

Given: A Benchmark

ARC-Easy

In the atomic model that includes a nucleus, the positive charge is...

What is the role of decomposers in a food chain?

A meteor becomes a meteorite when...

What type of inheritance controls blood types in humans?

⋮

Step 1: Partition the benchmark into clusters.

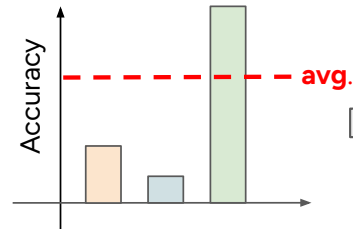
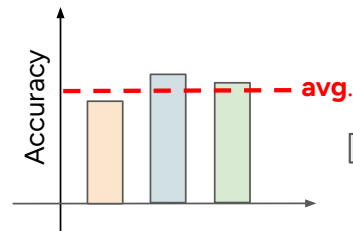
Cluster 1:
Chemical Properties

Cluster 2:
Fundamentals of
Astronomy

Cluster 3:
Biological
Concepts

⋮

Step 2: Gather the per-cluster accuracies.



Step 3: Measure *Harmony* via per-cluster distribution.



High Harmony: Aggregate score reflects overall competence.



Low Harmony: Aggregate score obscures strengths and weaknesses.