Performance Degradation Across Benchmark Tiers Premium Models 80 Mid-range Models Open-source Models 73% 70 65% 64% 60 58% 58% 56% 50 49% 48%

