Continuous Improvement Backlog Tracker

AI/ML Model Improvement Management System

📖 Quick Start Guide

WORKBOOK OVERVIEW:

This workbook helps you manage continuous improvement for AI/ML models through 5 integrated sheets:

- **Sheet 1: Instructions - This guide**
- **Sheet 2: Backlog Items - Main tracker for all improvement items**
- **Sheet 3: Priority Calculator - Interactive tool for scoring and prioritizing items**
- **Sheet 4: Quarterly Planning - Plan and track quarterly commitments (coming in Session 2)**
- **Sheet 5: Health Metrics - Monitor overall backlog health (coming in Session 2)**

HOW TO USE THIS WORKBOOK:

STEP 1: Add New Improvement Items

→ Go to 'Backlog Items' sheet

→ Add new row with Item ID (e.g., IMP-001)

→ Fill in Title, Description, Type, Owner, etc.

→ Use Priority Calculator to determine priority score

STEP 2: Calculate Priority Scores

→ Go to 'Priority Calculator' sheet

→ Enter scores 1-5 for each of 5 criteria

→ Automatic calculation shows total score and recommended priority

→ Copy results back to Backlog Items sheet

STEP 3: Manage Your Backlog

→ Review backlog monthly (grooming)

→ Update Status as work progresses

→ Archive completed items or move to separate 'Completed' section

→ Track Actual vs. Expected effort and value

STEP 4: Plan Quarterly Work

→ Filter by Priority (P1, P2 items)

→ Select items for upcoming quarter

→ Check total effort vs. team capacity

→ Assign target quarters and owners

KEY FEATURES:

✓ Automated Priority Scoring - Weighted calculation using 5 criteria
✓ Dropdown Validation - Consistent categories and statuses
✓ Conditional Formatting - Visual priority and status indicators
✓ Formula-Driven - Automatic calculations for scores and dates
✓ Flexible Filtering - Sort by priority, type, status, quarter

PRIORITY LEVELS:

- **P1 - Critical: Must do this quarter (Score ≥4.5)**
- **P2 - High: Should do if capacity (Score 3.5-4.49)**
- **P3 - Medium: Plan for next quarter (Score 2.5-3.49)**
- **P4 - Low: Future backlog (Score <2.5)**

IMPROVEMENT TYPES:

- **Performance Enhancement - Accuracy, speed, quality improvements**
- **New Feature/Capability - New functionality or use cases**
- **Cost Optimization - Reduce infrastructure or operational costs**
- **Technical Debt - Code quality, upgrades, refactoring**
- **Operational Improvement - Monitoring, deployment, processes**
- **User Experience - Usability, explainability, integration**
- **Scalability - Handle higher volume or new markets**
- **Compliance/Governance - Regulatory, ethical, documentation**

EFFORT SIZING:

- **XS (1-3 days) - Quick wins**
- **S (1 week) - Small projects**
- **M (2-4 weeks) - Medium projects**
- **L (1-2 months) - Large initiatives**
- **XL (2-4 months) - Major projects**
- **XXL (4+ months) - Epic initiatives**

STATUS VALUES:

- **Open - Not yet started**
- **In Progress - Active work ongoing**
- **Blocked - Cannot proceed (document blocker)**
- **Complete - Finished with outcomes documented**
- **Deferred - Deprioritized, not planned**
- **Cancelled - No longer relevant**

BEST PRACTICES:

1. Add all improvement ideas immediately - don't self-filter
2. Groom backlog monthly - update priorities and status
3. Use Priority Calculator consistently - objective scoring
4. Balance workload - 70% features, 20% tech debt, 10% innovation
5. Document outcomes - capture actual vs. expected results
6. Keep backlog fresh - archive items >6 months old not prioritized
7. Link to sources - reference performance reviews, A/B tests, incidents
8. Track velocity - monitor items completed per quarter
9. Celebrate wins - share completed items and value delivered
10. Review quarterly - align backlog with updated OKRs and strategy

COMMON PITFALLS TO AVOID:

✗ Everything marked P1 (use scoring to prioritize ruthlessly)
✗ Backlog grows indefinitely (archive stale items)
✗ No outcome documentation (capture learnings)
✗ Ignoring technical debt (allocate 20% capacity)
✗ Optimistic estimates (add 20-30% buffer)
✗ Bypassing backlog for urgent items (maintain as source of truth)

SUPPORT & CUSTOMIZATION:

- **Customize Types/Categories for your organization**
- **Add custom fields as needed (budget codes, teams, etc.)**
- **Adjust priority scoring weights based on your priorities**
- **Create views/filters for different stakeholder needs**
- **Export data for presentations or executive reports**

VERSION INFORMATION:

- **Template Version: 1.0**

- **Created: January 2026**
- **Based on: BABOK, MLOps, Agile best practices**

💡 TIP: Start small (10-15 items), grow over time as team adopts practice.

💡 TIP: Use this as living document - update weekly, review monthly, plan quarterly.

Item ID	Title	Description	Type	Priority	Priority Score	Status	Estimated Effort	Expected Value ($)	Owner	Target Quarter	Date Added	Last Updated	Source	Dependencies	Acceptance Criteria	Actual Effort	Actual Value ($)	Notes
IMP-001	Add behavioral velocity features to fraud model	Add transaction velocity features (transactions per hour, spending velocity, account age at first transaction) to capture behavioral fraud patterns. Analysis shows 15% of missed frauds exhibit high-velocity patterns these features would capture.	Performance Enhancement	P2	4.10	Open	S (1 week)	$150,000	Maria Rodriguez	Q2 2024	2024-03-10	2024-03-20	Performance Review	Blocks: IMP-015 (Real-time velocity scoring)	1. Five velocity features implemented (tx_per_hour, spend_velocity, account_age_at_tx, days_since_last_tx, tx_count_24h) 2. A/B test shows ≥3% recall improvement (p<0.05) 3. Precision maintained (no significant decrease) 4. Alert volume increase <10%			Feature engineering documented. Fraud team strongly supports. Targeting April completion.
IMP-002	Apply INT8 quantization to reduce latency	Apply INT8 quantization to fraud detection model to reduce inference latency from 120ms P95 to target <50ms. Analysis shows 100ms latency reduction translates to 2% conversion improvement worth $200K/year.	Performance Enhancement	P3	3.25	Open	M (2 weeks)	$200,000	Alex Kim	Q3 2024	2024-04-01	2024-04-05	User Feedback		1. INT8 quantized model deployed to production 2. P95 latency ≤50ms (validated over 1 week) 3. Accuracy degradation ≤1% (F1 score within 1% of FP32) 4. A/B test validates no negative fraud detection impact			Preliminary testing shows INT8 maintains 99.2% of FP32 accuracy. Product team strongly interested.
IMP-003	Build automated retraining pipeline	Automate model retraining to reduce cycle time from 2 weeks to 2 days and eliminate manual steps. Currently: data extraction, feature engineering, training, validation, deployment all require manual ML engineer intervention (10 hours/month).	Operational Improvement	P1	3.55	In Progress	L (6 weeks)	$48,000	Sarah Chen	Q1-Q2 2024	2024-01-15	2024-03-15	Technical Debt	Blocked by: None | Blocks: IMP-020 (Multi-model platform)	1. Automated pipeline executes weekly retraining 2. Human approval only for production deployment 3. Comprehensive validation checks before approval 4. Execution time ≤48 hours trigger to approval 5. Error handling and notifications implemented			Started Q1. Data pipeline complete (2 weeks). Training automation 80% complete. Infrastructure team approved deployment automation. On track for Q2.
IMP-004	Implement SHAP explanations for analysts	Add SHAP (SHapley Additive exPlanations) showing top 5 contributing features in fraud analyst UI. Currently analysts only see risk score, making triage difficult. Explanations expected to reduce manual review time 20%.	User Experience	P2	3.10	Open	M (3 weeks)	$75,000	Unassigned	Q3 2024	2024-03-20	2024-03-22	User Feedback	Requires: UI team support (2-3 days)	1. SHAP explanations generated for all predictions 2. Top 5 contributing features displayed in UI 3. Explanation latency <100ms (no workflow slowdown) 4. Analyst testing validates usefulness 5. Documentation explains SHAP interpretation			Fraud team very interested, willing to participate in testing. SHAP validated on test data, performance acceptable.
IMP-005	Upgrade TensorFlow from 2.8 to 2.15	Upgrade TensorFlow from 2.8 (Feb 2022, now 2 years old, unsupported) to 2.15 (latest stable). Provides: security patches, performance improvements (~10% inference speed), bug fixes, and maintains supported software compliance.	Technical Debt	P2	3.05	Open	S (1 week)	$0	Alex Kim	Q2 2024	2024-03-25	2024-03-26	Technical Debt	Enables: IMP-017 (Explore new TF 2.15+ features)	1. TensorFlow upgraded to 2.15.x in all environments 2. All unit tests pass 3. Model predictions bit-exact match 2.8 (validation set) 4. Performance metrics within 0.1% (accuracy,			Tested in dev environment, models load successfully. Security team requesting completion by end Q2 for compliance.
IMP-006	Extend fraud model to international markets	Expand fraud detection model to handle international (non-US) transactions for 50+ countries. Requires: currency conversion features, international merchant category mapping, country-specific fraud patterns. Enables $50M+ international market expansion.	New Feature/Capability	P4	3.25	Open	XXL (4-6 months)	$50,000,000	Unassigned	TBD	2024-02-15	2024-03-15	Strategic Initiative	Blocked by: Product international launch decision | Blocks: IMP-023 (Country-specific optimization)	1. Model supports target countries (CA, UK, AU, DE, FR initially) 2. Currency conversion features implemented 3. International merchant category mapping complete 4. Recall ≥80% on international test set			International launch under discussion, targeting Q4 2024 for CA/UK. Product will confirm by end Q2. Marked P4 until launch confirmed, then will become P1.

PRIORITY SCORE CALCULATOR

Interactive tool for objective, data-driven prioritization

HOW TO USE THIS CALCULATOR

1. Enter the Item ID and Title from your backlog item
2. Score each of the 5 criteria from 1 (lowest) to 5 (highest) using the guidance provided
3. The calculator automatically computes the weighted priority score
4. Use the final score to assign priority: P1 (≥4.5), P2 (3.5-4.49), P3 (2.5-3.49), P4 (<2.5)
5. Copy the Priority Score and Priority Level back to the Backlog Items sheet

TIP: Use this calculator consistently for all items to ensure objective, comparable prioritization

ITEM DETAILS

Item ID: | **Item Title:**

SCORING CRITERIA

Criterion	Weight	Your Score (1-5)	Weighted Score	Scoring Guidance
Business Value	0.35		**0.00**	5: >$500K/year or >10% improvement | 4: $100K-$500K or 5-10% | 3: $25K-$100K or 2-5% | 2: $5K-$25K or 1-2% | 1: <$5K or <1%
Effort (Lower = Higher Score)	0.25		**0.00**	5: XS (1-3 days) | 4: S (1 week) | 3: M (2-4 weeks) | 2: L (1-2 months) | 1: XL/XXL (2+ months)
Strategic Alignment	0.20		**0.00**	5: Critical for strategic OKR | 4: Strong alignment with strategy | 3: Moderate alignment | 2: Weak alignment | 1: No strategic relevance
Urgency / Risk	0.15		**0.00**	5: Critical issue causing problems now | 4: High risk if not addressed this quarter | 3: Moderate risk, address within 6 months | 2: Low risk, 6-12 months acceptable | 1: No time pressure
Feasibility	0.05		**0.00**	5: Proven approach, low risk | 4: Well-understood, moderate confidence | 3: Some unknowns, needs investigation | 2: Significant unknowns, high risk | 1: Experimental, very uncertain

TOTAL PRIORITY SCORE	0.00	*(out of 5.00)*

RECOMMENDED PRIORITY	P4 - Low

PRIORITY ASSIGNMENT RANGES		
P1 - Critical	Score ≥ 4.5	Must do this quarter - highest business value, strategic importance, or urgent need
P2 - High	Score 3.5 - 4.49	Should do this quarter if capacity available - significant value, good alignment
P3 - Medium	Score 2.5 - 3.49	Plan for next quarter - moderate value, can wait 3-6 months
P4 - Low	Score < 2.5	Future backlog - low value or high effort, revisit in future planning

EXAMPLE CALCULATION

```
EXAMPLE: "Add behavioral velocity features to fraud model"

Business Value: 4 (Expected $150K/year from 4% recall improvement)
  → 4 × 0.35 = 1.40

Effort: 4 (S - 1 week, ~40 hours - relatively quick win)
  → 4 × 0.25 = 1.00

Strategic Alignment: 5 (Critical for FY24 fraud reduction OKR)
  → 5 × 0.20 = 1.00

Urgency: 3 (Moderate - fraud rates acceptable but could improve)
  → 3 × 0.15 = 0.45

Feasibility: 5 (Proven features, low technical risk)
  → 5 × 0.05 = 0.25
```

QUARTERLY PLANNING DASHBOARD

Capacity planning, commitment tracking, and velocity analysis

CURRENT QUARTER OVERVIEW

Current Quarter: Q2 2024 **Team Capacity (** 480

CAPACITY PLANNING

Priority	# Items	Est. Hours	% of Total	Status
P1 - Critical	1	240	50%	In Progress
P2 - High	3	200	42%	Planned
P3 - Medium	0	0	0%	Not Planned
P4 - Low	0	0	0%	Not Planned
TOTAL COMMITTED	**4**	**440**	**92%**	
AVAILABLE CAPACITY		**40**	**8%**	

Capacity Utilization: ✗ **Over-committed**

ITEMS BY QUARTER

Quarter	P1 Items	P2 Items	P3 Items	P4 Items	Total Items	Est. Hours
Q1 2024	0	0	0	0	**0**	0
Q2 2024	1	3	0	0	**4**	440
Q3 2024	0	2	1	0	**3**	280
Q4 2024	0	0	0	1	**1**	400
TBD	0	0	0	0	**0**	0

IMPROVEMENT TYPE BALANCE (70-20-10 RULE)

Type	# Items	% Actual	% Target	Status
Performance & Features	4	67%	70%	✓ Balanced
Technical Debt	1	17%	20%	✓ Balanced
Innovation/Research	1	17%	10%	✓ Balanced

VELOCITY TRACKING (Items Completed per Quarter)				
Quarter	**Items Committed**	**Items Completed**	**Completion Rate**	**Velocity**
Q4 2023	5	4	80%	4
Q1 2024	6	5	83%	5
Q2 2024 (Current)	4	1	25%	In Progress
Q3 2024 (Planned)	3	0		
Average Velocity (last 2 quarters		**4.5**		

BACKLOG HEALTH METRICS DASHBOARD

Monitor overall backlog health, age, completion rates, and value delivered

OVERALL BACKLOG HEALTH

Health Score:	✓ HEALTHY	**Last Updated:**	2024-03-20

KEY METRICS SUMMARY

Total Active Items	**6**	*Items currently in backlog*
Items In Progress	**1**	*Active work ongoing*
Completion Rate (Q2)	**25%**	*1 of 4 committed items complete*
Average Item Age	**32 days**	*Time since items were added*
Stale Items (>90 days)	**0**	*Items not updated in 90+ days*

ITEMS BY STATUS

Status	Count	% of Total	Health Indicator
Open	5	83%	✓ Normal
In Progress	1	17%	✓ Normal
Blocked	0	0%	✓ No Blockers
Complete	0	0%	Add to completed section
Deferred	0	0%	✓ No Deferred
Cancelled	0	0%	✓ No Cancelled

PRIORITY DISTRIBUTION

Priority	Count	% of Total	Target Range	Status
P1 - Critical	1	17%	10-15%	⚠ Too Many P1
P2 - High	3	50%	20-25%	⚠ Review
P3 - Medium	1	17%	30-35%	Review Distribution
P4 - Low	1	17%	30-40%	Review Distribution

IMPROVEMENT TYPE DISTRIBUTION

Type	Count	% of Total
Performance Enhancement	2	33%
New Feature/Capability	1	17%
Cost Optimization	0	0%
Technical Debt	1	17%
Operational Improvement	1	17%
User Experience	1	17%
Scalability	0	0%
Compliance/Governance	0	0%

ITEM AGE ANALYSIS

Age Range	Count	% of Total	Health Indicator
< 30 days (Fresh)	3	50%	✓ Recent additions
30-60 days	2	33%	✓ Normal age
60-90 days	1	17%	⚠ Getting old
> 90 days (Stale)	0	0%	✓ No stale items

VALUE DELIVERED (COMPLETED ITEMS)

Quarter	Items Completed	Expected Value	Actual Value
Q4 2023	4	$425,000	$380,000
Q1 2024	5	$550,000	$520,000
Q2 2024 (To Date)	1	$150,000	$145,000
Total	**10**	**$1,125,000**	**$1,045,000**

Value Realization Rate: **93%** *93% of expected value delivered*

HEALTH INDICATORS

Priority Distribution	**✓ Healthy**	*P1/P2/P3/P4 distribution within normal ranges*

Type Balance	**✓ Balanced**	*70-20-10 rule maintained (66% features, 17% debt, 17% innovation)*
Capacity Utilization	**✓ Optimal**	*92% capacity committed for Q2 (good utilization)*
Staleness	**✓ Good**	*No items >90 days old without updates*
Blocked Items	**✓ Excellent**	*Zero blocked items*
Value Realization	**✓ Strong**	*93% value realization rate on completed items*