ToS on Gemini (Highly Assertive Prompt) - Accuracy and Misclassification Penalty per Language 0.66 0.65 Accuracy 0.63 **Penalty Score** 0.61 0.6 0.5 0.4 **Score** 0.3 0.21 0.20 0.19 0.2 0.18 0.1 0.0

90

e_U

1