AI Impact Measurement Scorecard

Track the 5 metrics that prove AI business value: a fill-in template for Forward Deployed Engineers.

How to Use This Scorecard

Most AI projects can show that the technology works. Far fewer can show that the business is better off because of it. This scorecard gives Forward Deployed Engineers (FDEs) one consistent way to measure, score and report AI impact across five metrics, so every conversation with stakeholders is grounded in evidence rather than impressions.

Use it on a single AI solution at a time. Fill it in at three points: before build (baseline and targets), during the pilot (early signals) and at the end of the measurement period (final scores and decision).

The three-step flow

Step	What you do	When
1. Define baseline	Record how the process performs today and agree targets and the definition of "verified outcome" with the business owner.	Before development starts
2. Measure	Collect the same data after deployment, for the same tasks and users, over a fixed period.	Pilot and measurement period
3. Score & decide	Score each metric 1–5, total the scores and use the decision bands to scale, optimise or reconsider.	End of measurement period

> **Golden rule**
>
> Measure before you build. Without a baseline, no metric in this scorecard is credible, and every result becomes a matter of opinion.

Choosing a measurement period

The period should be long enough to show stable behaviour, not just launch-week excitement:

- **30 days:** quick check for simple, high-volume workflows (for example, ticket tagging).
- **60 days:** the default for most pilots; long enough to see adoption settle.
- **90 days:** for complex or low-volume processes, or when ROI must be presented to finance.

Scoring scale and status colours

Every metric uses the same 1–5 score and a Red/Amber/Green status so results can be compared across projects:

Score	Meaning	Status
5	Clearly exceeds target with strong evidence	Green

4	Meets target	Green
3	Partly meets target; gaps identified	Amber
2	Well below target	Red
1	No measurable result or negative result	Red

> **About the worked example**
>
> Each metric section includes a worked example from Northwind Retail, a fictional company that deployed an AI support-ticket classifier. All figures are illustrative and shown to demonstrate method, not to set benchmarks.

Project Snapshot

Capture the essentials of the project before measuring anything. This page keeps everyone aligned on what problem is being solved, who owns the result and over what period it will be judged.

Field	Details
Project name	
Client / business unit	

Business problem being solved	
AI solution type	
FDE owner	
Business owner (signs off results)	
Other stakeholders	
Project start date	
Go-live date	
Measurement period	

Example: Northwind Retail

Field	Details
Business problem	Agents spend ~12 min per ticket manually tagging, routing and drafting replies; weekend overtime needed to clear backlog.
AI solution type	Ticket classifier + suggested-reply assistant integrated into the existing ticketing platform.
Business owner	Head of Support Operations
Measurement period	60 days post-launch (with 4-week pre-launch baseline)

Success looks like…

Write one or two sentences that the business owner would agree with.

Make it specific and measurable.

> **Your statement:**
>
> _____
>
> _____
>
> ____
>
> **Northwind example:** *"Within 90 days, cut average handling time by 25% and raise first-contact resolution to 70%, without lowering CSAT."*

METRIC 1 OF 5

Time to Value: How Quickly Does AI Deliver Results?

Time to value measures the gap between the day a project starts and the day the AI solution produces its first verified business outcome. A working demo, a successful model test or a go-live announcement are milestones, not value. The clock stops only when the solution measurably improves a real workflow for real users.

For FDEs, this metric matters because stakeholders judge momentum early. A project that shows a verified result in 10 weeks builds the trust and budget needed to scale; one that is still "almost ready" after six months usually loses both.

What it measures

- **Speed of delivery:** how long discovery, build, integration and deployment take end to end.
- **Speed of proof:** how long until the first outcome that the business owner agrees is real.
- **Speed to target:** how long until performance reaches the target agreed at kick-off.

Formula

> **TIME TO VALUE (DAYS)**
>
> **Date of first verified business outcome – Project start date**

Agree up front what counts as "verified". A good definition names the metric, the threshold and who signs it off, for example: "Average ticket handling time falls by at least 15% for two consecutive weeks, confirmed by the Support Operations Manager."

Fill-in worksheet

Input	Baseline	Current	Target	Status
Project start date				
First production deployment date				
First verified outcome date				
Time to value (days)				
Date target performance reached				

Worked example: Northwind Retail

Northwind Retail (a fictional mid-sized e-commerce company) asked its FDE to build an AI classifier that tags incoming support tickets and suggests replies. This example runs through all five metrics in this scorecard.

Input	Baseline	Current	Target	Status
Project start date	—	1 Jun	—	—
First production deployment	—	22 Jul	15 Jul	Amber
First verified outcome	—	14 Aug	31 Aug	Green
Time to value (days)	—	74	≤ 90	Green

Target performance reached	—	9 Sep	30 Sep	Green

> **What the example tells us**
>
> Deployment slipped one week because of ticketing-system API access, but the first verified outcome still landed 17 days ahead of target. Score: 4/5. The slip is worth noting so integration access is requested earlier on the next project.

Supporting indicators to track

The headline formula rarely tells the whole story. Tick the indicators you will track alongside it:

- ☐ Time to first working prototype
- ☐ Time to first production deployment
- ☐ Time lost waiting on data, access or approvals
- ☐ Time to reach the agreed performance target
- ☐ Number of scope changes after kick-off

Scoring guide

Score	What it looks like
5	Verified outcome delivered well ahead of target with no compromise on security or quality.
4	Verified outcome delivered on or before target; minor slips in interim milestones.
3	Verified outcome delivered up to 25% later than target.
2	Significantly late, or outcome only partially verified.
1	No verified outcome yet within the measurement period.

Red flags

- Counting the demo date or go-live date as "value".
- Shortcuts on security review or testing to hit a date.
- No agreed definition of a verified outcome before work starts.

Time to Value score

Score (circle one): 1 2 3 4 5 Status: ☐Red ☐Amber ☐Green

Notes:

_____

METRIC 2 OF 5

Productivity Improvement: Is AI Helping Teams Accomplish More?

Productivity improvement measures whether AI helps people finish tasks faster, handle more volume or spend less effort on repetitive work, without lowering quality. It is the most common reason enterprises adopt AI, which is exactly why it is often overstated.

The key discipline is comparing like with like. Measure the same task, of similar complexity, done by the same kind of user, before and after AI. Then check that quality held steady, because a task that is 30% faster but needs rework has not really improved.

What it measures

- **Time per task:** how long a typical unit of work takes.
- **Throughput:** how many units a person or team completes per day or week.
- **Quality:** error, rework or escalation rates that confirm speed did not cost accuracy.
- **Use of recovered capacity:** what the saved time is actually spent on.

Formula

PRODUCTIVITY IMPROVEMENT (%)

[(Baseline time per task – New time per task) ÷ Baseline time per task] × 100

Example: a task that took 40 minutes now takes 30 minutes with AI support. (40 – 30) ÷ 40 × 100 = 25% improvement, assuming comparable complexity and quality. Remember that time saved is not the same as money saved. Record what the recovered time is used for; that answer feeds directly into Metric 5 (ROI).

Fill-in worksheet

Input	Baseline	Current	Target	Status
Average time per task (min)				
Tasks completed per person per day				
Rework / error rate (%)				
Escalation rate (%)				
Productivity improvement (%)				

Worked example: Northwind Retail

Northwind measured 4 weeks of ticket handling before launch and 4 weeks after, using the same ticket categories and the same 60-agent team.

Input	Baseline	Current	Target	Status
Avg handling time (min)	12.0	8.0	9.0	Green
Tickets per agent per day	35	48	45	Green
Rework / reopen rate (%)	9%	7%	≤ 9%	Green
Escalation rate (%)	14%	15%	≤ 14%	Amber
Productivity improvement	—	33%	25%	Green

What the example tells us

Handling time fell from 12 to 8 minutes: (12 − 8) ÷ 12 × 100 = 33%. Quality held, with fewer reopened tickets. Escalations rose slightly, so the FDE reviewed which categories the classifier routes poorly. Recovered capacity was used to clear a backlog that previously needed weekend overtime. Score: 4/5.

Supporting indicators to track

The headline formula rarely tells the whole story. Tick the indicators you will track alongside it:

☐ Time per task (before vs. after)

☐ Volume handled per person

☐ Rework and error rates

☐ Review or approval effort

☐ Overtime or contractor hours

☐ Where recovered capacity is being used

Scoring guide

Score	What it looks like
5	Improvement of 30%+ with quality equal or better, and recovered capacity clearly redeployed.
4	Improvement meets target with quality steady.
3	Modest improvement (10–20%) or improvement with a small quality trade-off.
2	Marginal improvement, or gains offset by extra review work.
1	No measurable improvement, or quality has declined.

Red flags

- No baseline was recorded before launch.
- Comparing simple post-AI tasks with complex pre-AI tasks.

- Reporting minutes saved as rupees saved without checking how the time is used.

Productivity Improvement score

Score (circle one): 1 2 3 4 5 Status: ☐Red ☐Amber ☐Green

Notes:

_____

METRIC 3 OF 5

User Adoption: Are People Actually Using the AI Solution?

A technically excellent AI tool delivers nothing if people do not use it. User adoption measures whether the intended users rely on the solution regularly, as part of their real workflow, after the launch excitement fades.

Adoption is also an early-warning system. Usage that rises in week one and drops by week four usually signals inaccurate outputs, a confusing interface or extra manual steps. FDEs who watch adoption closely can fix these problems before they undermine the business case.

What it measures

- **Reach:** what share of intended users are active.
- **Stickiness:** whether users come back week after week.
- **Trust:** how often users accept, edit or override AI outputs.
- **Depth:** what share of eligible tasks are completed with AI assistance.

Formula

> **ADOPTION RATE (%)**
>
> **(Intended users actively using the solution ÷ Total intended users) × 100**

Define "active" before you measure, for example "used the tool on at least 3 working days in the week". Counting log-ins or sign-ups inflates adoption; counting meaningful use does not.

Fill-in worksheet

Input	Baseline	Current	Target	Status
Total intended users				
Weekly active users				
Adoption rate (%)				
AI suggestion acceptance rate (%)				

Override / heavy-edit rate (%)				
Tasks completed with AI (%)				

Worked example: Northwind Retail

Northwind tracked its 60 support agents for 8 weeks after launch. "Active" was defined as using AI suggestions on 3+ days per week.

Input	Baseline	Current	Target	Status
Total intended users	60	60	—	—
Weekly active users	—	52	48	Green
Adoption rate (%)	—	87%	80%	Green
Suggestion acceptance rate	—	71%	65%	Green
Override / heavy-edit rate	—	18%	≤ 20%	Green
Tickets handled with AI	—	64%	70%	Amber

> **What the example tells us**
>
> Adoption is strong at 52 ÷ 60 = 87%. But only 64% of tickets used AI, because agents skipped it for refund and billing tickets where

> suggestions were weaker. The FDE retrained the model on refund tickets and ran a 30-minute refresher session. Score: 4/5.

Supporting indicators to track

The headline formula rarely tells the whole story. Tick the indicators you will track alongside it:

- ☐ Weekly and monthly active users
- ☐ Repeat usage week over week
- ☐ Acceptance rate of AI outputs
- ☐ Edit and override rates
- ☐ Share of eligible tasks completed with AI
- ☐ User satisfaction score or feedback themes

Scoring guide

Score	What it looks like
5	85%+ adoption, stable or rising over time, high acceptance and positive feedback.
4	Adoption meets target; minor gaps in specific task types or teams.
3	50–70% adoption, or usage declining after launch.

2	Fewer than half of intended users active; frequent overrides.
1	Tool is available but largely unused.

Red flags

- Reporting sign-ups or launch-week numbers as adoption.
- High override rates that nobody investigates.
- Usage mandated by management but outputs ignored in practice.

User Adoption score

Score (circle one): 1 2 3 4 5 Status: ☐Red ☐Amber ☐Green

Notes:

_____

METRIC 4 OF 5

Business Outcome Improvement: Is AI Improving the Target Process?

Productivity shows that work got faster. Business outcome improvement shows whether the business actually got better. It ties the AI solution to the

KPI the organisation cares about: resolution rates, conversions, defect rates, claims processed on time or customer satisfaction.

This metric protects against hollow wins. A support assistant that cuts handling time but lowers customer satisfaction, or increases follow-up contacts, has not improved the business. Pick one primary KPI linked to the original problem, and one or two guard-rail KPIs that must not get worse.

What it measures

- **Primary KPI:** the outcome the AI was deployed to improve.
- **Guard-rail KPIs:** outcomes that must not decline as a side effect.
- **Unintended effects:** changes elsewhere in the process caused by the AI.

Formula

> **KPI IMPROVEMENT (%)**
>
> **[(Post-AI KPI – Baseline KPI) ÷ Baseline KPI] × 100**

For KPIs where lower is better (cost per interaction, error rate), reverse the sign: (Baseline − Post-AI) ÷ Baseline × 100. Where possible, compare against a pilot group or a team not yet using AI, so seasonal or market changes are not credited to the AI.

Fill-in worksheet

Input	Baseline	Current	Target	Status
Primary KPI				
Guard-rail KPI 1				
Guard-rail KPI 2				
Cost per transaction (₹)				
KPI improvement (%)				

Worked example: Northwind Retail

Northwind chose first-contact resolution (FCR) as its primary KPI, with customer satisfaction (CSAT) and cost per ticket as guard-rails.

Input	Baseline	Current	Target	Status
First-contact resolution	62%	71%	70%	Green
CSAT (out of 5)	4.1	4.2	≥ 4.1	Green
Repeat contacts within 7 days	11%	9%	≤ 11%	Green
Cost per ticket (₹)	₹96	₹71	₹75	Green
FCR improvement (%)	—	14.5%	13%	Green

What the example tells us

FCR rose from 62% to 71%: (71 − 62) ÷ 62 × 100 = 14.5%. CSAT and repeat contacts held or improved, so the speed gain did not damage the customer experience. A control team without AI saw FCR move only 1 point in the same period, which strengthens the case. Score: 5/5.

Supporting indicators to track

The headline formula rarely tells the whole story. Tick the indicators you will track alongside it:

☐ Customer service: first-contact resolution, CSAT, complaint rate, cost per interaction

☐ Sales & marketing: qualified leads, conversion rate, campaign performance, revenue

☐ Software engineering: delivery cycle time, defect rate, rework, release frequency

☐ Operations: processing time, error rate, throughput, missed SLAs

☐ Control or pilot group comparison in place

Scoring guide

Score	What it looks like
5	Primary KPI beats target, guard-rails steady or better, results confirmed against a control.
4	Primary KPI meets target; guard-rails steady.
3	Primary KPI improves but below target, or one guard-rail slips.
2	Little KPI change, or improvement cannot be separated from other factors.
1	Primary KPI flat or worse, or a guard-rail declines significantly.

Red flags

- Measuring AI output volume (responses, predictions) instead of the process outcome.
- No guard-rail KPIs, so side effects go unnoticed.
- Crediting AI for gains caused by seasonality, pricing or staffing changes.

Business Outcome Improvement score

Score (circle one): 1 2 3 4 5 Status: ☐Red ☐Amber ☐Green

Notes:

_____

METRIC 5 OF 5

Return on Investment: Does AI Generate More Value Than It Costs?

ROI translates everything above into the language leadership uses to decide whether to fund, scale or stop: money. It compares the measurable benefits of the AI solution with the full cost of building, deploying and running it over the same period.

A credible ROI figure is conservative. It counts all costs (not only the build), counts only benefits that have been verified, and keeps projected savings separate from realised savings. A modest ROI that finance can audit is far more useful than an impressive one nobody believes.

What it measures

- **Total investment:** development, integration, models and infrastructure, monitoring, maintenance, human review, training and support.
- **Realised benefits:** verified cost reductions, added revenue, avoided error costs and redeployed capacity that is actually used.
- **Projected benefits:** expected but not yet verified gains, reported separately.

Formula

> **ROI (%)**
>
> **[(Total measurable benefits – Total AI investment) ÷ Total AI investment] × 100**

Use the same measurement period for costs and benefits, and do not count a benefit twice. For example, if handling-time savings are already counted as reduced overtime, do not also count them as lower cost per ticket.

Fill-in worksheet

Input	Baseline	Current	Target	Status
Development & integration (₹)				
Model & infrastructure (₹)				
Monitoring & maintenance (₹)				
Human review, training & support (₹)				
Total investment (₹)				
Realised benefits (₹)				

ROI (%)				

Worked example: Northwind Retail

Northwind calculated ROI over the first 12 months, separating realised from projected benefits.

Input	Baseline	Current	Target	Status
Development & integration	—	₹4.5L	—	—
Model & infrastructure	—	₹2.5L	—	—
Monitoring & maintenance	—	₹1.5L	—	—
Human review, training, support	—	₹1.5L	—	—
Total investment	—	₹10L	—	—
Realised benefits (overtime, contractors, cost per ticket)	—	₹15L	₹13L	Green
ROI (%)	—	50%	30%	Green

> **What the example tells us**
>
> ROI = (₹15L − ₹10L) ÷ ₹10L × 100 = 50%. A further ₹4L of projected benefit (lower churn from faster resolution) was reported separately

> and not included, keeping the figure defensible. This is an illustrative calculation, not a guaranteed result. Score: 4/5.

Supporting indicators to track

The headline formula rarely tells the whole story. Tick the indicators you will track alongside it:

- ☐ All cost categories captured for the full period
- ☐ Benefits verified by finance or the business owner
- ☐ Projected and realised benefits reported separately
- ☐ No double-counting across benefit lines
- ☐ Payback period (months) calculated
- ☐ Cost per interaction trend over time

Scoring guide

Score	What it looks like
5	ROI well above target, fully costed, benefits verified by finance.
4	ROI positive and above target; small assumptions clearly stated.
3	ROI positive but below target, or relies partly on projected benefits.

2	Roughly break-even, or costs incompletely captured.
1	Negative ROI, or no credible calculation possible.

Red flags

- Leaving out monitoring, human review or retraining costs.
- Treating hours saved as payroll saved.
- Mixing projected and realised benefits in one number.

Return on Investment score

Score (circle one): 1 2 3 4 5 Status: ☐Red ☐Amber ☐Green

Notes:

_____

Technical Health Check

The five business metrics depend on a system that works reliably. Technical metrics do not prove value on their own, but a drop in accuracy, speed or safety usually shows up in adoption and outcomes a few weeks later. Track these alongside the scorecard as early-warning signals.

Technical metric	Target	Actual	Status
Accuracy / output quality (%)			
Latency (avg response time)			
Reliability / uptime (%)			
Cost per interaction (₹)			
Safety or policy incidents			
Data drift alerts			

Example: Northwind Retail

Technical metric	Target	Actual	Status
Classification accuracy	≥ 90%	93%	Green
Avg suggestion latency	≤ 2 s	1.4 s	Green
Uptime	≥ 99.5%	99.7%	Green
Cost per interaction	≤ ₹2.0	₹1.6	Green
Safety / policy incidents	0	1 (resolved)	Amber

> **Keep the link clear**

> Technical metrics confirm the system works. Business metrics confirm it matters. Report both, but lead with the business result when presenting to stakeholders.

Scorecard Summary Dashboard

Bring the five scores together at the end of the measurement period. Share this page with the business owner and leadership as a one-page summary of impact.

Metric	Score (1–5)	Status	Key finding
1. Time to Value			
2. Productivity Improvement			
3. User Adoption			
4. Business Outcome Improvement			
5. Return on Investment			

TOTAL (out of 25)			

Decision bands

Total score	Decision	What to do next
20–25	Scale	Expand to more teams, regions or workflows. Document the playbook so the next rollout is faster.
13–19	Optimise	Fix the lowest-scoring metric first (often adoption or a guard-rail KPI), then re-measure before scaling.
12 or below	Reconsider	Re-scope the use case, address root causes, or stop and redirect investment to a stronger opportunity.

Example: Northwind Retail

Metric	Score	Status	Key finding
Time to Value	4	Green	74 days to verified outcome; API access delayed deployment by a week.
Productivity	4	Green	Handling time down 33%; quality steady.
User Adoption	4	Green	87% adoption; refund tickets under-served.
Business Outcome	5	Green	FCR up 14.5% vs. 1 pt in control team.
ROI	4	Green	50% ROI in 12 months on realised benefits only.
TOTAL	21 / 25	Scale	Roll out to the returns and billing teams next quarter.

Next actions

Action	Owner	Due date
1.		
2.		
3.		

Sign-off: FDE owner _____ Business owner _____ Date _____

GSDC CERTIFIED FORWARD DEPLOYED ENGINEER

Measure, prove and scale AI impact like a certified FDE

This scorecard shows you what to measure. The certification shows employers and clients you can deliver it.

What the certification covers

- **Customer discovery:** turning vague business problems into measurable AI use cases with clear baselines.
- **Production AI delivery:** building, integrating and deploying AI inside real customer environments.
- **Evaluation and monitoring:** designing evaluation pipelines and tracking performance after launch.
- **Value realisation:** measuring adoption, business outcomes and ROI, and presenting them to leadership.

Who it is for

Software engineers, solutions engineers, consultants and technical professionals who want to move into customer-facing AI delivery roles, or formalise the FDE skills they already use.

GSDC
Global Skill Development Council

CERTIFIED FORWARD DEPLOYED ENGINEER

LEARN DIRECTLY FROM INDUSTRY EXPERTS, AI PRACTITIONERS, AND ENGINEERING LEADERS WHO ARE BUILDING AND DEPLOYING SCALABLE AI DRIVEN SOLUTIONS FOR ENTERPRISES WORLDWIDE.



ABOUT GSDC CERTIFICATION

EBOOK

Extensive and exclusive Ebook created by world's experts to help you with understanding core concepts.

CREATED BY EXPERTS

GSDC certifications are created and authored by world's leading experts in the field.

EBOOK

Extensive and exclusive Ebook created by world's experts to help you with understanding core concepts.

LEARNING MATERIALS

Get access to learning materials such as videos, ebooks, templates, and practice exams, which will help you clear the certification exam.

LEARNING OBJECTIVE

- Validate your Forward Deployed Engineering expertise.
- Increase your earning potential with an in-demand credential.
- Get complimentary career assistance after certification.

Enroll now with the code **LEARN20** To avail **20%** discount

Enroll Now

www.gsdcouncil.org