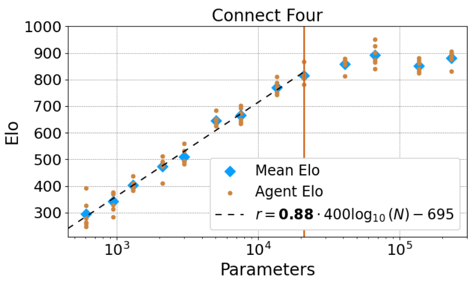
AlphaZero Neural Scaling and Zipf's Law: a Tale of Board Games and Power Laws

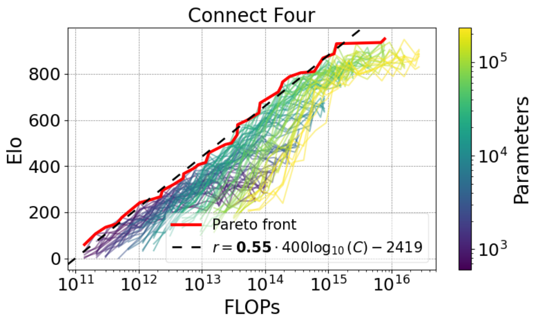
Oren Neumann, Claudius Gros
Goethe University Frankfurt

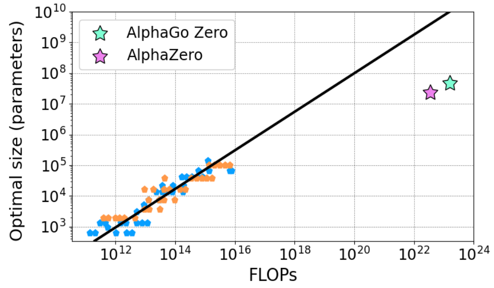
AlphaZero scaling laws

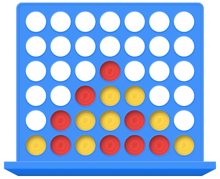
In a previous paper we show the AlphaZero exhibits power law scaling laws on two games, Connect Four and Pentago.

But where do these scaling laws come from?









The Quantization Model for LLM scaling

A possible explanation of LLM scaling laws: (*Michaud et al. 2024*)

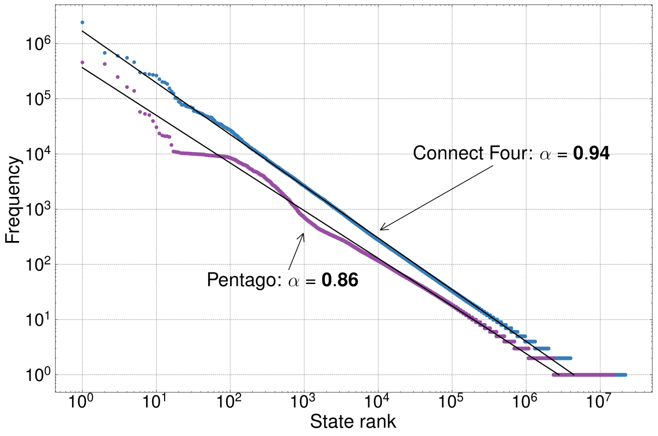
- Data frequencies follow Zipf’s law.
- LLMs learn tasks in discrete, independent *quanta*.
- Quanta are learned one after the other, in descending order of frequency.

↓

Performance power laws stem from Zipf’s law.

Human games are known to follow Zipf’s law (chess, go).

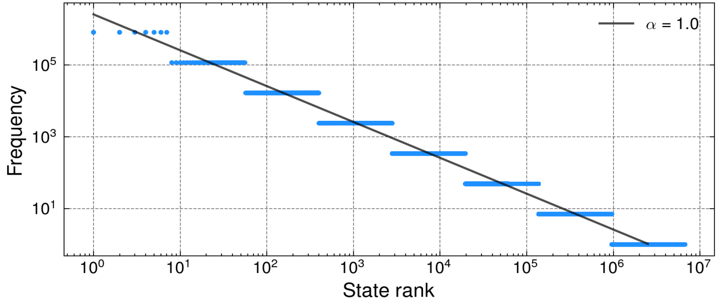
AlphaZero’s games follow Zipf’s law too:



Add strategy

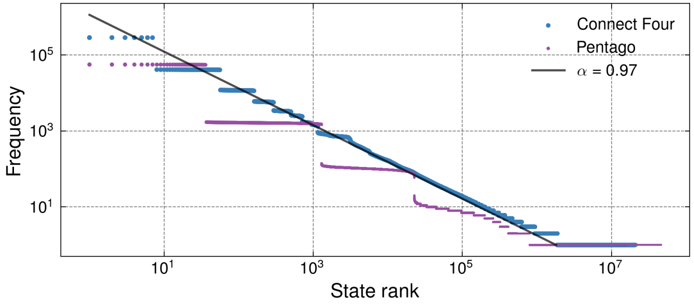
The power law arises from the branching-tree structure of board games:

A. Toy-model distribution



Add advanced rules

B. Random-game distribution



The Quantization Model for LLM scaling

A possible explanation of LLM scaling laws: (*Michaud et al. 2024*)

- ✓ Data frequencies follow Zipf's law.
- LLMs learn tasks in discrete, independent *quanta*.
- Quanta are learned one after the other, in descending order of frequency.

↓

Performance power laws stem from Zipf's law.

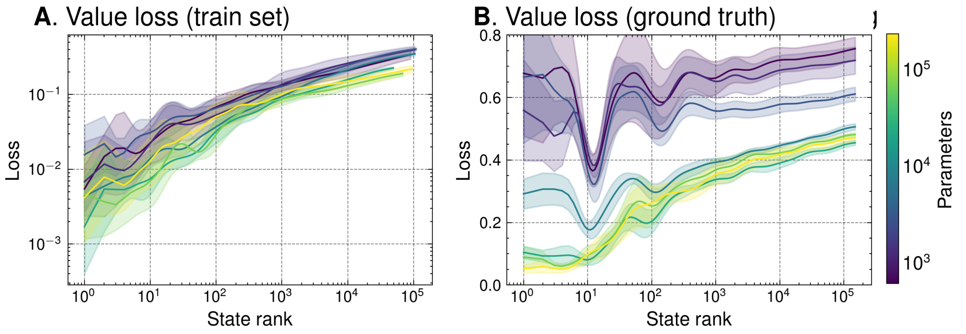
Value loss scaling

$$L_v = (z - v)^2$$

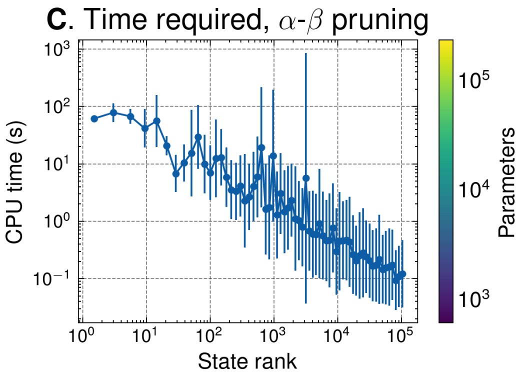
Game result (-1,1) Agent prediction

Value loss L_v is a good proxy for performance.
States (=board positions) are a proxy for ‚tasks‘.

AlphaZero agents learn board states in decreasing order of frequency.



This means agents learn board positions in REVERSE order of complexity, getting the *highest* accuracy on the *hardest* positions.



The Quantization Model for LLM scaling

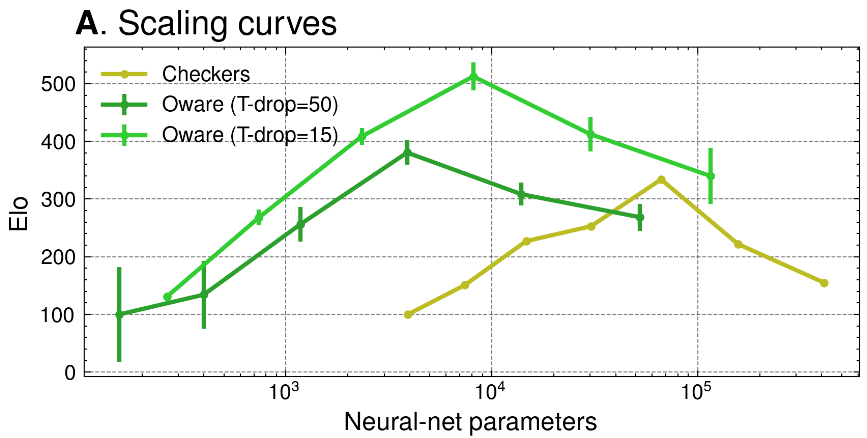
A possible explanation of LLM scaling laws: (*Michaud et al. 2024*)

- ✓ Data frequencies follow Zipf’s law.
- LLMs learn tasks in discrete, independent *quanta*.
- ✓ Quanta are learned one after the other, in descending order of frequency.

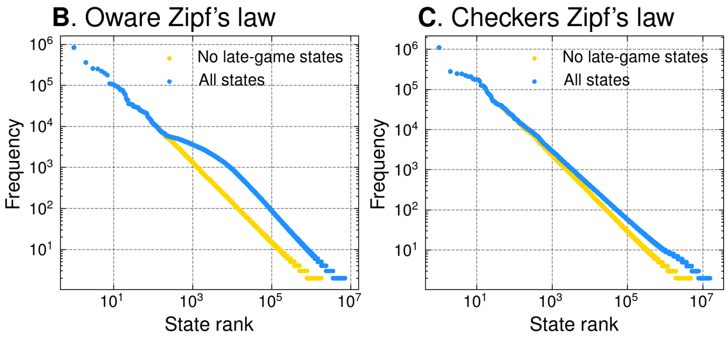
↓

Performance power laws stem from Zipf’s law.

In some games (checkers & Oware) power-law scaling laws break down, even scaling negatively with model size:

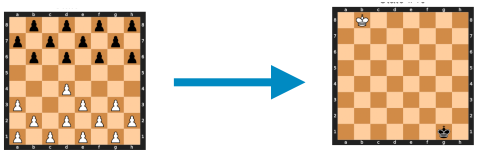


Coincidentally, these games also have strange deviations from Zipf's law.

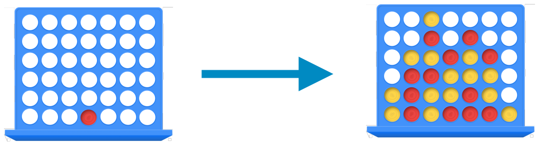


Turn distribution anomaly

The games displaying inverse scaling also share a specific tree structure:
The number of board configurations expands during play, but then contracts at the end of the game.

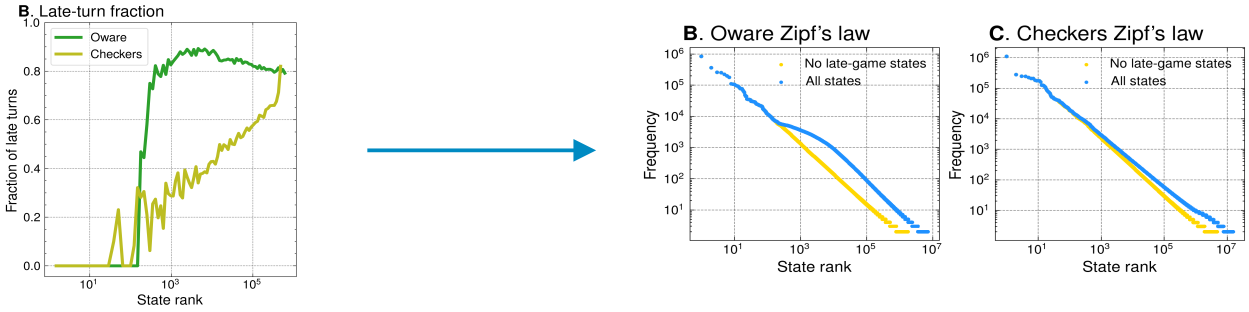


few openings few endings



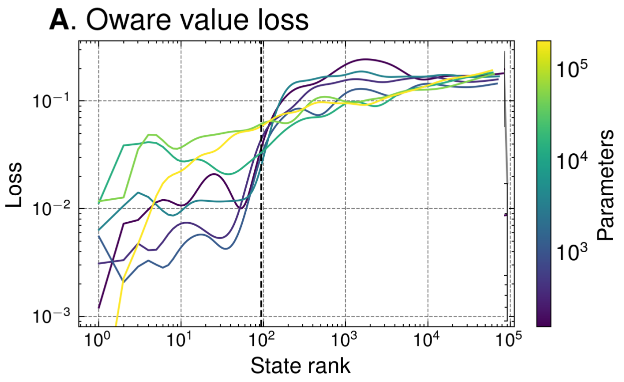
few openings many endings

Checkers and Oware are both capture-to-win games, where the game ends only by capturing *all* the opponent's pieces.
This anomaly skews the Zipf distribution:

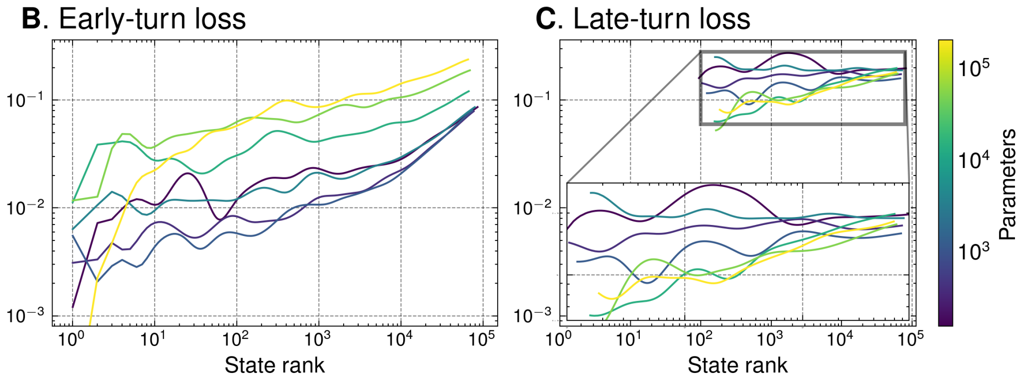


Turn distribution connection to inverse scaling

- Looking at the total loss, larger Oware models have lower loss.
- But smaller models' loss jumps at the transition to late-game states, why?



- Breaking down the loss to early- and late-game states, we see that larger models perform *worse* on openings, but *better* on end-game positions.



- Large models initially improve on early-game, then degrade.
  But they continuously improve on late-game.