

The Exponential Evolution of NLP Technology

Model Size:

1K

1M

1B

1T

10,000x

Improvement
in 30 years!

Performance
GPT-3

Claude-3

Relative Performance (log scale)

10^4

10^3

10^2

10^1

10^0

10^{-1}

1990

1995

2000

2005

2010

2015

2020

2025

Year

N-grams

Neural LM

Semantic Understanding
Unlocked

Word2Vec

LSTM

Attention Revolution
Begins

Transformer

BERT

GPT-2

GPT-3

Few-shot Learning
Emerges

Statistical Era

Early Neural Era

Embedding Era

Transformer Era

Performance
GPT-3

Claude-3

10,000x
Improvement
in 30 years!