Accuracy vs. Token Efficiency

5.1× fewer tokens

Legend:
- Accuracy (%)
- Tokens

DeepSeek-V3.2: 43.5%, 25,492

GPT-5.2 (Low): 41.5%, 5,003