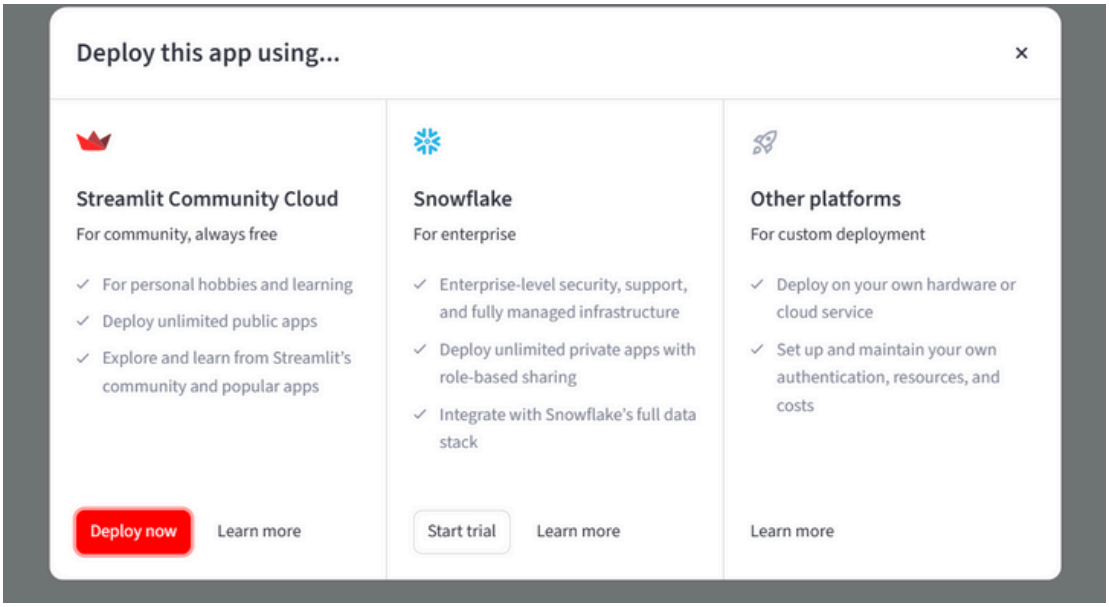
```
--- Running Predictions ---

------------------------------
Comment: You are so stupid and I will hurt you!
Result: ❌ Toxic labels detected:
  - TOXIC: 9.8/10 (Confidence: 0.9849)
  - OBSCENE: 5.2/10 (Confidence: 0.5217)
  - THREAT: 7.6/10 (Confidence: 0.7565)
  - INSULT: 6.9/10 (Confidence: 0.6884)

------------------------------
```

```
Comment: That was a nice effort, I appreciate your work.
Result: ✅ No toxic behavior detected.

------------------------------
```

```
Comment: This is the most obscene thing I have ever seen, you are an idiot.
Result: ❌ Toxic labels detected:
  - TOXIC: 9.9/10 (Confidence: 0.9908)
  - OBSCENE: 7.4/10 (Confidence: 0.7425)
  - INSULT: 9.4/10 (Confidence: 0.9419)
```



REFERENCES

1. Devlin et al., “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” 2019

.

2. Liu et al., “RoBERTa: Robustly Optimized BERT Approach,” 2020.

3. Kaggle, “Toxic Comment Classification Challenge,” 2018.

4. S. Lee, “Explainable AI for Toxic Comment Classification,” 2023.

5. T. Patel, “Multi-label Toxic Comment Detection Using Transformers,” 2023.

6. A. Singh, “Ensemble Learning for Content Moderation,” 2022.

7. C. Silva, “Offensive Language Detection in Multiple Languages,” 2023.

8. N. Gupta, “Contextual Understanding in Hate Speech Detection,” 2022.