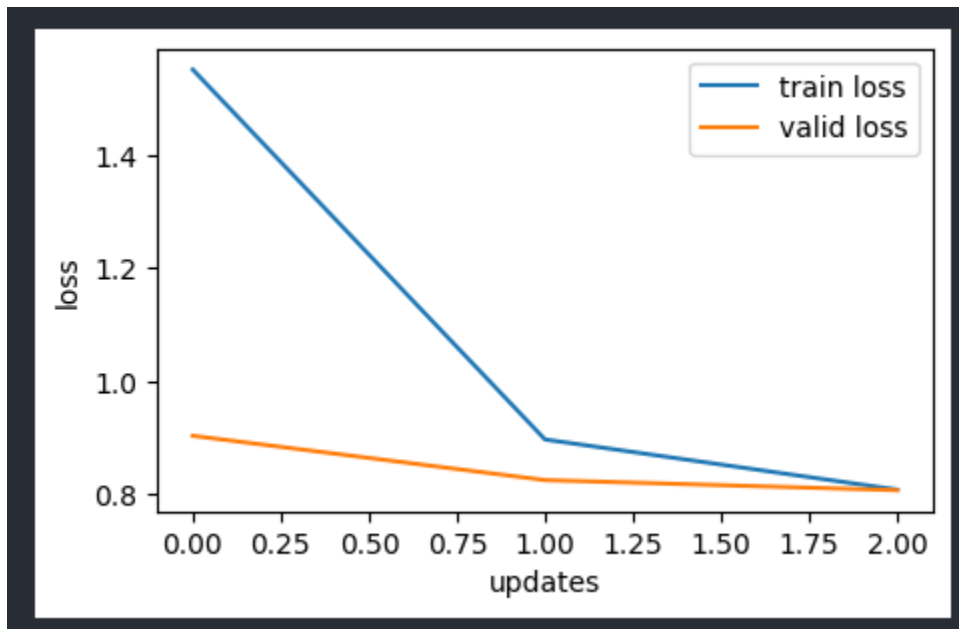


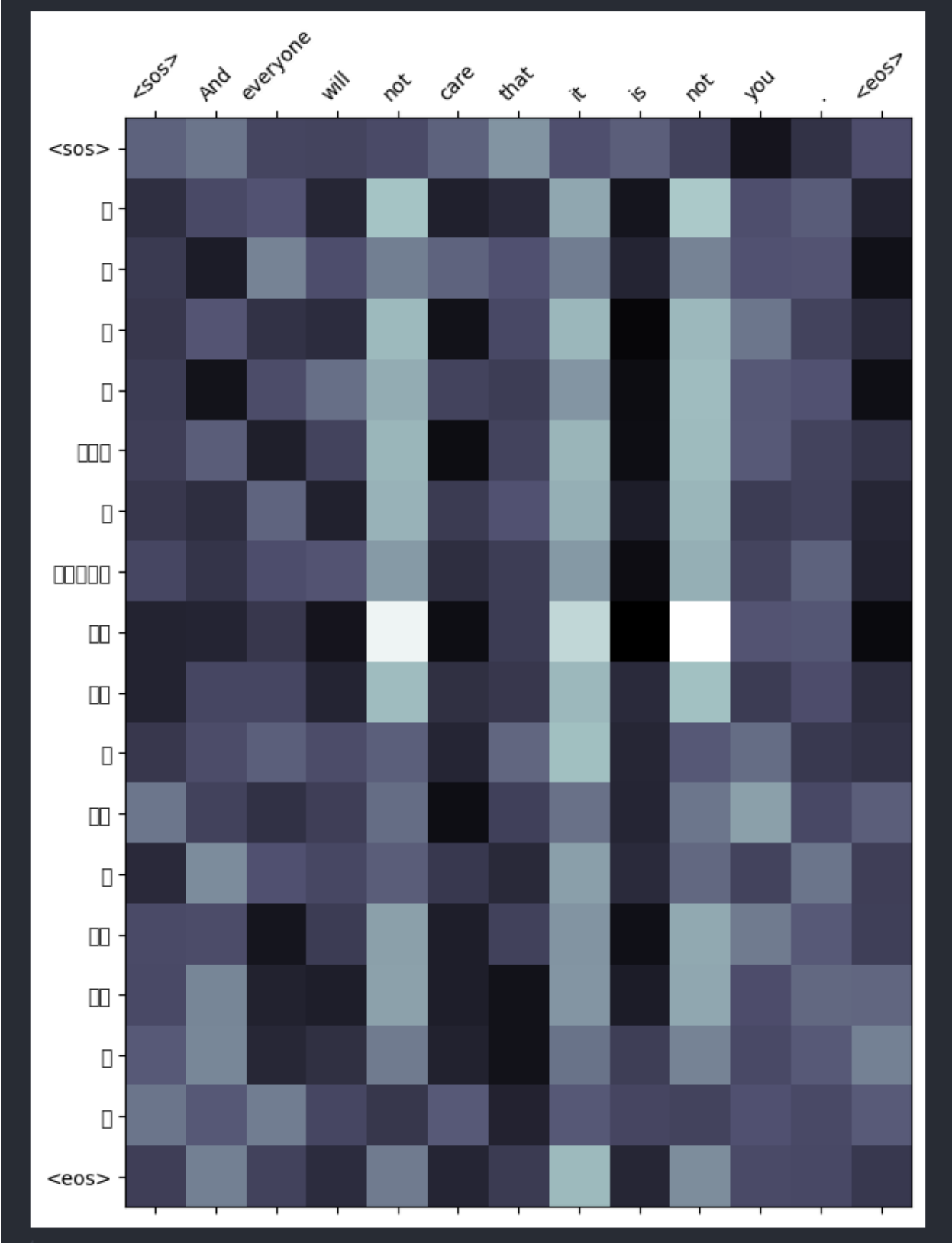
Attentions	Training Loss	Training Perplexity	Validation Loss	Validation Perplexity
General Attention	0.809	2.245	0.808	2.243
Multiplicative Attention	0.78	2.181	0.793	2.209
Additive Attention	0.758	2.134	0.795	2.216

Attentions	Training Time	Computational Efficiency
General Attention	32m 33s	Fast
Multiplicative Attention	37m 26s	Medium
Additive Attention	76m 1s	Slow

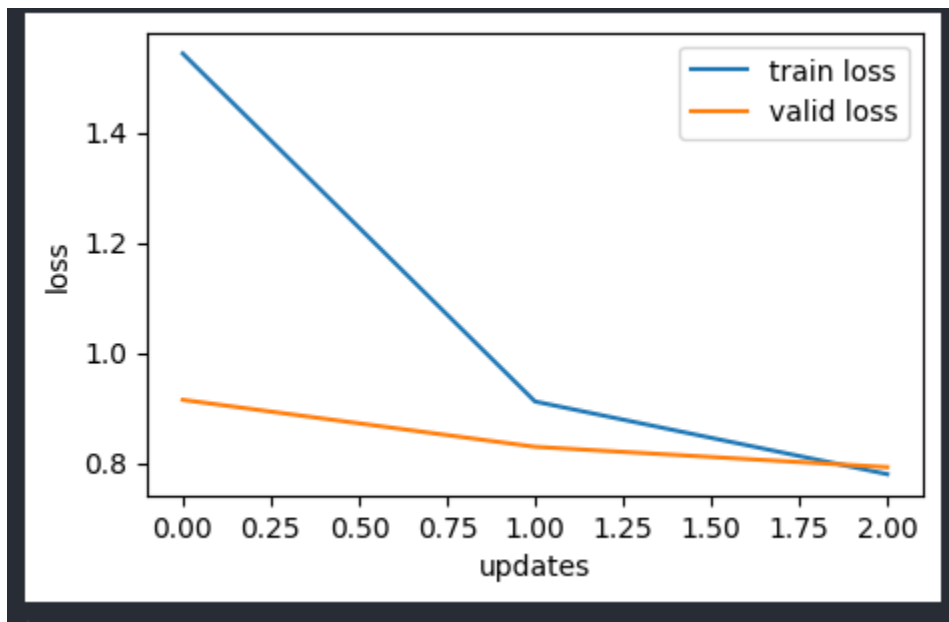
General Attention (Train & Valid Loss)



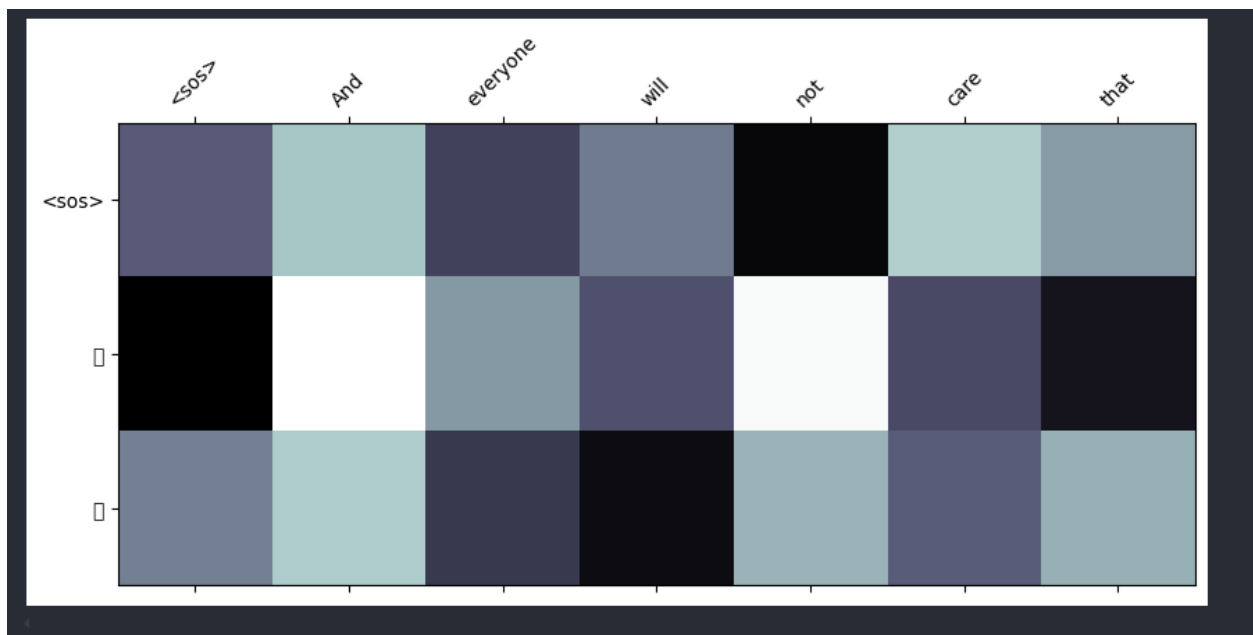
General Attention (Attention Map)



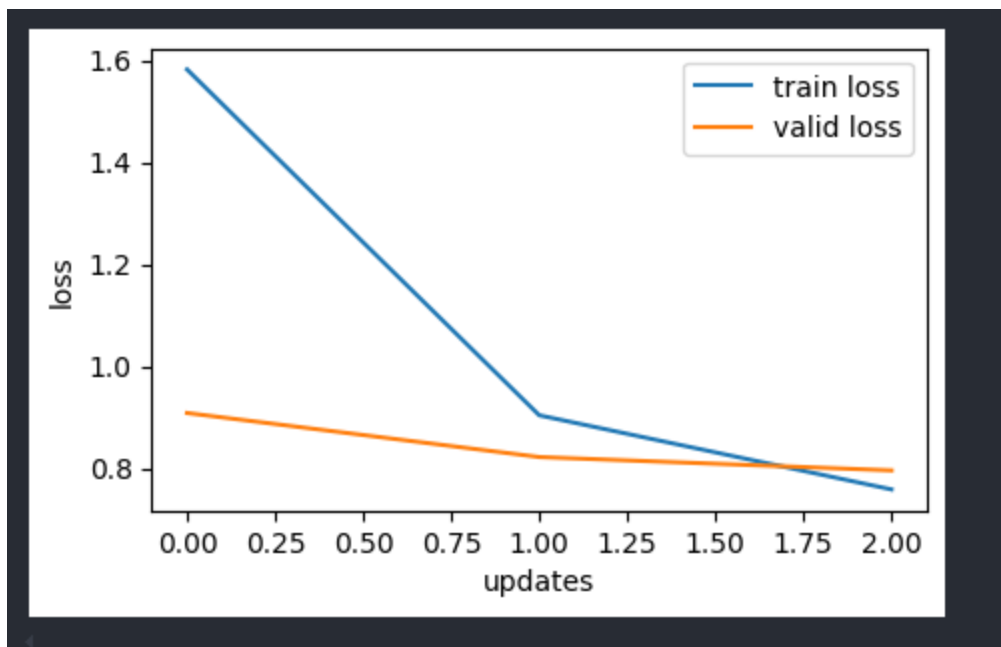
Multiplicative Attention (Train & Valid Loss)



Multiplicative Attention (Attention Map)



Additive Attention (Train & Valid Loss)



Additive Attention (Attention Map)

