

A decorative network diagram consisting of several blue circular nodes of varying sizes connected by thin black lines. The nodes are arranged in a non-linear fashion, with some having multiple connections. The diagram is positioned around the central text area, with nodes appearing in the top right, bottom left, and bottom right corners.

Week 7: Evals Cont.

Alfonso Vieyra, Denise Wei,
Raha Ghazi, Tom Gao

Evaluations

- Instead of using the pipeline function from the transformers library we are now loading the tokenizer and model manually
 - This allows for more control over what the input to the model is
- Continuing to refine evaluation process before we begin dedicated GPU hours of our computing cluster to running all the models
 - Now tracking how long running the model on each row takes and how much GPU utilization it demands

...

Summary Metrics

- Per Mo's feedback we will continue looking into all the ROUGE variants to better understand them and have a better idea of which to prioritize/utilize
 - **Also looking at ways to implement manual evaluation of a random sample**
- Also now looking at BERTScore along with other forms of Model-Graded Evaluation
- Looking at ways to evaluate the summary against the original dialogue and not just against the observed summary

...

Transcripts

- When we receive the transcripts, what should our main use of them be?
 - Save them for fine tuning?
 - Use them for further evaluation?
 - Use them for validation after fine tuning on current data?
- Will the transcripts come with summaries?

...