

SubsetEntry

September 4, 2026

SubsetEntry

Kind: Interface

Source: `atloria-monorepo/libs/agent-core/src/benchmark/swebench-subset.ts`

SWE-bench Verified — Curated Starter Subset

30 hand-picked tasks from the actual SWE-bench Verified dataset (500 tasks).
Balanced 10/10/10 across difficulty for clean comparison.
Selected for: well-defined test targets (FAIL_TO_PASS), pure Python,
and coverage across 9 repos.

Difficulty based on gold patch line count:
easy = < 20 lines
medium = 20-80 lines
hard = > 80 lines

IDs verified against princeton-nlp/SWE-bench_Verified (2026-03-25).

Properties

Property	Type
<code>id</code>	<code>string</code>
<code>repo</code>	<code>string</code>
<code>difficulty</code>	<code>`easy`</code>
<code>description</code>	<code>string</code>