DEVLINPECK.COM

AI Simulation Rollout Kit

Updated August 11, 2026

This kit pairs with my briefing on rolling out AI simulations. Print it, write in names and dates, and bring it to your kickoff. Every item maps to the roles, the 30-60-90 plan, the security review, the quality cadence, and the day-90 readout from the article.

Who owns what

Four roles, three of which you already have. The fourth is a reallocation, not a hire. Write in the names now: the rollout with a named owner on day 1 is the one that still has an owner on day 60.

Role	Who fills it	Name
Practice owner	An instructional designer on your team, usually the pilot champion	
Sponsor	You, the leader who greenlights the rollout	
SMEs	The best performers in the target role	
IT and security	Your existing function	

Practice owner

- [ ] Design scenarios and scoring criteria
- [ ] Read five to ten transcripts every week
- [ ] Tune scenarios weekly and ship the changes
- [ ] Keep the scenario registry current

Sponsor

- [ ] Pick the business metric and own the baseline
- [ ] Clear the security review
- [ ] Defend the budget and set usage guardrails
- [ ] Review the dashboard biweekly during the first cohort
- [ ] Recruit the target team's manager before launch

SMEs

- [ ] Supply the real conversations: the objections, the phrasing, what good actually sounds like
- [ ] Review the production scenario before the first cohort
- [ ] Join the quarterly refresh on what has changed in the field

IT and security

- [ ] Complete the one-time vendor review
- [ ] Set up SSO with automatic provisioning
- [ ] Sign off on the data flow

> **No named owner**
> the rollout that belongs to "the team" belongs to nobody by day 60. Formalize the practice owner's role in writing during the first 30 days.

The 30-60-90 plan

Days 1 to 30: productionize the pilot

- [ ] Finish the security review (question list in this kit)
- [ ] Set up SSO with automatic provisioning
- [ ] Formalize the practice owner's role in writing
- [ ] Rebuild the pilot scenario to production quality with SME input
- [ ] Define the scoring rubric and pass standard in writing
- [ ] Wire the simulation into the course or LMS path where the target audience already trains
- [ ] Set usage caps and alerts before anyone launches

> **Practice as a destination**
> if learners must leave their course or LMS to practice, most never do. Embed simulations where training already happens.

> **Unbounded usage budgets**
> simulation platforms price in usage-based AI credits, so you effectively pay per conversation, and voice consumes credits faster than text. Voice is also what makes the practice feel real, so budget for it rather than around it. Insist on per-simulation caps, usage alerts, and workspace-level credit visibility before the first cohort.

Days 31 to 60: first real cohort

- [ ] Launch to one full team with a manager who wants it
- [ ] Practice owner reads transcripts weekly and tunes the scenario
- [ ] Sponsor reviews the dashboard biweekly: completion, pass rates, attempts to mastery
- [ ] Collect the metric baseline you established in the pilot

> **Skipping the manager**
> if the target team's manager treats practice as optional, it is. Supervisor support has some of the strongest evidence among work-environment factors for whether training shows up on the job. Recruit the manager before the cohort launches.

Days 61 to 90: expand by use case, not by org chart

- [ ] Add the second and third scenario for the same audience before adding new audiences
- [ ] Keep scenarios in the sweet spot: character-based conversation practice, not vocational or safety walkthroughs
- [ ] Publish the first results readout against the business metric (template at the back of this kit)
- [ ] Set the renewal decision date: ____

Boiling the ocean
ten mediocre scenarios across five departments loses to three excellent ones for a single team. Expand from strength.

Weekly quality cadence

Simulations are living content. The maintenance model is a loop, not a launch.

Every week (practice owner)

- [ ] Read five to ten transcripts
- [ ] Flag scenarios that are too easy, characters that drift, and criteria that misfire
- [ ] Ship the fixes; with modern tools a saved change is live in about a minute, without republishing the course
- [ ] Update the scenario registry

Every quarter

- [ ] Ask top performers what has changed in the real conversations
- [ ] Update the affected scenarios the same week

Retirement check

- [ ] Archive any scenario nobody assigns or gets useful information from

> **Ignoring the transcripts**
> dashboards summarize; transcripts explain. Leaders who read a few transcripts a month make better calls than leaders who only watch pass rates.

> **Iteration speed is the quality system**
> course-shaped content gets one quality gate, review before publish. Simulation quality comes from the loop: read transcripts, tune the scenario, ship the change in minutes. If your tool makes iteration slow, no process will save quality.

Scenario registry

Every live simulation gets a named owner and a review date. Print this page again if you run more scenarios.

Scenario	Owner	Review date	Status (Live/Update/Archive)

Security review question list

Five questions cover what IT and security will ask. A clean vendor should be able to answer all five, in writing, in one email.

Question for the vendor	Vendor's written answer	Cleared (Y/N)
Are learner conversations used to train AI models, and under what terms do your model providers process conversation text?		
Where do transcripts and scores live, which subprocessors touch them, and how does data move between the simulation, our LMS, and our reporting stack?		
Do you hold SOC 2 or an equivalent attestation, with published security documentation our team can review without an NDA?		
Who can read learner transcripts and for how long? Can we set our own retention window?		
Do you support SSO through our identity provider, with automatic provisioning on first sign-in?		

> **Tip**
> transcripts are the raw material of your quality system, so the right retention answer is a deliberate window, not zero retention.

90-day results readout

Track three layers against the baseline you set in the pilot and put all three in the readout.

Example metrics per layer:

- **Usage:** completion rate, repeat attempts, practice minutes
- **Performance:** score trend against the rubric, attempts to mastery, pass rate against your own standard
- **Business:** the pilot's baseline metric, such as expert hours spent on live role plays, ramp time, or QA scores

Layer	Your metric	Baseline	Day-90 result	Decision it informs
Usage				Whether practice is happening where you embedded it
Performance				Whether the scenario is producing skill growth
Business				Whether to renew and expand

The pass standard is a design decision your practice owner makes for your context, not an industry constant. Treat pass-rate movement as feedback on the scenario and the cohort, not a comparison to an external benchmark.

Anchor the renewal decision on pass rate against your own standard, and ideally the metric the simulations were brought in to move. If the point was reducing the hours your experts spend running live role plays, this readout should show trainees performing as well or better while experts spend less time training them.

Anchor metric for the renewal decision: ____

Renewal decision date: ____

Decision: ____

Have any questions? Email me at devlin@devlinpeck.com