

Probabilities of predicting gendered nouns given attributes - ClinicalLongformer

probability

1.0
0.8
0.6
0.4
0.2
0.0

gender
female
male

Anger

Dangerousness

Fear

Coercion

Segregation

stigma_category

Avoidance

Help

Pity

Blame

