SMARTSUMM: COMPARING ABSTRACTIVE & EXTRACTIVE SUMMARIZATION TECHNIQUES

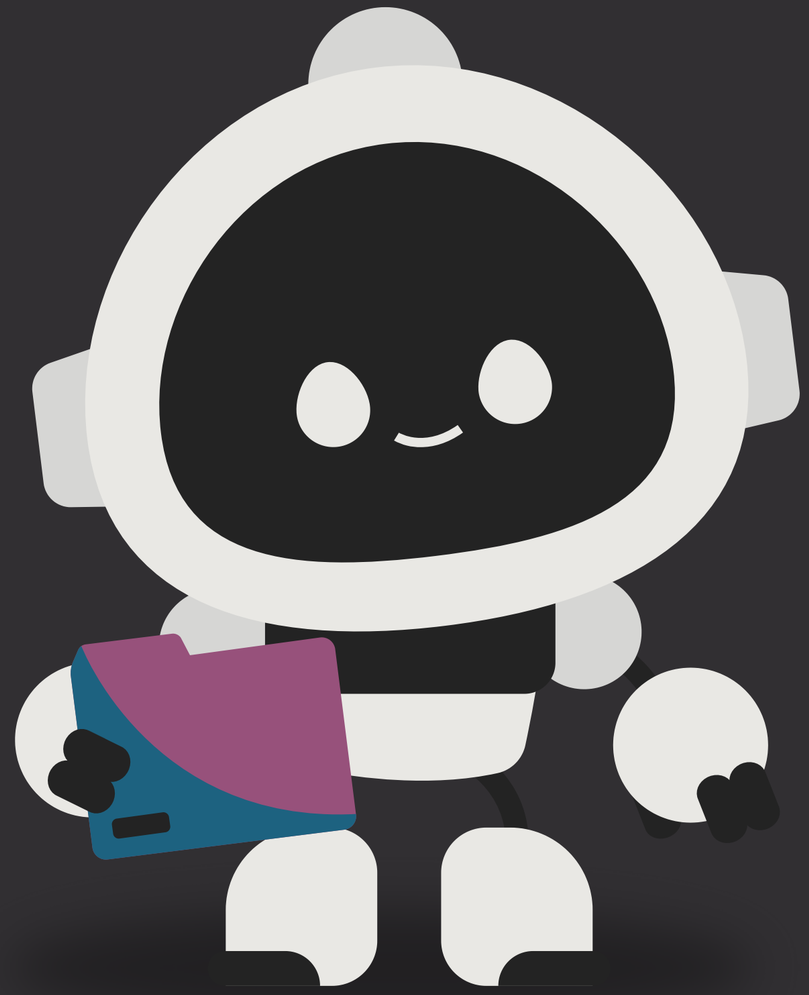
INTRODUCTION

Goal:

Compare modern abstractive models (BART, T5) with the classic extractive method (TextRank) using real news data.

Why this topic?

- Summarization is essential in NLP
- Transformers are popular—we tested if they’re better
- We wanted real-world experience with cutting-edge tools



NLP TECHNIQUE & DOMAIN

Domain: News Articles (XSum dataset)

Techniques Used:

- BART and T5 for abstractive summarization
- TextRank for extractive summarization

Why News?

- Articles are long, need concise summaries
- Great testbed for evaluating summarization models

SYSTEM FUNCTIONALITY & RESULTS

- Input: News article from XSum
- Models generate summaries
- Results compared using ROUGE scores
- Outputs saved in CSV for analysis

Model	ROUGE-1	ROUGE-2	ROUGE-L
BART	0.48	0.25	0.41
T5	0.2	0.04	0.14
TextRank	0.16	0.03	0.11

CHALLENGES & SOLUTIONS

Slow inference with Transformer models → used caching and batch processing

ROUGE limits → combined with manual review to check summary quality

Short or incomplete summaries → tuned generation parameters (min_length, max_length)

LESSONS LEARNED & FUTURE WORK

- TRANSFORMER MODELS DELIVER BETTER SUMMARIES BUT NEED TUNING
- EXTRACTIVE METHODS ARE FAST BUT LESS ACCURATE
- FUTURE PLANS:
- FINE-TUNE MODELS ON DOMAIN-SPECIFIC DATA
- EXPLORE NEWER MODELS LIKE PEGASUS
- BUILD A USER-FRIENDLY SUMMARIZATION APP

GITHUB: SMARTSUMM REPO

MARIAM ELRAFEI 22101442
OMNIA ADEL 22101925

THANK YOU