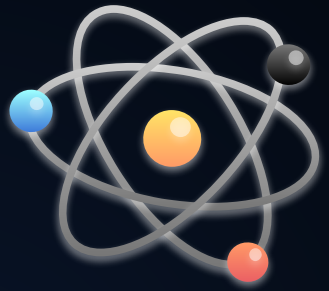
A COMPARISON OF TEXT SUMMARIZERS: Abstractive, Extractive, and GPT

Ari Kanevsky, Scott Nelson, Will Judy



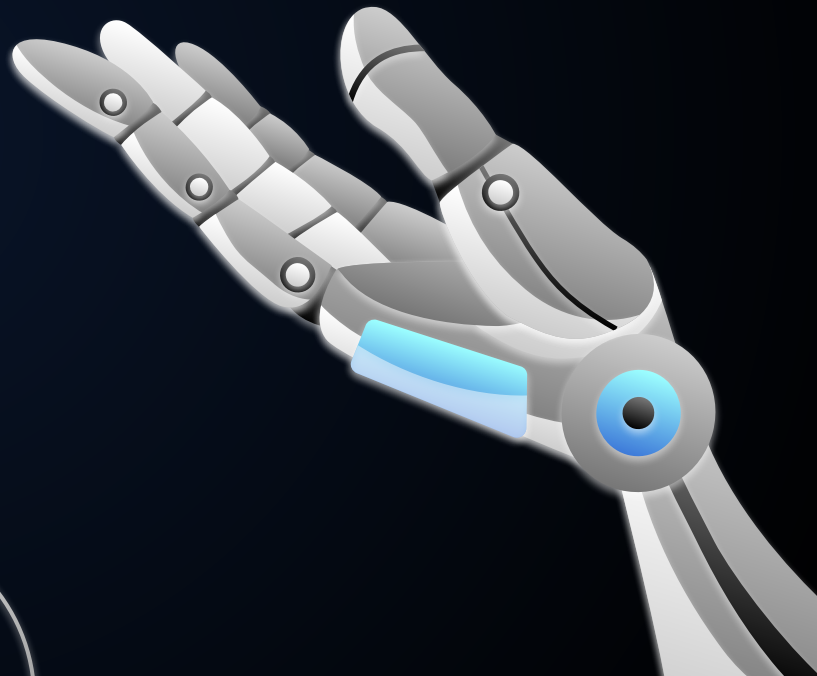
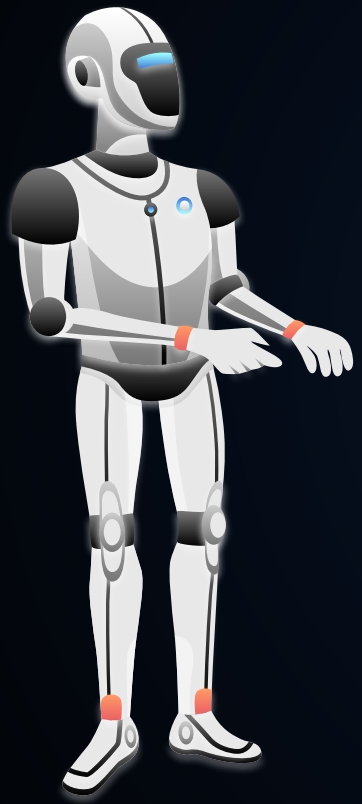


TABLE OF CONTENTS



01

OVERVIEW & HYPOTHESIS

Why Summarization?

- Very applicable and powerful across disciplines
- Great use of Transformers
- Useful for web scraping and consolidating content

Types of Summarization Models

Extractive

Simplest Approach
Identifies key phrases &
“extracts” to reuse in summary
LSA Model

Abstractive

Neural-Based
Finds ‘Big Picture’ and
generates **novel** summary
T5 Model

GPT

High-Complexity
Use of Transformers
Unknown Architecture
State of the Art Performance
GPT 3.5 Turbo

HYPOTHESIS

Using a ROUGE score for evaluation, across various complexity levels of text, we predict the following performance between our three summarizers:

Extractive < Abstractive < GPT

02

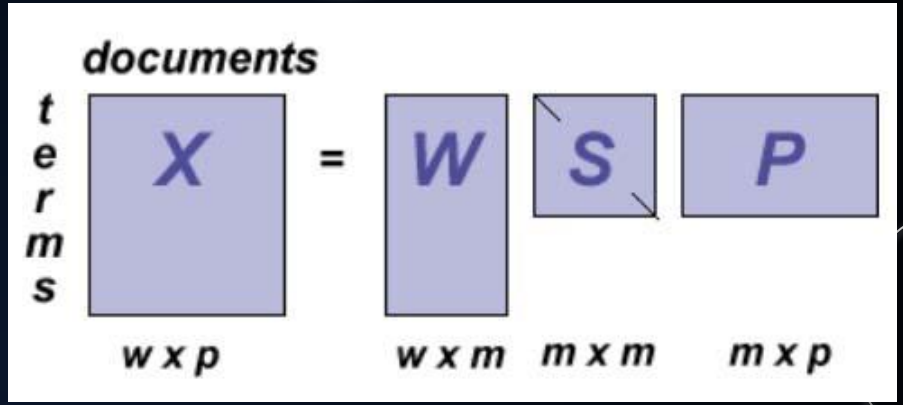
MODELS & METHODS

The Set-Up for Experimentation

LSA - Extractive

- LSA : Latent Semantic Analysis

- Constructs TF-IDF matrix



- LSA then finds low rank approximation of the matrix
  - Lowering rank is expected to merge dimensions associated with terms that have similar meanings - mitigates problems of identifying synonyms
  - Also known as Singular Value Decomposition

T5 - Abstractive

- T5 is an encoder-decoder model pre-trained on a multi-task mixture of unsupervised and supervised tasks and for which each task is converted into a text-to-text format.

- T5 works well on a variety of tasks out-of-the-box by prepending a different prefix to the input corresponding to each task, e.g., for translation: *translate English to German: …*, for summarization: *summarize: ….*

- Self-supervised training uses corrupted tokens, masking them similar to BERT

GPT - Transformer

- Uses 3.5 Turbo

- Give it prompt: “Please summarize this text: {text}”

- Given the wide range of training tasks, we expect GPT to provide the most helpful summaries

The Data - HuggingFace 🤗

BILLSUM

Collection of California & National Bills along with their summaries. Majority are from 2015-2016

CNN-DAILYMAIL

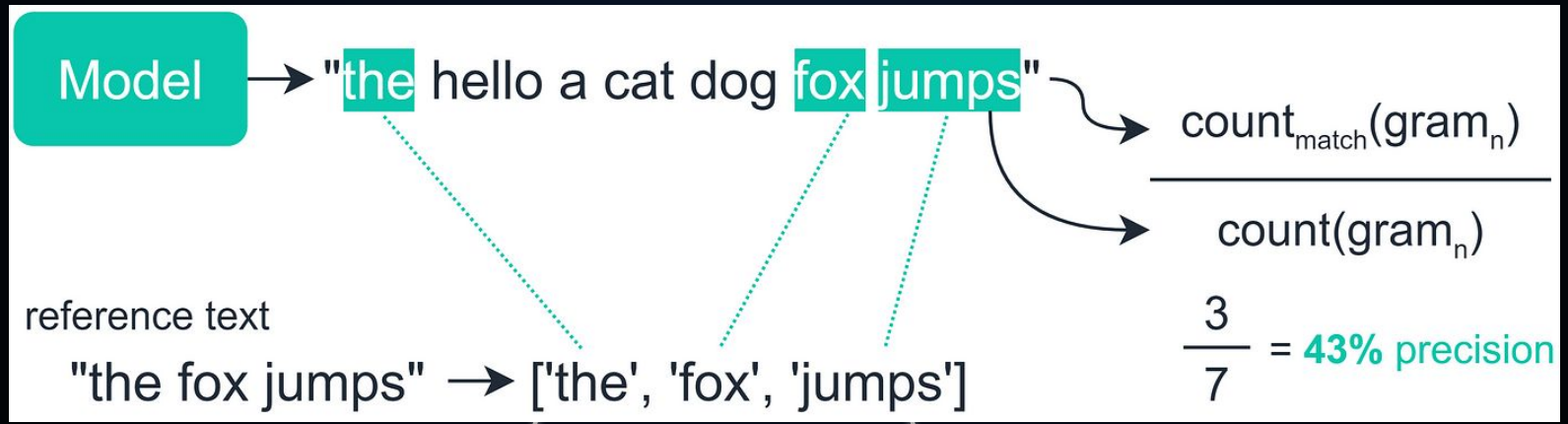
Huge collection of CNN & DailyMail News stories from 2007-2015, along with their summaries

ROUGE Scoring

N-Gram similarity scores

We used ROUGE Scores between the three models to compare the summaries to one another.

Computes Precision and Recall, then provides **F-Measure**

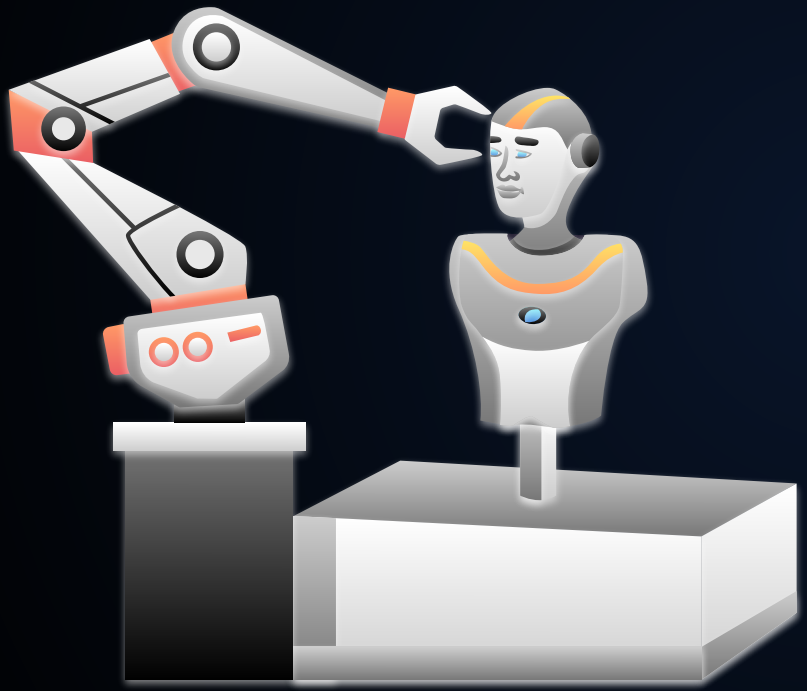


Putting it Together - Our Program

A) TKinter popup allows user to select text via medium of their choice

B) Summarizes chosen text using each model

C) Compares results to one another with a ROUGE score



03

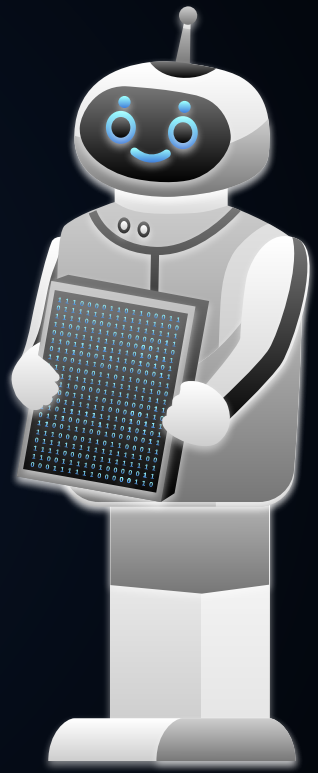
DEMO

See it all in action!

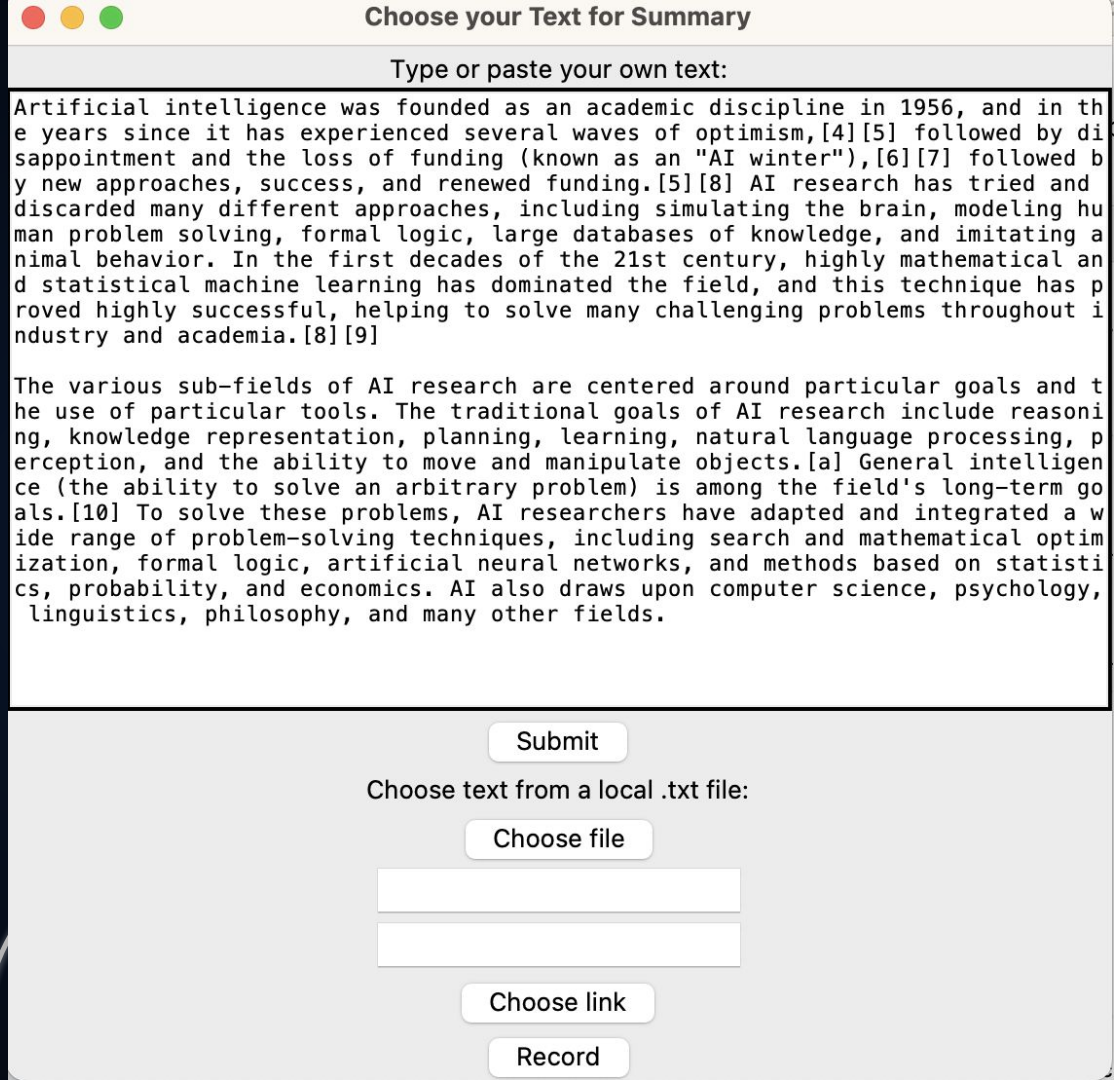
04

DISCUSSION

Results, Findings, Takeaways, Next Steps



Example: Wikipedia article on AI (Standard)



LSA Summarizer

AI research has tried and discarded many different approaches, including simulating the brain, modeling human problem solving, formal logic, large databases of knowledge, and imitating animal behavior.'

T5 Summarizer

Artificial intelligence was founded as an academic discipline in 1956, and in the years since it has experienced several waves of optimism, followed by disappointment and the loss of funding, followed by new approaches, success, and renewed funding. In the first decades of the 21st century, highly mathematical and statistical machine learning has dominated the field, and this technique has proved highly successful, helping to solve many challenging problems throughout industry and academia.

GPT-3.5-turbo

AI has already had a major impact on society, with applications in fields such as healthcare, finance, transportation, entertainment, and more. AI technologies are already being used to detect cancer, personalize medicine, prevent fraud, optimize transportation, and even create art and music. As AI continues to advance, it has the potential to transform industries and change the way we live and work.

However, there are also concerns about the risks and ethical implications of AI. Some worry that AI could become uncontrollable or be used to harm people or society. There are also concerns about AI taking over jobs traditionally done by humans and increasing economic inequality. As a result, there is a growing focus on ensuring that AI is developed in a responsible and ethical way, with appropriate safeguards and oversight to ensure its benefits are maximized while minimizing its risks.

ROUGE Evaluation (Standard)

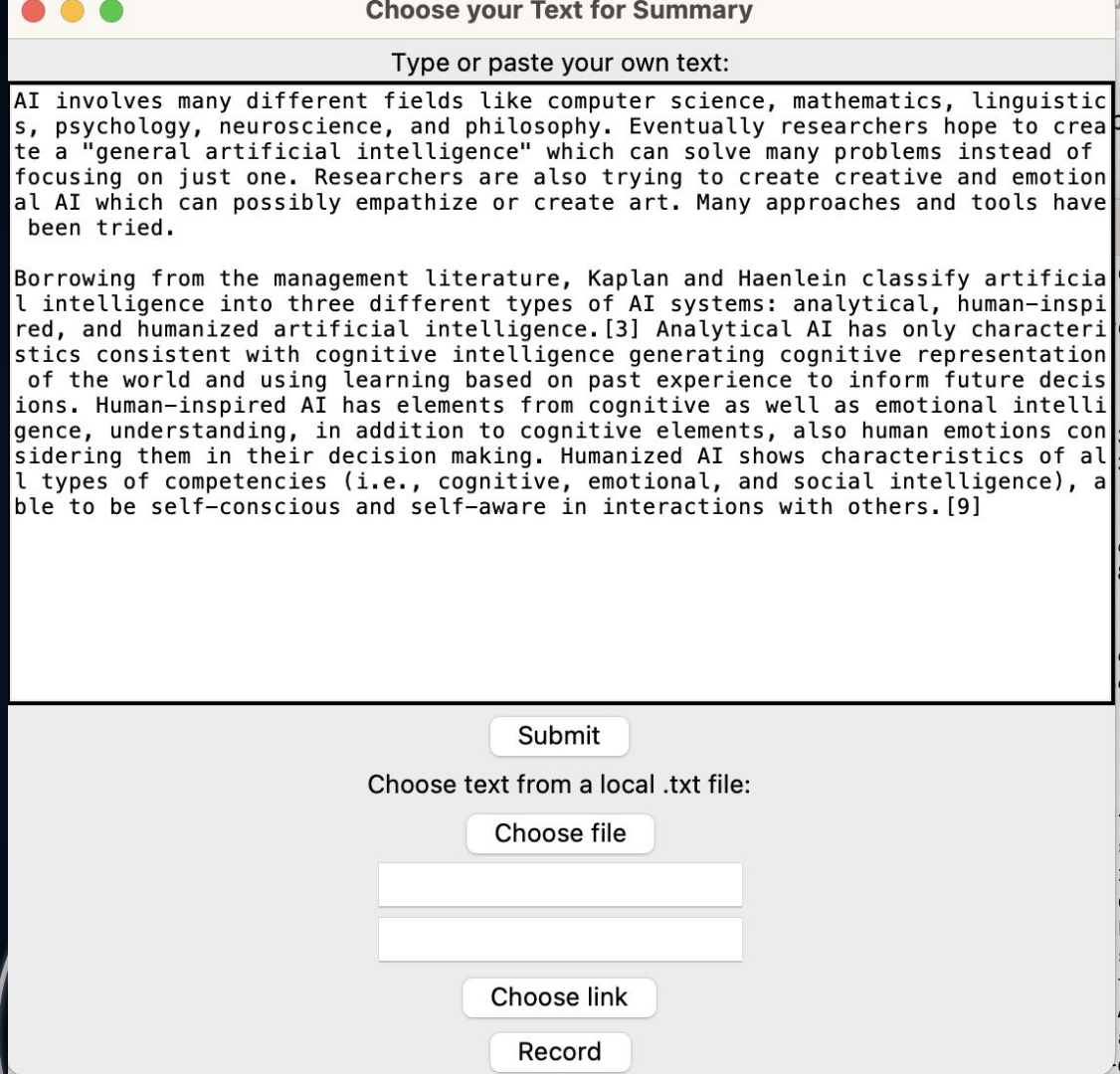
* Reported scores are “F-Measure”

LSA vs. T5
-Unigram similarity: 18%
-Bigram similarity: 0%
-Overall similarity: 11%

LSA vs. GPT
-Unigram similarity: 8%
-Bigram similarity: 0%
-Overall similarity: 7%

T5 vs. GPT
-Unigram similarity: 19%
-Bigram similarity: 2%
-Overall similarity: 12%

Example: Wikipedia article on AI (Simple)



LSA Summarizer

AI involves many different fields like computer science, mathematics, linguistics, psychology, neuroscience, and philosophy.

T5 Summarizer

AI involves many different fields like computer science, mathematics, linguistics, psychology, neuroscience, and philosophy. Eventually researchers hope to create a "general artificial intelligence" which can solve many problems instead of focusing on just one. Many approaches and tools have been tried.

GPT-3.5-turbo

Furthermore, AI systems can also be categorized based on their application areas. Some of the most common ones include:\n\n1. Natural Language Processing (NLP): NLP is a branch of AI that deals with the interaction between computers and humans using natural languages like English, French, or Chinese.\n\n2. Computer Vision (CV): CV is an AI technology that enables computers to interpret and understand the visual world through digital images and videos.\n\n3. Robotics: Robotics is the field of AI concerned with the development of robots that can perform tasks that would typically require human intervention.\n\n4. Game AI: Game AI is the area of AI that focuses on creating intelligent agents that can play games like chess, poker, or video games.\n\n5. Expert systems: Expert systems are AI systems that can simulate the decision-making ability of a human expert in a specific domain.\n\n6. Autonomous vehicles: Autonomous vehicles are self-driving cars, trucks, or drones that can operate without human intervention. \n\nOverall, the field of AI is vast and continually evolving, with researchers and engineers striving for new breakthroughs in technology that can drive innovative solutions to current and future problems.

ROUGE Evaluation (Simple)

* Reported scores are “F-Measure”

LSA vs. T5
-Unigram similarity: 33%
-Bigram similarity: 32%
-Overall similarity: 33%

LSA vs. GPT
-Unigram similarity: 5%
-Bigram similarity: 0%
-Overall similarity: 4%

T5 vs. GPT
-Unigram similarity: 15%
-Bigram similarity: 1%
-Overall similarity: 8%

Findings

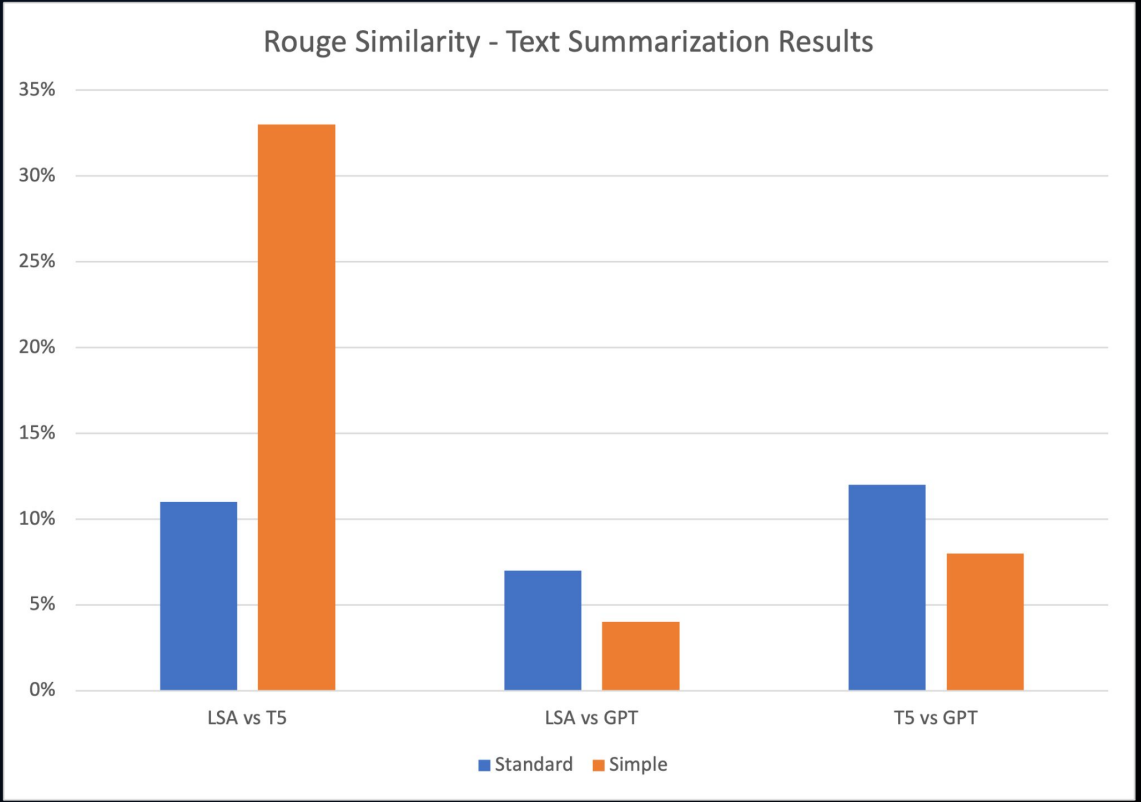
LSA & GPT are least similar regardless of task

The models with the highest similarity depends on the complexity of the task

- If summarizing simple text, T5 and LSA will be the most similar
- If summarizing text classified as standard to complexed, T5 and GPT will be the most similar

This emphasizes the strength of transformer-based models in complicated scenarios

Transformers are able to “understand” the underlying message



Conclusion

Hypothesis: Using a ROUGE score for evaluation, across various complexity levels of text, we predict the following performance between our three summarizers:

Extractive < Abstractive < GPT

Confirmed - ish … We used our common sense to say GPT provides the best results

Since we proved Abstractive (T5) is most similar to GPT, we believe it's a better method across all instances than Extractive (LSA)

Altogether - the most important lesson is that HuggingFace has a TON of awesome resources for any transformer-based projects. Datasets, Tutorials, Community Forums, etc. Check em out!

Next Steps

Implement the summarizers to gain knowledge!

- Summarize Scholarly articles
- Summarize Wikipedia articles
- Summarize News articles
    - **Hackathon - Scott & Ari - News Aggregator**
- Keep having fun :)

Sources

https://towardsdatascience.com/the-ultimate-performance-metric-in-nlp-111df6c64460

https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9328413

https://iv.cns.iu.edu/sw/lsa.html

https://research.google/pubs/pub46201/

https://huggingface.co/spaces/evaluate-metric/rouge

https://huggingface.co/SEBIS/code_trans_t5_base_code_documentation_generation_python


slidesgo