

PRODUCT STRATEGY · ANALYTICS · EARLY-STAGE VALIDATION

BERT KPI Tree & Measurement Framework

A practical measurement system connecting acquisition, journey completion, Blueprint quality, meaningful product value and responsible operating guardrails.

Nanda Kishore Reddy Mangunuri

Early beta

30 September 2026

Owner: Nanda Kishore Reddy Mangunuri
Product stage: Early beta
Measurement horizon: Beta to product-market validation
Last updated: 30 September 2026

1. Measurement objective

BERT should not optimise for downloads, account creation or daily opens in isolation. The product creates value when a user completes a Blueprint, meaningfully engages with it and uses it for reflection or collaboration.

The measurement system therefore connects five questions:

1. Can users reach their Blueprint?
2. Does the Blueprint feel specific, accurate and useful?
3. Do users act on or share the output?
4. Does the product remain reliable, trustworthy and privacy-conscious?
5. Can value be delivered at a sustainable operational cost?

2. North-star metric

Monthly Meaningful Blueprint Users (MMBU)

Definition: Unique verified users in a calendar month who generate or retrieve a completed Blueprint and complete at least one meaningful value action within seven days.

A meaningful value action is one of:

- reads at least three report sections;
- returns to the Blueprint on a later day;
- submits contextual strength feedback;
- creates a share link;
- completes the structured external feedback form.

**MMBU = distinct users with a completed Blueprint
AND at least one qualifying value action within 7 days**

Why this metric

- It measures experienced value rather than registration.
- It accommodates BERT's episodic usage; daily engagement is not inherently desirable.
- It connects product quality with action.
- It can grow through acquisition, completion, report quality, sharing or return behaviour.

Companion rate

Meaningful Blueprint Use Rate

= Meaningful Blueprint Users / Users who generated a Blueprint

The count measures scale. The rate identifies whether product quality is improving independently of acquisition.

Transitional beta metric

Report-section views and later-day returns are not fully instrumented today. Until they are available, use **Completed Blueprint Users** with strength feedback, sharing and external feedback as supporting signals. Do not present this approximation as the final north-star metric.

3. KPI tree

Monthly Meaningful Blueprint Users

- ├─ 1. Qualified acquisition
 - ├─ Store/website visitors
 - ├─ Install-to-account rate
 - └─ Verified-account rate
- ├─ 2. Journey activation
 - ├─ Profile and consent completion
 - ├─ Assessment start rate
 - ├─ Assessment completion rate
 - ├─ Blueprint generation success
 - └─ Median time to Blueprint
- ├─ 3. Blueprint quality
 - ├─ Perceived accuracy
 - ├─ Workplace usefulness
 - ├─ Genericness rate
 - ├─ Repetition rate
 - └─ Strength-fit rate
- ├─ 4. Value behaviour
 - ├─ Report depth
 - ├─ Strength-feedback participation
 - ├─ Share creation and share opens
 - ├─ Later-day return rate
 - └─ External-feedback completion
- └─ 5. Guardrails
 - ├─ Reliability and latency
 - ├─ AI groundedness and fallback
 - ├─ Privacy and deletion correctness
 - ├─ Accessibility
 - └─ Cost per completed/meaningful Blueprint

4. Current beta baseline

Metric	Baseline	Interpretation
Registered beta users	27	Small initial cohort
Completed Blueprints	20	74% registered-user completion
External feedback responses	12	60% of Blueprint recipients
Journey rated easy/very easy	83%	Positive usability signal
Completed within 15 minutes	92%	Journey duration broadly acceptable
Blueprint moderately/very accurate	92%	Encouraging, not validation
Blueprint felt generic	67%	Primary product-quality concern
Blueprint reported repetitive	42%	Report-content concern
At least moderately useful at work	100%	Positive perceived-value signal
Likely/very likely to share	67%	Promising collaboration signal
Recalled using in-app feedback	17%	Low discoverability or motivation
Reported technical issue	0%	Positive but based on 12 iOS respondents

These results are directional. The sample is small, iOS-only and subject to early-adopter and response bias.

5. Funnel framework

5.1 Canonical funnel

Store/website visit

→ App install/open

→ Account created

→ OTP verified

→ Consent completed

→ Birth profile completed

→ Assessment started

→ Assessment completed

→ Blueprint generated

→ Blueprint meaningfully used

→ Blueprint shared or revisited

5.2 Funnel definitions

Stage	Numerator	Denominator	Window
Account conversion	New accounts	First app opens	Same day
OTP verification	Successful verifications	OTP requests	30 minutes
Profile completion	Completed required profile	Verified accounts	24 hours
Assessment start	Created sessions with first answer	Completed profiles	24 hours
Assessment completion	Completed sessions	Started sessions	7 days
Blueprint generation	Successfully rendered reports	Completed assessments	30 minutes
Meaningful use	Users with qualifying value action	Generated Blueprints	7 days
Sharing	Users creating a share link	Meaningfully used Blueprints	7 days
Return	Users returning on a later day	Generated Blueprints	30 days

5.3 Segmentation

Segment funnel results by:

- platform: iOS/Android;
- app version and build;
- assessment engine/config version;
- report schema version;
- acquisition source when available;
- first-time versus returning user;
- environment for operational debugging only.

Avoid demographic segmentation unless it has a defined lawful, ethical and product-relevant purpose.

6. Metric dictionary

Acquisition and authentication

Metric	Formula	Decision supported
Verified-account rate	Verified users / accounts created	Is authentication preventing entry?
OTP first-attempt success	First-attempt verifications / users attempting verification	Is OTP understandable and reliable?
OTP median completion time	Median verification timestamp – request timestamp	Does OTP create avoidable delay?
Acquisition conversion	Verified accounts / attributed store or website visits	Are campaigns reaching suitable users?

Activation and assessment

Metric	Formula	Decision supported
Profile completion	Completed required profile / verified users	Is profile/birth onboarding trusted and usable?
Assessment start	Sessions with ≥ 1 answer / completed profiles	Does the value proposition motivate assessment?
Assessment completion	Completed sessions / sessions with ≥ 1 answer	Is the experience too long or confusing?
Median assessment time	Median completion – first answer	Is adaptive assessment becoming more efficient?
Median questions answered	Median response count for completed sessions	Does efficiency improve without quality loss?
Resume success	Resumed sessions reaching next valid state / resume attempts	Can interrupted users recover?

Blueprint quality

Metric	Formula	Decision supported
Generation success	Successful reports / completed assessments	Can users reach the core value?
Time to Blueprint	Report generated – assessment completed	Is generation wait acceptable?
Strength-fit rate	“Sounds like me” / all strength responses	Which report versions fit users better?
Feedback participation	Reports with ≥ 1 strength response / eligible reports	Is feedback visible and worthwhile?
Perceived accuracy	Survey distribution and mean	Does the report reflect the user?
Genericness rate	“Yes, generic” / survey responses	Is differentiation improving?
Repetition rate	Respondents selecting repetitive / survey responses	Is content synthesis improving?
Workplace usefulness	Survey rating distribution	Does BERT support the intended job?

Engagement, sharing and retention

Metric	Formula	Decision supported
Report depth	Users viewing ≥ 3 sections / report viewers	Are users exploring beyond the summary?
Later-day return	Users viewing on a later day / report owners	Does the Blueprint have durable value?
Share creation	Owners creating a link / report owners	Is the output worth sharing?
Share consumption	Active shared reports with ≥ 1 view / active shared reports	Do invited readers engage?
External feedback completion	Completed forms / issued feedback tokens	Are users willing to provide structured evidence?

Reliability and responsible-AI guardrails

Metric	Definition
Generation failure rate	Failed generation attempts / all attempts
Deterministic-fallback rate	Reports using fallback / generated reports
Generation latency	p50, p95 and p99 from start to persisted report
Assessment anomaly rate	Sessions with telemetry anomaly / started sessions
Grounding failure rate	AI outputs rejected for unsupported content / AI attempts
Schema rejection rate	Invalid AI payloads / AI attempts
PII incident count	Confirmed sensitive-data exposure events; target zero
Deletion completion	Successful deletions / valid deletion requests
Shared-link revocation	Disabled links no longer accessible / revoked links tested
Cost per completed Blueprint	Monthly attributable platform cost / completed Blueprints
Cost per meaningful user	Monthly attributable platform cost / MMBU

7. What can be measured today

Available now

- Registered and verified users.
- Stored onboarding status and inferred resume state.
- Assessment sessions, responses and completion state.
- V4 assessment creation, probe selection, response, stop-decision and summary events.
- Per-dimension confidence, uncertainty and contradiction telemetry.
- Report-generated events and persisted report views.
- Strength feedback and positive/negative fit.
- External feedback tokens and completed responses.
- Report share creation, active status and aggregate view count.
- App version, build and platform on V4 engine events.
- Admin feedback distributions with small-sample confidence intervals.

Partially measurable

- End-to-end funnel can be inferred from database state, but not every transition has a timestamped product event.
- Generation success can be inferred, but generation-start, stage timing and deterministic-fallback data need a consistent event contract.
- Sharing is persisted, but the full user interaction from viewing to native share completion is not observable.
- Resume state exists, but resume attempts and successful recoveries are not explicitly counted.

Missing instrumentation

- Store/website acquisition attribution.
- First app open and screen-level journey events.
- OTP request, delivery, entry attempt and verification timing.
- Birth-details explanation view and comprehension signal.
- Report section views and reading depth.
- Later-day report returns.

- Native share-sheet completion.
- Mobile crash-free session metrics linked safely to app version.
- Explicit generation start, stage duration, AI-validation outcome and fallback reason.

8. Event taxonomy

8.1 Common event envelope

Every new product event should contain:

- event name and schema version;
- UTC timestamp;
- pseudonymous user identifier where required;
- session/assessment/report identifiers where relevant;
- platform, app version and build;
- environment;
- client request ID for idempotency;
- allow-listed event properties.

Events must not contain email addresses, raw birth details, coordinates, OTP values, access/share tokens, raw assessment responses or report prose.

8.2 Proposed events

Journey	Event	Key safe properties
Acquisition	app_first_opened	platform, version, attributed source
Authentication	otp_requested	attempt number, delivery outcome
Authentication	otp_verification_completed	success, safe failure code, elapsed bucket
Profile	profile_step_completed	step key, elapsed bucket
Assessment	assessment_started	engine/config version, platform
Assessment	Existing V4 response/stop events	session and version metadata
Assessment	assessment_resumed	previous stage, elapsed bucket, success
Report	blueprint_generation_started	report schema and engine version
Report	Existing v4_report_generated	add duration, LLM/fallback and validation result
Report	blueprint_generation_failed	stage, safe error, recovery result
Engagement	blueprint_viewed	report version, audience, entry point
Engagement	blueprint_section_viewed	stable section key, report version
Feedback	Existing v4_strength_feedback_submitted	report/content version and response
Sharing	blueprint_share_created	report version and platform
Sharing	blueprint_share_opened	privacy-safe share ID and view bucket
Retention	blueprint_returned	elapsed-day bucket and report version

9. Dashboard design

9.1 Executive product dashboard

Displayed weekly and monthly:

- MMBU and Meaningful Blueprint Use Rate;
- registered users and completed Blueprints;
- end-to-end completion funnel;
- quality signals: accuracy, genericness, repetition and usefulness;
- sharing and later-day return;
- generation reliability and cost guardrails.

9.2 Journey dashboard

- conversion and median elapsed time per funnel step;
- largest absolute user loss and largest percentage drop;
- OTP attempts and failure reasons;
- assessment abandonment by answered-question band;
- resume attempts and recovery;
- generation waiting and failures.

9.3 Assessment and behavioural-quality dashboard

- median items answered;
- stop-reason distribution;
- confidence and uncertainty distributions;
- contradiction distribution;
- marginal information gain by probe;
- anomaly rate by engine/config release;
- “sounds like me” rate by domain and report version.

These operational behavioural metrics must not be interpreted as proof of psychometric validity.

9.4 Reliability and trust dashboard

- API and generation success;
- latency percentiles;
- AI validation/fallback rate;
- deletion success;
- revoked-link checks;
- privacy/security incidents;
- cost per completed and meaningful Blueprint.

10. Review cadence and ownership

Cadence	Review	Participants	Decisions
Daily during release	Reliability and funnel exceptions	Engineering/product	Incident response and rollback
Weekly	Product funnel and qualitative feedback	Product, design, engineering	Backlog and investigation priorities
Per release	Version comparison and guardrails	Product, engineering, behavioural reviewer	Promote, hold or rollback
Monthly	North-star, quality, sharing and cost	Founders/product	Roadmap and resourcing
Quarterly when scaled	Validation and fairness review	Product, behavioural science, privacy	Product boundaries and research plan

11. Decision rules

1. Do not optimise acquisition while completion or report quality is materially declining.
2. Do not promote an AI/report version that improves engagement but fails grounding, privacy or schema guardrails.
3. Do not infer a product trend from very small samples without counts and uncertainty intervals.
4. Pair quantitative movement with qualitative review when interpreting accuracy or usefulness.
5. Do not treat higher confidence scores as proof of real-world behavioural accuracy.
6. Investigate platform/build differences before assuming a general product problem.
7. Prefer cohort and version comparisons over cumulative lifetime averages.

12. Experimentation approach

Current beta stage

The cohort is too small for dependable high-powered A/B testing. Use:

- moderated usability sessions;
- pre/post release comparisons with raw counts and confidence intervals;
- report-level blinded human evaluation;
- event and failure-path analysis;
- qualitative coding of feedback.

Later stage

When sample size supports it:

- define the minimum meaningful effect before calculating sample size;
- randomise at user level and prevent report-version crossover;
- choose one primary outcome and a small number of guardrails;
- run long enough to capture delayed report use and sharing;
- report confidence intervals, not only whether a threshold was crossed.

Potential experiments:

1. Birth-data explanation variants → profile completion and comprehension.
2. OTP autofill/paste changes → verification success and time.
3. Progress indicator → assessment completion and perceived length.
4. Blueprint 2.0 versus current → genericness, accuracy and meaningful use.
5. Contextual feedback prompt → participation without report abandonment.

13. Initial measurement backlog

Order	Work item	Value
1	Define event names, owners, schemas and prohibited fields	Prevent inconsistent or unsafe analytics
2	Add explicit OTP and onboarding-step events	Locate the largest journey friction
3	Add generation start/fail/complete timing and fallback metadata	Measure core-value reliability
4	Add report and section view events	Enable the north-star metric
5	Add later-day return and resume-success events	Measure durable value and recovery
6	Version all report-quality and feedback records	Compare releases correctly
7	Build executive and journey dashboards in the admin console	Make evidence actionable
8	Establish monthly cost allocation	Track sustainability

14. Portfolio interpretation

This framework deliberately separates:

- **usage** from **meaningful product value**;
- **self-reported accuracy** from **scientific validation**;
- **AI output quality** from **behavioural score validity**;
- **available operational data** from **instrumentation still required**;
- **growth metrics** from **privacy, reliability and cost guardrails**.

The objective is not to create the largest possible dashboard. It is to create a decision system that tells the team where users lose value, whether a release improves the core experience and when growth should be slowed to protect trust.