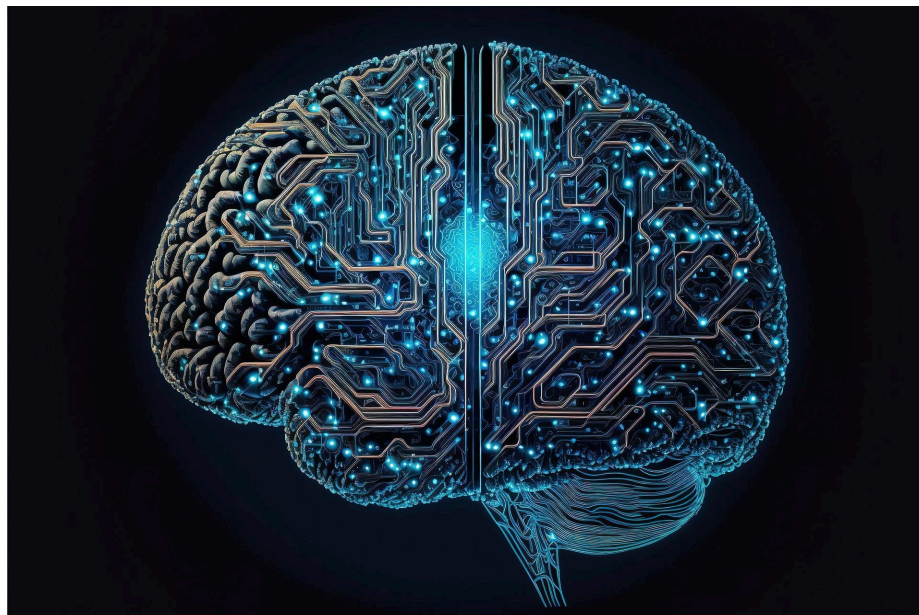


# Superficial Intelligence: An AI Odyssey with ChatGPT

by John M. Boyer, PhD, RFSPE, and Wanda Boyer, MC, PhD, RCC, ISPE Mentor



## 1. Introduction

Last summer, the ISPE book club selected Sam Kean's excellent book *The Disappearing Spoon: And Other True Tales of Madness, Love, and the History of the World from the Periodic Table of the Elements*. Early in the book, Kean comments that anxiety learning chemistry is understandable given that chemical notations can look “for all the world like computer error messages ([Xe]6s<sup>2</sup>4f<sup>1</sup>5d<sup>1</sup>)...”<sup>1</sup> Although Kean supplies many interesting facts and stories, we noticed (in August) that he didn't explain that parenthesized chemical formula. So, we performed a simple web search with the query, “*what do the parts of [Xe]6s24f15d1 mean?*” The top hit was a web page on a knowledge sharing site called *Quora*. The page's title was this longer question: “*Can somebody explain in details (sic) how Ce can have the electron configuration [Xe]6s24f15d1?*”<sup>2</sup> Quora had also just done an integration with ChatGPT via a button that, if pressed would give an answer from ChatGPT 3.5. Thus began an artificial intelligence odyssey.

## 2. The Medium Mangles the Meaning

As the section title implies, there wouldn't be much of a paper to read if it had gone well for ChatGPT. It didn't. The answer we got from Quora's ChatGPT did answer our question, but only because human intelligence enables us to ignore the parts of the answer that make it look dumber than a box of rocks. The mistakes were of such an elementary nature that we decided to take the conversation directly to ChatGPT on the OpenAI platform (again, the free version 3.5).

If you don't already know about electron energy levels and shells from Kean's book or a chemistry class you may have taken, ChatGPT's answer can (surprisingly) get you up to speed:

Certainly! Cerium (Ce) is an element with an atomic number of 58. The electron configuration of an element describes how its electrons are arranged in its atomic orbitals. The electron configuration of Ce can be broken down into its subshells:

- The first two electrons are in the 6s subshell, which can hold up to two electrons.
- The next fourteen electrons are in the 4f subshell, which can hold up to fourteen electrons.
- The fifteenth electron is in the 5d subshell, which can hold up to ten electrons. The 4f subshell is actually lower in energy than the 5d subshell, so electrons will fill the 4f subshell before the 5d subshell. This is why Ce has the electron configuration  $[Xe]6s24f15d1$ , rather than  $[Xe]6s24f14d2$ . The noble gas Xe represents the electron configuration of the previous element, xenon (atomic number 54). So, the  $[Xe]$  part of the electron configuration indicates that Ce's first 54 electrons have the same configuration as xenon. I hope that helps! Let me know if you have any further questions.

Superficially, this answer looooks *intelligent*, and it does convey many important facts about the notation. For example, it's clear that '6s2' means that the 6s subshell has 2 electrons and that '[Xe]' means that cerium shares xenon's configuration for 54 of its electrons. However, the answer also contains egregious errors that illustrate how little ChatGPT understands.

To start, consider the contradiction in the following two sentences extracted from bullets 2 and 3: "The next fourteen electrons are in the 4f shell... This is why Ce has the electron configuration  $[Xe]6s24f15d1$ ..." The first of these two sentences states that there are 14 electrons in the 4f shell, and then the second of these sentences presents a notation that includes 4f1, which indicates that only 1 electron is in the 4f shell.

Next, consider more of bullet 3: "The 4f subshell is actually lower in energy than the 5d subshell, so electrons will fill the 4f subshell before the 5d subshell. This is why Ce has the electron configuration  $[Xe]6s24f15d1$ , rather than  $[Xe]6s24f14d2$ ." They say the best lies contain some truths. Well, the 4f shell does have lower energy than 5d, but this has nothing to do with why  $[Xe]6s24f14d2$  isn't the notation for cerium. The '[Xe]' component denotes that much of cerium's electron configuration is the same as xenon (Xe), which has 54 electrons. In the notation  $[Xe]6s24f14d2$ , the components 6s2, 4f1, and 4d2 indicate a total of 5 additional electrons (2 in the s-shell of the 6th energy level, 1 in 4f, and 2 in 4d). So, we humans know  $[Xe]6s24f14d2$  does not describe cerium (Ce) because the notation clearly contains 2+1+2=5 more electrons than the 54 from xenon, so we used a highly specialized mathematical skill, known formally as addition to get a total of 59. This is one more than the 58 for cerium, which ChatGPT itself stated is "an element with an atomic number of 58."

We humans also have the highly specialized verbal skill of semantic processing, that is to say, we know what words mean. Hence, we can tell that the cited principle that "the 4f subshell is actually lower in energy than the 5d subshell" does not explain why  $[Xe]6s24f15d1$  is the correct notation for cerium because, as ChatGPT also explained in the same sentence, "electrons will fill the 4f subshell before the 5d subshell." Hence, the words ChatGPT used mean that 4f subshell having a lower energy level would explain why the notation should end with 4f2, *not* 4f15d1.

Why are those last two electrons spread across 4f and 5d when 4f has a lower energy level? That seemed like a good follow-up question for ChatGPT.

### 3. The Medium Fabricates a Falsehood

Before going to that question, though, we noted that ChatGPT's answer suggested a total of 71 electrons, rather than 58 (the 54 arranged as in xenon, 2 in the 6s shell, 14 in 4f, and then a last 1 in 5d). We guessed that ChatGPT was first misreading the '5' of 5d1 as also part of the number after 'f' in 4f1, which would give 15 electrons in 4f. Then, since an f subshell can only hold up to 14 electrons, perhaps it was changing the 15 to 14. After we interrogated ChatGPT further about the number of electrons, it changed its answer to one containing the following excerpt:

*Certainly, let's break down the electron configuration of cerium ( $[Xe]6s24f15d1$ ) in more detail, step by step.*

...

***4f15:** Following the 6s electrons, we move to the 4f subshell. The "4f15" indicates that there are 15 electrons in the 4f orbitals. The f orbitals can accommodate up to 14 electrons (with each orbital holding 2 electrons with opposite spins), so the presence of 15 electrons in the 4f orbitals suggests that one of the f orbitals is hosting an extra electron due to the unique electron filling patterns of some transition metals.*

***5d1:** Finally, we place the last electron in the 5d orbital. The "5d1" indicates that there's one electron in the 5d orbital. The d orbitals can hold up to 10 electrons (with each orbital holding 2 electrons with opposite spins), so having only one electron in the 5d orbital is consistent with the electron arrangement in cerium.*

*... This electron configuration is specific to the properties of cerium and is derived from the arrangement of its 58 electrons within its atomic structure.*

Rather than using the fact it knows that f shells can only have up to 14 electrons, ChatGPT fabricated a falsehood for why 4f should have 15 in this case: "due to the unique electron filling patterns of some transition metals." Except for those who do know that f subshells cannot ever contain more than 14 electrons, the explanation sounds believable, like an authority informing you that sometimes there could be 15 electrons in an f shell. And while this false explanation is a bit of a hand-wave, the explanation for 5d1 seems worse. Superficially, it also sounds authoritative, but presenting an explanation as a reason is misleading if the explanation is self-referential, such as cerium having "only one electron in the 5d orbital is consistent with the electron arrangement in cerium."

### 4. The Medium Knows Not That It Knows Not

After still more questions, ChatGPT politely agreed that the 4f shell contained neither 14 nor 15 electrons after all. Instead, it concluded that we (and Kean) gave an incorrect notation because it

violated “*the Aufbau principle, which states that electrons fill the lowest energy orbitals first before moving to higher energy orbitals.*” ChatGPT continued by informing us that the electron in 5d should instead be in 4f, i.e., that the notation should end with 4f2 rather than 4f15d1. We had to insist that the notation was correct and that we were asking *why* the notation ended with 4f15d1 rather than 4f2. In other words, why did cerium seem to violate the Aufbau principle? After one more round of ChatGPT putting 14 electrons in 4f again, it apologized profusely, agreed with how absolutely correct we (and Kean) were, and reworked its answer:

*In the case of cerium, it's more stable to have the configuration [Xe]6s<sup>2</sup>4f<sup>15</sup>d<sup>1</sup> rather than [Xe]6s<sup>2</sup>4f<sup>14</sup>d<sup>0</sup>, which you might expect based on the periodic table. This is due to the complex electronic structure and the subtle interplay of electron-electron interactions.*

So, ChatGPT agreed with the correct notation only by being spoon fed the answer insistently by users who already knew the answer. Even then, its answer isn't incorrect, but it is only barely correct in its ambiguity, referring to hand-wavy reasons like complex structures and subtle interplays. An answer that would have made ChatGPT appear as learned as it is loquacious would be: “The 4f, 5d, and 6s energy levels are very close to each other, and the transfer of one electron to the 5d shell is due to strong interelectronic repulsion in the compact 4f shell.”<sup>3</sup> It is mystifying that ChatGPT didn't not produce this information in its 5d1 explanation considering that this answer appears directly in the Wikipedia page for cerium.

## 5. Conclusion

Insofar as ChatGPT was recently reported to have a very high verbal IQ,<sup>4</sup> one simply must wonder how there could possibly be such simple errors in its output. One problem is that IQ tests measure generalized intelligence with only a small sample of questions, and the entity possessing the generalized intelligence is expected to fill in the huge conceptual gaps between those questions. In other words, the assumption of ‘if you can do this, you can do that’ may be typically true for human intelligence, but it is not so for artificial intelligence technologies like ChatGPT. One reason for this difference is that large language model (LLM) technologies like ChatGPT are not so much designed to give factually correct answers *as it is to give answers that seem consistent* with human language usage.<sup>5</sup> Hence, we must be very careful about using the answers we get from LLMs like ChatGPT 3.5 because they are subject to a phenomenon called AI hallucination, which is when generated text seems authoritative but is not grounded in facts.

AI hallucination is an active area of research in computer science, and a number of techniques are under development to help mitigate AI hallucinations in LLM output.<sup>6</sup> One such technique is called retrieval augmented generation (RAG), which combines the initial LLM generated text answers with similar text that the RAG method searches for either in its training data or in external sources, so that the final outputs are more grounded.<sup>7</sup>

Recently, ChatGPT-4 equipped with RAG was released more freely in Microsoft®'s Copilot for Windows®. It has three conversational modes ranging from more creative to more precise. Since we wanted factually correct information, we selected the more precise style. Its responses contained longer snippets of text that were directly attributed to several web-based sources of chemistry information. Seemingly due to needing less needle-and-thread work on fewer, longer

snippets to produce the final output text, we found it impossible to reproduce the AI hallucinations for that chemical formula that we produced with ChatGPT 3.5. As we write, ChatGPT-4o (“4 Omni”) has just been released with limitations for non-paying users of the OpenAI platform. It is also equipped with RAG, and it also made no mistakes answering the chemical formula question.

The most important takeaway here is that one should only use an LLM technology that is equipped with RAG because an LLM on its own will tend to make up the successive words in its output based on consistency with human language usage rather than consistency with facts expressed in its training data or in more up-to-date information sources. Interestingly, this means that the more efficacious results from the GPT-4 class technologies are not so much representative of better artificial intelligence as they are just *better search tools*, with the LLM helping to generate better search queries so that RAG can find better content. Citable content. Generated by human intelligence. To expect any greater level of human intelligence from today’s LLM technologies would be unwise.<sup>8</sup>

## NOTES

1. Sam Kean, *The Disappearing Spoon: And Other True Tales of Madness, Love, and the History of the World from the Periodic Table of the Elements*, (Boston, MA: Little, Brown and Company, 2010).
2. Quora chatbot, “Can somebody explain in details how Ce can have the electron configuration [Xe]6s24f15d1,” *Quora*, (June 27, 2023), <https://www.quora.com/Can-somebody-explain-in-details-how-Ce-can-have-the-electron-configuration-Xe-6s24f15d1>
3. Wikipedia contributors, “Cerium”, *Wikipedia, The Free Encyclopedia*, <https://en.wikipedia.org/wiki/Cerium>
4. Eka Roivainen, “AI’s IQ,” *Scientific American* 329, no. 1 (July/August 2023): 7. <https://www.scientificamerican.com/article/i-gave-chatgpt-an-iq-test-heres-what-i-discovered/>
5. E. V. Levine, “Cargo cult AI,” *Communications of the ACM* 66, no. 9 (Sep. 2023): 46-51. <https://doi.org/10.1145/3606945>
6. S. M. Towhidul Islam Tonmoy, S. M. Mehedi Zaman, J. Vinija, A. Rani, V. Rawte, A. Chadha, and A. Das, “A comprehensive survey of hallucination mitigation techniques in large language models,” *arXiv preprint* (Jan. 2024): 1-19. <https://doi.org/10.48550/arXiv.2401.01313>
7. Ibid.
8. John M. Boyer and Wanda Boyer, “A Strategic Roadmap for Mitigating Generative Artificial Intelligence Hallucinations,” *Cureus Journal of Computer Science* 1 (2024), <https://doi.org/10.7759/s44389-024-00364-7>