

DAILY ENGLISH READING

4 September 2026 - 30-minute active reading pack

TODAY'S QUESTION

How should an SP AI feature graduate from demo to production?

Difficulty: C1 | AI lifecycle | SP automation | shadow mode | validation

WHY SELECTED

A new open-access framework organizes AI evaluation into five phases: technical validation, operational robustness, controlled interaction, formal evidence, and real-world integration. Phase gates control promotion; drift, version updates or safety signals can trigger fallback and revalidation.

READING ORDER

0-3 min Preview: memorize the five phases and phase gates.
3-14 min Main: external validation, shadow mode, human-AI studies and fallback.
14-20 min FDA/EMA: context of use, risk, standards and lifecycle management.
20-25 min CDISC: apply the lifecycle to SDTM/ADaM, SAP and TFL automation.
25-30 min Output: build a five-phase validation plan for one SP AI feature.

SOURCES

Main: A five-phase evaluation framework for medical AI - npj Digital Medicine
<https://www.nature.com/articles/s41746-026-03155-7>
Support: FDA / EMA - Guiding Principles of Good AI Practice in Drug Development
Support: 2026 CDISC AI Innovation Challenge

BRIEF BACKGROUND

Phase 1: technical validation - core task performance.
Phase 2: operational robustness - external data, sites, conventions and edge cases.
Phase 3: controlled interaction - shadow mode and Human+AI evaluation.
Phase 4: formal evidence - quality, time, rework or other workflow outcomes.
Phase 5: real-world integration - monitoring, drift detection and controlled change.
SP path: benchmark -> external validation -> shadow -> supervised use -> monitoring.

KEY VOCABULARY

technical validation	技术验证	controlled core-task testing
operational robustness	运行稳健性	stability across realistic variation
controlled interaction validation	受控交互验证	test human-AI use before high stakes
clinical evidence validation	临床证据验证	formal evidence of meaningful benefit
real-world integration	真实世界集成	validation in routine operations
phase gate	阶段门槛	criterion required for promotion
fallback trigger	回退触发器	condition requiring earlier validation
shadow-mode testing	影子模式测试	run without controlling decisions
external validation	外部验证	test outside development setting
human-AI cooperation study	人机协作研究	evaluate joint human-AI performance
drift	漂移	change in data workflow or performance
version update	版本更新	change to model prompt code or rules
context of use	使用情境	defined role scope users and setting
risk-based performance assessment	基于风险的性能评估	validation scaled to failure risk
lifecycle management	全生命周期管理	monitor maintain and revalidate

USEFUL PHRASES

1. move from benchmark accuracy to lifecycle evidence
2. define explicit gates before production promotion
3. run the system in shadow mode before it influences decisions
4. re-enter an earlier phase after a material update
5. validate human-AI interaction, not the model alone
6. treat drift as a trigger for reassessment
7. align validation depth with the consequence of failure
8. collect evidence in the intended context of use
9. separate promotion criteria from rollback criteria
10. preserve an auditable record of every validated version

COMPREHENSION QUESTIONS

1. Why is benchmark performance insufficient for production approval?
2. How does robustness differ from controlled interaction validation?
3. Why is shadow mode useful in a regulated programming workflow?
4. Which changes should trigger fallback or revalidation?
5. How should context of use and failure risk affect validation depth?

RETELLING PROMPTS

30 sec: name the five phases and explain phase gates.

45 sec: benchmark -> external validation -> shadow -> supervised use -> monitoring.

60 sec: explain why a material update can invalidate earlier evidence.

5-MINUTE SPEAKING / OUTPUT TASK

Minute 1: choose one SP AI feature.

Minutes 2-3: define one test and one passing gate for each phase.

Minute 4: define drift, version, standards and safety fallback triggers.

Minute 5: give a production go/no-go recommendation.

ONE SENTENCE TO KEEP

Production readiness is accumulated evidence that a specific validated version performs acceptably with specific users, in a specific workflow, under a defined context of use.