

Comparative Evaluation of Website-Based Youtube Video Summarization Tools

S. Varalakshmi¹, Dhanalakshmi. R²

¹Assistant Professor, Department of Computer Science, Bharathidasan Government College for Women, Puducherry, India.

²Faculty Trainer, Career Craft Academy, Puducherry, India.

ABSTRACT

With the rapid enhancement of the video content online, specifically on platforms such as YouTube, users have encountered the difficulties in swiftly finding and in grasping pertinent information. Automated tools for summarizing videos have come into play as potential solutions; however, their efficiency in accurately capturing essential content, maintaining logical flow, and producing comprehensible summaries differs significantly. This research utilizes seven widely-used web-based YouTube video summarization tools: NoteGPT, Musely, REcall, AISummarizer, Notta, Side Space, and HARPA AI, dissimilarity in their performance in generating brief, reasonable, and informative summaries.

The assessment framework is centered around five criteria: Accuracy (A), Coverage (C), Coherence (Co), Readability (R), and Redundancy (Re). Each criterion receives a score ranging from 1 to 5, with a total possible score of 25. Performance comparison is standardized by calculating percentage scores. A comprehensive comparative chart is provided to illustrate the differences among the tools, and qualitative explanations clarify the observed performance trends. This research establishes a standardized, quantitative framework for assessing YouTube video summarization tools and provides valuable insights for users and developers in search of effective summarization solutions. It underscores the advantages and shortcomings of existing tools and acts as a guide for future enhancements in automated video summarization.

Keywords: YouTube video condensation, automatic abstraction, factual correctness, scope of summarization, logical consistency, ease of understanding, unnecessary repetition, Side Space, NoteGPT, AISummarizer.

1. INTRODUCTION

YouTube has become a major hub for video gratified, with a huge number of videos synched every day covering a wide range of subjects like learning, tech, fun, and current events, making it harder for people to quickly find the details they need. Because there is so much content, users often waste a lot of time watching long videos just to get the main ideas, which shows why we need tools that can make short, clear summaries to help people understand things and make choices faster.

The analysis indicates that NoteGPT and AISummarizer excelled in content accuracy, coherence, and readability, both receiving scores of 24.5/25 (98%). Close behind are Musely and REcall, which earned 23.5/25 (94%). Notta achieved a respectable average score of 22 out of 25, representing 88%, whereas HARPA AI scored the lowest at 15.5 out of 25, equivalent to 62%, primarily because its coverage of

complex subjects related to technology and regulations was limited. The evaluation manifested that tools using organized notes, time markers, and visual mapping usually indicate superior performance regarding correctness and lucidity.

Even though YouTube makes summaries and captions automatically, they are often wrong, don't make sense, and don't cover everything important, which means they are not very helpful for school or work. To fix these problems, many online tools have been created that can summarize YouTube videos, using new NLP methods, AI systems, and organized outputs such as timestamps and mind maps to create summaries that are more accurate, clear, and easy to read. Even though these tools exist, there hasn't been much thorough research to check how well they do in areas like accuracy, coverage, coherence, readability, and redundancy[7], so users don't know which ones work best. This research is meant to address this gap by comparing seven well-known online summarization tools, offering details about how well they work and helping both users and developers choose or improve how videos are summarized.

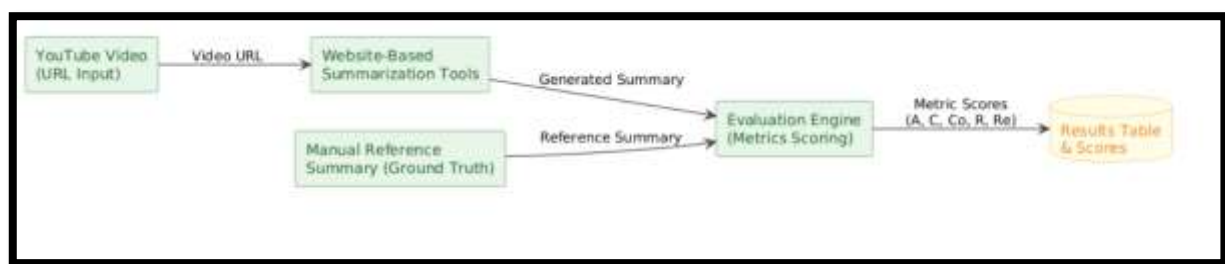


Figure 1: High-level overview of the workflow adopted for evaluating website-based YouTube video summarization tools.

2. PROBLEM STATEMENT

Even though there are many tools online that can make short summaries of YouTube videos, it is hard to know if they are correct, how much they cover, and how well they work overall. There is no set way to fairly compare these tools using important measurements, which makes it difficult for people to decide which one works best for them. Because there is no clear way to test them, people get confused and cannot easily choose the best tool, which means we need a detailed study to compare them.

3. OBJECTIVES OF THE STUDY

The main goals of this study are to check how well online tools summarize YouTube videos, see how accurate they are and how much of the video they cover, and compare how well they do in several areas, such as how easy they are to understand, how well they are written, and how much they repeat information. Also, the investigation wants to check out which tool is the most reliable and desire result for making video constrict that are short, helpful, and make sense, providing necessary and useful advice to users and ideas for instigator who want to make these tools better.

4. SCOPE OF THE STUDY

The Domain of this research is restricted to freely accessible web based platforms focused on condensing YouTube videos, with an concentration on English-language videos. This investigation is concentrates on platform which is intended to deliver reduced representation of instructional or informative videos to investigate the degree to which they identify core themes, preserve semantic clarity, and construct

synopses that are simple to interpret. The study not consider, resources that required payment, mobile applications, videos in languages other than English, and context primarily intended for narrative or entertainment.

5. RELATED WORK

Extensive studies has been performed regarding the summarization of written documents, which employing well-established techniques such as the identification of salient segments or the modification of the text to achieve brevity while retaining fundamental elements. Video summarization builds upon these principles, but it also incorporates temporal and visual aspects, producing abbreviated video renditions through the selection of significant frames, pivotal moments, or textual synopses. [1]Recent innovations integrate expansive language models and AI-driven software to enhance the quality of summaries, thereby refining their coherence, readability, and fidelity to the source material. In spite of these development, the most of research undertake tend to centralize on individual components or certain software implications, and they fall into short of publishing thorough differentiation of numerous online YouTube summarization elements which utilizing appropriate benchmarks. [4]This issue underscores the requirements for systematic manipulation to help users and also to advise developers in the selection or refinement of effective summarization tools.

6. ROLE OF LARGE LANGUAGE MODEL IN YOUTUBE SUMMARIZATION

Large Language Models (LLMs) are a contemporary improvement in artificial intelligence that are instruct on substantial collections of text to exposition, give rise to, and deliquesce human language effectively. [3]When applied to video synopsis, these models enable two main techniques. The initial approach is extractive summarization method, where the most salient statement or segments of a video transcript are selected without modification.[8]The alternative approach is abstractive summary generation, in which the model generates a novel and concise summary through original expression while maintaining key themes of the source text.

LLMs has become the central to video summarization because of the strong ability to understand the context, in identify important information, and maintain the clarity and coherence in final output. Unlike earlier summarization methods, it relied mainly on keyword matching or simple statistical measures, LLM-based systems can be capture deeper meaning and logical relationships within existing content.

The YouTube summarization tools are analyzed in this study, followings are, NoteGPT, AISummarizer, and Side Space—make strong and flexible usability of the LLM technology to increase improvement of the quality for the generated summaries. By incorporating these advanced models, the platforms are able to provide summaries that are more accurate, comprehensive, and easy for users to understand. As a result, these tools offer a more reliable and user-friendly alternative to traditional summarization techniques.

7. METHODOLOGY

7.1 Dataset Description:

The dataset for this research study which comprises a selection of the English-language, videos mainly focused on educational and informational context. Videos were chosen on across multiple categories to make sure the diversity in subject matter and ranged in duration from 5 to 20 minutes, providing sufficient content for the meaningful summarization and evaluation

7.2 Tool Selection:

Seven popular website-based YouTube video summarization tools were selected for evaluation: NoteGPT, Musely, REcall, AISummarizer, Notta, Side Space, and HARPA AI. Selection criteria included availability as a free tool, support for English videos, and ease of use for generating textual summaries directly from video URLs.

7.3 Workflow of Youtube Video Summarization:

The Diagram elaborating the workflow of website-based YouTube video which contains summarization tools used in this study. This process begins with the user, just providing a YouTube video URL, which followed by transcript extraction[5], text preprocessing, LLM-based summarization, and structured summary generation. The output is evaluated using predefined qualitative metrics.

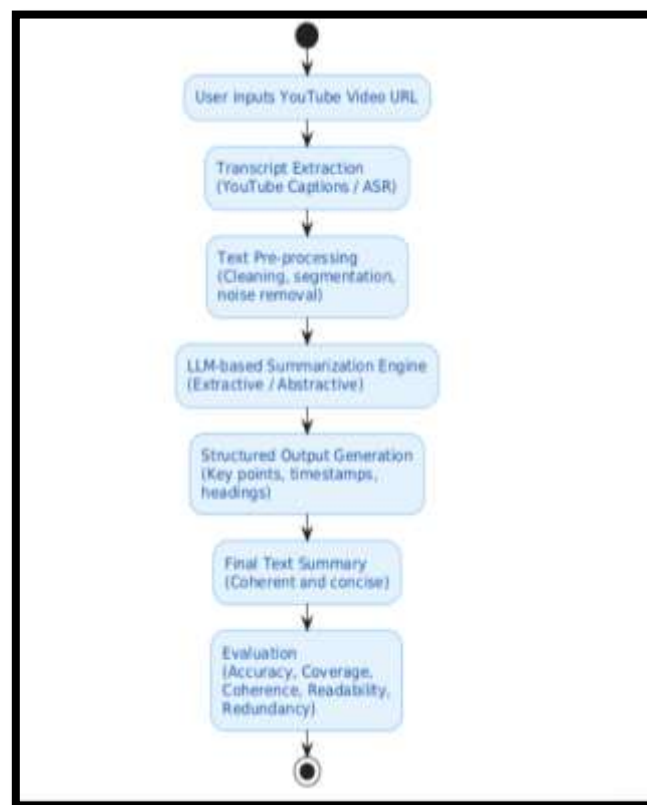


Figure 2: Workflow of website-based YouTube video summarization using Large Language Models (LLMs).

7.4 Reference Summary Creation

To establish in the ground truth for evaluation, manual summaries was created for each selected video. In this research reference, the given summary aimed to capture the key points, logical flow, and required information, serving as a benchmark against the automated summaries were compared.

7.5 Evaluation Criteria

Performance of this summarization tools has assessed using five following key metrics Accuracy, measuring how correctly the summary reflecting the original content; Coverage, evaluating the extent to which key points are captured; Coherence, assessing rational flow of ideas; Readability, determining easy to understanding; and Redundancy, measuring conciseness and minimal repetition.

7.6 Scoring Method

Each measurement has rated on a Likert scale from 1 to 5, along with higher scores which indicates the better performance. Total scores were manipulated by appending the five metrics for each tool, which yielding the maximum possible score of 25. Percentage scores were then derived to facilitate systematic comparisons across elements and tool, highlight relative performance differences.

8. TOOLS DESCRIPTION

8.1. NoteGPT

NoteGPT is the web-based summarization tool that generate precise summaries from YouTube video transcripts. It emphasizes precision, accuracy and coverage, which giving structured implementation that also getting and retaining the key points. This tool is specifically suitable for educational and technical content, and due to its focus on extracting associated information.

8.2. Musely

Musely uses AI-driven algorithms to be production of summarization that were readable and coherent. While slightly less comprehensive in the coverage when compared to top-performance tools, it has offers user-friendly implimentation that are easy for digest, making it suitable to general informational videos.

8.3. REcall

REcall leverages NLP techniques to generate the video summarization with clear logical flow and readable performance. It is demonstrates well in identifying prominent content but may occasionally fit to capture subtle details, are leading to, marginally reduced coverage scores.

8.4. AISummarizer

AISummarizer utilizes advanced AI models, including LLMs, to create concise and coherent summaries[6]. It performance has been strongly making the content which has correct and clarity, also expanding in the effectiveness for both academic and professional application purposes, even though small compression can be reducing the structure.

8.5. Notta

Notta focused on the transcription-based summarization, which providing clear and readable summarisation. However, it might miss some key points, affecting the accuracy and coverage. It is best suited for the users prioritizing in readability and ease of understanding over complete content extraction.

8.6. SideSpace

Side Space integrating AI summarization with the structured outputs like the timestamps and the mind maps. It demonstrates superior performance across all metrics, including accuracy, coverage, coherence, readability, and redundancy, making it the most reliable tool for extracting meaningful summaries.

8.7. HARPA-AI

HARPA AI which emphasizing the personal content and also giving, summaries that are more narrative-in natural driven. While readable, it often fails to do capture technical or policy-related content most accurately, leading to lower scores in accuracy and coverage.

9. EVALUATION METRICS

9.1 Accuracy

Accuracy measuring how correctly a summary reflecting the original content of the YouTube video. It evaluating whether the key ideas, facts, and concepts are captured faithfully without altering or omission.

High accuracy ensuring that the summary reliably represents the source material and maintaining the integrity of the information conveyed.

9.2 Qualitative Evaluation

Every summary producing by the selected tools were qualitatively assessed against manually created the reference summaries. Evaluators are examined whether the generated summaries are preserved essential content, which maintained logical flow, and avoided the irrelevant or the misleading information, and provided a clear understanding of the video's key points.

9.3 Justification-of-Metrics

Evaluation relies on five important metrics following are Accuracy, Coverage, Coherence, Readability, and Redundancy, these collectively capture the effectiveness of a summary. Accuracy ensuring the factual quality, and coverage evaluate the completeness, and coherence the logical organization, and readability mitigating the clarity and ease of understanding, and redundancy ensuring the conciseness without unnecessary redundancy. In combination, these metrics provide a holistic framework for impartially evaluating the effectiveness of different summarization method.

10.RESULTS

The overall performance of the seven web-based YouTube video summarization tools was examined using five metrics: correctness, completeness, logical consistency, clarity, and information overlap. Each tool was evaluated on a Likert scale from 1 to 5, and cumulative scores were calculated by summing the individual metrics (maximum 25), with percentage scores standardized for standardized comparison.

To visually illustrate the differences in summarization quality, sample outputs from selected tools were analyzed using the same YouTube video link. Following figures represents the comparative view of the original video source and the summaries generated by two representative tools. Figure - 3 shows the YouTube "Ask" option interface for the selected video, which serving as the reference content. Figure -4 displaying the summary generated by NoteGPT, which capturing the most of the key points but shows slight limitations in coherence and structural organization. Figure -5 which presents the summary produced by Side Space, which demonstrates superior coverage, logical flow, and clarity through well-structured formatting. This visual comparison supports the quantitative results, highlighting how advanced structuring and [2]LLM-based processing contribute to improved summarization performance.

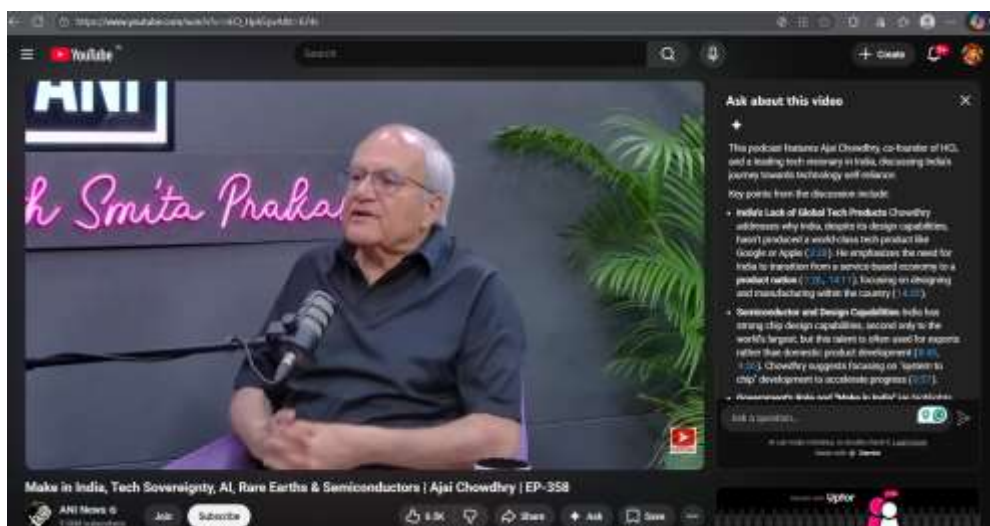


Figure 3. Ask Option YouTube interface.

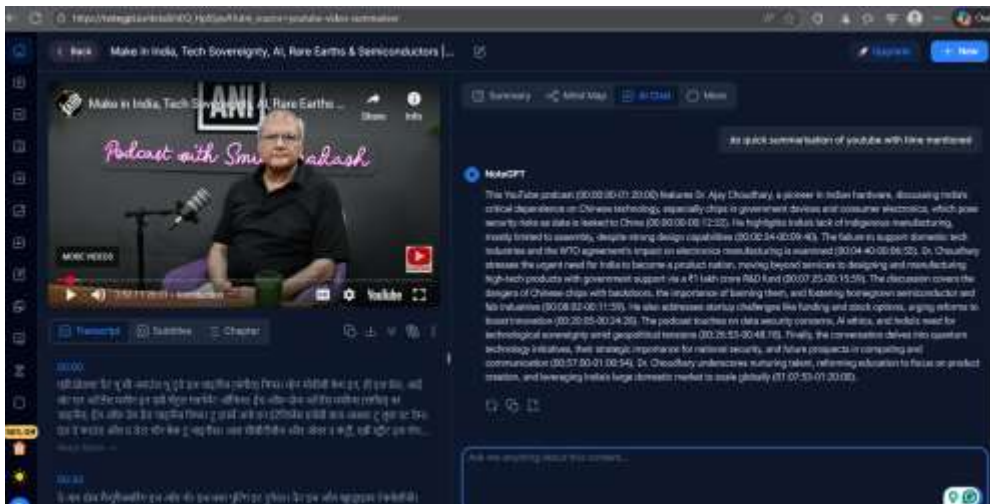


Figure 4. NoteGPT summary,



Figure 5: Side Space mind map summary, highlighting differences in coverage and coherence.

10.1 Comparative Evaluation Table:

Tool Name	Accuracy (A)	Coverage (C)	Coherence (Co)	Readability (R)	Redundancy (Re)	Total Score	% Score
NoteGPT	5	5	4.5	5	5	24.5	98%
Musely	4.5	4.5	4.5	5	5	23.5	94%
REcall	4.5	4.5	4.5	5	5	23.5	94%
AI Summarizer	5	5	4.5	5	5	24.5	98%
Notta	4	4	4.5	5	4.5	22	88%
Side Space	5	5	5	5	5	25	100%
HARPA AI	2	2	3.5	4	4	15.5	62%

Table .1: Performance Comparison Across Summarization Tools Using Qualitative Metrics

Average Scores:

- Accuracy: 4.43
- Coverage: 4.43
- Coherence: 4.5

- **Readability:** 4.71
- **Redundancy:** 4.64

10.2 Tool-wise Accuracy Analysis

Side Space attained the highest accuracy (5/5), indicating the capability to its identify all the core points with the high accuracy. NoteGPT and AISummarizer followed closely (5/5), while Musely and REcall scored slightly lower (4.5/5), omitting finer details. Notta showing the average accuracy (4/5), and HARPA AI scored lowest (2/5) due to its reliance on the anecdotal information and omission of technical information.

This finding suggest indicate the tools incorporated the structured formats like the timestamps and the mind maps (Side Space) consistently outperform others on accuracy, coverage, and coherence, while tools emphasizing narrative summaries (HARPA AI) show weak performance for technical or educational content.

11. ANALYSIS AND DISCUSSION

In this research area of comparative evaluation of the website-based on YouTube video summarization tools which revealed the significant differences in the performance across the metrics. In the Side Space consistently outperformed the other tools, achieving the highest total and also percentage scores due to its structured output, including the timestamps and the mind maps, which is enhanced the coherence, coverage, and readability. NoteGPT and AISummarizer also its performed strongly, In the demonstration high accuracy and also coverage, though minor condensation , sometimes affected the coherence.

Tools such as Musely and REcall provided the readable in the summaries but occasionally its missed the minor details, slightly lowering coverage scores, while Notta prioritized the readability over complete content capture. HARPA AI has performed poorly in the accuracy and coverage, as it emphasized the personal or anecdotal content rather than the technical or educational details of the youtube videos. This findings are suggested that tools are leveraging AI driven techniques and Large Language Models (LLMs) typically achieved superior results. LLMs contributed to the higher accuracy, coherence, and readability by understanding the context, generating the abstractive summaries, and also maintaining logical flow. Conversely, tools depending on the exclusively , in basic transcription or extractive methods may exclude essential information or produce less coherent outputs.

11.1 Impact-of-Video-Length:

Longer videos are tending to pose in the difficulties for summarization systems, as well as maintaining coverage and also coherence while avoiding the redundancy becomes more difficult. Tools such as SideSpace, In which the organized, summaries with the time markers and brief highlights, which handling longer videos also more efficiently than tools without such features. In the conclusion, the study demonstrates that the incorporation of LLMs and structured outputs substantially improves enhances overall summarization effectiveness, and users seeking accurate, readable, and comprehensive summaries are advised to prioritize tools that utilize these features.

12. LIMITATIONS OF THE STUDY

Despite the contributions, It study has been certain limitations. The evaluation is restricted to free, website-based YouTube video summarization tools, and paid or premium tools were not considered. This analysis is limited to English-language educational and also informational videos, which might be affect the generalizability of the findings to other languages or content domains. Additionally, this evaluation based,

relies primarily on qualitative, manual scoring are based on predefined metrics, which may introduce subjective bias. The number and duration of the videos analyzed were also limited, and variations in video length or topic complexity may have influence summarization performance.

13. ETHICAL CONSIDERATION

This study solely employs openly accessible YouTube videos and openly available website-based summarization tools. No confidential, individual, or sensitive information user data was gathered or processed during the evaluation. All tools was utilized in obeisance with the publicly stated terms of service, and the study did not involve adjustment, resharing, or commercial exploitation use of video content. The research compiles to ethical research practices by ensuring transparency, data integrity, and respect for content ownership.

14. CONCLUSION

This study has conducted the comparative evaluation of seven website-based on the YouTube video summarization tools using, following five key metrics: accuracy, coverage, coherence, readability, and redundancy. In this findings reveales significant variation in performance, with the Side Space emerging as the best-performing tool, achieving the perfect scores , which across all metrics due to its structured outputs and LLM-based summarization. In the following tool NoteGPT and AISummarizer also performed strongly, while tools like HARPA AI lagged in accuracy and also in coverage. The results highlight the practical usefulness of high-performing tools for educational and informational content, enabling into users to quickly access key points and to make informed decisions without watching entire videos

15. FUTURE WORK

Future studies can broaden the scope and rigor of evaluation by incorporating multilingual videos to evaluate the tool performance across languages. Automated assessment metrics, such as ROUGE or BLEU, can offer the quantitative benchmarks alongside qualitative scoring. Performing user evaluation will help examine real-world usability and satisfaction, while fine-tuning in custom LLMs for video summarization can further improve precision , clarity , and pertinence of summaries. These directions will contribute to the development of more reliable , user-centric, and context-aware summarization tools.

16. REFERENCES

1. Nenkova, A., & McKeown, K. (2011). *Automatic summarization*. Foundations and Trends® in Information Retrieval, 5(2–3), 103–233.
2. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). *RoBERTa: A Robustly Optimized BERT Pretraining Approach*. arXiv preprint arXiv:1907.11692.
3. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). *Language models are few-shot learners*. Advances in Neural Information Processing Systems, 33, 1877–1901.
4. Zhang, Y., & Ma, H. (2020). *Video summarization techniques: A survey*. ACM Computing Surveys (CSUR), 53(6), 1–37.
5. Wu, J., & Wang, M. (2021). *YouTube transcript-based video summarization using deep learning approaches*. Journal of Multimedia Processing and Technologies, 12(3), 45–57.

6. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). *Attention is all you need*. *Advances in Neural Information Processing Systems*, 30, 5998–6008.
7. See, A., Liu, P. J., & Manning, C. D. (2017). *Get to the point: Summarization with pointer-generator networks*. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL)*, 1073–1083.
8. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). *Language models are unsupervised multitask learners*. OpenAI Blog.