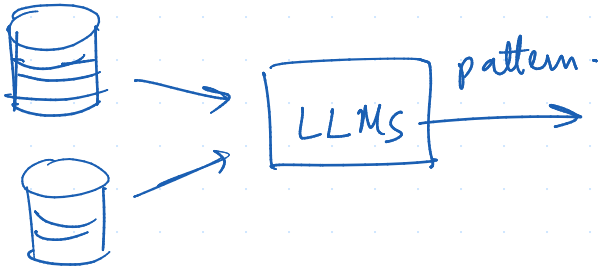
Lecture - 12

11/04/26.

Transformers. ⇔ LLMs.



$573281 \times 3567812 \neq \ldots$

LLMs are reasoning ?

is this true ?
when it is not able to do long simple calculations ?

Addition. ~ 30 examples.

$\left\{ \begin{array}{l} \begin{array}{r} 1 \\ 2 \\ \hline 3 \end{array} \qquad \begin{array}{r} 5 \\ 7 \\ \hline (1)2 \end{array} \\ \\ \begin{array}{r} 101 \\ 234 \\ \hline 335 \end{array} \qquad \begin{array}{r} {\scriptstyle 1} \\ 55551 \\ 219 \\ \hline 55770 \end{array} \end{array} \right\}$ examples.

Addition - Algorithm.

human

Given enough time

- n-length addition → do correctly
- If not → verify whether a addition is correct or not.

$253 + 756 + 395 = \cdots$

$(1011 + 723) + (527 + 635) + (931) =$

Similar algorithm → recursion

inherent recursion

$536712 + 3216 = \cdots$

$22 + 53 + 32 + 57 + 62 = \cdots$

Addition → Algorithm (X).

If a model is actually reasoning, it should perform well on 'logical (arithmetic) tasks'.

↳ which are easy for humans.

~50.

LLMs don't learn the same way as humans → (as of now)

They are good data compressors + Data (extrapolators)
+ interpolators.

(5)

Interpolation (1-3 digit)

= Max-5-digits

training.

3562
+ 253

{X, XX, XXX, XXXX, XXXXX}

XXXXX
+ YYYYY
ZZZZZZ

{Y, YY, YYY, YYYY, YYYYY}.

~200K samples.

X+Y, XX+Y, XXXXX + YYYYY.

90-10 train-val.

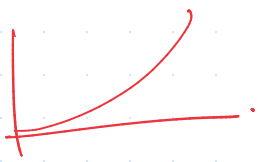
X+Y → (9x9) → 100

10.

2+1
2+2
9.

1+1
1+2
1+3
⋮
1+9
9.

X+YY → 9x99 → 1000



Interpolation → in dist → 1-3 digit.

✓

{x, xx, xxx} {y, yy, yyy} → 2000K.

✓ { Test data should not leak to the train dataset → (ensure) }

✓ {x, xx, xxx, xxxx, xxxxx} Separate

✓ {y, yy, yyy, yyyy, yyyyy} test set.

{xxxx xx} {yyy yyy} → extrapolation.

6 6.

7-digit. 7-digit

{xxxxxxxx} {yyyyyyyy}