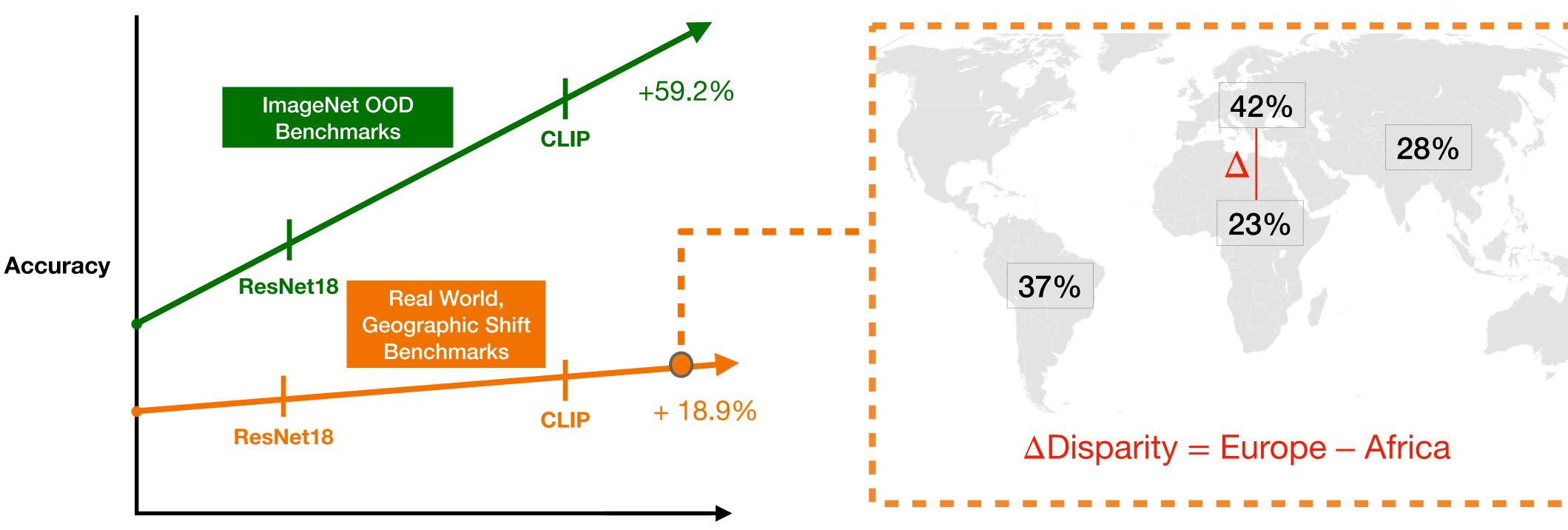
Benchmark progress is <u>2.5x</u> progress on

Geographic disparities have <u>tripled</u> between least performant models and today's best real-world, geographic shifts



Progress (Scaling, Multimodal Modeling)