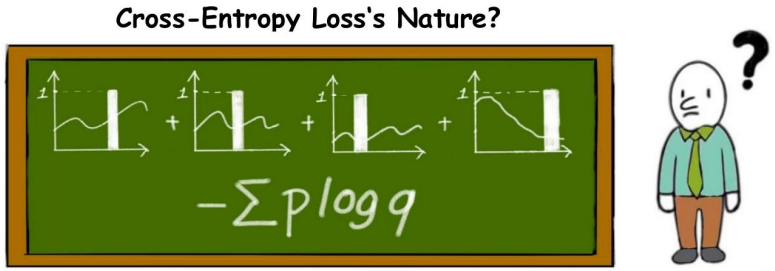
What Scales in Cross-Entropy Scaling Law?

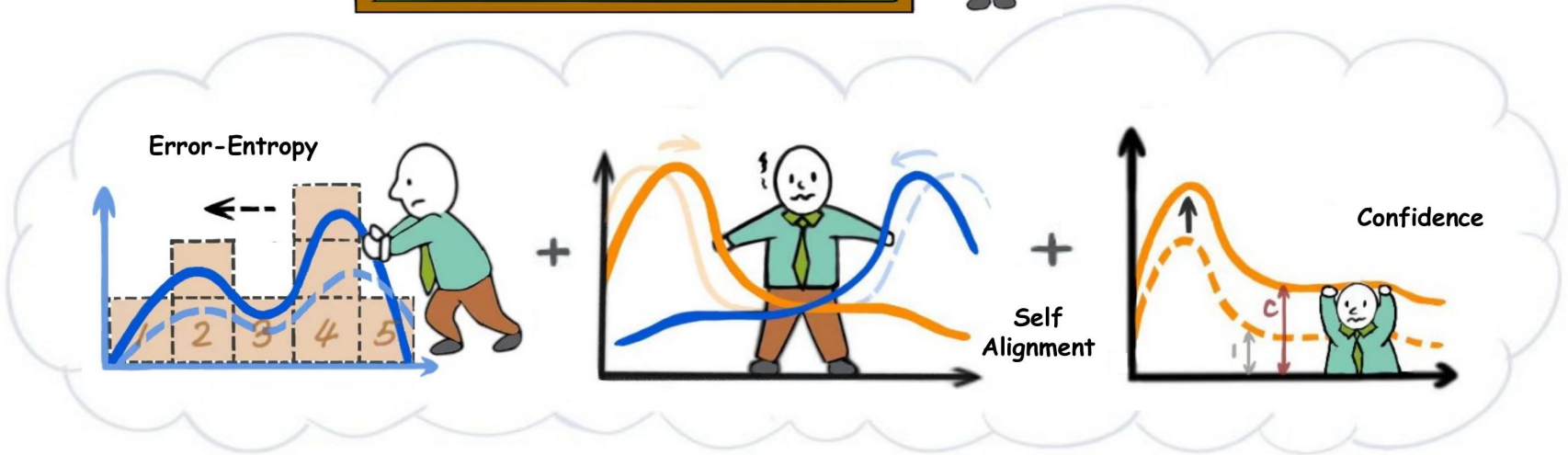
Junxi Yan, Zixi Wei, Jingtao Zhan, Qingyao Ai, Yiqun Liu

Tsinghua University

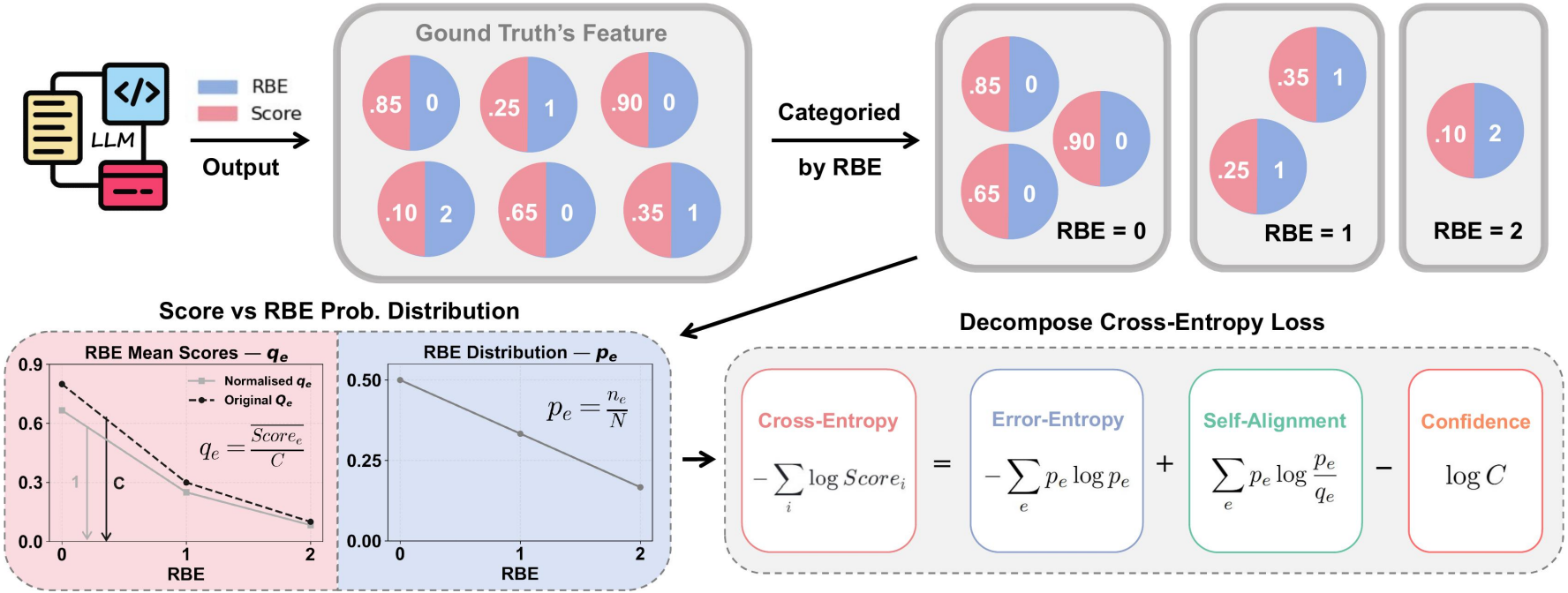
ICLR 2026

Motivation

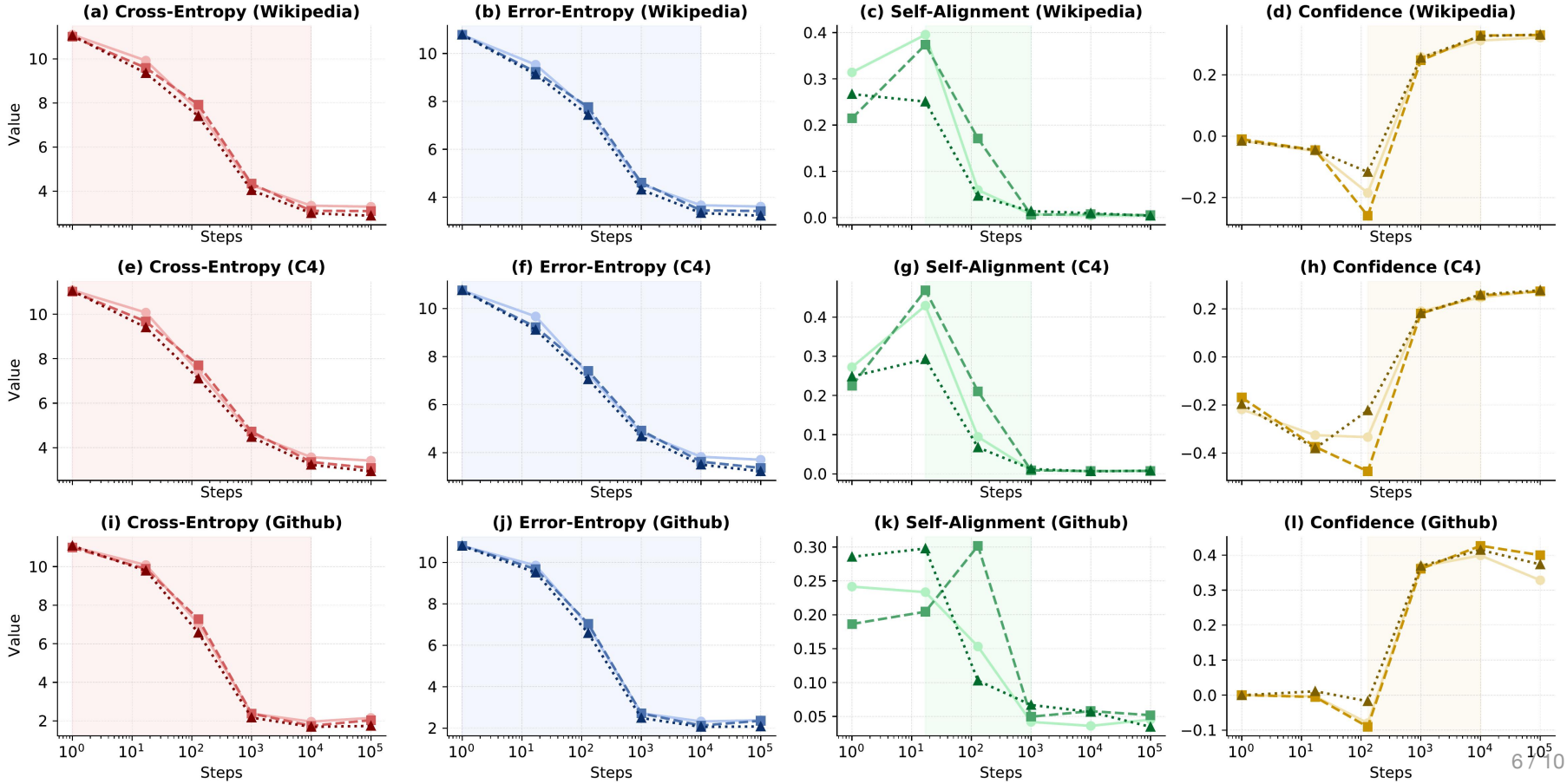

Cross-Entropy Loss's Nature?
$-\sum p \log q$


Error-Entropy
1
2
3
4
5
Self
Alignment
Confidence
c

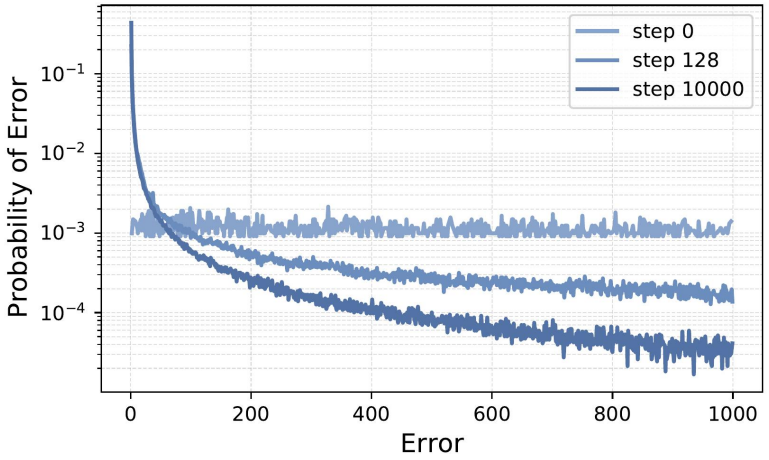
Decomposition Pipeline

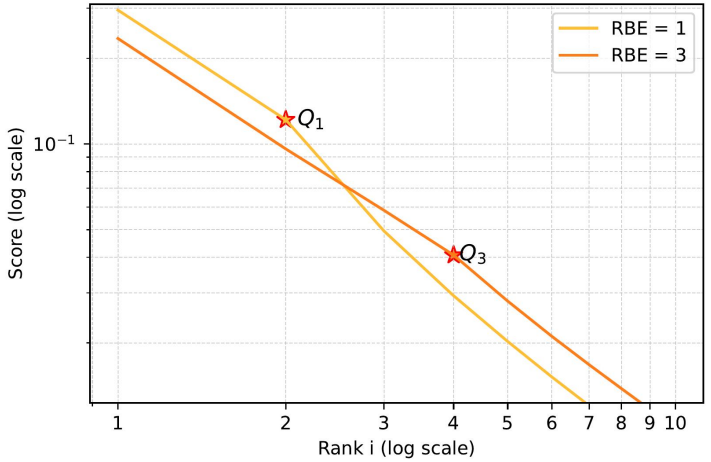


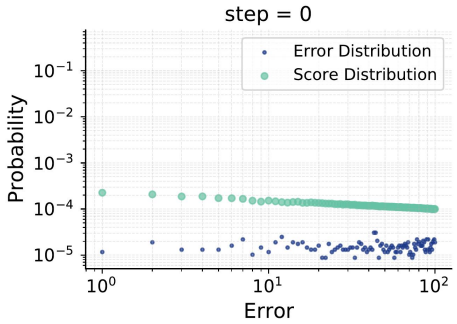
Training Dynamics

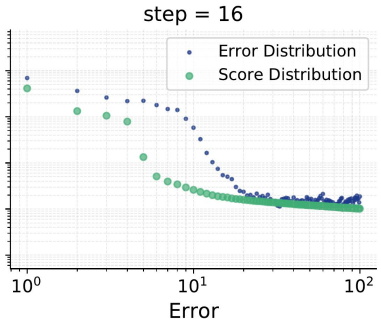


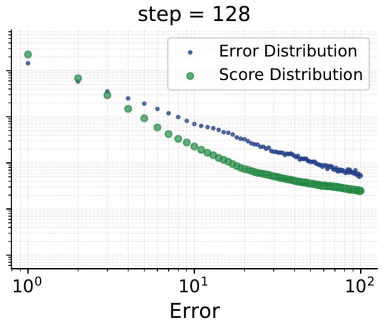
Component Dynamic Analysis

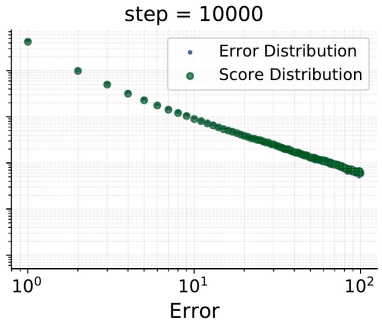




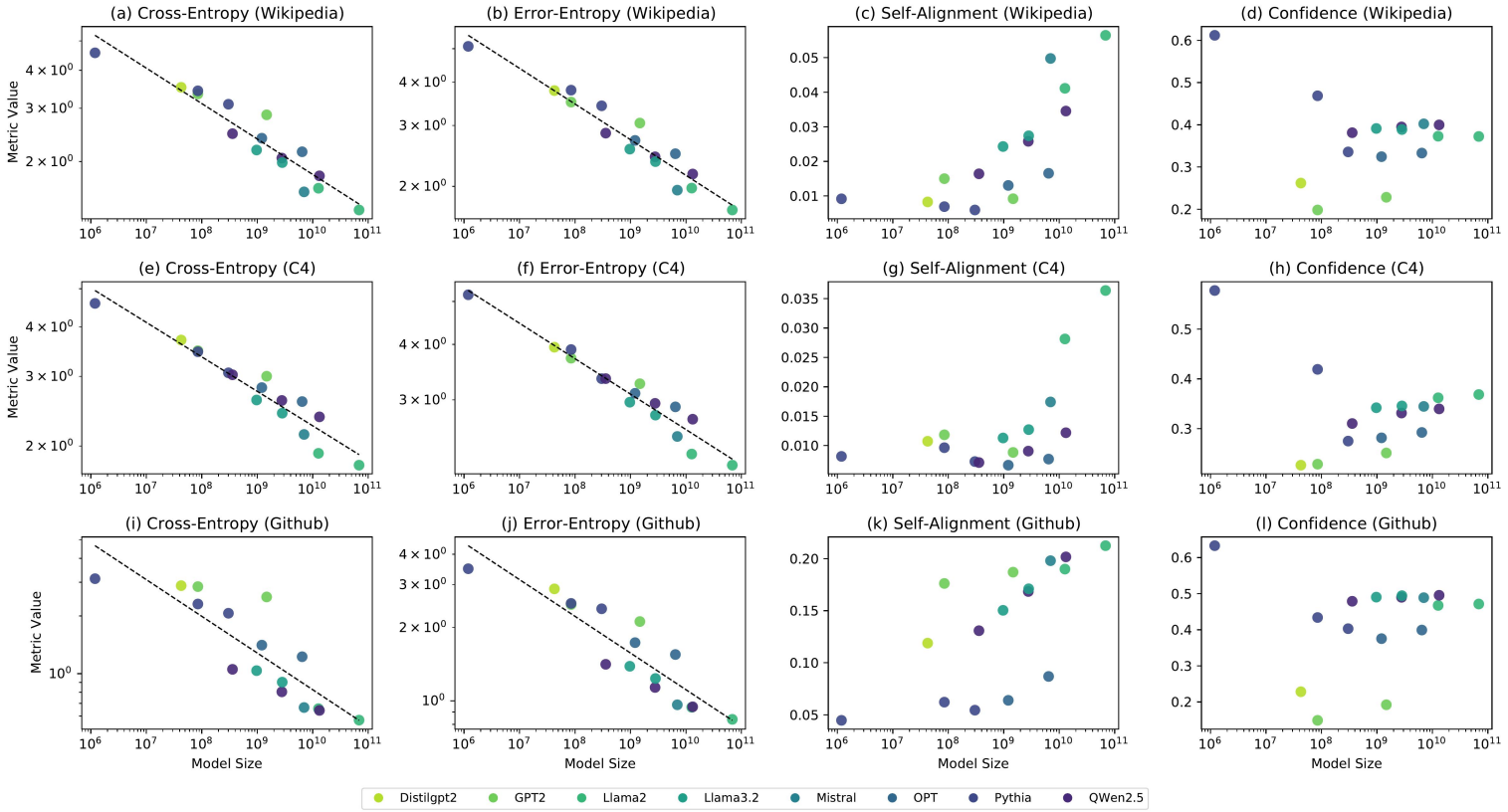




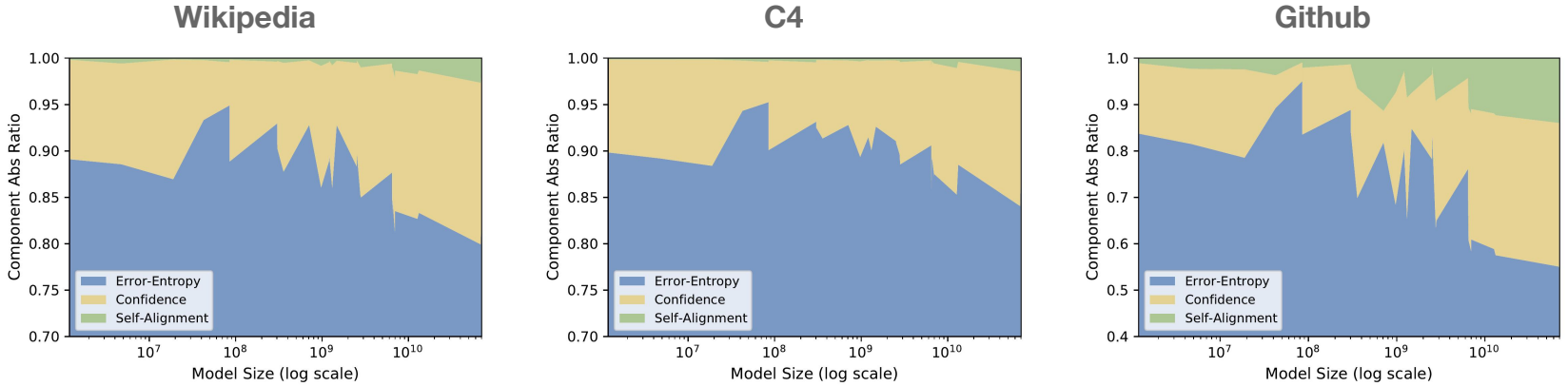




Scaling with Model Size



Component Share vs. Model Size



Error-Entropy dominates in small models but its share decreases as model size grows.

Key Findings

1. Cross-Entropy = Error-Entropy + Self-Alignment - Confidence
2. Only Error-Entropy follows a robust power-law scaling
3. Self-Alignment and Confidence are largely scale-invariant
4. Error-Entropy dominates in small models, diminishes in large ones
5. This explains why cross-entropy scaling breaks down at large scales

The Error-Entropy Scaling Law: a more accurate description of model behavior.