with the same level of resolution to which I am accustomed.

Iterative exploration—using feedback loops to refine and deepen understanding—is also crucial. This method is especially useful when dealing with abstract or complex concepts. Starting with a high-level question, receiving a response, and then diving deeper based on that response allows me to continually refine my queries until the layers of meaning or insight are fully uncovered. This approach turns the LLM interaction into a dynamic, inductive, and evolving conversation rather than a one-off query.

Self-reflection on communication is perhaps the most important strategy. The model's output directly reflects the input it receives. That means the better I am at articulating my ideas, the more effectively I can leverage the LLM. This isn't about using bigger words or more complex sentences—it's about clarity, precision, and structuring queries to get the most out of the model. It's like fine-tuning a musical instrument; the better the tuning, the more harmonious the result – and with a team LLMs and humans, the more triumphant a vibrant the orchestra.

Ultimately, these strategies underscore that the LLM is an incredibly powerful tool, but it remains a tool nonetheless. Its real strength lies in how well I engage with it, how deeply I push it, and how effectively I translate its outputs into something meaningful. By refining my approach, I'm not just getting better answers—I'm expanding the scope of what the LLM can reflect back.

Maximizing their capabilities isn't just about improving the model—it's about improving how I interact with it. Turing's Mirror reminds me that the depth and quality of my engagement ultimately determine the value of these interactions. As I continue to explore and push the boundaries of LLMs, keeping this in mind will be crucial to unlocking their full potential.

Conclusion: The Future of LLMs—Are We Asking the Right Questions?

As we stand on the brink of what many consider a revolutionary moment in language modeling, it's worth pausing to consider whether we're truly maximizing the potential of these models—or if we're simply projecting our limitations onto them. Turing's Mirror suggests that LLMs aren't independent agents with inherent intelligence but rather sophisticated reflectors, amplifying the complexity and nuance of the

inputs they receive like photomultipliers. This raises a critical question: Are the perceived limitations of these models truly inherent to their architecture, or are they simply a reflection of the limits of our queries?

Obviously, the debate is far from settled. On one hand, some argue that we've already hit the resolution limit of these LLM mirrors. The "pixels" are so finely tuned, so densely packed with data, that further increases in compute or model size might yield diminishing returns. If this is the case, the future of LLM development might not lie in building bigger models but in refining our ability to interact with them—learning to ask better, more sophisticated questions that push the boundaries of what these models can reflect back.

On the other hand, undeniable shortcomings in current LLMs suggest there's still room for improvement, perhaps even a necessity for it. The models we use today can exhibit biases—political, social, scientific, religious—that raise important ethical and practical concerns. Are these biases a result of the data they've been trained on, or are they a function of how they're fine-tuned? Or in the spirit of Turing's Mirror, are they a function of the questions that we choose to ask it? Could these biases be a reflection of the user queries, revealing more about us than about the LLM?

This ambiguity is precisely what makes the future of LLMs so compelling and uncertain. If the models are sensitive to user inputs in this way, then the path forward might involve developing better training data, refining fine-tuning processes, or creating mechanisms that allow users to better control the biases they encounter. However, if these biases and limitations are intrinsic to the architecture—deeply embedded in the way these models learn and generate responses—then the question becomes whether more compute, more data, or entirely new architectures are needed to overcome these challenges.

Moreover, the technical limitations we observe—like the difficulty LLMs have with reasoning, self-reference, or generating genuinely novel ideas—raise questions about the true nature of intelligence. Are these simply reflections of the current state of the art, or do they point to more fundamental constraints within the architecture of LLMs? Could it be that, much like increasing the resolution of an image beyond what the human eye can perceive, we've actually *already* reached a point where the "resolution" of these models exceeds our ability to fully leverage it? Maybe they are already capable far beyond our imagination, and we are simply limited in our capacity to pose appropriate queries. This would place the onus on us, where, as Douglas Adams told us, we need to evolve our queries, our understanding, and our methods of engagement to unlock new potentials and ask better questions.

As we look to the future, these ideas should shape the direction of LLM research and development. Whether we need more compute or better queries is a question that should define the next phase of LLM innovation. What's clear is that the way we engage with these models—how we frame our questions, how we interpret their responses—will be crucial in determining whether LLMs remain tools that reflect our current understanding or become catalysts for entirely new ways of thinking and reinventing the future.

In the end, Turing's Mirror doesn't just challenge us to reconsider the nature of LLMs—it challenges us to reconsider the nature of our own intelligence. As we refine our tools and our questions, we may find that the most profound insights come not from the LLM itself, but from how it reflects and accentuates our evolving understanding of the world.