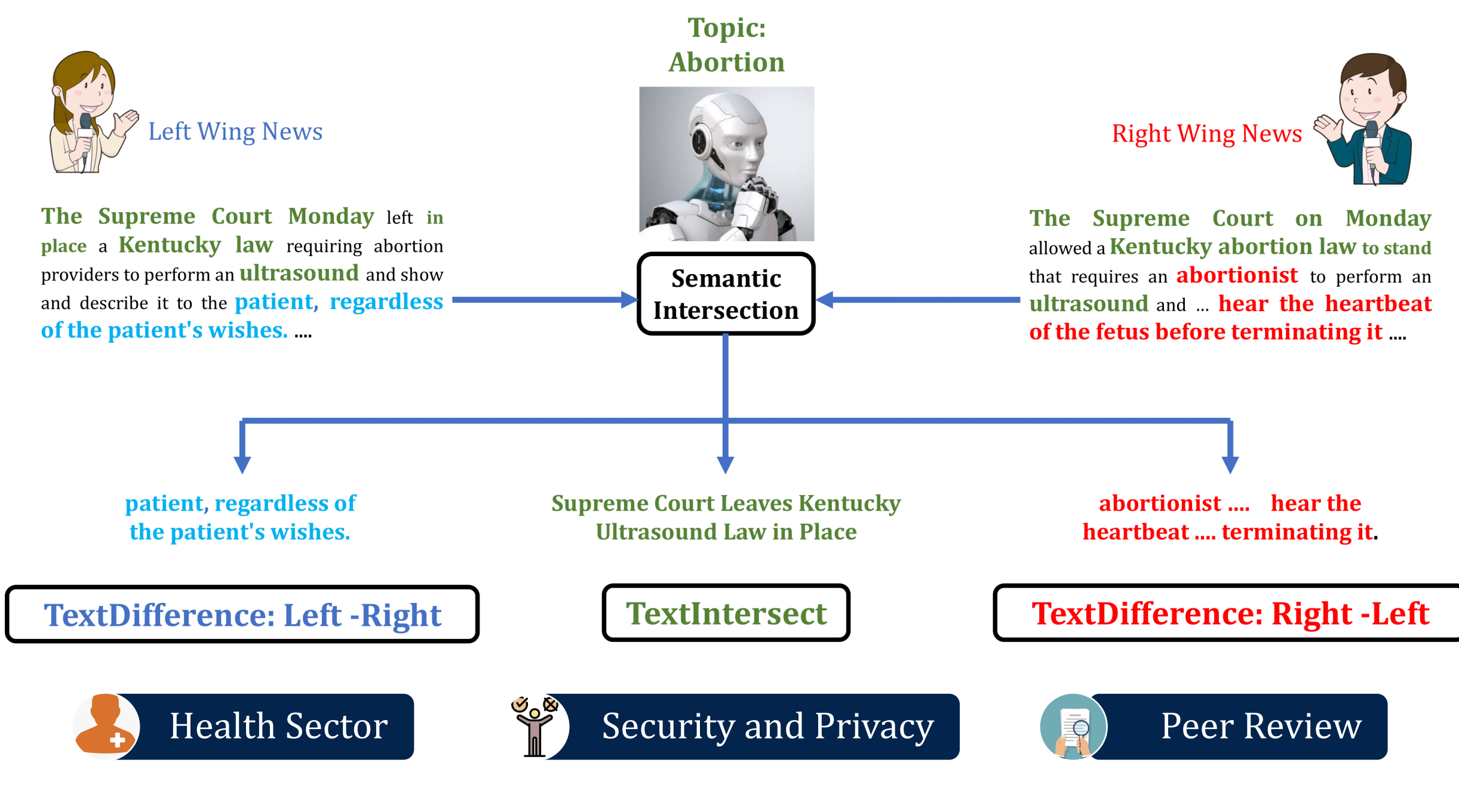


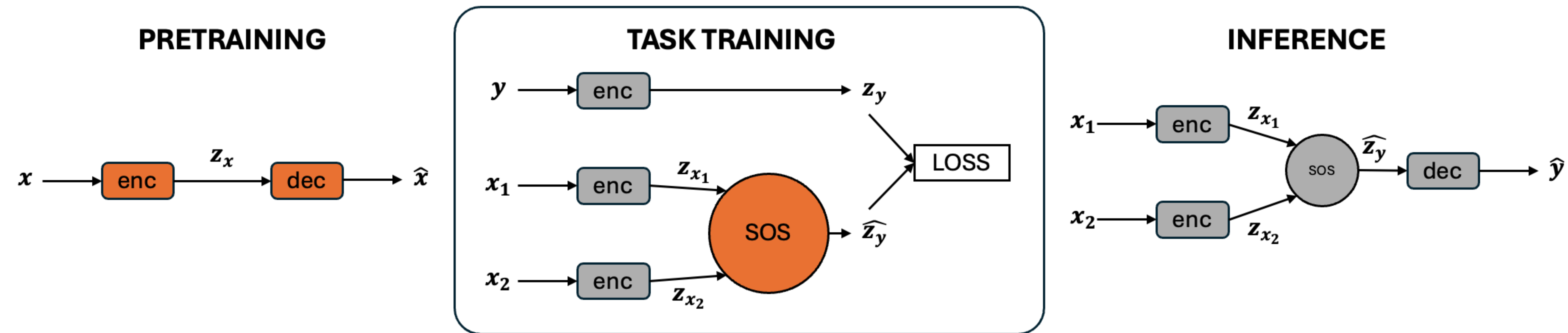
Semantic Overlap Summarization using Sentence Autoencoders

Naman Bansal, John Salvador, Shubhra Karmaker
Department of CSSE, Auburn University

Motivation



Proposed Framework



Semantic Intersection

Description

- Human-like summary
- Output must be present in both the input documents
- Can be extractive, abstractive or mixture of both
- Inspired by set-theory but still different from it
- Commutative

Challenges

- Virtually no data available in any domain
- Not clear how to apply un/semi supervised methods
- DL systems are data-hungry monsters
- Manual annotations not possible for humans

Benchmark Dataset



S_{prev}	S_o	S_{next}
He said he would continue to speak out on issues, and would do "some radio, some TV, but I don't anticipate full-time anchor duties."	Snow said he's received great satisfaction from talking to people about his illness.	Snow's cancer was diagnosed for the first time in February 2005.
S_1	S_2	
Snow stated he'll keep speaking out and doing some radio/TV, but won't take up full-time anchor duties; he found joy in discussing his illness.	Since his cancer diagnosis in February 2005, Snow found satisfaction talking about it.	

Autoencoder

Encoder	Decoder	R1	R2	R3
Roberta	1D	86.30	69.17	79.57
	3D	89.58	77.28	85.71
SBert	3D	76.70	54.39	70.22

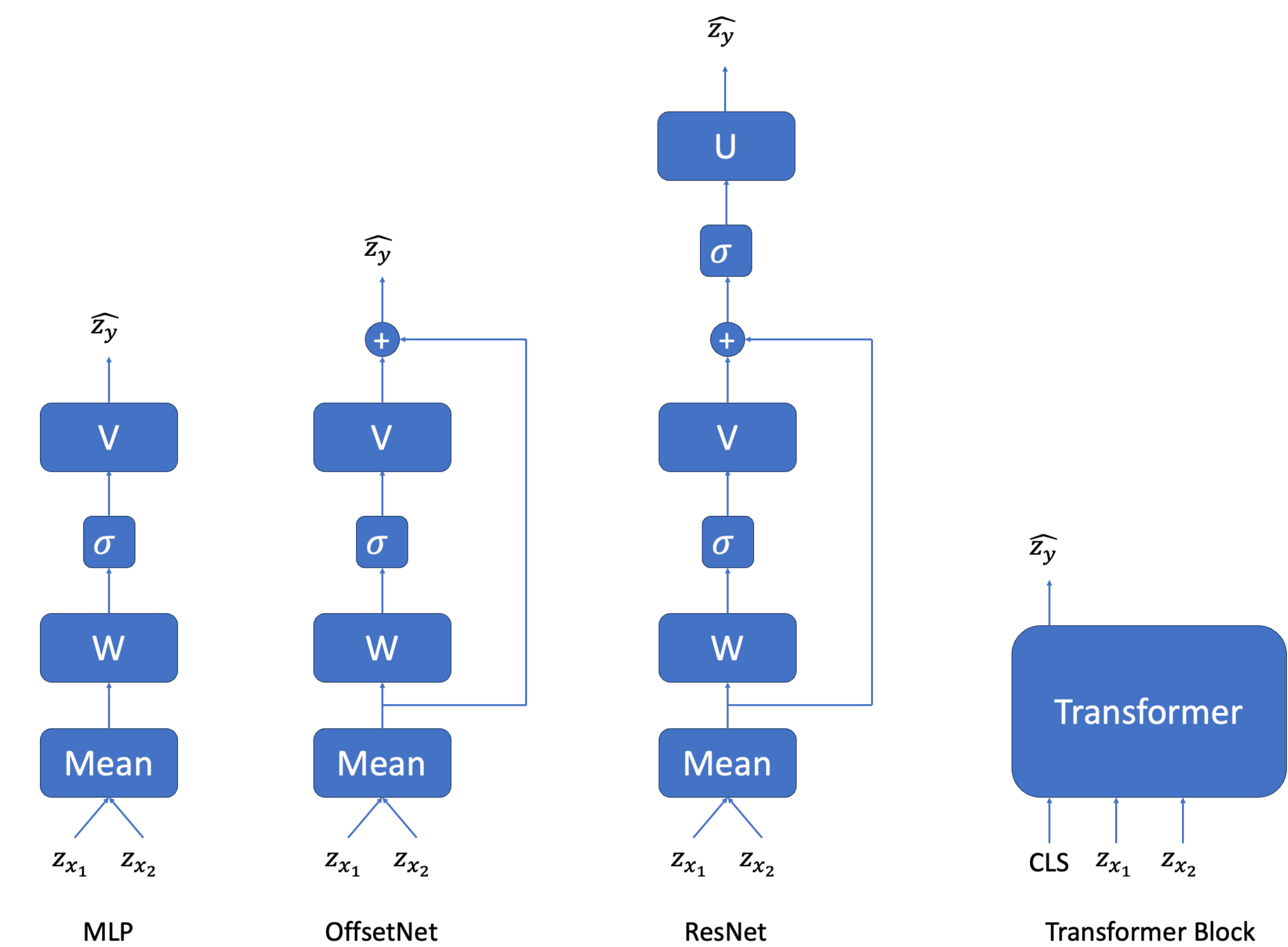
Table 1: Autoencoder Reconstruction Performance using ROUGE metric for autoencoders with 1-layer (1D) and 3-layer (3D) deep decoders.

Encoder	Pearson \uparrow	Spearman \uparrow	L1 \downarrow	L2 \downarrow
Roberta	0.54	0.56	0.46	0.30
SBert	0.86	0.85	0.14	0.03

Table 2: Correlation scores on STS benchmark for Encoders.

Trade-off between Reconstruction Performance and Semantic Capabilities

SOS



	R1	R2	R3
Roberta-1D	35.33	8.98	24.95
Roberta-3D	37.94	10.58	26.48
ChatGPT	47.19	25.92	39.50

Table 3: ROUGE scores of (upper) baselines models as compared to the proposed autoencoder approach. Autoencoder approach is $\sim 40\%$ as compared to the LLM based upper baseline in terms of ROUGE (R2). This is quite promising even though autoencoder model is approximately $\times 1000$ smaller than the LLM model.