

FUNDAMENTALS OF AI-AUGMENTED ENGINEERING: A CONVERSATIONAL APPROACH

Taner Onsay, Ph.D. & Sebnem Onsay
Authors
October 2026 · Revision 1.0

© 2026 AI First Engineering



The engineering workflow you will use throughout this book

CHAPTER - 7

PART 4 · ≈ 13 MIN

Fits ≠ Validated

Every number so far says “trust me.” Three attacks on the same 126 points — a better R^2 , a different data window, a different ruler — and one question: how do we know when a beautiful fit deserves our trust?

4

Problem

›

Model

›

Compute

›

Analyze

›

Validate

›

Improve



What more could we possibly want?

0.99982

R^2 · quadratic

126 measured points

3.9 %

$a_{y,fit}$ vs $-g$

close to gravity

2 mm

apex height

fit vs camera

6.81 m/s

release speed

physically plausible

VOTE NOW

Is the quadratic projectile model **validated**? Yes · no · it depends. Keep your answer — we come back to it.

THREE ATTACKS

- 1 a higher R^2
- 2 a different data window
- 3 a different ruler

The evidence looks overwhelming. Let us try to break it — the way a reviewer would.

Problem



Model



Compute



Analyze



Validate



Improve



More terms, higher R^2 . Which model is best?

Polynomial fitted to $y(t)$	Coefficients	R^2	RMS residual
degree 2 — quadratic	3	0.99982	6.1 mm
degree 3	4	0.99982	6.1 mm
degree 4	5	0.99991	4.3 mm
degree 6	7	0.99992	3.9 mm

SAME 126 POINTS, SAME LEAST SQUARES

Each extra term gives the curve more freedom to follow the measurements.

BEFORE YOU ANSWER

What physical meaning would we give the τ^3 , τ^4 , τ^5 , τ^6 terms in a basketball model?

Notice degree 3: one more term, **no improvement**. Physics predicted that.

More flexible $\neq$ more physical. More parameters $\neq$ more knowledge. Higher $R^2 \neq$ more valid.

Problem



Model



Compute



Analyze

**Validate**

Improve



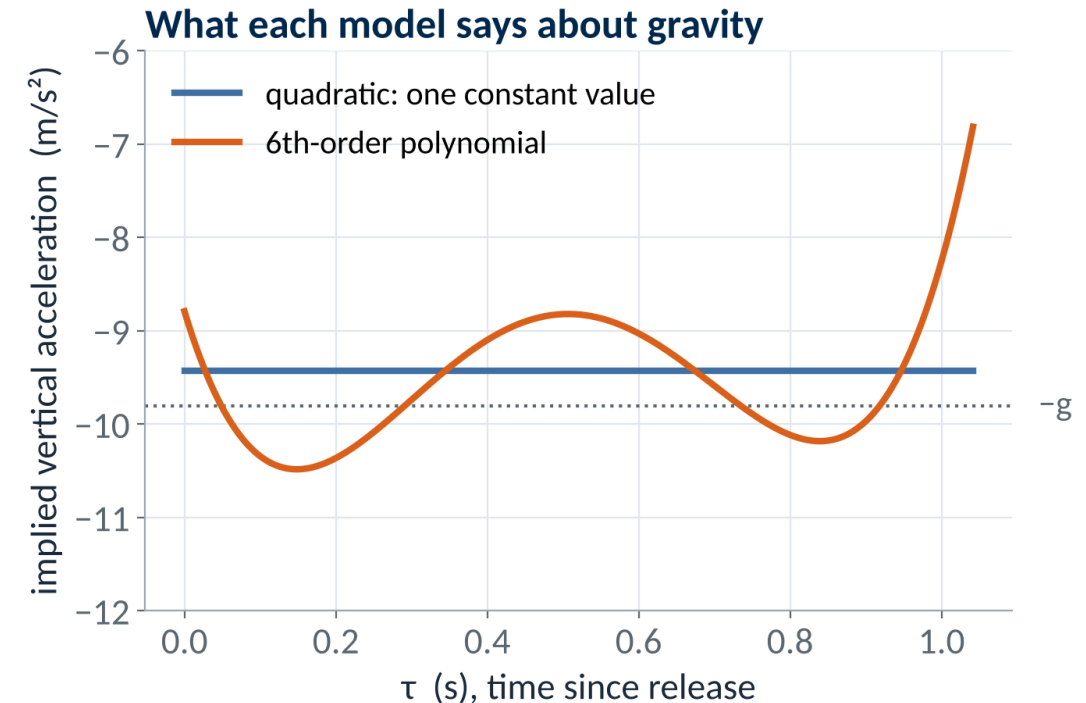
ACCEPT, MODIFY or REJECT?

AI ANALYST

“I tested quadratic, fourth-order and sixth-order polynomials. The sixth-order model has the highest R^2 , so it is the most accurate model of the basketball's motion.”

EVIDENCE TO WEIGH

- Read off the 6th-order curve, “gravity” drifts from -6.8 to -10.5 m/s^2 during one shot.
- Its RMS gain is about 2 mm — roughly half a video pixel (≈ 4 mm).
- R^2 measures agreement with THESE points — nothing more.



AI can optimize the metric we ask for. The engineer decides whether that metric answers the engineering question.

Problem



Model



Compute



Analyze

**Validate**

Improve



Same shot, same model. Did gravity change during the shot?

-9.43

all 126 points

$R^2 = 0.99982$

-9.90

first half only (63)

$R^2 = 0.99994$

-9.79

second half only (63)

$R^2 = 0.99974$

$a_{y,fit}$ in m/s^2 . Every one of these fits has R^2 above 0.9997.

OBVIOUSLY NOT

So why did the ESTIMATE change when we changed which measurements we used? List possible reasons — do not pick one yet.

WHAT A FITTED PARAMETER IS

An estimate derived from data — not a physical constant just because it carries units of m/s^2 .

A coefficient with units is still an estimate. It inherits everything — good and bad — from the data it was fitted to.

Problem



Model



Compute



Analyze



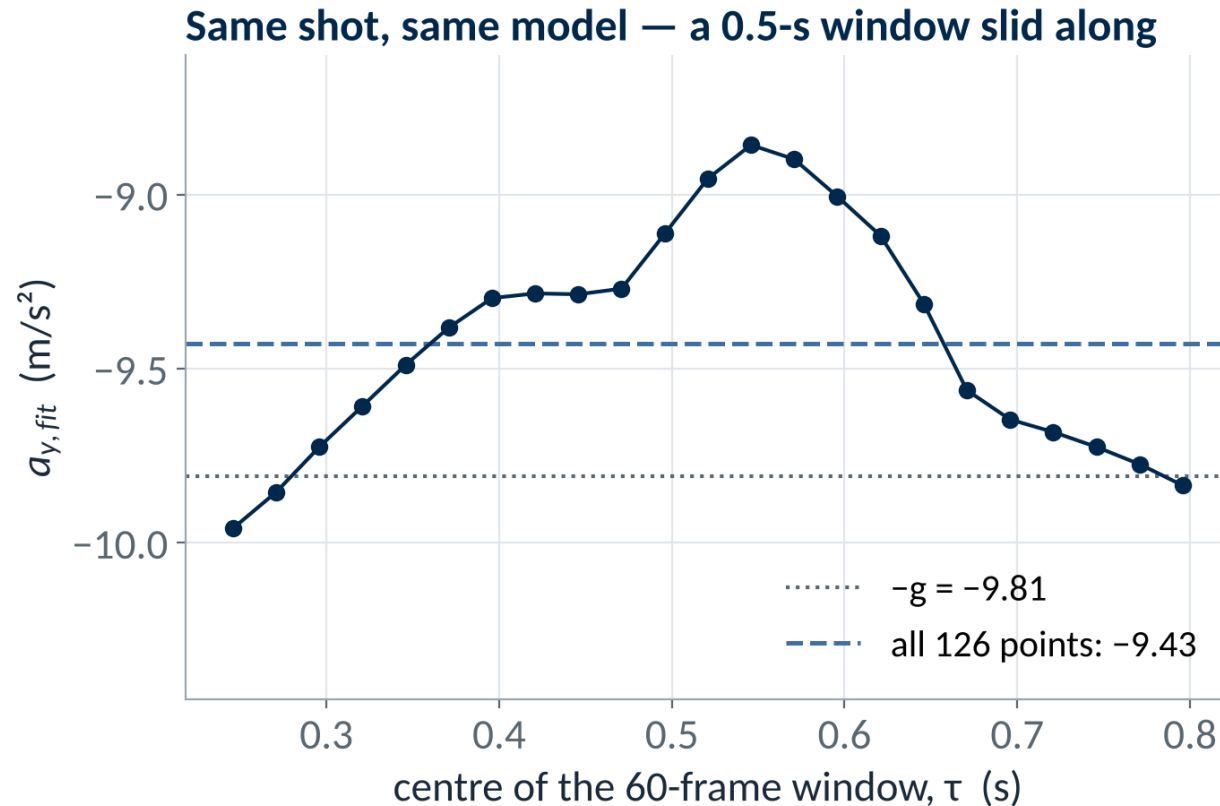
Validate



Improve



If the parameter is real, does it survive reasonable changes?



Reasonable analysis change	$a_{y,fit}$ (m/s^2)
all 126 points	-9.43
drop last 3 (ball at rim / net)	-9.42
drop last 8 frames	-9.41
middle 80 % of the flight	-9.31
any 0.5-s window (graph)	$-8.86 \dots -9.96$

READ IT AS

Small trims: stable. Short windows: not. Something deserves investigation.

Robust conclusions should not collapse under small, reasonable changes to the analysis.

Problem



Model



Compute



Analyze

**Validate**

Improve



What if the ruler were 4 % off?

Every measured length × ...	R^2	$a_{y,fit}$ (m/s ²)	apex (m)
1.00 (as measured)	0.999818	-9.43	4.00
1.04 (every length 4 % larger)	0.999818	-9.80	4.17
0.96 (every length 4 % smaller)	0.999818	-9.05	3.84

What-if: the real data rescaled on purpose. Not a claim that the calibration was wrong.

R^2 DOES NOT KNOW WHETHER YOUR RULER IS CORRECT.

REGRESSION CANNOT SEE

- the calibration length
- camera perspective
- axis orientation
- where the ball centre was clicked
- the time reference
- what the setup really measured

A model can fit biased data beautifully. Test data must be validated too.

Problem



Model



Compute



Analyze

**Validate**

Improve



When model and measurement disagree, what exactly is on trial?

AI ANALYST

“The fitted acceleration differs from gravity by 3.9 %, so the ideal-projectile model has been disproven.” **Accept · modify · reject?**

QUESTION THE MODEL

Does the mathematical / physical model adequately represent the system?

assumptions · omitted physics · parameters

QUESTION THE TEST

Are the measurements accurate enough to evaluate the model?

resolution · calibration · repeatability

QUESTION THE COMPARISON

Are model and test compared consistently?

coordinates · units · time origin · same quantity?

Disagreement exists. Its cause has not been established. Three suspects, not one.

Problem



Model



Compute



Analyze



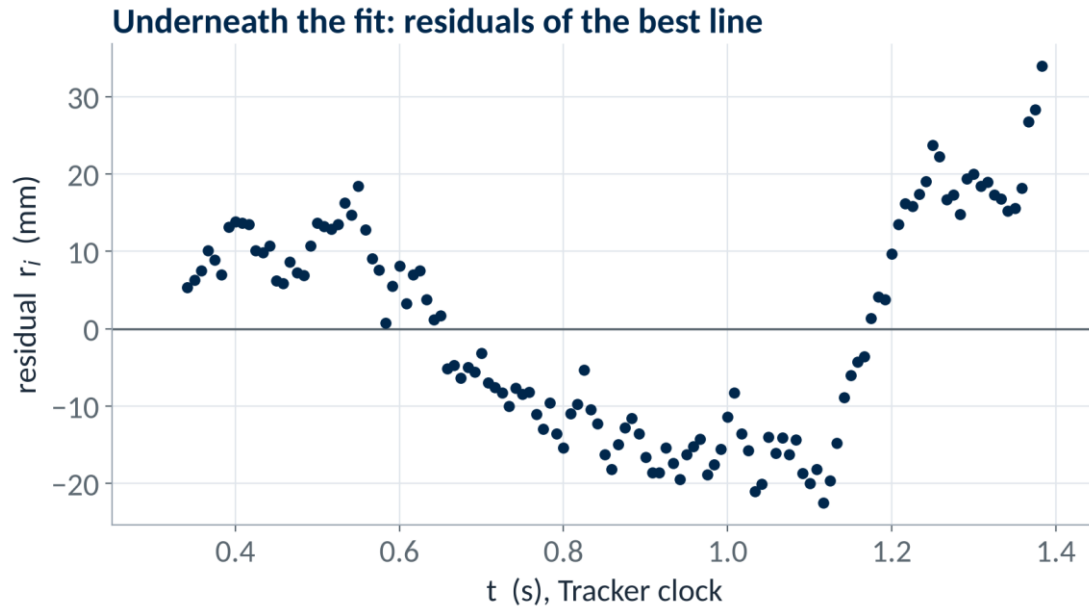
Validate



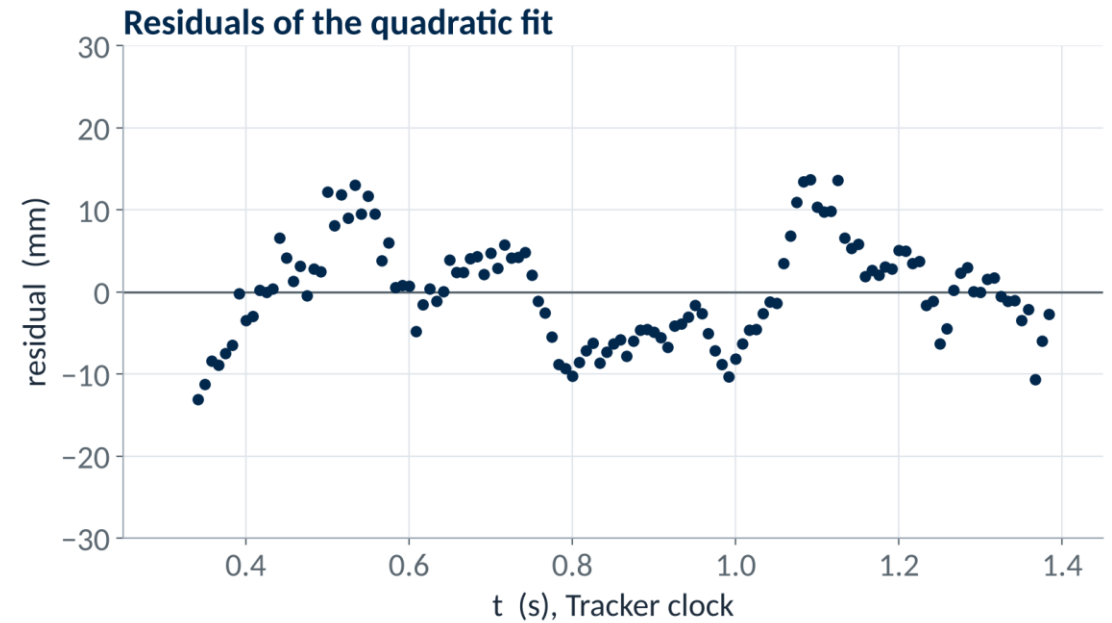
Improve



Random — or a pattern the model has not explained?



$x(t)$, straight line (Part 2): RMS 13.9 mm



$y(t)$, quadratic (Part 3): RMS 6.1 mm

Random small residuals say "measurement noise." Structured small residuals say "something systematic is left." **Same size — different message.**

Part 4 recognizes the question. Part 5 investigates the answers.

Problem



Model



Compute



Analyze



Validate



Improve



Small residuals are not automatically unimportant residuals

Engineering purpose	Accuracy it needs	Quadratic adequate?
show the overall trajectory	centimetres	yes
estimate release speed and angle	a few %	probably
predict apex height	centimetres	yes — within 2 mm here
predict where the ball crosses the hoop plane	a few cm (opening ≈ 0.11 m each side)	depends — on evidence not yet in hand
measure g	well under 1 %	not yet — 3.9 % unexplained
tell aerodynamic drag from measurement error	mm, with known test errors	not with what we have checked so far

A model is not "valid" or "invalid" in the abstract — it is adequate or inadequate for a purpose, a range and a required accuracy.

What decision are we trying to make? That question sets how much disagreement is acceptable.





Six questions before you trust a fitted model — any model, any data

IMG_1958 SO FAR

1	FIT	Does the model agree numerically with the measurements?	yes · R^2 0.99982
2	RESIDUALS	Is systematic structure left unexplained?	yes — structured
3	PHYSICS	Are the coefficients and behaviour physically plausible?	mostly — 3.9 % from g
4	STABILITY	Do reasonable analysis changes give similar conclusions?	partly — window-dependent
5	TEST QUALITY	Are the measurements accurate enough for the claim?	not yet examined
6	PURPOSE	Is the model accurate enough for the decision?	depends on the decision

Fit is item 1 of 6. Parts 2 and 3 answered it. Part 5 works on items 2 to 5.

Problem



Model



Compute



Analyze



Validate



Improve

**We have a model. We have test data.
They do not agree perfectly. What should we question?**

MODEL?

omitted physics · assumptions ·
parameters

DATA / TEST?

how the 126 points were produced

COMPARISON?

are we comparing like with like?

Part 5 — in lab: the IMG_1958 experiment goes on trial. Gravity, drag and spin as competing hypotheses; the Tracker screenshots as evidence; and an engineering decision.

Problem



Model



Compute



Analyze



Validate



Improve