ZenAI Behaviour Twin: Auditor Test Guide

Nov 19, 2026 · @Sunil Mysore Kempegowda

Purpose and how to use this guide

Auditors can confirm in about 20 minutes, with no sign-in, that the prototype helps a person recover from a lapse, keeps them safe, measures lasting change, and runs on Azure. Every step lists what it tests, why, and the output you should see.

Item	Detail
Live demo	https://behaviourtwin.zenaidigital.com (backup: Azure default address)
Health check	/api/health
Hosting	Azure Container Apps, UAE North, with Application Insights and Log Analytics
Release	v8.6; 109/109 rule scenarios and 82/82 browser tests pass on this exact build. The same 25 tests are built into the app: "For auditors: checklist (25 tests)" on the first screen
Time needed	Part A about 5 minutes; Part B about 15 minutes
Access	Public, no account, no password, no cookies
Data	Synthetic only. Nothing you type is stored after the page closes
Browser	Any current browser; a private window is best. Works on phones

Before you start: open the link in a private window. Check that the header chip reads "Azure: live · Azure Container Apps · UAE North". If it reads "Demo fallback mode", the page still works, but the Azure tests in Part B will show "not reachable". If you get lost at any point, click **Reset demo**.

The problem statement being tested

Challenge 09 asks for support that helps people *start, keep up and get back to* healthier behaviours, adapting to their goals, circumstances and setbacks over time. Our focus is the hardest part: getting back on track after a lapse, for movement.

The human problem: two in three UAE adults are insufficiently active (66.1%, the highest of 197 countries), and the benefit of step trackers roughly halves after six months. Most apps go quiet or reset a streak after a missed week; this prototype shrinks the plan, keeps progress and measures recovery.

Challenge 09 desired outcome	What the auditor should see	Tests
1. Adaptive personal plans	A plan built from the person's own baseline and chosen action; steps or active minutes	A1, A2, A3, B9
2. Continuous learning and re-engagement	Missed days detected, plan made smaller without blame, restart, recovery time recorded	A4, A5, A6, A8
3. Evidence-based behaviour support	Each message carries a technique and source; the assistant refuses medical questions	A6, B3, B4
4. User autonomy and inclusion	Keep or undo every change; consent per purpose; Arabic; wheelchair users; aggregates only for managers	A7, B2, B5, B6, B9
5. Sustained outcome measurement	Persistence at weeks 4, 8, 12, recovery, re-engagement and equity in an aggregate dashboard	A8, A9, B7
Safety and trust (implicit in the brief)	Red flags stop coaching; every decision traceable; same rules running in Azure	B1, A10, B8

Part A: the guided participant journey (about 5 minutes)

Part A proves the core of Challenge 09: a person lapses, the plan shrinks without blame, they decide, and recovery is measured. On the first screen, click **Start a movement-support demonstration** and then use **Next step**. The step number shows at the top of the tour panel.

ID	Do this	What it tests	Why it matters	Expected output	Result
A1	Read step 1	Persona and goal chosen by the person	Plans start from the person's own choice, not a template (outcome 1)	Noura, 41, Khalifa City (synthetic); goal "Walk with family · 15 minutes · after the evening meal"	

ID	Do this	What it tests	Why it matters	Expected output	Result
A2	**Next step** (step 2)	Seven-day baseline	The first goal comes from her own numbers, not a population target (BR-08)	Seven daily step values (4,700, 4,317, 2,951, 3,501, 3,093, 3,476, 4,452) and their average, 3,784 steps	
A3	**Next step** (step 3)	One small action, plan version 1	A small, feasible start lowers the barrier	"Plan v1" with the action, minutes and timing, and the reason it was set	
A4	**Next step**, then **Check in for 5 days**	Daily check-ins	Keeping up is recorded day by day (done, partly, not today, pause)	The active-day count rises by 5	
A5	**Next step**, then **Simulate 7 missed days**	Lapse detection	The brief asks for support after setbacks	Status changes to a paused or lapse state; no streak is broken	
A6	**Next step** (step 6); optionally click **Why did my plan change?**	The plan shrinks without blame	Smaller, feasible restarts keep people going (outcome 2); plain-language reason (outcome 3)	"Plan v3": 5 minutes; reason states a restart; "progress is never reset"; no guilt words	
A7	**Next step**, then **Keep the lighter plan** (or **Undo**)	User approval	The person stays in control of every change (outcome 4)	Keep: "Plan v4 … You kept the lighter plan". Undo: the previous plan returns	
A8	**Next step**, then **Simulate 5 days back on plan**	Recovery time and persistence for her	Recovery is the measured outcome, not downloads	"Back on plan 4 day(s) after the setback started" and "planned days done X of Y"	
A9	**Next step** (Executive view opens); scroll to Outcome measures	Aggregate dashboard	Managers see measurable outcomes without seeing individuals (outcome 5)	300 synthetic people. Persistence weeks 1–4 / 5–8 / 9–12: 87% / 57% / 46%; completion 51%; setbacks 10.1 per 100 person-weeks; median recovery 5 days; 76% recovered within 14 days; 95% re-engaged	

ID	Do this	What it tests	Why it matters	Expected output	Result
A10	**Next step** (step 10)	Trace of one decision	Auditable decisions: every output can be traced to a versioned rule	Trace panel with five parts: rules and versions (for example "DEC-02 v1.1.0"), output shown, plan version, KPI events, CloudEvents with a W3C traceparent. The Azure card appears below	

Figures in A9 come from a fixed random seed, so every auditor sees the same numbers.

Part B: safety, privacy, measurement and Azure (about 15 minutes)

Part B tests what makes the service safe to pilot: it stops on red flags, never gives medical advice, keeps managers to group figures, and runs the same rules live on Azure. Click **Reset demo** before you start.

ID	Do this	What it tests	Why it matters	Expected output	Result
B1	Click **Change persona**, pick P03, and answer **Yes** to the warning-symptom safety question	Safety screen before any plan (DEC-02)	A coaching app must never push activity on someone with warning symptoms	Coaching stops; approved guidance with 998 (ambulance) is shown; no plan is offered	
B2	**Reset demo**, then **Change persona**; during onboarding look at the permissions step	Consent per purpose	The person controls sharing (outcome 4)	Reminders and sharing with a professional are separate tick boxes that the person chooses; coaching works without sharing	
B3	In the Participant view, on the guide card, click **Should I take medicine for my knee?**	Assistant boundary	The assistant must not interpret symptoms or advise treatment	"Refused", pointing to a clinician	
B4	Click **I walked 20 minutes today**	Assistant acts within rules	Bounded help, with the tier shown	"Acted within rules"; the check-in is logged	
B5	Click **Change persona**, then choose **العربية**	Arabic, right to left	Inclusion for Arabic speakers (outcome 4)	The whole participant view switches to Arabic, right to left	

ID	Do this	What it tests	Why it matters	Expected output	Result
B6	**Reset demo**, then **Start with P15 (wheelchair user, active minutes)**	Inclusion for wheelchair users	Steps exclude some people; minutes do not	The plan and check-ins use active minutes, not steps	
B7	Open **Programme**, **Workplace lead** and **Executive** views; in Executive click **Copy CSV**	Aggregates only, with small-group suppression	Employers and managers never see individuals (BR-23)	Only groups of 20 or more appear; smaller groups are hidden; the CSV has metric, group and value columns, with no names or IDs; caveat "Synthetic people…"	
B8	Open **Rules & proof**; on the Azure card click **Run a lapse-and-restart in Azure**, then **Run a red-flag check in Azure**	The same versioned rules run in Azure	Proves Azure is doing real work, not just hosting a file	Restart: browser and Azure both "ON_TRACK", fingerprint **20d73377**, "Same rules, same answer". Red flag: both "SAFETY_STOP", fingerprint **ad1bf607**. Rule versions and a traceparent are listed	
B9	On the same card click **Programme KPIs from Azure**	Measurement computed in Azure	Outcome figures are reproducible on the server	300 synthetic people. Above own baseline weeks 9–12: 53%; recovered within 14 days: 76%; still checking in at week 12: 69%; persistence weeks 9–12: 46%; median recovery 5 days	
B10	Click **Bounded AI wording**	AI cannot bypass rules	Responsible AI: rules decide, the model may only reword	"Demo fallback mode" with the approved message, technique, rationale, source and limitation (no generative model: wording stays reviewed templates by design)	

ID	Do this	What it tests	Why it matters	Expected output	Result
B11	Open /api/health	Live Azure service	Independent check of hosting and telemetry	JSON with "service": "Azure Container Apps", "region": "UAE North", "release": "v8.6", "aiMatching": "Azure OpenAI embeddings (bt-embed)", six rule versions, telemetry "Application Insights (OpenTelemetry)"	
B12	Open the link on a phone	Mobile use	Most participants will use a phone	The page fits the screen with no sideways scrolling; the tour works	
B13	In Rules & proof, on the Azure card, under AI action matching, click the example "It was too hot after work and the kids need me", then Suggest restart actions	A live, bounded Azure AI call	AI proposes alternative actions; rules decide; the person chooses	"Azure OpenAI embeddings (text-embedding-3-small, UAE North) ranked the approved library"; three approved actions (family walk and indoor walk among them), each with technique, why, when it may not work, and source	
B14	Same card: click "My knee hurts when I walk", then Suggest restart actions	Rules stop the AI on pain or red flags	The model never sees text that needs a clinician	"Stopped by the rules (Pain or injury) before any AI call"; no suggestions	
B15	On the live link, open the "Architecture & evidence" tab (or press Go on B15); open both PDFs under Downloads	Architecture & evidence matches the live service	Claims about Azure should be checkable, not just stated	Container Apps, Azure OpenAI and Application Insights rows say "Live · confirmed by /api/health now"; the pilot row says "Designed for pilot"; the deck and this guide open as PDFs	

For the B8 fingerprints: the browser and Azure each compute one from the day-by-day statuses and every decision. If even one rule version differed, the fingerprints would not match.

Does it address the problem statement?

Yes, for one behaviour (movement): the tests show the full start, keep-up and get-back loop working, safely and measurably, on Azure. Two outcomes are met and three are partly met; the gaps are stated, not hidden.

Desired outcome	Verdict	Proven by	What is not yet proven
1. Adaptive personal plans	Partly met	A1–A3, B6	One behaviour only (movement); sleep and nutrition are designed, not built
2. Continuous learning and re-engagement	Met (rule-based)	A4–A8	Machine-learning personalisation runs on synthetic data only and is never shown to participants
3. Evidence-based behaviour support	Partly met	A6, B3, B4, B10	Incentives and peer challenges are designed for the pilot; expert behaviour-change review pending
4. User autonomy and inclusion	Met (reviews pending)	A7, B2, B5, B6, B7	Arabic, cultural and accessibility reviews by native speakers and users are pending
5. Sustained outcome measurement	Partly met	A8, A9, B7, B9	All figures are synthetic; no real behaviour change or health improvement is shown. A partner-measured biomarker is designed for the pilot
Safety and trust	Met for the demo	B1, A10, B8, B11	Clinical safety review by a DoH-licensed partner pending

What the auditor can conclude: the mechanism works, is safe by design, is traceable, and runs on Azure in the UAE. Whether it changes real behaviour is the question the proposed 12-week supervised pilot answers.

One result to read carefully: in the synthetic comparison, a fixed goal keeps more people above their own baseline (77% against 53%). Adaptive restart recovers more setbacks (76% against 67%) and keeps more people persisting (46% against 32%). We report this openly; the pilot is designed to settle which works better for real people.

Limitations, troubleshooting and reporting

The demo is a synthetic competition prototype: it is not a diagnostic, treatment or clinical decision system, and it does not claim proven health improvement.

Known limitations

- All people and figures are synthetic; nothing is stored after the page closes.
- The demo has no sign-in by design; the pilot adds sign-in, consent records and personal data controls.
- The tour text is in English; the participant view switches to Arabic (B5).
- Azure OpenAI embeddings are live for action matching (B13). There is no generative model: message wording stays reviewed templates by design, and the rewording endpoint shows a labelled fallback (B10).
- Only one small container runs, so the first request after a restart can take up to a minute.

Troubleshooting

Symptom	What to do
Header says "Demo fallback mode"	Reload the page. Part A still works; Part B Azure tests (B8–B11) need the live link
A step looks different from this guide	Click **Reset demo** and start the tour again; results are deterministic
Page loads slowly the first time	Wait up to a minute; the container may be starting
Text is too small	Use **Larger text** or **High contrast** in the header

How to report results: fill the Result column (Pass, Fail or Note) in Parts A and B, or add a comment on any row. Please include the test ID, what you saw, and a screenshot if a result differs.

Contact: ZenAIDigital Pty Ltd · zenaidigital.com