Distribution of Per-prompt Success Rates

Legend:
- Phi-3-mini-4k (n=871)
- LLaMA-2-7b-hf (n=863)
- DeepSeek-Coder (n=524)

X-axis: Per-prompt Success@OGF ≥ 1
Y-axis: Number of prompts