

Podcasts Track results — (podcast\_baselines) Baseline runs for Podcast track

Summary Statistics	
Run ID	bartpodcasts
Task	Summarization
Run type	automatic
Summarization model	Supervised neural or deep learning model; Transfer learning or other pre-trained model
Summarization method	Abstractive
Number of summaries	179

Overall measures	
Summaries rated "excellent"	25 (14.0%)
Summaries rated "good"	50 (27.9%)
Summaries rated "fair"	67 (37.4%)
Summaries rated "bad"	37 (20.7%)

ROUGE score compared to episode description	
ROUGE_L P score	0.2078 (95%-conf.int. 0.19668 - 0.21971)
ROUGE_L R score	0.2101 (95%-conf.int. 0.20063 - 0.22000)
ROUGE_L F score	0.1663 (95%-conf.int. 0.15924 - 0.17371)

Qualitative judgments	
Q1: Does the summary include names of the main people (hosts, guests, characters) involved or mentioned in the podcast?	51.4%
Q2: Does the summary give any additional information about the people mentioned (such as their job titles, biographies, personal background, etc)?	34.6%
Q3: Does the summary include the main topic(s) of the podcast?	78.2%
Q4: Does the summary tell you anything about the format of the podcast; e.g. whether it's an interview, whether it's a chat between friends, a monologue, etc?	49.2%
Q5: Does the summary give you more context on the title of the podcast?	65.4%
Q6: Does the summary contain redundant information?	14.0%
Q7: Is the summary written in good English?	86.6%
Q8: Are the start and end of the summary good sentence and paragraph start and end points?	62.0%

