Data Quality vs Quantity: The Surprising Truth

Success Rate on Hard Tasks (%) vs Training Data Size (log scale)

- Magma-R1 (400 samples)
- Magma-R1 (2400 samples)
- InfiGUI-R1 (31k samples)
- GUI-R1-7B (3k samples)
- GUI-R1-3B (3k samples)

Legend:
- Magma-R1
- Magma-R1
- GUI-R1-3B
- GUI-R1-7B
- InfiGUI-R1
- Hypothetical Trend