Plan Generation Accuracy: Ours vs. GPT-4 Baseline

Legend:
- GPT-4 (baseline)
- Ours (BDI-LLM)

Y-axis: Accuracy (%)

X-axis categories: Blocksworld, Logistics, Depots

Blocksworld: GPT-4 35%
Logistics: GPT-4 5%, Ours 99.6% (568/570)
Depots: GPT-4 4%, Ours 99.4% (497/500)