

CALIBRATION GUIDE

How to Calibrate Every Dimension of Your Agentic AI Scorecard

What a 2, a 5, and an 8 actually look like, so two people scoring the same project land on roughly the same number.

What's in this guide

- 01** Why the same project gets two different scores
- 02** Anchors for a 2, a 5, and an 8 on all six dimensions
- 03** Evidence-backed scores versus judgement calls
- 04** Two reviewers converging on a worked example
- 05** Why anchors should get harder to earn over time

An unanchored 1-to-10 scale is not a scoring method

Ask two people to rate the same project on "ease of implementation," and one comes back with a 3, the other with a 7.

Neither is being dishonest. Neither was told what a 5 is supposed to look like.

DIMENSION**A 2 LOOKS LIKE****AN 8 LOOKS LIKE****Strategic Fit**

Nobody above your level is tracking this problem

Named in this year's leadership priorities

Data Readiness

The data does not exist yet

Already accessible, structured, used for reporting

Ease of Implementation

Custom model work, multiple integrations

Off-the-shelf, one system, minimal retraining

DIMENSION	A 2 LOOKS LIKE	AN 8 LOOKS LIKE
Expected ROI	Value is theoretical, untested	A specific, defensible number
Speed to Value	Over six months to any result	A measurable result within four weeks
Autonomy Readiness	No logging, no approval step, no way to catch a bad action fast	Guardrails already built and tested for this exact level of autonomy

Score three dimensions against evidence, not confidence

Evidence-backed: Data Readiness, Speed to Value, Autonomy Readiness. Each traces to something concrete: an audit, a calendar count, a controls inventory.

Judgement calls: Strategic Fit, Ease of Implementation, Expected ROI. These deserve a second scorer and the confidence discount.

Two reviewers, before and after the rubric

AUTONOMY READINESS

NO RUBRIC

WITH RUBRIC

Reviewer A

6

3

Reviewer B

2

4

A four-point gap closed to one. Not identical, but close enough for a real conversation.



An unanchored 1-to-10 scale is not a scoring method. It is a survey question wearing a spreadsheet.

— Terence Kok

TERENCEKOK.COM · AGENTIC AI SERIES

DO THIS NEXT

09 / 10

1

Agree the anchors before you score anything

Five minutes as a group saves hours of arguing about what a 6 means.

2

Score judgement calls with more than one person

Strategic fit, ease of implementation, and ROI benefit most from a second opinion.

3

Revisit the rubric once a year

What earns an 8 today should get harder to earn as your programme matures.

TERENCEKOK.COM · AGENTIC AI SERIES

AGENTIC AI SCORING

10 / 10

**You do not need identical scores.
You need shared anchors.**

Score your candidate projects against this rubric, then get a ranked sequence, free at terencekok.com/resources.



Terence Kok
AI & Innovation Strategist

[SAVE](#) · [COMMENT](#) · [FOLLOW](#)