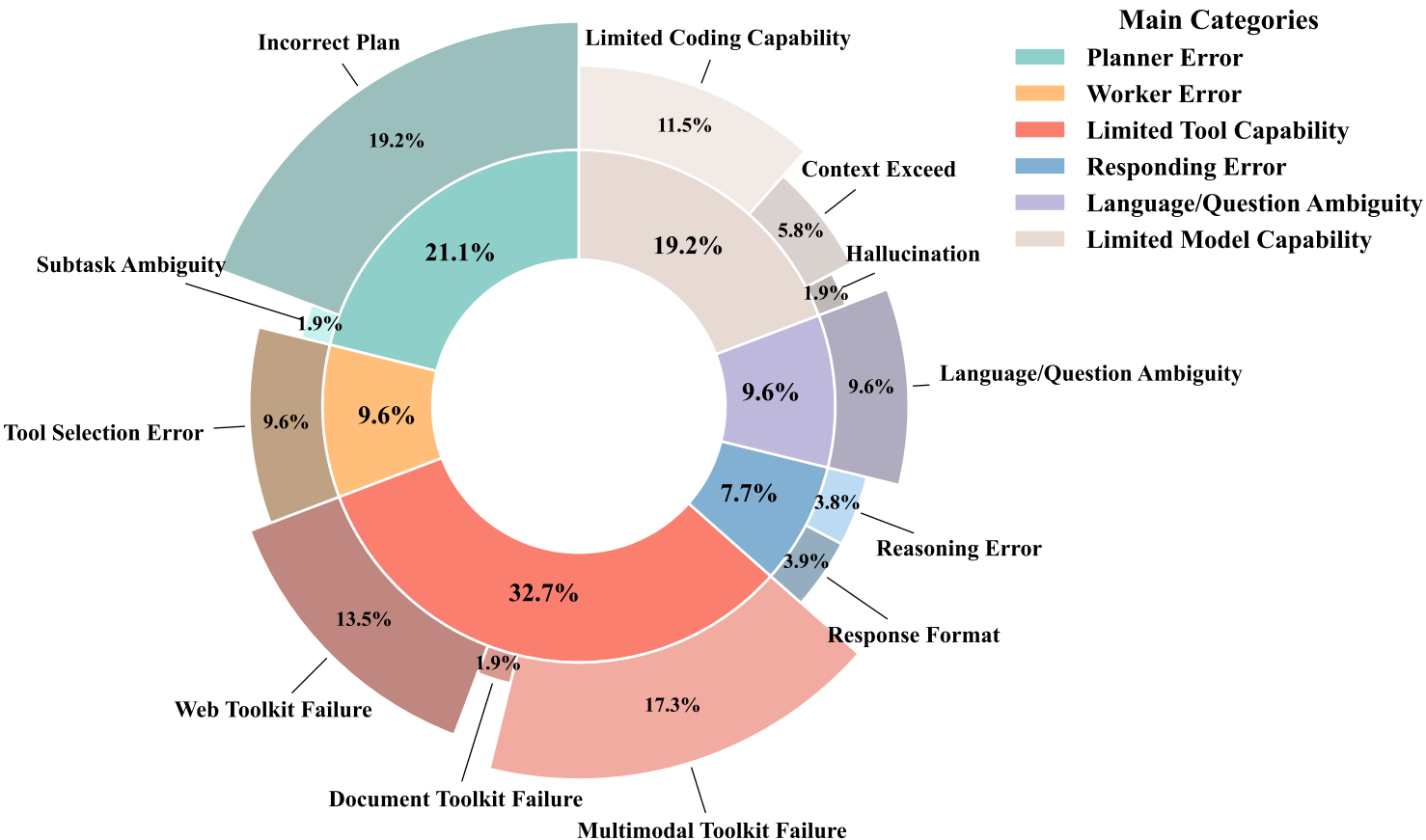Hierarchical Distribution of Failure Modes on the GAIA Benchmark