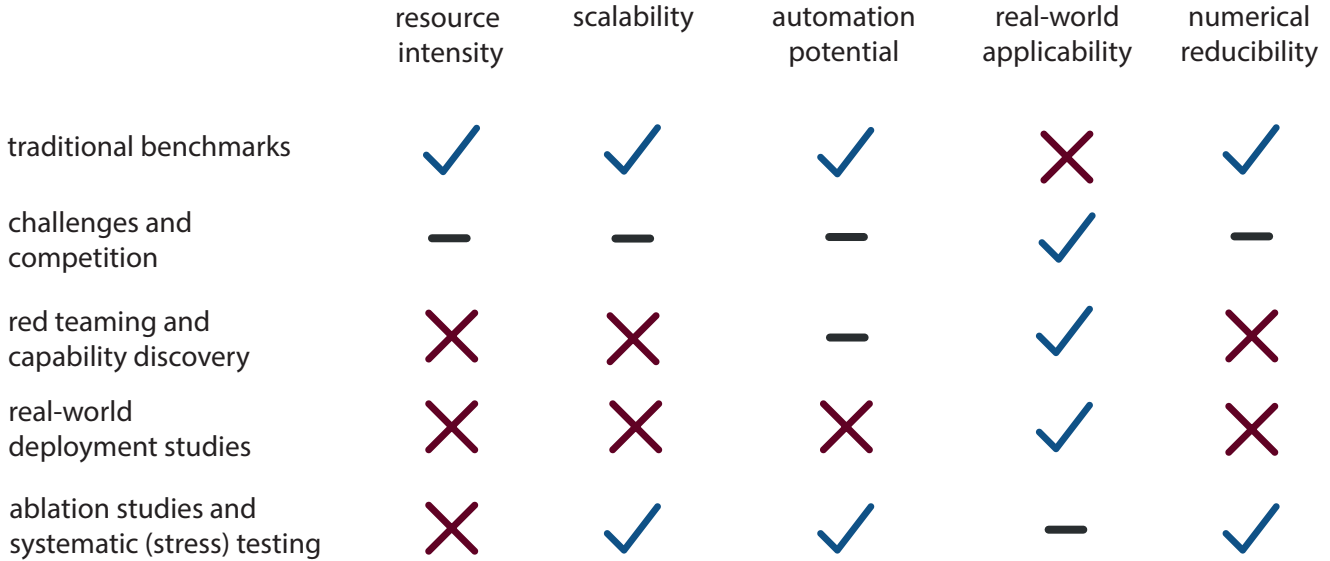
resource intensity
scalability
automation potential
real-world applicability
numerical reducibility
traditional benchmarks
challenges and competition
red teaming and capability discovery
real-world deployment studies
ablation studies and systematic (stress) testing