

**Figure 3: Comprehensive Performance Heatmap**  
**12 Models × 4 Benchmark Categories**

