One or Several Wolves?
CS6460 Fall 2023

Lila James
kjames64@gatech.edu

Abstract— A call for a new pedagogy is demanded by Generative Pre-trained Transformers. Prompts to AI are quickly becoming how students interface with the learning goals of a course. This paper examines the effects Generative AI has on the production of knowledge. By examining an AI agent's interaction with the goal state(s) of an online course, generative ensemble agents will be explored as a path to help bridge the education quality-gap between socio-economic classes through bespoke student tutors.

1 Introduction

The efficacy around providing new opportunities for disadvantaged communities came to the spotlight during the Covid-19 pandemic (Reich, J., et. Al: 2019, J. Shankar, et. al: 2021). For the first time in history, online education became the societal norm for the Western World (Zheng, Bender, & Lyon, 2021). This set the stage for a unique glimpse of how education, online, works in different communities, and the disparities between them.

Amazingly, this event overlaps with the explosion of Foundational AI models[1] (Bommasani et al: 2022). As such, this paper first presents a literature review of MOOCs as a flashpoint for analysis of disparities in education. The solution to many issues found in MOOCs will be presented by an accompanying analysis about the diversity of knowledge production within Foundational models, and how this may provide an *in* for revolutionary efficacy in education, where providing community and connection for education is paramount to teaching.

It is through connecting these two historical events that a unique opportunity presents itself. By merging the promises of MOOCs with the strengths of generative AI, educators may craft bespoke tools for the unique needs of all their students. The topic of this paper is measuring the diversity of knowledge produced from these tools to help ensure an entry point to learning for all students , of all abilities, and all skill levels.

[1] https://huggingface.co/sentence-transformers

2 MOOCs

MOOCs have had many promises including being a force of equality when access to education is concerned. The data behind these claims, however, needs to be examined in order to fully understand the nuances rippling into and from the societies MOOCs aim to serve. As such, the context of this *current work* analysis is during and after Covid-19. As this is seen as an inseparable change in the field of opportunities.

A *digital divide,* however, is replicated from the tangible inequalities found in society, which is mirrored in MOOCs and Foundational models alike. For example, not everyone has access to reliable internet, necessary devices, or environments that are safe or conducive to education- in fact, MOOCs may actually reify the educational divide amongst socio-economic classes, while providing opportunities to the already privileged (Zhenghao et. al: 2015, Reich, J& Ruipérez-Valiente, J. A: 2019). During Covid-19, however, these inequalities became center stage in the world of education as learning had to happen online, or not at all. There is emerging evidence that large undergraduate courses perform extremely well in MOOC settings, while smaller courses suffer (Reich, J& Ruipérez-Valiente, J. A: 2019). Again, painting a complex picture of the effects of MOOCs.

This study makes large assumptions, however, such as only analyzing currently enrolled undergraduate students. One could argue that this is studying only one side of the socio-economic divide, and that MOOCs in this context are simply designed for those already seeking higher education (Hansen & Reich, 2015). MOOCs are still young and a relatively new way of learning/teaching for the human species. As such, many equitable effects of MOOCs are explored to be not fully realized. These positives appear to be in their infant stages and their future expansion to all is ongoing.

With just using common logic, one can see the benefits of MOOCs: Asynchronous education allows for working adults to take courses, internet access is a door into any subject matter, and free MOOCs populate the internet. All these points, however, have been shown in the literature review to not be evenly spread amongst all.

The current state of MOOCs provides a new path upon which equitable education has started to form. While MOOCs themselves do not guarantee a democratizing force in education, they certainly provide a line of flight from the traditional education institution. The pros and cons of MOOCs highlight a promising space upon which to reflect and enact change. Borrowing from post-structuralist cultural theory, MOOCs have started a clear revolution in education, yet still are bound by traditional power structures and systemic failings plaguing western society, namely through the power structures and reification of socio-economic class (Foucault: 1978, Deleuze, G., & Guattari, F: 1987, Hansen & Reich, 2015, Reich et. al: 2019).

The potential remains high, yet many obstacles still need to be addressed. The two main points for improvement are listed below, and take their roots in *humanism*. That is, the power to create meaning and purpose stems from humans first, flowing into reality from us, of which MOOCs may provide (Sartre, 2007). The goal is to allow this meaning to spread equally amongst all classes.

1. Resource limitations, including time, money, and health, physical and mental (Reich, 2015, Reich et. al: 2019)
2. Pedagogical approaches need to be tailored to online education, and that not doing so may produce a negative MOOC experience (Shankar, K et. al: 2021)

These two points are apparent in foundational models as well. Which should not be surprising.

3 Foundational Models

One may easily look over to tools like ChatGPT, and think that the answer to education inequality has been presented: MOOCs & GPTs. This notion of hope is not misplaced, yet rather should come with a healthy understanding of the technologies behind Foundational Models, such as GPTs, and online education (Bommasani et al, 2021).

A Foundational model is considered self-supervising, such as using neural networks to create the embedding vectors for semantic understanding, are trained on a large corpus of data, and knowledge producing (Bommansani et al, 2021). The Large Language Model behind ChatGPT is a perfect example of a foundational model causing product troubles downstream. Luckily, these technologies at large use neural networks to essentially underpin the "thinking" of a model. It is difficult to separate Foundational Models from neural networks, which is worrisome, and yet also helpful as advances in this field will benefit countless.

Foundational models, however, were notoriously scrutinized in 2021 by teams[2] of researchers to be a force of homogenization- especially if not kept in check (Bommasani et al, 2021, Zheng: 2023, Zhang: 2022). Meaning, that without some grasp on how knowledge is produced and outputted to humans from machines, one may not hope for the best, but rather the worst. This is because the downstream effects of these models would be devastating as any and all processes

[2] CRFM, Stanford, and HAI

intertwined with a Foundational model would suffer. The inherent risk in Foundational models are quite simply summed up in four points presented in *On the Opportunities and Risks of Foundation Models 2022* and *Efficient Diversity-Driven Ensemble for Deep Networks*.

1: Foundational Models are built with and depended on neural networks

2: Neural Networks are largely homogeneous in their individual architectures[3], activation functions, and general *blackbox* knowledge production

3: Neural Networks at the scale of foundational models are infeasible, for now, except for those corporations with the resources to develop, maintain, and profit off said models.

4. Models can pass knowledge from one another, however, this too is largely a homogenous process.

In order to fully appreciate the risk of foundational models, one must understand how knowledge, diversity, and exchanges are produced and measured in the current state of neural networks.

4 A Neural Network Problem

Large neural networks are difficult and expensive to maintain and train. This is a possible cause for the Coporatization, and arguably neoliberalization, of said technologies away from Universities. Recent advances, however, have made evident that fine-tuning strategies used to bespoke networks may be replaced with Chains of Thought, CoT, as a prerequisite for achieving state-of-the-art (SOTA) benchmarks: for complex tasks requiring analysis, abstraction, and synthesis (Feng & Chen, 2023; Smola, Li, Zhang, & Zhang, 2022; Chi et al., 2022).

This opportunity presents itself as an educational equalizer as the base of an adaptable AI assistant for diverse and dynamic individuals. The goal is to avoid norm production in neural networks by intentionally introducing several *CoTs* into a single ensemble agent to help adjust said model to the unique user. The author here wishes to add to the definition of a *CoT,* most simply defined as prepending a prompt to elicit cohesion between questions/explanations[4], as also containing embedding parameters for large language models and source documents. To test this renewed definition, diversity within an ensemble agent dependent on Foundational models will be measured.

[3] Deep, wide, and deep-wide

[4] See appendix one for the Agent's explanation of Chains of Thoughts

To measure diversity in a neural network one may not rely on a single metric. For example, reliance solely on *Shannon Entropy* would fail to fully capture the nuances of connections. As what is lost from neuron to neuron, layer to layer, model to model is difficult to qualitatively assert, and it is even more difficult to scale this metric to every synaptic combination in a trillion parameter model (Bentti: 2023, Shannon: 1948). Unsurprisingly then the use of *Diversity Indices*[5], as found in Biology, are not reliable as they are predicated on the probability of a fixed number of categories, (Bentti: 2023). This however, in combination with other metrics may become a strength in a bounded-space. While these measurements may be argued to be further muddled when comparing the diversity of different agents with M, N, and O categories (Bettini: 2023, Morris: 2014); the shortcomings of this metric, however, are hypothesized to be mitigated when in combination with other metrics, and when a bounded space is created from probability distributions of all responses words. That is, unlike animals in a diversity study, we know precisely how many words an agent produces when prompted, and in combination.

To gain more insight into this dynamic categories problem, *System Neural Diversity*, SND is measured as the diameter of an ensemble network. SND forms from the *Wasserstein Metric* which compares probability distributions, and the *Gini coefficient*, a way to measure diversity in a group (Bentti: 2023, Zemel: 2019, Gini: 1921). As the metrics imply, an ensemble of Foundational Agents are represented by two networks where G = (Agent, Agent_Connections)

In the first network, SND is the measurement based on the average pairwise distance (Bentti: 2023). This first network is formed from differences between Agent pairs and may be dynamically set to KNN. What is being measured is the dot product of their two *State* vectors of metrics: Shannon Entropy, True Diversity, and Wasserstien scores. SND is then computed from this network.

The second network, which is still being created, is formed with the intention of passing knowledge from one agent to another. This is done by complex neuron representation. More on that in the solution section.

Now with a diversity metric being implemented, one seeks to better understand what it is being measured. The System Neural Diversity metric is just one tool that has helped uncover and support phenomena found in *wide* neural networks that make this architecture the natural choice when diversity is in mind (Bettini: 2023, Fan: 2023, Zhang: 2023, Kornblith: 2021, Lee: 2019). For highly focused tasks, deep neural networks are better at homogenizing output, which may be needed in some instances (Bettini: 2023, Zhang: 2023). Noticing these interplays is key

[5] Shannon's Entropy: $H =-\sum(pi \cdot ln(pi))$

in optimizing the cons & pros of each type of network. A brief breakdown of the main architectures may be found below.

Wide and Deep networks have been shown to produce different effects in output, with wide networks producing more diversity in output (Kornblith: 2023, Zhang: 2023, Levine: 2021, Lee: 2019, Cheng: 2016). It is thought that nodes in the first and second layers of a network play a disproportionately large role in producing knowledge, i.e., influencing heavily in deep networks (Zhang: 2023). Additionally, this may be exacerbated by the homogeneous activation functions typically found in neural networks (Fan: 2023, Pedersen: 2023).

It is then recommended that *SND* be used to analyze the diversity of an ensemble agent's outputs and to compute a response from this mixture of Foundational models using a RAG architecture. The networks created by the abstraction of each agent's state, will then be tested as deep, wide, and deep-wide networks. From this architecture, improvements should be expected in diversity while only using two hidden layers (Bettini: 2023, Fan: 2023, Zhang: 2023, Kornblith: 2021, Lee: 2019).

It would be pertinent to see how SDN is affected when using said architecture style on a range of Knowledge Intensive Learning Tasks (META: 2020). The level of complexity that a node should acquire is still an open question. The trend, however, is clearly that a more complex solution exists besides a simple scalar or float as representation, and that more diverse options, beyond random node weight distributions, are needed (Bentti: 2023, Pedersen: 2023, Zhang:2022, 2023).

5 Developmental Research & Limitations

Below, is a flowchart of the AI entity being developed and tested. It is an ensemble agent composed of N agents. Each agent has access to its own resources, has its own embedding parameters-including model, chunk size, overlap, etc. Additionally, Chains Of Thought[6], which reviews are outside the scope of this paper, may be used for an agent to further customize..

Once a prompt is asked, the metrics class measures the diversity of responses according to several metrics. These metrics compose a vector allowing for a KNN search for forming network connections, where G(E,V) = dot product of the two vectors being compared.

[6] Feng & Chen, 2023; Smola, Li, Zhang, & Zhang, 2022; Chi et al., 2022

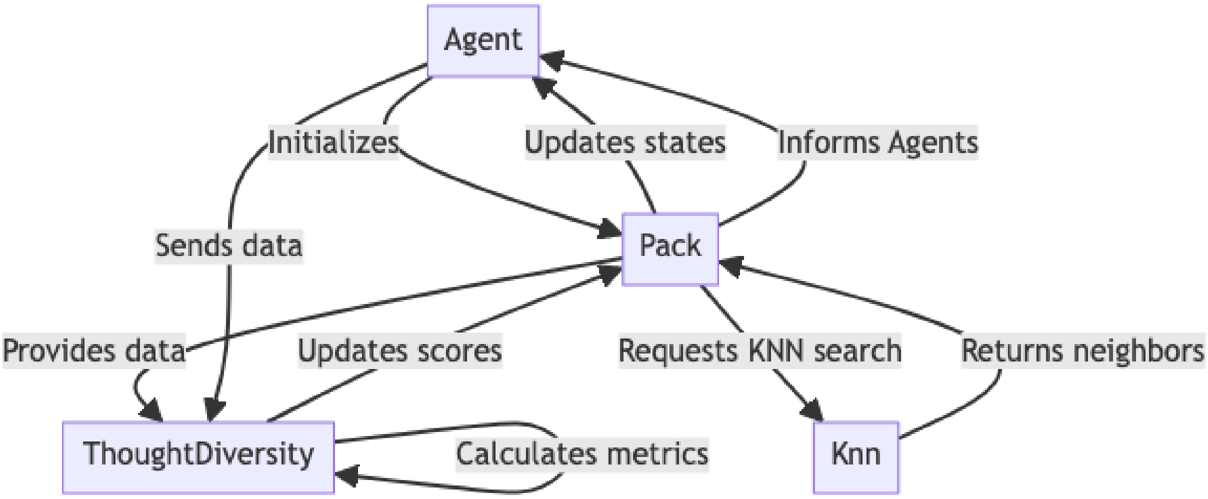


Figure 1—A high-level diagram of the ensemble entity being developed

After a prompt is given, the metrics class updates the states of individual agents based on Shannon entropy, True Diversity, and Wasserstien score. These metrics are the probability vectors of Agent X from the total Pack N response. That is, while research has shown each score alone has shortcomings, a vector of scores, however, is hypothesized to measure the downstreams effects of different embedding parameters, chains of thoughts, and source documentation. Allowing for new and interesting insights into semantic connections to be examined.

The pipeline for the Most Viable Product, MVP, has been set. Further testing is needed to help direct the direction of development, i.e., representation of diversity as measured by the metrics class and searched with the connections class. Below, a single Agent may be examined. One may see how each Agent is encapsulated and stand-alone. This allows for later recall of all agents. What this enables is for each Agent to be represented as a neuron inside a neural network, which is the topic of future work.

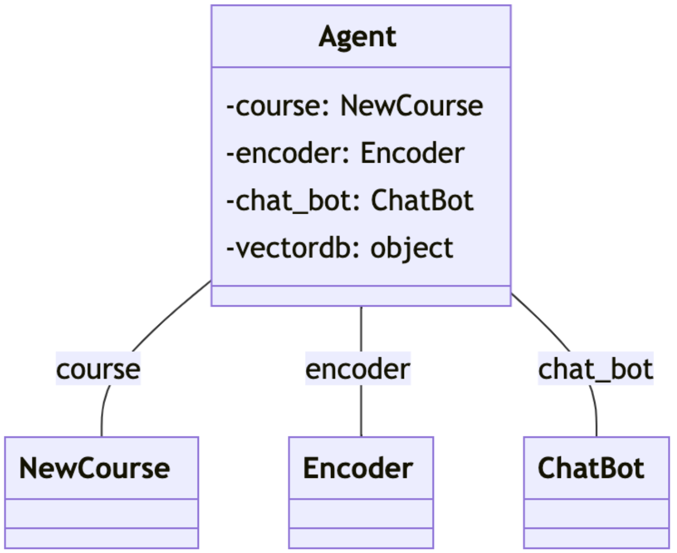


Figure 2—A single agent is an abstraction of three class working together

Once the metrics are calculated, the Pack of Agents forms a network through a customizable KNN search, which also incorporates a 30% chance of not connecting. Previous studies in neural networks have shown that randomization may allow for robust representation of data. As such, the state vectors of each Agent may be represented inside a Neuron- which is a move away from the traditional scalar representation. It is in this architecture that System Neural Diversity may be processed from not only the abstracted diversity scores, but from the agent responses themselves.

The next steps of this study are found in the Nets class. This is where complex neuron representation has begun. Now that a feed-forward neural network comprised of abstractions of each agent has been created several steps are needed

1. To incorporate a meaningful back-propagation method which utilizes diversity scores and diverse training examples
2. To measure System Neural Diversity once the back-propagation method is created
3. Begin testing with students to inform front-end design
4. Begin data analysis of Pack responses & user reported scores to direct back-end development.

Further limitations and potential downfalls are that measuring diversity coming from albeit different foundational models fine tuned in several ways may prove not enough to produce diverse knowledge. It may be the case that only the most lucrative companies have the resources to create, train, and maintain Foundational models. If this is the case, any insight into diversity, or lack of, in knowledge produced will at least be a signpost of how far we have strayed.

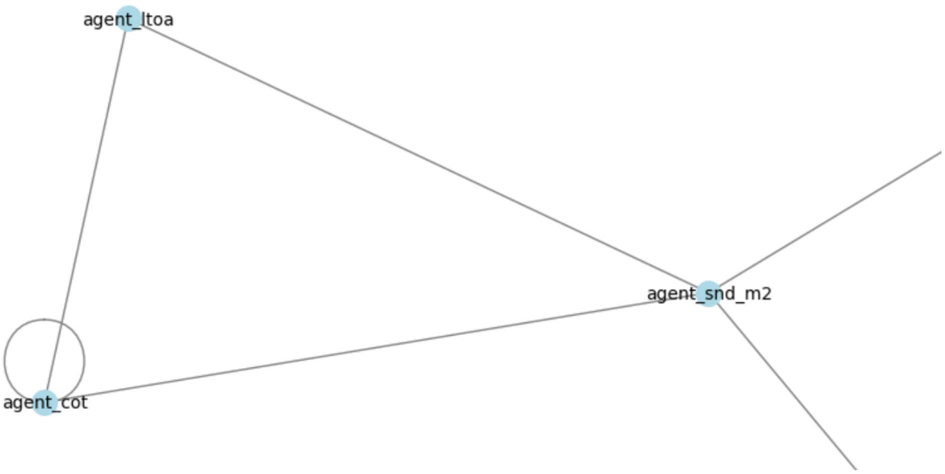
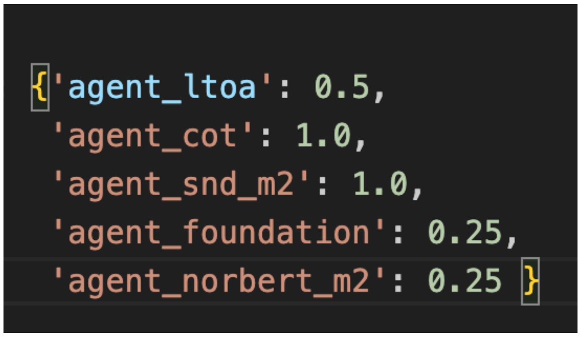


Figure 3—A high-level diagram of the Pack class represented by the wrappers it utilizes to form a network. Note, self connections are allowed.

Furthermore, the abstraction of Agents into a neural network may prove futile. While the intended goal is to produce a bespoke amalgamation of the Pack responses, this very well may prove to be fruitless. The initial results from this time and resource bound project have shown some signs of measuring diversity in answers, even if it may be at a superficial level.

The main concern for the metrics class is that the probability vectors of a word based on the entire Pack response will most likely not capture the semantic nuances. That is, how can this metrics class be tested to see if it is actually measuring diversity of thought rather than diversity of word choice?



```
{'agent_ltoa': 0.5,
 'agent_cot': 1.0,
 'agent_snd_m2': 1.0,
 'agent_foundation': 0.25,
 'agent_norbert_m2': 0.25 }
```

Figure 4—Agent centrality in the generated network of the ensemble agent.

A promising idea is that word choice is an indicator of the data being used to produce said Foundational models and responses. That is, by taking a weakness of Foundational models, circular creation[7], a canary in the coal mine might have appeared.

6. Conclusion

At scale, solutions are difficult. Metrics are the first step. Insight into data being used and closeness to other *Ground Truth Documents* is one step in the direction of controlling the breadth of insight one has into Foundational models. Awareness is half the battle. Meaning, metrics are neither enough nor trivial.

We are currently in an exciting and pivotal moment in human history. More research is needed to help build and deploy open-sourced Foundational models & tools that help produce diversity. At the very least, insight into how homogenized responses are should be mandatory reporting moving forward.

7 Resources:

1. Foucault, M. (1978). *The history of sexuality, Volume 1: An introduction*. Random House.

2. Deleuze, G., & Guattari, F. (1987). *A thousand plateaus: Capitalism and schizophrenia*. University of Minnesota Press.

3. Sartre, J.-P. (2007). *Existentialism is a humanism*. Yale University Press.

4. Shankar, K., Arora, P., & Binz-Scharf, M. C. (2021). Evidence on online higher education: The promise of COVID-19 pandemic data. *Volume 48, Issue 2*. Article.

5. Zheng, M., Bender, D., & Lyon, C. (2021). Online learning during COVID-19 produced equivalent or better student course performance as compared with pre-pandemic: Empirical evidence from a school-wide comparative study. *BMC Medical Education, 21*(495). https://doi.org/10.1186/s12909-021-02807-6

[7] AI creating content and then using said content to train on

6. Reich, J., & Ruipérez-Valiente, J. A. (2019). The MOOC pivot. *Science, 363*(6423), 130-131. https://doi.org/10.1126/science.aav7958

7. Hansen, J. D., & Reich, J. (2015). Democratizing education? Examining access and usage patterns in massive open online courses. *Science, 350*(6265), 1245-1248. https://doi.org/10.1126/science.aab3782

8. Bettini, M. (2023). System neural diversity: Measuring behavioral heterogeneity in multi-agent learning. Retrieved from https://arxiv.org/abs/2305.02128v1

9. Blesch, K., Hauser, O. P., & Jachimowicz, J. M. (2022). Measuring inequality beyond the Gini coefficient may clarify conflicting findings. *Nature Human Behaviour, 6*(11), 1525–1536. https://doi.org/10.1038/s41562-022-01430-7

10. Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Wang, W. (2022). On the opportunities and risks of foundation models. Retrieved from https://arxiv.org/abs/2108.07258v3

11. Cabañas, R. (2022). Diversity and generalization in neural network ensembles. Retrieved from https://arxiv.org/abs/2110.13786v2

12. Cheng, H.-T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H., Anderson, G., Corrado, G., Chai, W., Ispir, M., Anil, R., Haque, Z., Hong, L., Jain, V., Liu, X., & Shah, H. (2016). Wide & deep learning for recommender systems. Retrieved from https://arxiv.org/abs/1606.07792

13. Fan, F. (2023). Towards NeuroAI: Introducing neuronal diversity into artificial neural networks. Retrieved from https://arxiv.org/abs/2301.09245v2

14. Gini, C. (1921). Measurement of inequality of incomes. *The Economic Journal, 31*(121), 124-126.

15. Kornblith, S. (2021). Do wide and deep networks learn the same things? Uncovering how neural network representations vary with width and depth. Retrieved from https://arxiv.org/abs/2010.15327v2

16. Lee, J. (2019). Wide neural networks of any depth evolve as linear models under gradient descent. Retrieved from https://arxiv.org/abs/1902.06720v4

17. Levine, Y. (2021). The depth-to-width interplay in self-attention. Retrieved from https://arxiv.org/abs/2006.12467v3

18. Meta AI. (2020). Introducing KILT, a new unified benchmark for knowledge-intensive NLP tasks. Retrieved from https://ai.meta.com/blog/introducing-kilt-a-new-unified-benchmark-for-knowledge-intensive-nlp-tasks/

19. Morris, E. K., Caruso, T., Buscot, F., Fischer, M., Hancock, C., Maier, T. S., ... & Rillig, M. C. (2014). Choosing and using diversity indices: Insights for ecological applications from the German Biodiversity Exploratories. *Ecology and Evolution, 4*(18), 3514–3524. https://doi.org/10.1002/ece3.1155

20. Pedersen, J. (2023). Learning to act through evolution of neural diversity in random neural networks. Retrieved from https://arxiv.org/abs/2305.15945v2

21. Raji, I. D., & Buolamwini, J. (2020). A comprehensive study on face recognition biases beyond demographics. Retrieved from https://arxiv.org/abs/2006.12467

22. Seaborn, K., Barbareschi, G., & Chandra, S. (2023). Not only WEIRD but "Uncanny"? A systematic review of diversity in human–robot interaction research. *International Journal of Social Robotics*. https://doi.org/10.1007/s12369-023-00968-4

23. Wei, J. (2023). Chain-of-thought prompting elicits reasoning in large language models. Retrieved from https://arxiv.org/abs/2201.11903v6

24. Zemel, Y. (2019). Statistical aspects of Wasserstein distances. Retrieved from https://arxiv.org/abs/1806.05500v3

25. Zhang, A. (2022). Automatic chain of thought prompting in large language models. Retrieved from https://arxiv.org/abs/2210.03493v1

26. Zhang, X. (2023). Wider and deeper LLM networks are fairer LLM evaluators. Retrieved from https://arxiv.org/abs/2308.01862v1

7. Appendices

1. Sourced from Wei 2023 et. Al using the following embedding parameters

```
["facebook-dpr-ctx_encoder-multiset-base", 200, 25, 0.5]
```

'A chain of thought is a step-by-step reasoning process that a language model uses to arrive at an answer to a problem or question. It allows the model to decompose multi-step problems into intermediate steps, providing an interpretable window into the behavior of the model and suggesting how it might have arrived at a particular answer. This approach is used to facilitate reasoning in language models and is applicable to tasks such as math word problems, commonsense reasoning, and symbolic manipulation.'