











Initial Reading List
Independent Study - Winter 2023
Prof. Zach Wood-Doughty

S. No.	Title	Link	Comments
1.	Large language models in machine translation	 Large Language ...	One of the early papers on LLM by Google (2007)
2.	Attention is all you need	 Attention Is All ...	First transformers paper by Google Brain (2017)
3.	Improving language understanding by generative pre-training	 Improving Lang...	First GPT paper by OpenAI (2018)
4.	BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding	 BERT Pre-traini...	BERT by Google AI (2018)
5.	Transformer-XL: Attentive Language Models Beyond a Fixed-Length Context	 Transformer-XL ...	Transformers without fixed-length constraint by Google and CMU (2019)
6.	Language Models are Unsupervised Multitask Learners	 Language Model...	GPT-2 by OpenAI (2019)
7.	Language Models are Few-Shot Learners	 Language Model...	GPT-3 by OpenAI and JHU (2020)
8.	On the Dangers of Stochastic Parrots: Can Language Models be Too Big	 On the Dangers ...	Negative Impact of LLMs, keeping GPT in mind (2021)
9.	DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter	 DistilBERT, a di...	Distilled version of BERT (2020)
10.	Distilling the Knowledge in a Neural Network	 Distilling the Kn...	Explanation for DistilBert (2015)