
Project report template (Be sure to write less than 4 pages)¹

Lincoln Vann-Wakelin
Leaf

Ian Osorio Perez

Bhairavi Ahilan
Leaf

With
Sage & CeSIA & CivAI & Apart Research

Abstract

Over the last decade, we have seen the considerable traction Artificial Intelligence has gained, likewise we have also seen a cumulation in concerns about the nature of such a technology. Concerns about the risks of AI often extend themselves to questions about the future of labour, and the varying effects among industries. Among this discussion, consensus has not fully been achieved as to the extent AI will affect human researchers, and we aim to shed light on the current capabilities of AI in the realm of academic research.

Our application enables users to input an arXiv paper link, from which it automatically extracts a summary and the author's name. Leveraging generative AI, the program then creates a new summary of the same paper, that closely mirrors the style and intellectual voice of the original author(s), aiming to replicate their writing. As a means to quantify the AI's effectiveness, our application employs a similarity checker that performs line-by-line semantic analysis, providing a standardized metric to evaluate how accurately the AI has been able to replicate the original summary. Furthermore, we added features to tailor this demo towards a more general audience, where we ask users to guess which paper they believe has been AI generated.

Our findings indicate that AI can convincingly reproduce a summary of any given research paper summary from the arXiv collection.

¹ Research conducted at the AI risks and capabilities demonstrations jam, 2024

1. Introduction

The claim of the transformative potential of AI is one that goes without debate, albeit its presence has not been met entirely with ease, especially with regard to the issue of mass unemployment that the adoption of AI is typically met with. This stems the question of whether AI will eventually sideline human researchers for a number of reasons. Note that AI systems are designed to handle large volumes of data and thus will be able to identify patterns, trends and anomalies far quicker than any human counterpart. In a similar respect, one of the more time consuming aspects of research is retrieving relevant information, yet the Natural Language Processing capabilities of any LLM is again much more efficient than its human counterpart. This means that AI has the ability to sift through essentially all relevant studies, and we aim to prove that it can further replicate any author or paper henceforth changing the landscape behind research forever.

Testing whether AI is human-like is a multifaceted task, so note that our end-goal is to see if the current developments to such a technology have enabled it to achieve human-like consensus from a researcher’s perspective, that is can an AI effectively replace a researcher right now? An answer to this question is not binary bound, synthesis of human like response might not fare as well for social science studies or mathematics since AI presents an aggregate of information as opposed to newly discovered ideas. Our idea of quantifying the progress beneath NLP in models, at the time of writing, is further presented below.

2. Overview

Guess the Original Summary

Below are two summaries. One is the original arXiv summary, and the other is AI-generated. Can you guess which one is the original?

Summary A We transmitted 4x25 Gbit/s over 5-km standard G-AMF using four different mode groups in combination with direct detection and on-off keying format (OOK). Due to the square-law detection in the receiver MIMO processing is not effective. Therefore proper mode group selective multiplexing and de-multiplexing is essential.	Summary B We transmitted 4x25 Gbit/s over 5-km standard G-AMF using four different mode groups in combination with direct detection and on-off keying format. Due to the square-law detection in the receiver MIMO processing is not effective. Therefore proper mode group selective multiplexing and de-multiplexing is essential. We propose a novel method for mode group selective demultiplexing based on the use of a single-mode fiber Bragg grating (FBG) as a mode filter. The FBG is designed to selectively reflect the desired mode group while transmitting the other mode groups. The FBG is fabricated using a standard fiber drawing process and is integrated into a standard single-mode fiber pigtail. The FBG is then spliced to the multimode fiber using
--	---

Which summary do you think is the original?

☐ Summary A

☒ Summary B

[Submit Guess](#)

In effort to offer hard proof as to how capable AI is as of now, and implicitly how important it will be in future years, we have created an application that inputs a link to an arXiv paper, to then be met with an AI generated version of the summary of the same paper. We

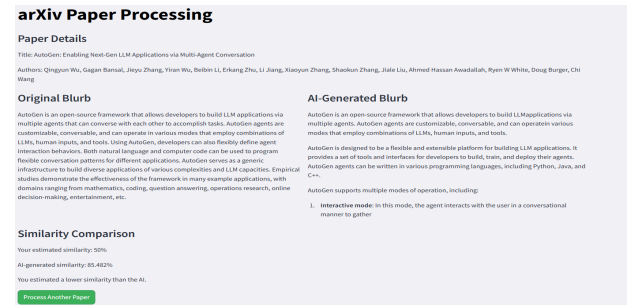
recognised the importance of a standard quantifiable metric upon comparison between the two, and so the AI version is then laid above the original summary, as presented in the paper, for a similarity check which is presented as a percent. The input format of a simple link, or a randomly generated paper as opposed to the author’s name and a summary was a means of further reinforcing the lack of human supervision needed for paper creation, given that all information is freely available on the internet.

As a means to further emphasize our point, we have incorporated an interactive feature whereby a paper is randomly generated and its AI counterpart is created. This is so we can ensure the user does not already know the author of the paper so that they will not be aware of their writing style. The two papers will be displayed

Apart Research Hackathon Submission

side by side, and the user has to choose which paper they believe was written by the human. This is more an ode to the NLP capabilities of AI, regardless of the domain or use case being applied, and thus is a general showcase of the fact that the capacity of AI and related technologies do not solely lie in the abundance of knowledge, but also the ability to pool this knowledge extremely quickly.

Here is our youtube video about the project: (<https://youtu.be/mczw8HNT3DU>)



3. Code

In the development stage of our API, we decided that we needed APIs to:

- Source ArXiv Papers
- Condense the blurb into a sentence or two
- Write a new blurb from the AI-generated condensed summary
- Compare text similarities

To do this, we decided to use the following APIs:

- For ArXiv papers we decided to use the native ArXiv api, given its FOSS nature and the fact that it is affiliated with ArXiv itself.
- For the AI segment, we decided to use HuggingFace's api, due to the free usage and the wide gamut of models available. We used a variety of models for each purpose, such as Mistral's 7B model to generate the new blurb and facebook's bart large cnn to condense the blurbs.

The front-end of the demo was written using streamlit, and is hosted on there right now ([\(\(ai-research--paper.streamlit.app\)\)](https://ai-research--paper.streamlit.app)). We chose streamlit as we felt it's always active nature and the ease of use was paramount to meeting the fast deadline.

The code is hosted at https://github.com/Masterlincs/ai_hackathon

4. Discussion and Conclusion

On average, our results of the similarities between the AI and human created papers tended towards an average of 75% similar, which is extremely high given that this was a general purpose AI. More time might have meant that we would look into potentially developing a research specific AI, be it through training a model with a variety of authors' styles prior to release, or creating our neural networks from which a model would learn to replicate any authors given style. In theory, this could tend to achieve an average above 85%.

We mainly found that between the two texts, finding the human version was a trivial task. Trials have left accuracy at guessing around 80%, which shows that the current AI is flawed at imitating a human. This is mainly due to our choice of LLM, that being Mistral 7B - a year old model whose only merit is its free cost at huggingface. To improve this a top tier model e.g. llama 7b or claude 3 would be trained specifically on research papers to allow its vocabulary to adapt.

To conclude, our findings indicate that AI is capable of imitating humans, beyond simple sentence formation, with upto 90% accuracy. This diminishes at the level of research papers however, as the in depth topics and little room for errors allow for AI to be singled out near immediately by a professional in a field. As AI reaches a level of sophistication that it can replicate a research paper, we anticipate many changes in the research process. This could be in forms of automation of the writing process (using models like ours, with altered parameters), or simply collaboration to spark innovation in all parts of research.