

Densing Law of LLMs

Chaojun Xiao¹, Jie Cai², Weilin Zhao¹, Guoyang Zeng², Biyuan Lin², Jie Zhou², Zhi Zheng²
 Xu Han¹, Zhiyuan Liu¹, Maosong Sun¹
¹Tsinghua University ²ModelBest Inc.
 xiaocj20@mails.tsinghua.edu.cn
 {han-xu, liuzy, sms}@tsinghua.edu.cn

Highlights

We introduce the concept of “capability density” to evaluate the training quality of large language models (LLMs) and describe the trend of LLMs that considers both effectiveness and efficiency.

(Relative) Capability Density. For a given LLM \mathcal{M} , its capability density is defined as the ratio of its **effective parameter size** to its actual parameter size, where the effective parameter size is the minimum number of parameters required for the reference model to achieve performance equivalent to \mathcal{M} .

We reveal an empirical law for the capability density of **open-source base LLMs** released since 2023.

Densing Law. The maximum capability density of LLMs exhibits an exponential growth trend over time.

$$\ln(\rho_{\max}) = At + B$$

Here, ρ_{\max} is the maximum capability density of LLMs at time t .

Figure 1 presents the capability density of popular LLMs, measured by their performance on 5 widely-used benchmarks. A trend is fitted between maximum capability density and release date, revealing that $A \approx 0.007$ with $R^2 \approx 0.93$. This indicates **the maximum capability density of LLMs doubles approximately every 3.3 months**¹. That means, around three months, it is possible to achieve performance comparable to current state-of-the-art LLMs using a model with half the parameter size.

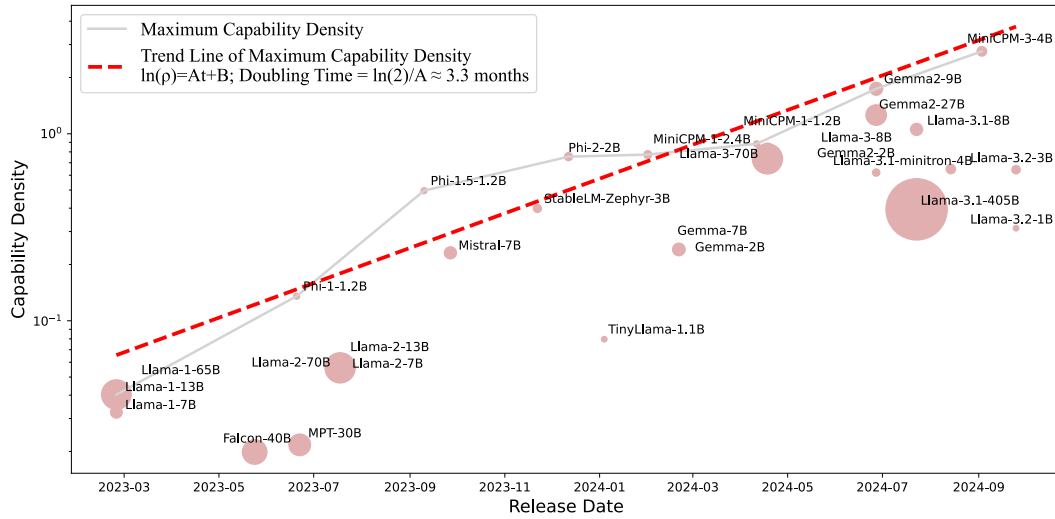


Figure 1: The estimated capability density of open-source base LLMs.

¹The capability density growth rate is affected by specific evaluation benchmarks and reference models.