

TL;DR -

WHY DO SUMMARIES MAKE SENSE
(OR DON'T)?

AUTHOR: EMILIA WIŚNIOS



“

No cards, no presents, and he would be spending the evening pretending not to exist. He gazed miserably into the hedge. He had never felt so lonely. More than anything else at Hogwarts, more even than playing Quidditch, Harry missed his best friends, Ron Weasley and Hermione Granger. They, however, didn't seem to be missing him at all. Neither of them had written to him all summer, even though Ron had said he was going to ask Harry to come and stay.

Why the summary looks like this?

Harry's feeling lonely and depressed on his birthday, because he has no friends and no one cares about him.



“

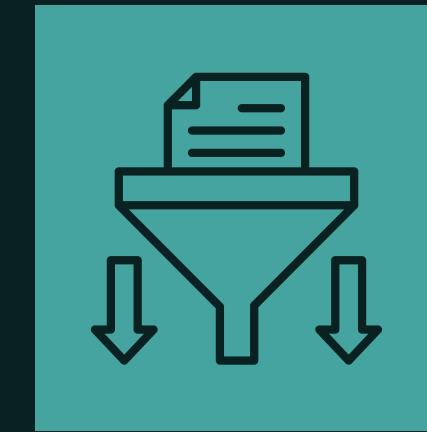
Ms. Heard published her op-ed with actual malice. She knew that Mr. Depp did not abuse her and that the domestic abuse allegations that she made against him in 2016 were false. She knew that the testimony and photographic "evidence" that she presented to the court and the supporting sworn testimony provided by her two friends were false and perjurious.

Is this a good summary?

Amber Heard is a fraudulent, manipulative, narcissistic sociopath.



Types of Summaries



Extractive

Choosing the most important sentences.



Abstractive

Making new text based on the given information.





WHY DO WE NEED XAI IN EXTRACTIVE SUMMARIZATION?

- Named entities
- First sentences case
- Long vs. short sentences



WHY DO WE NEED XAI IN ABSTRACTIVE SUMMARIZATION?

- Hallucinations
- Summary style

FINDING THE BALANCE



IS ATTENTION ALL YOU NEED?

> METHOD

The main idea is to visualise
attention weights.



IS ATTENTION ALL YOU NEED?

➤ PROBLEM

“

However, a debate has started to cast **doubt** on the **explanatory power** of attention in neural networks.



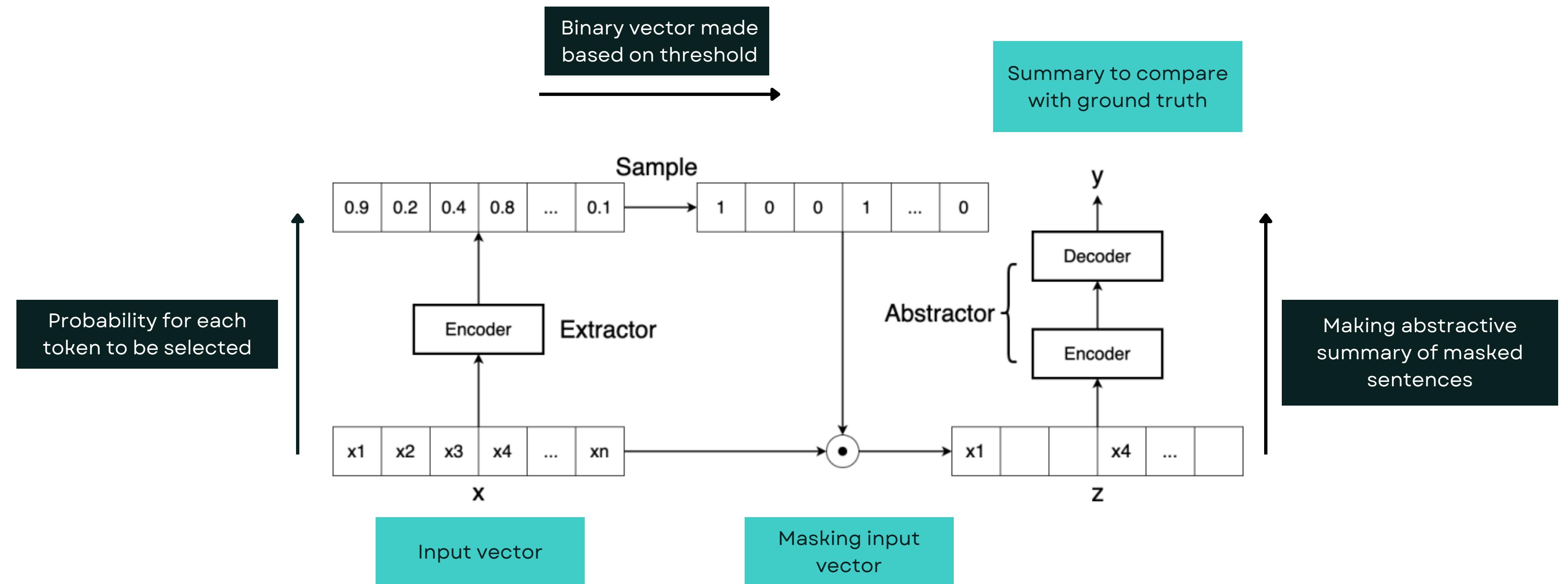
ABSTRACTION THROUGH EXTRACTION



METHOD

Choose the most informative sentences and make them abstract.





ABSTRACTION THROUGH EXTRACTION



PROBLEM

We still can't control the
abstractive part.



INTERPRETABILITY THROUGH CONTROL

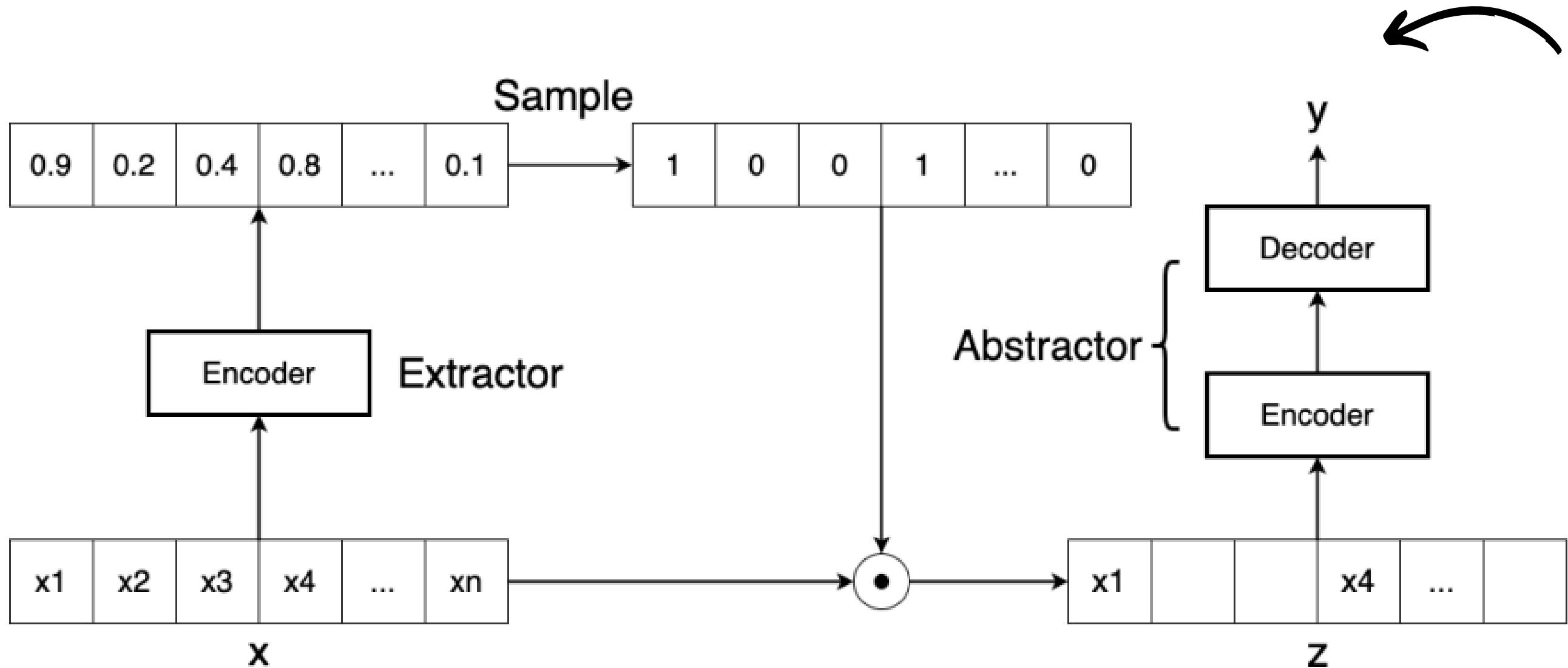


METHOD

Control the abstractor by adding rewards for specific behaviours.

MY APPROACH





Rewards for:
keeping the style,
correctness of
information,
ending the sentence

MY APPROACH



INTERPRETABILITY THROUGH CONTROL



BENEFITS

We achieve some level of interpretability in abstractive summaries.

MY APPROACH



Call for ideas



Evaluation

In unsupervised setting, how to say if the summary is good?

Features

What features would you expect for a summary to have?



CONTACT ME!



EMILIA WIŚNIOS



wisniosemilia@gmail.com



emilia-wisnios