|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Category** | **GPT Probability** | **GPT Reason** | **GCP Probability** | **GCP Reason** |
| **Profanity** |  |  |  |  |
| **Obscene** |  |  |  |  |
| **Insult** |  |  |  |  |
| **Hate** |  |  |  |  |
| **Toxic** |  |  |  |  |
| **Threat** |  |  |  |  |