Correlation Dimension of Autoregressive LLMs

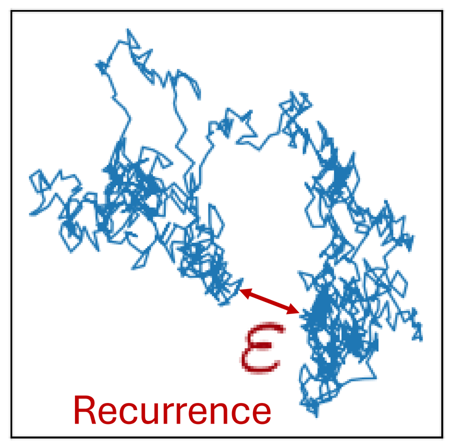
A Fractal-Geometric Metric for LLM Generalization, Degeneration, Hallucination

Xin Du, Kumiko Tanaka-Ishii
Waseda University

Correlation Dimension *d*

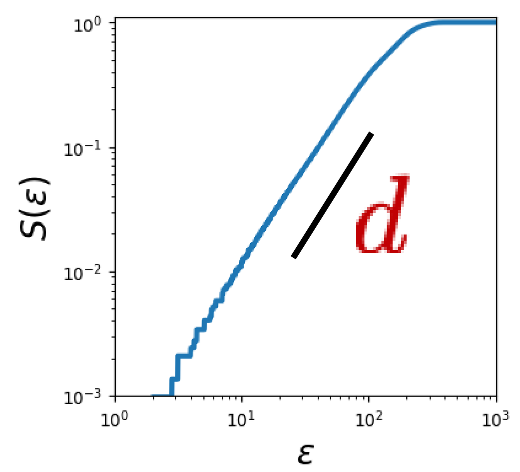
Vector sequence $x_1, x_2, \cdots, x_t \in \mathbb{R}^D$

Correlation Integral $S(\varepsilon) = \lim_{t\to\infty} \frac{2}{t(t-1)} \sum_{1\le i<j\le t} 1\{\|x_i - x_j\| < \varepsilon\}$



$S(\varepsilon) \propto \varepsilon^{d}$

\# recurrences — Distance threshold



Correlation Dimension *d*:

- A fractal dimension from trajectory of a dynamical system (Grassberger & Procaccia, 1983)
- Computationally tractable even in very high-dimensional spaces.

Measuring CorDim of LLMs

$x_t(\omega) = \log P_\theta(w_t = \omega \mid w_{t-1}, w_{t-2}, \cdots)$ $\quad \forall \omega \in \Omega$ vocabulary

Distance: Euclid distance, in this work

Alternatives: Fisher-Rao (Du&Tanaka-Ishii 2024) and JS Divergence

No need for Takens' embedding?

Both w/ and w/o produces similar results (Appendix E)

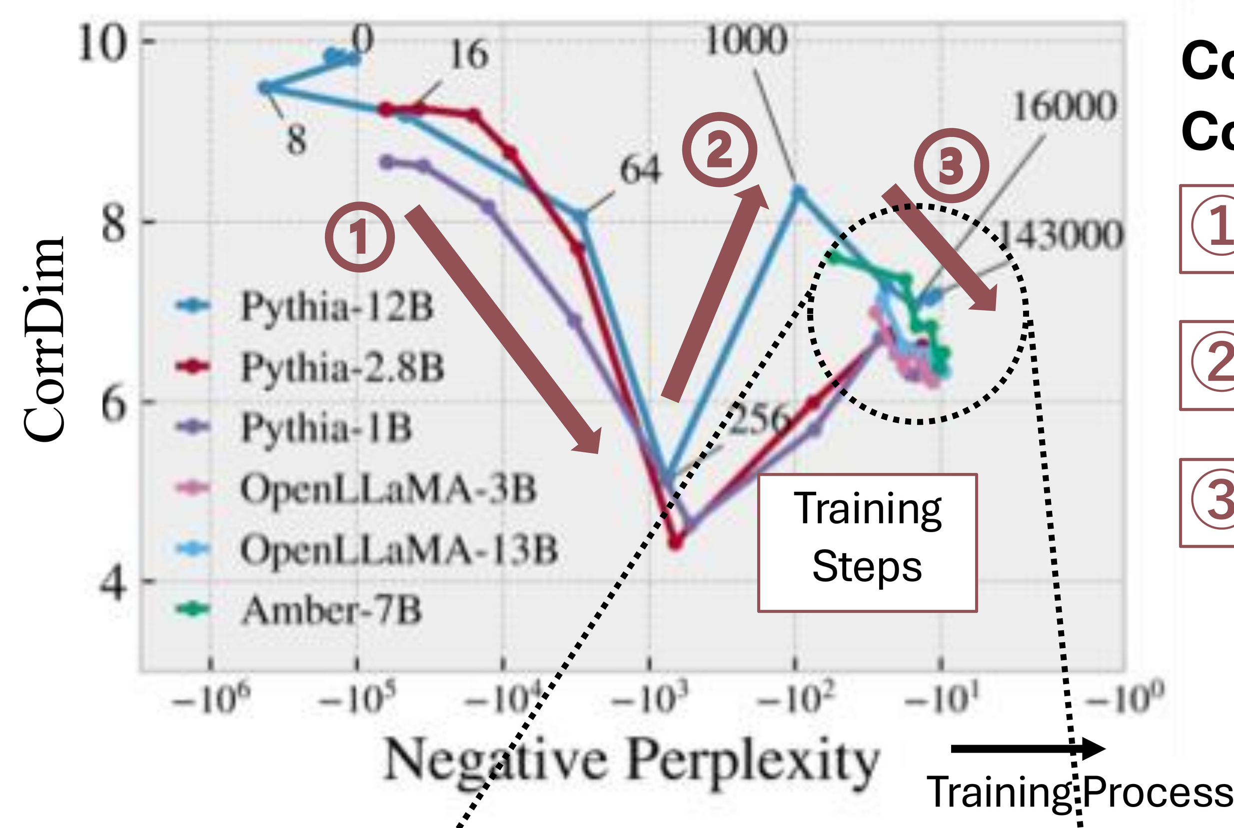
Questions

1. How does **d** evolve through training?
2. What is the ultimate **d** of LLM and natural language? How universal is **d**?
3. What does **d** show?

Related Works

1. History of studies of complexity and self-similarity of natural language
2. Only one previous work to calculate d of natural language: Doxas & Oliver (2017,10), based on simple word frequencies
3. Recent interest in macroscopic behavior of LLM
   - Intrinsic dimension of LLM (NeurIPS 2024, 2025)
   - Scaling laws for LLMs (J. Kaplan 2020)
4. Our previous work appeared in Physical Review Research (2024)

Three-Stage Transition in LLM Pre-Training



Consistent Transitions of CorDim

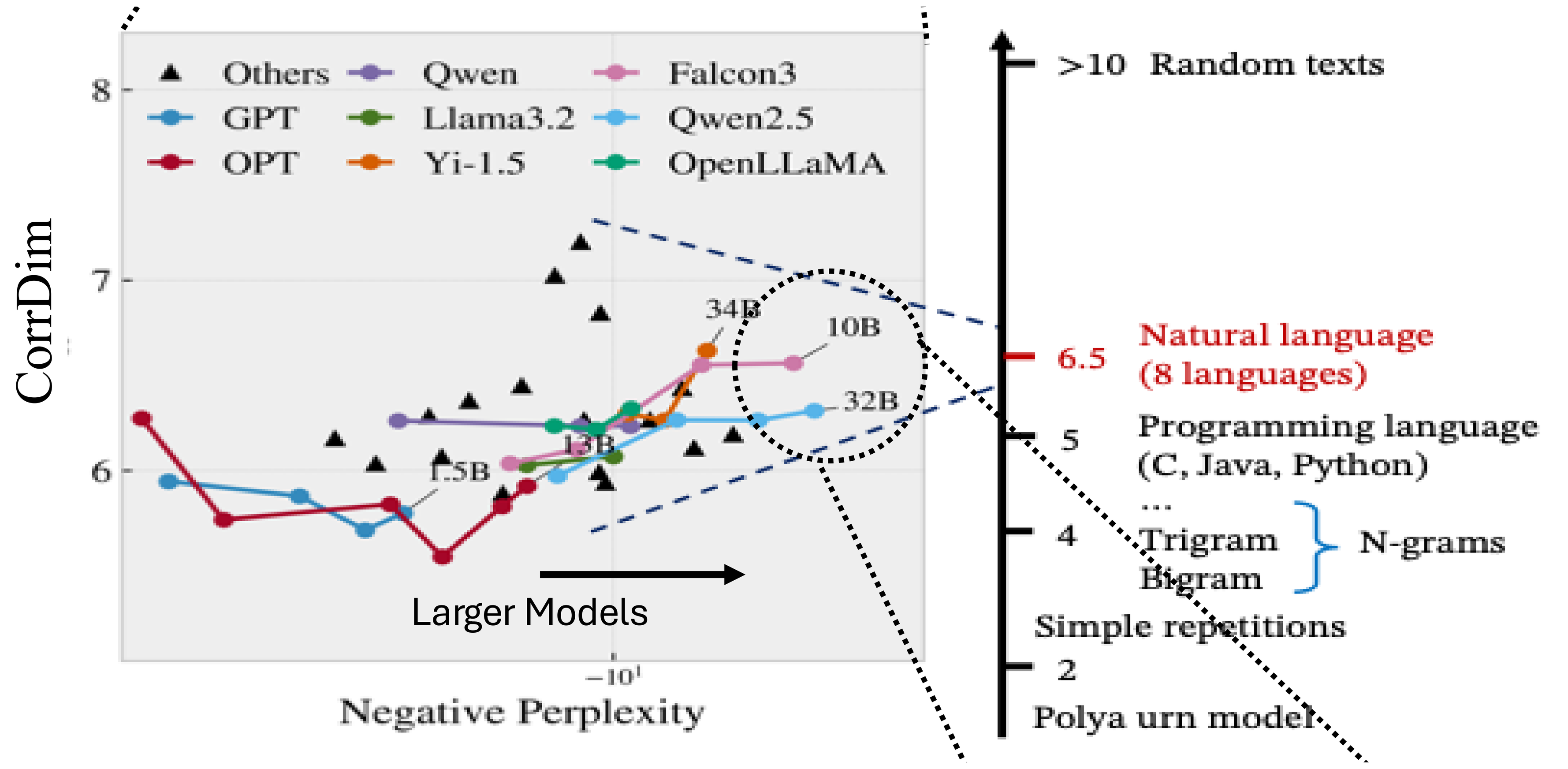
① short-range learning

② long-range learning

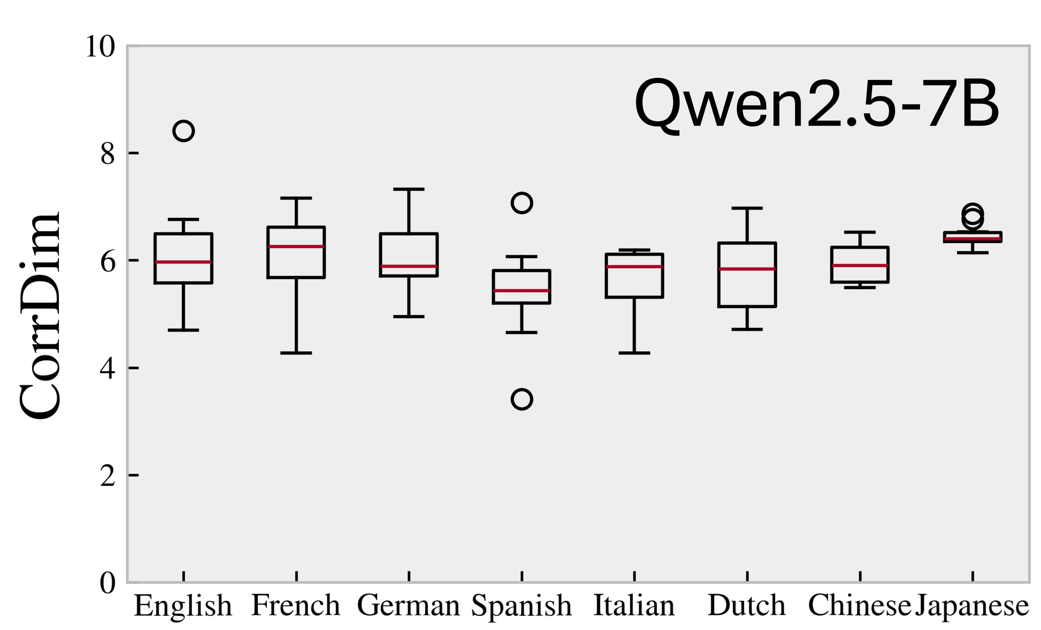
③ Generalization

Convergence at Perplexity Limit

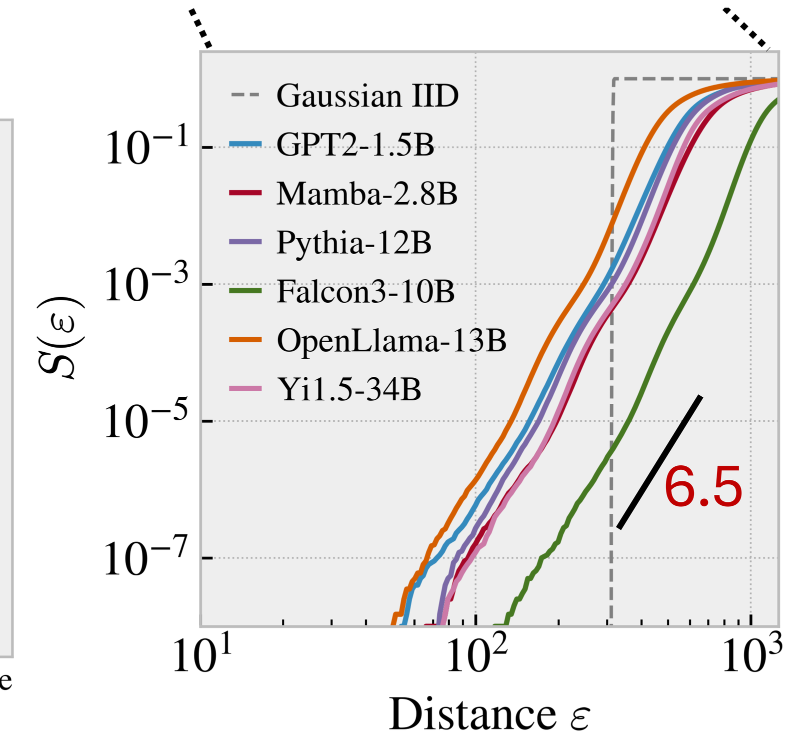
Dataset: Stanford Encyclopedia of Philosophy (60 articles)
Measuring with a variety of pretrained LLMs.



Universal across 8 languages
10 books in Project Gutenberg each language
Minor variations due to genre differences



Consistent across Architectures
Transformer & Mamba



CorrDim shows LLM Generation Qualities

Semantic Shortcuts (identified with distance of log-probs.)

Local Shortcuts

But then towards the paper's end, Newton added his new line of argument, which employed some ~~philosophical analysis together with some~~ experimental evidence to support the ...

Long-Range Shortcuts

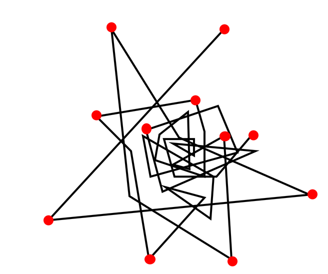
Newton's view that the ray actually contains a series of colors as its "qualities". Are these qualities "sensible" if their presence can be detected only through the use of one or more prisms but never through the inspection of the sunlight through ordinary means ? These are apt to strike us as ...

Repetitions, Incoherencies, Blands

d tends to decrease for such generation patterns

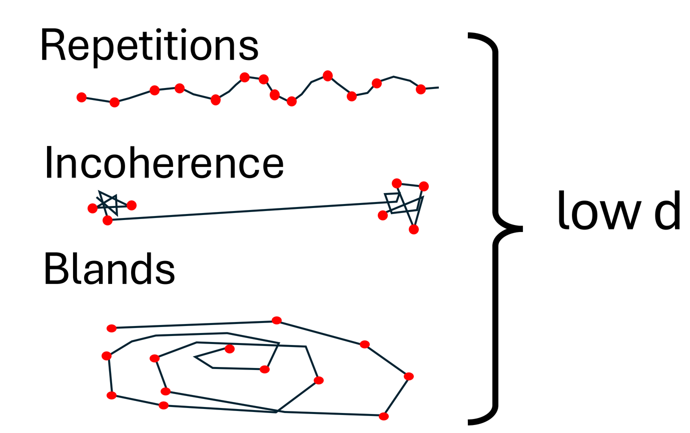
Normal text

$d \approx 6$



Random Walk has	$d = 2.0$
White noise has	$d \to \infty$

Texts generated with GPT4o, prompted to generate repetitions, incoherencies, blands



Group	CorrDim mean	CorrDim p-value
Normal	5.04	-
Repetitive	3.80	9.5E-7
Incoherent	3.96	2.9E-6
Bland	4.51	1.1E-3

Hallucination vs. Memorization

Low CorrDim on Context Predicts Risk of Hallucination

Article "process-theism"

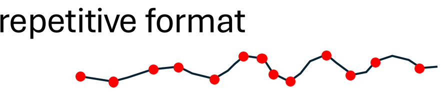
... Philosophers and theologians who have published a monograph defending some variety of process theism informed by Whitehead or Hartshorne include: Henry Nelson Wieman , Bernard Meland , Paul Weiss , Norman Pittenger , Daniel Day Williams , John Moskop, William L. Reese, John B. Cobb, Jr., Schubert Ogden , Edgar A. Towne , Eugene H. Peters , Bowman Clarke , Joseph Bracken, Burton Z. Cooper, Marjorie Hewitt Suchocki, Gene A. Reeves, Lewis S. Ford , André Gounelle, Rem B. Edwards, **(start generation)**

LLM continues generation of names

Model	Normal text (ave.)	Knowledge-intensive text	Recalling or Hallucinating
Qwen2.5-0.5B	5.88	3.32	hallucinate
Qwen2.5-7B	6.27	3.56	hallucinate
Qwen2.5-32B	6.32	4.42	hallucinate
Falcon3-1B	6.03	3.28	hallucinate
Falcon3-3B	6.11	3.14	hallucinate
Falcon3-7B	6.55	6.68	recall
Falcon3-10B	6.56	8.49	recall

drop → rise

Hallucination: continue the repetitive format



Qwen2.5-7B (CorrDim: 3.56):

Robert Schofield, Robert McAfee Brown, ... bert John Russell, John H. Cobb, Jr., Arthur Roberts , Robert ... Connolly , and Bruce L. Reichenbach .

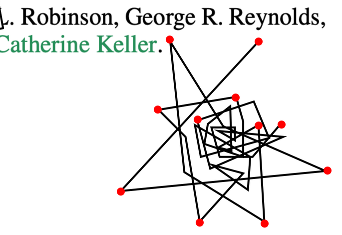
fabricated names

Falcon3-10B (CorrDim: 8.49):

Delwin Brown, David Ray Griffin, ... Robinson, George R. Reynolds, Franklin I. Gamwell, Alan I. Meg... , and Catherine Keller.

correct names

Memorization: continue the names



Insight: Hallucination requires far lower complexity

Take-away

Fractal is a key to understand LLM generation behaviors

CorDim is off-the-shelf:

- inference-time
- low memory overhead
- zero additional LLM calls
- robust to low-precision inference

NEURAL INFORMATION PROCESSING SYSTEMS 2025

NeurIPS arXiv lab