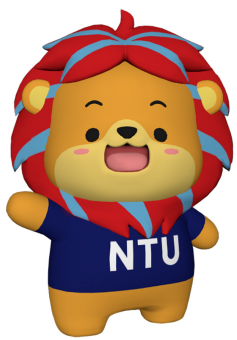
To model human linguistic prediction, make LLMs less superhuman

Byung-Doh Oh

Nanyang Technological University

HHU DLM Colloquium · 19 June 2026



Language processing in humans

The girl found the lamb ...

Language processing in humans

The girl found the lamb ...

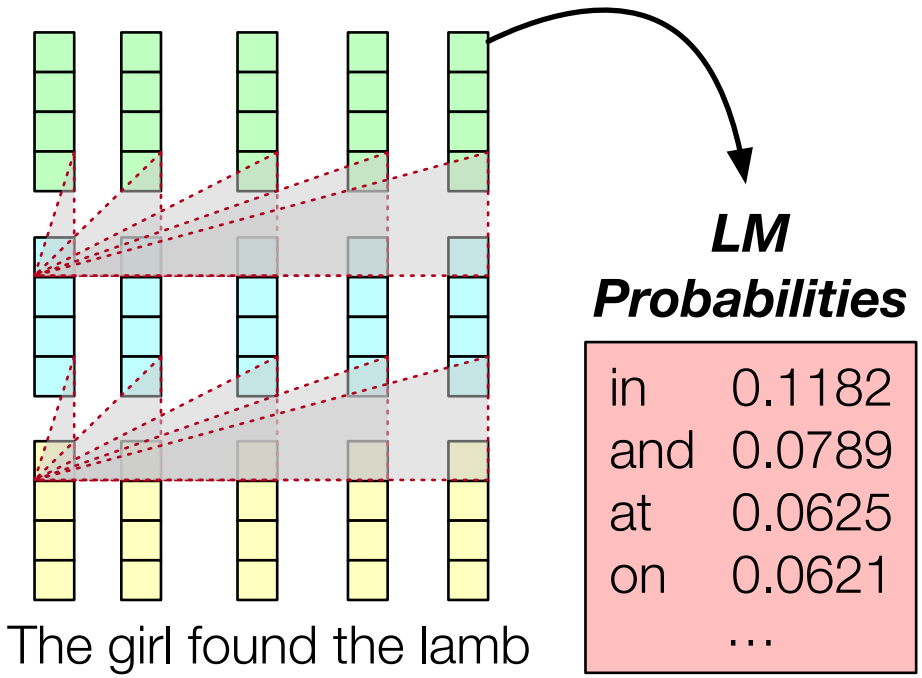
- **Incremental**: We don't wait until the end of the sentence.

Language processing in humans

The girl found the lamb ...

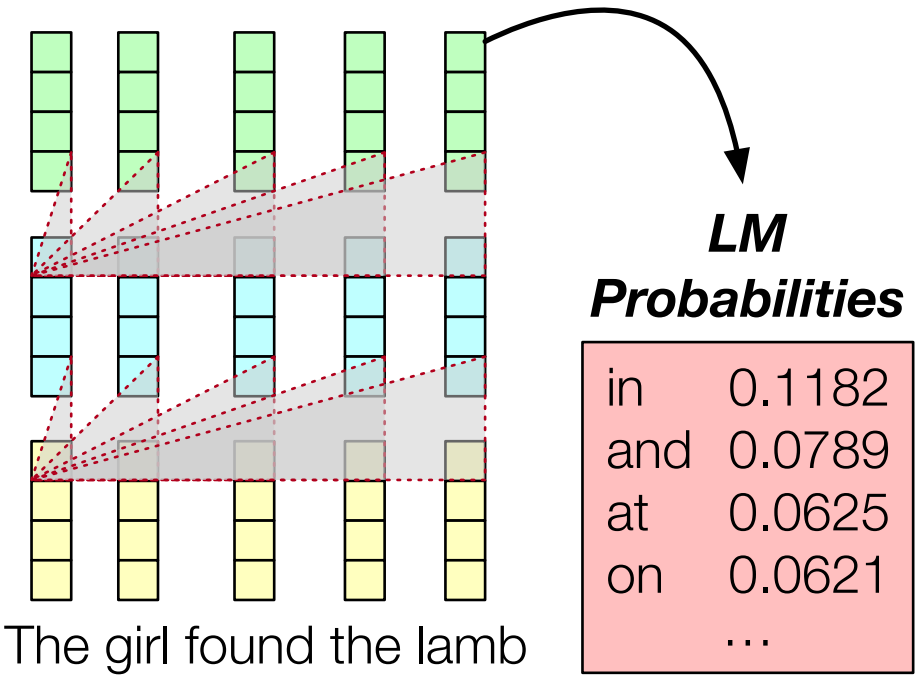
- **Incremental**: We don't wait until the end of the sentence.
- **Predictive**: We proactively anticipate the upcoming material.

Language processing in language models



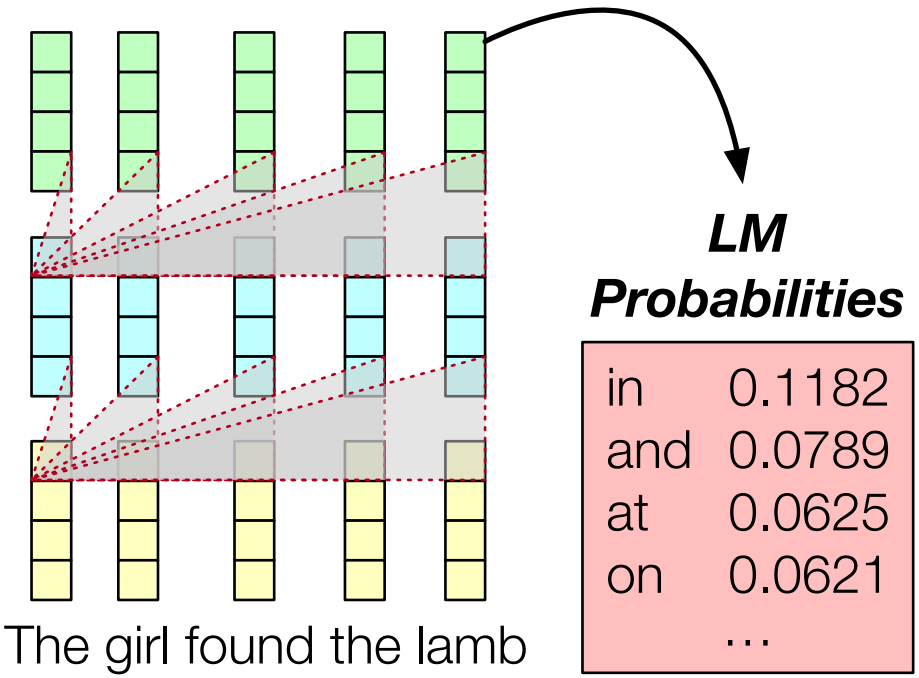
Language processing in language models

- **Incremental!**



Language processing in language models

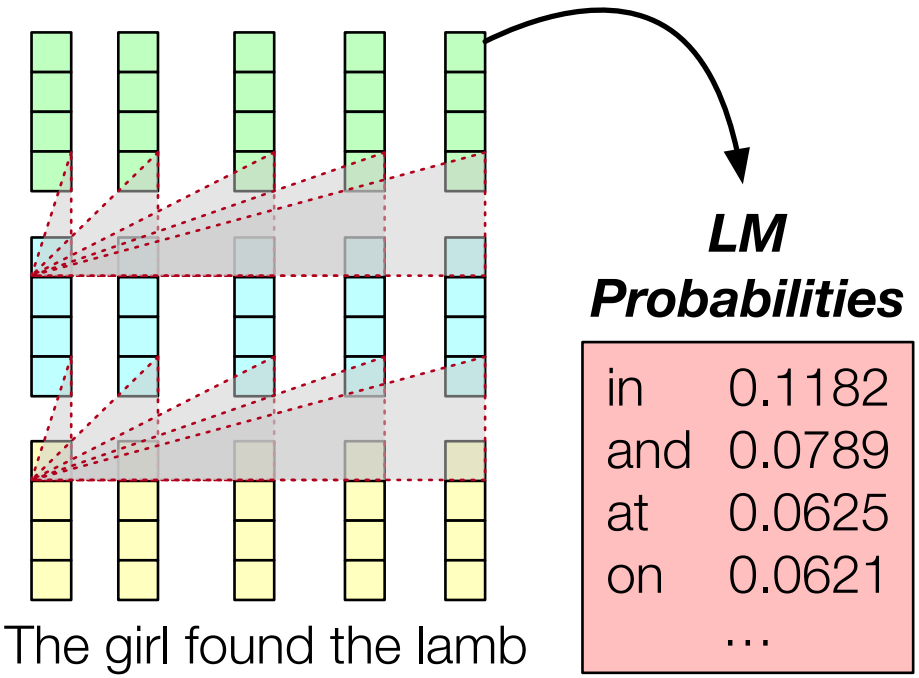
- **Incremental!**



- **Predictive!**

Language processing in language models

- **Incremental!**



- **Predictive!**

These 'shared principles' (Schrimpf et al., 2021) ground the use of LLMs in psycholinguistic research.

LLMs hold promise for psycholinguistics, but they need to be made less superhuman through language modeling techniques.

LLMs hold promise for psycholinguistics, but they need to be made less superhuman through language modeling techniques.

1. Background

LLMs hold promise for psycholinguistics, but they need to be made less superhuman through language modeling techniques.

1. Background
2. Superhuman long-term memory
3. Superhuman short-term memory

LLMs hold promise for psycholinguistics, but they need to be made less superhuman through language modeling techniques.

1. Background
2. Superhuman long-term memory
3. Superhuman short-term memory
4. What should we do about it?

LLMs hold promise for psycholinguistics, but they need to be made less superhuman through language modeling techniques.

1. Background
2. Superhuman long-term memory
3. Superhuman short-term memory
4. What should we do about it?
5. Conclusion

LLMs hold promise for psycholinguistics, but they need to be made less superhuman through language modeling techniques.

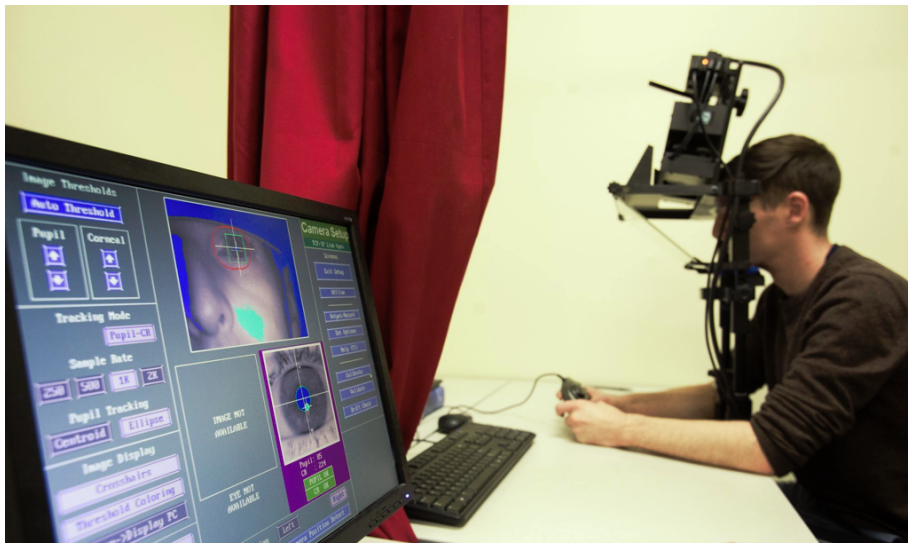
1. Background
2. Superhuman long-term memory
3. Superhuman short-term memory
4. What should we do about it?
5. Conclusion

Tal Linzen (NYU)

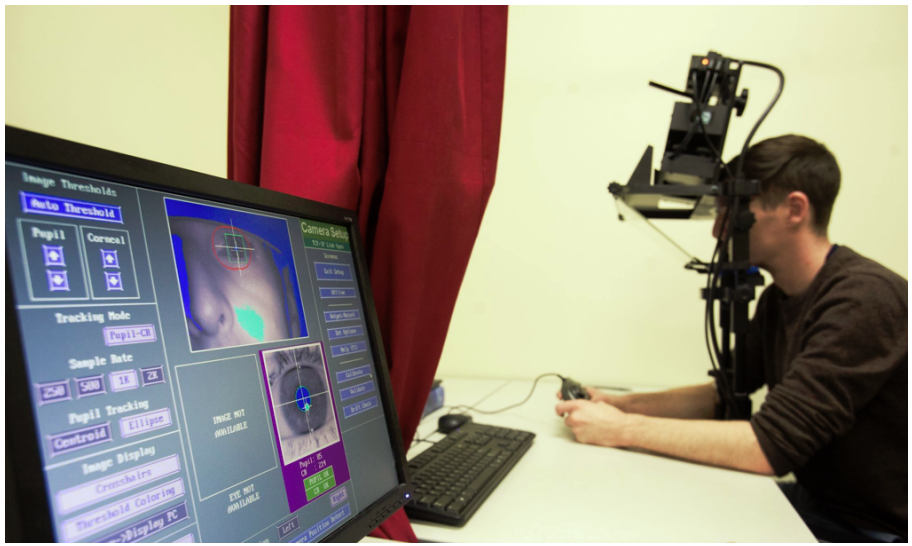
1. Background

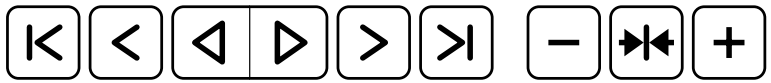
Human language processing is in part driven by prediction;
LLMs provide a valuable tool for studying this process.

Computation and behavior



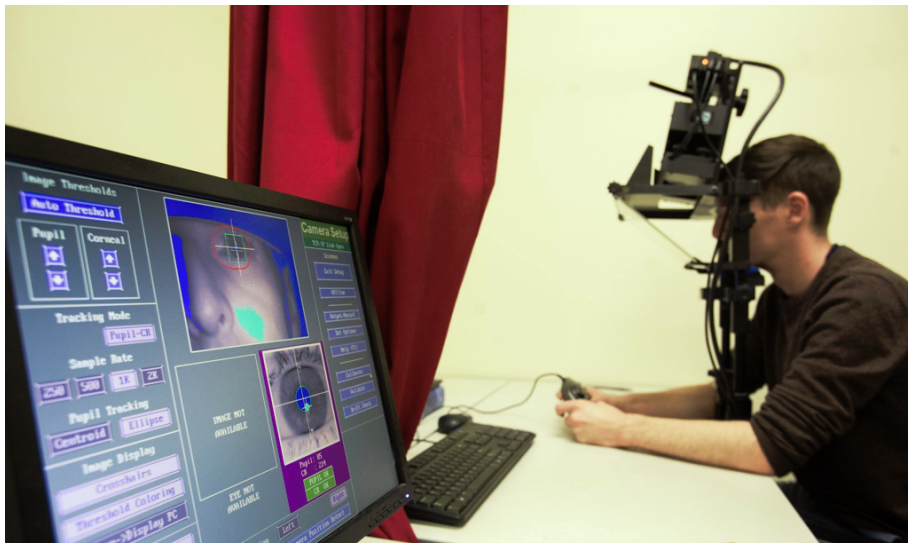
Computation and behavior

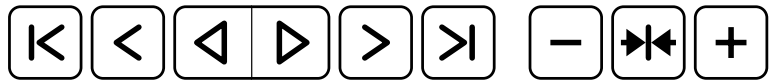




- By studying real-time behavior, psycholinguists aim to understand what goes on in the mind during language processing.

Computation and behavior

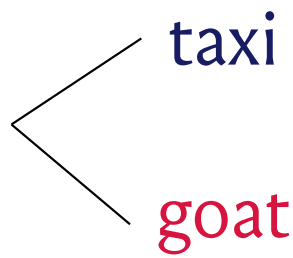




- By studying real-time behavior, psycholinguists aim to understand what goes on in the mind during language processing.
- **Assumption**: Longer reading times reflect more processing effort.

What have behavioral data told us about language processing?

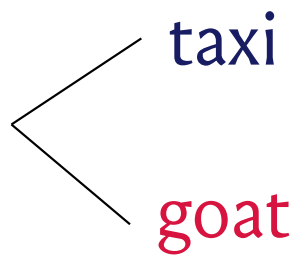
What have behavioral data told us about language processing?

I landed in Düsseldorf and took a 

Balota et al. (1985), Ehrlich and Rayner (1981), Kutas and Hillyard (1980)

What have behavioral data told us about language processing?

I landed in Düsseldorf and took a

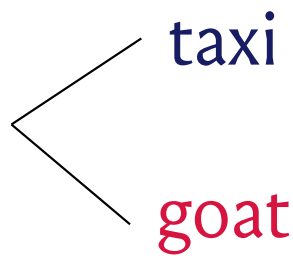


Balota et al. (1985), Ehrlich and Rayner (1981), Kutas and Hillyard (1980)

- **An idea**: Language processing is **inference under uncertainty**.

What have behavioral data told us about language processing?

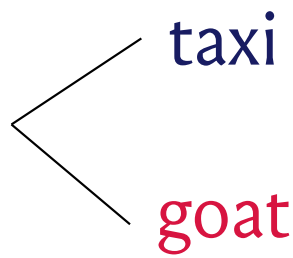
I landed in Düsseldorf and took a



Balota et al. (1985), Ehrlich and Rayner (1981), Kutas and Hillyard (1980)

- **An idea**: Language processing is **inference under uncertainty**.
- We maintain a **mental probability model** over all possible interpretations of the input.

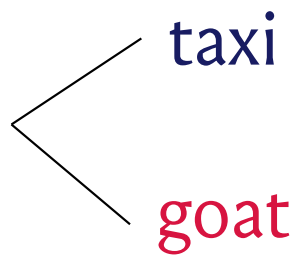
What have behavioral data told us about language processing?

I landed in Düsseldorf and took a 

Balota et al. (1985), Ehrlich and Rayner (1981), Kutas and Hillyard (1980)

- **An idea**: Language processing is **inference under uncertainty**.
- We maintain a **mental probability model** over all possible interpretations of the input.
- As each word arrives, we use it to update our probability model:

What have behavioral data told us about language processing?

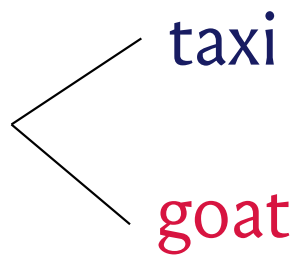
I landed in Düsseldorf and took a


Balota et al. (1985), Ehrlich and Rayner (1981), Kutas and Hillyard (1980)

- **An idea**: Language processing is **inference under uncertainty**.
- We maintain a **mental probability model** over all possible interpretations of the input.
- As each word arrives, we use it to update our probability model:

- **Predictable** words carry **less info** and **cause a smaller update**.

What have behavioral data told us about language processing?

I landed in Düsseldorf and took a 

Balota et al. (1985), Ehrlich and Rayner (1981), Kutas and Hillyard (1980)

- **An idea**: Language processing is **inference under uncertainty**.
- We maintain a **mental probability model** over all possible interpretations of the input.
- As each word arrives, we use it to update our probability model:

- **Predictable** words carry **less info** and **cause a smaller update**.
- **Unpredictable** words carry **more info** and **cause a larger update**.

Surprisal theory: Effects of context

I landed in Düsseldorf and took a **taxi** ...

Surprisal theory: Effects of context

I landed in Düsseldorf and took a **taxi** ...

$$\text{RT} \propto -\log \text{P}(w_t \mid w_1, w_2, \ldots, w_{t-1})$$

Hale (2001), Levy (2008)

Surprisal theory: Effects of context

I landed in Düsseldorf and took a **taxi** ...

$$\text{RT} \propto -\log \text{P}(w_t \mid w_1, w_2, \ldots, w_{t-1})$$

Hale (2001), Levy (2008)

- Readers are sensitive to probabilities over six orders of magnitude.

  (N. J. Smith & Levy, 2013)

Surprisal theory: Effects of context

I landed in Düsseldorf and took a **taxi** ...

$$\text{RT} \propto -\log \text{P}(w_t \mid w_1, w_2, \ldots, w_{t-1})$$

Hale (2001), Levy (2008)

- Readers are sensitive to probabilities over six orders of magnitude.
  (N. J. Smith & Levy, 2013)
- The effect of probability on RT is more of a slowdown at unpredictable words, rather than a boost at predictable words.
  (Shain et al., 2024)

Surprisal theory: Effects of context

I landed in Düsseldorf and took a **taxi** ...

$$\text{RT} \propto -\log \text{P}(w_t \mid w_1, w_2, \ldots, w_{t-1})$$

Hale (2001), Levy (2008)

- Readers are sensitive to probabilities over six orders of magnitude. (N. J. Smith & Levy, 2013)
- The effect of probability on RT is more of a slowdown at unpredictable words, rather than a boost at predictable words. (Shain et al., 2024)
- How probability is allocated across other alternative words does not seem to influence the actual word's RT. (Frisson et al., 2017)

Surprisal theory: Effects of context

I landed in Düsseldorf and took a **taxi** ...

$$\text{RT} \propto -\log \text{P}(w_t \mid w_1, w_2, \ldots, w_{t-1})$$

Hale (2001), Levy (2008)

Surprisal theory: Effects of context

I landed in Düsseldorf and took a **taxi** ...

$$\text{RT} \propto -\log \text{P}(w_t \mid w_1, w_2, \ldots, w_{t-1})$$

Hale (2001), Levy (2008)

- Traditional methods include the cloze task and n-gram models.

Surprisal theory: Effects of context

I landed in Düsseldorf and took a **taxi** ...

$$\text{RT} \propto -\log \text{P}(w_t \mid w_1, w_2, \ldots, w_{t-1})$$

Hale (2001), Levy (2008)

- Traditional methods include the cloze task and n-gram models.
- Large language models mitigate issues with data sparsity.

Surprisal theory: Effects of context

I landed in Düsseldorf and took a **taxi** ...

$$\text{RT} \propto -\log \text{P}(w_t \mid w_1, w_2, \ldots, w_{t-1})$$

Hale (2001), Levy (2008)

- Traditional methods include the cloze task and n-gram models.
- Large language models mitigate issues with data sparsity.
- Crucially, we want $\text{P}(w_t \mid w_1, w_2, \ldots, w_{t-1})$ to be human-like.

Evaluation: Regression modeling

I landed in Düsseldorf and took a **taxi** ...

$$\text{RT} \propto -\log \text{P}(w_t \mid w_1, w_2, \ldots, w_{t-1})$$

Hale (2001), Levy (2008)

Evaluation: Regression modeling

I landed in Düsseldorf and took a **taxi** ...

$$\text{RT} \propto -\log \text{P}(w_t \mid w_1, w_2, \ldots, w_{t-1})$$

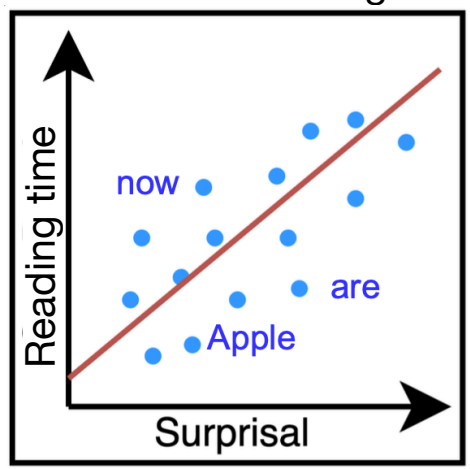
Hale (2001), Levy (2008)

Collecting word-by-word reading times

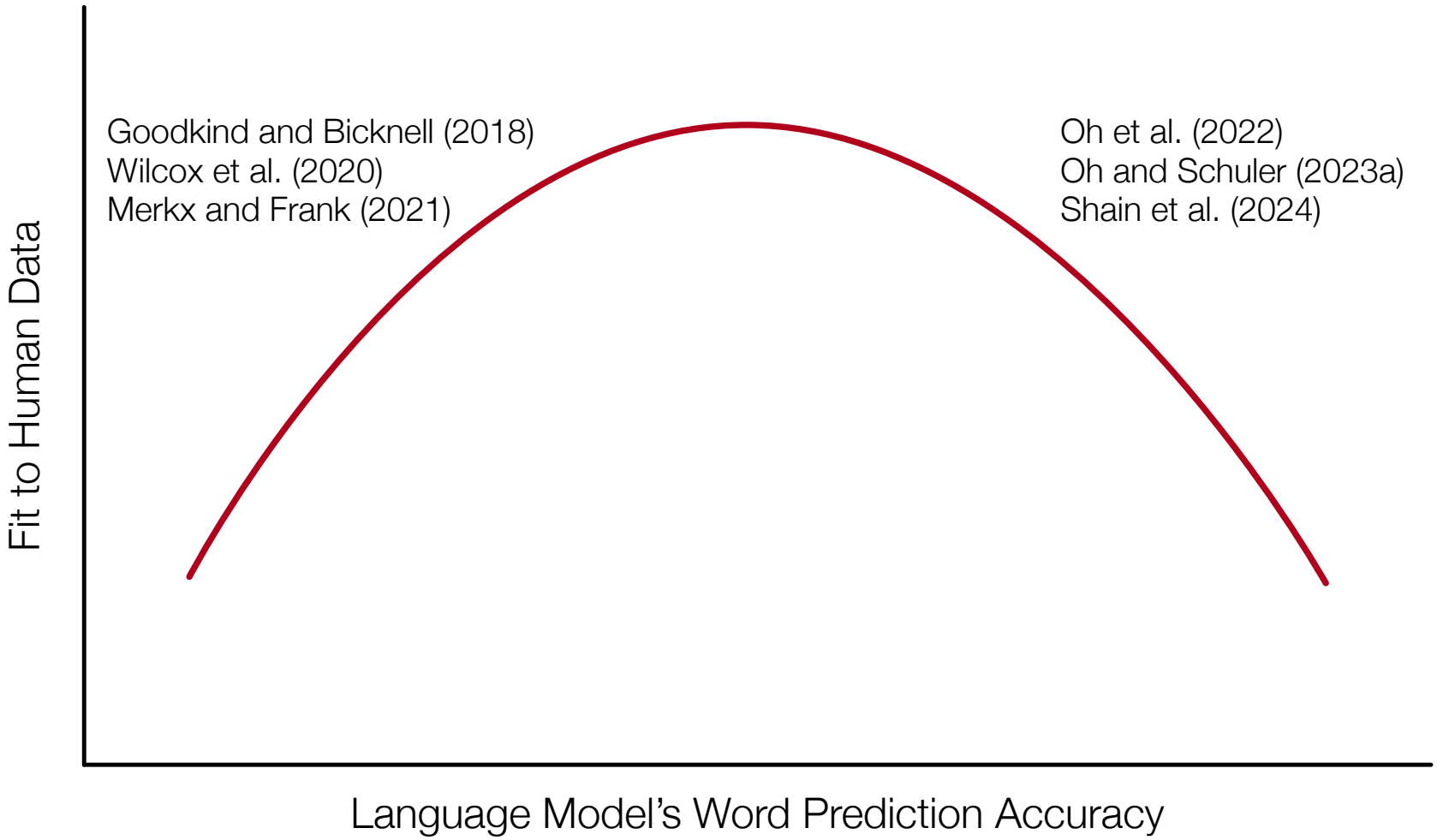
(e.g. Futrell et al. 2021)

Modeling approach

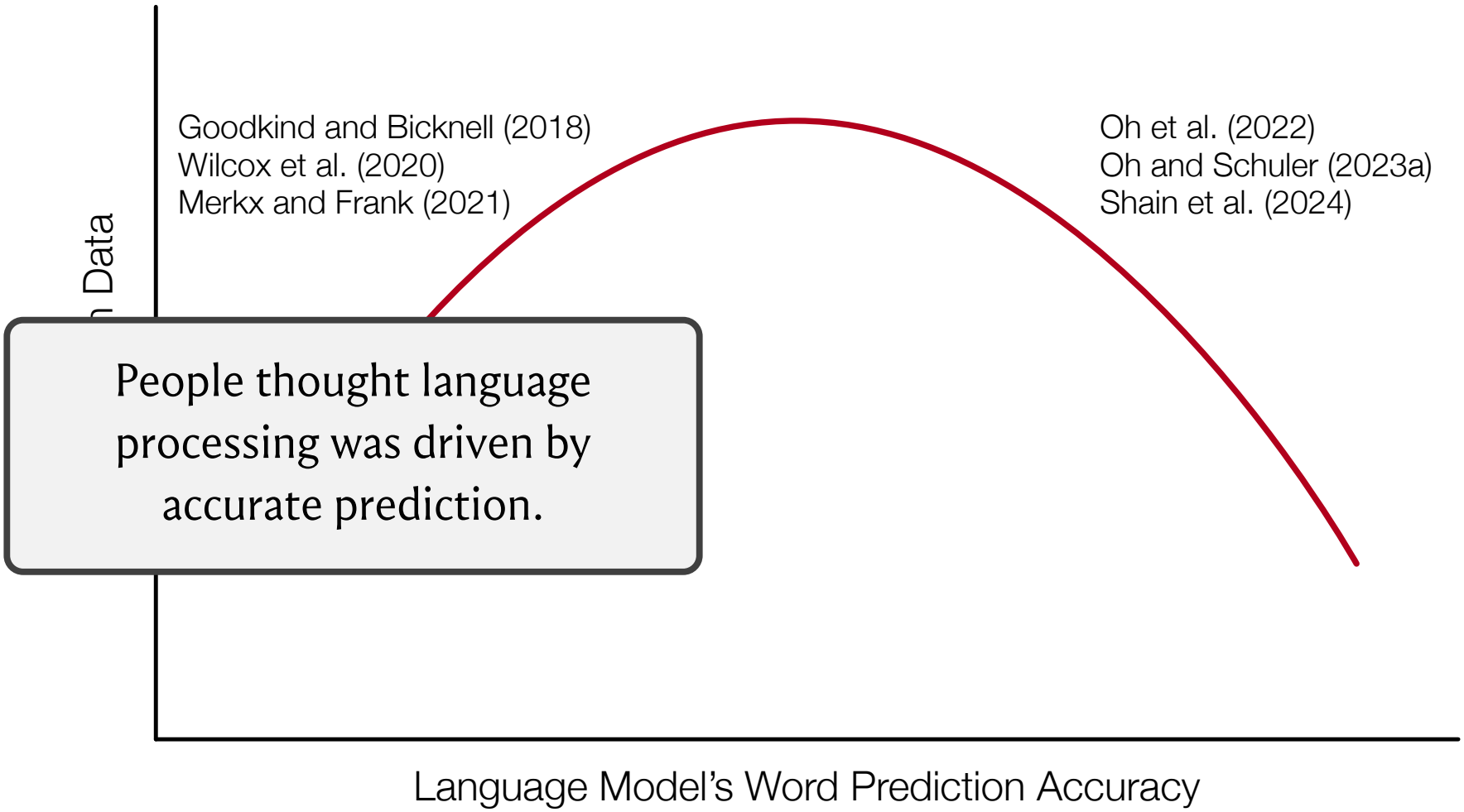
Linear fits to reading times



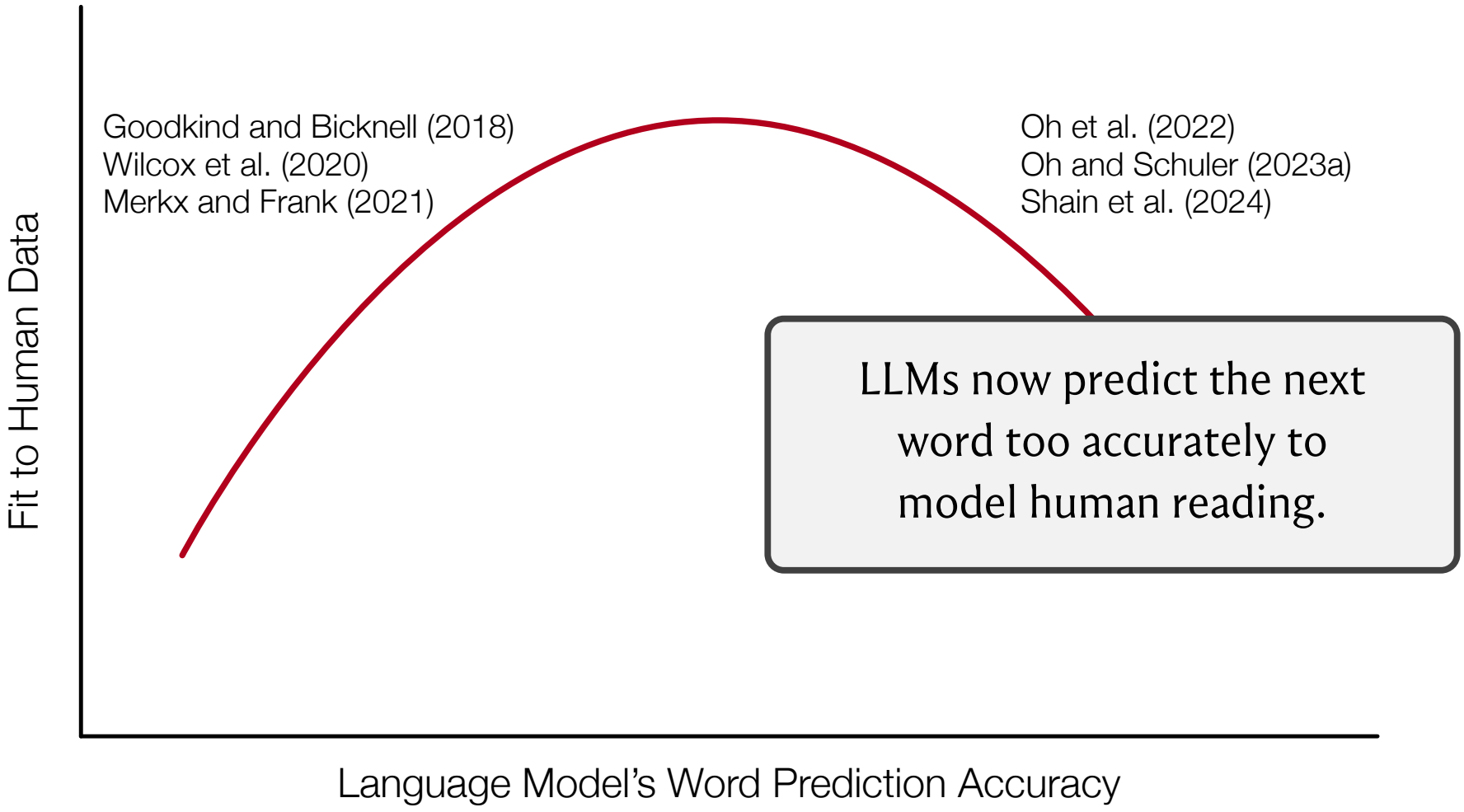
LLMs have become superhuman as models of prediction



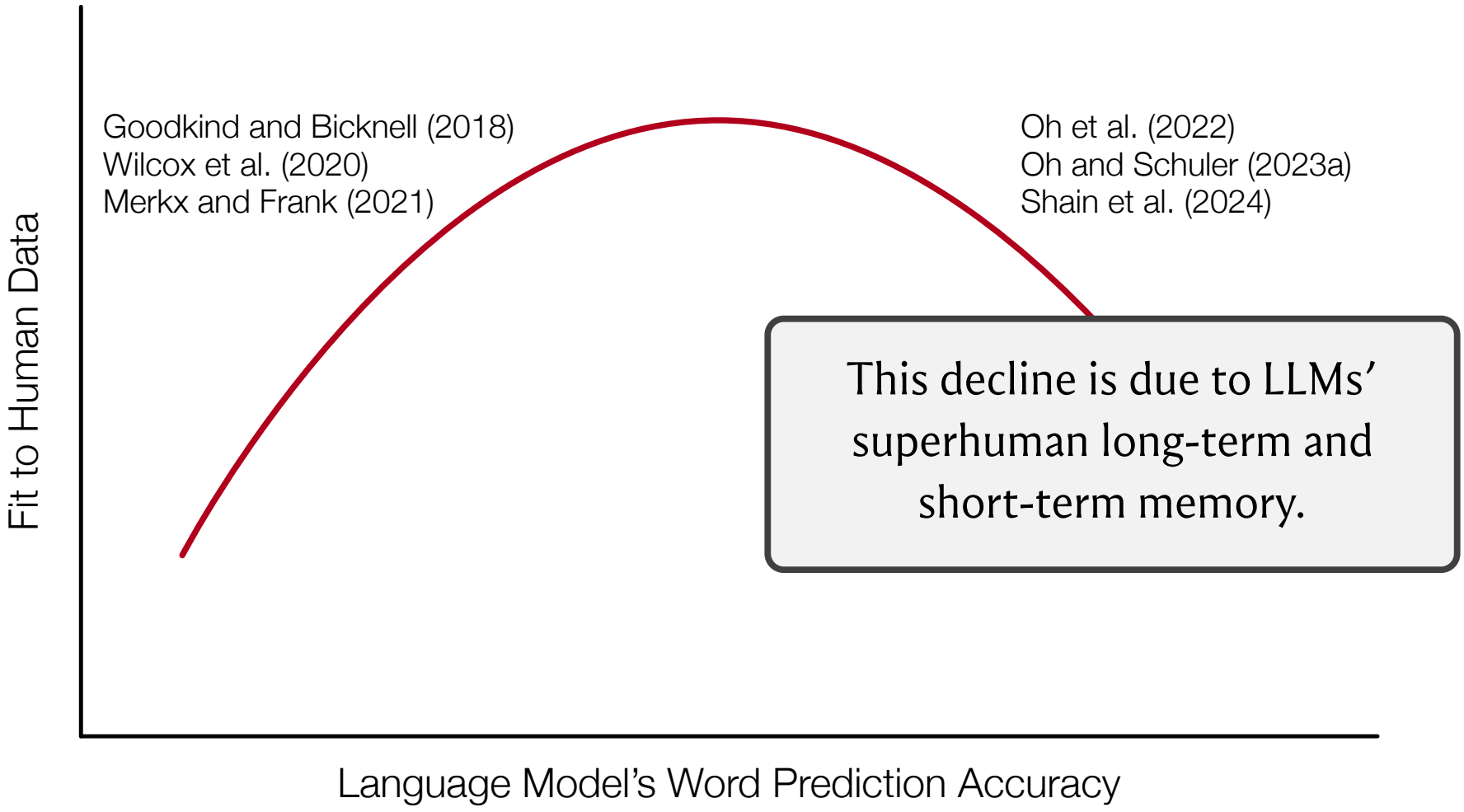
LLMs have become superhuman as models of prediction



LLMs have become superhuman as models of prediction



LLMs have become superhuman as models of prediction



2. Superhuman long-term memory

Humans naturally forget what they read or heard over time;
LLMs are unlikely to forget training examples they have seen.

The role of background knowledge in language processing

Joshua Kimmich was born in the city of _____

The role of background knowledge in language processing

Joshua Kimmich was born in the city of _____

- If you know his birthplace, very easy to predict; if not, very difficult.

The role of background knowledge in language processing

Joshua Kimmich was born in the city of _____

- If you know his birthplace, very easy to predict; if not, very difficult.
- Background knowledge plays a big role in language comprehension.

  (Kintsch, 1998; R. Smith et al., 2021)

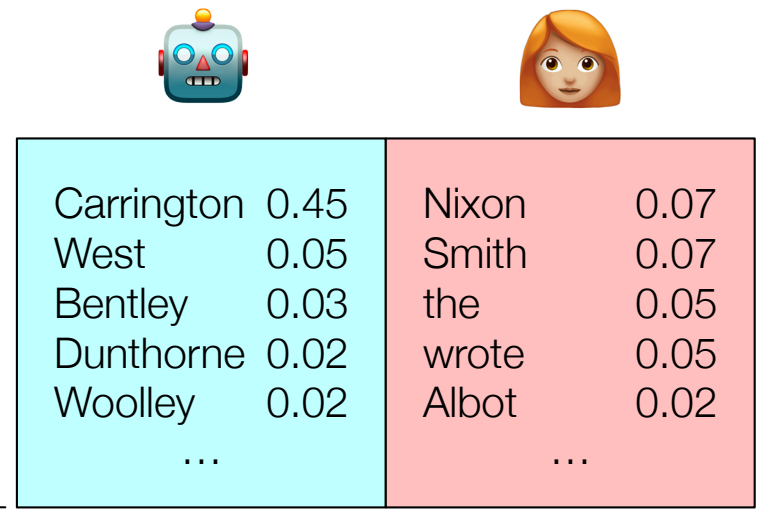
The role of background knowledge in language processing

Joshua Kimmich was born in the city of _____

- If you know his birthplace, very easy to predict; if not, very difficult.
- Background knowledge plays a big role in language comprehension.
  (Kintsch, 1998; R. Smith et al., 2021)
- We want a model that is consistent with a reader's knowledge.

LLMs know too much compared to humans

Two days later, the British astronomer Richard _____



LLMs know too much compared to humans

LLMs' training data contains a lot of factual information.

LLMs know too much compared to humans

LLMs' training data contains a lot of factual information.

> **Joshua Walter Kimmich** (German pronunciation: ['joːzʊ̯aː 'kɪmɪç];[5] born 8 February 1995) is a German professional footballer who plays as a defensive midfielder or right-back for Bundesliga club Bayern Munich and captains the Germany national team.[6] Known for his versatility, aggression and playmaking ability,[11] Kimmich has been compared with former Bayern Munich and Germany captain Philipp Lahm.[12]

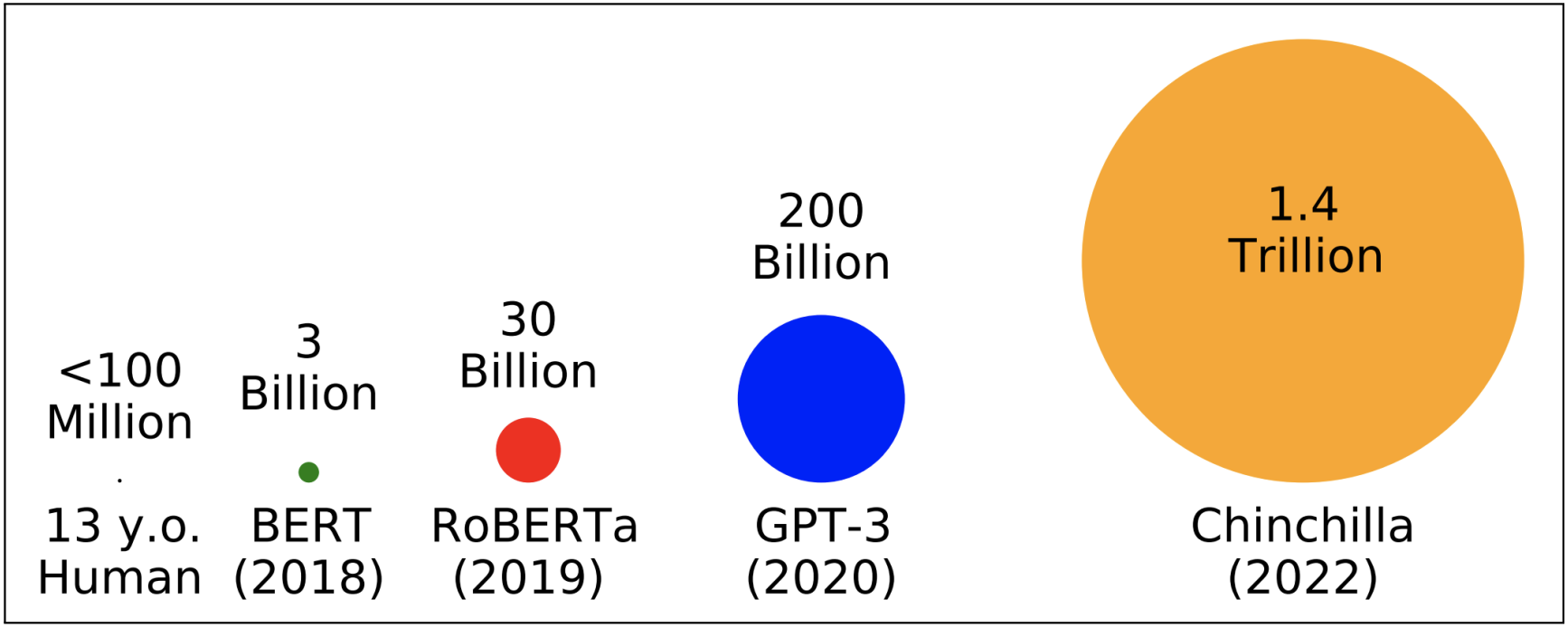
From Wikipedia

LLMs know too much compared to humans

LLMs are trained on massive amounts of such data.

LLMs know too much compared to humans

LLMs are trained on massive amounts of such data.



The BabyLM Challenge; Warstadt et al. (2023), Wilcox et al. (2025)

LLMs know too much compared to humans

LLMs do not easily forget training examples they encounter.

LLMs know too much compared to humans

LLMs do not easily forget training examples they encounter.

```
//
// Licensed under the Apache License,
// Version 2.0 (the "License");
// you may not use this file except in
// compliance with the License ...
```

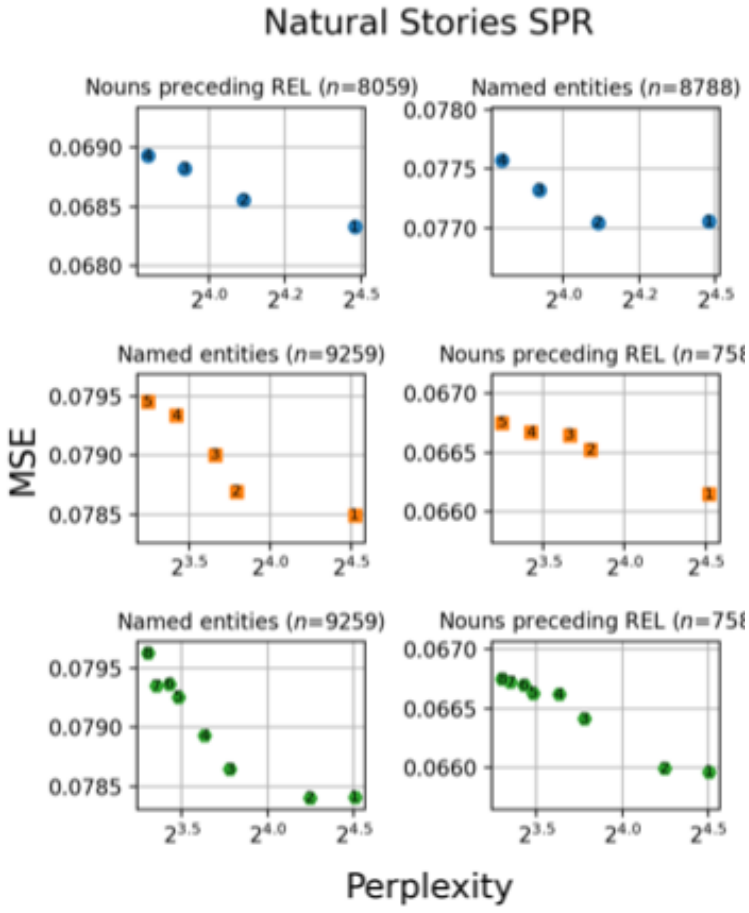
Example from Merrill et al. (2024)

Preliminary evidence that this is problematic

Misprediction of reading time by LLMs is the largest on named entities.

Preliminary evidence that this is problematic

Misprediction of reading time by LLMs is the largest on named entities.



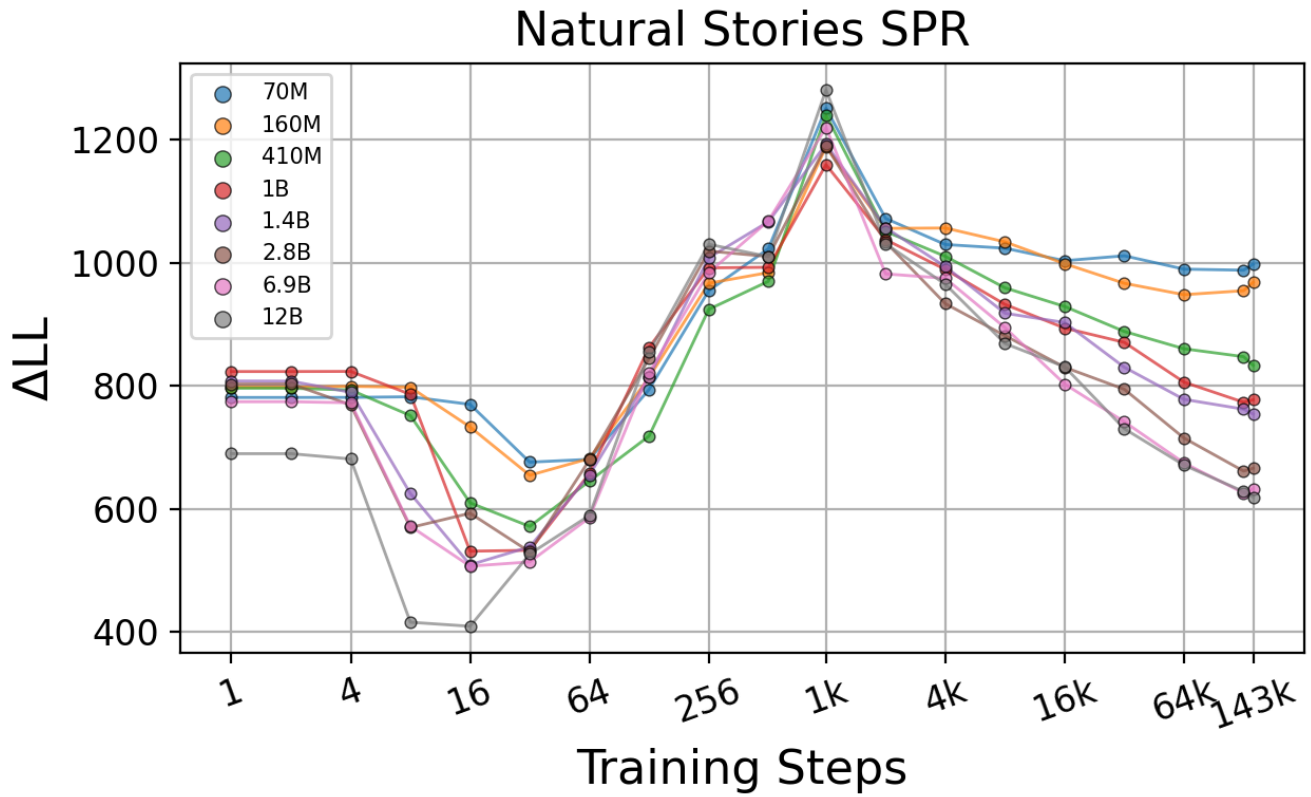
Oh and Schuler (2023a)

Preliminary evidence that this is problematic

LLMs trained on far less data provide better fits to reading time data.

Preliminary evidence that this is problematic

LLMs trained on far less data provide better fits to reading time data.



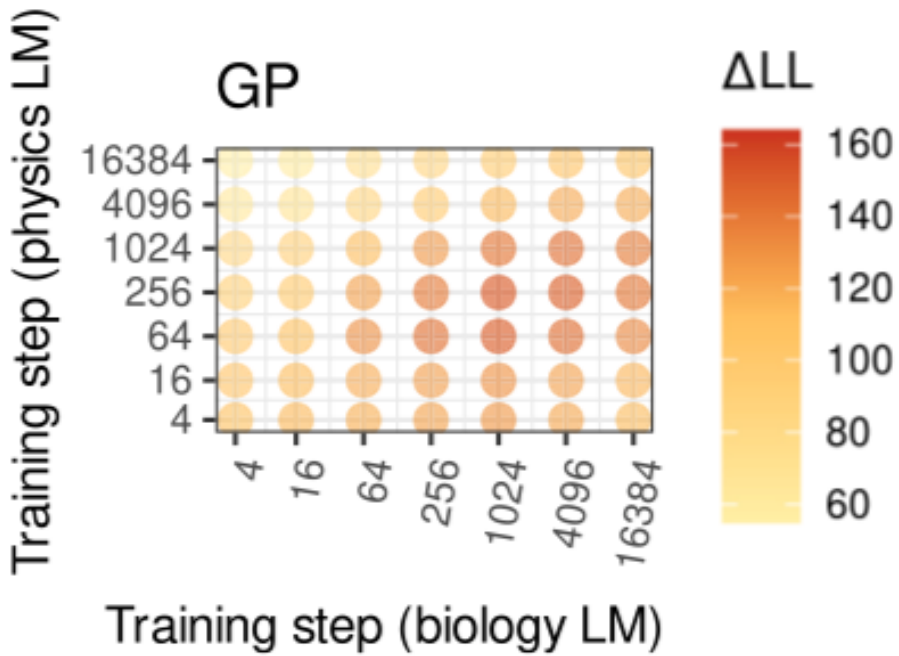
Oh and Schuler (2023b)

Preliminary evidence that this is problematic

LLMs that are aligned to readers' knowledge capture their reading behavior better, but too much training data has a deleterious effect.

Preliminary evidence that this is problematic

LLMs that are aligned to readers' knowledge capture their reading behavior better, but too much training data has a deleterious effect.



Škrjanec and Demberg (2026)

Other potentially problematic examples

Other potentially problematic examples

- Multiword expressions: *spill the beans*, *hit the sack*, ...

  (Rambelli et al., 2023)

Other potentially problematic examples

- Multiword expressions: *spill the beans*, *hit the sack*, ...

  (Rambelli et al., 2023)
- 'Famous' text: Harry Potter, Declaration of Independence, ...

Other potentially problematic examples

- Multiword expressions: *spill the beans*, *hit the sack*, ...

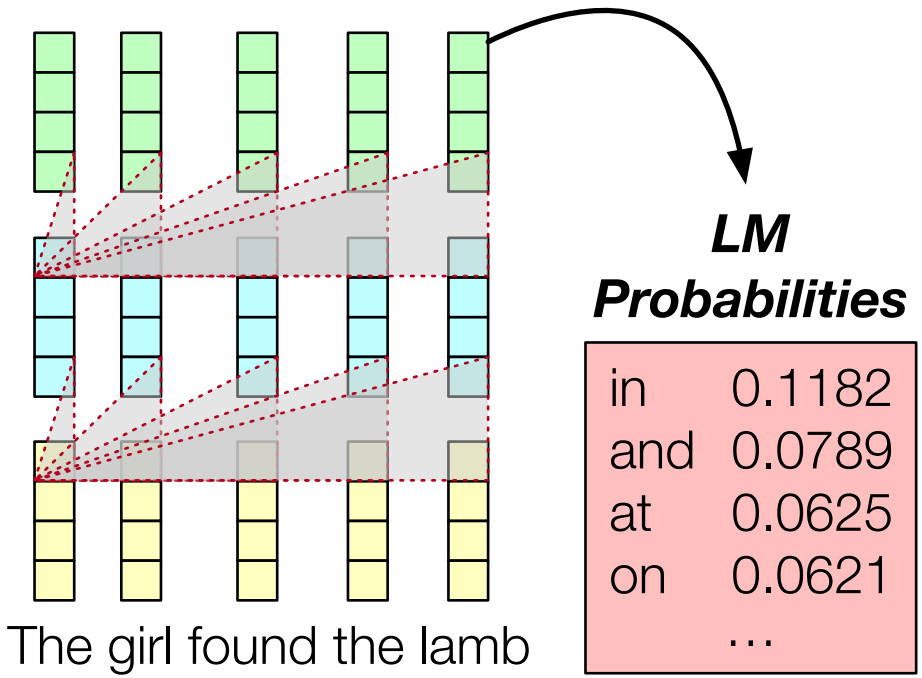
  (Rambelli et al., 2023)
- 'Famous' text: Harry Potter, Declaration of Independence, ...

- LLMs are likely to assign high probabilities to each word, which will inevitably underestimate readers' difficulty in processing them.

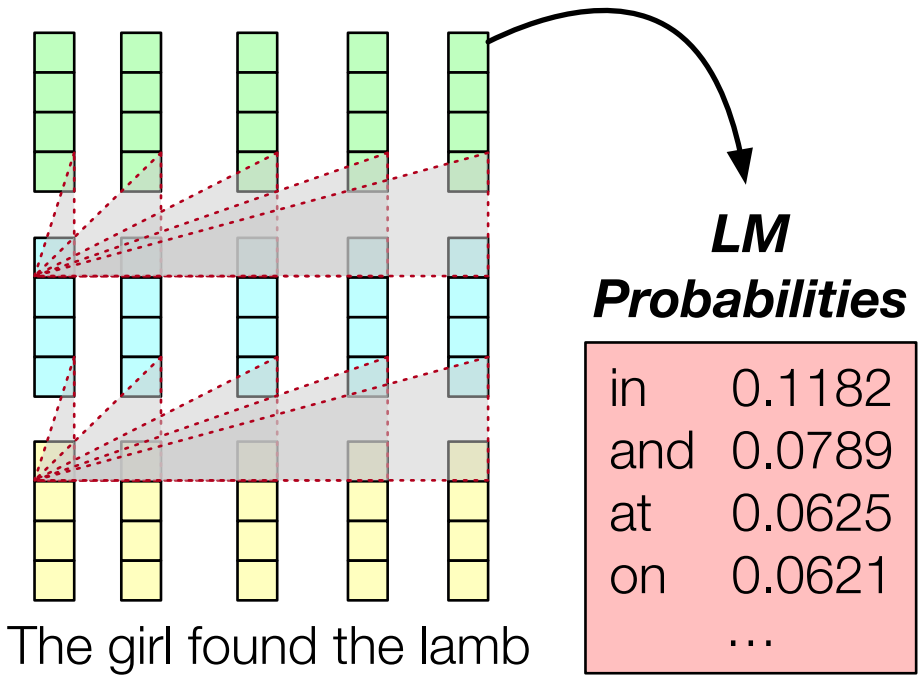
3. Superhuman short-term memory

Humans are subject to limitations in working memory;
LLMs have near-perfect access to large amounts of input text.

Language processing in language models



Language processing in language models



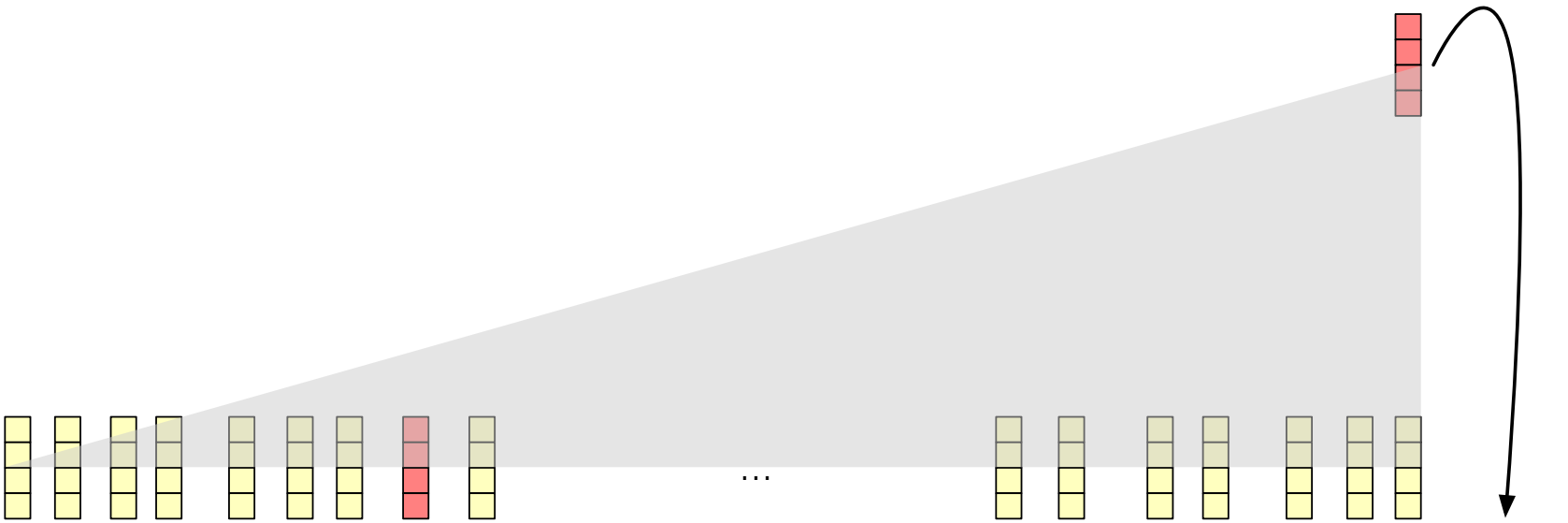
Transformers (Vaswani et al., 2017) can 'look at' all of the input to make next-word predictions.

A more extreme example

LLMs have such perfect access to tens of thousands to millions of words.

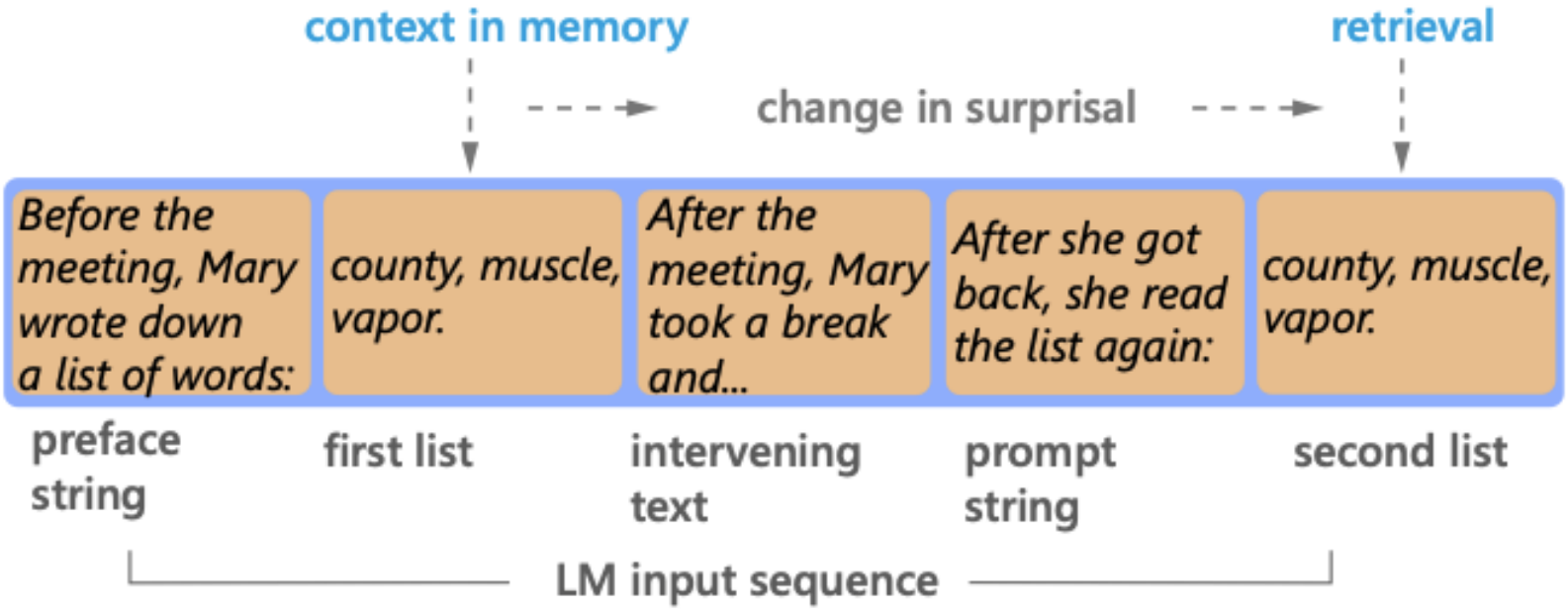
A more extreme example

LLMs have such perfect access to tens of thousands to millions of words.

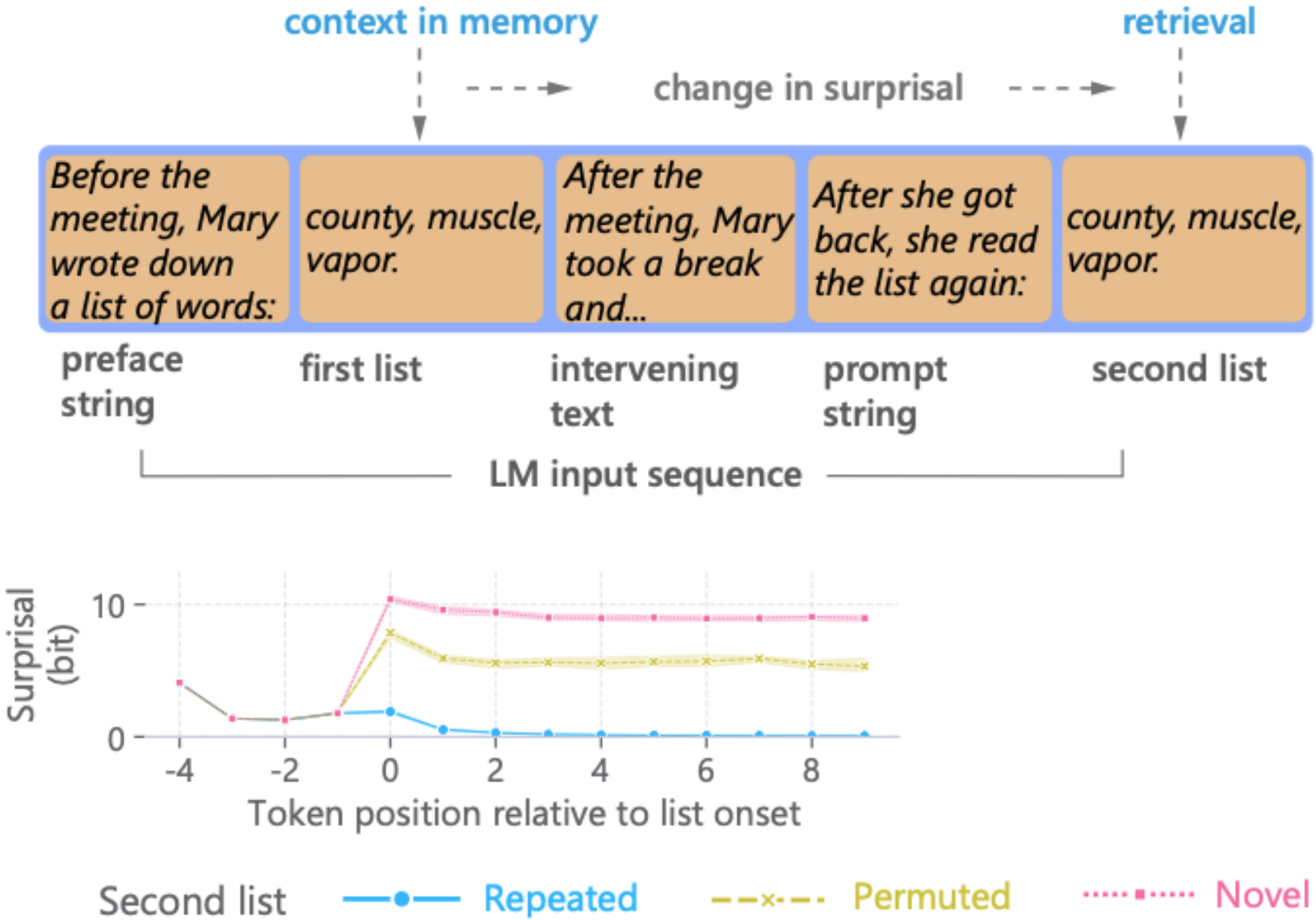


Transformers learn to copy earlier text

Transformers learn to copy earlier text



Transformers learn to copy earlier text



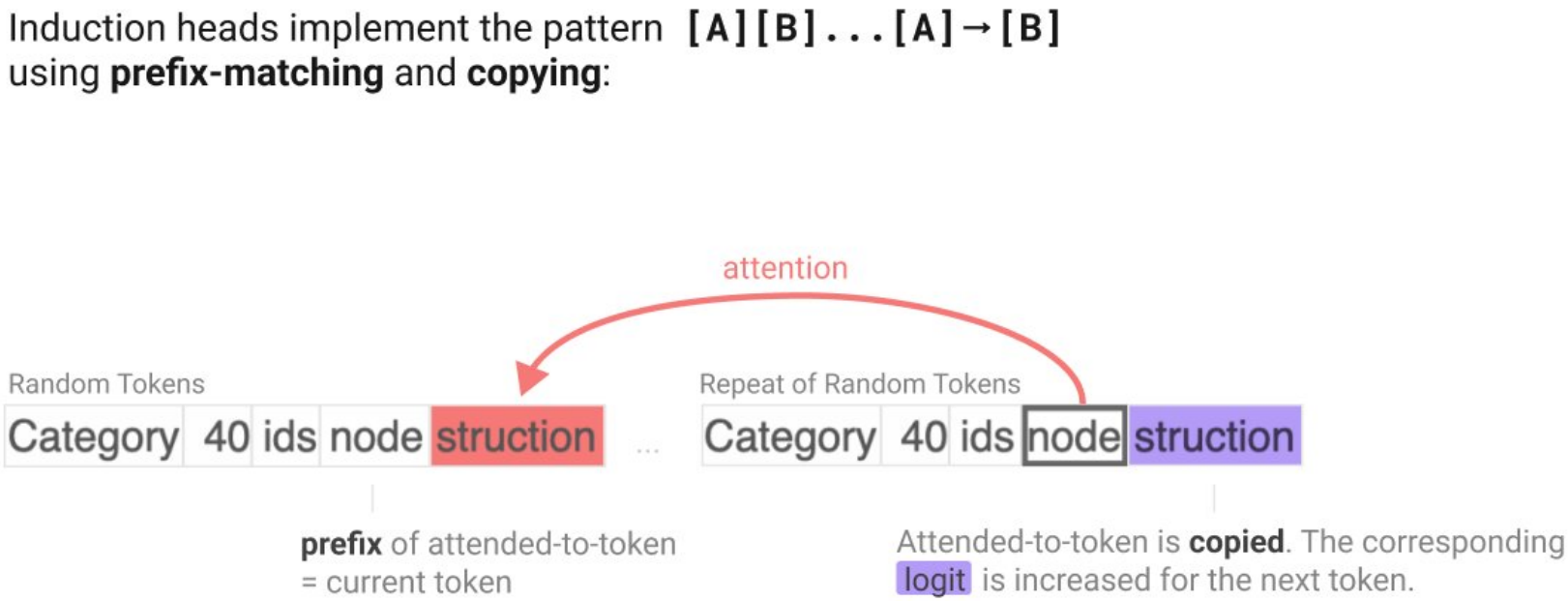
Armeni et al. (2022)

Transformers learn to copy earlier text

This is thought to be due to specialized ‘induction heads’.

Transformers learn to copy earlier text

This is thought to be due to specialized ‘induction heads’.



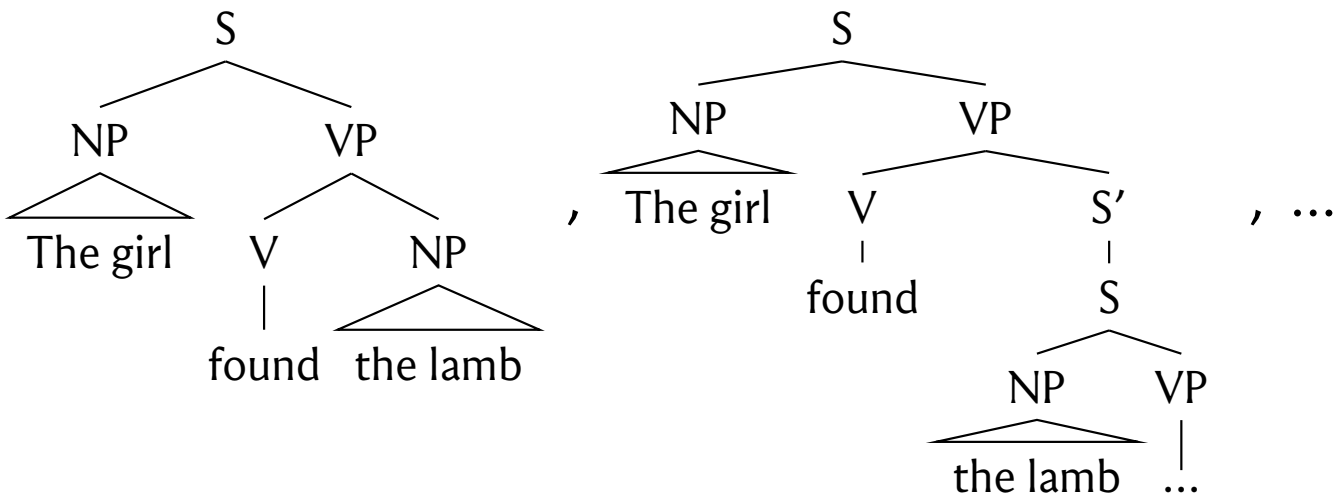
Olsson et al. (2022); see also Bietti et al. (2023) and Jelassi et al. (2024)

Limited vs. full parallelism

The girl found the lamb ...

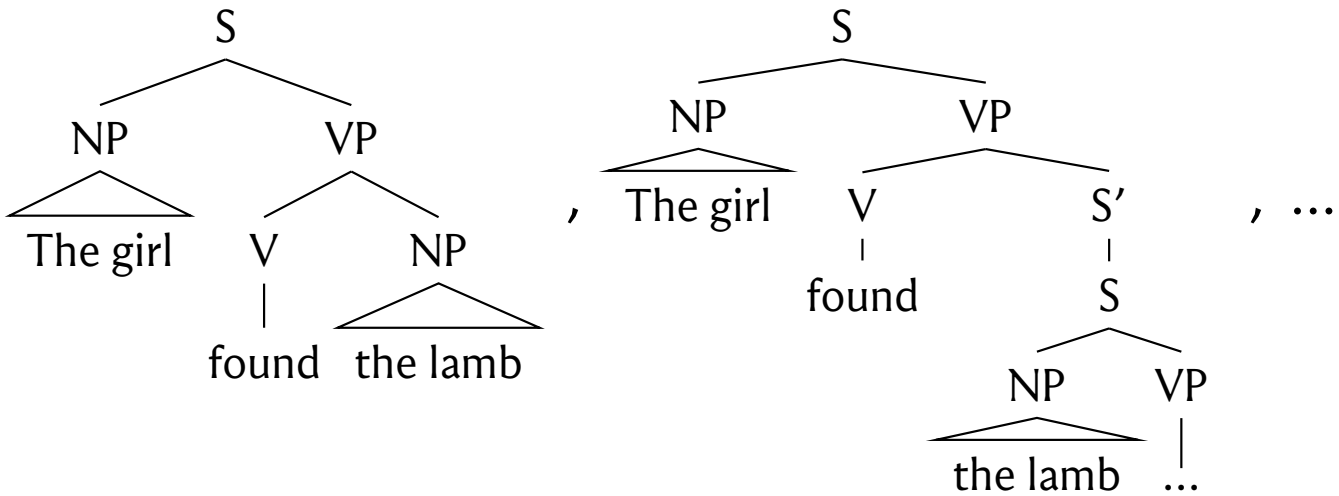
Limited vs. full parallelism

The girl found the lamb ...



Limited vs. full parallelism

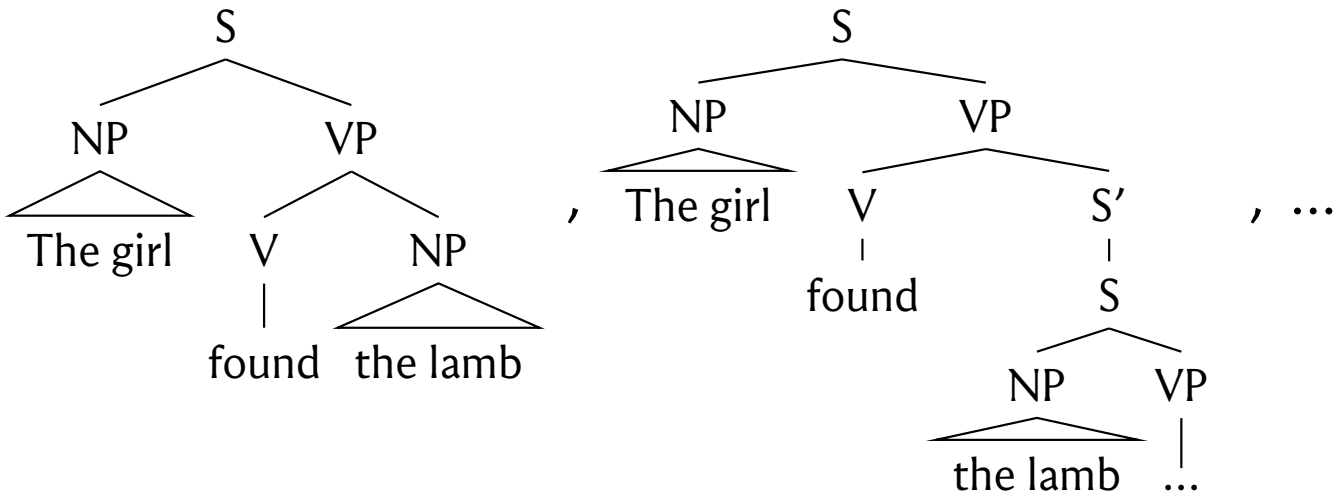
The girl found the lamb ...



- People cannot seem to hold many parses in memory at once.

Limited vs. full parallelism

The girl found the lamb …



- People cannot seem to hold many parses in memory at once.
- LLMs seem to be able to entertain many parses.

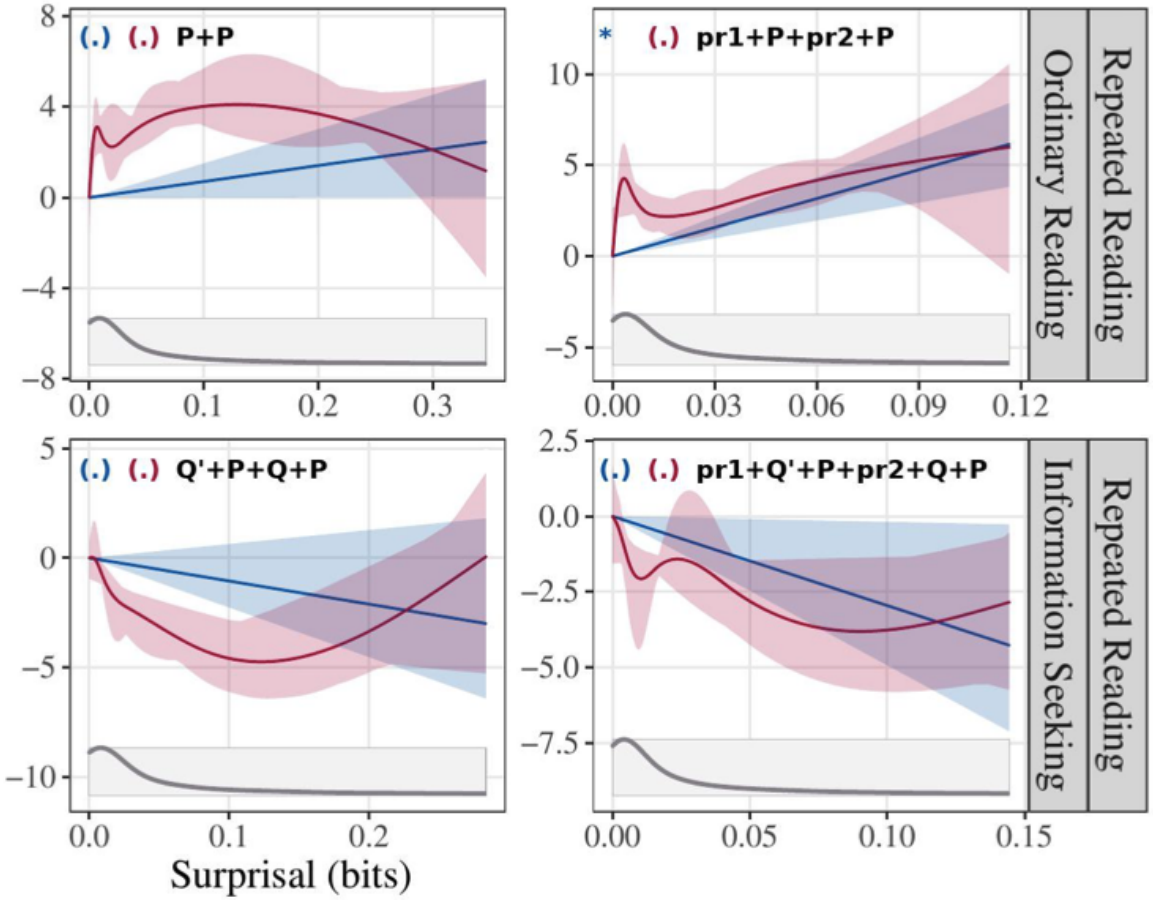
  (Huang et al., 2024; Timkey et al., 2025)

Preliminary evidence that this is problematic

LLMs completely fail to capture ‘repeated reading’ behavior, as they predict the repeated passage with near-perfect accuracy.

Preliminary evidence that this is problematic

LLMs completely fail to capture ‘repeated reading’ behavior, as they predict the repeated passage with near-perfect accuracy.



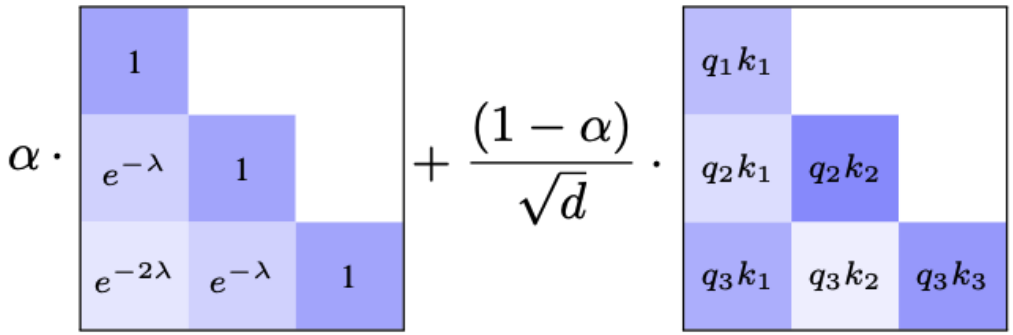
Gruteke Klein et al. (2024)

Preliminary evidence that this is problematic

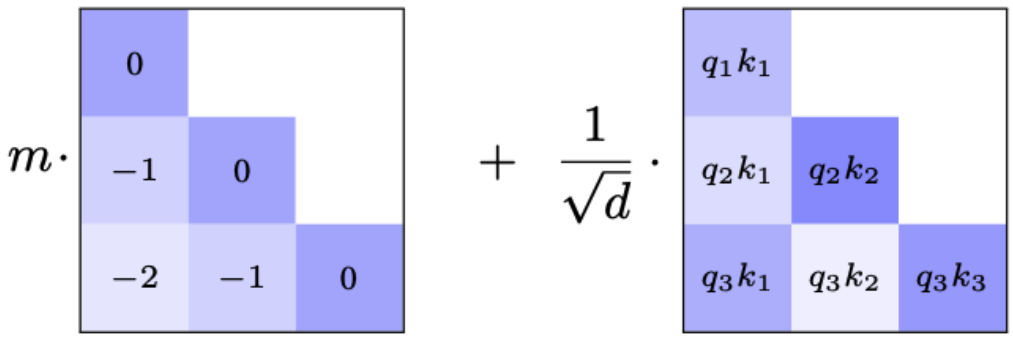
Incorporating a recency bias by intervening on attention weights improves fit to reading time data.

Preliminary evidence that this is problematic

Incorporating a recency bias by intervening on attention weights improves fit to reading time data.



(a) de Varda and Marelli (2024)



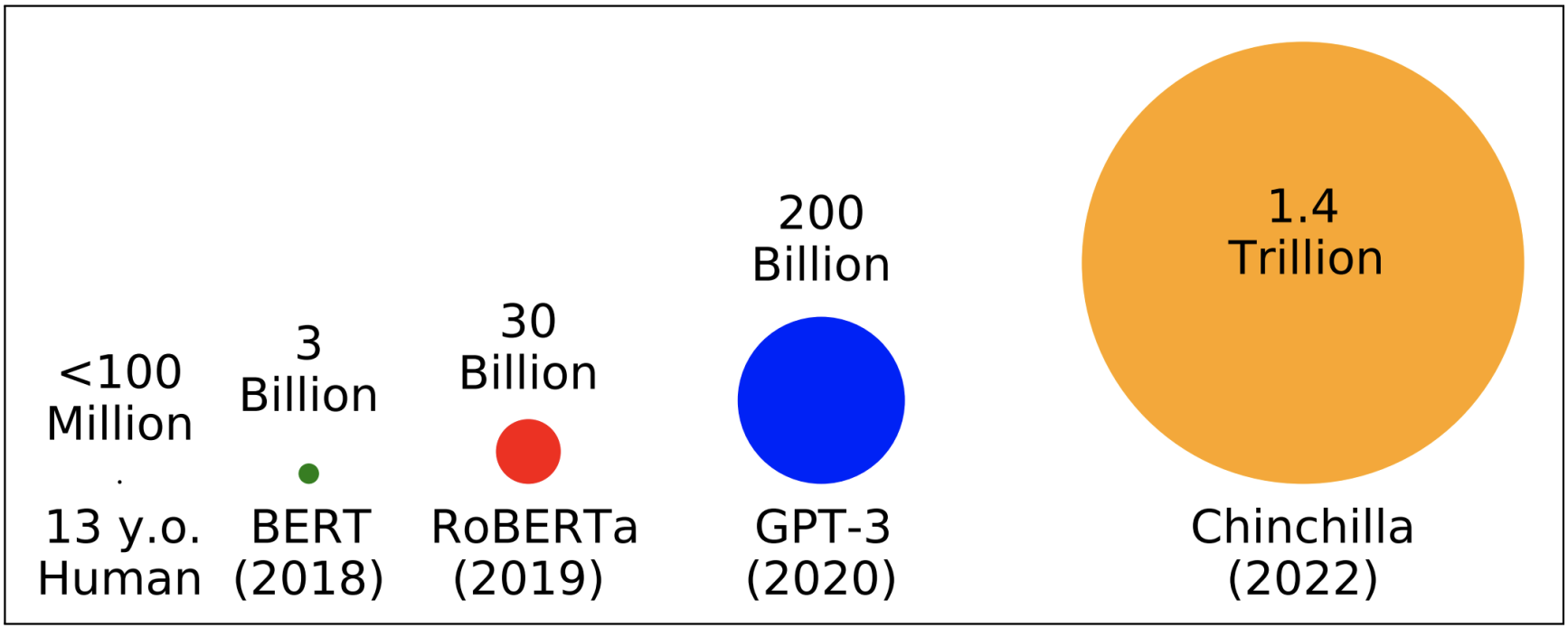
(b) Press et al. (2022)

Clark, Oh, and Schuler (2025)

4. What should we do about it?

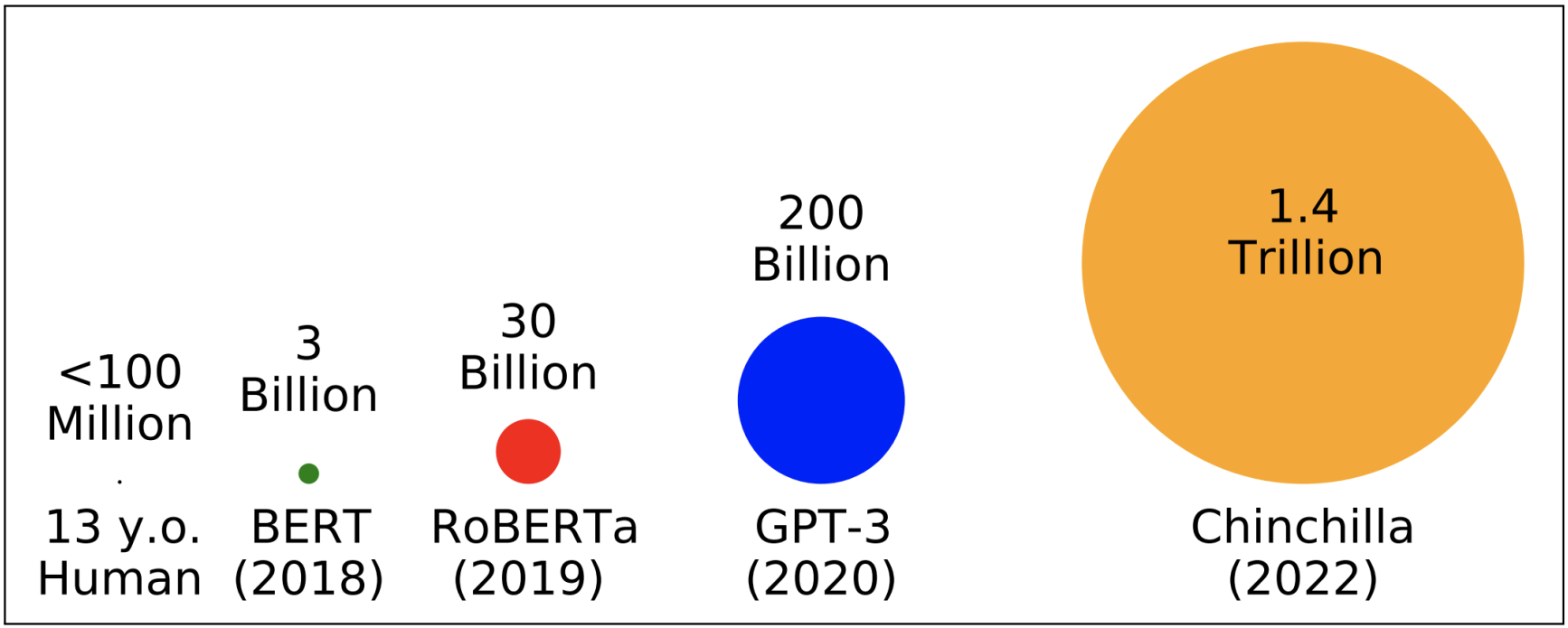
We can apply many language modeling techniques to address these issues—but in a different way from NLP applications.

Train models on human-scale training data



The BabyLM Challenge; Warstadt et al. (2023), Wilcox et al. (2025)

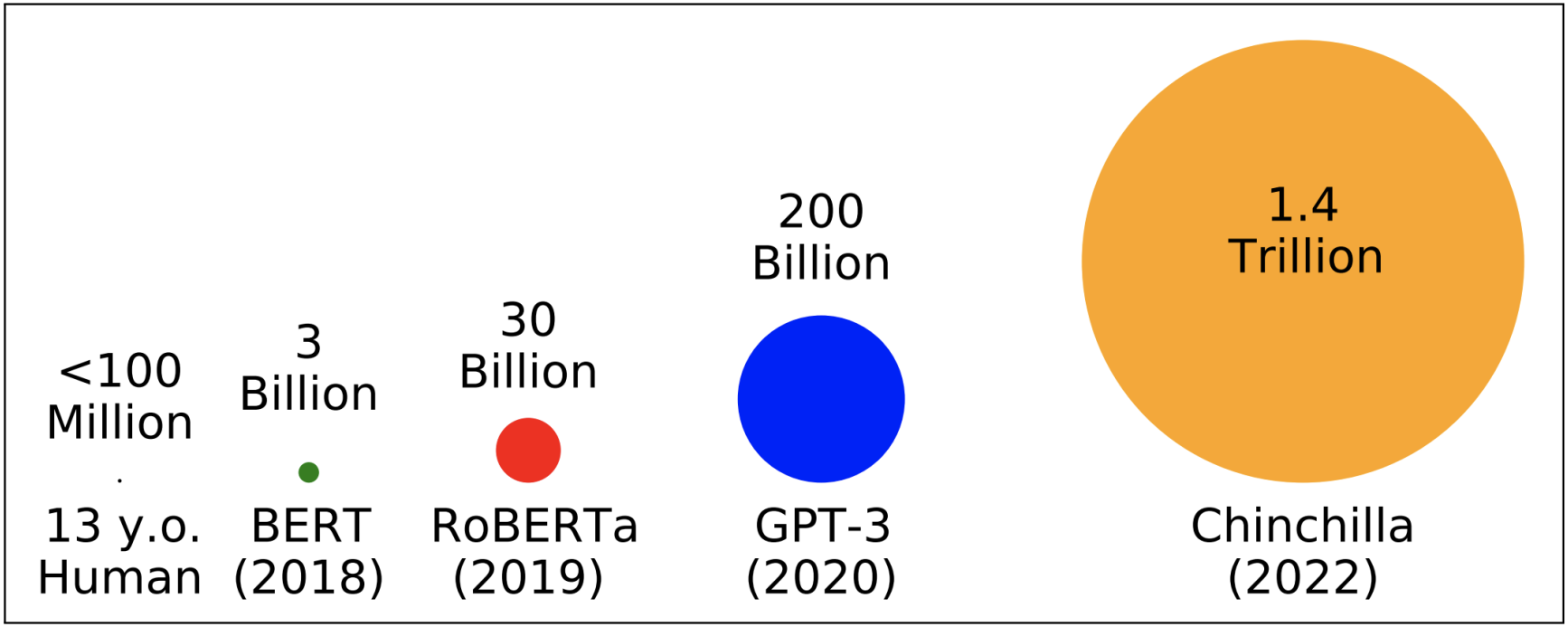
Train models on human-scale training data



The BabyLM Challenge; Warstadt et al. (2023), Wilcox et al. (2025)

- Simply decreasing the amount of data will probably not suffice.

Train models on human-scale training data



The BabyLM Challenge; Warstadt et al. (2023), Wilcox et al. (2025)

- Simply decreasing the amount of data will probably not suffice.
- Humans learn through other sensory experience. (Bender & Koller, 2020)

What about the training objective?

The training objective of LLMs incentivizes lexically sharp predictions.

What about the training objective?

The training objective of LLMs incentivizes lexically sharp predictions.

Joshua Kimmich was born in the city of Rottweil.

What about the training objective?

The training objective of LLMs incentivizes lexically sharp predictions.

Joshua Kimmich was born in the city of Rottweil.

- LLMs: **Rottweil** Villingen Ravensburg Überlingen

What about the training objective?

The training objective of LLMs incentivizes lexically sharp predictions.

Joshua Kimmich was born in the city of Rottweil.

- LLMs: **Rottweil** Villingen Ravensburg Überlingen
- Readers: Rottweil Villingen Ravensburg Überlingen

(Roland et al., 2012)

What about the training objective?

The training objective of LLMs incentivizes lexically sharp predictions.

Joshua Kimmich was born in the city of Rottweil.

- LLMs: **Rottweil** Villingen Ravensburg Überlingen
- Readers: Rottweil Villingen Ravensburg Überlingen

(Roland et al., 2012)

- We should experiment with alternative training objectives; currently, LLMs do nothing explicitly but prediction.

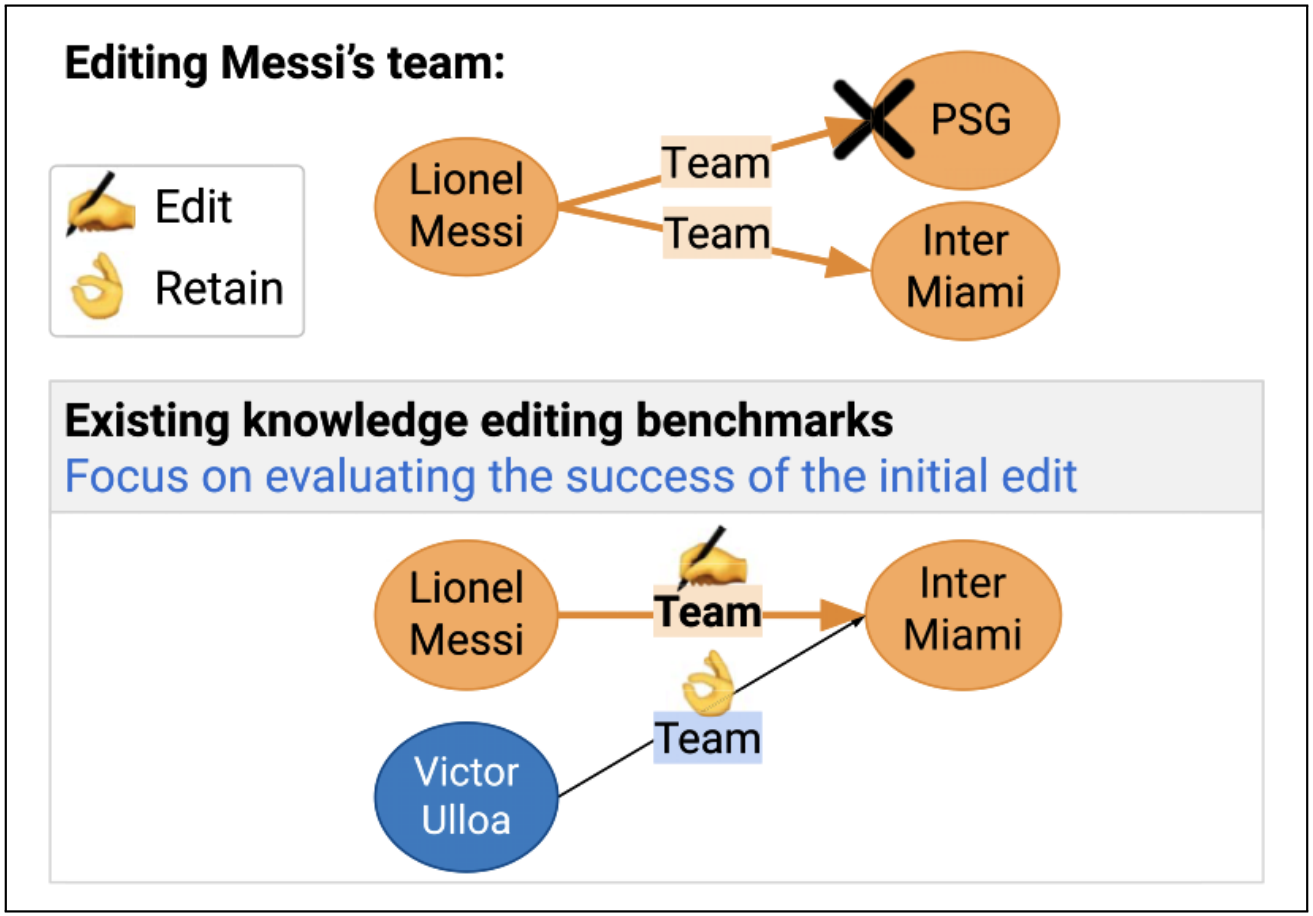
  (cf. structure building, Nair & Phillips, 2026; Slaats & Martin, 2025)

Apply post-hoc approaches

Knowledge editing can be applied to erase knowledge readers don't have.

Apply post-hoc approaches

Knowledge editing can be applied to erase knowledge readers don't have.

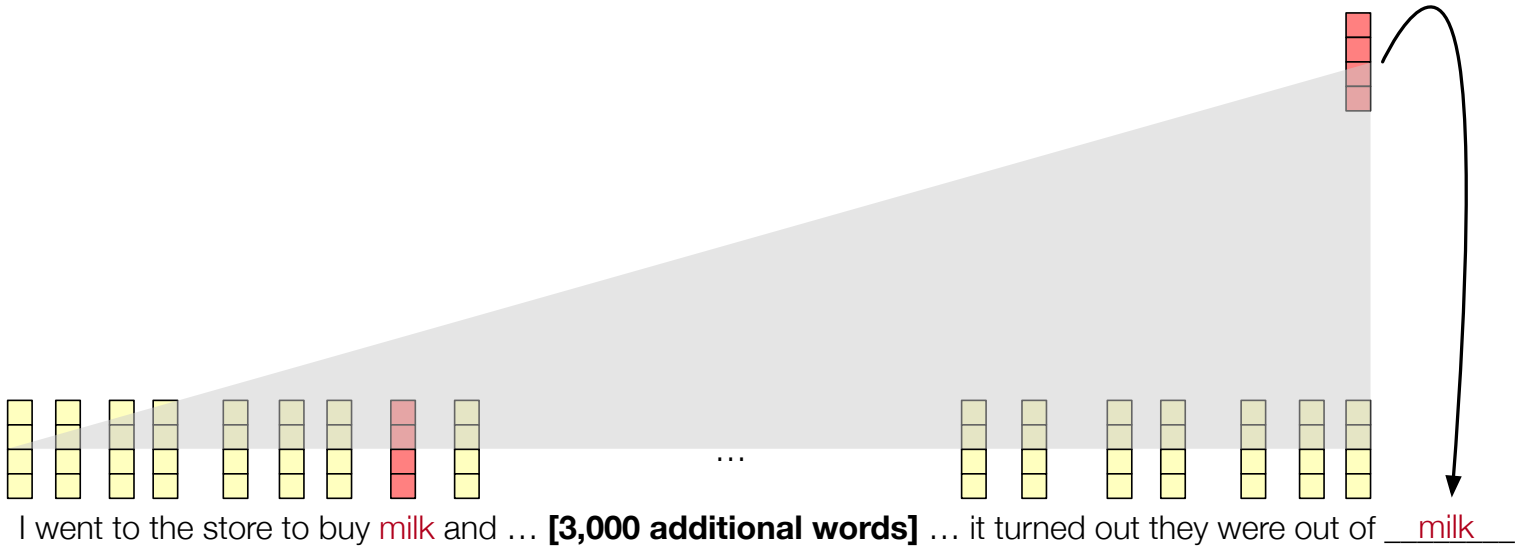


Cohen et al. (2024)

Train models with alternative architectures

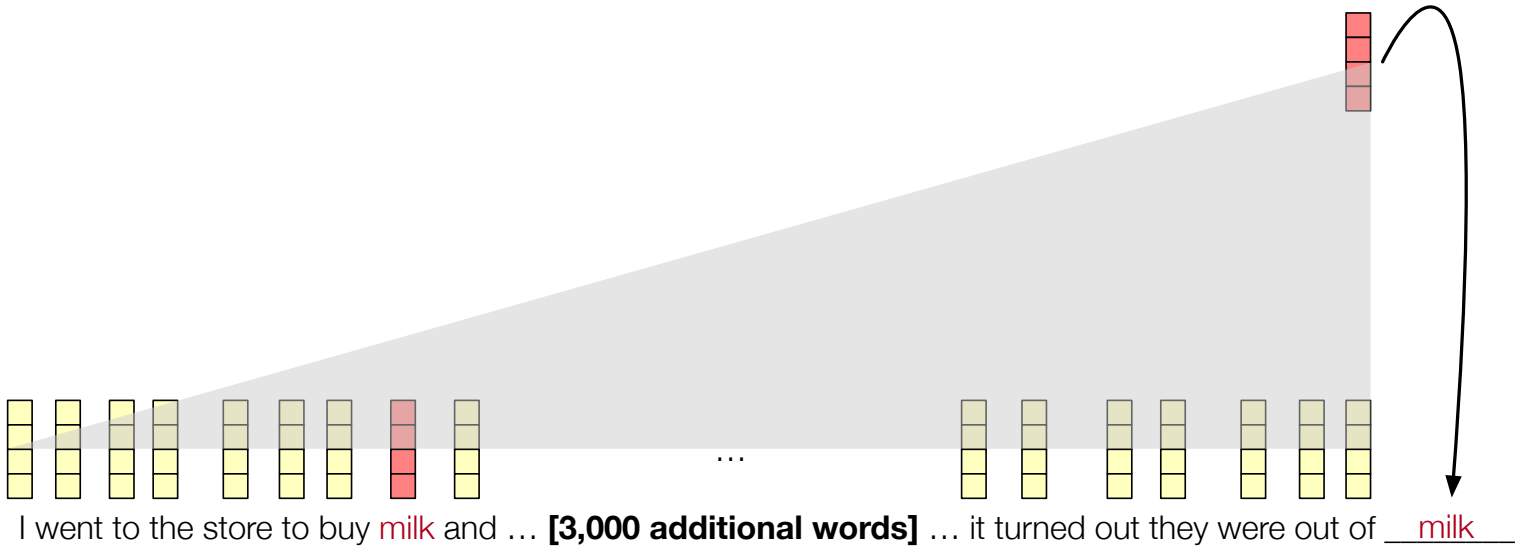
Train models with alternative architectures

We need to degrade the input in a way that resembles human memory.

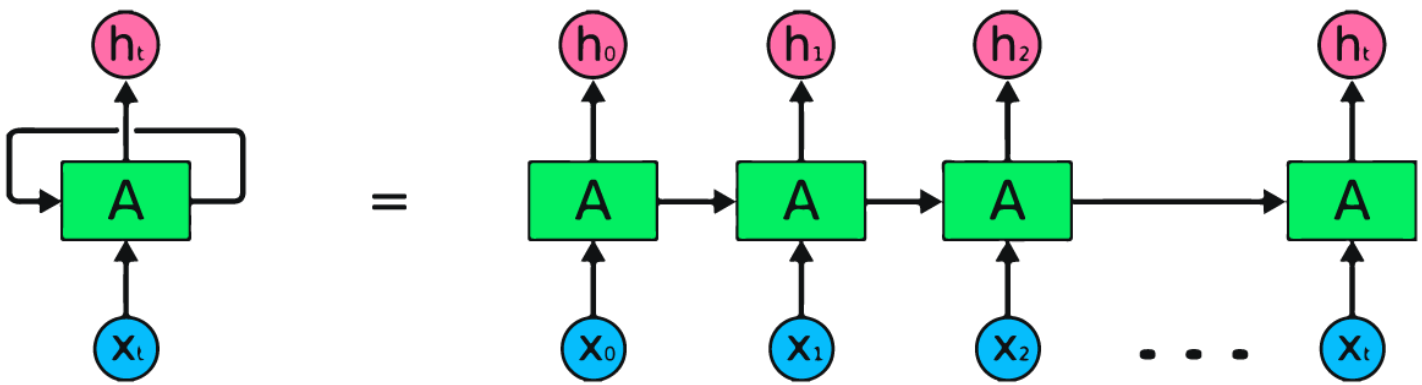


Train models with alternative architectures

We need to degrade the input in a way that resembles human memory.



Placing a renewed focus on recurrence could be one way.



Conduct targeted human experiments for benchmarking

Conduct targeted human experiments for benchmarking

- How likely are people to know that Kimmich was born in Rottweil? How does this influence their reading behavior?

Conduct targeted human experiments for benchmarking

- How likely are people to know that Kimmich was born in Rottweil? How does this influence their reading behavior?
- While reading an extended discourse, how do limitations in working memory influence reading behavior?

Conduct targeted human experiments for benchmarking

- How likely are people to know that Kimmich was born in Rottweil? How does this influence their reading behavior?
- While reading an extended discourse, how do limitations in working memory influence reading behavior?
- How can we study the linguistic predictions that people make more directly through the cloze task, brain responses, etc.?

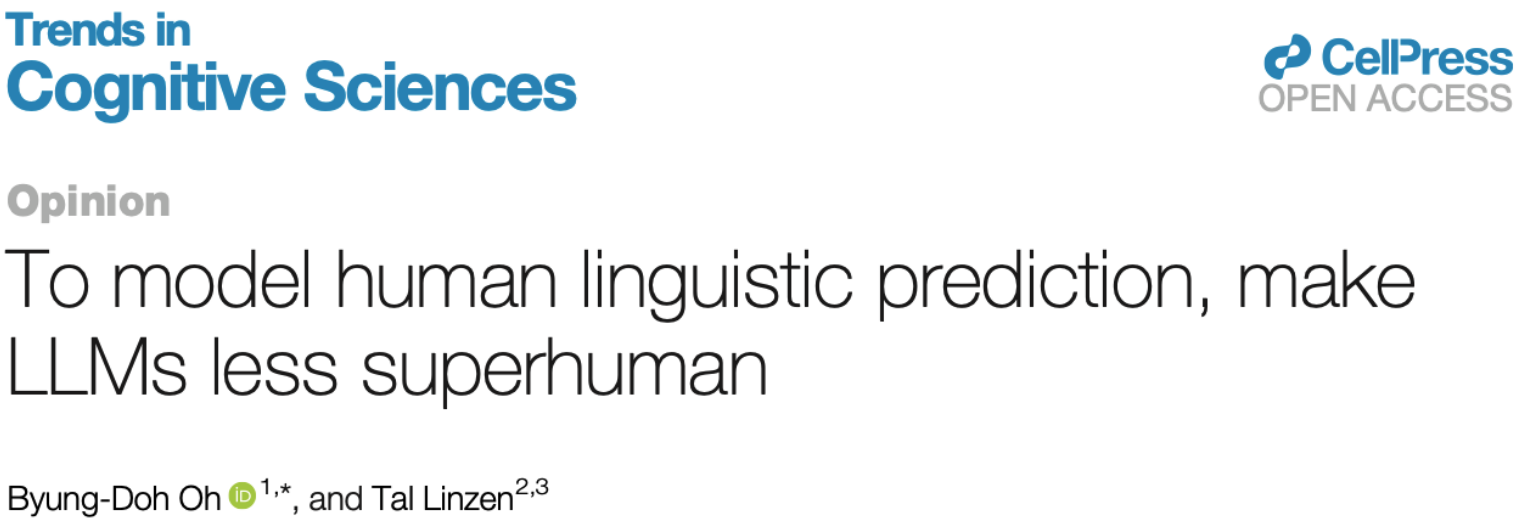
5. Conclusion

LLMs hold promise for psycholinguistics, but they need to be made less superhuman through language modeling techniques.

5. Conclusion

LLMs hold promise for psycholinguistics, but they need to be made less superhuman through language modeling techniques.

Thank you for listening!



References I

Armeni, K., Honey, C., & Linzen, T. (2022). Characterizing verbatim short-term memory in neural language models. *Proceedings of the 26th Conference on Computational Natural Language Learning*, 405–424.

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, *17*(3), 364–390.

Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5185–5198.

Bietti, A., Cabannes, V., Bouchacourt, D., Jegou, H., & Bottou, L. (2023). Birth of a transformer: A memory viewpoint. *Advances in Neural Information Processing Systems*, *37*, 1560–1588.

Clark, C., Oh, B.-D., & Schuler, W. (2025). Linear recency bias during training improves transformers' fit to reading times. *Proceedings of the 31st International Conference on Computational Linguistics*, 7735–7747.

Cohen, R., Biran, E., Yoran, O., Globerson, A., & Geva, M. (2024). Evaluating the ripple effects of knowledge editing in language models. *Transactions of the Association for Computational Linguistics*, *12*, 283–298.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, *20*(6), 641–655.

Frisson, S., Harvey, D. R., & Staub, A. (2017). No prediction error cost in reading: Evidence from eye movements. *Journal of Memory and Language*, *95*, 200–214.

References II

Gruteke Klein, K., Meiri, Y., Shubi, O., & Berzak, Y. (2024). The effect of surprisal on reading times in information seeking and repeated reading. *Proceedings of the 28th Conference on Computational Natural Language Learning*, 219–230.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics*.

Huang, K.-J., Arehalli, S., Kugemoto, M., Muxica, C., Prasad, G., Dillon, B., & Linzen, T. (2024). Large-scale benchmark yields no evidence that language model surprisal explains syntactic disambiguation difficulty. *Journal of Memory and Language*, *137*, 104510.

Jelassi, S., Brandfonbrener, D., Kakade, S. M., & Malach, E. (2024). Repeat after me: Transformers are better than state space models at copying. *Proceedings of the 41st International Conference on Machine Learning*, 21502–21521.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177.

Merrill, W., Smith, N. A., & Elazar, Y. (2024). Evaluating n-gram novelty of language models using Rusty-DAWG. *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, 14459–14473.

References III

Nair, S., & Phillips, C. (2026). Across the levels of analysis: Explaining predictive processing in humans requires more than machine-estimated probabilities. *arXiv preprint, arXiv:2604.09466*.

Oh, B.-D., & Schuler, W. (2023a). Why does surprisal from larger Transformer-based language models provide a poorer fit to human reading times? *Transactions of the Association for Computational Linguistics, 11*, 336–350.

Oh, B.-D., & Schuler, W. (2023b). Transformer-based language model surprisal predicts human reading times best with about two billion training tokens. *Findings of the Association for Computational Linguistics: EMNLP 2023*, 1915–1921.

Olsson, C., Elhage, N., Nanda, N., Joseph, N., DasSarma, N., Henighan, T., Mann, B., Askell, A., Bai, Y., Chen, A., Conerly, T., Drain, D., Ganguli, D., Hatfield-Dodds, Z., Hernandez, D., Johnston, S., Jones, A., Kernion, J., Lovitt, L., … Olah, C. (2022). In-context learning and induction heads. *arXiv preprint, arXiv:2209.11895*.

Rambelli, G., Chersoni, E., Senaldi, M. S. G., Blache, P., & Lenci, A. (2023). Are frequent phrases directly retrieved like idioms? An investigation with self-paced reading and language models. *Proceedings of the 19th Workshop on Multiword Expressions*, 87–98.

Roland, D., Yun, H., Koenig, J.-P., & Mauner, G. (2012). Semantic similarity, predictability, and models of sentence processing. *Cognition, 122*(3), 267–279.

References IV

Schrimpf, M., Blank, I. A., Tuckute, G., Kauf, C., Hosseini, E. A., Kanwisher, N., Tenenbaum, J. B., & Fedorenko, E. (2021). The neural architecture of language: Integrative modeling converges on predictive processing. *Proceedings of the National Academy of Sciences, 118*(45), e2105646118.

Shain, C., Meister, C., Pimentel, T., Cotterell, R., & Levy, R. (2024). Large-scale evidence for logarithmic effects of word predictability on reading time. *Proceedings of the National Academy of Sciences, 121*(10), e2307876121.

Škrjanec, I., & Demberg, V. (2026). Language models that match reader experience are better predictors of reading times. *Journal of Memory and Language, 146*, 104677.

Slaats, S., & Martin, A. E. (2025). What's surprising about surprisal. *Computational Brain & Behavior, 8*, 233–248.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition, 128*, 302–319.

Smith, R., Snow, P., Serry, T., & Hammond, L. (2021). The role of background knowledge in reading comprehension: A critical review. *Reading Psychology, 42*(3), 214–240.

Timkey, W., Huang, K.-J., Oh, B.-D., Prasad, G., Arehalli, S., Linzen, T., & Dillon, B. (2025). Eye movements reveal a dissociation between prediction and structural processing in language comprehension. *PsyArXiv, eq2ra_v1*.

References V


Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30*, 6000–6010.


Warstadt, A., Mueller, A., Choshen, L., Wilcox, E., Zhuang, C., Ciro, J., Mosquera, R., Paranjabe, B., Williams, A., Linzen, T., & Cotterell, R. (Eds.). (2023). *Proceedings of the BabyLM Challenge at the 27th Conference on Computational Natural Language Learning*.


Wilcox, E. G., Hu, M. Y., Mueller, A., Warstadt, A., Choshen, L., Zhuang, C., Williams, A., Cotterell, R., & Linzen, T. (2025). Bigger is not always better: The importance of human-scale language modeling for psycholinguistics. *Journal of Memory and Language, 144*, 104650.