AA-Omniscience v1.1 Review Report

Author: Benchmarks' Last Exam Lab 2026-11-02

AA-Omniscience Public v1.1 is the public subset of Artificial Analysis's knowledge benchmark, covering science and engineering, software, finance, health, humanities and social sciences, and law. Models answer short questions without external lookup and may abstain when they do not know.

Abstract

The stratified sample covers 100 questions from the 300-question public release, with 98 Accepted, 1 Doubtful and 1 Refused.

The confirmed defect is an accounting question whose cited source contains two objects matching the wording, while its reference accepts only one. One cardiac-development question retains a specific concern about source interpretation.

Outcome	Tasks	Meaning
Accepted	98	Within the scope of this review, we found no issue affecting evaluation validity and no specific concern awaiting verification.
Doubtful	1	We found a specific concern whose trigger, correct interpretation or effect on scoring still needs verification.
Refused	1	We confirmed at least one issue affecting evaluation validity.

Scope and version

We used AA-Omniscience Public v1.1, 1f9cbcec4720, stratified by domain and subtopic: 20 questions each in science/engineering/mathematics and software, and 15 each in finance, health, humanities/social sciences and law. Findings are current to November 6, 2026.

Issue distribution

Issue type	Refused	Doubtful	Total
C1 Incorrect answers or scoring logic	0	1	1
C2 Scoring requirements not stated in the task	1	0	1
Total	**1**	**1**	**2**

Representative cases

Two qualifying subtotals

C2 Scoring requirements not stated in the task

Question 122 **Refused**

The question asks for the new income-statement subtotal proposed alongside EBIT in the IASB's September 2018 project overview. The reference answer is Business profit.

Slide 10 lists Business profit; Profit before investing, financing and income tax; and EBIT. The first two both fit the description and are not short and long names for the same subtotal. Slide 12 assigns them values of 3,000, 3,500 and 3,900, reflecting differences including the treatment of integral associates and joint ventures.

The question provides no condition that rules out the second subtotal, while grading accepts only one.

Evidence 1

- IASB project overview, slides 10 and 12

The onset of cardiac septation

C1 Incorrect answers or scoring logic

Question 166 **Doubtful**

The question asks when the distal human embryonic cardiac outflow tract starts to separate and when distinct aortic and pulmonary trunks appear. The reference gives Carnegie stages 16 and 18; the model answers 14 and 16 and is graded incorrect.

The concern is the first part: a 2003 figure labels a separated specimen as stage 14, while 2012 and 2018 studies support actual channel separation at stage 16.

Evidence 3

- Webb et al., 2003
- Sizarov et al., 2012
- Yang et al., 2018

Partial credit for the metre answer

Question 220 **Accepted**

The question asks for the Persian metre named by code 4.5.11; the reference is khafif. The model supplies the fuller name Khafīf-e musaddas-e makhbūn-e maḥdhūf, plus metrical feet and a scansion, and receives partial credit.

The full name, feet and scansion are correct. However, the question asks only for the name, and the answering instruction requires just the answer without explanation. The additional feet and symbols go beyond what is needed to name the metre, providing a reasonable explanation for partial credit. We therefore withdraw the task-specific concern.

Evidence 2

- The Individuality of the Persian Khafīf Metre
- AA answering and grading rules

Evidence for cell-body shading

Question 5 **Accepted**

The question asks what provides shading during phototaxis in two dinoflagellate species without detectable eyespots. The reference is Cell body; the model answers Chloroplasts and is graded incorrect.

The textbook names both species and identifies the cell body as providing shading. The earlier plastid-shading concern came from a general discussion involving other dinoflagellates; it does not establish that “chloroplasts” should receive credit here. We therefore withdraw the concern.

Evidence 2

- Algae: Anatomy, Biochemistry, and Biotechnology, p. 97
- Jékely, 2009

Per-question results

Refused: 1 question

122

C2 Scoring requirements not stated in the task

Two distinct subtotals fit the question; it supplies no distinguishing condition, but the reference accepts only one.

Doubtful: 1 question

166

C1 Incorrect answers or scoring logic

Sources differ on the onset of septation; newer studies support the reference, so an error is not confirmed.

Evaluation impact and updates

Most questions in this sample showed no material defect.

Author-supplied sources, grading explanations and fixes will be used to update the relevant question-level findings.

Contact Us

benchmarks-last-exam.com

Benchmarks-Last-Exam@outlook.com

HarmonyOS Sans SC · Copyright 2021 Huawei Device Co., Ltd.