CHECKLIST

Learning Best Practices for MSPs Checklist

A practical checklist for deciding whether an AI or machine learning system is ready for MSP operations, with prerequisites, validation steps, rollout controls, and rollback boundaries.

Learning Best Practices for MSPs: Operational Readiness Checklist

Use this checklist to decide whether an AI or machine learning system is ready for MSP operations. It is designed for ticket triage, alert grouping, recommendation systems, and other learning workflows that affect service delivery.

---

1) Scope and boundary

- [ ] The use case is clearly defined.
- [ ] Ticket triage
- [ ] Alert grouping
- [ ] Recommendation generation
- [ ] Other: _____
- [ ] The system?s learning input sources are documented.
- [ ] The system?s allowed outputs are documented.
- [ ] The system?s non-negotiable limits are documented.
- [ ] The boundary states what the system may never do automatically.

- [ ] A reviewer can tell whether each action is:
- [ ] Autonomous
- [ ] Advisory
- [ ] Blocked

Boundary statement template

> The system may learn from _____, may produce _____, and must never automatically _____ without _____ approval.

---

2) Operational value

- [ ] One primary business metric is selected.
- [ ] One safety metric is selected.
- [ ] Both metrics are tied to real service operations.
- [ ] Baseline values are recorded.
- [ ] Target direction is defined.
- [ ] Review interval is set.

Common primary metrics

- [ ] Time to first useful action
- [ ] Recommendation acceptance rate
- [ ] Duplicate alert reduction
- [ ] Rework rate after remediation suggestion
- [ ] Time to correct queue classification

Common safety metrics

- [ ] High-severity misroute rate
- [ ] False suppression rate
- [ ] Manual override rate
- [ ] Escalation error rate

- [ ] Customer-impacting mistake rate

Metric definition template

Metric type	Name	Baseline	Target	Review interval
Primary	_____	_____	_____	_____
Safety	_____	_____	_____	_____

---

3) Prerequisites

Confirm the following before proceeding:

- [ ] A defined operational use case exists.
- [ ] Historical data is available.
- [ ] The data includes known-good outcomes.
- [ ] A baseline process exists.
- [ ] Inputs, predictions, human decisions, and outcomes can be logged.
- [ ] A rollback path exists.
- [ ] An accountable owner is assigned.
- [ ] Residual risk acceptance is documented.
- [ ] Exception review ownership is assigned.

Prerequisite notes

- Use historical data that reflects real operational conditions.
- Include both common cases and rare, high-impact cases.
- Do not rely only on easy examples.

---

4) Baseline review

- [ ] A representative sample of cases has been collected.
- [ ] Raw inputs or triggers are recorded.
- [ ] Human actions are recorded.
- [ ] Actual outcomes are recorded.
- [ ] Escalations and exceptions are recorded.
- [ ] Time required per case is recorded.
- [ ] Common and edge cases are both included.
- [ ] The sample is large enough to reflect expected production use.

Baseline report template

Case type	Input/trigger	Human decision	Outcome	Exception/escalation	Time to complete
Routine	_____	_____	_____	_____	_____
Ambiguous	_____	_____	_____	_____	_____
High-impact edge case	_____	_____	_____	_____	_____

---

5) Historical validation

- [ ] The workflow has been tested on historical data before production use.
- [ ] Routine case performance is reviewed.
- [ ] Ambiguous case performance is reviewed.
- [ ] Outlier handling is reviewed.
- [ ] Stability across teams or data sources is reviewed.
- [ ] Failure patterns are identified.
- [ ] Wrong-answer consequences are reviewed.
- [ ] Safety tradeoffs are explicitly considered.

Validation questions

- [ ] Where does the system help most?

- [ ] Where is it uncertain?
- [ ] Where should it remain advisory only?
- [ ] Which errors are acceptable, and which are not?

Validation summary template

> The system performs well on _____. It is uncertain on _____. It should remain advisory only for _____. The highest-risk failure mode is _____.

---

6) Learning over time

- [ ] Multiple review windows were compared.
- [ ] Performance improved in the intended direction.
- [ ] Manual correction rate decreased.
- [ ] High-confidence wrong suggestions decreased.
- [ ] Performance stayed stable after data changes.
- [ ] Severe error categories did not increase.
- [ ] If performance changed without operational improvement, learning was frozen for review.
- [ ] Data pipeline quality and label quality were checked if drift appeared.

Review rule

If the system changes but the operational result does not improve, pause retraining, inspect feature drift, and verify label quality before continuing.

---

7) Production rollout controls

- [ ] The rollout starts in shadow mode.
- [ ] Predictions are logged without affecting operations at first.

- [ ] A manual review sample is defined.
- [ ] Advisory mode is enabled only for low-risk cases initially.
- [ ] Expansion happens only after repeated validation.
- [ ] A manual override exists for high-severity workflows.
- [ ] Stop conditions are defined.
- [ ] Owners for each rollout phase are assigned.

Safe rollout phases

- [ ] Shadow mode
- [ ] Manual review sample
- [ ] Advisory mode for low-risk cases
- [ ] Expanded advisory use after validation
- [ ] Limited autonomy only if separately approved

Stop condition examples

- [ ] Misclassifications exceed the agreed threshold
- [ ] Data quality drops below the acceptable level
- [ ] Engineers override most outputs
- [ ] Severe incidents increase
- [ ] Unexpected behavior appears in critical workflows

---

8) Rollback and cleanup

- [ ] Automated decisions can be turned off quickly.
- [ ] The baseline workflow can be restored.
- [ ] Logs are preserved for incident review.
- [ ] Bad training data can be frozen or excluded.
- [ ] Retraining on corrupted or biased samples is prevented.
- [ ] The rollback runbook is documented.

- [ ] Version identifiers are recorded.
- [ ] Data snapshots are retained.
- [ ] Approval records are retained.

Rollback runbook template

- [ ] Disable model-driven actions.
- [ ] Revert to the baseline workflow.
- [ ] Confirm core service delivery is unaffected.
- [ ] Preserve operational logs and evidence.
- [ ] Freeze affected training data and artifacts.
- [ ] Review the incident and identify the failure mode.
- [ ] Approve retraining or cleanup only after review.

---

9) Governance and access control

- [ ] An owner is accountable for production use.
- [ ] Approval authority is documented.
- [ ] Access to training data is restricted.
- [ ] Access to configuration and model changes is restricted.
- [ ] Changes are logged with timestamps and reviewers.
- [ ] Exception handling is documented.
- [ ] Compliance or customer obligations are considered.
- [ ] Retention rules are defined for logs and snapshots.

Governance checklist prompt

Ask:

- [ ] Who can approve production use?
- [ ] Who can change the model or pipeline?
- [ ] Who can override a model recommendation?

- [ ] Who reviews exceptions and incidents?

---

10) Go-live readiness decision

Mark the result for each item:

- [ ] Scope is narrow and documented.
- [ ] Metrics are defined and measurable.
- [ ] Baseline evidence exists.
- [ ] Historical validation is complete.
- [ ] Learning behavior is stable over time.
- [ ] Rollout is staged.
- [ ] Rollback is tested and documented.
- [ ] Governance controls are in place.

Decision

- [ ] Ready for shadow mode
- [ ] Ready for advisory mode only
- [ ] Ready for limited autonomy with approval controls
- [ ] Not ready for production use

Final decision note

> Decision: _____ > > Conditions required before next step: _____ > > Reviewer: _____ > > Date: _____

Quick rule of thumb

A learning system is ready for MSP operations only when it has:

- a narrow boundary,

- measurable operational value,
- historical validation,
- visible learning behavior,
- staged rollout controls,
- and a rollback path that can be executed under incident pressure.

If any of those are missing, keep the system advisory only or block production use until the gap is closed.