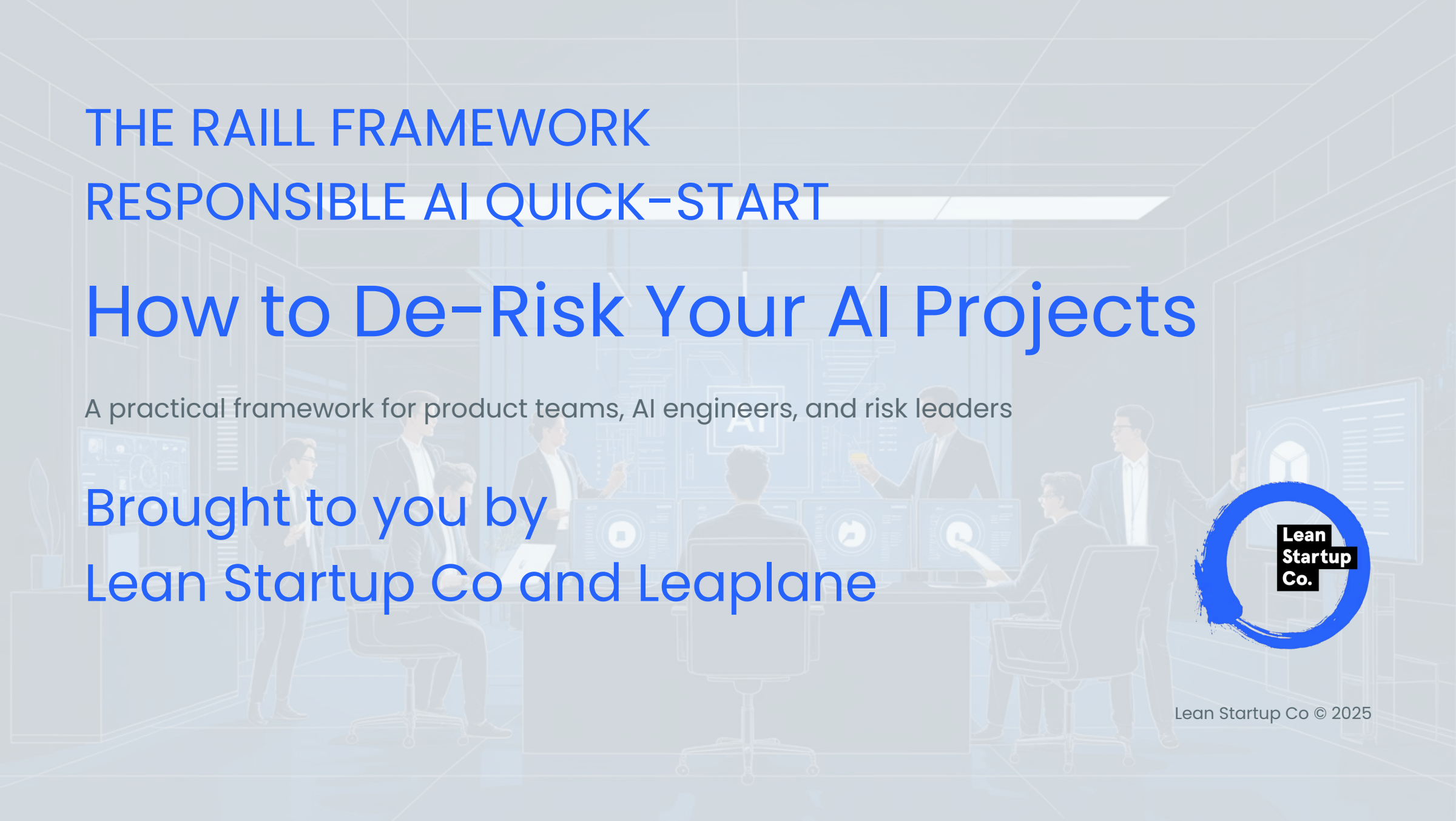
THE RAILL FRAMEWORK
RESPONSIBLE AI QUICK-START

How to De-Risk Your AI Projects

A practical framework for product teams, AI engineers, and risk leaders



Brought to you by
Lean Startup Co and Leaplane

Lean Startup Co © 2025

THE PROBLEM

The EU AI Act takes effect starting April 2025. Companies are scrambling to retrofit compliance into existing AI systems.

Most teams know they need "responsible AI" but don't know where to start. They're building AI faster than they're building guardrails.

The solution: Start with questions, not solutions.

Lean Startup Co © 2025

HOW TO USE THIS GUIDE

01

Work through each dimension with your team

30-60 minutes

02

Check the box for each question you can answer with confidence

03

Score your results

- ✓ 30+ boxes checked = Strong foundation
- ✓ 16-28 boxes checked = Significant gaps
- ✓ <16 boxes checked = Critical exposure

Use this as a self-assessment, sprint planning tool, or vendor evaluation criteria.

Lean Startup Co © 2025

THE 9 DIMENSIONS OF RESPONSIBLE AI

1	PRIVACY *Key question: Can you trace consent from data collection to model output?*
2	SECURITY *Key question: Have you tested for prompt injection and adversarial attacks?*
3	TRANSPARENCY *Key question: Can you explain how the model makes decisions to non-technical stakeholders?*
4	FAIRNESS *Key question: Have you tested performance across demographic groups?*
5	INCLUSION *Key question: Does your evaluation dataset reflect your actual user base?*
6	SAFETY *Key question: Have you stress-tested for harmful outputs?*
7	ROBUSTNESS *Key question: Can you predict how the model behaves under stress?*
8	EXPLAINABILITY *Key question: Can you explain decisions to people affected by them?*
9	GOVERNANCE Key question: Can you answer "who owns this?" in 5 seconds?

Lean Startup Co © 2025

THE CHECKLIST

Each dimension references the External Loop (Product) and the Internal Loop (Technical) questions

PRIVACY

External Loop (Product):

- □ What user data is truly necessary?
- □ What are the user's privacy expectations?
- □ What would cause user backlash or violate regulations?

Internal Loop (Technical):

- □ Is consent metadata tracked throughout the data pipeline?
- □ What sensitive data is used in training?
- □ Can the model leak or infer private information?

SECURITY

External Loop (Product):

- □ Who might maliciously misuse this feature?
- □ What is the worst-case scenario for misuse?
- □ What safety guidance do users need?

Internal Loop (Technical):

- □ Are there vulnerable endpoints or APIs?
- □ What is the blast radius of a breach?
- □ Are automated safeguards and blocking effective?

TRANSPARENCY

External Loop (Product):

- □ What do users believe the system does?
- □ Which processes or outcomes require visibility?
- □ What documentation is essential?

Internal Loop (Technical):

- □ Do the logs and metrics reflect the model's true behavior?
- □ Can we trace and audit key decisions?
- □ Are model cards auto-generated and updated?

FAIRNESS

External Loop (Product):

- □ Which user groups are most reliant on accurate outcomes?
- □ What are the potential signals of unequal treatment?
- □ What fairness guarantees must we provide?

Internal Loop (Technical):

- □ Are protected demographics missing from the data?
- □ Do metrics reveal performance gaps across groups?
- □ Are debiasing techniques applied and measured?

INCLUSION

External Loop (Product):

- □ Who is unable to use the system effectively?
- □ In what contexts or on which devices is the user experience poor?
- □ Is our feedback loop diverse and representative?

Internal Loop (Technical):

- □ What input patterns or languages confuse the model?
- □ Is there performance degradation for low-resource users?
- □ Are evaluation datasets sufficiently diverse?

SAFETY

External Loop (Product):

- □ What is the potential for physical, psychological, or social harm?
- □ How is the user experience designed with safety boundaries?
- □ How do we communicate safe and unsafe use cases?

Internal Loop (Technical):

- □ Does the model generate harmful, unethical, or unsafe outputs?
- □ How effective are the guardrails at blocking harmful content?
- □ Does the model understand context to avoid dangerous advice?

ROBUSTNESS

External Loop (Product):

- □ Under what conditions does the system fail or behave unpredictably?
- □ How does it perform in rare or extreme scenarios?
- □ What are the stability expectations for enterprise clients?

Internal Loop (Technical):

- □ Where are the edge-case failure points?
- □ How sensitive is the model to data or concept drift?
- □ Is there performance stability across model versions?

EXPLAINABILITY

External Loop (Product):

- □ What must the end-user understand about the system's output?
- □ Which specific decisions require a "why"?
- □ How simple must the explanations be for the audience?

Internal Loop (Technical):

- □ Can we identify which features drive the model's decisions?
- □ Can we retrieve attributions for specific outputs?
- □ Are explanations presented in a usable format?

GOVERNANCE & CONTROLLABILITY

External Loop (Product):

- □ Who is ultimately accountable for model outcomes?
- □ What are the required approval gates before launch?
- □ What is the recovery and rollback plan when failures occur?

Internal Loop (Technical):

- □ Is there clear ownership of the data, model, and pipeline?
- □ Are there automated risk and quality checks?
- □ Is there robust versioning and a reliable rollback mechanism?

Lean Startup Co © 2025

YOUR SCORE:

__ /54

30+ checked

Strong foundation - focus on continuous improvement

<28 checked

Significant gaps - prioritize high-risk dimensions

<8 checked

Critical exposure - need immediate action plan

Lean Startup Co © 2025

WHAT HAPPENS NEXT?

Most teams score between 10-24 on their first pass. That's normal—and fixable.

Three ways forward:

1

DIY

Use this as your roadmap and tackle gaps systematically

2

Workshop

2-hour facilitated session to align your team

3

Full Assessment

2-3 week deep-dive with roadmap

Lean Startup Co © 2025

Next Step

Let's jump on a 30-minute call to discuss your current projects and confirm your action plan.

Jonathan Bertfield, CEO

jonathan@leanstartup.co
(US)+1 347 495 2462

(EU)+30 698 033 3384

Lean Startup Co., LLC
440 N Barranca Ave, 2197
Covina, CA 91723

Lean Startup Co © 2025