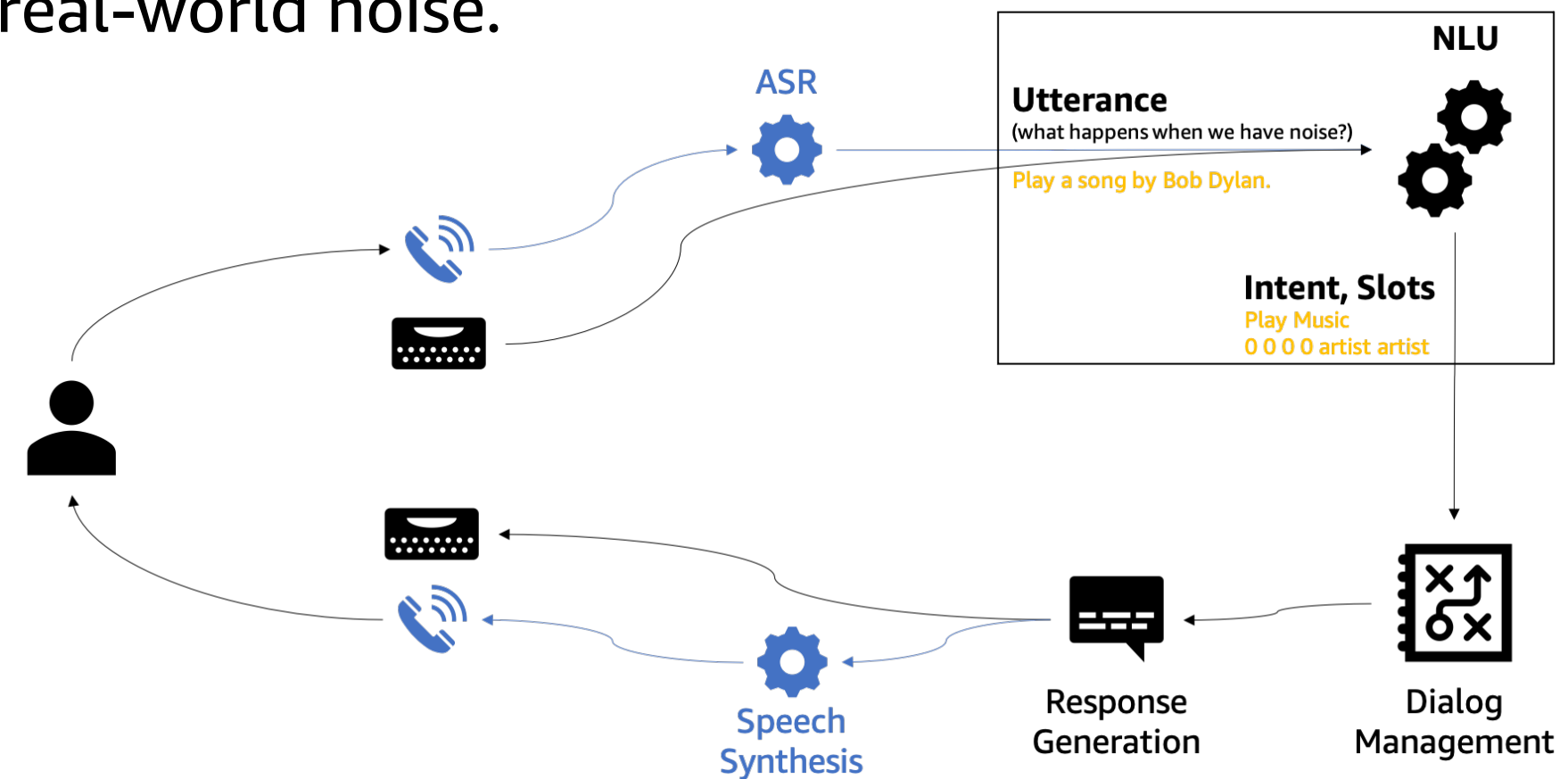# On the Robustness of Goal Oriented Dialog Systems

Jason Krone, Sailik Sengupta, Saab Mansour

AMAZON LEX

We investigate the robustness of BERT-based Intent Classification (IC) and Slot Labeling (SL) system to real-world noise.

We filter live conversations to identify the noise types.

Rule based and existing research to create training data for augmentation.

We leverage human experts to create noised version of ATIS and Snips test data.

Phenomena	Model	ATIS IC Acc. Orig.	Noisy	SL F1 Orig.	Noisy	Snips IC Acc. Orig.	Noisy	SL F1 Orig.	Noisy
Casing	BERT	**98.8**	73.1	**95.6**	16.2	**99.0**	23.3	**96.3**	5.0
	+Aug.$^{50\%}$	98.7	98.1	94.6	94.9	98.9	98.8	96.0	90.0
Spell.	BERT	**98.8**	97.3	**95.6**	87.0	99.0	98.7	96.3	91.2
	+Aug.$^{20\%}$	98.7	97.8	95.5	94.3	**99.1**	98.8	**96.4**	95.3
Abbreviations	BERT	96.0	86.5	**91.9**	75.9	97.9	96.9	95.8	85.2
	+Aug.$^{15\%}$	**96.6**	88.9	91.5	89.1	**98.3**	**98.3**	**96.6**	94.8
Morph.	BERT	**98.0**	97.1	92.3	93.5	98.0	96.9	95.8	85.2
	+Aug.$^{10\%}$	96.8	95.9	94.0	**94.1**	**98.3**	**98.3**	**97.2**	88.8
Synonyms	BERT	96.9	57.6	89.0	86.6	98.8	98.3	95.2	94.8
	+Aug.$^{10\%}$	97.2	90.4	**92.3**	90.2	99.2	98.7	**96.9**	96.4
Paraphrasing	BERT	**97.1**	90.9	88.3	86.7	98.8	98.5	95.5	93.7
	+Aug.$^{15\%}$	**97.1**	90.5	**88.8**	86.8	**99.0**	98.8	**95.9**	94.4

Real-world noise can affect the performance of NLU systems.

Training data augmentation can help boost the IC (avg 11.75%) and SL (avg 17.3) performance of NLU systems.

- Scope for improvement (eg. paraphrases, synonyms, morphological variants)

- One model to tackle all

Noise type	Training generation	Test generation	Examples
Casing	Rule Based	Rule Based	BOOK A FLIGHT FROM SAN JOSE TO NYC
Misspellings	Synthetic generation (Hasan et al., 2015)	Spell-Correction Pairs	*buk* a flight from San Jose to NYC / book a *glight* from San Jose to NYC
Synonyms	WordNet + LM Perplexity Filter	Human	*reserve* a flight from San Jose to NYC / book a *ticket* from San Jose to NYC
Paraphrases	Back Translation	Human	*can you book me* a flight from San Jose to NYC / *i'm looking for flights* from San Jose to NYC
Abbreviations	Rule Based	Human	book a flight from san jose 2 nyc
Morph. Variants	Internal Database + LM Perplexity Filter	Human	start *booking* a flight from san jose to nyc

Noised test data publicly available for benchmarking robustness of NLU systems!

{kronej, sailiks, saabm} @ amazon.com