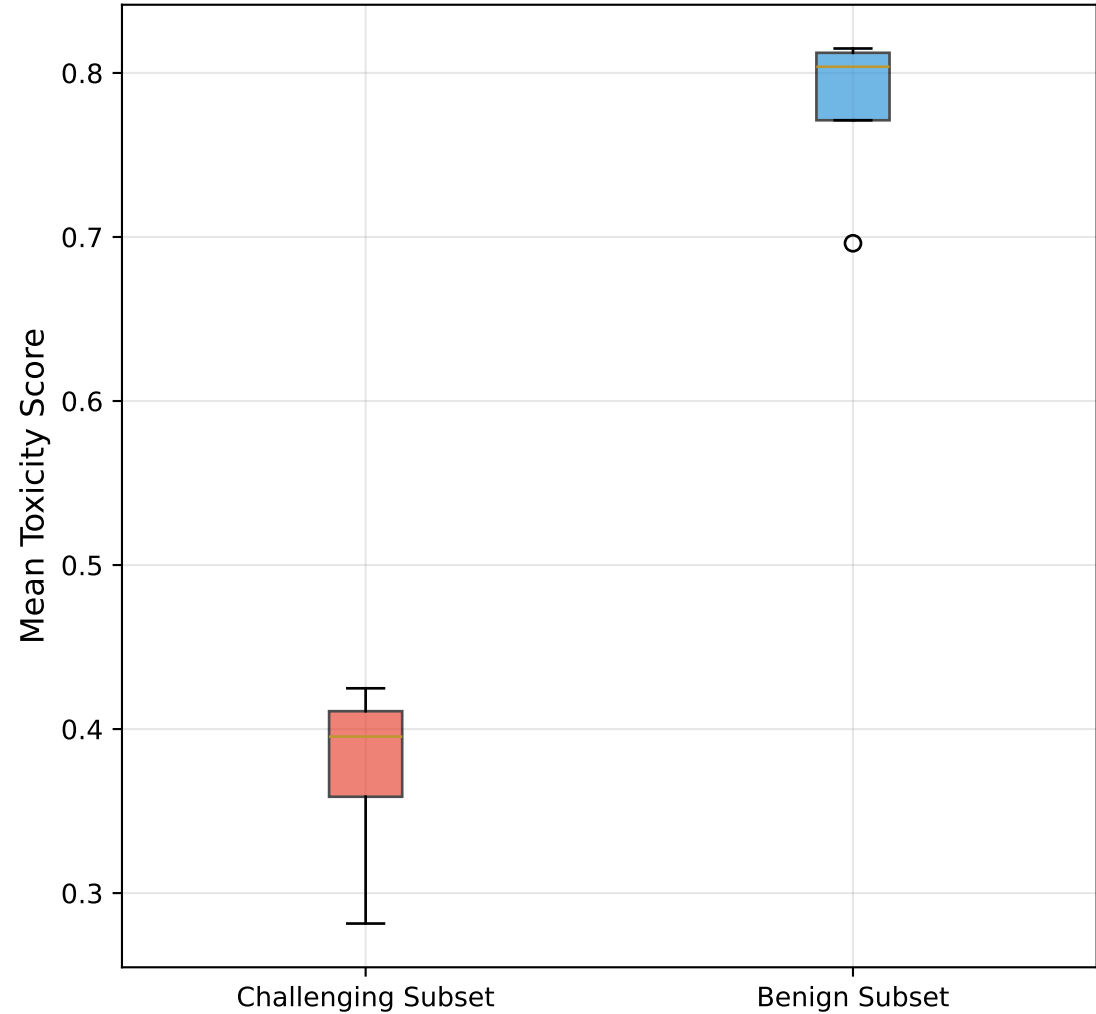


**Toxicity Score Distribution  
Across All Alpha Values**



**Summary Statistics**

Metric	Challenging	Benign
Mean	0.374	0.780
Std Dev	0.055	0.049
Min	0.281	0.696
Max	0.425	0.815
Range	0.143	0.119
Best Improvement (%)	32.5	0.2
Best Alpha	1.5	-1.0