

FUNDAMENTALS OF AI-AUGMENTED ENGINEERING

CHAPTER 3/LAB-3: Engineering Lab (3 Hours)



When Is the Water Ready?

Accuracy, resolution, convergence & model validation — how do we know a prediction is accurate ENOUGH?

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

© 2026 AI First Engineering

Problem



Model



Compute



Analyze



Validate



Improve

Today a mug of warm water becomes a defended accuracy claim



When Is the Water Ready?

What is this lab about? Predict when warm water in a mug reaches a target temperature — then defend, with evidence, how accurate that prediction is:

Measure → Model → Predict → Error → Refine → Validate → Decide
(DO › OBSERVE › RECORD › TALK TO AI › EVALUATE › DECIDE ...)

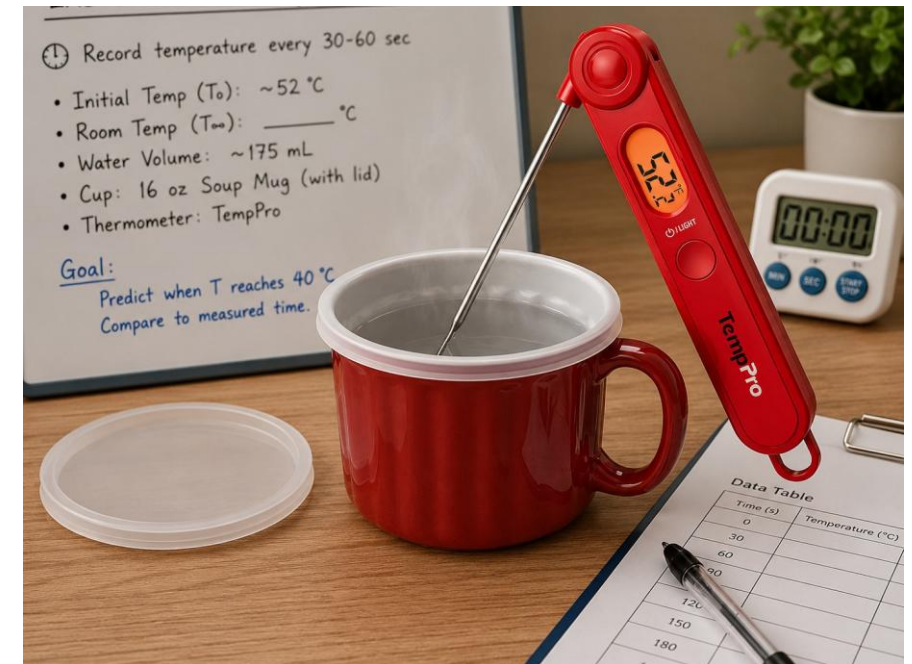
LEARNING OBJECTIVES

By the end of this lab, you will be able to:

- Collect a real cooling dataset with a standardized procedure
- Use a reduced-order model — and state what it leaves out
- Quantify error — absolute and percent, against a stated reference
- Run a refinement study and read convergence evidence honestly
- Judge PASS / FAIL against a stated engineering requirement
- Report with justified precision — digits the evidence supports

TOOLS YOU WILL USE

- Mug + warm water + TempPro thermometer — the physical system
- Excel + Microsoft Forms — the eight-station class dataset
- Python / Jupyter — model, prediction, refinement study
- AI assistant — engineering analyst & critic, never the judge
- Your engineering judgment — decide what is accurate enough





What you will do (step-by-step):

- 1 Readiness quiz — then verify Excel, Python, AI workspace
- 2 Meet the apparatus · safety briefing · standardize the setup
- 3 Commit predictions P1–P4 — before any data exists
- 4 Instrument check — eight “identical” thermometers, one water bath
- 5 Collect the cooling curve — lid on, a reading every 60 s · no dead time
- 6 Model the cooling · estimate k · predict the 40 °C crossing time
- 7 Watch the real crossing happen — then quantify the error
- 8 Refine the resolution · read the convergence evidence · apply the requirement
- 9 Class data checkpoint · BREAK THE AI — attack an accuracy claim
- 10 Accuracy budget · record → report handoff · human review, then sign

Would you confidently put your name on your accuracy claim? You will — literally — in the report's approval section.





MEASURE › PREDICT › DEFEND · CONVERGED ≠ CORRECT

In Chapter 3 the AI role advances — from coding companion to emerging engineering analyst — while you remain the lead engineer: the model, the evidence, and the accuracy claim stay yours.

THE RECURRING LOOP — INSIDE PROBLEM › MODEL › COMPUTE › ANALYZE › VALIDATE › IMPROVE



... and back to DO — the conversation evolves as the experiment evolves.



You, not AI: make the measurements, apply the engineering requirement, judge the convergence evidence, verdict every AI analysis — ACCEPT · MODIFY · REJECT — sign the claim.



AI, for you: model explanation, error arithmetic checks, plotting help, convergence-pattern description, discrepancy hypotheses — on the evidence you provide.

AI assists. Humans observe, verify, decide, and approve.

Never ask AI to invent what you did. Tell AI what you actually did.



AI as emerging engineering analyst — who does what



HUMAN TEAM — LEAD ENGINEER

- Own the experiment, the model, and the accuracy claim
- Perform every temperature measurement — honestly
- Commit predictions before the analysis
- Apply the engineering requirement — and defend the verdict
- Verdict every AI analysis: ACCEPT · MODIFY · REJECT
- Weigh numerical vs. measurement vs. model error
- Approve and sign the final engineering statement



AI — EMERGING ENGINEERING ANALYST

- Explain the reduced-order model and its assumptions
- Help organize data, plots, and the record
- Help implement the estimate, prediction, refinement
- Check error computations — E_{abs} , $E_{\%}$
- Describe convergence patterns in YOUR table
- Challenge unsupported digits · suggest discrepancy sources
- Draft report text from your evidence — never the verdict

AI analyzes the evidence. The judgment — and the responsibility — remain yours.



Create your AI workspace — the engineering record

ONE WORKSPACE · “LAB-3 — WHEN IS THE WATER READY?”

- Lab instructions + report template
- Temperature–time table · every reading
- Assumptions · predictions P1–P4 (committed)
- Estimated k · predicted vs. measured bracket
- Refinement table · convergence evidence · PASS / FAIL
- BREAK THE AI findings · your verdicts
- Forms submission · class comparison · decisions

The record documents what happened.

The report communicates what matters.

Feed the record as you go — don't reconstruct it at 2:45.

STARTER PROMPTS

PROMPT-1

“Here are the corresponding lecture notes, lab instructions and the lab report template. Just read them for later reference and wait for my command.” (drop or attach PDF versions of all 3 documents right into this chat)

PROMPT-2 *“You are our AI engineering analyst for the LAB-3 When Is the Water Ready? cooling experiment. We are the lead engineers. Explain the model, check our error arithmetic, help us plot and organize evidence, and describe convergence patterns when we ask — but never change our measured values, the engineering requirement, or our units without flagging the change explicitly. Never invent measurements, results, or conclusions, and never declare our result ‘accurate enough’ — that judgment is ours. This workspace is the engineering record for our final report.”*

Any capable AI platform works — a project/workspace feature (e.g., ChatGPT Projects) is convenient, but one persistent conversation is enough.



Evidence first — the number is only the beginning

0:00–0:10		Phase 0 · Readiness quiz + verify Individual quiz — concepts, not calculation · Excel, Python, AI workspace proven working.
0:10–0:20		Phase 1 · Apparatus · safety · the question Mug, thermometer, roles · safety briefing · when will the water reach the target — and how sure are we?
0:20–0:30		Phase 2 · Commit predictions P1–P4 Crossing time · cooling shape · station agreement · achievable accuracy — in writing, unedited.
0:30–0:45		Phase 3 · Instrument & resolution check Eight nominally identical thermometers, ONE water bath — how many digits can we honestly defend?
0:45–1:30		Phase 4+5 · Cooling curve + model Water poured, lid ON, probe in port · a reading every 60 s · meanwhile: model, estimate k — no dead time.
1:30–1:50		Phase 6 · Predict, then compare Predicted vs. measured 40 °C crossing · absolute & percent error, by hand first.
1:50–2:10		Phase 7 · Resolution & convergence Refine the prediction's time step · the change column · convergence is evidence, not proof.
2:10–2:20		Phase 8 · Requirement → PASS / FAIL The engineering requirement is given — ± 2.0 min · judge PASS / FAIL against it.
2:20–2:35		Phase 9 · Class data checkpoint Forms submission · eight stations, one Excel view · why don't we all agree?
2:35–2:45		Phase 10 · BREAK THE AI Attack the overconfident accuracy claim — digits, resolution, reference, tolerance.
2:45–3:00		Phase 11+12 · Budget · record · sign Spend one unit of effort wisely · evidence complete · human review — then sign.



Stations & logistics — eight parallel experiments

EACH STATION'S KIT — NOMINALLY IDENTICAL



Red 16-oz stoneware soup mug (+ lid)

Lid pre-drilled with a center probe port — installed immediately after pouring, and it stays on.



TempPro folding digital thermometer

~3.9-in probe through the lid's center port · read °C · note the display's smallest step.



~175 mL controlled warm water · ~50–52 °C

Poured ONLY by the instructor/TA — NEVER boiling · $\leq$ ~55 °C if directed · volume adjustable after pilot.



Timer + team laptops

Phone timer is fine · Excel · Python/Jupyter · AI assistant · Microsoft Forms access.

TEAM RULES

- Rotate roles: reader · timer · recorder · analyst.
- Everyone reads the thermometer at least once; everyone records.
- One engineering record per team — team number, members, date, station at the top NOW.
- Do not move the mug once logging starts — and never lean over another station's mug.

EXPERIMENTAL DESIGN: eight nominally identical stations run the SAME experiment in parallel — the class dataset IS part of the experiment.



Safety & conduct — controlled water, treated as hot



SAFETY — CONTROLLED WARM WATER

- Controlled warm water $\approx 50\text{--}52\text{ }^{\circ}\text{C}$ ($\leq \sim 55\text{ }^{\circ}\text{C}$ if directed) — NEVER boiling
- Warm is not harmless — treat every mug as hot; report any spill at once
- Only the instructor/TA prepares, pours, reheats, or tops up water
- Lid ON immediately after pouring — it stays on; probe stays in its port
- Mug stays on the table for the whole run · keep laptops and cables clear



HONEST DATA

- Read the display — record what you SAW, not what you expected
- A missed reading is a gap, not an invention — log it, move on
- Do NOT “calibrate away” instrument disagreement — observe it
- Variation between stations is evidence, not a scoreboard
- Failed readings and rejected AI advice belong IN the record

Honest records beat pretty records. The journey is the evidence.



Readiness check — concepts, not calculation



READINESS QUIZ

- ☐ 5 questions · 5 minutes
- ☐ Individual · closed notes · no AI
- ☐ Collected before the water arrives



EXCEL + PYTHON

- ☐ Excel opens and computes
- ☐ Notebook runs a test cell
- ☐ You can interrupt the kernel



AI WORKSPACE

- ☐ “LAB-3” workspace created
- ☐ Starter prompts pasted
- ☐ Instructions + template uploaded



If something is broken

- Ask your instructor/lab facilitator for help — don't burn 40 minutes alone
- Paste the exact error into AI and ask what it means (often faster than you'd think)
- Keep your team moving on one working machine meanwhile

*All nine green? CHECKPOINT ① — the TA reviews quiz answers before water is poured.
Experienced with another IDE? Fine — as long as you can run and modify a notebook and support yourself.*



The standardized setup — identical at every station

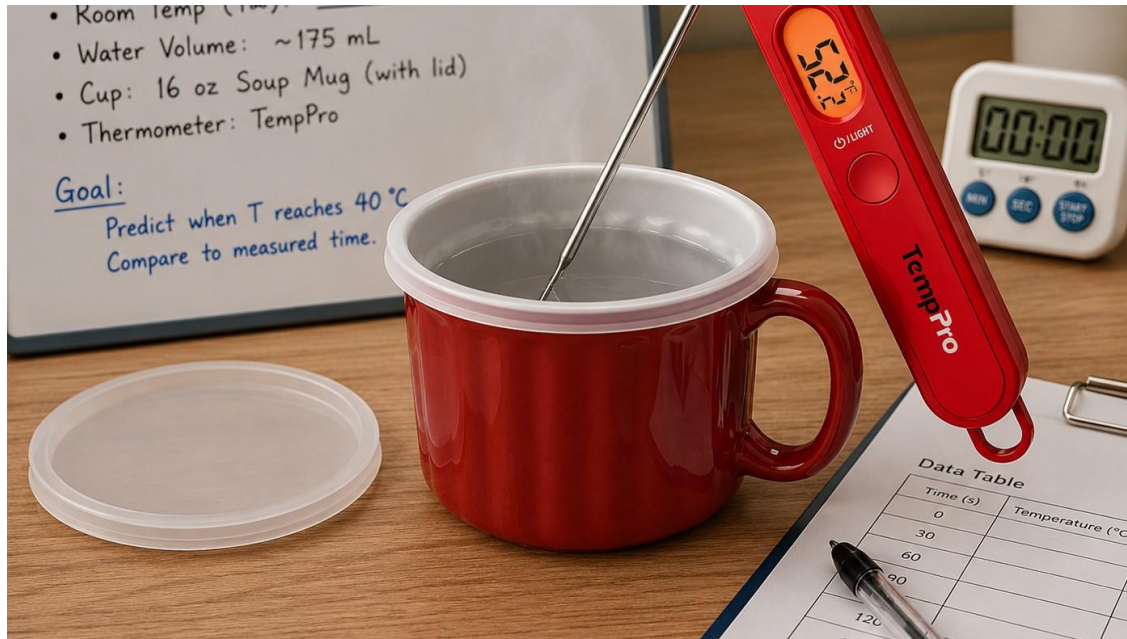
- 1 Mug on a stable, level spot — away from vents, windows, and sunlight
- 2 Record ambient T_{∞} BEFORE the pour — probe in air, or the posted room value
- 3 Thermometer on · reading in $^{\circ}\text{C}$ · note the display's smallest step
- 4 Instructor/TA pours the controlled warm water — lid goes ON immediately
- 5 Insert the probe through the lid's center port, to the marked depth
- 6 Probe near the CENTER of the water — touching neither wall nor bottom
- 7 Do NOT stir · do NOT lift the lid · do NOT re-seat the probe
- 8 One reader · one timer · one recorder — rotate roles during the run
- 9 Record T_0 at the moment logging starts — that instant is $t = 0$
- 10 One reading every 60 seconds — same probe, same spot, all run long



Why so strict? Every rule removes one source of error — that is what a standardized procedure is FOR. A lifted lid or a re-seated probe runs a DIFFERENT experiment.



A mug of warm water, a target temperature — and a deadline



FIND: the 40 °C crossing time t_{40} (min)

...and the evidence that says how accurate that prediction is — using YOUR station's measurements.

The model: $T(t) = T_{\infty} + (T_0 - T_{\infty}) \cdot e^{(-k \cdot t)}$

One parameter — k — carries all the physics. The accuracy JUDGMENT is yours — against the GIVEN ± 2.0 min engineering requirement.

Assumptions go in the record — including this one:

TALK TO YOUR AI TEAMMATE “Log our experiment inputs and assumptions: $T_0 =$ _____ °C, $T_{\infty} =$ _____ °C, $V \approx 175$ mL, lid on · probe in port, interval = 60 s, target ≈ 40 °C.”

ENGINEER DECIDES one uniform temperature for the whole mug is a MODELING choice — state it; revisit it in Phase 10.



Predict — in writing, right now

1

P1 — Minutes to 40 °C, a committed number

How long until the water crosses the target? Wrong is fine; uncommitted is not.

2

P2 — Will the cooling RATE stay constant?

Faster at the start, at the end, or steady — one word plus one sentence of physics.

3

P3 — Will all 8 stations get the same answer?

Within what spread — seconds or minutes? Circle one — you will check it at the class data checkpoint.

4

P4 — How accurately can YOU predict it?

± how many minutes? Commit it — Phase 8 compares it against the ±2.0 min engineering requirement.



TALK TO YOUR AI TEAMMATE “Here are our committed predictions P1–P4 — log them. Do not evaluate them yet.”

WHY PREDICT FIRST?

A prediction made after seeing the data isn't a prediction — it's a story.

Committing first is what makes the Phase 6 comparison honest — and it is how real engineering validation works.

Predictions go in the record BEFORE the warm water arrives — and are never edited afterward.



ENGINEER DECIDES the numbers you commit are yours — AI logs them, you own them.

Problem



Model



Compute



Analyze



Validate



Improve



Nominally identical instruments — do they agree?

1

Predict first — will eight thermometers agree?

Before comparing anything: write the spread you expect across eight thermometers reading the SAME water bath. ± 0.1 °C? ± 1 °C?

2

One bath — eight thermometers

The instructor/TA prepares ONE well-mixed, room-temperature water bath. Each station: probe in to the same depth, touching nothing · let it stabilize · read °C · post it.

3

Compare — and resist the urge to “fix” it

Write all eight readings side by side — min, max, spread. Do NOT calibrate anything away — observe and record. It returns in Phase 9.



TALK TO YOUR AI TEAMMATE “Here are eight readings of ONE water bath from eight identical thermometers — describe the spread and the defensible digits.”

INSTRUMENT NOTES

Resolution: the smallest step the display CAN show. Accuracy: how close a reading is to the truth — not printed on the box. Repeatability: does YOUR instrument repeat itself on the same bath? (T_{∞} is recorded BEFORE the pour — the probe never leaves the water afterward.)

If displays step by 0.1 °C but instruments disagree by 0.3 °C — which digits can you defend?



ENGINEER DECIDES the justified-digits call is made by the humans who saw the instruments — never by the display.

Problem



Model



Compute



Analyze



Validate



Improve



One reading every 60 seconds — no dead time

t (min)	0	1	2	3	4	5	...
T (°C) · ex.	52.0	51.3	50.7	50.0	49.4	48.8	...

Log T once every 60 s, from T_0 until the crossing is bracketed — then 3–5 confirming readings, and data collection may stop. Copy this table into your engineering record.

If $T(21 \text{ min}) > 40^\circ\text{C}$ and $T(22 \text{ min}) \leq 40^\circ\text{C}$, then $21 < t_{40} \leq 22$ — the crossing is BRACKETED within one 60-s interval · a missed reading is a gap marked “—”, never an invention · lid on, no stirring, mug stays put.

Record ALL readings — including odd ones. An inconvenient measurement is still a measurement; deleting it is how bias is born.



TALK TO YOUR AI TEAMMATE

“Here is our temperature–time table so far. Log it and plot T versus t . Flag anything inconsistent — but do not smooth, correct, or reinterpret the data.”



ENGINEER DECIDES

the water will not wait — while it cools (lid on, undisturbed): model, k estimate, plot, record. Cooling time is WORK time.

CHECKPOINT ② — $T_\infty \cdot T_0$ · running table · probe placement verified · parallel work under way.

Problem



Model



Compute



Analyze



Validate



Improve



One equation stands in for all the physics



The model: $T(t) = T_{\infty} + (T_0 - T_{\infty}) \cdot e^{(-k \cdot t)}$

Well-mixed water assumed — ONE temperature at any instant



Omitted physics — represented approximately by an effective k



Hand-check ONE point of your own data — on paper



Then automate: `cooling_T(t, T0, Tinf, k)` — your model function

minutes in, °C out — must reproduce your hand check

$$T(10) = 22 + 30 \cdot e^{(-0.223)} \approx 46.0 \text{ }^{\circ}\text{C}$$

(illustrative)

WHAT THE MODEL LEAVES OUT

- Convection in the water · evaporation at the surface · conduction into the stoneware · radiation · drafts.
- Probe placement and response time — the instrument is part of the system.
- Unstirred mugs grow warm tops and cool bottoms — one T is a modeling CHOICE.
- All of it is real physics — represented only approximately by one effective k . Useful $\neq$ complete.

A model this simple can still be USEFUL — if it is validated for its purpose. That is today's experiment.

Problem ›

Model ›

Compute ›

Analyze ›

Validate ›

Improve



From readings to a defended crossing time



Your T–t table — real measurements

DATA

Carry their own resolution: the instrument's, and the reading interval's.



Estimate k from a clean mid-run reading

PARAMETER

One logarithm — then checked against a second reading.



The model: $T(t) = T_{\infty} + (T_0 - T_{\infty}) \cdot e^{(-k \cdot t)}$

MODEL

A reduced-order stand-in for the real mug.



Step the model forward in time by dt

RESOLUTION

dt decides the finest answer the loop can give.



Stop at the first grid time at or past the crossing

STOPPING CRITERION

The condition — and the physics ($T_{\infty} < \text{target}$) that guarantees it triggers.



Report t_40 — with justified digits

OUTPUT

A number + its resolution + its evidence.

Every box adds its own error — measurement, parameter, model, resolution. Phase 6 measures the total; Phase 7 isolates the numerical share.



Python — the model gets a name first

WORKING METHOD — YOUR DATA DRIVES THE CODE

- AI may coach the code; the DATA stays yours
- Pick one clean mid-run reading — not the first, not the last
- Solve the model for k — one logarithm, no fitting library
- `cooling_T()` must reproduce your hand-checked point
- `estimate_k()` must invert `cooling_T()` — feed one into the other
- Check k with a SECOND reading — the spread is evidence
- Unexplainable code does not enter your program

Two small functions is right-sized.

A model and its inverse — both testable.

If two readings give slightly different k 's, that spread is information about the model — not a bug to hide.

THE CODE

```
import math
def cooling_T(t, T0, Tinf, k):
    return Tinf + (T0-Tinf) * math.exp(-k*t)

def estimate_k(t, T, T0, Tinf):
    return math.log((T0-Tinf)/(T-Tinf)) / t

# one clean mid-run reading — t in MINUTES:
k = estimate_k(10.0, 46.0, 52.0, 22.0)

print(k)      # 0.0223 (1/min)
# check with a SECOND reading:
print(estimate_k(15.0, 43.5, 52.0, 22.0))
```

Same signatures everywhere — call `cooling_T(t, T0, Tinf, k)` exactly that way inside the Phase 6 loop too.



The stepping loop predicts the crossing

THE PREDICTION — AUTOMATED

```
T0, Tinf, k, T_target = 52.0, 22.0, 0.0223, 40.0 # °C · 1/min
dt_seconds = 30 # the resolution knob
dt = dt_seconds / 60.0 # convert s → MINUTES — k is 1/min!
t = 0.0
while cooling_T(t + dt, T0, Tinf, k) > T_target:
    t = t + dt
print("Crossing at ~", t + dt, "min")
```

If the loop never stops: is T_∞ at or above the target? The model then says the water NEVER gets there —

interrupt the kernel and re-read your inputs; the stopping condition rests on physics.

BEFORE YOU RUN IT — WRITE DOWN:

- ☐ Your expected output — your cooling plot knows the neighborhood
- ☐ Why this loop must terminate — what if T_∞ were above 40 °C?
- ☐ What the loop tests: the NEXT grid time — $t + dt$ is the crossing
- ☐ Which line converts seconds → minutes — and why it must
- ☐ Then run — compare Python vs your plot vs prediction P1
- ☐ All three must tell one consistent story
- ☐ Any disagreement is evidence — chase it before Phase 7

The answer is a multiple of dt — that is what numerical resolution MEANS. Phase 7 turns that knob on purpose.



Two error numbers — by hand, before any code

THE ARITHMETIC — SHAPE ONLY

```

E_abs  = | t_predicted - t_ref |           [min]
E_%    = E_abs / t_ref × 100              [%]
t_ref  = MIDPOINT of the measured bracket — an ESTIMATE
        ...the 60-s bracket width remains part of the
        uncertainty...
REPORT  ...only digits the evidence supports...

```

The two classic gaps: percent error taken against the prediction instead of the measured reference —

and quoting seconds when the crossing is only bracketed to a 60-s interval.

YOUR ERROR STATEMENT MUST CONTAIN:

- ☐ Both error measures, with units
- ☐ The reference: bracket midpoint — labeled an estimate
- ☐ The reference's bracket — one 60-s interval wide
- ☐ The dt behind the predicted value
- ☐ Digits consistent with BOTH resolutions
- ☐ One sentence: is the error big? Compared to what?
- ☐ ...all computed by hand BEFORE any code checks it

MEASURE › MODEL › PREDICT › DEFEND — the defense starts here, with honest arithmetic.



Turn the resolution knob — on purpose

dt (s)	Predicted t_40 (min)	Change from previous (min)	Model evaluations	Change vs. the 60-s interval?
60	fill in	—	fill in	fill in
30	fill in	fill in	fill in	fill in
10	fill in	fill in	fill in	fill in
1	fill in	fill in	fill in	fill in

One experiment per row — record t_40, the change from the row above, and the evaluation count. In code: $dt = dt_seconds / 60.0$ — the table lists SECONDS, the model runs in MINUTES.

Record the sentence: “Below dt = ____ s, t_40 changes by less than ____ s — smaller than our 60-s reading interval.” That sentence is your convergence evidence.

Change ONLY dt — same model, same k, same target. One knob, one experiment per row. Never retype anything else.



TALK TO YOUR AI TEAMMATE “Here are my results for several time steps. Do not tell me which to accept — help me examine how the prediction changes as dt shrinks.”



ENGINEER DECIDES read the change column yourselves, before AI describes it — the shrinking changes ARE the convergence evidence.

Smaller dt → finer resolution → more evaluations. Where does refining stop buying anything?

Problem



Model



Compute



Analyze



Validate



Improve



Converged — toward what? A disagreement worth keeping

Evidence	Value (min)	Differs from refined by	What the gap tests
Refined loop (finest dt)	fill in	—	numerical resolution
Analytical model crossing	fill in	fill in	the stepping itself
Measured bracket	midpoint: fill in	fill in	the MODEL + measurement

$t_{\text{model,analytical}} = \ln((T_0 - T_\infty) / (T_{\text{target}} - T_\infty)) / k$ — exact for the MODEL and your k , not for the physics. One Python line.

Fill the table from your own runs — then say which gap is numerical and which is not. “Analytical” means exact for the reduced-order model — never exact physical truth.

The refined loop and the analytical MODEL crossing agree to seconds — exact for the model, not for the physics. The measured crossing BRACKET can still sit minutes away. The numerical error shrank; the model error did not.

 **TALK TO YOUR AI TEAMMATE** “Here is our table — refined, analytical, measured. Describe what each gap is evidence OF.” Then grade its answer.

 **ENGINEER DECIDES** when numerical changes are tiny next to measurement uncertainty, refinement is done — that judgment is yours.

Vocabulary today: convergence, tolerance, justified precision — and the chapter's law: CONVERGED ≠ CORRECT.

Problem



Model



Compute



Analyze



Validate



Improve



Tolerance is a requirement — not a Python setting

1

The requirement is GIVEN — read it

ENGINEERING REQUIREMENT: predicted vs. measured crossing (bracket midpoint) within ± 2.0 min — one requirement for all eight stations, set by the stakeholder.

2

BEFORE judging — revisit P4

You predicted your own achievable accuracy. Does the ± 2.0 min requirement demand more than your evidence can give — or less?

3

Judge: PASS / FAIL — from evidence

$|E_{\text{abs}}| \leq 2.0$ min, with the error known to defensible digits $\rightarrow$ PASS; otherwise FAIL. Record the verdict AND the evidence.



TALK TO YOUR AI TEAMMATE “Log the ± 2.0 min requirement and our PASS / FAIL evidence. Do not soften or restate the requirement.”

WHO OWNS THE TOLERANCE?

Not the computer, not AI, not team preference. The engineering REQUIREMENT owns it — here the instructor is the stakeholder. Would ± 2 min still serve a medical device? A production line? A safety-critical control? Same math — different requirement, different verdict.

You may PREDICT your accuracy (P4) — you do not get to pick the requirement.



ENGINEER DECIDES if the verdict is FAIL, that is a finding — diagnose which error source caused it before Phase 11.

Problem ›

Model ›

Compute ›

Analyze ›

Validate ›

Improve



CLASS DATA CHECKPOINT — eight stations, one dataset

Forms field	your entry	Forms field	your entry
Station number		Estimated k (1/min)	
$T_0 \cdot T_\infty$ (°C)		Predicted t_40 (min)	
Volume · bath check		Crossing bracket (min)	
Requirement	± 2.0 min	E_abs (min) · E_%	
Uncertainty note		PASS / FAIL	

Submit ONE standardized station summary — the fields below. The instructor's Excel view then compares all eight stations.

Then the question that matters: “Eight teams ran nominally the same experiment — why didn't we get the same answer?” Keep hypotheses; the next slide sorts them.

One submission per station — the recorder submits; the team confirms every number against the record first.



TALK TO YOUR AI TEAMMATE

“Before we submit: cross-check our Forms entries against our record — every value, every unit. Flag mismatches; change nothing.”



ENGINEER DECIDES CHECKPOINT ③ — the TA confirms your station's submission before the class comparison opens.

QR code / Forms link: [INSTRUCTOR: insert Microsoft Forms link / QR code here] · the instructor compiles all eight rows in Excel, live.

Problem



Model



Compute



Analyze



Validate



Improve



Eight answers — sort the disagreement, don't wave at it



Numerical resolution — the Δt each station chose

NUMERICAL

Bounded by your own refinement table — usually SECONDS here.



Instrument differences + reading resolution

INSTRUMENT

Your Phase 3 spread returns — nominally identical is not identical.



Initial conditions — T_0 , T_∞ , volume, start timing

INITIAL CONDITIONS

Small differences in the start propagate into the crossing time.



Procedure — probe depth, lid seal, timing

PROCEDURE

A re-seated probe or a lifted lid ran a slightly different experiment.



Environment — drafts, sun, bench location

ENVIRONMENT

Eight benches are eight micro-climates — the model has one T_∞ for each.



Model inadequacy — one k for a whole mug

MODEL

If ALL stations err the same direction, suspect the MODEL, not the stations.

Station-to-station variation is NOT automatically “numerical error.” Always name the source that could produce THIS disagreement — with evidence.



Phase 10 · 2:35–2:45 · Attack the accuracy claim — not the code



Interrogate the claim — pick at least three

Six decimal places vs. a 0.1 °C display · what dt — and did it say? · converged, by what evidence? · “accurate” against WHAT reference — and its 60-s bracket? · does it meet the ± 2.0 min requirement? · validated for THIS mug, THIS room?



Re-read the AI conversation

Did AI silently assume a k, a T_{∞} , a dt, or a tolerance you never authorized? Verdict what you find ›



Document every attack

What you asked, what it claimed, what you concluded — “the claim survived” is a finding too, if you show HOW you verified it.

VERDICT WHAT YOU FIND:



ACCEPT — with your reason



MODIFY — then retest



REJECT — with evidence



VERIFIED — also a result

BREAK THE AI: “Our model predicts the water reaches 40.000 °C at $t = 22.704183$ minutes — a highly accurate prediction.” Attack the CLAIM, then rewrite it into a statement you would sign.



Rebuild the claim — from evidence, not adjectives



1 · EVIDENCE

List what an accuracy claim needs: measured error, resolution, convergence change, the requirement, model validation. Which are present? Which are

missing?
Six decimals from a 0.1 °C display. That gap is data.



2 · INTERROGATE

Put your own evidence against the claim: your dt table, your crossing bracket, your instrument check, the ± 2.0 min requirement. Where exactly does the claim overreach?

“Highly accurate” — compared to WHAT?



3 · REWRITE → DEFEND

“At $dt = __ s$ the prediction changed $__ s$; error vs. the bracketed measured crossing $__ min (__ \%)$; within the $\pm 2.0 min$ requirement → PASS for THIS purpose.”

AI may draft the rewrite — it becomes YOURS when you can defend every number in it.

Never grade tone — grade EVIDENCE. A confident sentence with no reference, no resolution, and no tolerance is not accuracy; it is typography.

Printed digits are free — evidence is not. The number of digits you report is itself an engineering claim.



One more unit of effort — spend it or bank it



1 · OPTIONS

Finer dt? Better thermometer? More readings? Better k estimate? Better model — gradients, mixing? Or: nowhere — the result is already sufficient.

Every option has a cost — name it.



2 · CHOOSE

Compare YOUR error contributions: numerical change vs. instrument spread vs. station variation vs. model error. Spend where the biggest error lives — or declare sufficiency.

AI may propose; the choice is yours.



3 · JUSTIFY

Write the choice and the reason in the record. “Already adequate” is a legitimate engineering decision — when the evidence supports it.

One choice · one reason · in the record.

Watch for the reflex — “more decimals” is not automatically better engineering. Refining dt while thermometers disagree by 0.3 °C polishes the wrong thing. Spend effort where the evidence says the error lives.

Do not manufacture an improvement — “sufficient, and here is why” is a complete engineering answer.



Make — and defend — the accuracy claim

*“Under the stated assumptions ($T_0 = \underline{\hspace{1cm}}$ °C, $T_\infty = \underline{\hspace{1cm}}$ °C, $V \approx 175$ mL, lid on · undisturbed, $k = \underline{\hspace{1cm}}$ 1/min from our data), the model predicts the 40 °C crossing at $\underline{\hspace{1cm}}$ min ($dt = \underline{\hspace{1cm}}$ s, change from previous refinement $\underline{\hspace{1cm}}$ s). Measured crossing bracket: $\underline{\hspace{1cm}}$ – $\underline{\hspace{1cm}}$ min (midpoint $\underline{\hspace{1cm}}$ min — an estimate). Error vs. midpoint: $\underline{\hspace{1cm}}$ min ($\underline{\hspace{1cm}}$ %), against the ± 2.0 min engineering requirement $\rightarrow$ the model **PASSES** / **FAILS** for this purpose.”*

YOUR DEFENSE — one line of evidence on each rule:

The four lines: ① hand-checked model point ② convergence evidence — the change column ③ measured-crossing comparison ④ class-dataset consistency + stated limitations.

An engineering answer is a NUMBER + ASSUMPTIONS + EVIDENCE + JUDGMENT. A bare number is a guess wearing a lab coat.

Problem



Model



Compute



Analyze



Validate



Improve



Leave with the record complete — the report follows



EVIDENCE OF DOING

Instrument-check readings
T–t table — every reading
 $T_{\infty} \cdot T_0 \cdot \text{volume} \cdot \text{interval}$
Predictions P1–P4 —
unedited
Cooling plot + crossing
bracket



EVIDENCE OF THINKING

k estimate + hand check
Refinement table + change
column
Error arithmetic + BREAK
THE AI



DECISIONS RECORDED

± 2.0 min requirement applied
Justified reported digits
PASS / FAIL + evidence
Accuracy-budget choice



BEFORE YOU LEAVE

Forms submission confirmed
Missing report items listed
Assembly assigned — who, by
when
Station reset — mug emptied
& dried, lid rinsed, probe
folded
CHECKPOINT ④ — TA sign-off

Final formatting may be finished after lab — a complete, honest record tonight beats a polished report missing evidence.

Problem ›

Model ›

Compute ›

Analyze ›

Validate ›

Improve



Final report · human review & approval

“Before submission, we reviewed the completed report against our original engineering record — measurements, calculations, observations, code, plots, tests, validation evidence, and conclusions. We accept responsibility for the submitted engineering work.”

Human reviewed and approved.

EVERY TEAM MEMBER SIGNS — BEFORE
SUBMISSION (digital signature: create
once, copy-paste)

The recurring question, now literal: ***Would you confidently put your name on this result?***

AI assists. Humans observe. Humans verify. Humans decide. Humans approve.

Problem ›

Model ›

Compute ›

Analyze ›

Validate ›

Improve



One concise report — built from your record



FRAME IT

- 1 Problem / objective
- 2 Reduced-order model
- 3 Assumptions
- 4 Apparatus · safety · photo
- 5 Instrument check



GROUND IT

- 6 Conditions — $T_0 \cdot T_\infty \cdot V \cdot \Delta t$
- 7 Requirement — ± 2.0 min
- 8 Predictions P1–P4



BUILD IT

- 9 Data + plot + crossing bracket
- 10 k estimate + hand check
- 11 AI analysis · 12 Python
- 13 Refinement + convergence



OWN IT

- 14 BREAK THE AI — the claim
- 15 Validation $\times 4$ + precision
- 16 Class data checkpoint
- 17 Final statement · 18 budget + limits
- 19 Human review & approval — signed

A few honest sentences per section. Tables, plots, and photos carry the evidence — prose carries the judgment.

Problem



Model



Compute



Analyze



Validate



Improve

WHAT JUST HAPPENED

**You measured a real physical system,
stood a one-parameter model in for its physics,
predicted a crossing time — and measured the truth,
quantified the error, judged it against a requirement —
and signed the claim.**

MEASURE → MODEL → PREDICT → DEFEND. Your refinement study changed the answer by seconds; the measurement disagreed by minutes — you now know which error was worth chasing. And your loop still WALKED to the crossing, one dt at a time. Chapter 4 asks the question this laboratory earned: can we reach the required accuracy more efficiently?

Problem

›

Model

›

Compute

›

Analyze

›

Validate

›

Improve

FUNDAMENTALS OF AI-AUGMENTED ENGINEERING

CHAPTER 3 / LAB-3

Engineering Report Rubric

When Is the Water Ready? — Main LAB-3 Team Report

Suggested Assessment · 40 points

Used by the instructor and, where applicable, by an AI grading agent. Assessment rewards experimental evidence, honest error analysis, convergence reasoning, tolerance judgment, and human responsibility.

Criterion	What full credit looks like	Pts
Problem, model, assumptions & conditions	Clear objective; the reduced-order model stated with the well-mixed assumption named and k identified as an EFFECTIVE parameter; conditions (T_0 , T_∞ , $V \approx 175$ mL, lid on, 60-s interval) carry units; the target and the ± 2.0 min engineering requirement identified.	4
Instrument check & cooling data	Common-water-bath comparison readings recorded and reasoned about — never “calibrated away”; complete T – t table at the 60-s interval, lid on, undisturbed; measured crossing recorded as a one-interval BRACKET; labeled cooling plot.	6
Committed predictions	P1–P4 recorded before any data and left unedited; each revisited against the evidence afterward. Prediction correctness is not graded — commitment and reasoning are.	3
Model & k estimate	k estimated by the scaffolded method with one hand-checked point shown; <code>cooling_T()</code> reproduces the hand check; the second-reading consistency check performed and its spread discussed.	4

Continued

RUBRIC for the Lab Report

Prediction, error & justified precision	Predicted crossing compared against the bracketed measured crossing; E_abs and E_% computed by hand, any midpoint interpolation named as an added estimate; reported digits defended against the instrument resolution and the 60-s bracket.	6
Resolution & convergence study	dt table complete with the change column and evaluation counts, the seconds → minutes conversion explicit in the code; the trend read as evidence consistent with convergence, not proof; the numerical change weighed against measurement uncertainty.	6
Requirement, PASS / FAIL & final statement	The ±2.0 min engineering requirement applied as given and its ownership explained; the verdict follows from the explicit error-vs-requirement comparison; the final engineering statement complete, every blank filled, digits justified.	5
Class data checkpoint	Forms submission complete and confirmed; the eight-station comparison discussed; disagreement attributed to specific named sources — never “numerical error” by default.	3
BREAK THE AI, accuracy budget & signed approval	The overconfident claim attacked with at least three evidence-based reasons and rewritten defensibly; the accuracy-budget choice justified; limitations stated; human review signed by every member.	3
TOTAL		40

Problem



Model



Compute



Analyze



Validate



Improve

CHAPTER - 3

Lab 3 · ≈ 25 MIN

Lab Extension Challenge:

(Individual Assignment)

Students complete the analysis individually using the common lid-off dataset supplied by the instructor/TA. No at-home experiment is required.

Problem



Model



Compute



Analyze



Validate



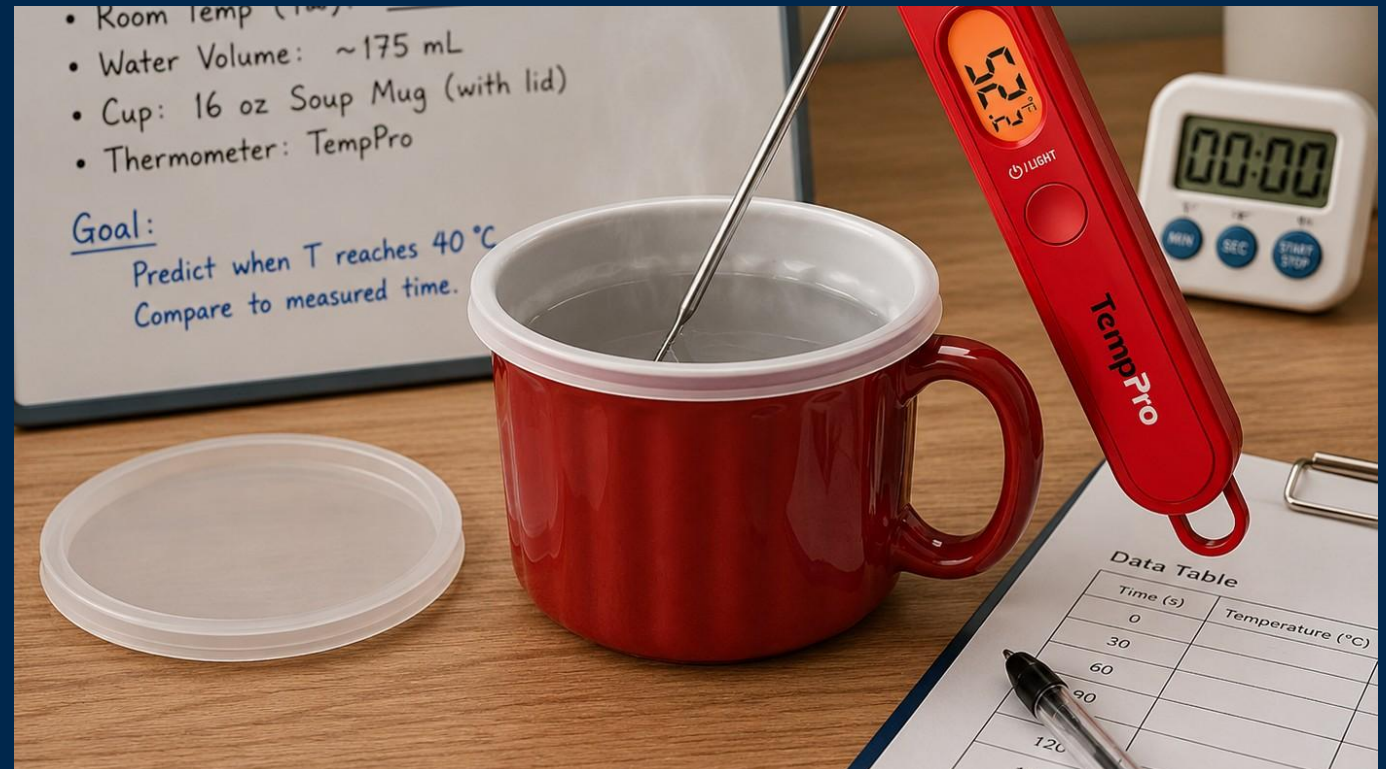
Improve



When Does the Model Stop Being Good Enough?

The numerics may still converge — is the model still correct enough for the purpose? Predict first — then measure.

An extension of the Lab-3 cooling study. You already built the model, the prediction, and the error toolkit — now taking the lid OFF puts the MODEL itself to the test.



Problem

>

Model

>

Compute

>

Analyze

>

Validate

>

Improve



An assessment mini-challenge — not another lab report

1

Designed as part of the Lab-3 assessment

The rubric is on the last slide of this section.

2

Separate from the 3-hour lab — analysis only

An individual follow-on, completed at your convenience under the instructor's arrangements — in-lab ONLY if time and approved equipment access permit.

3

Submit a separate short report

Download and use Chp3-Lab3-LabExtension-Challenge-WhenDoesTheModelStop-Report-Template.docx

THE ENGINEERING QUESTION

Take the lid OFF the same mug. The numerical method still converges — but is the baseline (lid-on) model still accurate enough for the purpose?

Same workflow as the whole lab:

**Predict → Measure → Compare → Explain
→ Ask AI → Critique → Validate → Decide**



ENGINEER DECIDES *the question is not whether the code converges — it is whether the MODEL still deserves your signature under a changed condition. AI assists. Humans observe, verify, decide, and approve.*

Problem



Model



Compute



Analyze



Validate



Improve



Commit to a prediction

1

The lid-off question

With the lid OFF, will the water reach 40 °C sooner, later, or at about the same time as the lid-on baseline? Circle one.

2

Sketch the shape

Will the lid-off cooling curve still look exponential — or change character entirely?

3

How wrong will baseline k be?

Predict the ratio $k_{\text{off}} / k_{\text{baseline}}$ — about 1? About 1.3? About 2? Commit a number.

4

Which physics does removing the lid change?

Evaporation · convection from the free surface · radiation — which of these paths was the lid blocking?

THE MODEL

$$T(t) = T_{\infty} + (T_0 - T_{\infty}) \cdot e^{(-k \cdot t)}$$

same model form · same mug · same volume —
held fixed

Look closely: the lid never appears in the equation.

So where — exactly — would its physics have to hide?



ENGINEER DECIDES record your predictions NOW — before opening/analyzing the supplied lid-off dataset; do not revise them afterward (correctness is not graded).

Problem

**Model**

Compute



Analyze



Validate



Improve



Lid off — one model on trial

Quantity	Baseline (lid on)	Lid OFF	Compare	Notes
$T_0 \cdot T_\infty$ (°C)	Lab-3	_____	$\Delta T_0 = \underline{\hspace{1cm}}$	
k (1/min)	Lab-3	_____	ratio = $\underline{\hspace{1cm}}$	
Pred. t_40 (base k)	Lab-3	_____	—	
Measured t_40 bracket	Lab-3	_____	E_abs = _____	
Verdict (± 2.0 min)	P / F	P / F	E_% = $\underline{\hspace{1cm}}$	

HELD FIXED

same mug
lid OFF

175 mL · 60 s · same depth ·
10–15 min is enough
SAFETY UNCHANGED:
controlled warm water,
*...never boiling · mug stays put ·
no leaning over it*

THEN — ONE PICTURE

Overlay the lid-off readings on your BASELINE (lid-on) model curve — one plot. A few readings are enough to show the divergence. Before reading on: does your committed prediction survive contact with the plot?



ENGINEER DECIDES *reuse your Lab-3 functions — test whether the same model FORM with a re-estimated k remains adequate; the DATA decide. If you are rewriting code, stop and ask why.*

Problem



Model



Compute



Analyze



Validate



Improve



Interrogate the change — don't assume it

THE k CHECK

Estimate k_{off} from the supplied lid-off readings — same method as Lab-3. Compute the ratio $k_{\text{off}} / k_{\text{baseline}}$.

Write the actual ratio next to your committed prediction.

THE CROSSING CHECK

Predict the lid-off 40 °C crossing with the BASELINE k . How far off is it against the lid-off data?

Read it off your overlay plot — the data, not your expectation.

CONVERGED ≠ CORRECT

Refine dt on the baseline- k prediction — it converges beautifully. It is also wrong.

Explain why convergence could not warn you the lid came off.

THE MODEL ASSUMPTION

Same model FORM, new k — does the exponential shape still describe the lid-off data?

What does that say about where the lid's physics lives?

✓ **ENGINEER DECIDES** *use your measured evidence — not what you expected to see. The data decide what changed.*

Problem

>

Model

>

Compute

>

Analyze

>

Validate

>

Improve



Ask AI for the analysis — then grade it

SUGGESTED PROMPT

“Here are two cooling datasets from the same standardized mug setup — my Lab-3 lid-on baseline and the TA-supplied lid-off dataset. The lumped model $T(t) = T_{\infty} + (T_0 - T_{\infty})e^{(-kt)}$ fit the baseline with $k = \underline{\hspace{1cm}}$. Analyze whether the same model form with a different k describes the lid-off data, propose physical reasons for the change, and list what evidence would justify still using the baseline k .”

Any capable AI assistant (your engineering analyst). Record a concise summary of its analysis. Do NOT copy the full answer.

NOW CRITIQUE THE AI RESPONSE

- 1 Is the analysis consistent with your two measured curves?
- 2 Does it match YOUR numbers — check one prediction against it?
- 3 Did it invent data, a k value, or a mechanism you did not measure?
- 4 Verdict: ACCEPT · MODIFY · REJECT — with a reason.
- 5 What one additional measurement would test its claim?

 **THE KEY DISTINCTION** *a fluent analysis is not necessarily a correct analysis. Your two measured curves outrank its prose.*

 **ENGINEER DECIDES** *AI analyzes; engineers verify the analysis against measured evidence.*

Problem

>

Model

>

Compute

>

Analyze

>

Validate

>

Improve



Validate the claim — then make the call

1 INDEPENDENT CHECK

Verify the important claim without repeating the analysis: check k_{off} with a SECOND reading, or hand-check one model point.

One line of Python or a calculator. Do the two estimates agree to the digits you defend? If not — which do you trust, and why?

2 ENGINEERING JUDGMENT

State a defensible conclusion: is the BASELINE model, with its baseline k , ACCEPTABLE or NOT ACCEPTABLE for predicting the lid-off crossing — against the ± 2.0 min requirement?

Explain physically why the behavior changed — which heat paths OPENED when the lid came off, and why one new k can (or cannot) absorb that change.

You are NOT expected to build an evaporation model — identifying the assumption and its consequence is the deliverable.

CONVERGED $\neq$ CORRECT — the numerics never noticed the lid come off. Validation did.

Problem

›

Model

›

Compute

›

Analyze

›

Validate

›

Improve



One compact report — 10 points

LAB EXTENSION CHALLENGE REPORT

separate from your main Lab-3 report · ½–1 page

- 1 **Committed predictions — lid-off question + k ratio**
- 2 **Conditions — TA-supplied lid-off dataset · same standardized mug setup**
- 3 **Supplied lid-off readings + overlay plot**
- 4 **k_{off} vs k_{baseline} + method**
- 5 **Baseline-model error at the lid-off crossing**
- 6 **Requirement verdict — ACCEPTABLE / NOT + physics**
- 7 **AI analysis + your ACCEPT / MODIFY / REJECT**
- 8 **Independent check + CONVERGED ≠ CORRECT note**

No formal Intro / Methods / Results / Discussion structure required.

Recommended Challenge Rubric	Points
Committed predictions (before measuring)	1
Lid-off data + overlay plot	3
k comparison · crossing error · requirement verdict	2
AI use AND critical evaluation	2
Independent check / validation	1
Final judgment — CONVERGED ≠ CORRECT	1
TOTAL :	10

Serious enough to take seriously — small enough that Lab-3 still belongs to the main lab.

Problem

›

Model

›

Compute

›

Analyze

›

Validate

›

Improve

RUBRIC for the Lab Extension Report

FUNDAMENTALS OF AI-AUGMENTED ENGINEERING

CHAPTER 3 / LAB-3

Lab Extension Challenge Rubric

When Does the Model Stop Being Good Enough? · Individual Assignment

Suggested Assessment · 10 points

Separate from the main LAB-3 team report. One compact individual challenge report. Assessment rewards the committed prediction, measured evidence, critical evaluation of AI, and engineering judgment.

Criterion	What full credit looks like	Pts
Committed predictions	The lid-off question and the k_{off} / k_{baseline} ratio committed before measuring, with brief physical reasoning. Correctness of the prediction itself is not graded.	1
Lid-off data + overlay plot	Sufficient lid-off readings at the 60-s interval; overlay against the baseline (lid-on) model curve with labeled axes and units; conditions recorded — same mug, volume, and stated probe placement.	3
k comparison, crossing error & requirement verdict	k_{off} estimated and compared to the baseline; the baseline-k prediction's error at the lid-off crossing computed; the ± 2.0 min requirement applied to an ACCEPTABLE / NOT ACCEPTABLE verdict.	2
AI use AND critical evaluation	The AI analysis summarized concisely and graded ACCEPT / MODIFY / REJECT against the measured curves, with a reason.	2
Independent check / validation	The key claim verified by a second reading or a hand-checked model point, reconciled to the digits being defended.	1
Final engineering judgment	A defensible CONVERGED $\neq$ CORRECT conclusion: why the numerics could not detect the model change, and what should change in the model or workflow.	1
TOTAL		10

Problem



Model



Compute



Analyze



Validate



Improve