Why do LLMs exhibit scaling laws

by Charles H. Martin, 9/5/25

𝗪𝗵𝘆 𝗱𝗼 𝗟𝗟𝗠𝘀 𝗲𝘅𝗵𝗶𝗯𝗶𝘁 𝘀𝗰𝗮𝗹𝗶𝗻𝗴 𝗹𝗮𝘄𝘀 ? I saw an interesting talk the other day using a little random matrix theory to provide an analogy as to why LLMs and NNs in general exhibit scaling laws. The paper addressed the question, why do we observe heavy-tailed power-law scaling when the current theory says the scaling is exponentially decayed? Here's the thing...

This kind of analysis was done in the late 90s, in theoretical physics literature, in the context of applied Renormalization Group (RG) theory. And the thing is, there are different mechanisms for the appearance of heavy-tailed power-law phenomena, like what we see with neural scaling laws.

In particular, see the book:

"𝘾𝙧𝙞𝙩𝙞𝙘𝙖𝙡 𝙋𝙝𝙚𝙣𝙤𝙢𝙚𝙣𝙖 𝙞𝙣 𝙉𝙖𝙩𝙪𝙧𝙖𝙡 𝙎𝙘𝙞𝙚𝙣𝙘𝙚𝙨 𝘾𝙝𝙖𝙤𝙨, 𝙁𝙧𝙖𝙘𝙩𝙖𝙡𝙨, 𝙎𝙚𝙡𝙛𝙤𝙧𝙜𝙖𝙣𝙞𝙯𝙖𝙩𝙞𝙤𝙣 𝙖𝙣𝙙 𝘿𝙞𝙨𝙤𝙧𝙙𝙚𝙧: 𝘾𝙤𝙣𝙘𝙚𝙥𝙩𝙨 𝙖𝙣𝙙 𝙏𝙤𝙤𝙡𝙨" by Didier Sornette ([2])

This is a great book if you want to understand more about heavy-tailed statistics, power-law phenomena, and RG theory. In fact, there is a whole chapter on "Mechanisms for Power Laws ."

Coming back to neural scaling laws, there are mechanisms based on multiplicative noise (like the paper talk I saw), and mechanisms based on strong correlations (i.e., RG theory). And it's not easy to distinguish one from the other.

(In fact, this is exactly why I created the SETOL approach, namely, to show that the power law phenomena we see in NNs is, in fact, associated with the Renormalization Group: [3], [4])

Sornette emphasizes this because many natural systems (earthquakes, financial crashes, turbulence, etc.) operate in finite regimes where the RG “hasn’t converged.” So the distributions we measure have fat tails—apparent Lévy-like or power-law behavior—even though the asymptotic theory would wash them away.

Now this is exactly what we see with neural scaling laws. The asymptotic theory says they should decay rapidly, and, yet, in practice, we observe power-law behavior.

So, it turns out, neural scaling laws are not just curve fitting— they are fingerprints of deep physical principles hiding inside our neural networks.

References

[1] [Why do LLMs exhibit scaling laws, Linkedin post, Charles H. Martin, 9/5/2025](https://www.linkedin.com/posts/charlesmartin14_%F0%9D%97%AA%F0%9D%97%B5%F0%9D%98%86-%F0%9D%97%B1%F0%9D%97%BC-%F0%9D%97%9F%F0%9D%97%9F%F0%9D%97%A0%F0%9D%98%80-%F0%9D%97%B2%F0%9D%98%85%F0%9D%97%B5%F0%9D%97%B6%F0%9D%97%AF%F0%9D%97%B6%F0%9D%98%81-%F0%9D%98%80%F0%9D%97%B0%F0%9D%97%AE%F0%9D%97%B9%F0%9D%97%B6%F0%9D%97%BB%F0%9D%97%B4-activity-7369379143718797316-oz7Q?utm_source=share&utm_medium=member_desktop&rcm=ACoAAAFZfUoBgPoGUucdnvtwuzPv79P8VHj6uvk)

[2] [Critical Phenomena in Natural Sciences - Chaos, Fractals, Self-Organization, and Disorder: Concepts and Tools, Didier Sornette, 2006](https://github.com/dimitarpg13/dynamical_systems_and_ergodicity/blob/main/literature/books/CriticalPhenomenainNaturalSciences-ChaosFractalsSelf-OrganizationandDisorder-ConceptsandTools-2ndEdition-2006.pdf)

[3] [SETOL: A Semi-Empirical Theory of (Deep) Learning, Charles H Martin, Christopher Hinrichs, 2025](https://github.com/dimitarpg13/deep_learning_and_neural_networks/blob/main/literature/articles/SETOL/SETOL-A_Semi-Empirical_Theory_of_Deep_Learning_Martin_2025.pdf)

[4] [Introduction to Renormalization, JF Melo, U. of Cambridge, 2019](https://github.com/dimitarpg13/dynamical_systems_and_ergodicity/blob/main/literature/articles/renormalization_group/Introduction_to_Renormalisation_Melo_2019.pdf)

[5] [Introduction to Renormalization Group Method, Roland Bauerschmidt, David C. Brydges, and Gordon Slade, 2019](https://github.com/dimitarpg13/dynamical_systems_and_ergodicity/blob/main/literature/articles/renormalization_group/bauerschmidt_brydges_slade_book.pdf)