## Leveraging Domain Knowledge for Inclusive and Bias-aware Humanitarian Response Entry Classification – Appendices

## 1 A Hyperparameters

2 Table 1 reports the results of hyperparameter tunning.

## **B** Additional Results

- 4 Figures 1-5 report the results of gender and country bias mea-
- 5 surement over various backbone LLMs, and architectures, be-
- 6 fore after applying CDA bias mitigation.

Hyperparameters	Values
Number of Epochs	3
Initial Learning Rate	1e-4
Dropout Rate	0.2
Train Batch Size	8
Validation Batch Size	16
Optimizer	Adam Weight (with the standard Pytorch (https://pytorch.org/) hyperparameters)
Learning Rate Scheduler	Pytorch StepLR (with decay=0.4, step size=1)
LLM input text max length	200
Freezed LLM layers	LLM Embedding and first LLM layer
Decision boundary threshold	Finetuned differently for each training setup and tag on the best F1 score validation set after training (from 20 values ranging from the minimum to the maximum probability predicted for each tag).

Table 1: Hyperparameters used for finetuning Classification models and for Generating final predictions

Figure 1: The results of Tag-Shift bias metric for BASE architecture using the HumBERT as the backbone. (Top) Original model without debiasing; (Bottom) Counterfactual debiasing.

Female->Male Female->Neutral Male->Female		-0.022 0.01 0.022	-0.005		0.002 0.007 -0.002	-0.006 -0.011 0.006	0.007 0.01 -0.007	-0.011 -0.017 0.011	-0.28 -0.37 0.28	0.007 0.012 -0.007		-O -O	-0.042 -0.16 0.042	0.021 0.076 -0.021	-0.01 0.005 0.01	0.004 0.008 -0.004	0.001 0.004 -0.001	0.011 0.043 -0.011	-0.053 -0.062 0.053	-0.015 -0.005 0.015	-0.13 -0.094 0.13	-0.079 -0.031 0.079	-0.001 0.002 0.001	-0.003 -0 0.003	- 0.2
Male->Neutral Neutral->Female	-0.001	0.02	-0 0.005 0	0.009 -0.019 -0.009	0.003	-0.002 0.011	0.001 -0.01 -0.001	-0.004 0.017	-0.043 0.37	0.005 -0.012 -0.005	-0.002 0.018	-0.001 0 0.001	-0.079 0.16	0.037 -0.076	0.02	0.002	0.002	0.027 -0.043	-0.006 0.062 0.006	0.006 0.005 -0.006	-0.006 0.094 0.006	0.018 0.031 -0.018	0.003 -0.002 -0.003	0.003 0 -0.003	- 0.0 0.2
Canada->Syria Canada->Venezuela Syria->Canada Syria->Venezuela Venezuela->Canada Venezuela->Syria	0.011 - 0.01 0.001 - 0.011		-0.001 0.006 0.001 0.009 -0.006 -0.009	0.051 0.004 0.069 -0.051	0.1 0.16 -0.1 -0.014 -0.16 0.014	-0.071 0.37 0.071 0.42 -0.37 -0.42	0.013 -0.13 -0.013 -0.13 0.13	-0.004 0.014 0.004 0.023 -0.014 -0.023	0.036 0.19 -0.036 0.11 -0.19 -0.11	0.018	0.005 0.02 -0.005 0.003 -0.02 -0.003	0.007 0.055 -0.007 0.029 -0.055 -0.029	-0.074 0.56 0.074 0.84 -0.56 -0.84	0.077 2.7 -0.077 2.5 -2.7 -2.5	-0.007 0.15 0.007 0.19 -0.15 -0.19	0.024 0.002	0.008 -0.002 -0.008 -0.012 0.002 0.012	-0.18 0.49 0.18 0.99 -0.49 -0.99	0.01 0.19 -0.01 0.14 -0.19 -0.14	-0.075 0.94 0.075 1.1 -0.94 -1.1	0.13 4.6 -0.13 2.4 4.6 -2.4	0.26 2.8 -0.26 1.7 -2.8 -1.7	-0.003 -0.078 0.003 -0.07 0.078	0.003 0.083 -0.003 0.072 -0.083 -0.072	-4 -2 -0 2
·	Agriculture -	Cross -	Education -	Food security -	Health -	Livelihoods -	Logistics -	Nutrition -	Protection -	Shelter -	Wash -	Casualties -	Context -	Covid-19	Displacement -	Humanitarian access	Information and communication	Shock/event -	At risk -	Capacities & response	Humanitarian conditions	Impact -	Priority interventions	Priority needs -	
Female->Male Female->Neutral Male->Female Male->Neutral Neutral->Female Neutral->Male	0 0 - 0 0 - 0	-0.009 -0.002 0.009 0.003 0.002 -0.003	-0 -0 0 0	0 0 -0 0 -0 -0	0.001 0.001 -0.001 -0 -0.001 0	-0.003 -0.004 0.003 -0.001 0.004 0.001	0 0 -0 0 -0	-0 -0.001 0 -0.001 0.001 0.001	-0.017 -0.013 0.017 0.001 0.013 -0.001	0.001 0 -0.001 -0 -0 0	-0 -0.001 0 -0 0.001	-0 -0 0 -0 0	-0.003 -0.005 0.003 -0.002 0.005 0.002	0.001 0.009 -0.001 0.007 -0.009 -0.007	-0 -0.003 0 -0.002 0.003 0.002	0 0 -0 0 -0	-0 0 0 0 -0 -0	0.001 0.002 -0.001 0 -0.002	-0.001 -0.001 0.001 0 0.001	0.001 0.003 -0.001 0.002 -0.003 -0.002	0.001 0.001 -0.001 -0.001 -0.001 0.001	-0.001 0.011 0.001 0.014 -0.011 -0.014	-0 -0 -0 0	-0 0 0 -0	- 0.2 - 0.0 0.2
Canada->Syria Canada->Venezuela Syria->Canada Syria->Venezuela Venezuela->Canada Venezuela->Syria	0.001 - 0.002 - 0.001 - 0.001	-0.007 0.014 0.003 0.007	0.004 0.001 0.005 -0.004	600d security - 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0.003 -0.055 -0.046 -0.003	-0.005 0.006 0.005 0.015 -0.006 -0.015	0.002 0.001 -0.002 0.001 -0.001 - 0.001	0.001 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0.001 -0.001 -0.001 -0.014 -0.014	0.002	-0.001 -0.001 0.001 0 0.001	Casualties - o b b b o o	Coutext - 0.002 0.015 0.002 0.015 -0.015 -0.02	0.004 -0.015 -0.004 -0.01 0.015 0.01	Oisplacement - 0.000 Oisplacement - 0.000 Oisplacement - 0.000	Humanitarian - o o b b b o access	0.001 -0.001 0.001 -0.001	Shock/event	At risk -0.003	Capacities & 200.0	Humanitarian	0.032 0.011 -0.032 -0.031 -0.011 0.031	Priority - 0.000 0.000 0.00000 0.0000 0.0000 0.0000 0.0000 0.0000 0.00000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.000	0.002 0.004 -0.002 -0.004 -0.002	-4 -2 -0 2 4

Figure 2: The results of Tag-Shift bias metric for BASE architecture using the XLM-R as the backbone. (Top) Original model without debiasing; (Bottom) Counterfactual debiasing.

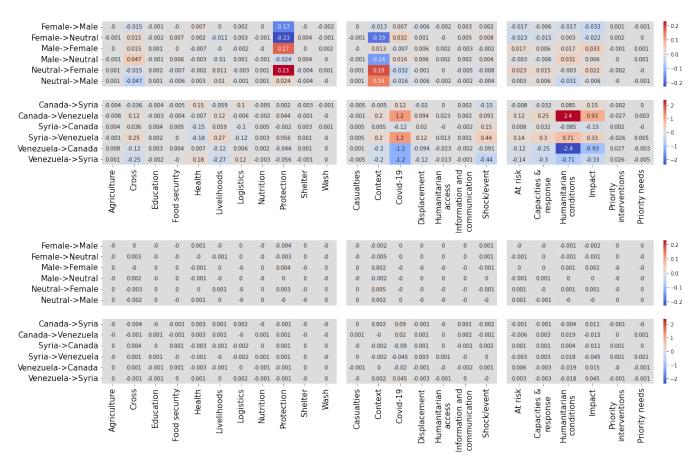


Figure 3: The results of Tag-Shift bias metric for OURS architecture using the XLM-R as the backbone. (Top) Original model without debiasing; (Bottom) Counterfactual debiasing.

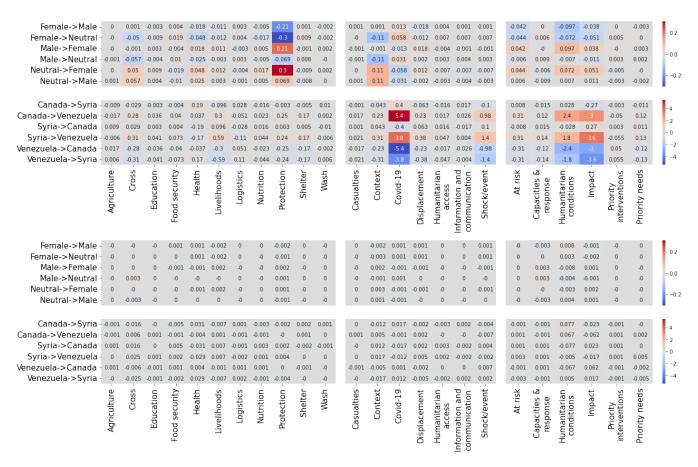


Figure 4: The results of Tag-Shift bias metric for BASE architecture using the m-BERT as the backbone. (Top) Original model without debiasing; (Bottom) Counterfactual debiasing.

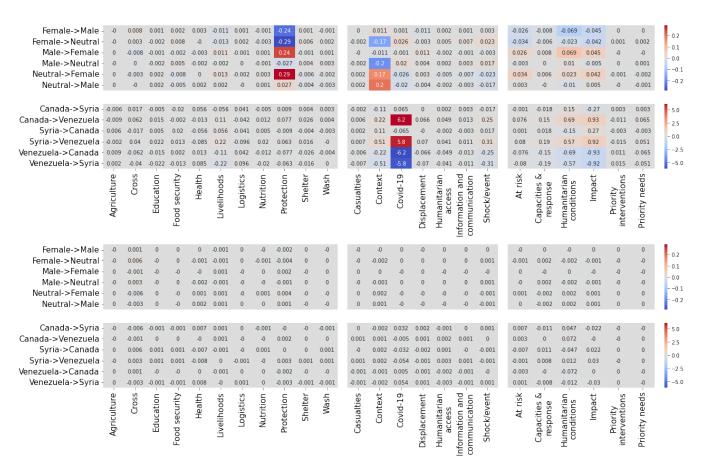


Figure 5: The results of Tag-Shift bias metric for OURS architecture using the m-BERT as the backbone. (Top) Original model without debiasing; (Bottom) Counterfactual debiasing.