**ADDZY: Slide 1, 2**

Hi, my name is Addzy, and together with my teammate Gaurav, we are going to be presenting our findings on the role of AI in combating fake news.

So, what is fake news?

Most of you must have already come across fake news on the internet. Be it Whatsapp forwards or articles forwarded to you by relatives, we have always been affected by it in some way or the other…but why?

Why does fake news spread 10 times faster than what is true? Because it is mostly something that is either too good to be true or something that makes you rage and share it with others who are like-minded.

**GAURAV: Slide 3**

Even though this seems like a minor thing at first, it can be disastrous when weaponized. One need not look further than the recent elections.

Lots of brilliant folks have tried to combat this issue with a method that seems most promising - AI. It’s mostly been portrayed as a classification problem which in simpler words means looking at some text and throwing it into either of two buckets: true or not true. But it isn’t as simple as it seems to be and there is to more to it than meets the eye. But why?

**ADDZY: Slide 4, 5**

A program cannot understand what is true and what is false, as it doesn’t have sufficient contextual knowledge of the world, like you or I.

In the next three minutes, we are going to showcase what we did to justify our claim.

We compared results obtained by running LSTM’s against state-of-the-art machine learning models like BERT and RoBERTa, which though sound like human names are actually textual analysis models that make use of Deep Learning.

Let’s briefly discuss one key aspect of these models: Self-Attention. These models use Self-Attention which encodes the information of how one word relates to another within the given context. Let’s a look at the image shown (Lecture Self-Attention Image):

**GAURAV: Slide 6**

We can see that certain patterns within the text can be utilized to extract features out of the sentence. If I were to say: “We were walking alongside the bank of a river.” Without the word river, the word bank could be construed in multiple ways. This tells us why context is important before assigning a value to a word. And context is exactly what self-attention gives us.

Nowadays, all the popular models utilize self-attention. Before this, we had LSTM’s which was the way to go…

You might be wondering what LSTM stands for…it means Long Short Term Memory. As the name suggests the model tries to maintain a memory of what it has seen in the recent past and tries to make sense of the words. But even though this method leads to good results on paper, it is lacking in one key aspect…can you guess what it is?

**ADDZY: Slide 6**

Context!!

As we discussed earlier, without understanding the context of what is being said, we cannot make an accurate judgment. Hence, Self-Attention comes into the picture, BERT and RoBERTa utilized this concept which helped us to develop better models. RoBERTa is a Robustly optimized and sufficiently trained model built on top of BERT.

**Slide 7**

Let’s look at some insights into the dataset these models were trained on. The image here shows the most common words or phrases used in fake news articles. As you can see the entirety of the dataset is biased towards certain topics, what does this tell us?

This tells us that, there can be no single dataset that can cover everything that is happening in the world at the moment.

**GAURAV: Slide 8**

So, onto the results…

While the models performed exceedingly well with what they were given, they were abysmal when it came to predicting anything that was a little out of the way from what they had come across. I know…it’s disappointing.

**Slide 9**

So, what are we saying? Are we saying that there is no solution to this problem? No, everything that we have seen so far tells us that this problem cannot have a definitive solution at this stage with the approaches we currently have. One cannot say whether a piece of text is “fake or not”, a program can only suggest how much of the article might be true based on the limited knowledge it has.

However, it will be possible when we achieve artificial general intelligence which is the singularity we hope to see in our lifetime.

**Slide 10**

Thank you.