

## RESEARCH PROPOSAL

### OVERVIEW

Much information has been digitalized and disseminated in high volume and velocity that have consequence for certain users with learning disabilities. In order to accommodate a society, the users have been used a variety of creative approaches to consume information efficiently. In not so doing, the most aspects of their life may be affected by, for example, missing due dates for college, personal projects, and career. Constant reliance on creativity may become increasingly stressful. For this specific reason, the automatic text summarization, which is one of Artificial Intelligence (AI) for Accessibility solutions, should further be optimized.

But since the objective of this algorithm is to summarize long document, and at the same time, conserve important context, the text summarization has been a challenging Natural Language Processing (NLP) AI problem. To the extent of what makes this proposal deliverable, the research proposal covers essential sections.

### PROBLEM IDENTIFICATION

How can the text summarization surpass the Recall-Oriented Understudy for Gisting Evaluation-1 (ROUGE-1) score of Hierarchical Learning for Generation with Long Source Sequences (HAT-BART)? Can the data analytics be readily available on or before December 1, 2021? Does this analytics include the qualitative evaluation?

### CONTEXT

A term *disability* is socially constructed. This term refers to the marginalized people facing accessibility barriers created by the society, rather than disability itself as an obstacle to overcome. Based on the World Health Organization assessment, at least one billion people in the world are identified as disabilities, and almost 200 million have significant difficulties to accommodate a society. This population consists of developmental, cardiovascular, motor, and learning disabilities, to name a few. Providing the text summarization for learning disability is a main research focus (WHO, 2011).

An underlying condition of learning disability may be due to either genetic or environmental factors that change neurological function, which, in turn, affects cognitive processes (i.e., dyslexia, autism, ADHD). The people with learning disabilities require some creative ways to circumvent accessibility barriers, and unfortunately, several youths who do not receive evaluation undergo difficulties with academics without knowing why (LDA, 2021). Even though the text summarization may be used as their cognitive and attentive aids, the inherent problem with AI is that, firstly, this

algorithm can only perform from what data it learns, and secondly, this algorithm has prediction errors.

The study demonstrates the prediction error in computer vision that has negative consequence for a person who is blind, especially, being overtrusting of AI to the extent of safety hazards. Change in metrics should be in consideration suggested by author (Morris, 2020). Pertaining to this ethical discourse, the text summarization should not only include quantitative but qualitative evaluation.

## CRITERIA FOR SUCCESS

Relevant criteria for successes are:

- The ROUGE-1 and ROUGE-2 metrics for the new text summarization are in use to compete against HAT-BART model as a benchmark (Paperswithcode, 2021).
- When the prediction errors occur, the errors should not be severe to the extent that could not be rectified even with human assistance.
- The reproducible research and presentation should be readily available on or before December 1, 2021.
- The data analytics should convince stakeholders and the model turn into deployable production available to the people with learning disabilities.

## SCOPE OF SOLUTION SPACE

This research focuses on developing the text summarization as AI solution for the people with learning disabilities. However, this research does not build user-friendly interface and it does not create the general AI. More specifically, this research focuses on developing the text summarization learned from PubMed and ArXiv documents and abstracts.

## CONSTRAINTS

- Processing the PubMed and ArXiv datasets containing over 40,000 scientific publications may be time- consuming, which requires efficient computation.
- The qualitative analysis is difficult, if not impossible, to completely evaluate due to large number of abstracts that may easily be overlooked.
- Developing the text summarization might not have ability to extrapolate ROUGE-1 and ROUGE-2 scores accurately on outside scientific publication abstracts.

## POTENTIAL STAKEHOLDERS

Program Manager

Lead Data Scientist

## DATA SOURCES

[ArXiv](#)

[PubMed](#)

[More Info](#)

## REFERENCES

“Natural Language Processing.” *Text Summarization*. Paperswithcode, Oct. 2021. <https://paperswithcode.com/task/text-summarization>

“Types of Learning Disabilities.” Learning Disabilities Association of America, Oct. 2021. <https://ldaamerica.org/types-of-learning-disabilities/>

“World Report on Disability.” *Disability – a global picture*. World Health Organization and World Bank, 2011. [https://www.who.int/disabilities/world\\_report/2011/report.pdf](https://www.who.int/disabilities/world_report/2011/report.pdf)

Morris, MR. “AI and Accessibility.” *Viewpoint*. Communications of the ACM, Jun. 2020. <https://cacm.acm.org/magazines/2020/6/245157-ai-and-accessibility/fulltext>