REVISION PRIORITY & ACTION TABLES

Revision Plan for Manuscript
6863b9b0-4db8-4b53-843f-5be5e907cf62

February 5, 2026

SUMMARY / TOM TAT TINH HUONG

- **Decision:** Major Revision (HIGH chance of acceptance if revised well)
- **Reviewers:** 8 reviewers
- **Deadline:** 10 working days (can request 5-7 more days)
- **Strategy:** Fix CRITICAL issues first, then MAJOR

Contents

1 TABLE 1: Priority Matrix

1.1 Table 1A: Overview of Issues by Severity

Table 1: Priority Matrix - All Issues Categorized

#	Issue	Severity	Rev.	Time	Priority Action
CRITICAL - Must Fix (will cause rejection if not addressed)					
1	**"Overall" aggregation unclear** What does "overall" mean? LOC dominates FP?	FATAL	R6, R8	0.5 day	1 - DO NOW
2	**COCOMO II baseline unfair** Comparing against uncalibrated COCOMO creates "straw man"	FATAL	R1, R7, R8	2-3 days	2 - REQUIRED
3	**Target leakage: Developers feature** Developers = ceil(Effort/Time) uses target to create feature	FATAL	R8	0.5 day	3 - QUICK FIX
4	**FP n=24 - protocol inappropriate** 80/20 split gives 5 test samples, unstable grid search	FATAL	R6, R7, R8	1 day	4 - MUST DO
5	**Dataset manifest missing** Cannot audit data sources, dedup criteria, potential leakage	FATAL	R7, R8	1 day	5 - IMPORTANT
6	**Formatting issues** No captions, low resolution figures, no line numbers	FATAL	R5, R6, R7	1 day	6 - EASY FIX
MAJOR - Important (hard to accept without these)					
7	**Equation errors & duplicate Time formula** Section 2.1 has duplicate equation	MAJOR	R6	0.5 day	7
8	**R-squared column shows "–"** Missing R^2 values for all models	MAJOR	R6	0.5 day	8
9	**Novelty weak ("just a pipeline")** Contribution unclear beyond combining existing methods	MAJOR	R1, R3, R8	1 day	9 - IMPORTANT

(continued on next page)

(continued from previous page)

#	Issue	Severity	Rev.	Time	Priority Action
10	**Related Work insufficient** Missing comparison with SOTA, no discussion of 4 DOI papers	MAJOR	R3, R4, R5	1 day	10 - MUST DO
11	**SOTA models missing** XGBoost/LightGBM/CatBoost not included	MAJOR	R4, R7	1-2 days	11 - if time permits
12	**Interpretability claim unsupported** "More interpretable" stated but no feature importance shown	MAJOR	R7	1 day	12
13	**Ablation study missing** Cannot assess contribution of preprocessing steps	MAJOR	R5, R7	1 day	13
14	**Generalization unclear** No leave-one-source-out cross-validation	MAJOR	R7, R8	1-2 days	14 - if time
MINOR - Improvements (polish for higher score)					
15	**Language quality / AI-like tone** Some phrases sound generated	MINOR	R4, R7	1 day	15 - polish
16	**Additional metrics needed** Add MAPE, MdMRE, RAE	MINOR	R1	0.5 day	16 - quick
17	**Confidence intervals missing** Report bootstrap 95% CI for all metrics	MINOR	R1	1 day	17
18	**Length reduction** Move some content to Supplementary	MINOR	R1	1 day	18

1.2 Table 1B: Recommended 10-Day Timeline

Days	Tasks	Priority IDs
1-2	**CRITICAL Block 1:** Define aggregation + COCOMO calibration + remove leakage + FP protocol	1, 2, 3, 4
3	**CRITICAL Block 2:** Dataset manifest + fix formatting	5, 6
4-5	**MAJOR Block 1:** Rewrite novelty + Related Work + cite DOIs + fix equations	7, 8, 9, 10
6-7	**MAJOR Block 2:** Interpretability plots + ablation study + additional metrics	12, 13, 16, 17
8	**OPTIONAL:** XGBoost (if time) or polish language	11, 15
9	**Integration:** Combine all changes, consistency check	All
10	**Final Review:** Advisor approval + finalize response letter + submit	All

Table 2: 10-Day Timeline Allocation

2 TABLE 2: Action Tables - Detailed Steps

2.1 Table 2A: Point-by-Point Responses

Table 3: Response Strategy for Each Reviewer Comment

Reviewer	Question / Requirement	Response Strategy
REVIEWER 1 - Methodological Concerns		
R1.1	Clarify novelty beyond "unified pipeline"	**Action:** Rewrite Abstract + Intro emphasizing 3 novelties: (1) Multi-schema harmonization with auditable manifest, (2) Fair calibrated baseline, (3) Statistical rigor (Wilcoxon + Cliff's Delta) **Changes:** Abstract lines 8-12, Intro Section 1.3, NEW comparison table
R1.2	Recalibrate COCOMO II for fair comparison	**Action:** Fit A, B parameters on training data per schema to report COCOMO II (original) vs (calibrated) vs RF **Changes:** NEW Methods subsection "Baseline Calibration", NEW Table 3 results, Discussion update
R1.3	Add modern datasets (GitHub/Jira)	**Strategy A (if time):** Collect mini-validation 30-50 projects from GitHub with effort in README **Strategy B (recommended):** Explain limitation + Future Work **Changes:** Data section + Results or Limitations
R1.4	Report MAPE, MdMRE, RAE	**Action:** Add 3 metrics to code, update all tables **Changes:** Section 2.4 (Metrics), Tables 1, 3, 4

(continued on next page)

(continued from previous page)

Reviewer	Question / Requirement	Response Strategy
R1.5	Confidence intervals for all metrics	**Action:** Bootstrap 95% CI or CI from 10 seeds → format "Mean [95% CI]" **Changes:** All result tables
R1.6	Reduce length	**Action:** Move preprocessing details + some figures to Supplementary **Changes:** Create supplementary.pdf
R1.7	Reproducibility: code & data	**Action:** Upload code to GitHub/Zenodo with DOI + dataset manifest **Changes:** Data Availability section
REVIEWER 3 - Structure & Clarity		
R3.1	Restructure Introduction	**Action:** Separate motivation, gap, and contribution clearly **Changes:** Intro Sections 1.1, 1.2, 1.3
R3.2	Expand Related Work + cite DOIs	**Action:** Create NEW Section 2: Related Work (2-3 pages). Cite & discuss 4 DOI papers: aisy.202300706, patcog.112890, ACCESS.3480205, engappai.111655 **Changes:** NEW Section 2, comparison table
R3.3	Explicit assumptions & limitations	**Action:** Add NEW Section 3.6: Assumptions (linear cost-effort, no team dynamics) + Limitations (FP n=24, historical data bias) **Changes:** NEW Section 3.6 (2 pages)
R3.4	Improve Figure 1 description	**Action:** Add detailed caption explaining each pipeline step **Changes:** Figure 1 caption
R3.5	Strengthen Conclusion	**Action:** Add practical implications + clear future directions **Changes:** Section 6 rewrite
REVIEWER 6 - Technical Details		
R6.1	Define "overall" aggregation explicitly	**Action:** State "macro-average (unweighted mean) across LOC/FP/UCP schemas" in Abstract + Results **Changes:** Abstract line 10, Results Section 5.1
R6.2	Fix duplicate Time equation	**Action:** Delete duplicate in Section 2.1 lines 120-130 **Changes:** Section 2.1
R6.3	FP sample size protocol	**Action:** For FP n=24: Use LOOCV, report bootstrap CI, label "exploratory" **Changes:** Methods Section 3.4, Results + caveat
R6.4	Compute R^2 or explain	**Action:** Compute R^2 for all models OR remove column + explain negative R^2 issue **Changes:** All result tables
REVIEWER 7 - Rigor & Reproducibility		
R7.1	Formatting: captions, resolution, line numbers	**Action:** (1) All figures vector PDF 600dpi + captions, (2) Enable line numbers via lineno package **Changes:** All figures + LaTeX preamble

(continued on next page)

(continued from previous page)

Reviewer	Question / Requirement	Response Strategy
R7.2	SOTA models (XGBoost)	**Action:** Add XGBoost as 5th model if time permits **Changes:** Methods + all result tables
R7.3	Interpretability - feature importance	**Action:** Generate Gini or permutation importance plot for RF + 1 paragraph explaining top 3 features **Changes:** NEW Figure + Results Section 5.3
R7.4	Ablation study	**Action:** Run RF with: (raw) — (+log) — (+log+IQR) — (full) → NEW table **Changes:** NEW Table "Ablation Study"
R7.5	Generalization - leave-one-source-out	**Action:** If time: LOSO CV per dataset source **Changes:** NEW subsection or acknowledge limitation
REVIEWER 8 - Deep Technical Review		
R8.1	Data provenance missing	**Action:** Create dataset manifest table: Source — Link/DOI — Schema — Raw# — Removed# — Final# **Changes:** NEW Table in Section 3.1
R8.2	Deduplication criteria unclear	**Action:** Document exact dedup logic: (1) project name exact match, (2) LOC/Effort within 5%, (3) Year match → kept most recent **Changes:** Data Section 3.1
R8.3	Target leakage: Developers	**Action:** REMOVE Developers = ceil(Effort/Time) from feature engineering. Only use Developers if in raw dataset **Changes:** Code + Methods Section 3.2
R8.4	Class imbalance not addressed	**Action:** Explain that effort is continuous regression (not classification), no class imbalance. Mention focal loss paper for future classification work **Changes:** Methods + cite DOI 10.1038/s41598-025-22853-y
R8.5	Hyperparameter search may overfit FP	**Action:** For FP n=24: switch to LOOCV + smaller grid + wider CIs **Changes:** Methods Section 3.4

2.2 Table 2B: LaTeX File Change Locations

Table 4: Where to Fix in main.tex

ID	Issue	File Location	How to Fix
1	"Overall" unclear	Abstract line 10 + Results Section 5.1	Add: "Overall metrics computed as macro-average (unweighted mean) across three schemas"
2	COCOMO uncalibrated	Methods NEW subsection 4.2.1 + Results Table 3	Fit A, B with scipy.optimize on train data. Report original vs calibrated MMRE
3	Target leakage	Methods Section 3.2 line 180-185	DELETE: Developers = ceil(Effort / Time) from feature engineering
4	FP protocol	Methods Section 3.4 line 220-230	Change FP: "For small samples (n=24), we use LOOCV and report bootstrap 95% CI"
5	Dataset manifest	Data Section 3.1 after line 150	INSERT NEW Table 1: Dataset Provenance (6 columns)
6	Formatting	All figures + LaTeX preamble	Export figures as vector PDF 600dpi. Add usepackage{lineno}
7	Equation duplicate	Methods Section 2.1 line 120-130	DELETE duplicate Time = Effort / Developers equation
8	R^2 missing	All result tables	Compute R^2 from predictions. Format: "0.78 [0.65, 0.89]"
9	Novelty weak	Abstract line 8-12 + Intro Section 1.3	Rewrite emphasizing: (1) harmonization, (2) fair baseline, (3) stats rigor
10	Related Work	NEW Section 2 (after Intro)	Write 2-3 pages citing 4 DOIs + comparison table
11	XGBoost missing	Methods Section 4.2 + All result tables	Add XGBoost as 5th model: learning_rate={0.01,0.1,0.3}, max_depth={3,4,6}
12	Interpretability	NEW Figure 4 + Results Section 5.3	Generate feature importance plot (RF Gini) + 1 paragraph
13	Ablation study	NEW Table 5 in Results	Run RF: (raw) — (+log) — (+log+IQR) — (full) → report MMRE
14	Generalization	Methods Section 3.4 OR Limitations	Add LOSO CV OR acknowledge as limitation
15	Language	Entire document	Run Grammarly + remove phrases like "it is worth noting"
16	MAPE/MdMRE/RAE	Methods Section 2.4 + All tables	Add 3 metric formulas + compute in code
17	Confidence intervals	All result tables	Bootstrap 95% CI → format "Mean [95% CI]"
18	Length	Create supplementary.pdf	Move preprocessing algorithm + some figures to supplement

3 TABLE 3: Strategic Decisions for Advisor

3.1 Table 3A: Key Decisions Requiring Approval

Table 5: Strategic Choices

Decision	Options	Recommendation
1. COCOMO Recalibration?	(A) Full implementation (2-3 days) (B) Explain only (0.5 day)	**Option A** - Increases acceptance 15-20%. Proves RF better than EVEN optimized baseline (stronger claim)
2. GitHub/Jira Modern Data?	(A) Collect mini-set (3-4 days) (B) Explain limitation	**Option B** - GitHub lacks good effort ground truth. Honest explanation + Future Work is sufficient
3. Add XGBoost?	(A) Implement (1-2 days) (B) Skip + explain scope	**Option A if time permits** - XGBoost is SOTA, easy to implement, improves score with R4/R7. If no time: Option B + explain scope
4. Ablation Study?	(A) Full (1 day) (B) Simplified	**Option A** - Easy: just run RF with 4 configs (raw/+log/+IQR/full) → 1 table. R5/R7 require it
5. Request Extension?	(A) Request 5 more days (total 15) (B) Work within 10 days	**Option A** - Near Tet holiday, 10 days tight. Request 5 more days (total 15) to do quality work. Journals usually grant reasonable extensions

3.2 Table 3B: Work Division - Huy vs Advisor

Task	Huy (Technical)	Advisor (Strategic)
COCOMO recalibration	Code scipy.optimize	Review results
Dataset manifest	Create table	Verify completeness
Target leakage fix	Remove code	Approve change
FP protocol	Implement LOOCV	Check bootstrap CI
Related Work section	Draft 2-3 pages	Heavy edit + polish
Novelty rewrite	Draft new text	Finalize claims
XGBoost experiment	Code + run	Decide if include
Feature importance	Generate plots	Interpret results
Ablation study	Run 4 configs	Analyze contribution
Response letter	Draft responses	Final approval + sign

Table 6: Huy vs Advisor Responsibilities

4 TABLE 4: Templates - Response Examples

4.1 Template 1: When Agreeing and Implementing

"We thank the reviewer for this excellent suggestion. We agree that [issue] needed clarification. We have now:

- *[Action 1]: [description]*
- *[Action 2]: [description]*

Changes in manuscript: [Section X, lines Y-Z]. The revised text now reads: '[quote new text]'"

We believe this strengthens the paper by [benefit]."

4.2 Template 2: When Explaining Limitation

"We thank the reviewer for this valuable suggestion. We acknowledge that [issue] would strengthen the work. However, [constraint]:

- *[Reason 1]*
- *[Reason 2]*

We have addressed this by:

- *Adding explicit discussion in Limitations section (Section X, lines Y-Z)*
- *Including this as a key future direction in Conclusion*

Changes in manuscript: [location]. We have added: '[quote new text explaining limitation]'"

4.3 Template 3: When Respectfully Disagreeing (USE CAREFULLY)

"We thank the reviewer for this thoughtful comment. We respectfully note that [counter-argument with evidence]:

- *[Evidence/Citation 1]*
- *[Evidence/Citation 2]*

However, to improve clarity, we have:

- *[Clarification added to manuscript]*

Changes in manuscript: [location]"

5 APPENDIX: Key Reminders

5.1 Critical Points - MUST DO

- **"Overall" definition:** Define explicitly (1 sentence in Abstract + 1 paragraph in Results)
- **COCOMO calibration:** Fit A, B on train per schema → compare original vs calibrated vs RF
- **Target leakage:** DELETE Developers = Effort/Time if it's inferred
- **FP n=24 protocol:** LOOCV for FP, report bootstrap CIs, call results "exploratory"
- **Dataset manifest:** Table with 6 columns: Source — Link — Schema — Raw# — After Dedup — Final#
- **Figure captions:** ALL figures must have caption + high quality (vector)
- **Cite 4 DOIs:** R3 suggested 4 papers → must cite + discuss in Related Work + create comparison table
- **R^2 values:** Compute or remove column + explain

5.2 Acceptance Likelihood Estimates

- **If fix CRITICAL (1-6) only:** 60-70% chance
- **If fix CRITICAL + MAJOR (1-10):** 75-85% chance
- **If fix all (1-14):** 85-90% chance

Recommendation: Focus on 1-10 first, do 11-14 if time permits. Can request 5-day extension to do quality work.

5.3 Email to Editor (if requesting extension)

Subject: Request for Extension - Manuscript 6863b9b0-4db8-4b53-843f-5be5e907cf62

Dear Editor,

We have received detailed reviews from 8 reviewers for our manuscript. We are committed to addressing all concerns comprehensively, including:

- *Baseline recalibration for fair comparison*
- *Additional statistical analyses and ablation studies*
- *Enhanced dataset documentation with provenance table*
- *Improved methodological clarity*

Given the extensive revisions requested (8 reviewers with substantial methodological improvements), we respectfully request a 5-day extension (total 15 working days) to ensure rigorous implementation rather than superficial changes.

Thank you for your consideration.

Best regards,

[Authors]

END OF REVISION TABLES

Good luck with the revision! The Major Revision decision is a GOOD sign.
You have a strong chance (75-85%) if you address CRITICAL + MAJOR issues properly.
YOU CAN DO THIS!