Are Emergent Abilities of Large Language Models a Mirage?

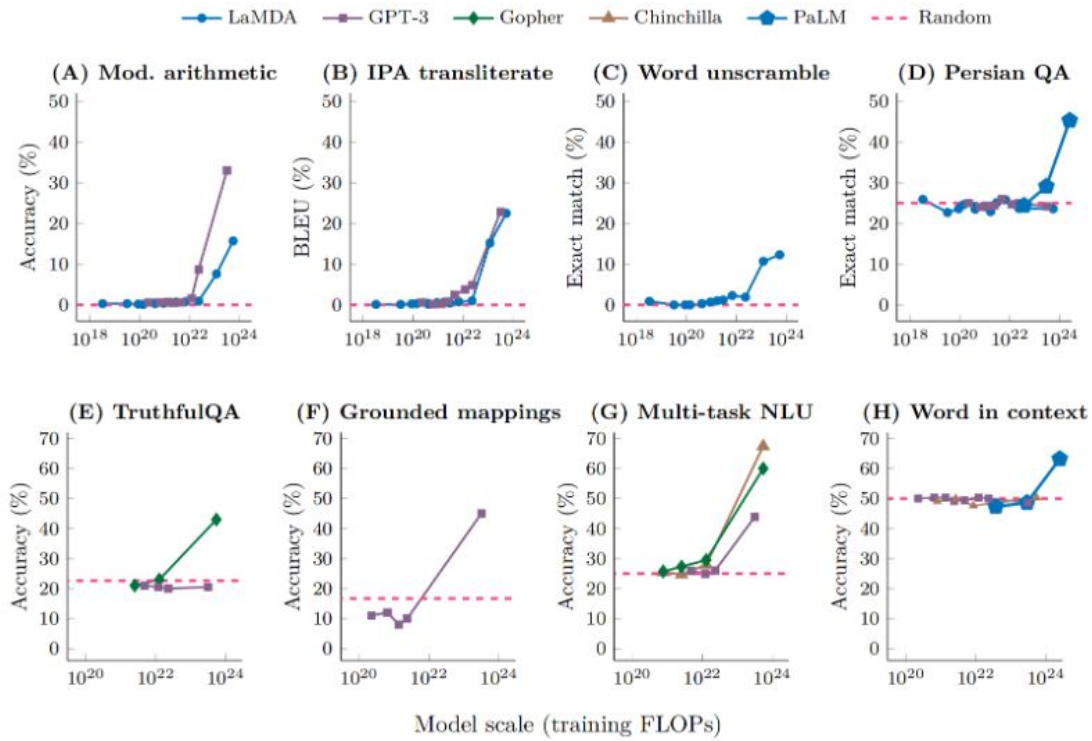
Сипаров Иван

Эмерджентные способности LLM

При увеличении масштаба модели, мы ожидаем, что ее производительность будет предсказуемо увеличиваться.

Но в некоторых задачах модель демонстрируют резкое увеличение производительности.

Пример задач, в которых модель демонстрирует резкое увеличение производительности.



Свойства эмерджентных способностей в LLM

- Резкость, мгновенный переход от отсутствия к присутствию
- Непредсказуемость, переход на кажущихся непредсказуемыми модельных масштабах

Альтернативное объяснение

- Выбором исследователем метрики измерения, которая является нелинейной или прерывистой.

- Наличия слишком малого количества тестовых данных для точной оценки производительности моделей меньшего размера

Математическая модель

Предположим, что у нас есть семейство моделей, потери кросс-энтропии которых уменьшаются по степенному закону,

$$\mathcal{L}_{CE}(N) = \left(\frac{N}{c}\right)^{\alpha}$$

где N - число параметров.

Математическая модель

Функция потерь имеет следующий вид:

$$\mathcal{L}_{CE}(N) \stackrel{\text{def}}{=} -\sum_{v \in V} p(v) \log \hat{p}_N(v)$$

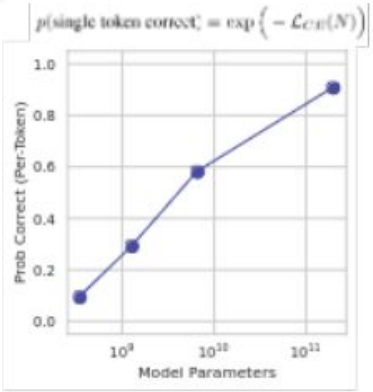
Но на практике, мы не знаем, распределение p, поэтому мы перепишем в следующем виде:

$$\mathcal{L}_{CE}(N) = -\log \hat{p}_N(v^*)$$

Математическая модель

Такая модель имеет вероятность выбора правильного токена

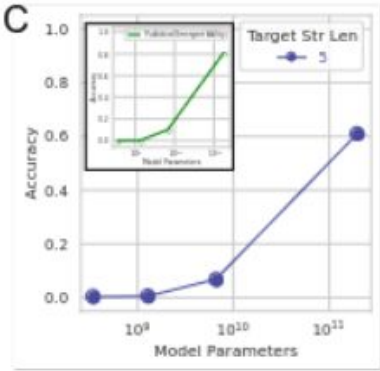
$$p(\text{single token correct}) = \exp\Big(-\mathcal{L}_{CE}(N)\Big) = \exp\Big(-(N/c)^{\alpha}\Big)$$



Математическая модель

Теперь предположим, что исследователь заменил метрику, на ту, которая требуют правильно выбрать L токенов, тогда

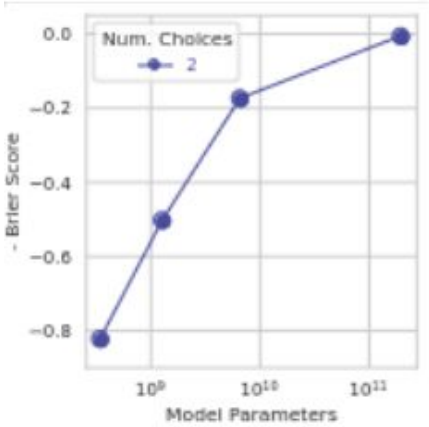
$$\text{Accuracy}(N) \approx p_N(\text{single token correct})^{\text{num. of tokens}} = \exp\left(-(N/c)^{\alpha}\right)^{L}$$



Математическая модель

Допустим, что мы теперь изменили метрику на псевдолинейную, тогда

$$\text{Token Edit Distance}(N) \approx L\left(1 - p_N(\text{single token correct})\right) = L\left(1 - \exp\left(-(N/c)^{\alpha}\right)\right)$$

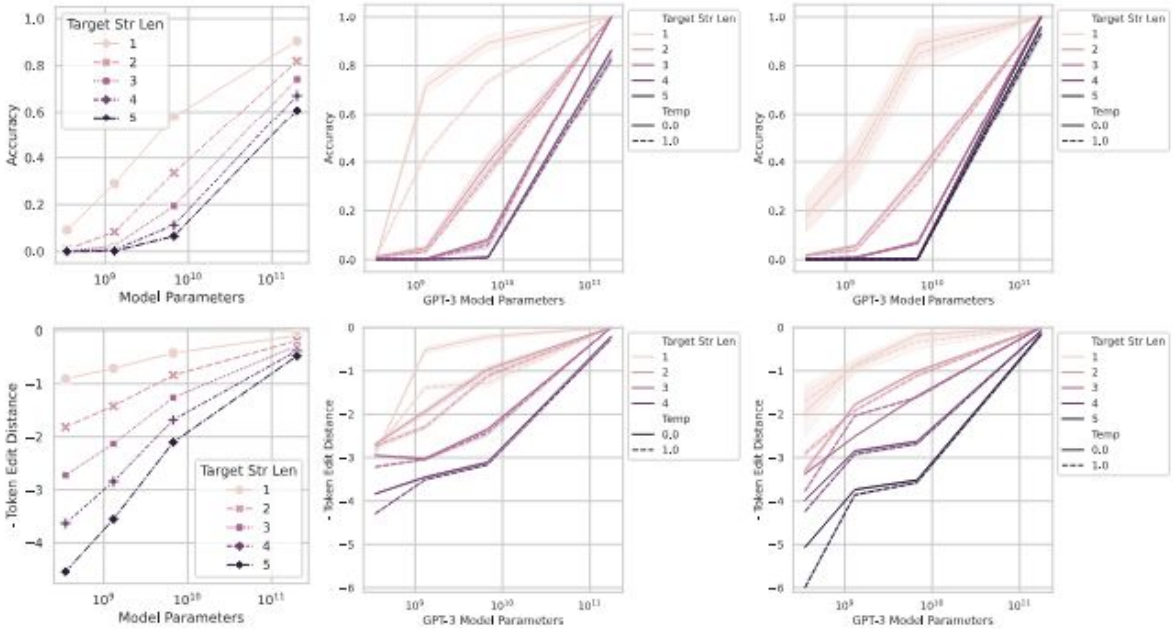


Эмерджентность семейство моделей GPT

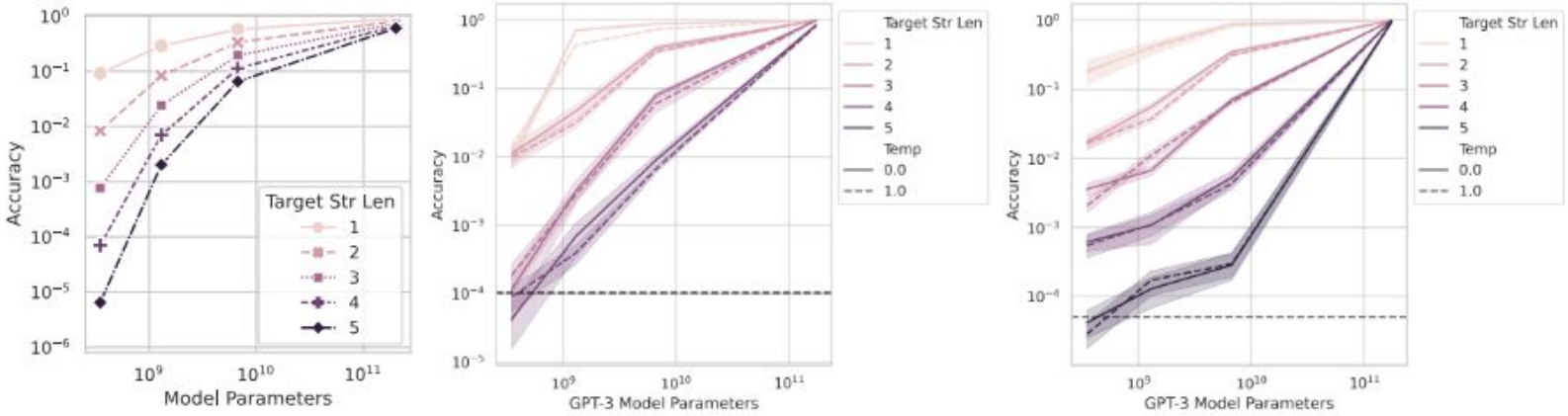
Проверим два альтернативных подходах, объясняющей эмерджентность

1. При замене метрики на линейную или непрерывную производительность модели будет плавной и предсказуемой.
2. При увеличении тестовых данных, для нелинейной метрики, мы будем наблюдать плавное и предсказуемое увеличение производительности модели.

Изменение производительности модели при замени метрики



Изменение производительности при увеличение тестовой выборки

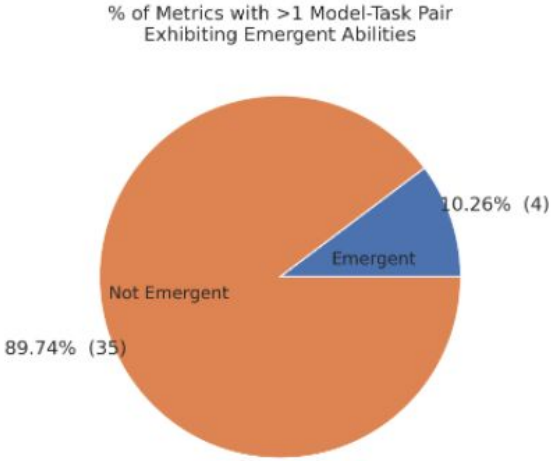


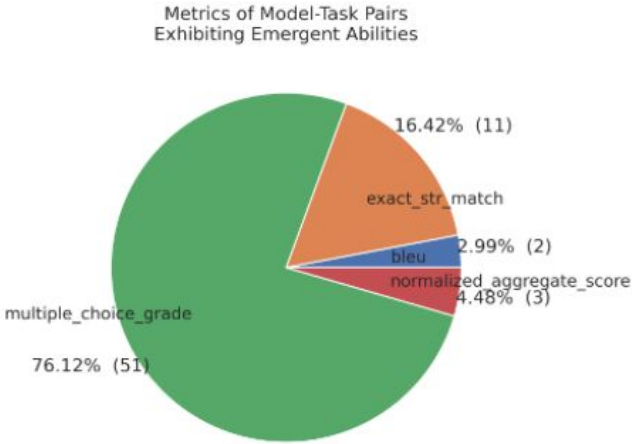
Метаанализ

Проверим два альтернативных подхода:

1. Эмерджентные способности появляются в конкретных метриках, а не в задачах.
2. При изменение метрики эмерджентные способности исчезают.

Метрики в которых проявляется эмерджентность





Изменение метрики в семействе LaMDA