Convert AI – Generated Text to Human – Written Text and Remove Plagiarism

[1]Najmul Hudha PA, [2]Nithya TK, [3]Sneha Sudhakaran, [4]Sruthi A
[1]Student, [2]Student, [3]Student, [4]Professor(CSE)
[1]Computer Science and Engineering Department,
[1]Nehru College of Engineering and Research Centre (NCERC), Thrissur, India

Abstract: The rapid integration of AI in text generation across various industries—such as content creation, education, and business—has raised concerns regarding the authenticity and originality of AI-generated content. While efficient and informative, such content often lacks the natural tone, flow, and engagement characteristic of human writing, making it appear artificial or detached. Moreover, due to the nature of AI training on extensive internet datasets, there is an increased risk of unintentional plagiarism through closely mimicked phrasing or structure. This study addresses these challenges by exploring methods to convert AI-generated text into more human-like language through refined syntax, tone adjustment, varied vocabulary, and thorough paraphrasing. It also examines techniques for detecting and mitigating plagiarism to ensure compliance with academic and professional standards. By applying advanced natural language processing techniques and stylistic enhancements, the research contributes to the ethical and effective use of AI in text generation, promoting originality and authentic communication.

Index Terms – AI – Generated, Natural Language Processing (NLP), Text Authenticity.

I.INTRODUCTION

In recent years, AI-generated text has become an increasingly popular tool across various fields, including academia, content creation, and digital marketing. With rapid advancements in machine learning and natural language processing, AI models can now produce coherent and contextually relevant content with minimal input. However, despite their efficiency, AI-generated texts often lack the nuance, authenticity, and originality that characterize human-authored writing. This presents a challenge for those who rely on AI tools: how to refine or transform AI-generated content so it resembles natural human writing while remaining free from plagiarism. Understanding the nature of AI-generated text is a crucial first step in this process. Models like ChatGPT or GPT-4 are trained on vast datasets, allowing them to mimic patterns in language without directly copying any specific source, yet the similarity of certain phrases to existing works can unintentionally raise plagiarism concerns. As such, the conversion of AI-generated material into more human-like writing begins with rephrasing, enhancing originality, and minimizing repetitive language. Injecting unique insights, adding a personal touch, and customizing the tone can make the content more reflective of genuine human thought. To further avoid unintentional plagiarism, writers should utilize detection tools to flag areas needing refinement. Another vital aspect of humanizing AI-generated content involves boosting its engagement factor. While AI tends to create clear and organized text, it often lacks the emotional resonance and subtlety of human writing. Adding anecdotes, concrete examples, and a clear personal perspective can make content more compelling. In addition, varying sentence structure and using rhetorical devices such as questions or vivid descriptions help break the mechanical rhythm often found in AI outputs. Finally, thorough proofreading ensures the writing flows naturally and eliminates any lingering signs of automation, such as awkward transitions or repeated phrases, ultimately resulting in a more polished and human-like piece.

II. LITERATURE REVIEW

[1] The application of artificial intelligence (AI) in translation teaching is transforming traditional approaches, enhancing language learning and translation skills in diverse ways. AI-driven tools like neural machine translation (NMT) systems, such as Google Translate and DeepL, offer students real-time feedback on translations, enabling them to understand nuances, idiomatic expressions, and contextual variations in different languages. By integrating AI into translation teaching, educators can introduce students to advanced machine translation algorithms, helping them grasp the underlying processes that power these technologies, including deep learning, neural networks, and data analysis.

[2] Translation as a form of Human-Computer Interaction (HCI) has emerged as a vital area of study, especially with the increasing globalization and need for seamless multilingual communication. Traditionally, translation was a purely human-driven process, relying heavily on linguistic expertise and cultural knowledge. However, with advancements in HCI, translation has evolved to incorporate various computational tools and algorithms, leading to what we now refer to as machine translation (MT).

[3] The rise of Artificial Intelligence (AI) brings transformative potential across various sectors, yet it also poses significant threats to human health and existence, sparking intense debate among researchers, ethicists, and policymakers. One primary concern is the potential for AI to exacerbate health inequalities; as AI technologies become more integrated into healthcare, discrepancies in access to advanced tools could widen the gap between affluent and underserved populations, leading to worse health outcomes for vulnerable groups.

[4] Artificial Intelligence (AI) has emerged as one of the most pivotal foundation technologies of our time, fundamentally transforming numerous sectors, including healthcare, finance, transportation, and agriculture. The literature surrounding AI spans decades, evolving from early theoretical frameworks to sophisticated algorithms that drive modern applications. Early AI research focused on symbolic reasoning and rule-based systems, exemplified by the development of expert systems in the 1970s and 1980s..

[5] The literature on Massive Open Online Courses (MOOCs) and their impact on translation elucidates a multifaceted relationship between technology and translation practices, reflecting the transformative effects of digital tools on both the accessibility and methodologies of translation. Researchers such as Pym (2010) and Garcia (2010) argue that the integration of machine translation (MT) within educational contexts enhances access to diverse linguistic resources, enabling learners to engage with content that transcends geographical and linguistic barriers. However, concerns about the limitations of MT in conveying cultural nuances and context have prompted scholars to advocate for the continued necessity of human translators in refining automated outputs, ensuring quality and contextual integrity.

[6] A corpus-based approach to plagiarism detection leverages large datasets of text to identify similarities and potential instances of copying between documents. This method involves creating a comprehensive corpus that includes various types of texts, which can range from academic papers to online articles. By employing techniques such as text mining, natural language processing, and statistical analysis, researchers can analyze linguistic features, including syntax, semantics, and structure, to detect not only direct copying but also paraphrasing and idea theft.

[7] The concept of "language models as few-shot learners" explores the capability of large pre-trained language models to adapt to new tasks with minimal task-specific data, a characteristic that has gained significant attention in recent years. Research demonstrates that models like GPT-3 and BERT can generalize from a limited number of examples, showcasing their potential to perform various tasks such as text classification, question answering, and summarization without extensive fine-tuning.

[8] The study of plagiarism detection through linguistic patterns, textual features, and detection methods encompasses a range of analytical approaches aimed at identifying and understanding instances of academic dishonesty. Research highlights various linguistic markers such as lexical choices, syntactic structures, and stylistic features that differentiate original texts from plagiarized content. Textual features, including semantic similarity, structural similarities, and coherence measures, are crucial in assessing the originality of a document.

III.PROBLEM STATEMENT

The increasing use of AI-generated text has transformed content creation, but it presents key challenges, particularly in maintaining originality, emotional depth, and human-like nuance. While AI can efficiently produce coherent content, it often lacks the creativity and authenticity of human writing, and may inadvertently replicate copyrighted material, raising concerns about plagiarism. Therefore, developing effective methods to humanize AI-generated content while ensuring its originality and legal compliance has become a critical need in today's digital landscape.

IV.PROPOSED SYSTEM

The proposed Text Conversion and Plagiarism Removal System aims to revolutionize the way users create and refine written content by integrating three essential functions: text summarizer, ai text detection, and paragraph humanizer . Unlike existing platforms that typically focus on either grammar correction or plagiarism checking, this innovative system bridges the gap by combining these functionalities into a single user-friendly interface. The system captures the natural flow, creativity, and emotional tone typical of human authorship, making the output more relatable and authentic.

V.MODULE DESCRIPTION

In project modules play an important role such that through modules we get a clear idea about the project. CONVERT AI – GENERATED TEXT TO HUMAN – WRITTEN TEXT AND REMOVE PLAGIARISM mainly has 6 modules which describe the site completely. The modules are:

1. USER LOGIN MODULE
2. ADMIN MODULE
3. TEXT PROCESSING MODULE
4. PLAGIARISM DETECTION MODULE
5. TEXT CONVERSION MODULE
6. SUMMARY GENERATION MODULE

1. USER LOGIN MODULE

New users can sign up by providing credentials (username, password, email, etc.). Optionally, email verification can be added. Authenticated login to ensure that only registered users can access the platform. Passwords should be encrypted for security. Once logged in, users can access the system features: upload files, view past results, check plagiarism reports, etc.

2. ADMIN MODULE

This module manages system operations, monitor usage, and handle users and content analytics. Admin can view, or delete user accounts. Also manages user activity and permissions. Admin oversees the performance of plagiarism detection and text conversion tasks—checking if modules are working correctly and efficiently.

3. TEXT PROCESSING MODULE
   This text processing module is used to pre-process and clean the uploaded text, making it suitable for further analysis. If the uploaded file is a document (PDF, DOCX), this module extracts the raw text. Breaks down the text into sentences or paragraphs to prepare it for rewriting or analysis.
4. PLAGIARISM DETECTION MODULE
   Plagiarism detection module compares input text against existing content and identify copied and identify copied or closely Matched segments.
5. TEXT CONVERSION MODULE
   Text conversion module is mainly used to rewrite or paraphrase the plagiarized sections to improve originality and readability. Uses Natural Language Processing (NLP) models or AI (like GPT) to restate sentences with the same meaning but different wording.
6. SUMMARY GENERATION MODULE
   To generate a concise version of the input text, highlighting the key points. Uses AI to generate a summary in new wording.

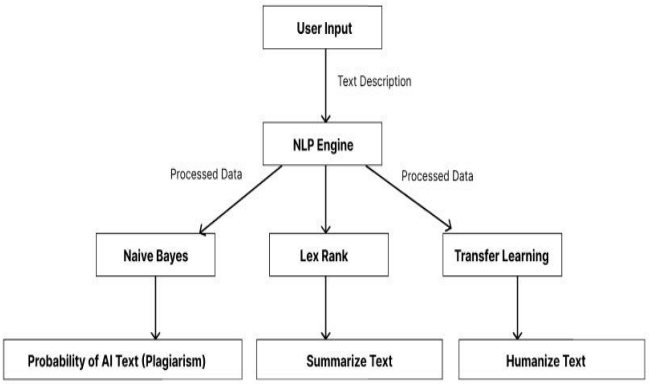
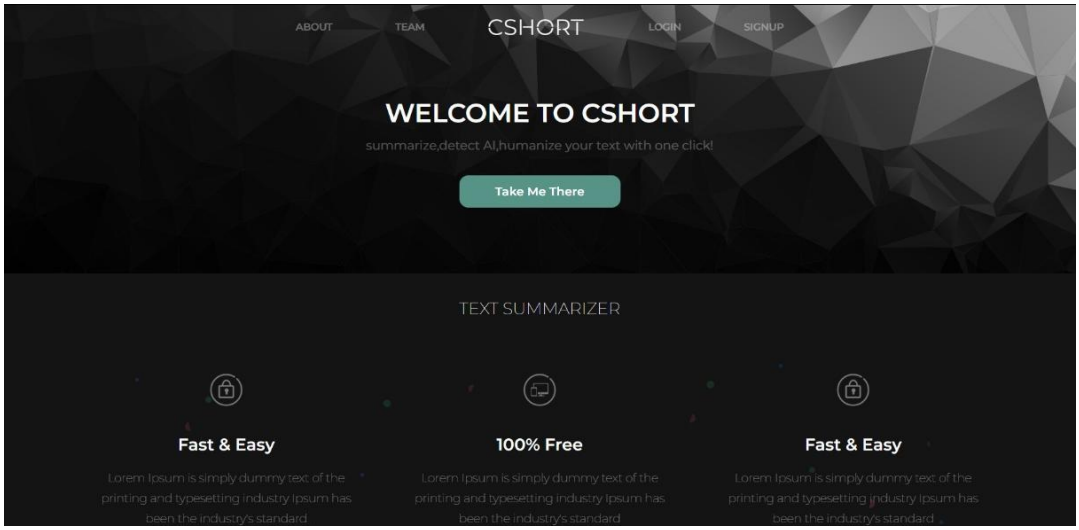
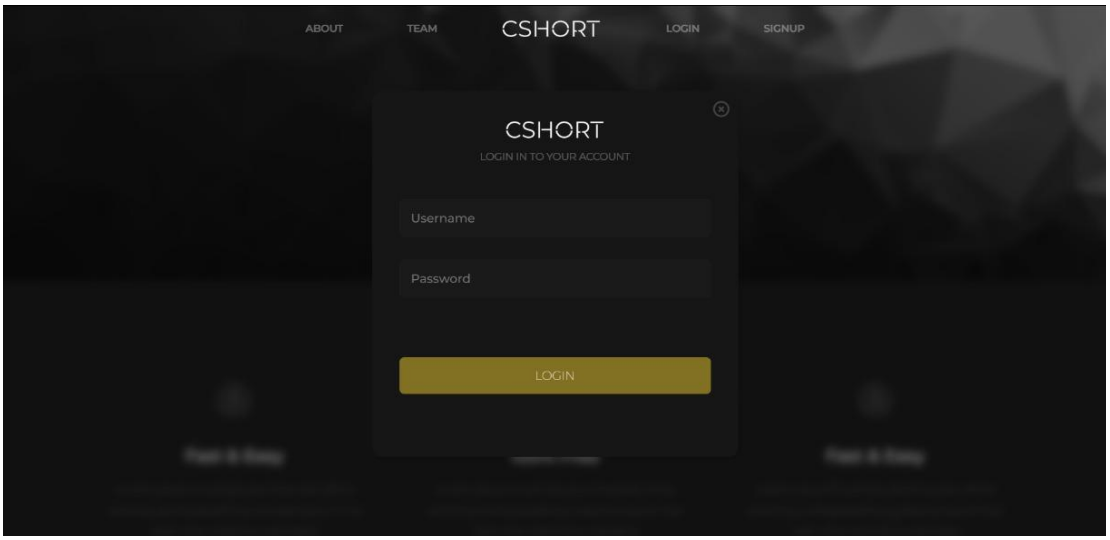
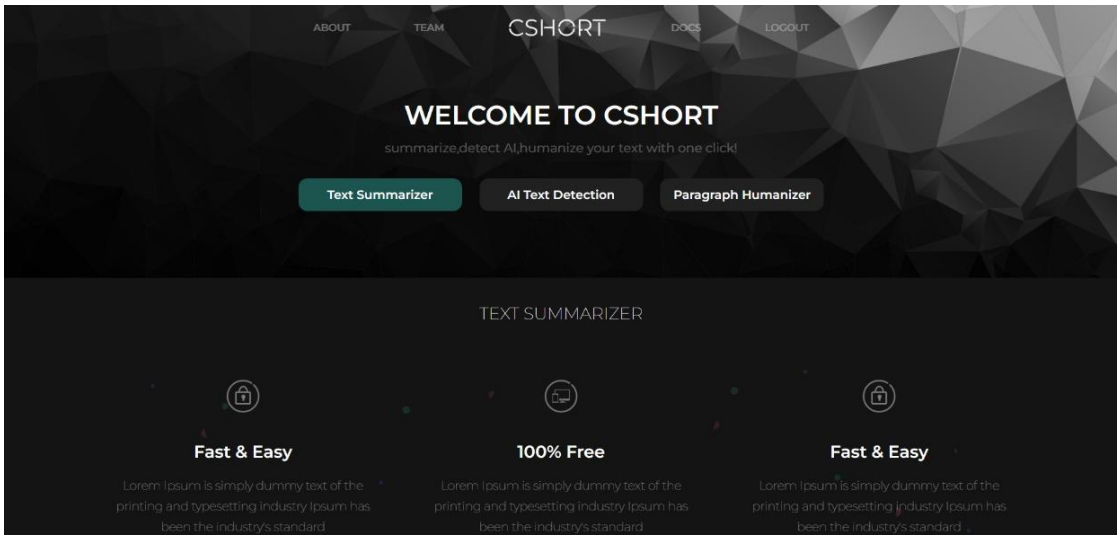


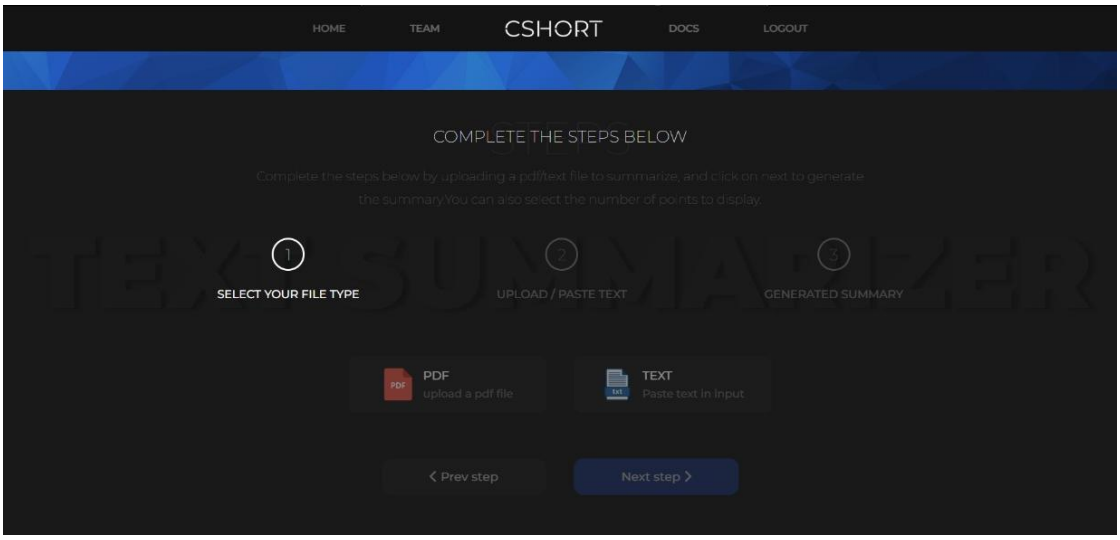
Fig 1 : System Architecture

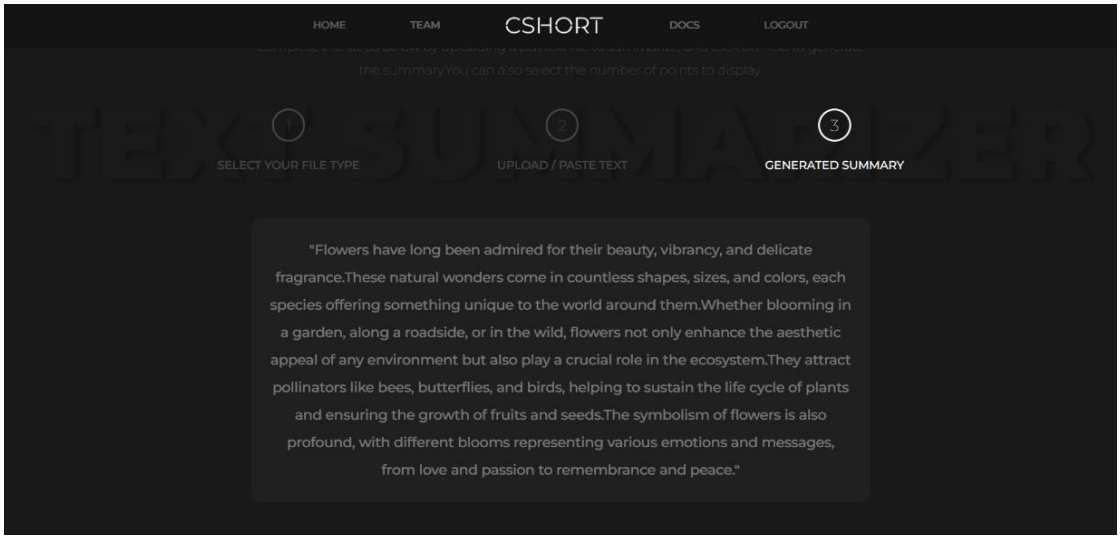
VI. RESULTS AND DISCUSSION

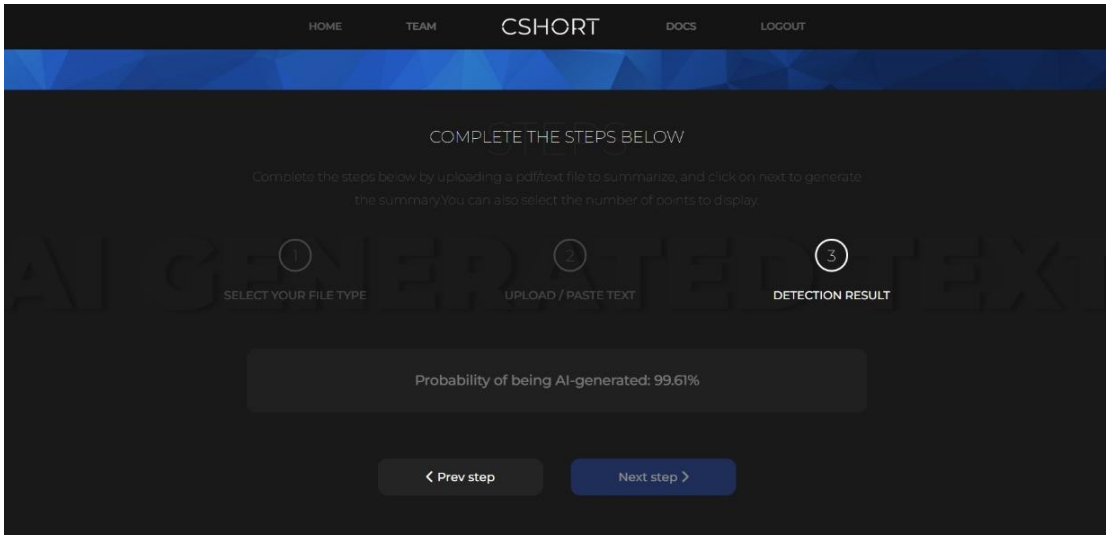



ABOUT
TEAM
CSHORT
LOGIN
SIGNUP
CSHORT
LOGIN IN TO YOUR ACCOUNT
Username
Password
LOGIN


ABOUT
TEAM
CSHORT
DOCS
LOGOUT
WELCOME TO CSHORT
summarize,detect AI,humanize your text with one click!
Text Summarizer
AI Text Detection
Paragraph Humanizer
TEXT SUMMARIZER
Fast & Easy
Lorem Ipsum is simply dummy text of the printing and typesetting industry Ipsum has been the industry's standard
100% Free
Lorem Ipsum is simply dummy text of the printing and typesetting industry Ipsum has been the industry's standard
Fast & Easy
Lorem Ipsum is simply dummy text of the printing and typesetting industry Ipsum has been the industry's standard


HOME
TEAM
CSHORT
DOCS
LOGOUT
COMPLETE THE STEPS BELOW
Complete the steps below by uploading a pdf/text file to summarize, and click on next to generate the summary.You can also select the number of points to display.
1
SELECT YOUR FILE TYPE
2
UPLOAD / PASTE TEXT
3
GENERATED SUMMARY
PDF
upload a pdf file
TEXT
Paste text in input
Prev step
Next step





VI.CONCLUSION

In conclusion, converting AI-generated text into human-written, plagiarism-free content represents an innovative and necessary step forward as AI becomes increasingly integrated into writing and content creation. This capability bridges the gap between leveraging AI's efficiency and maintaining the authenticity, creativity, and originality essential in academic, business, and creative contexts. By transforming AI drafts into unique, human-like expressions, these tools can support students, researchers, and content creators in producing work that is both ethically responsible and of high quality, ensuring compliance with academic and publishing standards. As AI continues to evolve, the demand for tools that refine and authenticate AI-generated text will only grow. s. By providing a reliable way to remove plagiarism and adapt AI text to sound genuinely human, these solutions can facilitate AI's productive use while preserving the integrity of human expression and upholding originality in all forms of content creation. Furthermore, in content marketing, publishing, and creative industries, these tools can enable professionals to use AI-generated drafts effectively while preserving brand voice and audience appeal.

REFERENCES

[1] C. Yang, The application of artificial intelligence in translation teaching, Proceedings of the 4th International Conference on Intelligent Science and Technology (2022, August) 56–60
[2] S. O'Brien, Translation as human-computer interaction, Translation Spaces 1 (1) (2012) 101–122
[3] H. Fetzer, What is artificial intelligence ?, in: Artificial Intelligence: its Scope and Limits. Studies in Cognitive Systems, vol. 4 Springer, Dordrecht, 1990 https:// doi.org/10.1007/978-94-009-1900-6_1.
[4] Dalal, F. Federspiel, R. Mitchell, A. Asokan, C. Umana, D. McCoy, Threats by artificial intelligence to human health and human existence, BMJ Glob. Health 8 (5) (2023) e010435.
[5]] T. Ray, Meta's Zuckerberg: 'AI Is Perhaps the Most Important Foundational Technology of Our Time, Zdnet, 2022, Feb 23. https://www.zdnet.com/article/ metas-zuckerberg-ai-is-perhaps-the-most-important-foundational-technology-of-our-time/.
[6] M. O'Hagan, Translations massively open translation: Unpacking the relationship between technology and translation in the 21st century, Int. J. Commun. 10 (2016) 18.
[7] Potthast, M., Stein, B., Barrón-Cedeño, A., and Rosso, P. (2010). A Corpus-based Approach to Plagiarism Detection. Accessed on 12th February 2024.
[8] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., and Sutskever, I. (2019). Language Models are Few-Shot Learners. Accessed on 12th February 2024.