PUBLIC-SAFE PACKET | PROOF BEFORE PRODUCTION | FOLIUM SYSTEMS

AI RISK AND LAUNCH STANDARD

Folium AI Risk Launch Standard

This packet is meant to be printed only when the paper is worth it. It gives owners, operators, technical reviewers, and leadership a practical launch standard for AI workflows that need to become useful without becoming uncontrolled.

AUDIENCE	PURPOSE	UPDATED
Owners, operators, IT, security, staff leaders, executive sponsors	Define the gates that make AI work safer, faster, and easier to operate	May 2026

- The right launch standard protects speed by making risk visible early.
- AI evaluation must test workflow behavior, source grounding, permissions, and user journeys.
- A proof should not become a dependency until owners, blockers, support, and rollback are clear.

01

LAUNCH THESIS

Fast AI work needs better stop signs, not slower ambition.

The Folium AI Risk and Launch Standard is built for businesses that want to move quickly without letting AI become a mystery dependency. The standard names the gates that protect speed: govern, map, measure, manage, monitor.

EVIDENCE | BOUNDARY | ACTION

GOVERN

Name authority before use

Owners, permissions, review points, live-action limits, escalation, blocked actions, and decision rights.

MAP

Draw the workflow before automating

Systems, users, data sources, provider handoffs, runtime placement, failure modes, and dependencies.

MEASURE

Test behavior, not only presentation

Evaluate retrieval, answer quality, tool routing, browser paths, refusals, latency, accessibility, and known limits.

MANAGE

Operate after launch

Support, incidents, rollback, release notes, training, source freshness, monitoring, and improvement loops.

02

RISK REGISTER

The risks that matter most are operational.

Most AI risk discussions focus only on model behavior. Folium expands the view to the full operating system around the model.

DECISION GRID | REVIEW LENS | NEXT GATE

RISK	HOW IT APPEARS	LAUNCH CONTROL
Wrong answer	AI gives inaccurate or stale information with confidence.	Source-grounding checks, eval cases, refusal rules, review path.
Wrong action	AI updates, sends, routes, or triggers something it should not.	Permission table, human gate, blocked actions, audit trail.
Wrong data	Private, regulated, or secret data enters prompts, logs, or external tools.	Data boundary map, redaction, retention rules, secrets custody.
Wrong runtime	Sensitive work is placed in a runtime that does not fit privacy, latency, cost, or control needs.	Runtime placement decision, fallback, portability, vendor-exit plan.
Wrong owner	No one owns quality, support, rollback, source freshness, or staff adoption.	Owner map, support model, incident route, training packet.
Wrong launch	A proof becomes a production dependency without evidence.	Launch blockers, go/no-go gate, known-limits record, pilot criteria.

03

LAUNCH GATES

A workflow should earn each next level of authority.

Folium uses launch gates to decide when a workflow is safe to inspect, safe to sandbox, safe to pilot, or ready for production planning.

DECISION GRID | REVIEW LENS | NEXT GATE

GATE	MINIMUM EVIDENCE	DECISION
Gate 1: Public proof	Public-safe page, packet, screenshot, or workflow sketch.	Continue only if the value is clear enough to scope.
Gate 2: Scoped workflow	Business process, users, systems, data classes, owners, and exclusions.	Continue only if the first proof is narrow and safe.
Gate 3: Sandbox behavior	Clickable flow, redacted sources, eval cases, known limits, staff review.	Continue only if behavior is useful and inspectable.
Gate 4: Architecture review	Runtime map, permissions, logs, secrets, support, fallback, rollback.	Continue only if reviewers can approve pilot conditions.
Gate 5: Controlled pilot	Limited users, monitored use, training, incident path, repair cadence.	Continue only if evidence supports expanded dependency.
Gate 6: AI operations	Ongoing monitoring, cost control, source maintenance, change review, adoption metrics.	Operate only with clear ownership and improvement rhythm.

04

EVALUATION DISCIPLINE

Evaluation should test the job the workflow is supposed to do.

A useful AI launch standard measures whether the system helps the business perform the task safely, not whether a single answer sounds polished.

EVIDENCE | BOUNDARY | ACTION

Behavior cases

Realistic prompts, messy user language, edge cases, bad inputs, missing data, and role-specific questions.

- Happy path
- Messy path
- Boundary path
- Refusal path

Source and retrieval checks

Verify that answers come from approved sources, stale sources are flagged, and unsupported claims are avoided.

- Approved sources
- Freshness
- Citation need
- Unsupported claim handling

Tool and route checks

Confirm that AI chooses the right tool, refuses blocked tools, escalates sensitive actions, and does not invent authority.

- Allowed tool
- Blocked tool
- Human gate
- Escalation

User journey proof

Check the actual user path in browsers and devices, including mobile, tablet, desktop, forms, downloads, and visible states.

- Desktop
- Tablet
- Mobile
- Download and print

05

LAUNCH BLOCKERS

These failures should stop the launch.

Blockers are not paperwork. They protect the customer from turning a promising proof into a brittle dependency.

CHECKLIST	OWNER PATH	RELEASE SIGNAL

- The AI claims it can perform live actions that are not approved.
- Private data, secret labels, internal source names, or credentials leak into public output.
- The system cannot explain what source supports a factual answer.
- No owner exists for source freshness, support, rollback, incident response, or launch signoff.
- Staff cannot explain what the AI is allowed to do and when to escalate.
- The workflow affects money, customers, access, compliance, or reputation without a human review gate.
- The buyer cannot see known limits, failed cases, repair decisions, or acceptance criteria.
- The launch path has no degraded mode if a model, retrieval source, API, database, or provider fails.

06

HUMAN ADOPTION

AI launch risk includes people.

A workflow that is technically impressive can still fail if staff do not understand it, trust it, correct it, or know when to override it.

EVIDENCE | BOUNDARY | ACTION

Role clarity

Staff should know which part of the job AI supports and which part remains human responsibility.

Training packet

Users need simple examples, limits, escalation steps, and what-good-looks-like guidance.

Feedback loop

Staff need a way to report wrong answers, missing sources, confusing routes, and adoption friction.

Manager visibility

Leaders need operational signals: usage, quality, savings, errors, support load, and readiness to expand.

Job strengthening

Folium frames AI as capacity expansion: reduce repetitive work, preserve human judgment, and strengthen staff capability.

Fear reduction

People fear what they cannot see. The launch standard makes the workflow, limits, and support path visible.

07

OPERATING CADENCE

After launch, the system still needs care.

AI work changes after release: sources age, staff learn, costs move, edge cases appear, tools update, and customers ask new questions. The standard defines how the system stays healthy.

DECISION GRID | REVIEW LENS | NEXT GATE

CADENCE	REVIEW FOCUS	OUTPUT
Weekly early pilot	Usage, friction, wrong answers, failed routes, support tickets, staff comments.	Repair list, training note, source update, or gate decision.
Monthly operations	Quality trend, cost trend, source freshness, incidents, adoption, role changes.	Improvement backlog and release note.
Quarterly executive review	Business value, risk posture, expansion candidates, vendor/runtime fit, staffing impact.	Continue, expand, refactor, retire, or redesign decision.
Incident-triggered	Unsafe output, wrong action, data issue, provider failure, user harm, regulatory concern.	Rollback, communication, repair, root cause, and relaunch gate.

08

NEXT STEP

Launch discipline is how Folium moves fast without pretending risk disappeared.

Use this standard to decide whether one AI workflow is ready for a scoped proof, a sandbox, a pilot, or a production plan. The answer should come from evidence, not excitement.

Bring the workflow

Name the business process, the systems involved, the people affected, and the decision this packet should support.

Separate proof from production

Keep public proof, sandbox review, pilot access, and production dependency in separate gates with clear owners.

Ask for the evidence

Request screenshots, browser checks, known limits, launch blockers, support plans, and the next approval path.