

Benchmarks 101 Cover Page

Target Audience

Non-technical members of the public who want a brief but nuanced understanding of AI benchmarks, with a focus on journalists, educators, and policymakers and their staff. Particularly, those working on technology, science, health, defense, and consumer protection policy.

Expected Read Time

5-7 minutes

Brief Description

Goal. This policy primer provides a clear and accessible foundation for understanding how artificial intelligence is evaluated through benchmarks. Its goal is to equip educators, policymakers, and journalists with the knowledge to interpret AI performance claims critically and use this understanding to inform effective governance.

Content. This primer:

- Defines benchmarks and situates them within the AI lifecycle
- Explains why benchmarks matter and how they shape the research ecosystem
- Guides readers through interpreting a benchmark with questions to ask
- Illustrates three common limitations of benchmarking with examples

Resources and Pedagogical Approach. The material combines concise bullet points with tables, accessible explanations and real-world examples. It is designed to be colorblind-accessible and printer-friendly, and it supports grayscale printing for use in classroom settings. For readers who wish to “dive deeper,” the primer provides links to additional technical and policy resources.

Value for Policymakers. The primer equips policymakers to:

- Critically interpret AI performance claims in the media and hearings
- Ask targeted oversight questions about AI system evaluation
- Identify opportunities to promote benchmarks aligned with public interest goals

Why This Matters Now. In a fast-moving AI landscape, credible evaluation is essential to protect the public interest, but policymakers are caught between the extremes of rosy product announcements or catastrophic failures. This primer offers a practical, evidence-based framework for interpreting AI benchmarks in order to make better-informed decisions.