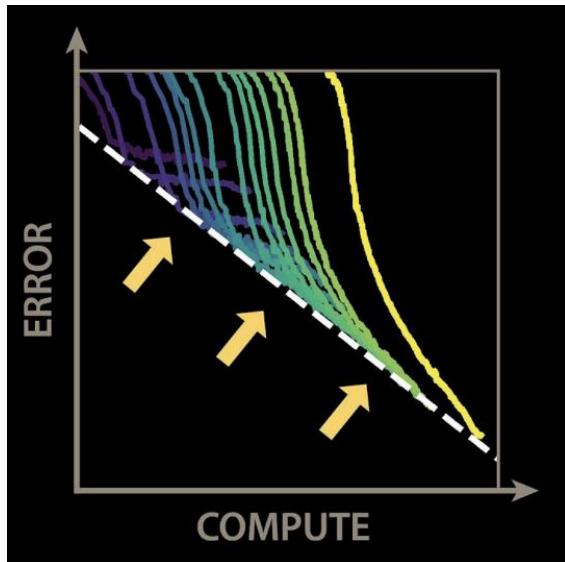


AI Tidak Dapat Melewati Garis Ini: Memahami Hukum *Scaling* dalam Model Bahasa Neural



Grafik yang menunjukkan bahwa dengan peningkatan komputasi, error menurun, namun terjadi penurunan manfaat (*diminishing returns*) seiring bertambahnya komputasi. (Sumber: [Welch Labs](#))

Dalam beberapa tahun terakhir, perkembangan dalam model bahasa neural (*neural language models*) telah mengubah cara kita memahami bahasa alami. Dari mesin pencari hingga *chatbot* yang cerdas, model bahasa kini memainkan peran sentral dalam berbagai aplikasi kecerdasan buatan. Salah satu aspek penting yang mendasari perkembangan ini adalah konsep *scaling laws* atau hukum skala. Beberapa karya penelitian yang relevan dalam memahami dinamika ini adalah makalah oleh Kaplan et al. (2020) berjudul “*Scaling Laws for Neural Language Models*” dan karya oleh Henighan et al. (2020) berjudul “*Scaling Laws for Autoregressive Generative Modeling*”. Kedua makalah ini mengungkapkan hukum scaling (atau hukum skala) yang mengatur bagaimana performa model bahasa neural meningkat dengan bertambahnya sumber daya, serta batas-batas yang mungkin menghalangi AI mencapai tingkat performa yang diharapkan.

Apa Itu Hukum Skala?

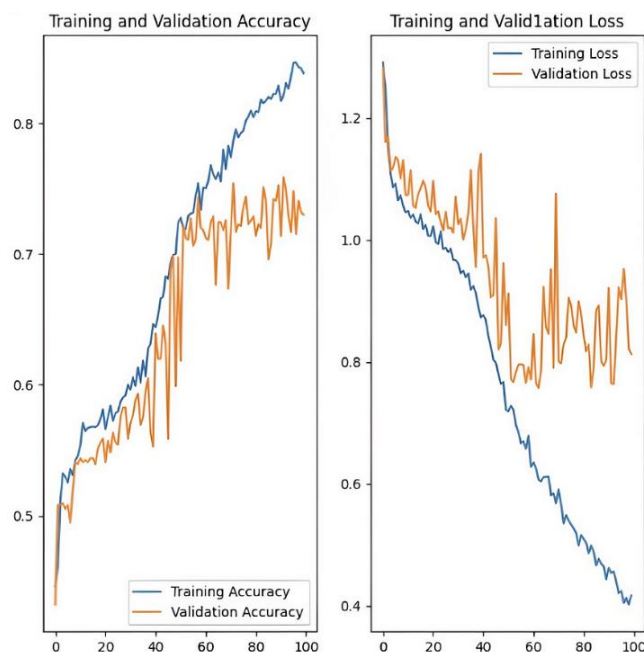
Dalam makalah “*Scaling Laws for Neural Language Models*” (Kaplan et al., 2020), ditemukan bahwa kinerja model bahasa meningkat secara konsisten seiring dengan bertambahnya ukuran parameter model, jumlah data pelatihan, dan kekuatan komputasi yang

digunakan. Dalam studi ini, peneliti mengeksplorasi hubungan antara parameter-parameter tersebut, dan menemukan bahwa performa model mengikuti *power-law* (hukum daya), di mana peningkatan signifikan dalam kinerja bisa diperoleh dengan memperbesar model.

Menurut Kaplan et al. (2020), ada tiga faktor utama yang harus ditingkatkan untuk memperoleh performa yang lebih baik:

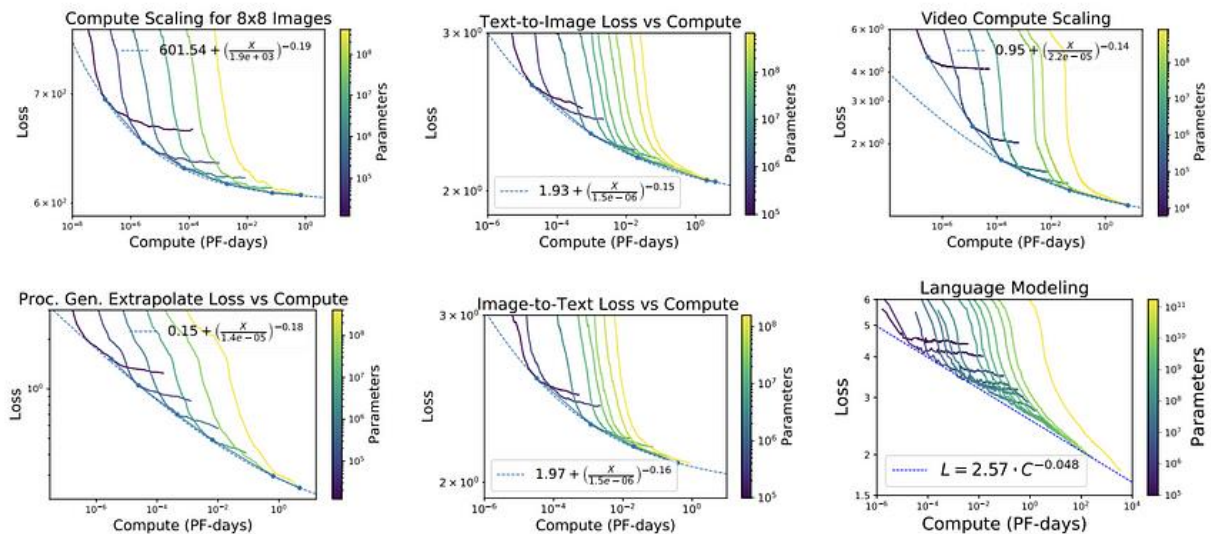
1. **Ukuran Model (Parameter):** Model yang lebih besar cenderung mampu menangkap lebih banyak pola dalam data.
2. **Ukuran Dataset:** Semakin besar data yang digunakan, semakin baik model dalam generalisasi dan menangani beragam konteks.
3. **Sumber Daya Komputasi:** Latihan model yang besar membutuhkan komputasi yang besar, dan ada hubungan yang mendasar antara jumlah komputasi yang digunakan dan hasil kinerja model.

Kaplan dan timnya menemukan bahwa untuk mencapai performa optimal, perlu ada keseimbangan di antara ketiga faktor ini. Jika salah satu faktor ditingkatkan secara berlebihan tanpa mengimbangi yang lain, hasil yang diperoleh akan kurang optimal. Contohnya adalah *overfitting*, yang terjadi ketika model *machine learning* memberikan prediksi akurat untuk data pelatihan, tetapi tidak untuk data baru.



Grafik kiri menunjukkan peningkatan akurasi model selama pelatihan, sementara grafik kanan menunjukkan penurunan loss. Perbedaan signifikan antara training dan validation menunjukkan terjadinya *overfitting*. (Sumber: dokumen pribadi)

Batasan yang Tidak Dapat Dilintasi oleh AI



Grafik-grafik ini menggambarkan bagaimana loss berkurang seiring peningkatan komputasi dan parameter model dalam tugas yang berbeda, dengan pola yang konsisten dan dapat diprediksi. (Sumber: Kaplan et al. (2020))

Meskipun hukum scaling memberikan wawasan yang optimistis, Henighan et al. (2020) dalam makalah mereka “*Scaling Laws for Autoregressive Generative Modeling*” mengemukakan bahwa ada batas-batas praktis dan teoritis yang menghambat performa AI, terutama dalam skenario penggunaan nyata. Ada beberapa garis yang, menurut mereka, sulit dilewati oleh AI tanpa terobosan baru dalam teknologi:

- Batasan Energi dan Sumber Daya:** Kebutuhan komputasi yang meningkat secara eksponensial untuk setiap peningkatan kecil dalam performa menjadi kendala besar. Dibutuhkan energi yang luar biasa besar dan biaya yang tinggi untuk menjalankan model bahasa neural besar seperti GPT-3 atau GPT-4. Akibatnya, ini menjadi kendala utama dalam mengembangkan model yang lebih cerdas dan lebih besar.
- Keterbatasan Data Berkualitas:** Salah satu masalah yang dihadapi dalam pengembangan AI adalah ketersediaan data berkualitas. Model memerlukan sejumlah besar data pelatihan yang representatif, dan pada akhirnya, AI mungkin mencapai batas di mana tidak ada lagi data baru yang dapat dimanfaatkan untuk meningkatkan kinerja.
- Generalisasi dan Ketergantungan pada Pembelajaran Terbimbing:** AI saat ini sangat bergantung pada data yang diberikan manusia, dan masih sulit bagi mereka untuk benar-benar generalisasi melampaui data yang telah mereka pelajari.

Kemampuan untuk belajar seperti manusia yang secara mandiri dan dari sedikit contoh masih menjadi tantangan besar yang belum teratasi.

Bahri et al. (2021) dalam makalah mereka “*Explaining Neural Scaling Laws*” juga mengemukakan pentingnya mempertimbangkan proses dinamis dalam pembelajaran AI. Mereka menemukan bahwa selain dari ukuran model dan data, waktu pelatihan juga memainkan peran penting dalam bagaimana kinerja model meningkat. Penelitian mereka menunjukkan bahwa penambahan waktu pelatihan secara signifikan memperbaiki hasil yang dicapai oleh model-model besar.

Garis yang Tidak Bisa Dilewati

Seiring dengan perkembangan model bahasa yang semakin besar dan kuat, komunitas riset AI mulai mempertanyakan sejauh mana model-model ini bisa terus meningkat. Kaplan et al. (2020) menyebutkan bahwa meskipun skala parameter dan data dapat memberikan manfaat yang signifikan, ada batas praktis yang tidak bisa diabaikan. Model dengan ukuran tertentu akan membutuhkan peningkatan komputasi yang eksponensial untuk setiap peningkatan kecil dalam performa. Ini menunjukkan bahwa meskipun AI bisa terus berkembang, ada titik di mana biaya komputasi dan sumber daya yang dibutuhkan menjadi tidak praktis.

Henighan et al. (2020) juga menyebutkan bahwa selain kendala fisik dan finansial, AI mungkin menghadapi “garis yang tidak bisa dilalui” dalam hal kemampuan kognitif. Model bahasa yang ada saat ini tidak bisa memahami dunia dengan cara yang sama seperti manusia. AI mungkin mampu menghasilkan teks yang tampak seperti hasil pemikiran manusia, tetapi pemahaman sejati tentang konteks, niat, atau makna sering kali masih jauh dari sempurna.

Kesimpulan

Penelitian tentang hukum scaling dalam model bahasa neural menunjukkan bahwa seiring dengan meningkatnya ukuran model, data, dan sumber daya komputasi, performa AI terus meningkat. Namun, ada batas-batas besar yang mungkin akan sulit untuk dilewati oleh AI dalam jangka panjang. Biaya sumber daya yang sangat besar, keterbatasan data, dan ketidakmampuan untuk benar-benar generalisasi seperti manusia menjadi penghambat utama. Untuk AI mencapai terobosan yang lebih besar, dibutuhkan solusi inovatif yang dapat melampaui “garis” ini, melainkan juga mengubah cara AI belajar dan berinteraksi dengan dunia.

Referensi:

- Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., ... & Amodei, D. (2020). *Scaling Laws for Neural Language Models*. [arXiv:2001.08361](https://arxiv.org/abs/2001.08361).
- Henighan, T., Kaplan, J., Katz, M., Chen, M., Hesse, C., Jackson, K., ... & McCandlish, S. (2020). *Scaling Laws for Autoregressive Generative Modeling*. [arXiv:2010.14701](https://arxiv.org/abs/2010.14701).
- Bahri, Y., Dyer, E., Kaplan, J., Lee, J., Sharma, U. (2021). *Explaining Neural Scaling Laws*. [arXiv:2102.06701](https://arxiv.org/abs/2102.06701).