چارچوب ارزیابی توانمندی AI / GenAI

چارچوبی پژوهش‌محور برای سازمان‌های DevOps در صنایع تحت‌نظارت

پیش‌نویس جهت بازبینی · تهیه‌شده با Claude Code

فهرست مطالب

1. خلاصه مدیریتی
2. یافته‌های پژوهشی: سایر شرکت‌ها و چارچوب‌ها چه می‌کنند
3. الگوبرداری از شرکت‌های فناوری
4. معیارسنجی بانکداری و خدمات مالی تحت نظارت
5. مقایسه ساختارهای سطح‌بندی رایج در بازار
6. ابعاد رایج مهارتی AI / GenAI
7. ابعاد مهارتی صنعت بانکداری / صنایع تحت نظارت
8. مدل ارزیابی پیشنهادی و توجیه آن
9. نردبان پیشنهادی قابلیت AI / GenAI در ۱۰ سطح
10. نگاشت ۱۰ سطح به مدل‌های رایج ۴/۵/۶ سطحی
11. نگاشت به خانواده‌های شغلی و سطوح ارشدیت
12. راهنمای ارزیابی ویژه هر تیم
13. سنجه امتیازدهی منابع انسانی (HR Scoring Rubric)
14. فرایند ارزیابی مبتنی بر شواهد
15. آزمون‌های عملی بر حسب نقش
16. راهنمای ارتقا و مسیر شغلی
17. سطوح هدف ۴ماهه و ۱۲ماهه
18. برنامه آموزشی بر اساس سطح
19. ملاحظات حاکمیت و ریسک
20. برنامه پیاده‌سازی منابع انسانی (HR)
21. قالب‌ها (Templates)
22. ده اقدام نخست پیشنهادی
23. منابع و مآخذ

چارچوبی پژوهش‌محور و متناسب با صنایع تحت نظارت، برای تیم‌های Observability، Platform، SRE، DevOps، Release، Security، QA، DataOps، Service Desk و مدیریت فنی.

شیوه‌ی مطالعه‌ی این گزارش: بخش‌های 2 تا 7 بهترین رویه‌های پژوهش‌شده هستند؛ بخش‌های 8 تا 22 توصیه‌های ترکیب‌شده و یکپارچه برای سازمان شما به شمار می‌روند. نتیجه‌گیری‌های استنباطی به‌صراحت برچسب‌گذاری شده‌اند.

---

1. خلاصه مدیریتی

توصیه: به‌کارگیری یک **نردبان ترکیبی ۱۰ سطحی تسلط بر هوش مصنوعی (L0–L10)** در کنار یک **روبریک ۱۶ بُعدی با امتیازدهی ۰ تا ۵**، به‌گونه‌ای که **کنترل‌های نظارتی از همان ابتدا اعمال شوند** (مدیریت داده، ریسک مدل و دروازه‌های تأیید/بازگردانی از سطح L2 به بعد ارزیابی می‌شوند، نه آنکه در بالاترین سطوح به‌صورت الحاقی افزوده شوند). ارزیابی **مبتنی بر شواهد** است (مصنوعات کاری، بازبینی همتایان و کار اثبات‌شده — نه صرفاً خودارزیابی)، همراه با **اهداف ۴ ماهه و ۱۲ ماهه** برای هر نقش.

چرا کارآمد است: این رویکرد رفتار (نردبان — اینکه فرد چه کاری را به‌صورت ایمن می‌تواند انجام دهد) را از جزئیات قابلیت (روبریک — اینکه چقدر خوب، در ۱۶ بُعد) جدا می‌کند؛ بنابراین واحد منابع انسانی برچسب‌های سطح‌بندی شفافی به دست می‌آورد و در عین حال مدیران از یک رد امتیازدهی قابل دفاع و قابل ممیزی برخوردار می‌شوند. اعمال زودهنگام کنترل‌های انطباق با واقعیت یک بانک تحت نظارت همخوانی دارد: ریسک هنگامی بیشترین مقدار را دارد که کاربران تازه‌کار با داده‌های واقعی سروکار دارند، نه زمانی که معماران ارشد سکوهای دارای حاکمیت می‌سازند.

پاسخ‌های کلیدی:

پرسش	پاسخ
(الف) دیگران چه می‌کنند	بیشتر چارچوب‌های عمومی از **۴ تا ۶ مرحله گسترده** استفاده می‌کنند (برای نمونه: آگاهی ← پذیرش ← یکپارچه‌سازی ← رهبری). تعداد اندکی مختص بانک‌ها هستند و تقریباً هیچ‌کدام دروازه‌های ریسک مدل/انطباق را از ابتدا اعمال نمی‌کنند. (برداشت‌شده از مطالب عمومی.)
(ب) آیا ۱۰ سطح مناسب است؟	**برای کاربرد ما مناسب است** — اما تنها به این دلیل که L0–L10 هم‌زمان هم به‌عنوان مسیر رشد فردی و هم به‌عنوان نقشه قابلیت سازمانی عمل می‌کند. برای سطح‌بندی صرفاً منابع انسانی، این ۱۱ نقطه به **۵ باند منابع انسانی** (در ادامه) فروکاسته می‌شوند تا از دقت کاذب پرهیز شود.
(ج) مدل پیشنهادی منابع انسانی	برای گفت‌وگوهای مربوط به جبران خدمات/ارتقا از **جمع‌بندی ۵ باندی** استفاده کنید؛ و برای برنامه‌ریزی توسعه و شواهد ممیزی، **مجموعه کامل L0–L10 به‌همراه روبریک ۱۶ بُعدی** را حفظ کنید.
(د) نگاشت به ۱۰ تیم	هر تیم یک **باند هدف بر اساس نقش** دریافت می‌کند، نه یک سطح یکسان برای همه — بیشتر تیم‌ها در محدوده L3 تا L6 قرار می‌گیرند، تیم‌های سکو/خودکارسازی هدف L7 تا L9 را دنبال می‌کنند و یک هسته کوچک حاکمیتی در L10 جای می‌گیرد.

جمع‌بندی باندهای منابع انسانی (برای تصمیم‌های سطح‌بندی):

باند	سطوح نردبان	معنای ساده
A - بنیادی	L0–L2	آگاه نسبت به GenAI و قواعد داده؛ هنوز کاربری قابل اتکا نیست
B - کاربر عملی	L3–L4	استفاده روزمره ایمن و مؤثر همراه با راستی‌آزمایی
C - یکپارچه‌ساز	L5–L6	هوش مصنوعی را در جریان‌های کاری جای می‌دهد؛ سطح تیم را ارتقا می‌دهد
D - سازنده	L7–L9	خودکارسازی‌ها، راهکارها و سکوهای دارای حاکمیت را می‌سازد
E - رهبر	L10	مالک راهبرد هوش مصنوعی سازمانی، حاکمیت و ریسک نظارتی است

نتیجه نهایی: این چارچوب آماده تصمیم‌گیری است. واحد منابع انسانی سطح‌بندی را بر پایه ۵ باند پیش می‌برد؛ مدیران توسعه و ممیزی را بر اساس نردبان ۱۱ سطحی و روبریک ۱۶ بُعدی اجرا می‌کنند؛ و بانک به کنترل‌های ریسک زودهنگام و قابل اثبات و نیز پیشرفت سنجش‌پذیر ۴ ماهه/۱۲ ماهه برای هر تیم دست می‌یابد.

2. یافته‌های پژوهشی: سایر شرکت‌ها و چارچوب‌ها چه می‌کنند

این بخش جمع‌بندی می‌کند که شرکت‌های فناوری، مؤسسات مشاوره، فروشندگان AIOps، نهادهای استانداردگذار عمومی و بانک‌های تحت نظارت، امروز مهارت‌های AI/GenAI را چگونه ارزیابی می‌کنند. ابتدا بهترین رویه‌های مستندِ پژوهش‌شده بیان می‌شود؛ تفسیر ما به‌صورت جداگانه برچسب‌گذاری شده است.

بهترین رویه‌های پژوهش‌شده (برگرفته از منابع عمومی)

چگونه مهارت‌های AI را ارزیابی می‌کنند. هیچ شرکت فناوری بزرگ، مؤسسه مشاوره یا بانکی به‌صورت عمومی یک نردبان منابع انسانی (HR) داخلی منتشر نمی‌کند که شایستگی‌های مشخص AI را به سطوح مهندسی نگاشت کند. آنچه آن‌ها به‌جای آن منتشر می‌کنند، در پنج الگوی قابل‌استفاده مجدد جای می‌گیرد:

رویکرد	چه کسی از آن استفاده می‌کند (نمونه‌های عمومی)	در واقع چه چیزی را می‌سنجد
مدل‌های بلوغ (در سطح سازمان/تیم)	Gartner (۵ مرحله)، پذیرش Copilot منتشرشده توسط شرکای Microsoft (۵ سطح)، پذیرش سازمانی Anthropic (۴ مرحله)، خوشه‌های عملکردی Google/DORA (۴ خوشه)	اینکه سازمان تا چه اندازه پیشرفت کرده، نه درجهٔ یک فرد
نردبان‌های مسئولیت/شایستگی	SFIA 9 (۷ سطح)، e-CF / EN 16234-1 (۵ سطح، نگاشت‌شده به EQF 3-8)	مسئولیت فردی بر اساس خودگردانی، اثرگذاری، پیچیدگی و دانش
رده‌های گواهینامه	AWS، Google Cloud، IBM (Foundational ← Associate ← Professional/Specialty)	دانش آزمون‌شده از طریق آزمون‌های تحت نظارت
امتیازدهی حوزه‌های قابلیت	«rewired McKinsey» (حوزه‌ها، نه یک نردبان شماره‌گذاری‌شده)، CNCF Platform Engineering (بلوغ به ازای هر جنبه)	قوت در حوزه‌های نام‌گذاری‌شده به‌جای یک عدد واحد
دروازه‌های چرخهٔ حیات / خودگردانی	ServiceNow (build ← benchmark ← monitor)، GitLab Duo (دروازه‌های Read/Write/Delete)، رده‌های AIOps در PagerDuty	اینکه چقدر به یک سیستم یا فرد اعتماد می‌شود که اقدام کند

از چه مراحلی استفاده می‌کنند. تعداد مراحل عمدتاً حول ۴ و ۵ متمرکز است. DORA و Anthropic از ۴ استفاده می‌کنند؛ Gartner، پذیرش Microsoft و بیشتر مقیاس‌های بلوغ به سبک ITIL از ۵ استفاده می‌کنند. SFIA برای مسئولیت فردی از ۷ و e-CF از ۵ بهره می‌برد. نردبان‌های خودگردانی فروشندگان بین ۳ تا ۵ متغیر است (Atlassian ۳: Rovo؛ AIOps تجمیع‌شده: تقریباً L0-L5). هیچ تعداد استاندارد واحدی در صنعت وجود ندارد.

چه ابعادی را ارزیابی می‌کنند. در میان همهٔ منابع، ابعاد بر این موارد همگرا می‌شوند: سواد AI/GenAI؛ سواد prompt و ابزار؛ عمق فنی (ML/MLOps، مهندسی داده)؛ RAG/grounding؛ طراحی agent و استفاده از ابزار؛ AI مسئولانه (انصاف، شفافیت، hallucination، حریم خصوصی)؛ پیامدهای بهره‌وری؛ و سواد حاکمیت/ریسک مدل. استانداردهای عمومی یک لایهٔ امنیتی (OWASP LLM Top 10 2025، با prompt injection در رتبهٔ نخست) و یک لایهٔ AI قابل‌اعتماد (ویژگی‌های NIST AI RMF) را نیز می‌افزایند.

چه شواهدی را می‌پذیرند. قوی‌ترین رویهٔ عمومی، شواهد عینی را بر خودارزیابی ترجیح می‌دهد:

نوع شواهد	نمونه منابع
قبولی در گواهینامهٔ تحت نظارت	AWS، Google Cloud، IBM
تله‌متری درون‌محصول	نرخ پذیرش پیشنهاد، درصد کد نوشته‌شده توسط AI، میزان استفادهٔ فعال
مطالعات کنترل‌شده	RCT گیت‌هاب با ~۵۵٪ سرعت بیشتر؛ آزمایش سازمانی Accenture

نوع شواهد	نمونه منابع
سنجه‌های تحویل	تحرک DORA / SPACE / DevEx (داده‌های عینی CI/CD و رخدادها)
مصنوعات حاکمیتی	فهرست‌های موجودی مدل، ارزیابی‌های AI مسئولانه، factsheetها، شواهد کنترلی (IBM watsonx.governance، PwC، KPMG)
مجموعه‌داده‌های Benchmark/ ground-truth	ServiceNow Data Kit؛ playbookهای اعتبارسنجی‌شده با رخداد در PagerDuty

چگونه مهارت‌ها را به ارشدیت، استخدام و آموزش گره می‌زنند. SFIA و e-CF روشن‌ترین نمونه‌های عمومی برای گره‌زدن مهارت به ارشدیت از طریق سطوح مسئولیت هستند. گواهینامه‌ها استخدام و واجد شرایط بودن برای نقش را مشروط می‌کنند (فروشندگان cloud). آکادمی‌های یادگیری (Deloitte AI Academy، Accenture LearnVantage، Capital One Tech College / برنامهٔ ۱۶۰ساعتهٔ MLETP) پیشرفت آموزشی را ساختارمند می‌کنند. EU AI Act اکنون یک تعهد سواد AI برای کارکنان وضع کرده و سواد پایه را به‌جای یک گزینهٔ اختیاری، به یک الزام انطباقی تبدیل می‌کند.

چگونه با حاکمیت و ریسک مدل برخورد می‌کنند. حاکمیت به‌عنوان یک مسیر موازی در نظر گرفته می‌شود، نه یک حاشیه. NIST AI RMF (Govern، Map، Measure، Manage) بر کمیته‌های ریسک بانک و سه خط دفاعی نگاشت می‌شود. ISO/IEC 42001 تنها سیستم مدیریت AI قابل‌گواهی‌شدن است (۳۸ کنترل در پیوست A). بانک‌ها رهنمود ریسک مدل SR 11-7 را با تمام قوت بر AI/ML/GenAI اعمال می‌کنند و اعتبارسنجی مستقل، تأیید شورا (مانند HSBC AI Review Councils) و استفادهٔ اجباری از دروازهٔ مجاز (JPMorgan LLM Suite، Goldman GS AI Assistant) را الزامی می‌کنند. فروشندگان اقدامات را بر اساس ردهٔ ریسک مشروط می‌کنند (GitLab Duo: Read = همیشه مجاز، Write/Delete = همیشه نیازمند پرسش).

تفسیر (استنتاج‌شده از منابع عمومی)

- هرگونه نگاشت شایستگی‌های AI به سطوح مهندسی مشخص در منابع انسانی، **استنتاجی** است؛ هیچ شرکتی چنین نگاشتی را منتشر نمی‌کند، بنابراین یک نردبان داخلی باید با پیوندزدن شایستگی‌های AI بر یک چارچوب مسئولیت موجود (SFIA/e-CF) ساخته شود.
- شواهد عمومی دربارهٔ ارزش AI **آمیخته است و باید هیجان را تعدیل کند**: DORA در سال ۲۰۲۴ دریافت که پذیرش AI با کاهش پایداری تحویل همبسته بوده است (دسته‌های بزرگ‌تر، churn بیشتر)، و پژوهشی که به MIT NANDA استناد شده گزارش داد که ~۹۵٪ از پایلوت‌های GenAI هیچ اثر مالی قابل‌اندازه‌گیری نشان نداده‌اند که این به مسائل گردش‌کار و انگیزه نسبت داده شده، نه کیفیت مدل. به همین دلیل ارزیابی باید پیامدهای اندازه‌گیری‌شده را پاداش دهد، نه شیفتگی به ابزار را.
- بهترین رویه برای یک سازمان تحت نظارت آن است که یک **مسیر بلوغ حاکمیتی را به‌موازات** یک مسیر قابلیت اجرا کند، تا رشد مهارت فردی هرگز از کنترل‌ها پیشی نگیرد (توصیهٔ ترکیب‌شده).

---

3. الگوبرداری از شرکت‌های فناوری

شرکت‌های فناوری و Cloud

شرکت‌های بزرگ فناوری و ارائه‌دهندگان Cloud هم‌زمان هم تولیدکننده ابزارهای GenAI هستند و هم پُرمطالعه‌ترین به‌کارگیرندگان آن. مستندات عمومی این شرکت‌ها به‌طور غیرمعمولی برای الگوبرداری مفید است، زیرا سه چیزی را که یک سازمان خدمات مالی تحت نظارت باید از هم جدا نگه دارد تفکیک می‌کند: (1) **گواهینامه‌ها/مهارت‌افزایی** که توانمندی فردی را می‌سنجد، (2) چارچوب‌های **اندازه‌گیری بهره‌وری/به‌کارگیری** که می‌سنجد آیا توانمندی به نتیجه تبدیل می‌شود یا خیر، و (3) چارچوب‌های **حاکمیت/ریسک مدل** که استفاده ایمن از AI را تضمین می‌کنند. الگوی زیر کاملاً از منابع عمومی استخراج شده است.

(a) این سازمان‌ها در عمل برای ارزیابی و مهارت‌افزایی توانمندی AI چه می‌کنند

سازمان	آنچه انجام می‌دهند (عمومی)	سازوکار اصلی
Microsoft	یک **Skilling Center** برای Copilot به‌همراه مسیرهای یادگیری ساختاریافته، رویدادهای داخلی «Copilot Expo»/مهارت‌افزایی مجازی، و مدل Center of Excellence (CoE) را برای پیشبرد و سنجش به‌کارگیری اجرا می‌کند.	یادگیری مبتنی بر نقش + حاکمیت CoE
GitHub	پژوهش‌های بهره‌وری داوری‌شده (RCTها) منتشر می‌کند و تله‌متری درون‌محصولی (نرخ پذیرش، پیشنهادهای استفاده‌شده) عرضه می‌کند که مشتریان برای سنجش ارتقای توانمندی از آن بهره می‌برند.	اندازه‌گیری مبتنی بر شواهد
Google / Google Cloud	پژوهش **DORA State of DevOps** و گواهینامه‌های Cloud Skills Boost / Google Cloud (شامل مسیرهای یادگیری Generative AI) را تألیف می‌کند. DORA اکنون به‌صراحت اثر AI بر تحویل را مطالعه می‌کند.	پژوهش صنعتی + گواهینامه
AWS	برنامه‌های یادگیری **Skill Builder** + گواهینامه‌های مبتنی بر نقش: **AI Practitioner (AIF-C01)** و **Generative AI Developer – Professional**؛ آزمایشگاه‌های عملی Bedrock/Amazon Q.	نردبان گواهینامه‌ای لایه‌بندی‌شده
OpenAI / Anthropic	**راهنماها/playbookهای به‌کارگیری** سازمانی (چارچوب چهارمرحله‌ای به‌کارگیری Anthropic؛ راهنمای سازمانی OpenAI) و مستندات بهترین‌روش prompt/agent منتشر می‌کنند؛ برای توانمندسازی با شرکت‌های مشاوره (Accenture، BCG، McKinsey) همکاری می‌کنند.	playbookهای به‌کارگیری + توانمندسازی شریک
IBM	گواهینامه **watsonx.governance Lifecycle Advisor** و آموزش حاکمیتی که مستقیماً به نقش‌های ریسک مدل/انطباق گره خورده است.	اعتبارنامه‌های ویژه حاکمیت
Meta / Netflix / Uber / LinkedIn	افشاگری‌های مهندسی-بلاگ درباره ابزارهای داخلی AI و اندازه‌گیری بهره‌وری توسعه‌دهنده؛ هیچ «نردبان AI» رسمیِ عمومی ندارند.	ابزارهای داخلی + معیارهای DevEx (استنتاج‌شده از مواد عمومی)

(b) ساختارهای سطح/بلوغ که به‌کار می‌برند

ساختار	مالک	شکل
بلوغ پنج‌سطحی به‌کارگیری Copilot (اغلب چنین نقل می‌شود: Initial ← Experimenting ← Operational ← Strategic ← Transformational)	اکوسیستم Microsoft / شرکا	بلوغ سازمانی، نه رتبه فردی (منتشرشده توسط شرکا؛ به‌عنوان استنتاج‌شده علامت‌گذاری شود)
مدل بلوغ Agentic AI (امنیت و حاکمیت)	Microsoft Learn	نردبان بلوغ حاکمیتی
رده‌های گواهینامه (Associate ← Foundational ← Professional/Specialty)	AWS، Google Cloud، IBM	نردبان توانمندی فردی
چارچوب چهارمرحله‌ای به‌کارگیری سازمانی (تقریباً: explore ← transform ← scale ← pilot)	راهنمای سازمانی Anthropic	مراحل به‌کارگیری سازمانی
خوشه‌های عملکرد DORA (High ← Medium ← Low ← Elite)	Google/DORA	عملکرد تحویل تیم

توجه: ارائه‌دهندگان Cloud **مدل‌های بلوغ را به سازمان** و **رده‌های گواهینامه را به فرد** اعمال می‌کنند. یک سازمان خدمات مالی باید این تفکیک را بازتاب دهد، نه آنکه آن‌ها را در هم بیامیزد.

(c) ابعاد مهارتی که ارزیابی می‌شوند

بُعد	جایی که نمایان می‌شود
استفاده مناسب / انتخاب موارد کاربرد	حوزه اصلی AWS AI Practitioner («چه زمانی از AI/ML/GenAI استفاده کنیم»)
مدل‌های پایه و prompt engineering	مستندات AWS، Google Cloud، OpenAI/Anthropic
Retrieval-Augmented Generation (RAG) و grounding	Amazon Q Business، Bedrock، IBM watsonx
طراحی agent / tool-use	Microsoft Copilot Studio، راهنمای agent در Anthropic/ OpenAI
AI مسئولانه: سوگیری، انصاف، توهم (hallucination)، حریم خصوصی	Bedrock Guardrails، Microsoft Responsible AI، IBM watsonx.governance
نتایج بهره‌وری (توان عملیاتی، کیفیت، جریان)	DORA، SPACE، DevEx
سواد حاکمیت / ریسک مدل	IBM watsonx Governance Lifecycle Advisor

(d) شواهدی که می‌طلبند

نوع شواهد	منبع	آنچه اثبات می‌کند
قبولی در آزمون گواهینامه با نظارت (proctored)	AWS، Google Cloud، IBM	دانش فردی آزموده‌شده
تله‌متری درون‌محصولی (نرخ پذیرش پیشنهاد، درصد کد نگاشته‌شده توسط AI، استفاده فعال)	داشبوردهای GitHub Copilot، Microsoft Copilot	استفاده واقعی، نه خوداظهاری
آزمایش‌های تصادفی‌سازی‌شده کنترل‌شده / شبه‌آزمایش‌ها	GitHub (≈55% تکمیل سریع‌تر وظیفه در RCT سرور HTTP با JavaScript؛ آزمایش سازمانی Accenture)	ارتقای بهره‌وری علّی
حرکت معیارهای DORA / SPACE / DevEx	پژوهش Google/DORA	تحویل و تجربه در سطح تیم
معیارهای نتیجه‌محور برنامه به‌کارگیری (اثر کسب‌وکاری، کیفیت، نرخ به‌کارگیری)	راهنمای CoE مایکروسافت	نتیجه در برابر فعالیت

هشدار مهم برای یک سازمان تحت نظارت: شواهد عمومی آمیخته است، نه یکدست مثبت. پژوهش DORA سال 2024 گوگل دریافت که به‌کارگیری AI **با کاهش پایداری تحویل همبسته بود** (اندازه دسته‌های بزرگ‌تر، churn بیشتر)، و رقمی که توسط MIT NANDA نقل شده، حدود 95% از آزمایش‌های GenAI را در نشان‌دادن اثر مالی قابل‌اندازه‌گیری ناموفق دانست — شکست‌هایی که به **یکپارچه‌سازی جریان کار و انگیزه‌ها نسبت داده شد، نه کیفیت مدل**. بنابراین شواهد توانمندی باید اثبات مهارت فردی را با تحویل تیمی + معیارهای ریسک جفت کند، و هرگز یکی به‌تنهایی کافی نیست.

(e) چگونگی پیوند توانمندی AI با ارشدیت / ترفیع / استخدام / آموزش

پیوند	الگوی عمومی	اطمینان
غربالگری استخدام	گواهینامه‌های ارائه‌دهندگان (IBM، Google Cloud GenAI، AWS AI Practitioner watsonx Governance) به‌عنوان سیگنال‌های شناخته‌شده استخدام به‌رسمیت شناخته می‌شوند.	بالا (عمومی)
تخصص‌گرایی نقش	مسیرهای متمایز پدیدار می‌شوند: کاربرِ AI (گواهینامه AI Practitioner) در برابر مهندس سازنده AI (GenAI Developer Professional) در برابر مشاور حاکمیت AI (IBM).	بالا (عمومی)
ترفیع / سطح‌بندی	این شرکت‌ها «پله‌های AI» داخلی را روی نردبان‌های مهندسی خود منتشر **نمی‌کنند**. هر ادعایی مبنی بر اینکه «سطح X مستلزم مهارت AI نوع Y است» در داخل Microsoft/ Google/Meta **(استنتاج‌شده از مواد عمومی)** است و نباید به‌عنوان واقعیت ارائه شود.	استنتاج‌شده — به‌وضوح علامت‌گذاری شود
مسیر آموزش	به‌شدت صریح: مسیرهای یادگیری ← گواهینامه ← آزمایشگاه‌های عملی (Skill Builder، Skilling Center، Cloud Skills Boost).	بالا (عمومی)

(f) مدیریت ایمنی / حاکمیت / امنیت / ریسک مدل

حوزه کنترل	سازوکار عمومیِ نماینده
ایمنی محتوا و خروجی	Amazon **Bedrock Guardrails** (فیلتر محتوا، مسدودسازی موضوع، کاهش توهم/PII)؛ اصول Microsoft Responsible AI
شناسایی ریسک پیش از استقرار	«Responsible AI risk radar» مایکروسافت؛ مدل بلوغ Agentic AI
چرخه حیات مدل و ریسک مدل	**IBM watsonx.governance** — ثبت موارد کاربرد، رهگیری چرخه حیات، factsheetها/متادیتا، پایش انصاف/سوگیری/drift؛ با OpenPages Model Risk Governance یکپارچه می‌شود (مستقیماً مرتبط با رژیم‌های ریسک مدل بانکی مانند نظارت به‌سبک SR 11-7)
نگهبانی داده / ریسک ارائه‌دهنده	بررسی عمومی سیاست‌های نگه‌داشت داده (مثلاً مناقشه سیاست 30روزه Anthropic) نشان می‌دهد سازمان‌ها اکنون **promptها، خروجی‌ها، logها و traceها را به‌عنوان داده تحت حاکمیت** در نظر می‌گیرند
حاکمیت Center of Excellence	مدل CoE مایکروسافت سیاست، استانداردهای استفاده و اندازه‌گیری را متمرکز می‌کند

جمع‌بندی برای الگوبرداری: الگوی معتبر و قابل‌انتقال از این دسته یک **مدل سه‌لایه** است — (1) گواهینامه‌های فردیِ لایه‌بندی‌شده (professional ← foundational ← متخصص حاکمیت)، (2) اندازه‌گیری نتیجه از طریق DORA/ SPACE/DevEx به‌علاوه تله‌متری ابزار، و (3) یک ستون فقرات حاکمیت/ریسک مدل (guardrails + حاکمیت چرخه حیات + CoE). آنچه به‌طور عمومی در دسترس **نیست** — و نباید جعل شود — هر نردبان منابع انسانی داخلیِ شرکتی است که توانمندی‌های مشخص AI را به سطوح مهندسی مشخص نگاشت دهد.

شرکت‌های مشاوره / پژوهش / مشاوران راهبردی

شرکت‌های مشاوره، پژوهش و مشاوران راهبردی شیوه تفکر بیشتر سازمان‌های بزرگ (از جمله بانک‌های تحت نظارت) درباره توانمندی AI را شکل می‌دهند. آن‌ها هم‌زمان دو نقش بازی می‌کنند: ارزیابی‌های بلوغ AI و برنامه‌های مهارت‌افزایی را به مشتریان می‌فروشند، و برنامه‌های گسترده فلوئنسی (روانی) داخلی AI را برای کارکنان خود اجرا می‌کنند. هر دو الگوهای مفیدی هستند. نکته کلیدی برای یک سازمان تحت نظارت: این شرکت‌ها چارچوب‌ها و مراحل بلوغ منتشر می‌کنند، اما عموماً روبریک‌های داخلیِ رده‌به‌رده منابع انسانی را که مهارت AI را به باندهای پرداخت یا دروازه‌های ترفیع مشخص نگاشت

می‌دهد منتشر نمی‌کنند. جایی که نگاشتی تلویحی است اما رسماً منتشر نشده، با **(استنتاج‌شده از مواد عمومی)** علامت‌گذاری شده است.

(a) این شرکت‌ها در عمل برای ارزیابی و مهارت‌افزایی توانمندی AI چه می‌کنند

شرکت	رویکرد توانمندسازی (عمومی)	برنامه نیروی کار داخلی
Deloitte	چارچوب‌بندی «AI fluency» در سراسر نیروی کار؛ راهبرد مهارت‌افزایی/بازمهارت‌سازی و بازطراحی شغل را از طریق پژوهش State of AI in the Enterprise به مشتریان می‌فروشد.	**Deloitte AI Academy** — مهارت‌افزایی داخلی ساختاریافته برای کارکنان فنی و غیرفنی؛ مسیرهای یادگیری و گواهینامه‌های مبتنی بر نقش.
McKinsey	توانمندسازی AI به‌عنوان یک practice رسمی؛ تفکر مدل عملیاتی «rewired» که استعداد، فناوری و به‌کارگیری را جفت می‌کند. **QuantumBlack** بازوی AI/تحلیل است که کهن‌الگوهای نقش فنی (data scientist، ML engineer، data engineer، طراح، مترجم) را تعریف می‌کند.	آکادمی‌های داخلی و خط‌لوله‌های استعداد «Forward»/QuantumBlack Labs؛ استفاده گسترده از کهن‌الگوهای نقش.
Gartner	پژوهش و مشاوره: **AI Maturity Model** پرکاربرد (5 مرحله) و Hype Cycle را منتشر می‌کند؛ درباره مدل عملیاتی، حاکمیت AI و شکاف‌های مهارتی مشاوره می‌دهد.	کاربردی ندارد (مدل مشاوره‌ای، نه نیروی کار تحویل به همان معنا).
Accenture	«LearnVantage» در مقیاس بزرگ و تحول استعداد AI را می‌فروشد؛ راهبرد استعداد حول «بازآفرینی (reinvention)» و تیم‌های انسان + AI.	بازمهارت‌سازی داخلی انبوه در مقیاس (صدها هزار نفر کارمند) با یادگیری AI مبتنی بر نقش.
BCG	توانمندی تحول AI از طریق **BCG X** (واحد ساخت/طراحی فناوری آن)؛ قاعده «10-20-70» (10% الگوریتم‌ها، 20% فناوری/داده، 70% افراد/فرایند) جهت سرمایه‌گذاری در توانمندی را تعیین می‌کند.	آکادمی‌های داخلی AI؛ کهن‌الگوهای فنی در BCG X.
PwC	جعبه‌ابزار **Responsible AI** و بازمهارت‌سازی نیروی کار («New skills. New world.»)؛ زاویه قوی حاکمیت/اطمینان‌بخشی.	مهارت‌افزایی AI در سطح کل شرکت و برنامه‌های citizen-developer.
KPMG	چارچوب **AI governance / Trusted AI** و توانمندی «digital workforce»/خودکارسازی؛ منظر کنترل و اطمینان‌بخشی.	مهارت‌افزایی داخلی AI گره‌خورده به چارچوب اعتماد آن.
Forrester	پژوهش و مشاوره: مدل‌های بلوغ AI و به‌کارگیری AI؛ سازمان‌ها را بر اساس وضعیت به‌کارگیری بخش‌بندی می‌کند (مثلاً experimenters در برابر transformers) **(استنتاج‌شده از مواد عمومی — برچسب‌های دقیق رده‌ها بسته به گزارش Forrester متفاوت است)**.	کاربردی ندارد (مدل مشاوره‌ای).

(b) ساختارهای سطح/بلوغ که به‌کار می‌برند

چارچوب	مراحل / سطوح	آنچه موجب پیشروی به مرحله بعد می‌شود
Gartner AI Maturity Model	1. Awareness ← 2. Active ← 3. Transformational ← 4. Systemic ← 5. Operational	حرکت از گفت‌وگوی غیررسمی (1) به آزمایش‌ها (2)، به دست‌کم یک سیستم تولیدی با تأمین مالی و تخصص درون‌سازمانی (3)، به جاگیری AI در سراسر فرایندها (4)، به بازشکل‌دهی مدل کسب‌وکار توسط AI (5).
بلوغ / به‌کارگیری AI در Forrester	رده‌های به‌کارگیری از آزمایش اولیه تا استفاده مقیاس‌یافته/تحول‌آفرین **(استنتاج‌شده از مواد عمومی)**	گستره استقرار، انضباط حاکمیتی و ارزش کسب‌وکاری قابل‌اندازه‌گیری.

چارچوب	مراحل / سطوح	آنچه موجب پیشروی به مرحله بعد می‌شود
«rewired» / توانمندی در McKinsey	کمتر یک نردبان شماره‌گذاری‌شده؛ در سراسر حوزه‌ها ارزیابی می‌شود — راهبرد، استعداد، مدل عملیاتی، فناوری/داده، به‌کارگیری و مقیاس‌دهی، حاکمیت	توانمندی به‌ازای هر حوزه امتیازدهی می‌شود؛ مقیاس‌دهی و به‌کارگیری دشوارترین دروازه‌ها هستند.
BCG 10-20-70	یک مدل سطحی نیست — یک اصل تخصیص سرمایه‌گذاری است	تأکید می‌کند که 70% ارزش از افراد، فرایند و به‌کارگیری می‌آید، نه خودِ مدل.
fluency در Deloitte	«AI fluency» در سراسر نیروی کار؛ چارچوب‌بندی applied ← foundational ← advanced **(استنتاج‌شده از مواد عمومی)**	گستره (چه میزان از نیروی کار روان هستند) به‌علاوه عمق (بازطراحی شغل حول AI).

این مدل‌های بلوغ **سازمانی** هستند، نه فردی. یک سازمان تحت نظارت باید آن‌ها را به‌عنوان ابزار «واحد ما کجای منحنی است» تلقی کند و یک نردبان مهارت فردیِ جداگانه را روی آن بسازد.

(c) ابعاد مهارتی که ارزیابی می‌کنند

بُعد	آنچه پوشش می‌دهد	تأکیدشده توسط
روانی AI/GenAI (fluency)	درک اینکه AI چه می‌تواند/نمی‌تواند بکند؛ سواد prompt و ابزار؛ قضاوت درباره کیفیت خروجی	Deloitte، Accenture، PwC
عمق فنی	مهندسی ML/MLOps، مهندسی داده، توسعه مدل، ارزیابی	McKinsey/QuantumBlack، BCG X
پایه داده	کیفیت داده، خطلوله‌ها، حاکمیت، مدیریت feature	همه
ترجمه / محصول	تبدیل مسائل کسب‌وکاری به راه‌حل‌های AI؛ نقش «مترجم» / محصول AI	McKinsey، BCG
به‌کارگیری و تغییر	پیشبرد استفاده، بازطراحی مشاغل و جریان‌های کاری حول AI	Deloitte، Accenture، BCG (همان «70»)
AI مسئولانه / قابل‌اعتماد	انصاف، شفافیت، پاسخگویی، نظارت انسانی	PwC، KPMG
حاکمیت و ریسک AI	ریسک مدل، کنترل‌ها، هم‌راستایی با مقررات	KPMG، Gartner، PwC

(d) شواهدی که می‌طلبند / توصیه می‌کنند

نوع شواهد	چگونه نمایان می‌شود
سیستم‌های تولیدی در حال استفاده	مرحله «Operational» در Gartner دست‌کم یک سیستم AI در تولید با تأمین مالی اختصاصی را می‌طلبد — نه اسلاید.
ارزش کسب‌وکاری قابل‌اندازه‌گیری	هم Forrester و هم McKinsey بلوغ بالاتر را بر ارزش/ROI اثبات‌شده مشروط می‌کنند، نه بر آزمایش‌ها.
گواهینامه‌های مبتنی بر نقش	Deloitte AI Academy و Accenture LearnVantage گواهی تکمیل/گواهینامه گره‌خورده به مسیرهای یادگیری صادر می‌کنند.

نوع شواهد	چگونه نمایان می‌شود
شواهد portfolio / ساخت	کهن‌الگوهای QuantumBlack و BCG X با مدل‌های عرضه‌شده، کد و راه‌حل‌های مستقرشده اثبات می‌شوند.
مصنوعات حاکمیتی	PwC/KPMG انتظار ارزیابی‌های مستندِ Responsible-AI، فهرست‌های مدل و شواهد کنترل را دارند.
معیارهای به‌کارگیری	درصد نیروی کار آموزش‌دیده/فعال، مشاغل بازطراحی‌شده (Deloitte اشاره می‌کند که 84% شرکت‌ها مشاغل را بازطراحی نکرده‌اند — یک نشانه بلوغ).

(e) چگونگی پیوند توانمندی AI با ارشدیت / استخدام / ترفیع / آموزش

سازوکار	الگوی عمومی
کهن‌الگوهای نقش، استخدام را پیش می‌برند	McKinsey/QuantumBlack و BCG X بر اساس کهن‌الگوهای فنیِ نام‌دار (data scientist، ML engineer، data engineer، مترجم، طراح) استخدام و سطح‌بندی می‌کنند که هر کدام رده‌های ارشدیت دارند.
آکادمی‌های یادگیری به‌عنوان ستون فقرات آموزش	Deloitte AI Academy، Accenture LearnVantage و آکادمی‌های خاص هر شرکت، سازوکار استاندارد آموزش هستند؛ مسیرهای مبتنی بر نقش دسترسی به کارهای AI-محور را دروازه‌بندی می‌کنند.
پیوند با ترفیع	شرکت‌ها به‌طور عمومی روانی AI را به «بازآفرینی» نقش‌ها گره می‌زنند، اما روبریک صریحِ «برای رسیدن به Senior Manager به مهارت AI نوع X نیاز دارید» **منتشر نشده است** — هر چنین نگاشتی را به‌عنوان **(استنتاج‌شده از مواد عمومی)** تلقی کنید.
تغییر راهبرد استعداد	پژوهش Deloitte دریافت حدود 75% سازمان‌ها انتظار دارند طی دو سال راهبردهای استعداد خود را به‌دلیل GenAI تغییر دهند — نشانه‌ای از اینکه توانمندی AI در حال تبدیل‌شدن به عامل استخدام و پیشرفت است، نه اختیاری.

(f) مدیریت ایمنی / حاکمیت / امنیت / ریسک مدل

شرکت	چارچوب‌بندی حاکمیتی	آنچه یک سازمان تحت نظارت می‌تواند وام بگیرد
PwC	جعبه‌ابزار **Responsible AI**: انصاف، شفافیت، پاسخگویی، استواری (robustness)، نظارت انسانی؛ اطمینان‌محور.	یک بُعد مهارتی برای «Responsible AI» با رفتارهای نام‌دار و قابل‌ارزیابی.
KPMG	چارچوب **Trusted AI / AI governance** و کنترل‌های «digital workforce»؛ میراث ممیزی و اطمینان‌بخشی.	ذهنیت کنترل‌وشواهد که به‌طور تمیز با انتظارات ریسک مدل بانکی (مثلاً به‌سبک SR 11-7) نگاشت می‌شود.
Gartner	AI TRiSM (Trust, Risk and Security Management) به‌عنوان یک رشته؛ بلوغ حاکمیت را در کنار بلوغ توانمندی توصیه می‌کند.	یک مسیر بلوغ حاکمیت را با مسیر توانمندی جفت کنید تا رشد مهارت هرگز از کنترل پیشی نگیرد.
McKinsey/BCG/ /Deloitte Accenture	اصول Responsible-AI تعبیه‌شده در کار تحول؛ ریسک به‌عنوان دروازه مقیاس‌دهی تلقی می‌شود.	«آیا این می‌تواند ایمن و در چارچوب سیاست مستقر شود» را به یک توانمندی صریحِ سطح‌ترفیع برای نقش‌های فنی ارشد تبدیل کنید.

نکته کاربردی برای یک سازمان خدمات مالی تحت نظارت: از یک مدل پنج‌مرحله‌ایِ به‌سبک Gartner برای الگوبرداری سازمان استفاده کنید؛ کهن‌الگوهای نقش McKinsey/BCG را برای تعریف سطوح فنی فردی وام بگیرید؛ آکادمی + مسیرهای یادگیری مبتنی بر نقش Deloitte/Accenture را به‌عنوان ستون فقرات آموزش بپذیرید؛ و یک مسیر Responsible-AI و ریسک مدل به‌سبک PwC/KPMG به آن متصل کنید تا توانمندی و حاکمیت با هم پیش بروند. هیچ روبریک ترفیع داخلیِ هیچ شرکتی را چنان‌که گویی منتشر شده وارد نکنید — که نشده است.

منابع: - Deloitte: State of AI in the Enterprise 2026 - Gartner AI Maturity Model (BMC summary) - Deloitte AI Academy - Deloitte: Generative AI shaping the future workforce

ارائه‌دهندگان ITSM / DevOps / Observability / AIOps

این دسته، مرتبط‌ترین الگوبرداری از منظر عملیاتی برای یک سازمان فناوریِ خدمات مالی تحت نظارت است، زیرا این ارائه‌دهندگان AI را مستقیماً در ابزارهایی تعبیه می‌کنند که تیم‌های ITSM، platform، SRE و DevSecOps شما همین حالا از آن‌ها استفاده می‌کنند. مستندات عمومی آن‌ها دو کار مفید انجام می‌دهد: (1) **سطوح بلوغ خودگردانی/خودکارسازی** را تعریف می‌کند که می‌توانید مستقیماً به‌عنوان ستون فقرات یک نردبان توانمندی وام بگیرید، و (2) **کنترل‌های human-in-the-loop (HITL)** را مستند می‌کند که به‌طور تمیز با الزامات ریسک مدل و مدیریت تغییر نگاشت می‌شوند. بیشتر این ارائه‌دهندگان یک «روبریک مهارت data scientist» منتشر نمی‌کنند — آن‌ها یک مدل عملیاتی منتشر می‌کنند برای اینکه چه میزان به AI اجازه عمل می‌دهید، و چه اثباتی پیش از بالابردن درجه خودگردانی لازم است. آن مدل عملیاتی دقیقاً همان چیزی است که یک نردبان منابع انسانی/مهارت باید آن را رهگیری کند.

(a) این سازمان‌ها در عمل برای ارزیابی و مهارت‌افزایی توانمندی AI چه می‌کنند

الگوی مشترک، **خودگردانی تدریجیِ دروازه‌بندی‌شده با شواهد** است، نه یک گواهینامه یک‌باره. توانمندی با توانایی یک اپراتور (یا تیم) برای بالابردن ایمنِ یک جریان کاری در مقیاس خودگردانی نشان داده می‌شود.

ارائه‌دهنده / محصول	آنچه ارزیابی می‌کنند / چگونه افراد را مهارت‌افزایی می‌کنند	استنتاج‌شده در برابر منتشرشده
ServiceNow Now Assist	Now Assist **Skill Kit** به سازندگان اجازه می‌دهد promptها/مهارت‌های سفارشی بسازند و منتشر کنند و مجموعه‌داده‌های «ground-truth» بسازند (Now Assist **Data Kit**) تا دقت را پیش از استقرار محک بزنند. **Now Assist Analytics** به‌کارگیری، استفاده و عملکرد را اندازه می‌گیرد. بنابراین مهارت‌افزایی به مهارت‌های ساختن، محک‌زدن و پایش گره خورده است، نه فقط استفاده از آن‌ها.	منتشرشده
Atlassian Intelligence / Rovo + JSM	توانمندی با ساختن/سفارشی‌سازی **Rovo agents**، grounding آن‌ها در دانش، و پیشروی از چت به تیکت‌های واگذارشده تا خودکارسازیِ مبتنی بر تریگرِ خودگردان نشان داده می‌شود. Atlassian یک مدل «governance-in-the-loop» را چارچوب‌بندی می‌کند که در آن انسان‌ها کار آماده‌شده توسط AI را ممیزی/اعتبارسنجی می‌کنند.	منتشرشده
PagerDuty AIOps	اپراتورها از پاسخ دستی ← تریاژ/عیب‌یابی به‌کمک AI ← «پیشنهاد سپس عمل با تأیید» ← playbookهای هوشمندِ تولیدشده از رخدادهای گذشته پیش می‌روند. توانمندی با مدون‌کردن خودکارسازی‌ها و runbookهای قابل‌اتکا نشان داده می‌شود.	منتشرشده
Datadog AI (Bits AI)، Dynatrace Davis، New Relic AI، Splunk AI/ITSI، Elastic AI Assistant، Grafana AI	توانمندی با پیکربندی دستیار در برابر تله‌متری شما، اعتبارسنجی خروجی‌های ریشه‌یابی/ناهنجاری آن در برابر رخدادهای شناخته‌شده، و تنظیم آشکارسازها/SLOها نشان داده می‌شود. «Davis» در Dynatrace، AI علّی/پیش‌بین را با AI مولد جفت می‌کند — اپراتورها باید پیش از اقدام بفهمند چرا یک توصیه ارائه شده است.	عمدتاً استنتاج‌شده از موضع‌گیری عمومیِ محصول

ارائه‌دهنده / محصول	آنچه ارزیابی می‌کنند / چگونه افراد را مهارت‌افزایی می‌کنند	استنتاج‌شده در برابر منتشرشده
Harness AI، GitLab Duo، HashiCorp/Terraform AI	توانمندی با بهره‌برداری ایمن از جریان‌های agentic در SDLC/IaC نشان داده می‌شود: GitLab Duo ابزارها را به Read/Write/Delete دسته‌بندی می‌کند و Write/Delete را پشت تأیید دروازه‌بندی می‌کند؛ HITL به‌طور پیش‌فرض در حالت تعاملی فعال است. اپراتورها با پیکربندی صحیح سیاست‌های حاکمیت در سطح ابزار، صلاحیت خود را اثبات می‌کنند.	منتشرشده (GitLab)؛ استنتاج‌شده (Harness، HashiCorp)

(b) ساختارهای سطح/بلوغ که به‌کار می‌برند

هم‌گرایی قوی بر سر یک **نردبان خودگردانی 3 تا 5 پله‌ای** وجود دارد. این پُرکاربردترین مصنوع قابل‌استفاده‌مجدد برای نردبان خودِ شماست.

سطح	برچسب رایج (ترکیب‌شده از الگوهای PagerDuty، GitLab، Atlassian)	انسان چه می‌کند	AI چه می‌کند
L0	دستی / بدون AI	همه‌چیز	هیچ‌چیز
L1	**یاری‌گر / مکالمه‌ای**	می‌پرسد، می‌خواند، تصمیم می‌گیرد	خلاصه می‌کند، پیش‌نویس تهیه می‌کند، پاسخ می‌دهد، زمینه را آشکار می‌سازد
L2	**توصیه (فقط پیشنهاد)**	هر پیشنهاد را بازبینی و تأیید می‌کند	علت ریشه‌ای، رفع اشکال، کد، runbook پیشنهاد می‌دهد
L3	**تأیید-برای-عمل (دروازه HITL)**	پیش از اجرا، یک اقدام مشخص را تأیید می‌کند	فقط پس از تأیید صریح انسان اجرا می‌کند
L4	**خودگردانی تحت نظارت**	پایش می‌کند، نرده‌های محافظ (guardrail) تعیین می‌کند، استثناها را مدیریت می‌کند	به‌طور خودگردان در دامنه‌های محدود و کم‌ریسک عمل می‌کند؛ موارد لبه‌ای را تشدید می‌کند
L5	**خودگردان (محدود)**	پس از واقعه ممیزی می‌کند؛ مالک guardrailها/سیاست است	جریان‌های کاری کامل را در چارچوب سیاست تریگر و تکمیل می‌کند

- **Atlassian Rovo** یک مدل سه‌سطحی منتشر می‌کند: مکالمه‌ای ← عضو یکپارچه تیم (واگذاری تیکت به agentها) ← خودکارسازیِ مبتنی بر تریگرِ خودگردان. (منتشرشده)
- **PagerDuty** پیشروی از عیب‌یابی به‌کمک AI به «اجرای خودکار رفع اشکال پس از تأیید» تا playbookهای هوشمند آموخته‌شده را توصیف می‌کند. (منتشرشده)
- **GitLab Duo** دروازه را در دسته‌بندی ابزار رمزگذاری می‌کند: Write/Delete، Read = Always Allow = Always Ask. (منتشرشده)
- جدول تجمیع‌شده L0–L5 بالا **(استنتاج‌شده / ترکیب‌شده از مواد عمومی)** است — هیچ ارائه‌دهنده‌ای به‌تنهایی این مقیاس دقیق 6پله‌ای را منتشر نمی‌کند، اما یک ترکیب وفادار است.

(c) ابعاد مهارتی که ارزیابی می‌کنند

بُعد	«خوب» چه شکلی است	جایی که به‌طور عمومی نمایان می‌شود
نگارش prompt / agent	می‌تواند مهارت‌ها/agentهای سفارشی بسازد و نسخه‌بندی کند و آن‌ها را در داده‌های مورد اعتماد ground کند	ServiceNow Skill Kit؛ Atlassian Rovo agents؛ GitLab Duo flows
ارزیابی و محک‌زنی	مجموعه‌داده‌های ground-truth می‌سازد، دقت را پیش از عرضه اندازه می‌گیرد	ServiceNow Data Kit (منتشرشده)؛ برای دیگران استنتاج‌شده
تله‌متری / grounding دامنه	AI را به داده‌های observability درست متصل می‌کند؛ خروجی‌های RCA را اعتبارسنجی می‌کند	Datadog، Dynatrace Davis، New Relic، Splunk، Elastic، Grafana (استنتاج‌شده)
مهندسی خودکارسازی	playbook/runbookها را مدون می‌کند، شعاع تأثیر (blast radius) ایمن تعیین می‌کند	PagerDuty Runbook/Process Automation (منتشرشده)
پیکربندی guardrail / سیاست	تعیین می‌کند کدام اقدام‌ها مجاز/نیازمند پرسش/مسدود هستند؛ دروازه‌های تأیید را تعریف می‌کند	حاکمیت ابزار GitLab؛ guardrailهای ServiceNow (منتشرشده)
قضاوت / نظارت	می‌داند چه زمانی انسان را در حلقه نگه دارد؛ توهم/اعتماد بیش‌ازحد را تشخیص می‌دهد	«governance-in-the-loop» در Atlassian؛ تشدید HITL در ServiceNow (منتشرشده)
پایش و تنظیم پیوسته	به‌کارگیری، drift، دقت و هزینه را پس از استقرار رهگیری می‌کند	ServiceNow Now Assist Analytics (منتشرشده)

(d) شواهدی که می‌طلبند

نوع شواهد	مصنوع عینی	منبع
مجموعه‌داده‌های ground-truth / محک	مجموعه‌داده‌های برچسب‌خورده برای پیش‌بینی و راستی‌آزمایی دقت پیش از go-live	ServiceNow Data Kit (منتشرشده)
معیارهای دقت / به‌کارگیری / عملکرد	داشبوردهای استفاده، دقت، ارزش	ServiceNow Now Assist Analytics (منتشرشده)
رد ممیزیِ تأییدها	تأییدهای انسانیِ ثبت‌شده برای اقدام‌های Write/Delete/رفع اشکال	GitLab Duo HITL؛ PagerDuty approve-to-act (منتشرشده)
agentها / runbookها / IaC نسخه‌بندی‌شده	مهارت‌های منتشرشده، playbookهای هوشمند، ماژول‌های Terraform در کنترل نسخه	ServiceNow، PagerDuty، HashiCorp (منتشرشده/استنتاج‌شده)
پیکربندی‌های guardrail / سیاست	قاعده‌های ابزار و پیکربندی ایمنیِ محتوا به‌عنوان مصنوعات قابل‌بازبینی	تنظیمات حاکمیت GitLab؛ ServiceNow Guardian (منتشرشده)
آموخته‌های رخداد	playbookهای خودکار تولیدشده و اعتبارسنجی‌شده از رخدادهای گذشته	PagerDuty (منتشرشده)

(e) پیوند توانمندی AI با ارشدیت / ترفیع / استخدام / آموزش

این ارائه‌دهندگان نردبان منابع انسانی منتشر نمی‌کنند، بنابراین نگاشت زیر **(استنتاج‌شده / ترکیب‌شده برای زمینه شما)** است اما به‌خوبی در مدل‌های خودگردانی آن‌ها ریشه دارد:

سیگنال ارشدیت	توانمندی نشان‌داده‌شده (با استفاده از سطوح خودگردانی ارائه‌دهنده)
Junior / IC	در L1–L2 عمل می‌کند: از دستیار استفاده می‌کند، پیشنهادها را بازبینی می‌کند، هرگز به‌طور خودکار اجرا نمی‌کند
Senior / IC	در L3 عمل می‌کند: agentها/runbookها را می‌نویسد، دروازه‌های HITL را پیکربندی می‌کند، دقت را با داده‌های ground-truth اعتبارسنجی می‌کند
Lead / Staff	در L4 عمل می‌کند: دامنه‌های خودگردانیِ تحت نظارت را طراحی می‌کند، blast radius تعیین می‌کند، مالک خط‌لوله‌های ارزیابی و پایش است
Principal / Manager	مالک سیاست L5 است: تصمیم می‌گیرد خودگردانی کجا مجاز است، guardrailها را تأیید می‌کند، مسئول وضعیت ریسک مدل است

مسیرهای آموزشی در این دسته، ابزار-محور هستند: گواهینامه‌های ارائه‌دهنده + آزمایشگاه‌های عملی (مثلاً ServiceNow Now Learning، Atlassian University، مستندات و گواهینامه‌های PagerDuty/GitLab). توصیه کاربردی: **ترفیع را مشروط بر حرکت نشان‌داده‌شده و ممیزی‌شده در نردبان خودگردانی کنید** — نه صرفاً بر سابقه یا آشنایی با ابزار.

(f) مدیریت ایمنی / حاکمیت / امنیت / ریسک مدل

این، قوی‌ترین و مستقیماً قابل‌انتقال‌ترین بخش این دسته برای یک سازمان به‌سبک بانکی است.

کنترل	سازوکار	ارائه‌دهنده (منتشرشده)
دروازه‌بندی اقدام بر اساس ریسک	Read = خودکار، Write/Delete = نیازمند تأیید؛ رفع اشکال فقط با تأیید اجرا می‌شود	GitLab Duo؛ PagerDuty
guardrailهای محتوا/ایمنی	پایش و کاهش محتوای توهین‌آمیز، آسیب‌پذیری‌های امنیتی، افشای داده‌های حساس	ServiceNow **Now Assist Guardian**
تشدید موضوع حساس	تعاملات حساس را تشخیص می‌دهد و به‌جای پاسخ‌دادن به انسان مسیردهی می‌کند	ServiceNow (نمونه چت‌بات منابع انسانی)
کاهش prompt-injection / تهدید agentic	تشخیص داخلی prompt-injection؛ تهدیدهای امنیتی agentic مستندشده؛ هویت ترکیبی برای حداقل‌امتیاز و قابلیت‌ممیزی	GitLab Duo
human-in-the-loop به‌طور پیش‌فرض	بدون اقدام بدون تأیید در حالت تعاملی؛ اعتبارسنجی «governance-in-the-loop»	GitLab Duo؛ Atlassian Rovo
هم‌راستایی با مقررات	اشاره صریح به observability از نوع HITL به‌سبک EU AI Act و تأیید چندنفره برای تصمیم‌های خودکار	GitLab Duo
پایش / دیده‌بانی drift	تحلیل‌های به‌کارگیری، استفاده، دقت و عملکرد پس از استقرار	ServiceNow Now Assist Analytics

نکته کلیدی برای سازمان: نردبان خودگردانی (b) را به‌عنوان ستون فقرات روبریک توانمندی AI خود بپذیرید، شواهد بخش (d) را به‌عنوان دروازه بالابردن خودگردانی الزامی کنید، و کنترل‌های (f) را به‌عنوان الزامات غیرقابل‌مذاکره مدیریت تغییر تلقی کنید. ارائه‌دهندگان هم مقیاس بلوغ و هم زیرساخت حاکمیتی را در اختیار شما می‌گذارند؛ وظیفه نردبان منابع انسانی این است که گواهی دهد افراد می‌توانند هر پله را ایمن بهره‌برداری کنند و آن را با مصنوعات قابل‌ممیزی اثبات کنند.

چارچوب‌ها و استانداردهای عمومی

چارچوب‌های عمومی به‌ندرت «مهارت GenAI» را به‌عنوان یک امتیاز واحد می‌سنجند. در عوض، آن‌ها سه بلوک سازنده قابل‌استفاده‌مجدد را عرضه می‌کنند که یک سازمان خدمات مالی می‌تواند وام بگیرد: (1) **سطوح مسئولیت** (یک فرد چه میزان ارشد/خودگردان است)، (2) **ابعاد مهارت یا بلوغ** (چه چیزی ارزیابی می‌شود)، و (3) **کنترل‌ها و شواهد** (چه چیزی باید نشان داده شود، به‌ویژه برای ایمنی، امنیت و ریسک مدل). جدول زیر هر چارچوب را به آنچه واقعاً عرضه می‌کند نگاشت می‌دهد.

(a) این چارچوب‌ها چه می‌کنند و چه چیزی عرضه می‌کنند

چارچوب	آنچه واقعاً انجام می‌دهد	عرضه اصلی
SFIA (Skills Framework for the Information Age)، نسخه 9	مهارت‌های IT/دیجیتال را در برابر 7 «سطح مسئولیت» عمومی تعریف می‌کند. SFIA 9 (2024) مهارت‌های صریح AI/data-science و راهنمایی درباره تقسیم کار بین انسان و AI را افزود.	نردبان ارشدیت + توصیف‌گرهای سطح به‌ازای هر مهارت؛ مستقیماً برای ترفیع/استخدام قابل‌استفاده
European e-Competence Framework (e-CF / EN 16234-1)	استاندارد EU که 41 شایستگی ICT را به 5 سطح مهارت (e-1 تا e-5) نگاشت می‌دهد و با European Qualifications Framework (EQF) هم‌راستا است.	مدل شایستگی با رده‌های مهارت؛ قابل‌انتقال در زمینه‌های مقرراتی EU
معیارهای DORA / Accelerate	معیارهای عملکرد تحویل نرم‌افزار با پشتوانه پژوهشی؛ تیم‌ها را به‌عنوان Low / Medium / High / Elite محک می‌زند.	معیارهای نتیجه‌محور در سطح تیم (نه مهارت فردی) برای اعتبارسنجی اینکه آیا ابزار AI واقعاً تحویل را بهبود می‌دهد
ITIL 4	چارچوب مدیریت خدمات؛ practiceها و یک مدل «چهار بُعد» + Service Value System را تعریف می‌کند.	زبان توانمندی عملیاتی و بلوغ فرایند برای تیم‌های run/ops
CNCF Platform Engineering Maturity Model	پلتفرم‌های توسعه‌دهنده داخلی را در سراسر سطوح بلوغ ارزیابی می‌کند.	رده‌های بلوغ برای تیم‌های platform/توانمندساز که ابزار AI را به‌طور داخلی عرضه می‌کنند
NIST AI RMF (AI 100-1)	چارچوب مدیریت ریسک داوطلبانه که حول 4 کارکرد ساختار یافته است: Govern، Map، Measure، Manage (به‌علاوه یک GenAI Profile، NIST AI 600-1).	کارکردهای حاکمیت و دسته‌های کنترل؛ ستون فقرات حاکمیت ریسک AI
ISO/IEC 42001:2023	استاندارد قابل‌گواهیِ AI Management System (AIMS)؛ پیوست A، 38 کنترل ویژه AI را در 9 هدف فهرست می‌کند.	کنترل‌های سیستم‌مدیریتیِ قابل‌ممیزی و الزامات شواهد
OWASP Top 10 for LLM Applications (2025)	فهرست رتبه‌بندی‌شده مهم‌ترین ریسک‌های امنیتی برنامه‌های LLM/GenAI.	شایستگی‌های امنیتی عینی و چک‌لیست تهدید برای توسعه به‌کمک AI
چارچوب‌های Responsible-AI / سواد AI (مثلاً OECD AI Principles، تکلیف سواد AI در EU AI Act، ویژگی‌های trustworthy-AI در NIST)	اصول و یک تعهد سواد پایه را برای کارکنانی که AI می‌سازند یا استفاده می‌کنند تعریف می‌کنند.	خط پایه سواد + ابعاد اخلاقی؛ به‌طور فزاینده یک الزام قانونی

چارچوب	آنچه واقعاً انجام می‌دهد	عرضه اصلی
Secure-SDLC for AI (مثلاً NIST SSDF SP 800-218 + پروفایل GenAI در 800-218A، راهنمای SAFECODE/ OWASP)	practiceهای توسعه ایمن را به مؤلفه‌های AI/ GenAI گسترش می‌دهد.	کنترل‌های فرایندی و practiceهای توسعه‌دهنده برای عرضه ایمن AI

(b) ساختارهای سطح/بلوغ

چارچوب	ساختار سطح	یادداشت‌ها
SFIA 9	7 سطح: 1 Follow، 2 Assist، 3 Apply، 4 Enable/، 5 Ensure/Advise، 6 Initiate Influence، 7 Set strategy/Inspire/Mobilise	هر سطح با Autonomy، Influence، Complexity، مهارت‌های کسب‌وکاری/رفتاری، و Knowledge تعریف می‌شود
e-CF	5 سطح مهارت (e-1 پایین‌ترین تا e-5 بالاترین)، نگاشت‌شده به سطوح EQF 8-3	مبتنی بر شایستگی به‌جای مبتنی بر نقش
DORA	4 خوشه عملکرد: Low / Medium / High / Elite	باندهای نتیجه‌محور، نه سطوح مهارت
ITIL 4	بلوغ توانمندیِ practice (به‌ازای هر practice ارزیابی می‌شود)؛ بدون یک مقیاس ثابت واحد	اغلب با یک مدل توانمندی 5نقطه‌ای جفت می‌شود
CNCF Platform Eng. Maturity	4 سطح بلوغ (Provisional ← Operational ← Optimizing ← Scalable)، به‌ازای هر جنبه ارزیابی می‌شود	پایین‌تر = تاکتیکی، بالاتر = راهبردی؛ قرار نیست همه‌جا به حداکثر برسد (استنتاج‌شده: برچسب‌های دقیق سطح بسته به نسخه منبع متفاوت است)
NIST AI RMF	بدون سطوح بلوغ؛ 4 کارکرد × دسته‌ها/زیردسته‌ها	Profileها/رده‌ها توسط سازمان تعریف می‌شوند
ISO/IEC 42001	منطبق / نامنطبق (قابل‌گواهی)؛ درجه‌بندی‌نشده	بلوغ به‌طور تلویحی از پوشش کنترل استنباط می‌شود
OWASP LLM Top 10	فهرست ریسک رتبه‌بندی‌شده (LLM01–LLM10)؛ بدون مقیاس بلوغ	به‌عنوان چک‌لیست پوشش به‌کار می‌رود

(c) ابعاد مهارت / ارزیابی

چارچوب	ابعاد ارزیابی‌شده
SFIA 9	Autonomy، Influence، Complexity، مهارت‌ها/رفتارهای کسب‌وکاری، Knowledge — به‌ازای هر مهارت اعمال می‌شود (شامل مهارت‌های AI/data)
e-CF	5 حوزه شایستگی: Plan، Build، Run، Enable، Manage
DORA	فراوانی استقرار (Deployment frequency)، Lead time for changes، Change failure rate، Failed-deployment recovery time (+ قابلیت‌اطمینان)
ITIL 4	چهار بُعد: Organisations & People، Information & Technology، Partners & Suppliers، Value Streams & Processes
CNCF Platform Eng.	Investment، Adoption، Interfaces، Operations، Measurement
NIST AI RMF	ویژگی‌های قابل‌اعتمادبودن: معتبر و قابل‌اتکا، ایمن، امن و تاب‌آور، پاسخگو و شفاف، قابل‌توضیح و قابل‌تفسیر، حریم‌خصوصی‌محور، منصفانه (سوگیری مدیریت‌شده)

چارچوب	ابعاد ارزیابی‌شده
ISO/IEC 42001	سیاست AI، سازمان/نقش‌های داخلی، منابع، چرخه حیات سیستم AI، کیفیت و تبارشناسی داده، نظارت انسانی، زنجیره تأمین/شخص ثالث
OWASP LLM Top 10	Prompt injection، افشای اطلاعات حساس، زنجیره تأمین، مسموم‌سازی داده/مدل، مدیریت خروجی، عاملیت بیش‌ازحد (excessive agency)، نشت system-prompt، ضعف‌های vector/embedding، اطلاعات نادرست، مصرف بی‌حد (unbounded consumption)
Responsible-AI / سواد	انصاف، شفافیت، پاسخگویی، نظارت انسانی، حریم خصوصی، به‌علاوه سواد پایه AI برای کارکنان

(d) شواهد لازم

چارچوب	شواهد متداول
SFIA 9	مسئولیت‌های نشان‌داده‌شده در سطح، محصولات کاری، ارزیابی مدیر/همتا، گواهینامه‌های نگاشت‌شده به مهارت‌ها
e-CF	portfolio شایستگی نگاشت‌شده به EQF؛ مدارک و خروجی‌های نشان‌داده‌شده
DORA	تله‌متری از سیستم‌های CI/CD، رخداد و استقرار (عینی، نه خوداظهاری)
ITIL 4	سوابق فرایند، جریان‌های ارزش تعریف‌شده، تعاریف نقش، به‌کارگیریِ ممیزی‌شده practiceها
CNCF Platform Eng.	معیارهای به‌کارگیری پلتفرم، رابط‌های self-service، داده‌های عملیاتی/SLO، بازخورد کاربر
NIST AI RMF	فهرست ریسک، ارزیابی‌های اثر، نتایج اندازه‌گیری، فعالیت‌های مستندِ Govern/Map/Measure/Manage
ISO/IEC 42001	Statement of Applicability، ارزیابی‌های اثر AI، سوابق تبارشناسی داده، logهای نظارت، پایش drift، رد ممیزی (قابل‌گواهی توسط شخص ثالث)
OWASP LLM Top 10	مدل‌های تهدید، نتایج red-team / pen-test، اقدام‌های کاهشی به‌ازای هر ریسک، شواهد پیکربندی ایمن
Secure-SDLC for AI	SBOM/AI-BOM، سوابق منشأ (provenance)، شواهد کدنویسی و بازبینی ایمن، نتایج آزمون/اسکن

(e) چگونگی پیوند توانمندی AI با ارشدیت / ترفیع / استخدام / آموزش

- **SFIA و e-CF تنها دو موردی هستند که مستقیماً تصمیمات منابع انسانی را پیش می‌برند.** توصیف‌گرهای سطح SFIA به‌طور رایج برای نوشتن پروفایل‌های نقش، تعیین معیارهای ترفیع و ساختاردهی باندهای پرداخت به‌کار می‌روند؛ مهارت‌های AI/data افزوده‌شده در SFIA 9 به یک سازمان اجازه می‌دهد بگوید، برای مثال، «از یک ML engineer سطح L5 انتظار می‌رود درباره استقرار مدل Ensure/Advise کند.» e-CF به مدارک رسمی نگاشت می‌شود و جایی که قابلیت‌انتقال EU/مقرراتی اهمیت دارد ترجیح داده می‌شود.
- **DORA، CNCF و ITIL در سطح تیم متصل می‌شوند، نه فرد.** بهترین کاربرد آن‌ها اعتبارسنجیِ بازده سرمایه‌گذاری AI است (مثلاً آیا کدنویسی به‌کمک AI تیم را از High به Elite برد، یا آیا نرخ شکست تغییر بدتر شد؟)، و دروازه‌بندی هزینه توانمندسازی/آموزش.
- **NIST AI RMF، ISO 42001، OWASP و چارچوب‌های responsible-AI شایستگی‌ها و سواد لازم را تعریف می‌کنند** که باید به‌عنوان انتظار در شرح نقش‌ها و به‌عنوان برنامه درسی آموزش ظاهر شوند — به‌طور فزاینده یک پیش‌نیاز استخدام/onboarding، نه یک «خوب است داشته باشیم». (استنتاج‌شده از مواد عمومی: هیچ‌کدام از

این چارچوب‌ها یک نردبان ترفیع تجویز نمی‌کنند؛ سازمان‌ها معمولاً شایستگی‌های خود را روی سطوح SFIA پیوند می‌زنند.)

(f) ایمنی / حاکمیت / امنیت / ریسک مدل

دغدغه	چارچوب(های) متناسب	آنچه عرضه می‌کند
حاکمیت و مدیریت ریسک AI	NIST AI RMF (Govern/Map/Measure/Manage)؛ ISO/IEC 42001	کارکردها + کنترل‌های قابل‌گواهی؛ به‌طور تمیز روی کمیته‌های ریسک به‌سبک بانکی نگاشت می‌شود
مدیریت ریسک مدل	NIST AI RMF + ISO 42001 (و، استنتاج‌شده، هم‌راستایی با practiceهای ریسک مدل بانکی مانند اعتبارسنجی به‌سبک SR 11-7)	اعتبارسنجی، پایش، تشخیص drift، نظارت انسانی
امنیت برنامه/LLM	OWASP Top 10 for LLM Applications (2025)	فهرست تهدید اولویت‌بندی‌شده و اقدام‌های کاهشی برای برنامه‌های به‌کمک AI و دارای AI تعبیه‌شده
توسعه ایمن AI	NIST SSDF (SP 800-218 + پروفایل GenAI در 800-218A)؛ راهنمای OWASP/SAFECODE	practiceهای Secure-SDLC گسترش‌یافته به داده‌های آموزش، مدل‌ها و promptها
Responsible AI / انصاف / سواد	OECD AI Principles؛ ویژگی‌های trustworthy-AI در NIST؛ تکلیف سواد AI در EU AI Act	ابعاد اخلاقی + یک خط پایه اجباری سواد کارکنان
تاب‌آوری عملیاتی	ITIL 4؛ معیارهای بازیابی/قابلیت‌اطمینان DORA	مدیریت رخداد و انتظارات بازیابی برای خدمات AI

نکته کاربردی برای یک سازمان تحت نظارت: از **SFIA 9** (یا e-CF) برای نردبان افراد، از **NIST AI RMF + ISO 42001** برای ستون فقرات حاکمیت و ریسک مدل، از **OWASP LLM Top 10 + NIST SSDF** برای شایستگی‌های امنیتی، و از **DORA/CNCF/ITIL** به‌عنوان شواهد نتیجه‌محور استفاده کنید که نشان دهد توانمندی AI واقعی است و صرفاً ادعا نشده.

منابع: SFIA 9 AI skills framework · CNCF Platform Engineering Maturity Model · OWASP Top 10 for LLM Applications 2025 · ISO/IEC 42001 Annex A controls

---

4. معیارسنجی بانکداری و خدمات مالی تحت نظارت

4.0 شرکت‌های فناوری در برابر بانک‌ها / مؤسسات مالی

حوزه	شرکت‌های فناوری	بانک‌ها / مؤسسات مالی	تأثیر بر چارچوب ارزیابی
آزمایشگری در هوش مصنوعی	اغلب سریع‌تر و بازتر	کنترل‌شده‌تر و مبتنی بر سیاست	دانش ایمنی و سیاست‌ها باید زودتر ارزیابی شود
استفاده از داده	آزمایشگری داخلی گسترده‌تر	کنترل‌های سخت‌گیرانه طبقه‌بندی داده و حریم خصوصی	مدیریت داده باید بخشی از سطوح پایه باشد
خودکارسازی	ممکن است خودمختاری بالاتری مجاز باشد	تأیید انسانی و حسابرسی الزامی است	سطوح خودکارسازی باید مستلزم تأیید و قابلیت بازگشت باشند

حوزه	شرکت‌های فناوری	بانک‌ها / مؤسسات مالی	تأثیر بر چارچوب ارزیابی
حاکمیت	محصول/پلتفرم‌محور	درگیری گسترده ریسک، انطباق، حقوقی و حسابرسی	شایستگی حاکمیتی باید صریح باشد
ارزیابی	متمرکز بر بهره‌وری و نوآوری	متمرکز بر توانمندی تعدیل‌شده با ریسک و کنترل	ارزیابی باید میان بهره‌وری و ریسک توازن برقرار کند
Agentic AI	پایلوت‌های سریع‌تر	محدودیت‌های شدید بر اقدامات در محیط عملیاتی	سطوح Agentic مستلزم کنترل‌های قوی‌تر هستند
شواهد	خروجی مهندسی	خروجی مهندسی به‌علاوه شواهد انطباق	شواهد HR باید شامل قابلیت حسابرسی باشد
مهارت‌ها	مهارت‌های سازندگی هوش مصنوعی	سازندگی هوش مصنوعی به‌علاوه مهارت‌های ریسک/کنترل هوش مصنوعی	نردبان باید شامل مهارت‌های ریسک و حاکمیت باشد

معیارسنجی بانکداری و خدمات مالی تحت نظارت

شرکت‌های خدمات مالی تحت نظارت، توانمندی هوش مصنوعی را به‌شکلی بسیار متفاوت از شرکت‌های صرفاً فناوری‌محور دنبال می‌کنند. از آنجا که این شرکت‌ها تحت رژیم‌های ریسک مدل، حمایت از مصرف‌کننده و تاب‌آوری عملیاتی فعالیت می‌کنند (مانند SR 11-7، GDPR، EU AI Act، FCA/PRA، MAS)، «مهارت هوش مصنوعی» را ترکیبی از توانایی فنی **و** انضباط کنترلی در نظر می‌گیرند. توانمندی به‌ندرت تنها بر پایه استعداد خام مدل‌سازی ارتقا می‌یابد؛ بلکه مشروط به این است که آیا فرد می‌تواند هوش مصنوعی را به‌شیوه‌ای حاکمیت‌شده، قابل‌حسابرسی و تحت نظارت انسانی بسازد، مستقر کند و راهبری نماید.

تقریباً هیچ‌یک از این شرکت‌ها «نردبان مهارت هوش مصنوعی» داخلی منتشر نمی‌کنند. موارد زیر از بیانیه‌های مطبوعاتی عمومی، آگهی‌های شغلی/مسیر شغلی، گزارش‌های سالانه و اظهارات CIO و وبلاگ‌های مهندسی بانک‌ها استخراج شده‌اند. هر موردی که به‌طور مستقیم به‌صورت عمومی بیان نشده باشد با برچسب **(استنباط‌شده از منابع عمومی)** مشخص شده است.

(a) آنچه این سازمان‌ها در عمل برای ارزیابی و ساخت توانمندی هوش مصنوعی انجام می‌دهند

شرکت	اقدام عمومی	آنچه درباره ساخت توانمندی نشان می‌دهد
JPMorgan Chase	عرضه «LLM Suite» به حدود ۱۴۰ تا ۲۵۰ هزار کارمند؛ تجمیع مدل‌های OpenAI/Anthropic؛ بیش از ۴۵۰ مورد کاربرد GenAI؛ برگزیده «نوآوری سال» ۲۰۲۵ نشریه American Banker	توانمندسازی در مقیاس انبوه برای تمام کارکنان، نه فقط مهندسان؛ یک دروازه داخلی حاکمیت‌شده به سطح اصلی ساخت مهارت بدل می‌شود
Morgan Stanley	«AI @ Morgan Stanley Assistant» (۲۰۲۳) و «Debrief» برای مشاوران ثروت؛ «AskResearchGPT» (GPT-4o) برای اوراق بهادار نهادی، شامل بیش از ۷۰ هزار گزارش پژوهشی	دستیارهای مختص نقش؛ مشاوران آموزش می‌بینند که بر خروجی هوش مصنوعی نظارت کنند، نه آنکه از صفر تولید نمایند
Goldman Sachs	«GS AI Assistant» در سطح کل شرکت پس از پایلوت با حدود ۱۰ هزار کارمند؛ نسخه ویژه توسعه‌دهندگان؛ مدل‌های LLM میزبانی‌شده داخلی تأییدشده (GPT-4، Gemini، متن‌باز)؛ Copilot در محیط‌های امن	دستیارهای لایه‌بندی‌شده و مختص نقش؛ مهندسان روی copilotها در محیط‌های تأییدشده مهارت می‌یابند

شرکت	اقدام عمومی	آنچه درباره ساخت توانمندی نشان می‌دهد
Capital One	«Tech College» (از ۲۰۱۷، شامل مسیر Machine Learning & AI)؛ برنامه ۱۶۰ ساعته آموزش مهندسی یادگیری ماشین (MLETP)؛ ادبیات عمومی پررنگ درباره Responsible/Explainable AI در آگهی‌های شغلی ML	آکادمی داخلی رسمی + تبدیل ساختاریافته مهندسان نرم‌افزار/داده به مهندسان ML
HSBC	«AI Review Councils» تعبیه‌شده در سراسر گروه (نیمه اول ۲۰۲۵)؛ Chief AI Officer؛ آگهی نقش‌های «Group Responsible AI Governance Manager»؛ سرمایه‌گذاری سنگین در آموزش	حاکمیت‌محور؛ توانمندی تا حدی بر اساس توانایی گذراندن بازبینی شورا سنجیده می‌شود
DBS	چارچوب «PURE» (هدفمند، بدون شگفتی، محترمانه، قابل‌توضیح)؛ موضع صریح «co-pilot، انسان در حلقه»؛ آموزش کارکنان برای تفسیر توصیه‌های هوش مصنوعی	مهارت = تفسیر و نظارت ایمن بر هوش مصنوعی، نه استفاده خودمختار
Citi	تشریح عمومی برنامه آموزشی هوش مصنوعی برای کارکنان در سطح کل شرکت؛ استقرار دستیارهای توسعه/کدنویسی و ابزارهای GenAI داخلی برای ده‌ها هزار نفر	حرکت گسترده برای سواد نیروی کار در کنار ابزارهای مهندسان

الگوی مشترک: یک **دروازه LLM داخلی امن** (مستقل از مدل و ثبت‌شونده) محل اصلی به‌کارگیری و مشاهده توانمندی است. استفاده عمومی شخصی از ChatGPT/Claude معمولاً محدود می‌شود؛ ابزار مجاز همان سطح سنجش است.

(b) ساختارهای سطح / بلوغی که به کار می‌برند

ساختار	محل مشاهده عمومی	یادداشت‌ها
سه خط دفاعی (Three Lines of Defense)	راهنمای OCC / SR 11-7، اعمال‌شده بر مدل‌های AI/GenAI	توسعه ← اعتبارسنجی مستقل ← حسابرسی داخلی؛ این همان عدسی غالب بلوغ در بانکداری است
AI Review Council / دروازه‌های تأیید	HSBC (AI Review Councils)، نقش‌های شورای هوش مصنوعی مسئولانه در بانک‌ها	موارد کاربرد از طریق بازبینی لایه‌بندی‌شده بر اساس سطح ریسک پیش می‌روند
لایه‌های دستیار مختص نقش	Goldman، Morgan Stanley، JPMorgan	گونه‌های توسعه‌دهنده / بانکدار / تحلیل‌گر / مشاور از همان پلتفرم
پیشرفت در آکادمی داخلی	مسیرهای Capital One Tech College؛ اتمام MLETP	پیشرفت ساختاریافته مبتنی بر دوره **(استنباط‌شده از منابع عمومی که اتمام دوره شرط واجد شرایط بودن برای نقش ML است)**
طبقه‌بندی مدل بر اساس لایه‌بندی ریسک	فهرست مدل SR 11-7 + لایه‌بندی بر اساس اهمیت	مدل‌های با اهمیت بالا اعتبارسنجی و پایش سنگین‌تری دریافت می‌کنند

در این شرکت‌ها **هیچ «نردبان توانمندی هوش مصنوعی» شماره‌گذاری‌شده‌ای (L1–L5) به‌طور عمومی افشا نشده است**. مدل بلوغ عملی، چرخه عمر ریسک مدل به‌علاوه تأیید شورا است، نه یک سنجه شایستگی HR. **(استنباط‌شده از منابع عمومی.)**

(c) ابعاد مهارتی که ارزیابی می‌کنند

بُعد	شواهد ارزیابی‌شدن آن	تضاد با شرکت‌های صرفاً فناوری‌محور
مهندسی LLM (ML/GenAIها، RAG، vector storeها، چرخه عمر کامل ML)	آگهی‌های شغلی ML شرکت Capital One؛ دستیار توسعه‌دهنده Goldman	مشابه شرکت‌های فناوری
Responsible & Explainable AI	انتظارات نقش ML در Capital One؛ DBS PURE؛ نقش‌های حاکمیتی HSBC	وزن بسیار سنگین‌تر نسبت به شرکت‌های فناوری متعارف
سواد ریسک مدل و اعتبارسنجی	ساختار سه خط دفاعی SR 11-7	تا حد زیادی در شرکت‌های صرفاً فناوری‌محور غایب است
استفاده امن/حاکمیت‌شده از ابزار (مدیریت داده، بدون نشت)	مدل‌های میزبانی‌شده داخلی Goldman؛ دروازه JPMorgan LLM Suite	سخت‌گیرانه‌تر؛ صرفاً ابزار مجاز
قضاوت در حالت انسان در حلقه / نظارت	موضع co-pilot شرکت DBS؛ «قضاوت انسانی در هسته» HSBC؛ گردش‌کارهای مشاوران Morgan Stanley	متمایز شرکت‌های تحت نظارت
سواد هوش مصنوعی برای غیرمهندسان	عرضه به بیش از ۱۴۰ هزار نفر در JPMorgan؛ برنامه آموزشی Citi	سواد گسترده‌تر و الزامی از سوی هیئت‌مدیره
قابلیت حسابرسی / مستندسازی	الزامات مستندسازی SR 11-7 **(به‌طور خاص برای GenAI استنباط‌شده)**	بسیار قوی‌تر

(d) شواهدی که مطالبه می‌کنند

نوع شواهد	مبنای عمومی
اتمام آموزش داخلی (دوره/برنامه)	Capital One Tech College، MLETP (۱۶۰ ساعت)
اثبات استفاده ایمن درون دروازه LLM مجاز	JPMorgan LLM Suite، Goldman GS AI Assistant، دستیارهای Morgan Stanley
مستندات مدل، سوابق اعتبارسنجی، گزارش‌های پایش	الزامات ریسک مدل SR 11-7 (اعمال‌شده بر AI/ML/GenAI)
تأییدیه شورا / هیئت بازبینی بر یک مورد کاربرد	HSBC AI Review Councils
پایبندی به یک چارچوب هوش مصنوعی مسئولانه نام‌برده	DBS PURE؛ اصول هوش مصنوعی مسئولانه منتشرشده توسط شرکت
پیامدهای بهره‌وری/کیفیت (زمان صرفه‌جویی‌شده، دقت)	JPMorgan به صرفه‌جویی حدود ۳ تا ۶ ساعت در هفته اشاره می‌کند؛ کاهش حجم پرس‌وجوهای پژوهشی در Morgan Stanley **(سنجه‌های پیامد، گزارش‌شده توسط شرکت)**

(e) چگونگی پیوند توانمندی هوش مصنوعی با ارشدیت / استخدام / ارتقا / آموزش

- **استخدام:** آگهی‌های شغلی مهندس ML/AI (مثلاً Capital One) به‌صراحت تجربه Responsible & Explainable AI و ML با چرخه عمر کامل را الزامی می‌کنند — یعنی مهارت هوش مصنوعی حاکمیت‌شده یک معیار غربالگری است، نه یک امتیاز اضافی. **(شواهد آگهی شغلی عمومی.)**

- **مسیرهای رهبری اختصاصی:** ایجاد نقش‌های Chief AI Officer (HSBC) و آگهی‌های «Responsible AI Governance Manager» نشان می‌دهد که حاکمیت هوش مصنوعی مسیر شغلی ارشد مستقل خود را دارد. **(عمومی.)**
- **آموزش به‌عنوان توانمندساز، نه هنوز دروازه ارتقای منتشرشده:** آکادمی‌های داخلی (Capital One Tech College/MLETP، برنامه برنامه‌ریزی‌شده Citi) توانمندی می‌سازند، اما شرکت‌ها **به‌طور عمومی اعلام نمی‌کنند** که اتمام دوره یک الزام رسمی ارتقاست. هرگونه چنین پیوندی باید **(استنباط‌شده از منابع عمومی)** تلقی شود.
- **دسترسی به ابزار به‌عنوان نشانگر توانمندی:** دریافت دستیار ویژه توسعه‌دهندگان (Goldman) یا research GPT (Morgan Stanley) مختص نقش است — خودِ دسترسی بازتاب نیاز/اعتماد ارزیابی‌شده است. **(استنباط‌شده از منابع عمومی.)**

(f) چگونگی مدیریت ایمنی، حاکمیت، امنیت و ریسک مدل

حوزه کنترل	رویه تحت نظارت (عمومی)	چرا با شرکت‌های صرفاً فناوری‌محور متفاوت است
مدیریت ریسک مدل	SR 11-7 (Fed/OCC) اعمال‌شده بر AI/ML/GenAI؛ سه خط دفاعی؛ اعتبارسنجی مستقل؛ فهرست و لایه‌بندی مدل	شرکت‌های صرفاً فناوری‌محور به‌ندرت اعتبارسنجی مستقل الزامی یا بازرسی‌های نظارتی دارند
حاکمیت هوش مصنوعی مسئولانه	HSBC AI Review Councils؛ چارچوب DBS PURE؛ اصول هوش مصنوعی مسئولانه منتشرشده	شوراهای رسمی با اختیار وتو بر استقرار
استفاده امن از LLM	مدل‌های میزبانی‌شده/تأییدشده داخلی (Goldman)؛ دروازه تجمیعی دروازه‌بندی‌شده (JPMorgan LLM Suite)؛ محدودیت بر چت‌بات‌های عمومی	کنترل‌های نشت داده و محرمانگی الزامی و حسابرسی‌شده هستند
انسان در حلقه	«همیشه یک انسان در حلقه» DBS؛ «قضاوت، تصمیم‌گیری و پاسخگویی انسانی در هسته» HSBC؛ بازبینی خروجی توسط مشاوران Morgan Stanley	پاسخگویی باید بر عهده یک انسان نام‌برده باشد
قابلیت حسابرسی و قابلیت توضیح	انتظارات مستندسازی، پایش و قابلیت توضیح تحت SR 11-7؛ «آسان برای توضیح» در DBS PURE	ناظران می‌توانند استدلال قابل‌بازتولید مطالبه کنند
ریسک تأمین‌کننده / مدل	فهرست‌های تأیید چندمدلی (Goldman: GPT-4/Gemini/متن‌باز)؛ مدل شخص ثالث به‌عنوان ریسک تأمین‌کننده تلقی می‌شود	هوش مصنوعی شخص ثالث تحت رژیم‌های ریسک تأمین‌کننده و ریسک تمرکز قرار می‌گیرد
سوگیری / انصاف	انتظارات Responsible/Explainable AI در نقش‌های ML (Capital One)؛ پوشش وام‌دهی منصفانه برای مدل‌های مصرف‌کننده **(برای هوش مصنوعی اعتبار مصرف‌کننده استنباط‌شده)**	قانون حمایت از مصرف‌کننده (مانند ECOA/وام‌دهی منصفانه) مواجهه حقوقی می‌افزاید

نتیجه‌گیری کلی برای یک سازمان تحت نظارت: عامل تمایزبخش این نیست که آیا کارکنان می‌توانند از GenAI استفاده کنند، بلکه این است که آیا می‌توانند آن را درون یک دروازه کنترل‌شده به کار گیرند، مستند و اعتبارسنجی کنند، یک انسان را پاسخگو نگه دارند و از یک شورای حاکمیتی عبور کنند. بنابراین یک نردبان توانمندی هوش مصنوعی در این بخش باید به **تحویل حاکمیت‌شده، قابل‌حسابرسی و تحت نظارت انسانی** دست‌کم به اندازه مهارت فنی خام وزن دهد — که این تأکید درست برعکس اغلب نردبان‌های شرکت‌های صرفاً فناوری‌محور است.

منابع: CNBC — JPMorgan LLM Suite، CIO Dive، وبلاگ JPMorganChase؛ Morgan Stanley AskResearchGPT — بیانیه مطبوعاتی، CNBC؛ CNBC — Goldman Sachs GS AI Assistant، Fortune؛

Capital One MLETP، Tech College؛ HSBC AI؛ DBS Responsible AI؛ Fed SR 11-7 / راهنمای ریسک مدل OCC.

5. مقایسه ساختارهای سطح‌بندی رایج در بازار

بازار از نردبان‌های خودمختاری ۳ سطحی گرفته تا چارچوب‌های مسئولیت ۷ سطحی، و همچنین مدل‌های حوزه‌ای بدون شماره‌گذاری، از همه چیز استفاده می‌کند. جدول زیر گزینه‌های ساختاری اصلی را مقایسه می‌کند.

ساختار	چارچوب‌های واقعی که از آن استفاده می‌کنند	نقطه قوت	نقطه ضعف
۳ سطح	خودمختاری Atlassian Rovo (مکالمه‌ای ← یکپارچه ← خودمختار)	ساده، سریع در انتقال مفهوم، مناسب برای کنترل خودمختاری ابزارها	برای کاربردهای منابع انسانی بیش از حد درشت‌دانه است؛ نمی‌تواند تفاوت‌های ظریف میان نیروهای تازه‌کار و ارشد را تفکیک کند
۴ سطح	خوشه‌های DORA (Low/Medium/High/Elite)؛ پذیرش Anthropic (explore/pilot/scale/transform)؛ CNCF (Provisional/Operational/Scalable/Optimizing)	تمیز و مناسب برای بلوغ سازمان/تیم و مراحل پذیرش؛ روایت اجرایی آسان	برای نیروی کار بزرگ با گستره‌ای از سطوح ارشدیت، گام‌های کافی ندارد؛ فاصله ارتقاها زیاد به نظر می‌رسد
۵ سطح	بلوغ AI گارتنر (Awareness ← Transformational)؛ پذیرش Microsoft Copilot؛ e-CF (e-1 تا e-5)؛ ITIL به ازای هر practice	رایج‌ترین، به‌خوبی شناخته‌شده، توازن میان جزئیات و سادگی	همچنان گستره وسیعی از ارشدیت فردی را در یک گام فشرده می‌کند
۶ سطح	خودمختاری ترکیبی AIOps (L0-L5: دستی ← خودمختار محدود)	مناسب برای خودمختاری کنش و کنترل ریسک با یک کف صریح «دستی»	برای خودمختاری سامانه بهینه شده، نه برای رشد شغلی انسان
۷ سطح	۹ سطح مسئولیت SFIA (Follow ← Set strategy)	استاندارد صنعتی برای مسئولیت فردی؛ ویژگی‌های غنی خودمختاری/تأثیرگذاری/پیچیدگی؛ نگاشت به EQF	عمومی برای کل حوزه IT؛ بدون پیوند زدن، مختص AI نیست
۱۰ سطح	رایج در نردبان‌های منابع انسانی شرکت‌های فناوری بزرگ (استنتاج‌شده از منابع عمومی؛ تعاریف مشخص سطوح منتشر نشده‌اند)	گام‌های ریزدانه برای سازمان‌های بزرگ مناسب است؛ ارتقاهای مکرر و انگیزه‌بخش؛ سطح‌بندی دقیق حقوق و دستمزد	تعدد گام‌ها خطر بیش‌مهندسی و بحث‌های اغراق‌آمیز بر سر سطوح مجاور را به دنبال دارد
باندهای شایستگی نقش‌محور	حوزه‌های «McKinsey rewired»؛ سواد AI دلویت (Deloitte)؛ کهن‌الگوهای BCG X / QuantumBlack	تنوع واقعی نقش‌ها را بازتاب می‌دهد؛ از تحمیل یک مقیاس واحد به همه پرهیز می‌کند	مقایسه میان نقش‌ها دشوارتر است؛ سیگنال ارتقای واحد ضعیف‌تری دارد

ساختار	چارچوب‌های واقعی که از آن استفاده می‌کنند	نقطه قوت	نقطه ضعف
ماتریس‌ها / روبریک‌های مهارتی	زیردسته‌های NIST AI RMF؛ OWASP LLM Top 10؛ کنترل‌های Annex A استاندارد ISO 42001	ریزدانه، مبتنی بر شواهد، قابل ممیزی؛ ایده‌آل برای امتیازدهی به شایستگی‌های مشخص	یک روایت نیست؛ به‌تنهایی به فرد نمی‌گوید «من در چه سطحی هستم»
مراحل بلوغ	Gartner، Forrester، Microsoft، DORA	بهترین گزینه برای پیگیری سازمان/برنامه و گزارش‌دهی به هیئت‌مدیره	سازمان را توصیف می‌کند، نه فرد را
مسیرهای گواهینامه	رده‌های AWS / Google Cloud / IBM؛ آکادمی‌های Deloitte/Accenture	عینی، قابل تأیید بیرونی، کنترل‌کننده استخدام	دانش را می‌سنجد، نه کاربست در محیط کار یا قضاوت را

آیا ۱۰ سطح زیاد، کم یا مناسب است؟

برای یک سازمان بزرگ خدمات مالی تحت نظارت با گستره‌ای وسیع از ارشدیت (از کارآموزان و تحلیلگران تا مهندسان اصلی، معماران و مدیر ارشد هوش مصنوعی)، **۱۰ سطح مناسب است** — به شرط آنکه با یک روبریک همراه شود. استدلال چنین است:

- **چرا کمتر نه.** یک نردبان ۴ یا ۵ گامی، یک توسعه‌دهنده تازه‌کار و یک مهندس staff را در سطل‌های مجاور قرار می‌دهد و فاصله‌های بزرگ و دلسردکننده‌ای در ارتقا و سطح‌بندی نادقیق حقوق ایجاد می‌کند. مدل‌های بلوغ (Gartner 5، DORA 4) سازمان را توصیف می‌کنند، نه افراد را، و از این رو برای یک درجه‌بندی فردی نامناسب‌اند.
- **چرا بیشتر نه.** فراتر از حدود ۱۰ سطح، تمایز سطوح مجاور دشوار می‌شود و بحث‌های کالیبراسیون بر منفعت آن می‌چربد.
- **چرا رویکرد ترکیبی بهترین است.** نردبان شماره‌گذاری‌شده به‌تنهایی، قضاوت‌های ذهنی و مبتنی بر هیاهو را دعوت می‌کند («به نظر در کار با AI خوب است»). روبریک به‌تنهایی هیچ روایت شغلی ارائه نمی‌دهد. ترکیب یک **نردبان ۱۰ سطحی (برای روایت منابع انسانی، ارتقا و حقوق)** با یک **روبریک ۰-۵ به ازای هر بُعد (برای امتیازدهی عینی و مبتنی بر شواهد)** هم یک داستان روشن و هم یک سنجش قابل دفاع و قابل ممیزی به دست می‌دهد — دقیقاً همان چیزی که یک سازمان تحت نظارت به آن نیاز دارد. این ۱۰ سطح را می‌توان با ۷ سطح مسئولیت SFIA و با مدل‌های بلوغ ۴/۵ مرحله‌ای برای محک‌زنی بیرونی و گزارش‌دهی به هیئت‌مدیره متناظر کرد.

6. ابعاد رایج مهارتی AI / GenAI

این ۱۶ بُعد در بیشتر نقش‌ها کاربرد دارند. هر بُعد یک معنای ساده و قابل‌فهم را با شیوه مشاهده آن در عمل (با ترجیح شواهد عینی بر خودارزیابی) همراه می‌کند.

#	بُعد	معنای آن	نحوه مشاهده
1	سواد AI	درک اینکه AI چیست، چه کاری می‌تواند و نمی‌تواند انجام دهد و محدودیت‌های آن چیست	ارزیابی کوتاه؛ می‌تواند یک مورد کاربرد و ریسک‌های آن را به زبان ساده توضیح دهد
2	سواد GenAI	درک مدل‌های مولد، توکن‌ها، context و رفتار توهم‌زایی (hallucination)	کوئیز یا مصاحبه؛ تشخیص زمانی که خروجی احتمالاً غیرقابل اتکاست
3	مهندسی پرامپت	نوشتن پرامپت‌های روشن و ساختارمند؛ تکرار برای بهبود خروجی	نمونه پرامپت‌ها/خروجی‌ها؛ آثار پرامپت نسخه‌بندی‌شده؛ بازبینی همتایان

#	بُعد	معنای آن	نحوه مشاهده
4	استفاده از ابزارهای AI	استفاده مؤثر از ابزارهای مجاز AI و در چارچوب سیاست‌ها	تله‌متری درون‌محصول (استفاده فعال، نرخ پذیرش)؛ گزارش‌های ابزار
5	کاربست مختص شغل	به‌کارگیری AI در وظایف واقعی نقش خود	محصولات کاری؛ تغییر اندازه‌گیری‌شده در توان عملیاتی یا کیفیت وظایف
6	تفکر انتقادی / اعتبارسنجی	بررسی خروجی AI از نظر دقت، سوگیری و مرتبط بودن پیش از استفاده	یادداشت‌های بازبینی؛ نمونه‌های خطای کشف‌شده؛ رد خروجی نامناسب
7	بازبینی / تأیید انسانی	حفظ پاسخگویی یک انسان در قبال تصمیمات یاری‌شده با AI	سوابق تأیید؛ ردپای امضای نهایی روی کارهای متأثر از AI
8	توانایی خودکارسازی	ساخت خودکارسازی‌ها و عامل‌های (agent) قابل اتکا و ایمن	runbookها/playbookها/agentهای نسخه‌بندی‌شده؛ محدودیت‌های شعاع تأثیر (blast-radius)
9	درک داده و RAG	درک grounding، بازیابی (retrieval)، کیفیت داده و تازگی آن	پیکربندی‌های RAG؛ توانایی توضیح منبع grounding و حالت‌های شکست
10	درک AI عامل‌محور (Agentic)	درک عامل‌های چندمرحله‌ای، استفاده از ابزار و ریسک خودمختاری	طراحی/راهبری یک agent با حفاظ‌های (guardrail) مناسب
11	ایمنی و امنیت AI	تشخیص prompt injection، نشت داده، عاملیت بیش‌ازحد (OWASP LLM Top 10)	مدل‌های تهدید؛ نتایج red-team/آزمون؛ پیکربندی‌های امن
12	حاکمیت و انطباق	کار در چارچوب سیاست AI، الزامات ثبت گزارش و تأیید	شواهد رعایت؛ یافته‌های ممیزی پاک؛ گواهی‌های سیاستی
13	آگاهی از ریسک مدل	درک اعتبارسنجی مدل، پایش و رانش (drift)	مشارکت در اعتبارسنجی/پایش؛ ریسک مدل مستندشده
14	همکاری / اشتراک دانش	گسترش رویه‌های خوب AI؛ راهنمایی دیگران	پرامپت‌ها/قالب‌های قابل استفاده مجدد به‌اشتراک‌گذاشته‌شده؛ منتورینگ؛ ارائه‌های داخلی
15	تأثیر کسب‌وکاری	پیوند استفاده از AI به ارزش قابل اندازه‌گیری (زمان، کیفیت، هزینه)	معیارهای ROI/پیامد؛ سنجش‌های قبل/بعد، نه حکایت‌ها
16	رهبری و استراتژی	تعیین جهت‌گیری، استانداردها و اولویت‌های سرمایه‌گذاری AI	مالکیت کار استانداردها/CoE؛ آثار استراتژی؛ تصمیمات سبد (portfolio)

7. ابعاد مهارتی صنعت بانکداری / صنایع تحت نظارت

در یک سازمان خدمات مالی تحت نظارت، ابعاد زیر بر فراز ابعاد رایج قرار می‌گیرند. وزن بیشتری دارند و — به‌طور حیاتی — باید **زودتر** از یک شرکت صرفاً فناوری‌محور در نردبان اثبات شوند.

#	بُعد FSI	معنای آن	نحوه مشاهده
1	آگاهی از AI مسئولانه	به‌کارگیری انصاف، شفافیت، پاسخگویی و نظارت انسانی (مثلاً DBS PURE)	رعایت چارچوب نام‌برده؛ بررسی‌های مستند AI مسئولانه

#	بُعد FSI	معنای آن	نحوه مشاهده
2	طبقه‌بندی داده / محرمانگی	شناخت رده‌های حساسیت داده و مدیریت صحیح هر کدام	طبقه‌بندی درست در عمل؛ بازبینی‌های پاک مدیریت داده
3	مدیریت داده مشتری و PII/PCI	حفاظت از داده‌های نظارت‌شده مشتری در تمام کاربردهای AI	نبود PII/PCI در پرامپت‌ها/گزارش‌ها؛ شواهد DLP و gateway
4	آگاهی از ریسک مدل (SR 11-7)	درک فهرست مدل‌ها، رده‌بندی اهمیت (materiality)، اعتبارسنجی مستقل	سوابق اعتبارسنجی؛ طبقه‌بندی اهمیت؛ گزارش‌های پایش
5	آگاهی از prompt injection	تشخیص و کاهش تزریق (OWASP LLM #1)	مدل‌های تهدید؛ اقدامات کاهشی آزمون‌شده؛ مدیریت امن پرامپت
6	ریسک ابزارهای AI شخص ثالث	برخورد با LLMهای بیرونی تحت رژیم ریسک تأمین‌کننده و ریسک تمرکز	استفاده از فهرست مدل‌های تأییدشده؛ ارزیابی‌های ریسک تأمین‌کننده
7	ثبت گزارش ممیزی AI	اطمینان از ثبت و نگهداری تعاملات AI	شواهد گزارش/ردیابی (trace)؛ انطباق در نگهداری داده
8	تبیین‌پذیری / قابلیت ردیابی	توانایی توضیح اینکه یک تصمیم متأثر از AI چگونه اتخاذ شده است	آثار تبیین‌پذیری؛ مستنداتی که از بازبینی عبور می‌کنند
9	تأیید انسانی برای کنش‌های عملیاتی	الزام امضای نهایی پیش از آنکه AI روی Write/Delete/اصلاح اقدام کند	ردپای ممیزی تأیید (دروازه‌هایی به سبک GitLab Duo / PagerDuty)
10	تفکیک وظایف	جداسازی نقش‌های ساخت، اعتبارسنجی و ممیزی (سه خط دفاعی)	شواهد جداسازی نقش؛ اعتبارسنجی مستقل توسط یک طرف متفاوت
11	یکپارچگی با مدیریت تغییر	هدایت تغییرات AI از مسیر کنترل تغییر استاندارد	تیکت‌های تغییر؛ تأییدهای پیوندخورده به استقرارهای AI
12	آگاهی از انطباق قانونی	شناخت الزامات قابل اعمال (SR 11-7، سواد EU AI Act، اعطای وام منصفانه)	گواهی‌ها؛ تکمیل آموزش انطباق
13	تولید شواهد ممیزی	تولید آثاری که ناظران می‌توانند بازرسی کنند	بیانیه‌های قابلیت اعمال (Statements of Applicability)، ارزیابی‌های اثر، factsheetها
14	گزارش حوادث AI	شناسایی و گزارش‌دهی خطاها و موارد نزدیک به شکست AI	گزارش‌های حادثه ثبت‌شده؛ پیروی از مسیر تشدید (escalation)
15	مدیریت ریسک تأمین‌کننده / مدل	مدیریت چرخه عمر مدل، تأیید و ریسک تمرکز	فهرست‌های تأیید چندمدلی؛ پایش چرخه عمر و رانش
16	SDLC امن برای کد یاری‌شده با AI	بازبینی و امن‌سازی کد تولیدشده توسط AI (NIST SSDF 800-218A)	بازبینی کد، SBOM/AI-BOM، نتایج اسکن/آزمون
17	خودکارسازی متناسب با ریسک	تنظیم خودمختاری متناسب با ریسک کنش و شعاع تأثیر	دروازه‌های لایه‌ای (Read خودکار / تأیید برای Write-Delete)؛ دامنه محدود

چرا این ابعاد باید زودتر در یک نردبان تحت نظارت ظاهر شوند

در یک شرکت صرفاً فناوری‌محور، انتظار می‌رود مهارت‌های حاکمیت و ریسک مدل تنها در سطوح ارشد بروز کنند. در بانکداری، هزینه یک اشتباه — تخطی از اعطای وام منصفانه، نشت داده مشتری، یا یک مدل اعتبارسنجی‌نشده در محیط عملیاتی — نظارتی و اعتباری است، نه صرفاً فنی. SR 11-7، الزام سواد در EU AI Act و انتظارات سه خط دفاعی به این

معناست که **هر** فردی که با AI سروکار دارد باید از همان نخستین سطح خود، نظم پایه‌ای در استفاده مسئولانه، مدیریت داده و انضباط تأیید را به نمایش بگذارد. توانمندی توسط تحویل حاکمیت‌شده، قابل ممیزی و تحت نظارت انسانی مهار می‌شود — نه صرفاً با استعداد خام ساخت مدل. در عمل، ابعاد 1، 2، 3، 5، 7 و 9 باید در سطوح ورودی الزامی باشند؛ ابعاد عمیق‌تر ریسک مدل و اعتبارسنجی (4، 10، 13، 15) با ارشدیت مقیاس می‌گیرند اما آگاهی نسبت به آن‌ها زود آغاز می‌شود.

8. مدل ارزیابی پیشنهادی و توجیه آن

پیشنهاد: یک مدل ترکیبی. از یک **نردبان توانمندی ۱۰ سطحی** برای روایت منابع انسانی (داستان شغلی، ارتقا، سطح‌بندی حقوق) و از یک **روبریک ۱۶ بُعدی ۰-۵** برای امتیازدهی عینی استفاده کنید، که هر دو با مدل‌های بلوغ استاندارد ۴/۵ مرحله‌ای برای محک‌زنی و گزارش‌دهی به هیئت‌مدیره متناظر شده‌اند.

چگونگی کارکرد دو بخش در کنار هم

- **نردبان (L1-L10)** به پرسش «در چه سطحی هستم و گام بعدی چیست» به زبان ساده پاسخ می‌دهد — مناسب گفت‌وگوهای منابع انسانی، تصمیمات ارتقا و جبران خدمات.
- **روبریک (۱۶ بُعد، هر یک با امتیاز ۰-۵)** به پرسش «آیا این فرد واقعاً می‌تواند، با شواهد، این کار را انجام دهد» پاسخ می‌دهد — مناسب ارزیابی فنی و ممیزی. هر سطح روی نردبان با یک نیم‌رخ روبریک الزامی پشتیبانی می‌شود (مثلاً حداقل امتیازها در ابعاد ایمنی، حاکمیت و بازبینی انسانی)، به‌گونه‌ای که هر ارتقا با شواهد پشتیبانی می‌شود نه با برداشت ذهنی.

نگاشت به سایر مدل‌ها

باند نردبان	مناسب برای	نگاشت به بلوغ (Gartner 5 / DORA 4)	نگاشت به SFIA 7
L1-L3 (بنیادین)	مناسب منابع انسانی؛ سواد + استفاده ایمن	Awareness/Active؛ Low/Medium	Follow / Assist / Apply
L4-L6 (کاربر حرفه‌ای)	مناسب ارزیابی فنی	Operational؛ High	Enable / Ensure-Advise
L7-L8 (ارشد/راهبر)	ارزیابی فنی + حاکمیت	Systemic؛ Elite	Initiate-Influence
L9-L10 (اصلی/اجرایی)	مناسب منابع انسانی + استراتژی	Transformational	Set strategy

چرا ۱۰ سطح برای یک سازمان بزرگ تحت نظارت با ترکیب گسترده‌ای از ارشدیت

- کل جمعیت را پوشش می‌دهد — از تحلیلگران و غیرمهندسانی که تنها به سواد AI نیاز دارند تا مهندسان اصلی، معماران و مدیر ارشد هوش مصنوعی — بدون آنکه مسیرهای شغلی متمایز را به یک گام واحد تحمیل کند.
- گام‌های مکرر و خوش‌تعریف، ارتقاها را انگیزه‌بخش و سطح‌بندی حقوق را دقیق نگه می‌دارند.
- سطوح پایین‌تر (L1-L3) **مناسب منابع انسانی** هستند (سواد، استفاده ایمن از ابزار) و می‌توان آن‌ها را به‌سبکی ارزیابی کرد؛ سطوح میانی و بالایی (L4-L8) **مناسب ارزیابی فنی** هستند و به شواهد و — برای ابعاد حاکمیتی — بازبینی مستقل نیاز دارند؛ سطوح بالا (L9-L10) **مناسب استراتژی/رهبری** هستند.

چگونه کنترل‌های نظارتی نردبان را جابه‌جا می‌کنند

ابعاد حاکمیت، ریسک مدل و AI مسئولانه **در ابتدا متمرکز می‌شوند**: نظم پایه‌ای در استفاده مسئولانه، مدیریت داده، بازبینی انسانی و انضباط تأیید از L1 الزامی است، نه موکول‌شده به سطوح ارشد. یک **مسیر بلوغ حاکمیتی را به‌موازات** مسیر توانمندی اجرا کنید تا رشد مهارت هرگز از کنترل‌ها پیشی نگیرد. استقرار ایمن، منطبق با سیاست و قابل ممیزی را یک **شایستگی فنی صریح در سطح ارشد** بسازید — در L7 به بالا، «تحویل ایمن تحت SR 11-7 / سه خط دفاعی» به اندازه مهارت ساخت مدل اهمیت دارد.

چگونه مدیران از ارزیابی هیاهو در برابر توانمندی واقعی اجتناب می‌کنند

این مرکزی‌ترین ریسک است و شواهد عمومی، احتیاط را الزامی می‌کنند: در سال 2024، DORA پذیرش AI را به کاهش ثبات تحویل پیوند داد، و پژوهشی که به نقل از MIT NANDA منتشر شد نشان داد حدود ۹۵٪ از پایلوت‌های GenAI هیچ تأثیر مالی قابل اندازه‌گیری نشان ندادند — شکست‌هایی در جریان کار و انگیزه‌ها، نه در کیفیت مدل. برای محافظت در برابر هیاهو، الزام کنید:

1. **شواهد بر شور و شوق.** روبریک را تنها در برابر آثار امتیازدهی کنید — تله‌متری، محصولات کاری، ردپای تأیید، سوابق اعتبارسنجی، اسناد حاکمیتی — هرگز بر اساس هیجان خوداظهاری‌شده نسبت به ابزار.
2. **معیارهای پیامد، نه فعالیت.** تأثیر کسب‌وکاری اندازه‌گیری‌شده (زمان صرفه‌جویی‌شده، کیفیت، کاهش نقص) را با داده‌های قبل/بعد پاداش دهید، نه حجم استفاده از AI را.
3. **ابعاد اعتبارسنجی و بازبینی انسانی به‌عنوان دروازه.** یک امتیاز بالای بهره‌وری نمی‌تواند امتیاز پایین در تفکر انتقادی/اعتبارسنجی، بازبینی انسانی یا ایمنی را جبران کند — تا «سریع اما ناایمن» به‌عنوان ارشد قلمداد نشود.
4. **کالیبراسیون مستقل.** برای ابعاد حاکمیت و ریسک مدل، امضای نهایی فردی خارج از خط تحویل نامزد را الزام کنید (سه خط دفاعی)، به‌گونه‌ای که اعتبارسنجی مستقل SR 11-7 را بازتاب دهد.
5. **عبارت‌پردازی مستقل از ابزار.** ابعاد را حول قضاوت و پیامدها تعریف کنید، نه محصولات مشخص، تا نردبان در برابر تغییرات تأمین‌کننده و مدل پایدار بماند.

---

9. نردبان پیشنهادی قابلیت AI / GenAI در ۱۰ سطح

این نردبان، شایستگی‌های AI/GenAI را بر روی یک ستون فقرات مبتنی بر مسئولیت (به سبک SFIA: خودگردانی، اثرگذاری، پیچیدگی) پیوند می‌زند؛ همان‌گونه که در بخش ۲ توصیه شد. هیچ شرکتی به‌صورت عمومی نگاشتی از AI به سطوح HR منتشر نکرده است، بنابراین هم‌ترازسازی سطح با گرید **(استنتاج‌شده از منابع عمومی)** است؛ شایستگی‌های زیربنایی، انواع شواهد و کنترل‌ها از چارچوب‌های عمومی برگرفته شده‌اند (SFIA 9، e-CF، NIST AI RMF، OWASP LLM Top 10 2025، ISO/IEC 42001، SR 11-7، DORA/SPACE).

راهنمای مطالعه. سطوح ۱ تا ۳ مربوط به سواد همگانی و استفاده ایمن است (مبنای الزامی برای همه کارکنان بر اساس تعهد سواد در EU AI Act). سطوح ۴ تا ۶ به استفاده مولد در سطح کارورز اختصاص دارد. سطوح ۷ تا ۹ ساخت و مهندسی است. سطح ۱۰ رهبری راهبرد و حاکمیت است. کنترل‌های نظارتی از همان ابتدا تعبیه شده‌اند: مدیریت داده و ایمنی از L2 تا L3 ظاهر می‌شوند؛ دروازه‌های تأیید و بازگشت (rollback) برای خودکارسازی از L7 به بعد پدیدار می‌شوند؛ و مالکیت کامل حاکمیت در L10 قرار دارد.

L1 — بدون مواجهه

میدان	محتوا
منطق پژوهش‌محور	مدل‌های بلوغ (Gartner، Anthropic) از یک کفِ «کاوش/هیچ» آغاز می‌شوند؛ یک نردبان به یک صفرِ صریح نیاز دارد تا سازمان بتواند بدهیِ سواد پایه را در برابر الزام سوادآموزیِ EU AI Act اندازه‌گیری کند.
خلاصه	از ابزارهای GenAI برای کار استفاده نکرده است؛ هیچ مدل ذهنیِ کارآمدی از کارکرد آن‌ها یا مخاطراتشان ندارد.
رفتار مورد انتظار	در کار روزمره از AI استفاده نمی‌کند. ممکن است کنجکاو یا بدبین باشد. هیچ مواجهه‌ای با ابزارهای مجاز ندارد.
مهارت‌های فنی	هیچ پیش‌نیازی ندارد.
مهارت‌های مهندسی پرامپت (prompt-engineering)	ندارد.
استفاده از ابزار AI	ندارد.
کاربرد ویژه نقش شغلی	هنوز هیچ.
مهارت‌های اتوماسیون	ندارد.
مهارت‌های Data/RAG	ندارد.
مهارت‌های Agentic-AI	ندارد.
مهارت‌های ایمنی و حاکمیت AI	ندارد — این شکاف و مخاطره اصلی است.
آگاهی از مخاطرات بانکی/مقرراتی	از این آگاه نیست که جای‌گذاری (paste) داده در ابزارهای عمومی نقض کنترل‌هاست.
شواهد مورد نیاز برای HR	ندارد (وضعیت آغازین پیش‌فرض).
نمونه ارزیابی عملی	یک آزمون کوتاه سنجش سواد، نتیجه‌ای نزدیک به صفر بازمی‌گرداند.
نمونه‌های ویژه نقش	تازه‌واردِ پیش از onboarding؛ کارکنانِ نقش‌هایی که هیچ استقرار AI در آن‌ها انجام نشده است.
سیگنال ارتقا	ماژول اجباریِ onboarding سواد AI را تکمیل می‌کند.
آموزش مورد نیاز برای سطح بعدی	سواد AI سازمانی ۱۰۱: GenAI چیست، ابزارهای مجاز کدام‌اند، و قواعدِ «داده را paste نکنید».
ضدالگوها	استفاده از ابزارهای مصرفیِ غیرمجاز «صرفاً برای امتحان کردن».
پرچم‌های قرمز	جای‌گذاری داده مشتری یا داده محرمانه در چت‌بات‌های عمومی.

L2 — آگاهی نسبت به هوش مصنوعی (AI Awareness)

حوزه	محتوا
منطق مبتنی بر پژوهش	EU AI Act سواد پایه را به یک الزام انطباقی (compliance) تبدیل کرده است، نه یک گزینه‌ی اختیاری؛ مدل‌های بلوغ نیز آگاهی را به‌عنوان دروازه‌ای پیش از هرگونه استفاده‌ی عملی در نظر می‌گیرند.

حوزه	محتوا
خلاصه	درک می‌کند که GenAI چه کارهایی می‌تواند و چه کارهایی نمی‌تواند انجام دهد، از وجود ابزارهای مجاز آگاه است و قواعد مدیریت داده را **پیش از** کار با هر ابزار می‌داند.
رفتار مورد انتظار	می‌تواند GenAI را به زبان ساده توضیح دهد؛ توهم (hallucination) را به‌عنوان یک حالت شکست واقعی تشخیص می‌دهد؛ می‌داند کدام ابزار برای کدام طبقه از داده تأیید شده است.
مهارت‌های فنی	صرفاً مفهومی: پرامپت/پاسخ، مدل، توهم (hallucination)، پنجره‌ی زمینه (context window).
مهارت‌های مهندسی پرامپت	پرامپت خوب را از بد تشخیص می‌دهد؛ اما هنوز این مهارت را به‌صورت عملی کسب نکرده است.
استفاده از ابزارهای AI	از درگاه مجاز (مثلاً مجموعه‌ی داخلی LLM) آگاه است؛ اما هنوز کاربر منظم آن نیست.
کاربرد ویژه‌ی شغل	می‌تواند ۱ تا ۲ مورد را نام ببرد که AI ممکن است در نقش او کمک‌کننده باشد.
مهارت‌های اتوماسیون	ندارد.
مهارت‌های Data/RAG	به‌صورت مفهومی درک می‌کند که چرا مبنا قرار دادن داده (grounding) اهمیت دارد.
مهارت‌های Agentic-AI	ندارد.
مهارت‌های ایمنی و حاکمیت AI	**قواعد طبقه‌بندی داده را می‌داند**: اینکه چه چیزی عمومی/داخلی/محرمانه/محدود است و اینکه داده‌ی محدود هرگز بدون تأیید به یک مدل داده نمی‌شود.
آگاهی از ریسک بانکی/مقرراتی	می‌داند که خروجی دارای توهم را نمی‌توان بدون بررسی در زمینه‌های ارتباط با مشتری یا مقرراتی به کار برد.
شواهد مورد نیاز برای HR	قبولی در ارزیابی اجباری سواد AI + استفاده‌ی قابل‌قبول (acceptable-use).
نمونه‌ی ارزیابی عملی	چندگزینه‌ای: «کدام ابزار را می‌توانید برای داده‌ی محرمانه‌ی مشتری استفاده کنید؟» + شناسایی یک ریسک توهم.
نمونه‌های ویژه‌ی نقش	کارکنان عملیات شعبه، تحلیلگر انطباق (compliance analyst)، توسعه‌دهنده‌ی تازه‌کاری که فرایند آشناسازی (onboarding) را تکمیل کرده است.
نشانه‌ی ارتقا	گواهی استفاده‌ی قابل‌قبول را دریافت می‌کند؛ نخستین پرامپت‌های خود را در ابزار مجاز اجرا می‌کند.
آموزش مورد نیاز برای سطح بعد	آشنایی عملی با ابزار چت مجاز؛ مبانی پرامپت‌نویسی ایمن.
الگوهای نامطلوب (Anti-patterns)	«درباره‌اش خوانده‌ام» بدون اینکه هرگز ابزار مجاز را باز کرده باشد.
پرچم‌های قرمز (Red flags)	باور به اینکه خروجی AI مرجع و قطعی است؛ ندانستن قواعد داده پس از گذراندن آموزش.

L3 — کاربر پایه هوش مصنوعی (Basic AI User)

میدان	محتوا
مبنای پژوهشی	نردبان‌های پذیرش ارائه‌شده توسط فروشندگان (مدل ۵سطحی Microsoft Copilot) کاربردِ روزمرهٔ یاری‌شده را در مراحل آغازین قرار می‌دهند؛ مطالعات کنترل‌شده (سرعت حدود ۵۵٪ بیشتر در GitHub) نشان می‌دهند ارزش‌آفرینی از همان سطحِ استفادهٔ شایستهٔ پایه آغاز می‌شود.
خلاصه	از GenAI مجاز برای کارهای روزمرهٔ ساده و کم‌ریسک استفاده می‌کند و قواعد مدیریت داده و راستی‌آزمایی را به‌طور قابل‌اتکا به‌کار می‌بندد.
رفتار مورد انتظار	با کمک هوش مصنوعی پیش‌نویس تهیه می‌کند، خلاصه می‌سازد و بازنویسی می‌کند؛ همواره پیش از استفاده خروجی را بررسی می‌کند؛ هرگز دادهٔ محدودشده را وارد (paste) نمی‌کند.
مهارت‌های فنی	رابط کاربری مجاز را با اطمینان به‌کار می‌گیرد؛ محدودیت‌های token/context را در عمل درک می‌کند.
مهارت‌های مهندسی پرامپت (prompt-engineering)	پرامپت‌های تک‌مرحله‌ای روشن همراه با نقش و زمینه می‌نویسد.
استفاده از ابزار هوش مصنوعی	استفادهٔ روزانه از چت/دستیار تأییدشده برای پیش‌نویس‌نویسی و پرسش‌وپاسخ.
کاربرد ویژهٔ شغل	پیش‌نویس ایمیل، خلاصهٔ جلسات، بازنویسی اسناد، توضیح کد.
مهارت‌های اتوماسیون	ندارد — صرفاً دستی و درون‌ابزاری.
مهارت‌های Data/RAG	از قابلیت بارگذاری سند داخلی ابزار در چارچوب کلاس‌های دادهٔ تأییدشده استفاده می‌کند.
مهارت‌های هوش مصنوعی عامل‌محور (Agentic-AI)	ندارد.
مهارت‌های ایمنی و حاکمیت هوش مصنوعی	**هر خروجی را پیش از استفاده راستی‌آزمایی می‌کند؛ دادهٔ حساس را پنهان‌سازی می‌کند یا از آن پرهیز می‌کند؛ رفتار غیرعادی را گزارش می‌دهد.** هوش مصنوعی را پیش‌نویس می‌داند، نه منبع.
آگاهی از ریسک بانکی/مقرراتی	می‌داند پیش‌نویس‌های تولیدشده با هوش مصنوعی برای مشتریان یا نهادهای ناظر نیازمند بازبینی و تأیید نهایی انسانی است.
شواهد لازم برای HR	گواهی مدیر مبنی بر استفادهٔ ایمن و روزمره؛ گواهینامهٔ مقدماتیِ فروشنده (اختیاری).
نمونهٔ ارزیابی عملی	خلاصه‌سازی یک سند خط‌مشی ۳صفحه‌ای و مشخص‌کردن نقاطی که خلاصه ممکن است نادرست باشد.
نمونه‌های ویژهٔ نقش	تحلیلگری که گزارش‌ها را خلاصه می‌کند؛ PMی که تیکت پیش‌نویس می‌کند؛ کارشناس پشتیبانی که پاسخ‌ها را پیش‌نویس می‌کند.
نشانهٔ ارتقا	استفادهٔ پیوسته، ایمن و روزانه؛ آغاز به تکرار و بهبود پرامپت‌ها برای دستیابی به نتایج بهتر.
آموزش لازم برای سطح بعد	مهندسی پرامپت ساختاریافته: few-shot، تجزیهٔ مسئله، قالب‌بندی خروجی.
الگوهای نامطلوب (Anti-patterns)	کپی‌وپیست عینی خروجی هوش مصنوعی در محصولات قابل‌تحویل.
پرچم‌های قرمز	نادیده‌گرفتن راستی‌آزمایی؛ نشت گاه‌وبیگاهِ دادهٔ حساس.

L4 — کاربر اثربخش پرامپت (Effective Prompt User)

فیلد	محتوا
توجیه پژوهش‌محور	در همه منابع، سواد کار با پرامپت/ابزار به یک بُعد همگرا تبدیل شده است؛ در این سطح، تله‌متری درون‌محصول (نرخ پذیرش) معنادار می‌شود.
خلاصه	با به‌کارگیری سنجیده فنون پرامپت‌نویسی به نتایج قابل‌اتکا و باکیفیت می‌رسد و می‌داند چه زمانی AI ابزار مناسبی نیست.
رفتار مورد انتظار	پرامپت‌ها را به‌صورت روشمند بازنگری و اصلاح می‌کند؛ از مثال‌ها، محدودیت‌ها و خروجی ساختاریافته بهره می‌گیرد؛ الگوهای توهم (hallucination) را تشخیص می‌دهد و خود را اصلاح می‌کند.
مهارت‌های فنی	با خروجی‌های ساختاریافته (جداول، JSON) و پرامپت‌نویسی در سطح system/role در ابزار به‌راحتی کار می‌کند.
مهارت‌های مهندسی پرامپت	few-shot، chain-of-thought، تجزیه مسئله، کنترل قالب، و بازنگری تکرارشونده پرامپت.
استفاده از ابزار AI	استفاده حرفه‌ای از ابزارهای چت و دستیارهای درون-IDE در جایی که کاربرد دارد.
کاربرد متناسب با شغل	پیش‌نویس‌هایی نزدیک به نسخه نهایی تولید می‌کند؛ از AI برای ساخت چارچوب تحلیل و دریافت نکات بازبینی کد استفاده می‌کند.
مهارت‌های اتوماسیون	قالب‌های پرامپت را ذخیره و بازاستفاده می‌کند؛ اتوماسیون برنامه‌نویسی‌شده ندارد.
مهارت‌های Data/ RAG	با چسباندن/بارگذاری زمینه (context) تأییدشده، پرامپت‌ها را به‌خوبی مستند و زمینه‌مند می‌کند؛ درک می‌کند که زمینه‌مندسازی (grounding) توهم را کاهش می‌دهد.
مهارت‌های Agentic-AI	از پالایش چندمرحله‌ای (multi-turn) استفاده می‌کند؛ عامل‌های خودمختار به کار نمی‌برد.
مهارت‌های ایمنی و حاکمیت AI	پرامپت‌هایی طراحی می‌کند که **درخواست منابع/بیان عدم‌قطعیت** می‌کنند؛ استنادهای ساختگی را تشخیص داده و کنار می‌گذارد؛ انضباط طبقه‌بندی داده را حفظ می‌کند.
آگاهی از ریسک بانکی/ مقرراتی	در سطح کاربر می‌داند که prompt injection وجود دارد (OWASP LLM #1)؛ نسبت به محتوای چسبانده‌شده غیرقابل‌اعتماد محتاط است.
شواهد لازم برای HR	پورتفوی پرامپت (قبل/بعد)؛ تله‌متری‌ای که نرخ بالای پذیرش پیشنهاد همراه با بازکاری اندک را نشان می‌دهد.
نمونه ارزیابی عملی	با دریافت یک نیازمندی نامرتب، از طریق ۳ بار بازنگری پرامپت یک مشخصات ساختاریافته تولید کند و هر تغییر را توضیح دهد.
نمونه‌های متناسب با نقش	توسعه‌دهنده‌ای که از AI برای کدهای قالبی (boilerplate) و تست‌ها استفاده می‌کند؛ تحلیلگری که مدل‌های اولیه می‌سازد؛ افسر ریسکی که روایت‌های کنترلی را پیش‌نویس می‌کند.
نشانه ارتقا	دیگران از او می‌پرسند «چطور به این نتیجه رسیدی؟»؛ پرامپت‌های قابل‌بازاستفاده پدیدار می‌شوند.
آموزش موردنیاز برای سطح بعدی	یکپارچه‌سازی AI در گردش‌کارهای چندمرحله‌ای؛ زنجیره‌سازی ابزارها؛ اسکریپت‌نویسی سبک.
ضدالگوها	پرامپت‌نویسی بیش‌ازحد برای کارهای پیش‌پاافتاده؛ به‌کارگیری ترفندهای پرامپت به‌جای دانش حوزه‌ای.
پرچم‌های قرمز	پذیرش استنادهای ساختگی؛ نادیده‌گرفتن نشانه‌های عدم‌قطعیت.

L5 — کاربر هوش مصنوعی در جریان کاری (Workflow AI User)

فیلد	محتوا
مبنای پژوهشی	شاخص‌های تحویل DORA/SPACE زمانی شروع به حرکت می‌کنند که هوش مصنوعی در جریان‌های کاری واقعی تنیده شود، نه در گفت‌وگوهای پراکنده و موردی؛ اینجاست که بهره‌وری قابل‌اندازه‌گیری پدیدار می‌شود.
خلاصه	هوش مصنوعی را در سراسر یک جریان کاری چندمرحله‌ای فردی/تیمی جای می‌دهد و نتیجه را اندازه‌گیری می‌کند.
رفتار مورد انتظار	هوش مصنوعی را در مراحل مختلف زنجیر می‌کند (پژوهش ← پیش‌نویس ← بازبینی ← قالب‌بندی)؛ promptها را برای یک فرایند تکرارشونده استانداردسازی می‌کند؛ زمان صرفه‌جویی‌شده و کیفیت را رصد می‌کند.
مهارت‌های فنی	ابزارها را به هم متصل می‌کند (دستیار IDE + چت + اسناد)؛ اسکریپت‌نویسی سبک یا اتصال‌دهی no-code.
مهارت‌های prompt-engineering	کتابخانه‌ها و قالب‌های prompt قابل‌استفاده مجدد برای یک جریان کاری می‌سازد.
استفاده از ابزار هوش مصنوعی	چند ابزار هوش مصنوعی به‌صورت همزمان برای یک فرایند منسجم به کار گرفته می‌شوند.
کاربرد ویژه شغل	کمک سرتاسری (end-to-end) روی یک خروجی تحویلی تکرارشونده (برای مثال، بازبینی PR + تولید تست؛ خط لوله تولید گزارش).
مهارت‌های اتوماسیون	مونتاژ جریان کاری no-code/low-code؛ promptهای زمان‌بندی‌شده؛ هنوز اتوماسیون در سطح production نیست.
مهارت‌های داده/RAG	از بازیابی (retrieval) فراهم‌شده توسط ابزار روی پایگاه‌های دانش تأییدشده استفاده می‌کند؛ اثرهای chunking/grounding را درک می‌کند.
مهارت‌های Agentic-AI	از دستیارها/ابزارهای تک‌مرحله‌ای استفاده می‌کند؛ با مفاهیم agent آشنا است.
مهارت‌های ایمنی و حاکمیت هوش مصنوعی	**نقاط بازبینی انسانی را در دل جریان کاری** تعبیه می‌کند؛ مستند می‌سازد که کدام مراحل با کمک هوش مصنوعی انجام شده‌اند؛ داده‌های حساس را از مسیرهای (hop) تأییدنشده دور نگه می‌دارد.
آگاهی از ریسک بانکی/مقرراتی	می‌داند که مراحل جریان کاری با کمک هوش مصنوعی ممکن است قابل‌حسابرسی باشند؛ سابقه‌ای از جایی که هوش مصنوعی به یک خروجی تحویلیِ مشمول مقررات دست زده است نگه می‌دارد.
شواهد لازم برای HR	جریان کاری مستندشده + شاخص‌های پیش/پس (زمان چرخه، نرخ نقص، دوباره‌کاری).
نمونه ارزیابی عملی	یک کار تکرارشونده را پیش از هوش مصنوعی در برابر پس از آن نشان دهد، همراه با زمان صرفه‌جویی‌شدهٔ اندازه‌گیری‌شده و یک مرحله بررسی کیفیت.
نمونه‌های ویژه نقش	توسعه‌دهنده با جریان کاری PR/تست به‌همراه هوش مصنوعی در حلقه (AI-in-loop)؛ تحلیلگر با یک جریان استانداردشدهٔ پژوهش‌به‌گزارش مبتنی بر هوش مصنوعی.
سیگنال ترفیع	جریان کاری او توسط هم‌تیمی‌ها اتخاذ می‌شود؛ دستاورد قابل‌اندازه‌گیری در سطح تیم.
آموزش لازم برای سطح بعدی	اشتراک‌گذاری/استانداردسازی الگوهای هوش مصنوعی برای دیگران؛ مربی‌گری؛ روش‌های ارزیابی سبک.
ضدالگوها	بهینه‌سازی سرعت فردی در حالی که دوباره‌کاری در مراحل پایین‌دستی افزایش می‌یابد (همان هشدار پایداری DORA).

فیلد	محتوا
پرچم‌های قرمز	دسته‌های بزرگ‌تر و شلخته‌تر؛ ورود خروجی بازبینی‌نشدهٔ هوش مصنوعی به مصنوعات مشترک.

L6 — کاربر حرفه‌ای هوش مصنوعی (AI Power User)

حوزه	محتوا
مبنای پژوهشی	بُعد اثرگذاری (influence) در SFIA: در این سطح فرد بر شیوهٔ کار یک تیم اثر می‌گذارد، نه صرفاً بر خروجی شخصی خود؛ مدل‌های مبتنی بر حوزه‌های قابلیت، توانمندی فرابخشی را پاداش می‌دهند.
خلاصه	متخصص شناخته‌شدهٔ هوش مصنوعی در تیم/اسکواد که استفادهٔ مؤثر همگان از AI را ارتقا می‌دهد و هنجارهای ایمن محلی را تعریف می‌کند.
رفتار مورد انتظار	منتورینگ می‌کند، کتابخانه‌های prompt را گردآوری و سازمان‌دهی می‌کند، قراردادها و عرف‌های AI تیم را تعریف می‌کند، ابزارهای جدید را ارزیابی می‌کند و از انضباط راستی‌آزمایی (verification) پشتیبانی می‌کند.
مهارت‌های فنی	تسلط عمیق بر کل مجموعه‌ابزار مجاز؛ توانایی مقایسهٔ مدل‌ها/ابزارها روی وظایف واقعی.
مهارت‌های مهندسی prompt	طراحی الگوهای prompt مقاوم، قابل استفادهٔ مجدد و مقاوم در برابر تزریق (injection-resistant) برای تیم.
استفاده از ابزار AI	ابزارها را ارزیابی و معرفی می‌کند؛ پیش‌فرض‌ها و حفاظ‌ها (guardrails) را برای تیم تنظیم می‌کند.
کاربرد ویژهٔ شغل	کیفیت/سرعت خروجی کل یک تیم را بالا می‌برد؛ نشست‌های نمایش و آموزش داخلی (show-and-tell) برگزار می‌کند.
مهارت‌های اتوماسیون	اتوماسیون‌های مشترک no-code/low-code را همراه با دروازه‌های بازبینی (review gates) می‌سازد.
مهارت‌های Data/ RAG	پایگاه‌های دانش تیمی را برای بازیابی (retrieval) گردآوری می‌کند؛ مصالحه‌های (trade-off) کیفیت زمینه‌سازی (grounding) را درک می‌کند.
مهارت‌های Agentic-AI	قابلیت‌های دستیار/agent را به‌صورت کنترل‌شده برای تیم آزمایش (pilot) می‌کند.
مهارت‌های ایمنی و حاکمیت AI	**مالک هنجارهای استفادهٔ ایمن در سطح تیم است**؛ از رعایت قواعد راستی‌آزمایی و داده اطمینان حاصل می‌کند؛ نخستین مرجع پاسخ به پرسش‌های «آیا این مجاز است؟»؛ موارد را به حاکمیت (governance) ارجاع می‌دهد.
آگاهی از ریسک بانکی/مقرراتی	می‌داند کدام موارد کاربرد نیازمند دخالت واحد ریسک/تطبیق (compliance) هستند و آن‌ها را به‌درستی مسیریابی می‌کند.
شواهد لازم برای HR	محصولات تیمی پذیرفته‌شده (کتابخانهٔ prompt، قراردادها)؛ بهبود اندازه‌گیری‌شده در سطح تیم؛ سابقهٔ منتورینگ.
نمونهٔ ارزیابی عملی	برگزاری یک کارگاه (clinic) ۳۰ دقیقه‌ای که جریان کاری دو همکار را به‌طور قابل‌اندازه‌گیری بهبود دهد.
نمونه‌های نقش‌محور	توسعه‌دهندهٔ ارشد به‌عنوان سرپرست AI اسکواد؛ تحلیل‌گر ارشدی که استفاده از AI را در یک desk استانداردسازی می‌کند.
سیگنال ارتقا	به مرجع اصلی (go-to) تبدیل می‌شود؛ الگوهای او فراتر از تیم بلافصل گسترش می‌یابند.

حوزه	محتوا
آموزش لازم برای سطح بعدی	یکپارچه‌سازی برنامه‌نویسی‌شدهٔ AI (API)ها، چارچوب‌های ارزیابی (eval harnesses)، مهندسی RAG، مبانی MLOps).
ضدالگوها	تبلیغ ابزار بدون اندازه‌گیری؛ ترویج ابزارهایی که آشفتگی (churn) را افزایش می‌دهند.
پرچم‌های قرمز	تشویق سرعت به‌جای کنترل‌ها؛ گسترش میان‌بُرهای غیررسمی در مدیریت داده.

L7 — مشارکت‌کننده در خودکارسازی هوش مصنوعی (AI Automation Contributor)

حوزه	محتوا
مبنای پژوهشی	نردبان‌های خودمختاری ارائه‌شده توسط فروشندگان (دروازه‌های Read/Write/Delete در GitLab Duo؛ مسیر build→benchmark→monitor در ServiceNow) گذار از حالت کمک‌کننده به سیستم‌های کنش‌گر را به‌مثابه آستانه‌ای کنترل‌شده در نظر می‌گیرند؛ به همین دلیل دروازه‌های تأیید/بازگشت (rollback) در همین سطح وارد می‌شوند.
خلاصه	خودکارسازی‌های مبتنی بر هوش مصنوعی را در محیط تولید (production) با استفاده از APIها و با کنترل‌های صریح تأیید و بازگشت می‌سازد.
رفتار مورد انتظار	کدی می‌نویسد که مدل‌ها/ابزارها را برای انجام کار واقعی فراخوانی می‌کند؛ طراحی را برای مدیریت خطا، ثبت رویداد (logging) و بازگشت‌پذیری انجام می‌دهد؛ خروجی را پشت دروازه‌های بازبینی منتشر می‌کند.
مهارت‌های فنی	یکپارچه‌سازی LLM API، فراخوانی ساختارمند/تابعی (structured/function calling)، مدیریت خطا، ثبت رویداد، مدیریت اسرار (secrets).
مهارت‌های مهندسی پرامپت	پرامپت‌نویسی برنامه‌محور، طرح‌واره‌ها/اعتبارسنجی خروجی، تنظیم پرامپت مبتنی بر ارزیابی (eval-driven).
استفاده از ابزار هوش مصنوعی	SDKها و APIهای پلتفرم؛ یکپارچه‌سازی با CI.
مهارت‌های خودکارسازی	خودکارسازی‌های تولیدی با **تأیید انسان در حلقه (human-in-the-loop) برای کنش‌های نوشتن/حذف و مسیر بازگشت آزمون‌شده**؛ خاصیت تکرارپذیری بی‌اثر (idempotency)؛ پایش.
مهارت‌های Data/RAG	خطوط لوله پایه RAG (دریافت، تکه‌سازی، جاسازی، بازیابی) را روی منابع تأییدشده می‌سازد.
مهارت‌های هوش مصنوعی عامل‌محور (Agentic-AI)	دستیارهای ابزارمحور تک‌منظوره با دامنه محدود؛ بدون خودمختاری باز و بی‌انتها.
مهارت‌های ایمنی و حاکمیت هوش مصنوعی	پیاده‌سازی **اقدامات کاهشی OWASP LLM Top 10** (به‌ویژه تزریق پرامپت در رتبه #1)، فیلترینگ ورودی/خروجی، دسترسی ابزار با کمترین امتیاز (least-privilege)، ثبت رویداد ممیزی؛ ثبت هر کنش خودکار.
آگاهی از ریسک بانکی/مقرراتی	می‌داند خودکارسازی‌هایی که فرایندهای تحت‌نظارت را لمس می‌کنند پیش از ورود به تولید نیازمند تأیید ریسک مدل/حاکمیت هستند؛ هرگز در گام‌های مؤثر بر مشتری بدون دروازه به‌صورت خودکار کنش نمی‌کند.
شواهد مورد نیاز برای HR	خودکارسازی منتشرشده در محیط تولید همراه با دروازه‌های تأیید مستند، بازگشت، گزارش‌ها (logs) و یک نتیجه ارزیابی/سنجش (eval/benchmark).

حوزه	محتوا
نمونه ارزیابی عملی	ساخت یک خودکارسازی که پیش‌نویس تهیه و مسیریابی می‌کند (نه ارسال خودکار) همراه با یک مسیر بازگشت و یک آزمون مقاومت در برابر تزریق.
نمونه‌های نقش‌محور	مهندسی که یک ربات دسته‌بندی (triage) مبتنی بر هوش مصنوعی با تأیید انسانی می‌سازد؛ تحلیلگر داده‌ای که یک خط لوله گزارش مبتنی بر داده‌های مستند (grounded) را خودکار می‌کند.
نشانه ارتقا	خودکارسازی‌ها به‌طور قابل‌اتکا در محیط تولید اجرا می‌شوند و از بازبینی کنترلی عبور می‌کنند.
آموزش لازم برای سطح بعدی	طراحی سیستم/راه‌حل؛ معماری بازیابی؛ چارچوب‌های ارزیابی؛ تصمیم‌گیری درباره بده‌بستان‌ها (trade-off) در میان تیم‌ها.
ضدالگوها	اجرای خودکار کنش‌های نوشتن/حذف بدون تأیید؛ خودکارسازی در سطح «نمایشی» (demo-grade) در محیط تولید.
پرچم‌های قرمز	نبود بازگشت، نبود ثبت رویداد، مجوزهای ابزار بیش از حد گسترده، سطح حمله تزریق پرامپت بدون اقدام کاهشی.

L8 — طراح راهکار هوش مصنوعی (AI Solution Designer)

حوزه	محتوا
مبنای پژوهشی	بر اساس SFIA، پیچیدگی و میزان اثرگذاری تا سطح معماری راهکارها افزایش می‌یابد؛ مدل‌های مبتنی بر حوزه‌های قابلیتی (مدل «rewired» مک‌کینزی/McKinsey) به مالکیت سرتاسری راهکار در میان حوزه‌های مختلف پاداش می‌دهند.
خلاصه	راهکارهای سرتاسری هوش مصنوعی را برای یک مسئله کسب‌وکاری طراحی می‌کند، الگو، مدل و کنترل‌های مناسب را انتخاب می‌نماید و ارزش آن را با ارزیابی (evaluation) اثبات می‌کند.
رفتار مورد انتظار	میان build/buy و میان RAG در برابر fine-tune در برابر prompt انتخاب می‌کند، eval‌ها و SLA‌ها را تعریف می‌نماید، سطح کنترل (control surface) را طراحی می‌کند و موازنه‌های هزینه/تأخیر/ریسک را توجیه می‌سازد.
مهارت‌های فنی	معماری در سطح مدل‌ها، بازیابی (retrieval)، ارکستراسیون، ارزیابی، رصدپذیری (observability) و مهندسی هزینه/تأخیر.
مهارت‌های مهندسی پرامپت	راهبرد پرامپت و زمینه (context) در سطح سیستمی؛ مبتنی بر eval و نه به‌صورت موردی و بداهه.
استفاده از ابزارهای هوش مصنوعی	تصمیم‌گیری در سطح پلتفرم؛ استانداردهایی را تعیین می‌کند که دیگران بر اساس آن‌ها می‌سازند.
مهارت‌های اتوماسیون	الگوهای تأیید/بازگشت (approval/rollback) و پایش را تعریف می‌کند که L7 آن‌ها را پیاده‌سازی می‌نماید.
مهارت‌های Data/RAG	معماری RAG در سطح production: کیفیت داده، تازگی داده، ارزیابی بازیابی، تضمین‌های مبتنی‌سازی (grounding) و مجموعه‌داده‌های مرجع (ground-truth).
مهارت‌های هوش مصنوعی عامل‌محور (Agentic-AI)	طراحی راهکارهای عاملی چندابزاره و محدودشده همراه با guardrailها و مسیرهای ارجاع (escalation).

حوزه	محتوا
مهارت‌های ایمنی و حاکمیت هوش مصنوعی	چارچوب **NIST AI RMF (Govern/Map/Measure/Manage)** را در دل طراحی می‌گنجاند؛ ارزیابی Responsible-AI، شناسنامه مدل (model factsheet) و شواهد ارزیابی را به‌عنوان خروجی‌های تحویل‌دادنی تولید می‌کند.
آگاهی از ریسک بانکی/مقرراتی	طراحی را به‌گونه‌ای انجام می‌دهد که از اعتبارسنجی ریسک مدل **SR 11-7** عبور کند: مستندسازی، قلاب‌های اعتبارسنجی مستقل (independent-validation hooks)، پایش و راهکار جایگزین (fallback)؛ و شورای بازنگری هوش مصنوعی را زودهنگام درگیر می‌سازد.
شواهد مورد نیاز برای منابع انسانی	یک راهکار تحویل‌داده‌شده به‌همراه سند معماری، نتایج eval/benchmark، ارزیابی Responsible-AI و بازنگری موفق ریسک مدل.
نمونه ارزیابی عملی	تولید یک طراحی برای یک کاربرد تحت‌مقررات، شامل انتخاب الگو، برنامه eval، کنترل‌ها و نحوه عبور آن از SR 11-7.
نمونه‌های نقش‌محور	مهندس ارشد/معمار (staff engineer/architect) که یک دستیار مشاوره‌ای مبتنی‌سازی‌شده (grounded) طراحی می‌کند؛ سرپرست ML که یک راهکار امتیازدهی (scoring) تحت پایش طراحی می‌نماید.
نشانه ارتقا	طراحی‌ها به‌عنوان معماری‌های مرجع بازاستفاده می‌شوند؛ راهکارها در همان بار نخست از حاکمیت عبور می‌کنند.
آموزش لازم برای سطح بعدی	مهندسی پلتفرم/چارچوب عاملی (agent-framework)؛ ارکستراسیون چندعاملی (multi-agent)؛ توانمندسازی بازاستفاده‌پذیر در مقیاس.
الگوهای نامطلوب	انتخاب fine-tuning/عامل‌ها به‌خاطر اعتبار و نام‌آوری به‌جای تناسب با مسئله؛ طراحی بدون برنامه eval یا برنامه کنترل.
پرچم‌های قرمز	نبود راهبرد اندازه‌گیری؛ افزودن کنترل‌ها به‌صورت الصاقی پس از طراحی؛ نادیده‌گرفتن الزامات اعتبارسنجی.

L9 — سازندهٔ پلتفرم/Agent هوش مصنوعی (AI Agent / Platform Builder)

حوزه	محتوا
مبنای پژوهشی	بلوغ platform-engineering در CNCF و سطح‌بندی AIOps، پلتفرم‌های قابل استفادهٔ مجدد و خودمختاری قابل اعتماد را به‌عنوان پیشرفته‌ترین قابلیت مهندسی در نظر می‌گیرند؛ این بالاترین سطحِ ساخت در نقش مشارکت‌کنندهٔ فردی است.
خلاصه	چارچوب‌های agentic و پلتفرم‌های مشترک هوش مصنوعی را می‌سازد که بسیاری از تیم‌ها بر پایهٔ آن کار می‌کنند، با ایمنی و حاکمیت تعبیه‌شده در طراحی.
رفتار مورد انتظار	پلتفرم‌های قابل استفادهٔ مجددِ agent/orchestration، لایه‌های guardrail، چارچوب‌های eval و مجموعه‌دادگان مرجع (golden datasets) را مهندسی می‌کند که در سطح کل سازمان به‌کار می‌روند.
مهارت‌های فنی	orchestration چندعاملی (multi-agent)، چارچوب‌های ابزار/مجوز، پایداری پلتفرم، زیرساخت مقیاس‌پذیر RAG/vector و MLOps/LLMOps در مقیاس بالا.
مهارت‌های Prompt-engineering	سامانه‌های prompt/guardrail و لایه‌های دفاع در برابر injection را به‌عنوان قابلیت‌های پلتفرم می‌سازد.

حوزه	محتوا
استفاده از ابزار هوش مصنوعی	پلتفرم/SDK داخلی را فراهم می‌کند که سطوح L4 تا L8 از آن مصرف می‌کنند.
مهارت‌های اتوماسیون	**چارچوب سطح‌سازمانیِ تأیید، kill-switch، rollback و audit** را می‌سازد که همهٔ اقدامات خودکار را بر اساس سطح ریسک کنترل می‌کند.
مهارت‌های داده/RAG	پلتفرم بازیابی سازمانی: دسترسی به دادهٔ حاکمیت‌شده، lineage، مجموعه‌دادگان ارزیابی و SLAهای تازگی/کیفیت.
مهارت‌های Agentic-AI	سامانه‌های چندعاملی عملیاتی با خودمختاری محدود، escalation، قابلیت مشاهده‌پذیری و مجوزهای اقدام سطح‌بندی‌شده (Read خودکار / Write+Delete کنترل‌شده).
مهارت‌های ایمنی و حاکمیت هوش مصنوعی	**کنترل‌های منطبق با ISO/IEC 42001** را به‌عنوان قابلیت‌های پلتفرم پیاده‌سازی می‌کند؛ موجودی مدل‌ها (model inventory)، تولید factsheet و تله‌متری شواهد کنترل را می‌سازد.
آگاهی از ریسک بانکی/ مقرراتی	**gateway مجاز (sanctioned-gateway)** و قلاب‌های سه‌خطی دفاع (three-lines-of-defense) را می‌سازد؛ تضمین می‌کند که هر تیم مصرف‌کننده به‌صورت پیش‌فرض کنترل‌های منطبق‌برمقررات را به ارث می‌برد.
شواهد مورد نیاز برای HR	یک چارچوب پلتفرم/agent که توسط چند تیم به‌کار گرفته شده باشد، همراه با guardrailهای مستند، audit/تله‌متری و تأییدیهٔ حاکمیتی.
نمونهٔ ارزیابی عملی	نمایش یک پلتفرم چندعاملی با کنترل سطح‌بندی‌شدهٔ اقدام، kill-switch، رد ممیزی (audit trail) و یک eval دفاع در برابر injection.
نمونه‌های مختص نقش	مهندس ارشد (Principal engineer) مالک پلتفرم داخلی هوش مصنوعی؛ سرپرست چارچوب agent.
سیگنال ارتقا	پلتفرم به بستر پیش‌فرض تبدیل می‌شود؛ کنترل‌ها در سطح کل سازمان مقیاس می‌یابند.
آموزش لازم برای سطح بعدی	استراتژی سازمانی، سیاست‌گذاری، تعامل مقرراتی، ریسک سبد (portfolio risk) و نفوذ اجرایی.
الگوهای نامطلوب	خودمختاری بدون kill-switch؛ پلتفرمی که به تیم‌ها اجازهٔ دور زدن کنترل‌ها را می‌دهد.
نشانه‌های هشدار	مجوزهای نامحدود برای agent؛ نبود رد ممیزی؛ نبود موجودی مدل.

L10 — رهبر استراتژی و حاکمیت هوش مصنوعی (AI Strategy / Governance Leader)

فیلد	محتوا
مبنای پژوهشی	بخش «Govern» در NIST AI RMF و رویه‌های بانکی (شوراهای بازنگری هوش مصنوعی در HSBC و دروازه‌های مجاز) استراتژی سازمانی هوش مصنوعی و حاکمیت ریسک مدل را به‌عنوان یک کارکرد رهبری مجزا و فراتر از هر نقش ساخت‌وساز در نظر می‌گیرند.
خلاصه	استراتژی سازمانی هوش مصنوعی را تعیین می‌کند و چارچوب حاکمیت، ریسک و سیاست‌گذاری را که به سازمان اجازه می‌دهد هوش مصنوعی را به‌صورت ایمن گسترش دهد، در اختیار دارد.

فیلد	محتوا
رفتار مورد انتظار	استراتژی و سرمایه‌گذاری در حوزه هوش مصنوعی را تعریف می‌کند؛ ریاست یا مشاوره شورای بازنگری هوش مصنوعی را بر عهده دارد؛ مالک مسیر بلوغ حاکمیتی است؛ و میان نوآوری و ریسک نظارتی توازن برقرار می‌کند.
مهارت‌های فنی	سواد فنی گسترده‌ای که برای به‌چالش‌کشیدن طراحی‌ها و اعتبارسنج‌ها کافی باشد؛ نقش عملی و دست‌به‌کد ندارد.
مهارت‌های مهندسی پرامپت	در سطح وظیفه کاربردی ندارد (N/A)؛ استانداردها را تعیین می‌کند.
استفاده از ابزارهای هوش مصنوعی	سیاست ابزارهای مجاز و استانداردهای تدارکات را تعریف می‌کند.
مهارت‌های اتوماسیون	مالک سیاست رده‌بندی ریسک است که تعیین می‌کند چه چیزی اساساً می‌تواند خودکارسازی شود.
مهارت‌های داده/RAG	مالک حاکمیت داده سازمانی و سیاست داده هوش مصنوعی است.
مهارت‌های هوش مصنوعی عامل‌محور (Agentic AI)	سیاست اعتماد به خودمختاری و شرایطی را که تحت آن عامل‌ها مجاز به اقدام هستند تعیین می‌کند.
مهارت‌های ایمنی و حاکمیت هوش مصنوعی	**مالک کل چارچوب حاکمیتی است**: برنامه NIST AI RMF، سیستم مدیریتی ISO/IEC 42001، سیاست هوش مصنوعی مسئولانه (Responsible-AI)، انطباق با سواد هوش مصنوعی (EU AI Act) و پاسخ به رخداد.
آگاهی از ریسک بانکی/نظارتی	مالک **همسویی با SR 11-7، سه خط دفاع (three-lines-of-defense)، تعامل با نهاد ناظر، تأیید شورا و فهرست موجودی مدل‌ها** در مقیاس سازمانی است.
شواهد مورد نیاز برای HR	استراتژی سازمانی هوش مصنوعی و چارچوب حاکمیتی تصویب‌شده؛ نتایج ممیزی/نظارتی؛ وضعیت ریسک سبد (portfolio)؛ سابقه رهبری شورا.
نمونه ارزیابی عملی	ارائه یک استراتژی سازمانی هوش مصنوعی همراه با نقشه راه موازی بلوغ حاکمیتی و تبیین اینکه چگونه نیازهای نهادهای ناظر و SR 11-7 را برآورده می‌کند.
نمونه‌های ویژه نقش	رئیس هوش مصنوعی / مدیر ارشد هوش مصنوعی (Chief AI Officer)؛ مدیر ریسک و حاکمیت هوش مصنوعی؛ مالک سبد هوش مصنوعی در سطح CTO/CIO.
سیگنال ارتقا	استراتژی در سراسر سازمان پذیرفته می‌شود؛ هوش مصنوعی بدون نقض کنترل‌ها یا مشاهدات نظارتی گسترش می‌یابد.
آموزش لازم برای سطح بعدی	سطح نهایی — رشد بیشتر در گرو گستره اختیارات و نفوذ بر هیئت‌مدیره و نهادهای ناظر است.
الگوهای نادرست	استراتژی‌ای که اجازه می‌دهد قابلیت‌ها از کنترل‌ها پیشی بگیرند؛ حاکمیتی که صرفاً کاغذبازی است و نه اجرای واقعی.
پرچم‌های قرمز	نبود فهرست موجودی مدل‌ها، نبود برنامه پاسخ به رخداد، گسترش قابلیت‌ها پیش از بلوغ حاکمیتی.

10. نگاشت ۱۰ سطح به مدل‌های رایج ۴/۵/۶ سطحی

نردبان ۱۰ سطحی، ضمن آنکه دانه‌بندی دقیقی برای ارتقا فراهم می‌کند، به‌خوبی روی مدل‌های کوتاه‌تری که بازار عملاً به کار می‌برد (بخش ۵) فشرده می‌شود. نگاشت‌های زیر به منابع انسانی، مدیران ارشد و سنجش‌گری بیرونی (benchmarking) امکان می‌دهند هر کدام نردبان را در چارچوب موردپسند خود بخوانند. همه نگاشت‌ها **(استنتاج‌شده از منابع عمومی)** هستند؛ یعنی سطوح ما را با شکل چارچوب‌های عمومی همتراز می‌کنند، نه با هیچ نردبان منتشرشده‌ای از یک شرکت خاص.

سطح ما	۴ سطحی (Aware/Adopt/Scale/Lead)	بلوغ ۵ سطحی به سبک Gartner	باند شایستگی ۶ سطحی	سطح مسئولیت SFIA	باند Low/Med/High
L1 بدون مواجهه	Aware	1 — آگاهی	Band 0 — هیچ	(پایین‌تر از SFIA 1)	Low
L2 آگاهی نسبت به AI	Aware	1 — آگاهی	Band 1 — بنیادی	SFIA 1 — Follow	Low
L3 کاربر پایه AI	Adopt	2 — فعال (آزمایش‌گری)	Band 2 — مقدماتی	SFIA 2 — Assist	Low
L4 کاربر کارآمد Prompt	Adopt	2 — فعال	Band 3 — ماهر	SFIA 3 — Apply	Medium
L5 کاربر AI در گردش‌کار	Adopt ← Scale	3 — عملیاتی	Band 3 — ماهر	SFIA 3 — Apply	Medium
L6 کاربر حرفه‌ای AI	Scale	3 — عملیاتی	Band 4 — پیشرفته	SFIA 4 — Enable	Medium
L7 مشارکت‌کننده در خودکارسازی AI	Scale	4 — سامانمند (مدیریت‌شده)	Band 4 — پیشرفته	SFIA 4-5 — Enable/Ensure	High
L8 طراح راهکار AI	Scale ← Lead	4 — سامانمند	Band 5 — خبره	SFIA 5 — Ensure, advise	High
L9 سازنده Agent/ Platform هوش مصنوعی	Lead	5 — تحول‌آفرین	Band 6 — مرجع	SFIA 6 — Initiate, influence	High
L10 رهبر راهبرد/ حاکمیت AI	Lead	5 — تحول‌آفرین	Band 6 — مرجع	SFIA 7 — Set strategy, inspire	High

چگونه آن را بخوانیم. - **۴ سطحی (Aware/Adopt/Scale/Lead)** — مناسب‌ترین گزینه برای روایت مدیریتی و گزارش‌دهی پذیرش (بازتاب‌دهنده explore/pilot/scale/transform در Anthropic و خوشه‌های DORA). - **۵ سطحی به سبک Gartner** — مناسب‌ترین گزینه برای داشبوردهای بلوغ سازمانی/سبدی. - **باند شایستگی ۶ سطحی** — مناسب‌ترین گزینه برای باندبندی حقوق/رتبه در منابع انسانی، جایی که شش باند از پیش وجود دارد. - **سطح SFIA** — مناسب‌ترین گزینه برای پیوند خوردن به یک چارچوب مسئولیتی موجود و برای قابلیت مقایسه بیرونی. - **Low/Med/High** — سریع‌ترین باند غربالگری: Low = سواد/استفاده ایمن (L1-L3)، Medium = عمل‌کننده مولد (L4-L6)، High = سازنده/رهبر (L7-L10).

مسیر حاکمیتی به‌موازات پیش می‌رود (توصیه ترکیبی). سطح قابلیت هرگز نباید از بلوغ حاکمیتی سازمان برای همان مورد کاربرد فراتر رود. یک قاعده عملی: هر فردی که در L7+ فعالیت می‌کند باید در چارچوب کنترل‌هایی کار کند که مالکیت آن‌ها در L9-L10 قرار دارد، تا رشد مهارت فردی هیچ‌گاه از محیط کنترلی پیشی نگیرد (توصیه بخش ۲).

11. نگاشت به خانواده‌های شغلی و سطوح ارشدیت

این بخش کهن‌الگوهای ارشدیت را برای ۱۰ تیم به نردبان نگاشت می‌کند و برای هر کدام یک **کف** (حداقل الزامی برای ایفای نقش) و یک **هدف** (وضعیت پایدار مورد انتظار) تعیین می‌کند. کف‌ها خط پایه را تضمین می‌کنند: ذیل الزام سواد هوش مصنوعی در EU AI Act و قواعد سازمان در زمینه مدیریت داده، **هر نقشی کف سختگیرانه L2 (آگاهی از AI) دارد**؛ هر نقشی که در فرایند تحویل از AI استفاده می‌کند کف **L3 (استفاده روتین ایمن همراه با راستی‌آزمایی)** دارد. هدف‌ها **(استنتاج‌شده از منابع عمومی)** هستند — کالیبره‌شده با سطوح مسئولیت SFIA، نه با هیچ نردبان منتشرشده‌ی شرکتی.

کف + هدف ارشدیت (ستون فقرات عمومی).

ارشدیت	کف	هدف	ملاحظات
Junior	L3	L4	پیش از استفاده بدون نظارت از AI باید در حالت ایمن و راستی‌آزمایی‌شده باشد.
Mid	L4	L5-L6	انتظار می‌رود AI را در جریان‌های کاری جای دهد و آن را بسنجد.
Senior	L5	L6-L7	انتظار می‌رود تیم را ارتقا دهد و/یا اتوماسیون‌های دارای دروازه (gated) را عرضه کند.
Lead / Architect	L6	L8-L9	مالک طراحی راهکار و الگوهای مرجع است.
Manager (مدیر افراد)	L4	L6	باید AI را آن‌قدر خوب بفهمد که بتواند مربیگری کند، شواهد را بازبینی کند و کنترل‌ها را اعمال نماید؛ نیازی به ساختن ندارد.
Expert / Principal (IC)	L7	L9- L10	بالاترین اختیار ساخت؛ تأثیرگذاری بر پلتفرم و استراتژی.

سطح مورد انتظار به تفکیک تیم (کف ← هدف). کف‌های پایین‌تر از ستون فقرات عمومی مجاز نیستند؛ جایی که پروفایل ریسک یک تیم اقتضا کند، کف بالاتر برده می‌شود (با ذکر).

تیم	Junior	Mid	Senior	Lead/ Architect	Manager	Expert/ Principal
1. مهندسی نرم‌افزار (Software Engineering)	L3 → L4	L4 → L6	L5 → L7	L7 → L9	L4 → L6	L8 → L10
2. علم داده / ML (Data Science / ML)	L4 → L5	L5 → L7	L7 → L8	L8 → L9	L5 → L7	L9 → L10
3. مهندسی داده (Data Engineering)	L3 → L4	L4 → L6	L6 → L7	L7 → L8	L4 → L6	L8 → L9
4. پلتفرم / DevOps / SRE	L3 → L4	L4 → L6	L6 → L7	L7 → L9	L4 → L6	L8 → L10
5. مدیریت محصول (Product Management)	L3 → L4	L4 → L5	L5 → L6	L6 → L7	L4 → L6	L7 → L8

تیم	Junior	Mid	Senior	Lead/ Architect	Manager	Expert/ Principal
6. QA / مهندسی تست (Test Engineering)	→ L3 L4	→ L4 L5	→ L5 L7	L6 → L8	L4 → L5	L7 → L8
7. ریسک / ریسک مدل / انطباق (Risk / Model Risk / Compliance)	→ L3 L4	→ L4 L5	→ L5 L7	L7 → L9	L4 → L6	L8 → L10
8. امنیت / InfoSec (Security)	→ L3 L4	→ L4 L6	→ L6 L7	L7 → L9	L4 → L6	L8 → L10
9. تحلیلگران کسب‌وکار / عملیات (Business / Operations Analysts)	→ L3 L4	→ L4 L5	→ L5 L6	L6 → L7	L4 → L5	L7 → L8
10. طراحی / UX (Design)	→ L3 L4	→ L4 L5	→ L5 L6	L6 → L7	L4 → L5	L6 → L7

حداقل‌ها و قواعد. - **کف فراگیر: L2** برای هر کارمند (سواد + استفاده قابل‌قبول)، صرف‌نظر از نقش یا میزان استفاده از AI. - **کف تحویل: L3** برای هرکس که در محصول کاری خود از AI استفاده می‌کند (استفاده روتین ایمن + راستی‌آزمایی + انضباط در طبقه‌بندی داده). - **کف‌های بالاتر در نقاط تمرکز ریسک:** علم داده/ML از L4 آغاز می‌شود (آن‌ها مدل‌هایی می‌سازند که مشمول ریسک مدل‌اند)؛ ارشدهای ریسک/انطباق و امنیت باید به L7 برسند تا بتوانند به‌طور معتبر سامانه‌های AI تحت حاکمیت خود را بازبینی و به چالش بکشند. - **دروازه اتوماسیون:** هیچ‌کس بالاتر از L6 عمل نمی‌کند (یعنی اتوماسیون‌های عملیاتی، L7+، را عرضه کند) مگر در چارچوب کنترل‌های متعلق به L9-L10 (تأیید، بازگشت به عقب، حسابرسی). توانمندی هرگز از حاکمیت پیشی نمی‌گیرد. - **مدیران:** کف L4 تا بتوانند شواهد AI را بخوانند و کنترل‌ها را اعمال کنند؛ ملزم به ساختن نیستند (الزام L7+ ندارند) مگر آنکه مأموریت دوگانه IC/فنی داشته باشند. - **هدف‌ها وضعیت پایدار آرمانی‌اند**، نه تضمین ارتقا؛ ارتقا همچنان مستلزم شواهد مشخص‌شده برای هر سطح در بخش ۹ است.

---

12. راهنمای ارزیابی ویژه هر تیم

12.x DevOps / CI-CD

این زیربخش تعریف می‌کند که توانمندی هوش مصنوعی برای مهندسانی که خطوط لوله یکپارچه‌سازی پیوسته / تحویل پیوسته (CI-CD)، خودکارسازی استقرار و ابزارهای انتشار را در یک محیط بانکی تحت‌نظارت می‌سازند و بهره‌برداری می‌کنند، چه شکلی دارد. تأکید بر استفاده ایمن، قابل‌حسابرسی و قابل‌اندازه‌گیری از GenAI است — هرگز صرفاً بر سرعت. در یک بافت تحت‌نظارت، یک تغییر خط لوله که توسط هوش مصنوعی تولید شده، همچنان یک تغییر تولیدی است: باید از همان کنترل‌های بازبینی، تأیید، تفکیک وظایف (SoD) و بازگشت (rollback) عبور کند که هر تغییر نوشته‌شده توسط انسان از آن‌ها عبور می‌کند.

نگاهی کلی به نوارهای توانمندی

نوار	بازه نردبان	چه شکلی در DevOps / CI-CD دارد	وظایف معمول هوش مصنوعی انجام‌شده
LOW	L1-L3	از هوش مصنوعی برای نگارش مجزا و کم‌ریسک با راستی‌آزمایی کامل انسانی استفاده می‌کند. هرگز اسرار، نام میزبان تولیدی یا داده مشتری را در یک ابزار وارد نمی‌کند. خروجی پیش از آنکه به یک مخزن برسد بازبینی می‌شود.	پیش‌نویس یک مرحله Jenkinsfile یا یک قطعه YAML؛ توضیح یک خط لاگ خطادار؛ تولید اسکریپت‌های استقرار قالبی؛ پیش‌نویس اولیه یادداشت‌های انتشار از یک فهرست commit.
MEDIUM	L4-L6	از طریق پرامپت‌نویسی هدفمند نتایج قابل‌اتکا می‌گیرد و هوش مصنوعی را در کار چندمرحله‌ای خط لوله جای می‌دهد. هنجارهای ایمن برای تیم تعیین می‌کند و دستاوردها را می‌سنجد. هوش مصنوعی به تشخیص و تحلیل کمک می‌کند، اما انسان‌ها همچنان مالک تصمیم merge و استقرار هستند.	تشخیص ساختاریافته خرابی خط لوله؛ خلاصه‌های ریسک تغییر برای CAB؛ تحلیل و روایت معیارهای DORA؛ تحلیل گزینه‌های rollback؛ قالب‌سازی اجزای خط لوله قابل‌استفاده مجدد با پرامپت‌ها.
HIGH	L7-L10	خودکارسازی و پلتفرم‌های مجهز به هوش مصنوعی در سطح تولیدی می‌سازد که دروازه‌های تأیید، حفاظ‌ها و rollback در آن‌ها تعبیه شده است. مالک حاکمیت نحوه استفاده از هوش مصنوعی در زنجیره ابزار تحویل است یا آن را شکل می‌دهد.	خودکارسازی‌های مجهز به هوش مصنوعی پشت دروازه‌های تأیید/rollback؛ تولید policy-as-code راستی‌آزمایی‌شده در برابر کاتالوگ کنترل؛ ابزار agent/پلتفرم برای خودترمیمی خط لوله با حفاظ‌ها؛ استراتژی سازمانی AI-in-SDLC و وضعیت نظارتی.

نمونه‌های نگاشت‌شده به سطوح نردبان

سطح	نمونه عینی DevOps / CI-CD	کنترل ایمنی / داده موردنیاز
L1	هرگز از GenAI برای کار خط لوله استفاده نکرده؛ بی‌اطلاع از اینکه واردکردن یک `Jenkinsfile` با اعتبارنامه‌های تعبیه‌شده یک ریسک نشت داده است.	کاربرد ندارد — هدف برای ارتقا.
L2	می‌تواند توضیح دهد چه چیزی را می‌توان و نمی‌توان به یک ابزار هوش مصنوعی تأییدشده فرستاد (بدون اسرار، بدون IP تولیدی، بدون داده مشتری، بدون منطق کنترلی اختصاصی).	پیش از تایپ یک پرامپت، طبقه‌بندی داده و فهرست ابزارهای تأییدشده را می‌داند.
L3	یک اصلاح YAML lint یا یک گام استقرار shell ساده تولید می‌کند، آن را خط‌به‌خط می‌خواند و پیش از commit در یک sandbox آزمایش می‌کند.	خروجی را راستی‌آزمایی می‌کند؛ در محیط غیرتولیدی اجرا می‌کند؛ ورودی‌ها را پنهان‌سازی می‌کند.
L4	یک خرابی متناوب خط لوله را با دادن یک لاگ پاکسازی‌شده + بافت خط لوله به هوش مصنوعی و تکرار پرامپت‌ها برای جداسازی علت ریشه‌ای تشخیص می‌دهد.	لاگ‌ها از نام میزبان‌ها، توکن‌ها و شناسه‌های حساب پاکسازی شده‌اند.
L5	هوش مصنوعی را در سراسر یک گردش‌کار انتشار جای می‌دهد: یادداشت‌های انتشار را از تاریخچه commit پاکسازی‌شده تولید می‌کند، یک خلاصه ریسک تغییر پیش‌نویس می‌کند و زمان چرخه صرفه‌جویی‌شده را ردگیری می‌کند.	هر خروجی هوش مصنوعی از بازبینی PR موجود و تأیید CAB عبور می‌کند؛ دستاوردها سنجیده می‌شوند.
L6	به‌عنوان رهبر هوش مصنوعی تیم عمل می‌کند: قالب‌های پرامپت برای تولید Jenkinsfile/YAML و تحلیل rollback منتشر می‌کند، هنجار «بدون اسرار در پرامپت‌ها» را تعیین می‌کند و همکاران را مربیگری می‌کند.	مالک هنجارهای استفاده ایمن در سطح تیم؛ استفاده تیم را حسابرسی می‌کند.

سطح	نمونه عینی DevOps / CI-CD	کنترل ایمنی / داده موردنیاز
L7	یک خودکارسازی مجهز به هوش مصنوعی می‌سازد (مثلاً تولید یک طرح rollback یا PR) که فقط پشت تأیید دستی و یک دروازه rollback خودکار اجرا می‌شود.	دروازه تأیید + rollback خودکار + لاگ حسابرسی + SoD اعمال‌شده.
L8	یک راه‌حل سرتاسری مجهز به هوش مصنوعی را معماری می‌کند (مثلاً امتیازدهی ریسک تغییر که دروازه استقرار را تغذیه می‌کند) که از راستی‌آزمایی ریسک مدل عبور می‌کند.	راستی‌آزمایی مستند ریسک مدل؛ بازنویسی انسانی (override)؛ پایش.
L9	یک پلتفرم محافظت‌شده و قابل‌استفاده مجدد برای خودترمیمی خط لوله یا تولید policy-as-code/agent می‌سازد که در چند تیم استفاده می‌شود.	حفاظ‌های سراسری سازمان، لاگ‌گیری متمرکز، سیاست نسخه‌بندی‌شده، کلید قطع (kill switch).
L10	مالک استراتژی سازمانی برای هوش مصنوعی در زنجیره ابزار تحویل و وضعیت نظارتی/حاکمیتی مرتبط است.	حاکمیت سازمانی هوش مصنوعی، کنترل‌های رو به ناظر، مالکیت ریسک.

پوشش توانمندی به وظیفه

وظیفه هوش مصنوعی	LOW (L1-L3)	MEDIUM (L4-L6)	HIGH (L7-L10)
تولید Jenkinsfile / YAML	پیش‌نویس مراحل منفرد، راستی‌آزمایی دستی	تولید قالب‌های قابل‌استفاده مجدد و پارامتری	قالب‌سازی خودکار از طریق ابزار محافظت‌شده
تشخیص خط لوله	توضیح یک خط لاگ	علت ریشه‌ای ساختاریافته از لاگ‌های پاکسازی‌شده	تریاژ خودکار با دروازه انسانی
اسکریپت‌های استقرار	قالبی، آزمایش‌شده در sandbox	چندمحیطی، بازبینی‌شده	تولیدشده پشت تأیید + rollback
یادداشت‌های انتشار	پیش‌نویس اولیه از commitها	پرداخت‌شده، تنظیم‌شده برای مخاطب، تکرارپذیر	خودکارشده در خط لوله انتشار
تحلیل ریسک تغییر	خلاصه‌سازی یک تغییر واحد	روایت ریسک آماده برای CAB	امتیازدهی ریسک که دروازه استقرار را تغذیه می‌کند
تحلیل DORA	خواندن داشبورد	تحلیل روندها + روایت	طراحی/خودکارسازی اندازه‌گیری
توصیه rollback	پیشنهاد یک گزینه برای راستی‌آزمایی	مقایسه گزینه‌ها با مصالحه‌ها	تولید طرح rollback راستی‌آزمایی‌شده پشت دروازه
خودکارسازی CI/CD	کاربرد ندارد (فقط مصرف‌کننده)	نمونه‌سازی اسکریپت‌های کمکی	خودکارسازی تولیدی با حفاظ‌ها
استفاده ایمن از کد هوش مصنوعی	بازبینی هر خط	تعیین هنجارهای بازبینی تیم	اعمال سیاست راستی‌آزمایی سراسری سازمان
Policy-as-code	خواندن سیاست موجود	پیش‌نویس/ویرایش سیاست با بازبینی	تولید سیاست راستی‌آزمایی‌شده در مقیاس پلتفرم

شواهدی که HR می‌تواند راستی‌آزمایی کند

HR نیازی به قضاوت درباره کیفیت کد ندارد. HR راستی‌آزمایی می‌کند که شواهد وجود دارند، تاریخ‌دار هستند و نشان می‌دهند کنترل‌های درست اعمال شده‌اند.

نوع شواهد	به دنبال چه باشید	تأیید می‌کند
PRها / merge requestهای ادغام‌شده	تغییر مجهز به هوش مصنوعی با تأیید بازبین ثبت‌شده؛ رعایت SoD (نویسنده ≠ تأییدکننده)	نگارش ایمن L3+
قالب‌های پرامپت / runbookها	دارایی‌های پرامپت مشترک و تحت‌کنترل نسخه که توسط مهندس نوشته شده‌اند	تعیین هنجار L6
سوابق خط لوله / CAB	خلاصه‌های ریسک تغییر، یادداشت‌های انتشار، طرح‌های rollback منتسب به کار مجهز به هوش مصنوعی، با مسیر تأیید	استفاده گردش‌کاری L5-L7
داشبوردها / گزارش‌های DORA	معیارهای قبل/بعد درباره فراوانی استقرار، زمان پیش‌رو (lead time)، نرخ شکست تغییر، MTTR همراه با تحلیل مهندس	دستاوردهای سنجیده‌شده L5+
پیکربندی خودکارسازی	حضور دروازه‌های تأیید، دروازه‌های rollback، لاگ‌گیری حسابرسی در هر خودکارسازی متأثر از هوش مصنوعی	خودکارسازی محافظت‌شده L7+
اسناد ریسک مدل / حاکمیت	تأیید راستی‌آزمایی، کلید قطع، پایش برای راه‌حل‌ها/پلتفرم‌های هوش مصنوعی	L8-L10
گواهی استفاده از ابزار / مدیریت داده	بدون اسرار، شناسه‌های تولیدی یا داده مشتری در پرامپت‌ها؛ فقط ابزارهای تأییدشده	همه سطوح — کنترل پایه

نمونه‌های آزمون عملی

آزمون 1 — تشخیص خط لوله و نگارش ایمن (هدف L4) به مهندس یک اجرای خط لوله خطادار با یک لاگ واقعی اما از پیش پاکسازی‌شده و `Jenkinsfile` /YAML مرتبط بدهید. از او بخواهید از ابزار هوش مصنوعی تأییدشده برای شناسایی علت ریشه‌ای و پیشنهاد یک اصلاح استفاده کند. معیار قبولی: به‌درستی تشخیص می‌دهد که ورودی‌ها از پیش پاکسازی شده‌اند و تأیید می‌کند هیچ اسرار بیشتری افشا نشده است؛ از پرامپت‌نویسی هدفمند و تکراری استفاده می‌کند؛ یک اصلاح درست تولید می‌کند؛ بیان می‌کند که اصلاح باید از بازبینی PR عبور کند و ابتدا در محیط غیرتولیدی اجرا شود. مردود اگر اعتبارنامه‌ها/نام میزبان‌های خام را وارد کند یا خروجی هوش مصنوعی را بدون راستی‌آزمایی آماده merge بداند.

آزمون 2 — خودکارسازی rollback محافظت‌شده (هدف L7) یک سناریو فراهم کنید: یک استقرار در حال تخریب یک سرویس است. از مهندس بخواهید یک توصیه/خودکارسازی rollback مجهز به هوش مصنوعی طراحی کند. معیار قبولی: هوش مصنوعی طرح rollback را پیشنهاد می‌دهد، اما اجرا پشت یک دروازه تأیید دستی، یک دروازه rollback/سلامت خودکار، لاگ‌گیری حسابرسی کامل و SoD اعمال‌شده قرار دارد؛ بازنویسی انسانی صریح است. مردود اگر به هوش مصنوعی اجازه داده شود به‌طور خودمختار یا بدون مسیر حسابرسی روی تولید عمل کند.

سطوح هدف توصیه‌شده

جمعیت	هدف پس از 4 ماه	هدف پس از 12 ماه
همه مهندسان DevOps / CI-CD (پایه)	L3 — استفاده روتین ایمن و راستی‌آزمایی‌شده؛ صفر رخداد مدیریت داده	L4-L5 — پرامپت‌نویسی قابل‌اتکا جای‌گرفته در گردش‌کارهای انتشار با دستاوردهای سنجیده‌شده
مهندسان ارشد / رهبران	L4-L5	L6-L7 — تعیین هنجار تیم و خودکارسازی تولیدی محافظت‌شده
متخصصان پلتفرم / ابزار	L5	L7-L8 — خودکارسازی محافظت‌شده و راه‌حل‌های هوش مصنوعی راستی‌آزمایی‌شده
مدیر یا اصلی (principal) DevOps / پلتفرم	L5-L6	L8+ — طراحی راه‌حل و مشارکت در حاکمیت AI-in-SDLC

نکته: پیشرفت فراتر از L7 مستلزم انضباط اثبات‌شده در دروازه تأیید، rollback و حسابرسی است، نه صرفاً مهارت فنی. در یک بستر تحت‌نظارت، رفتار کنترلی معیار تعیین‌کننده است، نه هوشمندی خودکارسازی.

(برداشت‌شده از منابع عمومی): تعاریف نردبان و سرعت سطح هدف از رویه عمومی در دسترس DevOps/DORA (فراوانی استقرار، زمان پیش‌رو، نرخ شکست تغییر، MTTR) و کنترل‌های متداول مدیریت تغییر صنایع تحت‌نظارت (تفکیک وظایف، تأیید CAB، مسیرهای حسابرسی، راستی‌آزمایی ریسک مدل) ترکیب شده‌اند. هیچ نردبان HR داخلی یا اختصاصی شرکتی استفاده نشده است.

x SRE / Operations.12

هوش مصنوعی در مهندسی قابلیت اطمینان سایت (SRE) با سیستم‌های تولیدی، دسترس‌پذیری رو به مشتری و تله‌متری حساس سروکار دارد. در یک بافت خدمات مالی تحت‌نظارت، آستانه بالاتر از اکثر تیم‌هاست: یک پیشنهاد هوش مصنوعی که یک اقدام اصلاحی ناامن را راه‌اندازی کند، می‌تواند موجب قطعی یا نقض کنترل شود. توانمندی در اینجا کمتر بر اساس پرامپت‌نویسی هوشمندانه و بیشتر بر اساس **کنترل منضبط انسان‌درحلقه (human-in-the-loop)، آگاهی از شعاع انفجار (blast radius) و بهداشت داده در لاگ‌ها و معیارها** سنجیده می‌شود.

توانمندی چه شکلی دارد

نوار	رفتار در عمل	سطح نردبان معمول
LOW	هشدارها یا خطوط لاگ خام را در یک ابزار GenAI وارد می‌کند تا «بپرسد چه اشکالی هست»، بدون هیچ آگاهی از اینکه لاگ‌ها ممکن است حاوی PII، شماره حساب، اسرار یا نام میزبان‌های داخلی باشند. حدس‌های RCA را بدون راستی‌آزمایی در برابر تله‌متری واقعی می‌پذیرد. هیچ تصوری از ابزارهای تأییدشده ندارد. از خروجی هوش مصنوعی برای عمل مستقیم روی تولید استفاده می‌کند.	L0-L2
MEDIUM	از دستیارهای هوش مصنوعی تأییدشده برای خلاصه‌سازی رخدادها، خوشه‌بندی هشدارها و پیش‌نویس postmortemها استفاده می‌کند و همیشه پیش از اتکا به خروجی، آن را در برابر داشبوردهای منبع راستی‌آزمایی می‌کند. نمونه‌های لاگ را پاکسازی می‌کند یا از نمونه‌های پاکسازی‌شده استفاده می‌کند. از هوش مصنوعی برای پیشنهاد فرضیه‌های RCA و تطبیق runbook استفاده می‌کند، سپس آن‌ها را دستی تأیید می‌کند. هر پیشنهاد اصلاحی هوش مصنوعی را به‌عنوان یک پیشنهاد نیازمند تأیید انسانی تلقی می‌کند.	L3-L5
HIGH	گردش‌کارهای عملیاتی مجهز به هوش مصنوعی را با دروازه‌های تأیید و rollback می‌سازد و بهره‌برداری می‌کند: خط زمانی رخداد خودکار خلاصه‌شده، تحلیل نرخ سوختن (burn-rate)، پیش‌بینی ظرفیت و پیشنهادهای اصلاح ایمن که از طریق کنترل تغییر مسیریابی می‌شوند. هنجارهای ایمن برای تیم on-call تعیین می‌کند، تعریف می‌کند چه داده‌ای می‌تواند وارد پرامپت‌ها شود و دیگران را ارتقا می‌دهد. در بالاترین سطح، خودکارسازی محافظت‌شده‌ای را معماری می‌کند که از راستی‌آزمایی ریسک مدل و مدیریت تغییر عبور می‌کند.	L6-L10

نمونه‌های نگاشت‌شده به سطوح نردبان

حوزه توانمندی	L3-L4 (پایه / مؤثر)	L5-L6 (گردش‌کار / قدرتمند)	L7-L9 (خودکارسازی / پلتفرم)
خلاصه‌سازی رخداد و هشدار	از هوش مصنوعی می‌خواهد یک رشته هشدار پاکسازی‌شده را خلاصه کند، در برابر منبع هشدار بررسی می‌کند	یک قالب پرامپت استاندارد که کل چرخش on-call از آن استفاده می‌کند تدوین می‌کند؛ زمان تریاژ صرفه‌جویی‌شده را می‌سنجد	خط لوله به‌طور خودکار خط زمانی رخداد را از جریان رویدادها با پنهان‌سازی در زمان ingest پیش‌نویس می‌کند

حوزه توانمندی	L3-L4 (پایه / مؤثر)	L5-L6 (گردش‌کار / قدرتمند)	L7-L9 (خودکارسازی / پلتفرم)
تحلیل لاگ/معیار	یک قطعه لاگ پاکسازی‌شده را برای تشخیص الگوها وارد می‌کند، در پلتفرم لاگ‌گیری راستی‌آزمایی می‌کند	از هوش مصنوعی برای همبسته‌سازی ناهنجاری‌های معیار در چند سرویس استفاده می‌کند، در برابر داشبوردها اعتبارسنجی می‌کند	یک دستیار تحلیل می‌سازد که به تله‌متری فقط‌خواندنی با کنترل دسترسی متصل است
پیشنهادهای RCA	علل ریشه‌ای کاندید را می‌خواهد، هر یک را دستی تأیید می‌کند	پرامپت‌نویسی ساختاریافته RCA را اجرا می‌کند و فرضیه‌ها را بر اساس شواهد رتبه‌بندی می‌کند	agent یک RCA رتبه‌بندی‌شده با شواهد پیوندخورده پیشنهاد می‌دهد؛ انسان‌ها پیش از هر اقدام تأیید می‌کنند
توصیه runbook	می‌پرسد «کدام runbook با این علامت تطبیق دارد»، آن را باز و می‌خواند	پرامپت + فهرست runbook را تدوین می‌کند تا تطبیق‌ها سازگار باشند	سیستم بازیابی، runbook تأییدشده را نمایان می‌کند، هرگز آن را به‌طور خودمختار اجرا نمی‌کند
پیش‌نویس postmortem	یک postmortem بدون سرزنش (blameless) را از یادداشت‌ها پیش‌نویس می‌کند، برای دقت ویرایش می‌کند	postmortemها را قالب‌بندی می‌کند و کیفیت اقلام اقدام را ردگیری می‌کند	به‌طور خودکار پیش‌نویس را از داده رخداد برای نگارش و تأیید انسانی سرهم می‌کند
نرخ سوختن SLO	از هوش مصنوعی می‌خواهد یک رقم نرخ سوختن را تفسیر کند	یک روایت تکرارپذیر نرخ سوختن برای ذی‌نفعان می‌سازد	هشدار خودکار نرخ سوختن با بافت تولیدشده توسط هوش مصنوعی، آستانه‌های تنظیم‌شده توسط انسان
پیش‌بینی ظرفیت	یک پیش‌بینی را در برابر تاریخچه بررسی منطقی می‌کند	از هوش مصنوعی برای مدل‌سازی سناریوهای تقاضا استفاده می‌کند، مفروضات را اعتبارسنجی می‌کند	پیش‌بینی ظرفیت تولیدی‌شده با راستی‌آزمایی ریسک مدل و پایش
اصلاح ایمن	اصلاحات هوش مصنوعی را فقط به‌عنوان پیشنهاد تلقی می‌کند؛ هرگز خودکار اعمال نمی‌کند	تعریف می‌کند کدام اصلاحات واجد شرایط هوش مصنوعی هستند و کدام هرگز نیستند	خودکارسازی اصلاح را با دروازه‌های تأیید، rollback و لاگ‌گیری حسابرسی می‌سازد
IR انسان‌درحلقه	همیشه یک تصمیم‌گیرنده انسانی را در حلقه نگه می‌دارد	نقاط بازرسی HITL را در فرآیند IR تیم تثبیت می‌کند	یک پلتفرم IR طراحی می‌کند که در آن هوش مصنوعی کمک می‌کند اما انسان‌ها هر تغییر را مجاز می‌کنند

شواهدی که HR می‌تواند راستی‌آزمایی کند

نوع شواهد	به دنبال چه باشید
سوابق رخداد	postmortemها و تیکت‌های رخداد که در آن‌ها کمک هوش مصنوعی افشا شده، با درج راستی‌آزمایی انسانی
قالب‌های پرامپت / runbookها	قالب‌های پرامپت و فهرست‌های runbook مشترک و تحت‌کنترل نسخه که تیم واقعاً از آن‌ها استفاده می‌کند
مسیر مدیریت تغییر	سوابق تأیید و rollback که نشان می‌دهد اصلاحات پیشنهادی هوش مصنوعی از کنترل تغییر عبور کرده‌اند، نه ویرایش مستقیم تولید

نوع شواهد	به دنبال چه باشید
اثبات مدیریت داده	شواهد پاکسازی لاگ/معیار یا استفاده فقط از ابزارهای تأییدشده و دارای کنترل دسترسی (بدون PII/اسرار خام در پرامپت‌ها)
دستاوردهای قابل‌اندازه‌گیری	معیارهای قبل/بعد: زمان ترياژ، میانگین زمان تا خلاصه، زمان چرخش postmortem، دقت پیش‌بینی
توانمندسازی همکاران (+L6)	اسناد آنبوردینگ، brown-bagها یا هنجارهایی که فرد برای تیم on-call نوشته است
محصولات تولیدی (+L7)	خودکارسازی‌های مستقرشده با لاگ‌های حسابرسی، حفاظ‌ها و تأیید ریسک مدل/راستی‌آزمایی

نمونه‌های آزمون عملی

آزمون 1 — ترياژ ایمن رخداد (هدف L3-L5). به مهندس یک طوفان هشدار واقع‌گرایانه اما مصنوعی به‌علاوه نمونه لاگ‌هایی بدهید که عمداً حاوی یک شماره حساب جعلی و یک نام میزبان داخلی هستند. از او بخواهید از یک ابزار هوش مصنوعی تأییدشده برای خلاصه‌سازی رخداد و پیشنهاد علل ریشه‌ای محتمل استفاده کند. - نشانه‌های قبولی: فیلدهای حساس را پنهان می‌کند یا از واردکردن آن‌ها امتناع می‌کند، فقط از یک ابزار تأییدشده استفاده می‌کند، خروجی RCA را به‌عنوان فرضیه تلقی می‌کند، در برابر داشبورد ارائه‌شده راستی‌آزمایی می‌کند و هرگز عمل روی تولید بدون تأیید را پیشنهاد نمی‌دهد. - نشانه‌های مردودی: داده حساس خام را وارد می‌کند، RCA هوش مصنوعی را به‌عنوان واقعیت می‌پذیرد، یا یک تغییر تولیدی فوری پیشنهاد می‌دهد.

آزمون 2 — طراحی اصلاح ایمن (هدف L6-L8). یک رخداد تکرارشونده ارائه دهید و از مهندس بخواهید یک اصلاح مجهز به هوش مصنوعی طراحی کند. ارزیابی کنید که آیا طراحی شامل این موارد است: یک دروازه تأیید، یک rollback تعریف‌شده، یک محدودیت صریح شعاع انفجار، لاگ‌گیری حسابرسی، فهرستی از اصلاحاتی که هرگز واجد شرایط هوش مصنوعی نیستند، و اینکه چگونه از مدیریت تغییر و بازبینی ریسک مدل عبور می‌کند.

کنترل‌های ایمنی و داده تحت‌نظارت تعبیه‌شده

- **بدون داده حساس در پرامپت‌ها:** لاگ‌ها/معیارها باید پاکسازی شوند یا از طریق ابزار تأییدشده و دارای کنترل دسترسی دسترسی یابند؛ هرگز شماره حساب، اسرار، PII یا شناسه‌های مشتری را وارد نکنید.
- **انسان هر تغییر تولیدی را مجاز می‌کند:** هوش مصنوعی می‌تواند پیشنهاد دهد، خلاصه کند و تحلیل کند؛ یک انسان از طریق مدیریت تغییر تأیید و اجرا می‌کند. بدون اصلاح خودمختار بدون تأیید صریح و دروازه‌دار.
- **تأیید + rollback + حسابرسی به‌صورت پیش‌فرض:** هر خودکارسازی متأثر از هوش مصنوعی نیازمند دروازه‌های تأیید، یک مسیر rollback آزمایش‌شده و لاگ‌گیری حسابرسی تغییرناپذیر است.
- **راستی‌آزمایی ریسک مدل (+L8):** راه‌حل‌های هوش مصنوعی که بر تصمیمات دسترس‌پذیری یا ظرفیت اثر می‌گذارند، باید پیش از استفاده تولیدی از راستی‌آزمایی ریسک مدل و مدیریت تغییر سازمان عبور کنند.
- **فقط ابزارهای تأییدشده:** از سرویس‌های GenAI مجاز با تنظیمات مناسب اقامت داده (data-residency) و نگه‌داری استفاده کنید؛ ابزارهای سایه (shadow) برای داده عملیاتی مجاز نیستند.

سطوح هدف توصیه‌شده

جمعیت	پس از 4 ماه	پس از 12 ماه
همه مهندسان SRE/Ops	**L3** — استفاده روزانه ایمن و راستی‌آزمایی‌شده روی وظایف کم‌ریسک با مدیریت داده درست	**L5** — هوش مصنوعی جای‌گرفته در گردش‌کارهای ترياژ، postmortem و تحلیل با دستاوردهای سنجیده‌شده
SREهای ارشد / رهبران on-call	**L4-L5**	**L6** — متخصص هوش مصنوعی تیم که هنجارهای ایمن تعیین می‌کند و دیگران را توانمند می‌سازد

جمعیت	پس از 4 ماه	پس از 12 ماه
SREهای ابزار/پلتفرم	**L5**	**L7-L8** — خودکارسازی‌های تولیدی با دروازه‌های تأیید/rollback که از راستی‌آزمایی ریسک مدل عبور می‌کنند

اهداف فرض می‌کنند دسترسی به ابزار تأییدشده و زمانِ اختصاص‌یافته برای توانمندسازی وجود دارد. پیشرفت فراتر از L5 باید مشروط به انضباط ایمنی اثبات‌شده باشد، نه صرفاً دستاوردهای توان عملیاتی (throughput).

x Platform / Kubernetes.12

این راهنما به افراد غیرمتخصص کمک می‌کند ارزیابی کنند یک مهندس Platform / Kubernetes چقدر خوب از GenAI برای انجام ایمن و مؤثر کارش استفاده می‌کند. کار Kubernetes با پایداری تولید، مرزهای امنیتی و هزینه سروکار دارد — بنابراین استفاده از هوش مصنوعی در اینجا نه‌فقط بر اساس سرعت، بلکه بر اساس اینکه آیا مهندس انسانی را برای هر چیزی که یک خوشه زنده را تغییر می‌دهد در حلقه نگه می‌دارد، سنجیده می‌شود. در یک محیط تحت‌نظارت، با هر خوشه به‌عنوان مجاور-تولید (production-adjacent) رفتار کنید.

توانمندی HIGH / MEDIUM / LOW چه شکلی دارد

نوار	چه شکلی در عمل دارد	سطح نردبان معمول
LOW	از هوش مصنوعی برای پیش‌نویس یک manifest YAML یا توضیح یک خطای `kubectl` استفاده می‌کند، اما خروجی را بدون بررسی namespaceها، تگ‌های image یا محدودیت‌های منابع کپی می‌کند. هیچ آگاهی از اینکه کدام داده خوشه برای واردکردن در یک ابزار ایمن است ندارد. نمی‌تواند تشخیص دهد چه زمانی توصیه هوش مصنوعی نادرست است.	L1–L3
MEDIUM	با بافت خوشه پنهان‌شده، به‌طور قابل‌اتکا برای manifestها، مقادیر Helm و تشخیص پرامپت می‌دهد. پیشنهادها را پیش از اعمال در برابر وضعیت واقعی خوشه و سیاست راستی‌آزمایی می‌کند. از هوش مصنوعی در سراسر یک گردش‌کار استفاده می‌کند (مثلاً تشخیص ← پیشنهاد اصلاح ← نوشتن PR) و می‌تواند زمان صرفه‌جویی‌شده را تقریباً بیان کند.	L4–L5
HIGH	هنجارهای ایمن هوش مصنوعی را برای تیم پلتفرم تعیین می‌کند، خودکارسازی‌های تأییدشده (دستیارهای سلف‌سرویس، بررسی‌های سیاست، گزارش‌های ظرفیت) را با دروازه‌های تأیید و rollback می‌سازد، و ابزار قابل‌استفاده مجدد و محافظت‌شده‌ای طراحی می‌کند که از بازبینی امنیتی و ریسک مدل عبور می‌کند.	L6–L9

نمونه‌های توانمندی نگاشت‌شده به سطوح نردبان

وظیفه پلتفرم	LOW (L1-L3)	MEDIUM (L4-L5)	HIGH (L6-L9)
تولید manifest	یک Deployment تولیدشده را همان‌طور که هست وارد می‌کند	manifest تولید می‌کند، سپس محدودیت‌ها، probeها، securityContext را اصلاح و با linter / `kubectl --dry-run` اعتبارسنجی می‌کند	قالب‌های manifest مجهز به هوش مصنوعی را با پیش‌فرض‌های سازمانی تعبیه‌شده نگه‌داری می‌کند
کمک Helm	از هوش مصنوعی می‌خواهد «یک Helm chart بنویسد»، آن را آزمایش‌نشده عرضه می‌کند	از هوش مصنوعی برای بازآرایی `values.yaml` و templateها استفاده می‌کند، با `helm template` + diff راستی‌آزمایی می‌کند	یک کتابخانه chart مشترک و پرامپت‌هایی می‌سازد که تیم از آن‌ها استفاده مجدد می‌کند

وظیفه پلتفرم	LOW (L1-L3)	MEDIUM (L4-L5)	HIGH (L6-L9)
تشخیص Pod/ node	لاگ‌های خام (که ممکن است حاوی اسرار باشند) را برای دریافت توضیح وارد می‌کند	لاگ‌ها را پنهان می‌کند، از هوش مصنوعی برای محدودکردن علت ریشه‌ای استفاده می‌کند، در برابر معیارها/رویدادها تأیید می‌کند	تشخیص‌های متداول را در یک runbook / دستیار هوش مصنوعی تثبیت می‌کند
توصیه محدودیت منابع و autoscaling	حدس requests/limits هوش مصنوعی را می‌پذیرد	داده مصرف واقعی را تغذیه می‌کند، پیشنهادهای HPA/VPA را در برابر ظرفیت آزاد بررسی منطقی می‌کند	پیشنهادهای right-sizing را به‌عنوان PRهای بازبینی‌شده خودکار می‌کند، هرگز خودکار اعمال نمی‌شود
ظرفیت خوشه	به‌طور کلی از هوش مصنوعی می‌پرسد «آیا خوشه من به‌اندازه کافی بزرگ است»	داده استفاده واقعی (پاکسازی‌شده) را برای یک روایت ظرفیت به هوش مصنوعی می‌دهد	گزارش‌دهی ظرفیت هوش مصنوعی تکرارپذیر را با تأیید انسانی می‌سازد
عیب‌یابی service-mesh	پیکربندی mesh را کورکورانه کپی می‌کند	از هوش مصنوعی برای تفسیر خطاهای mTLS/مسیریابی استفاده می‌کند، ابتدا در محیط غیرتولیدی اعتبارسنجی می‌کند	الگوهای عیب‌یابی mesh را در ابزار تیم رمزگذاری می‌کند
دستیار سلف‌سرویس	کاربرد ندارد	یک دستیار Q&A داخلی برای تیم‌های اپلیکیشن آزمایشی راه‌اندازی می‌کند	یک دستیار پلتفرم سلف‌سرویس محافظت‌شده می‌سازد و حاکمیت می‌کند
بهینه‌سازی هزینه	توصیه عمومی «از spot instances استفاده کن» را تکرار می‌کند	از هوش مصنوعی روی داده هزینه واقعی برای یافتن صرفه‌جویی‌های مشخص استفاده می‌کند، اثر را اعتبارسنجی می‌کند	مالک گزارش‌دهی هزینه مبتنی بر هوش مصنوعی با صرفه‌جویی‌های سنجیده‌شده و منتسب است
Policy-as-code	از OPA/Kyverno بی‌اطلاع است یا سیاست بررسی‌نشده می‌نویسد	از هوش مصنوعی برای پیش‌نویس/پالایش سیاست‌ها استفاده می‌کند، در برابر manifestهای واقعی آزمایش می‌کند	خط لوله policy-as-code مجهز به هوش مصنوعی را در سطح سازمان طراحی می‌کند

شواهدی که HR می‌تواند راستی‌آزمایی کند

برای تأیید این موارد نیازی به دانش Kubernetes ندارید. محصولات را بخواهید و به دنبال نشانگرهای زیر باشید.

شواهد	یک مورد خوب چه شکلی است
PRهای ادغام‌شده	manifestها/تغییرات Helm مجهز به هوش مصنوعی که از CI، linting و بازبینی همتا عبور کرده‌اند — نه ویرایش مستقیم روی خوشه
تاریخچه پرامپت یا چت	شناسه‌های خوشه، اسرار و داده مشتری پیش از فرستادن به ابزارهای هوش مصنوعی پنهان یا حذف شده‌اند
سوابق تغییر	هر تغییر تولیدی پیشنهادی هوش مصنوعی یک تیکت، یک تأییدکننده و یک گام rollback مستند دارد
معیارهای قبل/بعد	اعداد بیان‌شده و قابل‌بررسی: MTTR رخداد، زمان تا manifest، درصد کاهش هزینه، دقت ظرفیت
محصولات تیمی (فقط HIGH)	قالب‌ها، runbookها، سیاست‌ها یا یک دستیار مشترک که دیگران در تیم واقعاً از آن استفاده می‌کنند
تأیید ابزار	فقط ابزارهای هوش مصنوعی تأییدشده شرکت استفاده شده؛ بدون واردکردن در چت‌بات‌های عمومی غیرمجاز

نمونه‌های آزمون عملی

آزمون 1 — تشخیص و اصلاح (نشانه MEDIUM). به مهندس یک Deployment خراب بدهید (مثلاً یک pod در `CrashLoopBackOff` با یک probe بد و یک محدودیت منابع ناموجود) و یک قطعه لاگ پاکسازی‌شده. از او بخواهید از هوش مصنوعی برای تشخیص و تولید یک manifest اصلاح‌شده استفاده کند. - قبولی: هر چیز حساس را پیش از پرامپت پنهان می‌کند، علت ریشه‌ای واقعی را شناسایی می‌کند، محدودیت‌ها/probeها را اصلاح می‌کند و پیش از پیشنهاد با dry-run یا lint اعتبارسنجی می‌کند — نه‌فقط واردکردن پاسخ هوش مصنوعی.

آزمون 2 — Right-size با یک دروازه (نشانه HIGH). داده استفاده CPU/حافظه پاکسازی‌شده فراهم کنید و یک توصیه autoscaling و محدودیت منابع بخواهید که به‌صورت یک تغییر قابل‌بازبینی ارائه شود. - قبولی: یک پیشنهاد موجه HPA/requests/limits به‌عنوان یک PR با یادداشت‌های rollback تولید می‌کند و صریحاً آن را خودکار روی یک خوشه زنده اعمال نمی‌کند و به دروازه تأیید استناد می‌کند.

کنترل‌های ایمنی و داده تحت‌نظارت (همیشه الزامی)

- **بدون تغییرات خودکار روی خوشه زنده.** هوش مصنوعی می‌تواند پیشنهاد دهد؛ یک انسان تأیید و اعمال می‌کند. تغییرات تولیدی از مدیریت تغییر با یک طرح rollback عبور می‌کنند.
- **حداقل‌سازی داده.** بدون اسرار، توکن‌ها، kubeconfigها، داده مشتری، نام میزبان‌های داخلی یا بازه‌های IP در پرامپت‌ها. ابتدا لاگ‌ها و پیکربندی‌ها را پنهان کنید.
- **فقط ابزارهای تأییدشده.** از ابزار هوش مصنوعی سازمانی مجاز با تضمین‌های مدیریت داده استفاده کنید؛ هرگز داده خوشه را در ابزارهای عمومی غیرمجاز وارد نکنید.
- **قابلیت حسابرسی.** تغییرات تولیدی متأثر از هوش مصنوعی به یک تیکت، یک تأییدکننده و منطق پرامپت/خروجی قابل‌ردگیری هستند.
- **دروازه‌های سیاست و امنیت.** manifestها و سیاست‌های تولیدشده پیش از merge از policy-as-، linting-code (مثلاً OPA/Kyverno) و بازبینی امنیتی عبور می‌کنند.

سطوح هدف توصیه‌شده

افق	هدف	منطق
پس از 4 ماه	**L4–L5** برای اکثر مهندسان پلتفرم	استفاده قابل‌اتکا و راستی‌آزمایی‌شده از هوش مصنوعی جای‌گرفته در گردش‌کارهای روزمره (manifestها، تشخیص، Helm) با کنترل‌های داده عادت‌شده
پس از 12 ماه	**L6–L7**، با 1–2 مهندس در **L8+**	هنجارهای تیم توسط کاربران قدرتمند تعیین می‌شود، به‌علاوه خودکارسازی‌های تأییدشده و دروازه‌دار (دستیار سلف‌سرویس، policy-as-code، گزارش‌دهی هزینه) ساخته‌شده توسط کارکنان ارشد

(برداشت‌شده از منابع عمومی: نوارهای نردبان و سطوح هدف توصیه‌های ترکیبی نگاشت‌شده به چارچوب L1–L10 ارائه‌شده و رویه متداول Kubernetes هستند؛ از هیچ نردبان داخلی شرکتی برگرفته نشده‌اند.)

12. Observability / Monitoring x

این زیربخش مهندسان و SREهایی را پوشش می‌دهد که پشته‌های پایش (Prometheus/PromQL، Loki/LogQL، tracing، Grafana، Elasticsearch توزیع‌شده) را می‌سازند و بهره‌برداری می‌کنند. GenAI به‌طور فزاینده برای پیش‌نویس پرس‌وجوها، تولید داشبوردها، توصیه قواعد هشدار، کاهش نویز هشدار، همبسته‌سازی سیگنال‌ها و تحلیل traceها استفاده می‌شود. در یک بافت خدمات مالی تحت‌نظارت، قید الزام‌آور این نیست که آیا هوش مصنوعی می‌تواند یک پرس‌وجو یا داشبورد تولید کند — بلکه این است که آیا اپراتور انسانی می‌تواند خروجی را در برابر سیگنال‌های زنده اعتبارسنجی کند و از نقاط کور در هشداردهی تولیدی پرهیز کند.

توانمندی LOW / MEDIUM / HIGH چه شکلی دارد

نوار	نردبان	چه شکلی در کار observability دارد	وضعیت ریسک
LOW	L1-L3	از هوش مصنوعی برای پیش‌نویس پرس‌وجوهای منفرد PromQL/LogQL/ES و توضیح معیارهای ناآشنا استفاده می‌کند. خروجی را با اجرای پرس‌وجو و بررسی منطقی نتایج راستی‌آزمایی می‌کند. هنوز برای آستانه‌های هشدار یا داشبوردهای تولیدی به هوش مصنوعی اعتماد نمی‌کند.	راستی‌آزمایی دستی روی هر خروجی؛ شعاع انفجار پایین
MEDIUM	L4-L6	به‌طور قابل‌اتکا برای داشبوردهای سیگنال طلایی، مجموعه‌های قواعد هشدار و منطق کاهش نویز پرامپت می‌دهد. از هوش مصنوعی برای همبسته‌سازی معیارها/لاگ‌ها/traceها در حین رخدادها استفاده می‌کند. هنجارهای تیم برای استفاده ایمن هوش مصنوعی در پایش تعیین می‌کند و پرس‌وجوهای تولیدشده دیگران را بازبینی می‌کند.	هوش مصنوعی جای‌گرفته در گردش‌کار؛ انسان مالک آستانه‌ها و اثر runbook است
HIGH	L7-L10	خودکارسازی‌های تأییدشده می‌سازد: تشخیص ناهنجاری مجهز به هوش مصنوعی، داشبوردهای خودکار تولیدشده از تعاریف سرویس، خطوط لوله تنظیم هشدار با rollback. راه‌حل‌های monitoring-as-code را معماری می‌کند که از بازبینی ریسک مدل عبور می‌کنند و حفاظ‌ها را در سطح سازمان تعبیه می‌کنند.	خودکارسازی تولیدی با دروازه‌های تأیید، rollback و راستی‌آزمایی

نمونه‌های نگاشت‌شده به سطوح نردبان

نردبان	رفتار نماینده observability
L2	می‌داند که نباید نام میزبان‌های تولیدی، شناسه‌های مشتری یا خطوط لاگ خام با PII را در یک ابزار هوش مصنوعی عمومی وارد کند؛ قواعد داده sandbox در برابر زنده را می‌فهمد.
L3	از هوش مصنوعی می‌خواهد یک پرس‌وجوی PromQL برای تأخیر p99 بنویسد، آن را اجرا می‌کند و پیش از استفاده تأیید می‌کند که نتیجه با مقدار شناخته‌شده داشبورد تطبیق دارد.
L4	از پرامپت‌های ساختاریافته (نام معیارها، طرح برچسب، بازه زمانی، قصد) استفاده می‌کند تا فیلترهای صحیح LogQL و Elasticsearch DSL را در تلاش اول یا دوم به‌دست آورد؛ نام معیارهای ساختگی هوش مصنوعی را می‌گیرد.
L5	یک گردش‌کار تکرارپذیر می‌سازد: هوش مصنوعی داشبورد سیگنال طلایی (تأخیر، ترافیک، خطاها، اشباع) را برای هر سرویس پیش‌نویس می‌کند، مهندس پنل‌ها را در برابر ترافیک واقعی اعتبارسنجی می‌کند، کاهش زمان ساخت داشبورد را می‌سنجد.
L6	کتابخانه پرامپت و فهرست بازبینی تیم برای قواعد هشدار تولیدشده توسط هوش مصنوعی را تعیین می‌کند؛ دیگران را در اعتبارسنجی آستانه‌ها و پرهیز از هشدارهای منفی کاذب مربیگری می‌کند.
L7	یک خودکارسازی تأییدشده کاهش نویز هشدار عرضه می‌کند (مثلاً گروه‌بندی/حذف تکرار مجهز به هوش مصنوعی) پشت یک feature flag با rollback و یک دروازه تأیید انسانی.
L8	یک راه‌حل سرتاسری تشخیص ناهنجاری برای یک لایه سرویس طراحی می‌کند که از راستی‌آزمایی ریسک مدل عبور می‌کند، با حالت‌های شکست مستند و بازگشت به آستانه‌های ایستا.
L9	یک agent/پلتفرم قابل‌استفاده مجدد و محافظت‌شده می‌سازد که از فراداده سرویس، داشبوردها و قواعد هشدار را برای چند تیم تولید می‌کند.
L10	مالک استاندارد حاکمیت برای استفاده از هوش مصنوعی در سراسر کارکرد observability است، از جمله تعهدات نظارتی و حسابرسی.

شواهدی که HR می‌تواند راستی‌آزمایی کند

سطح ادعاشده	شواهد قابل‌راستی‌آزمایی (بدون نیاز به دانش تخصصی)
L2-L3	تکمیل آموزش داخلی مدیریت داده/سیاست هوش مصنوعی؛ نمونه‌های پرس‌وجوهای پیش‌نویس‌شده با هوش مصنوعی همراه با یادداشتی درباره نحوه راستی‌آزمایی هر یک.
L4-L5	قالب‌های پرامپت ذخیره‌شده؛ شمارش زمان ساخت داشبورد یا تکرار پرس‌وجو قبل/بعد؛ تأیید همتا که خروجی‌های هوش مصنوعی در برابر داده زنده اعتبارسنجی شده‌اند.
L6	کتابخانه پرامپت تیم یا فهرست بازبینی نوشته‌شده توسط فرد؛ سابقه جلسات مربیگری یا آنبوردینگ دیگران.
L7	pull request ادغام‌شده برای یک خودکارسازی که شامل یک دروازه تأیید و یک مسیر rollback مستند است؛ ارجاع تیکت مدیریت تغییر.
L8-L10	اسناد تأیید ریسک مدل/راستی‌آزمایی؛ سوابق بازبینی معماری؛ سیاست یا استاندارد حاکمیت با نام فرد به‌عنوان مالک.

نکته: HR باید وجود و نویسندگی محصولات (PRها، تیکت‌ها، فهرست‌های بازبینی، تأییدها) را راستی‌آزمایی کند و اثر را با رهبر فنی تأیید کند، نه آنکه درباره صحت فنی پرس‌وجوها قضاوت کند.

نمونه‌های آزمون عملی

آزمون 1 — تولید پرس‌وجو با راستی‌آزمایی (هدف L3-L4). به نامزد یک نمونه طرح معیار/برچسب بدهید و بپرسید: «یک پرس‌وجوی PromQL برای تأخیر درخواست در صدک 95اُم سرویس X طی 15 دقیقه اخیر، و یک پرس‌وجوی LogQL برای یافتن پاسخ‌های 5xx همان سرویس بنویس. نشان بده چگونه صحت هر یک را تأیید کردی.» معیار قبولی: توابع و برچسب‌های صحیح، بدون نام معیارهای ساختگی، و یک گام راستی‌آزمایی صریح (آن را اجرا کرد، با یک مقدار شناخته‌شده مقایسه کرد). نامزدهای قوی هر برچسب حدسی را علامت می‌زنند و یادآور می‌شوند که آستانه‌ها باید در برابر داده واقعی تأیید شوند.

آزمون 2 — کاهش نویز هشدار با استدلال ایمنی (هدف L5-L7). یک مجموعه هشدار پرنویز ارائه دهید (مثلاً یک هشدار CPU پرنوسان و هشدارهای پایین‌دستی تکراری) و از نامزد بخواهید از هوش مصنوعی برای پیشنهاد یک رویکرد کاهش نویز استفاده کند. معیار قبولی: گروه‌بندی/حذف تکرار مجهز به هوش مصنوعی یا تنظیم آستانه پیشنهاد می‌شود، اما نامزد صریحاً ایمنی منفی کاذب را حفظ می‌کند (هشدارهایی را که می‌توانند قطعی‌ها را پنهان کنند سرکوب نمی‌کند)، یک rollback تعریف می‌کند و تغییر را پیش از تولید از یک دروازه تأیید عبور می‌دهد. نشانه درستی، امتناع از خاموش‌کردن هشدارها بدون اعتبارسنجی است.

کنترل‌های ایمنی و داده تحت‌نظارت تعبیه‌شده

کنترل	الزام
مدیریت داده	بدون PII، شناسه‌های حساب/مشتری، اسرار یا نام میزبان‌های تولیدی واردشده در ابزارهای هوش مصنوعی غیرتأییدشده؛ فقط از ابزار سازمانی تأییدشده و نمونه‌های پاکسازی‌شده استفاده کنید.
پاسخگویی انسانی	یک مهندس نام‌برده مالک هر آستانه هشدار، داشبورد و پرس‌وجوی تولیدشده توسط هوش مصنوعی است که به تولید می‌رسد. خروجی هوش مصنوعی یک پیش‌نویس است، هرگز مرجع نهایی نیست.
حفاظت در برابر منفی کاذب	کاهش نویز هشدار هرگز نباید سیگنال‌هایی را که می‌توانند رخدادهای اثرگذار بر مشتری را پنهان کنند سرکوب کند؛ تغییرات نیازمند اعتبارسنجی در برابر رخدادهای تاریخی هستند.
دروازه‌های تغییر	خودکارسازی‌های تولیدی نیازمند تأیید، کنترل نسخه، مسیر حسابرسی و یک مسیر rollback آزمایش‌شده هستند.

کنترل	الزام
همسویی با ریسک مدل	تشخیص ناهنجاری و سایر منطق تشخیص مبتنی بر هوش مصنوعی باید از راستی‌آزمایی ریسک مدل با حالت‌های شکست مستند و بازگشت به آستانه ایستا عبور کنند.

سطوح هدف توصیه‌شده

افق	هدف	منطق
پس از 4 ماه	**L4 (کاربر مؤثر پرامپت)**، با کارکنان ارشد در **L5**	اکثر مهندسان observability باید به‌طور قابل‌اتکا پرس‌وجوها و داشبوردهای سیگنال طلایی را تولید و راستی‌آزمایی کنند؛ رهبران هوش مصنوعی را در گردش‌کارهای رخداد و داشبورد جای می‌دهند.
پس از 12 ماه	**L5-L6**، با حداقل یک **L7** در هر تیم	کارکرد، هوش مصنوعی را در گردش‌کارهای چندمرحله‌ای با دستاوردهای سنجیده‌شده بهره‌برداری می‌کند؛ یک کاربر قدرتمند هنجارها را تعیین می‌کند؛ حداقل یک مهندس یک خودکارسازی تأییدشده و محافظت‌شده با rollback عرضه می‌کند.

سطوح هدف بالا توصیه‌های ترکیبی برای این کارکرد هستند، نه یک استاندارد داخلی موجود. تعاریف L1-L10 مرجع نردبان ارائه‌شده هستند؛ نگاشت‌ها به وظایف observability (برداشت‌شده از منابع عمومی) درباره استفاده از Prometheus/Grafana/Loki/Elasticsearch و رویه استاندارد SRE هستند.

x Service Desk / IT Support.12

Service Desk / IT Support حجم بالایی از تیکت‌ها، تعاملات مشتری/کارمند و درخواست‌های دسترسی را مدیریت می‌کند. در یک محیط خدمات مالی تحت‌نظارت، استفاده از هوش مصنوعی در اینجا قدرتمند اما حساس است: نمایندگان به‌طور روتین با PII مشتری، داده حساب و سوابق مرتبط با حسابرسی سروکار دارند. ارزیابی باید دستاوردهای سرعت و سازگاری را پاداش دهد و در عین حال ثابت کند که قواعد مدیریت داده، بازبینی انسانی و قابلیت ردگیری هرگز دور زده نمی‌شوند.

کنترل‌های ایمنی و داده تحت‌نظارت (در هر سطح اعمال می‌شوند): - هرگز PII مشتری، شماره حساب، داده کارت، اعتبارنامه‌ها یا متن کامل تیکت‌ها را در ابزارهای GenAI غیرتأییدشده/عمومی وارد نکنید. فقط از ابزارهای مجاز و دارای حاکمیت داده استفاده کنید. - پاسخ‌های مشتری پیش‌نویس‌شده توسط هوش مصنوعی پیش از ارسال توسط یک انسان بازبینی می‌شوند؛ انسان مالک پیام نهایی است. - تصمیمات طبقه‌بندی/مسیریابی تیکت توسط هوش مصنوعی قابل‌حسابرسی و توسط یک فرد قابل‌بازنویسی هستند. - اقدامات چت‌بات و خودکارسازی که بر حساب‌ها اثر می‌گذارند نیازمند محدودیت‌های دامنه تعریف‌شده، لاگ‌گیری و مسیرهای rollback/تشدید (escalation) هستند. - هر خودکارسازی که با داده مشتری سروکار دارد، پیش از تولید از تأیید کنترل تغییر عبور می‌کند.

توانمندی LOW / MEDIUM / HIGH چه شکلی دارد

نوار توانمندی	بازه نردبان	چه شکلی در Service Desk دارد
LOW	L1-L3	قواعد داده را می‌داند؛ از هوش مصنوعی تأییدشده برای خلاصه‌سازی تیکت‌های پاکسازی‌شده، پیش‌نویس پاسخ‌های روتین و جستجوی KB استفاده می‌کند و همیشه پیش از اقدام یا ارسال راستی‌آزمایی می‌کند.
MEDIUM	L4-L6	پرامپت‌نویسی قابل‌اتکا برای طبقه‌بندی/مسیریابی و توصیه‌های KB؛ هوش مصنوعی را در سراسر چرخه عمر تیکت جای می‌دهد، دستاوردهای deflection/زمان رسیدگی را می‌سنجد و هنجارهای ایمن برای تیم تعیین می‌کند.

نوار توانمندی	بازه نردبان	چه شکلی در Service Desk دارد
HIGH	L7-L10	خودکارسازی‌های تأییدشده (مسیریابی خودکار، حل مسائل تکراری، پیش‌بینی نقض SLA) و چت‌بات‌های L1 با حفاظ‌ها، منطق تحویل (handoff) و rollback می‌سازد که از راستی‌آزمایی ریسک مدل عبور می‌کنند.

نمونه‌های نگاشت‌شده به سطوح نردبان

سطح	نمونه Service Desk
L1	هرگز از GenAI برای کار پشتیبانی استفاده نکرده؛ بی‌اطلاع از اینکه کدام ابزارها تأییدشده‌اند یا چه داده‌ای نباید به اشتراک گذاشته شود.
L2	می‌تواند پیش از کار با هر ابزار هوش مصنوعی، قواعد مدیریت داده (بدون PII در ابزارهای غیرتأییدشده) را بیان کند.
L3	از هوش مصنوعی تأییدشده برای خلاصه‌سازی یک تیکت پاکسازی‌شده یا پیش‌نویس یک پاسخ روتین بازنشانی رمز عبور استفاده می‌کند، سپس پیش از ارسال راستی‌آزمایی و ویرایش می‌کند.
L4	از پرامپت‌های هدفمند برای طبقه‌بندی تیکت‌ها بر اساس دسته/اولویت و توصیه مقاله KB درست به‌طور قابل‌اتکا و با کیفیت سازگار استفاده می‌کند.
L5	هوش مصنوعی را در سراسر جریان کامل تیکت جای می‌دهد (خلاصه‌سازی تا طبقه‌بندی تا مسیریابی تا پیش‌نویس پاسخ) و بهبود حل در تماس اول و زمان رسیدگی را ردگیری می‌کند.
L6	به‌عنوان متخصص هوش مصنوعی تیم عمل می‌کند: کتابخانه پرامپت و راهنمای عملی استفاده ایمن را می‌نویسد، همکاران را در پیش‌نویس ایمن از نظر PII آموزش می‌دهد و کیفیت خروجی هوش مصنوعی را حسابرسی می‌کند.
L7	یک خودکارسازی تأییدشده مسیریابی خودکار یا مسائل تکراری می‌سازد (مثلاً حل خودکار تیکت‌های خطای شناخته‌شده) با دروازه‌های لاگ‌گیری، تأیید و rollback.
L8	یک راه‌حل سرتاسری پیش‌بینی نقض SLA یا ترياژ خودکار طراحی می‌کند که از راستی‌آزمایی ریسک مدل عبور می‌کند و پایش خود را تعریف می‌کند.
L9	یک پلتفرم چت‌بات/agent L1 قابل‌استفاده مجدد با حفاظ‌های سراسری سازمان، کنترل‌های داده مشتری و منطق استاندارد تحویل به انسان می‌سازد.
L10	مالک استراتژی هوش مصنوعی پشتیبانی، اهداف deflection، حاکمیت و وضعیت ریسک نظارتی در سراسر سازمان خدمات است.

شواهدی که HR می‌تواند راستی‌آزمایی کند

نوع شواهد	چه چیزی بخواهید
ابزار	تأیید اینکه فرد فقط از ابزارهای هوش مصنوعی مجاز/تأییدشده استفاده می‌کند (بدون PII در ابزارهای عمومی).
محصولات کاری	پرامپت‌های ذخیره‌شده، کتابخانه پرامپت یا راهنمای عملی استفاده ایمن نوشته‌شده یا استفاده‌شده توسط فرد.
معیارها	ارقام قبل/بعد زمان رسیدگی، حل در تماس اول، نرخ deflection یا دقت مسیریابی منتسب به استفاده از هوش مصنوعی.
بازبینی‌ها/تأییدها	تأییدهای کنترل تغییر، تأیید راستی‌آزمایی ریسک مدل و طرح‌های rollback برای هر خودکارسازی یا چت‌باتی که ساخته‌اند.

نوع شواهد	چه چیزی بخواهید
توانمندسازی همکاران	جلسات آموزشی برگزارشده، همکاران آنبوردشده یا هنجارهای مستندشده (برای L6+).
مسیر حسابرسی	لاگ‌هایی که نشان می‌دهند تصمیمات مسیریابی/خودکارسازی هوش مصنوعی ثبت و توسط انسان قابل‌بازنویسی هستند.

نمونه‌های آزمون عملی

آزمون 1 — خلاصه‌سازی، طبقه‌بندی، مسیریابی ایمن (هدف L3-L4): یک تیکت واقع‌گرایانه حاوی PII (نام، شماره حساب) بدهید. از نامزد بخواهید با استفاده از یک ابزار تأییدشده خلاصه کند، اولویت/دسته را طبقه‌بندی کند و یک مسیر توصیه کند. قبولی: پیش از هر استفاده از ابزار، PII را پاکسازی یا پنهان می‌کند (یا از ابزار دارای حاکمیت به‌درستی استفاده می‌کند)، یک خلاصه دقیق و طبقه‌بندی صحیح تولید می‌کند و بیان می‌کند که تصمیم مسیریابی قابل‌بازبینی است.

آزمون 2 — پیش‌نویس یک پاسخ مشتری با قضاوت تحویل (هدف L4-L6): یک شکایت مشتری ناراضی درباره یک تراکنش ناموفق ارائه دهید. یک پیش‌نویس پاسخ مجهز به هوش مصنوعی بخواهید. قبولی: یک پیش‌نویس دقیق، با لحن مناسب و سازگار با مقررات تولید می‌کند؛ آن را برای بازبینی انسانی پیش از ارسال علامت می‌زند؛ به‌درستی تشخیص می‌دهد چه زمانی موضوع باید به یک انسان تشدید شود (مثلاً اختلاف، تقلب مشکوک) به‌جای رسیدگی توسط هوش مصنوعی/چت‌بات.

سطوح هدف توصیه‌شده

افق	سطح هدف	منطق
پس از 4 ماه	L3-L4	اکثر نمایندگان روزانه به‌طور ایمن از هوش مصنوعی تأییدشده برای خلاصه‌سازی، طبقه‌بندی و پیش‌نویس استفاده می‌کنند، با پرامپت‌نویسی قابل‌اتکا و خروجی راستی‌آزمایی‌شده.
پس از 12 ماه	L5 (تیم) / L6- L7 (رهبران)	تیم‌ها هوش مصنوعی را در سراسر چرخه عمر تیکت جای می‌دهند و دستاوردها را می‌سنجند؛ رهبران هنجارها را تعیین می‌کنند و شروع به ساخت خودکارسازی‌های تأییدشده تحت حاکمیت ریسک مدل و کنترل تغییر می‌کنند.

(برداشت‌شده از منابع عمومی) معیارهای مشخص deflection و زمان رسیدگی باید در برابر خطوط پایه خود سازمان تعیین شوند؛ ارقام منتشرشده صنعتی به‌شدت متفاوت‌اند و نباید به‌عنوان اهداف متعهدشده تلقی شوند.

x Data / DataOps.12

این راهنما نحوه ارزیابی توانمندی هوش مصنوعی برای مهندسان و تحلیلگرانی را پوشش می‌دهد که روی خطوط لوله داده، انبارها، کیفیت و حاکمیت کار می‌کنند. در یک بافت خدمات مالی تحت‌نظارت، کار داده با سوابق حساس مشتری (PII، حساب، داده تراکنش) سروکار دارد، بنابراین هر استفاده از هوش مصنوعی باید قواعد اقامت داده، ماسک‌گذاری و حداقل امتیاز را رعایت کند. آستانه «آیا هوش مصنوعی SQL تولید کرد» نیست — بلکه «آیا فرد نتایج داده صحیح، تحت‌حاکمیت و قابل‌حسابرسی را سریع‌تر تولید کرد در حالی که ریسک داده را کنترل می‌کرد» است.

توانمندی LOW / MEDIUM / HIGH چه شکلی دارد

نوار	نردبان	چه شکلی در Data / DataOps دارد	وضعیت ریسک داده
LOW	L1-L2	از هوش مصنوعی استفاده نمی‌کند، یا داده تولیدی/PII خام را در ابزارهای عمومی وارد می‌کند تا «از آن بخواهد پرس‌وجوی من را اصلاح کند». نمی‌تواند توضیح دهد چرا این ناامن است. خروجی در برابر طرح واقعی راستی‌آزمایی نشده.	اگر بدون نظارت باشد خطرناک است — ممکن است داده حساس را نشت دهد.

نوار	نردبان	چه شکلی در Data / DataOps دارد	وضعیت ریسک داده
MEDIUM	L3-L5	از هوش مصنوعی تأییدشده برای پیش‌نویس SQL، توضیح یک خرابی خط لوله یا پیشنهاد قواعد کیفیت داده استفاده می‌کند؛ همیشه در برابر کاتالوگ راستی‌آزمایی می‌کند و ابتدا روی محیط غیرتولیدی اجرا می‌کند. هوش مصنوعی را در گردش‌کارهای تکرارشونده جای می‌دهد (مثلاً تریاژ DAGهای خطادار) و زمان صرفه‌جویی‌شده را ردگیری می‌کند. هرگز ردیف‌های واقعی مشتری را وارد نمی‌کند.	برای وظایف روتین و کم‌تامتوسط‌ریسک با راستی‌آزمایی ایمن است.
HIGH	L6- L10	هنجارهای تیم برای استفاده ایمن از هوش مصنوعی در داده تعیین می‌کند؛ خودکارسازی‌های تأییدشده (هشدارهای ناهنجاری/تازگی، آشکارسازهای schema-drift) را با rollback و تأیید انسانی می‌سازد؛ راه‌حل‌های داده هوش مصنوعی راستی‌آزمایی‌شده طراحی می‌کند؛ مالک حاکمیت هوش مصنوعی-روی-داده در سراسر سازمان است.	مورد اعتماد برای طراحی کنترل‌هایی که دیگران به آن‌ها متکی‌اند.

نمونه‌های توانمندی نگاشت‌شده به سطوح نردبان

توانمندی	L3-L4 (پایه / مؤثر)	L5-L6 (گردش‌کار / قدرتمند)	L7-L8 (خودکارسازی / راه‌حل)	L9-L10 (پلتفرم / حاکمیت)
تولید SQL	یک پرس‌وجو را از یک پرامپت پیش‌نویس می‌کند، شمارش ردیف و joinها را راستی‌آزمایی می‌کند، فقط از نمونه‌های ماسک‌شده/مصنوعی استفاده می‌کند	پرامپت‌ها را با بافت طرح قالب‌بندی می‌کند؛ SQL هوش مصنوعی را پیش از اجرا برای هزینه و افشای PII بازبینی می‌کند	الگوهای پرس‌وجوی مجهز به هوش مصنوعی بازبینی‌شده را با حفاظ‌های هزینه به خطوط لوله عرضه می‌کند	استاندارد سازمانی برای بازبینی SQL تولیدشده توسط هوش مصنوعی و فهرست ابزارهای تأییدشده را تعیین می‌کند
تشخیص خرابی خط لوله	از هوش مصنوعی می‌خواهد یک stack trace/لاگ را تفسیر کند، علت ریشه‌ای را دستی تأیید می‌کند	یک جریان تریاژ تکرارپذیر می‌سازد که اجراهای خطادار را از لاگ‌ها خلاصه می‌کند	تشخیص اولیه را با اصلاح تأییدشده توسط انسان خودکار می‌کند	playbook هوش مصنوعی پاسخ به رخداد و مسیر حسابرسی را تعریف می‌کند
تولید قاعده کیفیت داده	بررسی‌های null/بازه/یکتایی را پیشنهاد می‌دهد، در برابر توزیع واقعی اعتبارسنجی می‌کند	مجموعه قواعد را در جداول تولید می‌کند و نرخ گرفتن نقص را می‌سنجد	قواعد پیشنهادی هوش مصنوعی را پشت تأیید به چارچوب DQ مستقر می‌کند	چرخه عمر قاعده DQ و آستانه‌های مثبت کاذب را در سطح سازمان حاکمیت می‌کند
پایش تازگی	یک بررسی تازگی برای یک جدول پیش‌نویس می‌کند	SLAهای تازگی و هشداردهی را با کمک هوش مصنوعی استاندارد می‌کند	هشدار تازگی را با پیکربندی ایمن از نظر rollback خودکار می‌کند	مالک سیاست SLA تازگی و حاکمیت استثناهاست
تشخیص schema-drift	یک تغییر طرح تشخیص‌داده‌شده و اثر آن را توضیح می‌دهد	drift-watch را در جداول حیاتی می‌سازد، به مالکان مسیریابی می‌کند	تشخیص drift را با تأیید مهاجرت دروازه‌دار خودکار می‌کند	سیاست سازمانی تغییر طرح/قرارداد را تعیین می‌کند
توضیح lineage	از هوش مصنوعی می‌خواهد بالادست/پایین‌دست یک فیلد را توضیح دهد، در کاتالوگ راستی‌آزمایی می‌کند	به‌طور روتین خلاصه‌های lineage را برای تحلیل اثر تولید می‌کند	خودکارسازی آگاه از lineage را به مدیریت تغییر متصل می‌کند	مالک حاکمیت lineage و قابلیت ردگیری گزارش‌دهی نظارتی است

توانمندی	L3-L4 (پایه / مؤثر)	L5-L6 (گردش‌کار / قدرتمند)	L7-L8 (خودکارسازی / راه‌حل)	L9-L10 (پلتفرم / حاکمیت)
عیب‌یابی ETL/ELT	از هوش مصنوعی برای پیشنهاد اصلاحات یک transform خراب استفاده می‌کند، در محیط غیرتولیدی آزمایش می‌کند	پرامپت‌ها/runbookهای عیب‌یابی قابل‌استفاده مجدد می‌سازد	اصلاح خودکار تأییدشده را با rollback عرضه می‌کند	استانداردهای خودکارسازی ایمن برای transformها را تعریف می‌کند
تشخیص ناهنجاری	ناهنجاری‌های علامت‌گذاری‌شده توسط هوش مصنوعی را بازبینی می‌کند، مثبت/منفی واقعی را تأیید می‌کند	آستانه‌های ناهنجاری را تنظیم می‌کند و دقت (precision) را ردگیری می‌کند	مدل‌های ناهنجاری راستی‌آزمایی‌شده را با انسان‌درحلقه مستقر می‌کند	مالک پلتفرم ناهنجاری راستی‌آزمایی‌شده با ریسک مدل است
دستیار کاتالوگ	از هوش مصنوعی برای یافتن/توصیف مجموعه‌داده‌ها استفاده می‌کند، فراداده را راستی‌آزمایی می‌کند	فراداده کاتالوگ را در مقیاس با کمک هوش مصنوعی بهبود می‌دهد	غنی‌سازی کاتالوگ را با دروازه‌های بازبینی خودکار می‌کند	دقت کاتالوگ-هوش مصنوعی و کنترل‌های دسترسی را حاکمیت می‌کند
مدیریت داده حساس	می‌داند چه چیزی PII است؛ هرگز آن را به هوش مصنوعی نمی‌دهد؛ از داده ماسک‌شده استفاده می‌کند	هنجارهای ماسک‌گذاری/داده مصنوعی را در تیم اعمال می‌کند	ماسک‌گذاری به‌صورت پیش‌فرض را در ابزار داده هوش مصنوعی می‌سازد	مالک سیاست حریم خصوصی داده برای همه هوش مصنوعی-روی-داده است

شواهدی که HR می‌تواند راستی‌آزمایی کند

HR نیازی به قضاوت درباره کیفیت SQL ندارد — راستی‌آزمایی می‌کند که محصولات وجود دارند و کنترل‌ها رعایت شده‌اند. هر یک از موارد زیر را بخواهید (با محتوای حساس پنهان‌شده):

نوع شواهد	به دنبال چه باشید	نگاشت به
Pull requestها / merge requestها	اصلاحات SQL یا خط لوله مجهز به هوش مصنوعی که بازبینی، در محیط غیرتولیدی آزمایش و تأیید شده‌اند	L3-L5
اسناد تریاژ / runbook	یک گردش‌کار تشخیص مجهز به هوش مصنوعی مستند با زمان صرفه‌جویی‌شده سنجیده‌شده	L5-L6
استفاده از ابزار تأییدشده	استفاده فقط از ابزار هوش مصنوعی سازمانی مجاز؛ بدون شواهد واردکردن داده در ابزار عمومی	+L2 (دروازه)
خودکارسازی‌های مستقرشده	کارهای ناهنجاری/تازگی/drift با دروازه‌های تأیید و rollback، به‌علاوه تیکت‌های تغییر	L7-L8
سوابق راستی‌آزمایی	تأیید ریسک مدل یا حاکمیت روی یک راه‌حل داده هوش مصنوعی	+L8
محصولات سیاست / هنجار	استانداردهای تیم، آموزش یا اسناد حاکمیت نوشته‌شده توسط فرد	-L6, L9 L10
معیارها	نرخ گرفتن نقص، MTTR روی رخدادهای خط لوله، پایبندی به SLA تازگی قبل/بعد از هوش مصنوعی	+L5

نمونه‌های آزمون عملی

آزمون 1 — تولید و راستی‌آزمایی SQL ایمن (هدف L3-L5). به نامزد یک طرح ماسک‌شده یا مصنوعی و یک پرسش تجاری بدهید (مثلاً «حساب‌های فعال ماهانه بر اساس بخش، 6 ماه اخیر»). مشاهده کنید: - آیا فقط از ابزار تأییدشده و داده مصنوعی/ماسک‌شده استفاده می‌کنند (هرگز PII واقعی)؟ - آیا بافت طرح را در پرامپت ارائه می‌دهند به‌جای حدس‌زدن نام ستون‌ها؟ - آیا نتیجه را راستی‌آزمایی می‌کنند — شمارش ردیف، fan-out در join، حالت‌های لبه‌ای — پیش از آنکه آن را صحیح اعلام کنند؟ - آیا به هزینه پرس‌وجو / partition pruning توجه می‌کنند؟ در L3 قبول اگر ایمن و راستی‌آزمایی‌شده باشد؛ در L4-L5 اگر از یک رویکرد پرامپت‌نویسی قابل‌استفاده مجدد و خوش‌ساختار استفاده کرده و به‌دقت اعتبارسنجی کرده باشند.

آزمون 2 — تشخیص خرابی خط لوله با کنترل‌ها (هدف L5-L7). یک لاگ کار خطادار پنهان‌شده ارائه دهید (مثلاً یک خرابی ELT شبانه: تغییر طرح بالادست به‌علاوه یک انفجار کلید null). مشاهده کنید: - آیا می‌توانند از هوش مصنوعی برای خلاصه‌سازی لاگ و پیشنهاد یک علت ریشه‌ای رتبه‌بندی‌شده استفاده کنند، سپس آن را دستی تأیید کنند؟ - آیا یک اصلاح را پیشنهاد می‌دهند که ابتدا در محیط غیرتولیدی آزمایش شده و دارای طرح rollback است؟ - در L6-L7: آیا این را به یک runbook تریاژ قابل‌استفاده مجدد یا یک خودکارسازی دروازه‌دار تبدیل می‌کنند به‌جای یک اصلاح موردی؟ در L5 قبول برای یک گردش‌کار دستی سالم و کنترل‌شده؛ در L7 اگر یک خودکارسازی تأییدشده با rollback طراحی کنند.

سطوح هدف توصیه‌شده

افق	هدف	منطق
پس از 4 ماه	**L4** برای مشارکت‌کنندگان فردی؛ **L5** برای مهندسان داده ارشد/رهبر	همه به‌طور ایمن و قابل‌اتکا SQL، تشخیص و قواعد DQ مجهز به هوش مصنوعی راستی‌آزمایی‌شده را با استفاده فقط از ابزارهای تأییدشده و داده ماسک‌شده تولید می‌کنند. ارشدها هوش مصنوعی را در گردش‌کارهای تکرارشونده با دستاوردهای سنجیده‌شده جای می‌دهند.
پس از 12 ماه	**L5-L6** برای ICها؛ **L7+** برای رهبران/معماران	اکثر ICها گردش‌کارهای داده چندمرحله‌ای هوش مصنوعی را اجرا می‌کنند و هم‌تیمی‌ها را ارتقا می‌دهند؛ رهبران خودکارسازی‌های تحت‌حاکمیت (ناهنجاری، تازگی، drift) را با تأیید و rollback عرضه می‌کنند، و قوی‌ترین‌ها به‌سوی طراحی راه‌حل راستی‌آزمایی‌شده حرکت می‌کنند.

کنترل‌های ایمنی / داده تحت‌نظارت تعبیه‌شده (غیرقابل‌مذاکره در هر سطح)

- **بدون PII واقعی یا داده تولیدی مشتری/تراکنش در هیچ ابزار هوش مصنوعی** — فقط از داده ماسک‌شده یا مصنوعی استفاده کنید.
- **فقط ابزار هوش مصنوعی سازمانی تأییدشده** — بدون چت‌بات‌های عمومی/مصرفی برای داده کاری.
- **پیش از اعتماد راستی‌آزمایی کنید** — خروجی هوش مصنوعی یک پیش‌نویس است؛ در برابر کاتالوگ/طرح زنده اعتبارسنجی کنید، هرگز SQL هوش مصنوعی بازبینی‌نشده را روی تولید اجرا نکنید.
- **ابتدا غیرتولیدی** — همه transformها/اصلاحات پیشنهادی هوش مصنوعی را در یک محیط غیرتولیدی آزمایش کنید.
- **تأیید انسانی و rollback** — هر خودکارسازی که روی داده عمل می‌کند نیازمند یک دروازه تأیید، یک تیکت تغییر و یک مسیر rollback است.
- **حداقل امتیاز** — دسترسی مجهز به هوش مصنوعی باید مجوزهای داده مبتنی بر نقش موجود را رعایت کند؛ هوش مصنوعی استحقاق‌ها (entitlements) را دور نمی‌زند.
- **قابلیت حسابرسی** — یک مسیر از تغییرات مجهز به هوش مصنوعی برای بازبینی ریسک مدل و نظارتی نگه دارید.

(تعاریف سطح نردبان چارچوب L1-L10 بیان‌شده سازمان هستند؛ نگاشت‌های توانمندی و طراحی‌های آزمون بالا توصیه‌های ترکیبی هستند، نه توصیف رویه داخلی هیچ شرکت خاصی.)

x Security / Governance.12

این تیم کارکرد کنترلی برای پذیرش هوش مصنوعی در یک محیط خدمات مالی تحت‌نظارت است. توانمندی در اینجا کمتر بر اساس بهره‌وری شخصی و بیشتر بر اساس توانایی اجازه‌دادن به بقیه سازمان برای استفاده ایمن از هوش مصنوعی سنجیده می‌شود. یک مهندس قوی که سریع عرضه می‌کند اما نمی‌تواند کنترل‌ها را مستند کند، برای این تیم رتبه LOW است.

دامنه ارزیابی‌شده: بازبینی ریسک هوش مصنوعی، حفاظت از اسرار، دفاع در برابر prompt-injection، بازبینی کد هوش مصنوعی، کنترل‌های OWASP LLM، آگاهی از NIST AI RMF و policy-as-code، ISO/IEC 42001، تولید شواهد حسابرسی، پاسخ به رخداد هوش مصنوعی و ریسک vendor/مدل.

LOW / MEDIUM / HIGH چه شکلی دارد

نوار	بازه نردبان	چه چیزی در عمل می‌بینید	وضعیت کنترل تحت‌نظارت
LOW	L1-L3	از GenAI برای پیش‌نویس روتین و جستجوی کد استفاده می‌کند. می‌داند که نباید اسرار، PII یا داده تولیدی را در ابزارهای عمومی وارد کند. هنوز نمی‌تواند ریسک هوش مصنوعی را برای دیگران ارزیابی کند یا کد تولیدشده توسط هوش مصنوعی را برای مسائل injection یا نشت داده بازبینی کند.	از کنترل‌ها پیروی می‌کند؛ آن‌ها را طراحی یا اعمال نمی‌کند. به نظارت روی هر چیز رو به مشتری یا تولید نیاز دارد.
MEDIUM	L4-L6	بازبینی‌های ساختاریافته ریسک هوش مصنوعی را اجرا می‌کند، الگوهای prompt-injection و افشای اسرار را در کد مجهز به هوش مصنوعی می‌گیرد، هنجارهای استفاده ایمن برای یک تیم تعیین می‌کند و کار را به OWASP LLM Top 10 نگاشت می‌کند. از قصد NIST AI RMF و ISO/IEC 42001 آگاه است.	کنترل‌ها را اجرا می‌کند و همکاران را ارتقا می‌دهد. می‌تواند استفاده کم/متوسطریسک هوش مصنوعی را تأیید کند؛ تصمیمات پرریسک و ریسک مدل را تشدید می‌کند.
HIGH	L7-L10	policy-as-code و خطوط لوله خودکار شواهد حسابرسی می‌سازد، پاسخ به رخداد هوش مصنوعی را طراحی می‌کند، ارزیابی ریسک vendor/مدل را رهبری می‌کند و پلتفرم‌های agent محافظت‌شده‌ای را معماری می‌کند که از راستی‌آزمایی ریسک مدل عبور می‌کنند. مالک چارچوب حاکمیت هوش مصنوعی است.	کنترل‌ها را در سطح سازمان تعریف و راستی‌آزمایی می‌کند. پاسخگو به ریسک مدل، حسابرسی و ناظران است.

نمونه‌های نگاشت‌شده به سطوح نردبان

سطح	توانمندی نماینده برای این تیم
L2-L3	قواعد طبقه‌بندی داده و اینکه کدام ابزارهای هوش مصنوعی برای کدام لایه‌های داده تأییدشده‌اند را می‌داند؛ هرگز داده محدودشده را به endpointهای غیرتأییدشده نمی‌فرستد.
L4	پرامپت‌های هدفمند برای اسکن یک diff به دنبال اسرار سیم‌کشی‌شده (hardcoded)، sinkهای injection و مدیریت ناامن خروجی می‌نویسد؛ یافته‌ها را پیش از اقدام راستی‌آزمایی می‌کند.
L5	بررسی‌های امنیتی هوش مصنوعی را در یک گردش‌کار بازبینی چندمرحله‌ای جای می‌دهد (اسکن اسرار pre-commit، بازبینی diff مجهز به هوش مصنوعی، ثبت شواهد) و نرخ مثبت کاذب/گرفتن را می‌سنجد.
L6	متخصص امنیت هوش مصنوعی تیم: هنجارهای استفاده ایمن را منتشر می‌کند، یک خودارزیابی OWASP LLM Top 10 را برای یک محصول اجرا می‌کند، مهندسان را در دفاع از prompt-injection مربیگری می‌کند.
L7	یک خودکارسازی تأییدشده می‌سازد (مثلاً بررسی policy-as-code یا تولیدکننده شواهد حسابرسی) با دروازه‌های تأیید انسانی و rollback پیش از آنکه با یک کنترل سروکار پیدا کند.

سطح	توانمندی نماینده برای این تیم
L8	یک راه‌حل سرتاسری امنیت هوش مصنوعی طراحی می‌کند (مثلاً دریافت (intake) ریسک vendor/مدل مجهز به هوش مصنوعی) که از راستی‌آزمایی رسمی ریسک مدل عبور می‌کند و ریسک باقیمانده را مستند می‌کند.
L9	یک agent یا پلتفرم بازبینی قابل‌استفاده مجدد و محافظت‌شده می‌سازد که در تیم‌ها استفاده می‌شود، با لاگ‌گیری، کنترل دسترسی و کلید قطع.
L10	مالک حاکمیت سازمانی هوش مصنوعی است: برنامه را به NIST AI RMF و ISO/IEC 42001 نگاشت می‌کند، اشتهای ریسک را تعیین می‌کند و به حسابرسی و ناظران پاسخ می‌دهد.

شواهدی که HR می‌تواند راستی‌آزمایی کند

سطح ادعاشده	محصول قابل‌راستی‌آزمایی (نه فقط خوداظهاری)
L3-L4	آموزش تأییدشده استفاده از هوش مصنوعی و مدیریت داده تکمیل‌شده؛ نمونه بازبینی‌های امنیتی مجهز به هوش مصنوعی با یادداشت‌های راستی‌آزمایی.
L5	یک گردش‌کار مستند هوش مصنوعی-در-بازبینی با معیارهای قبل/بعد (مثلاً اسرار گرفته‌شده، زمان بازبینی، نرخ مثبت کاذب).
L6	هنجارهای تیم یا runbook منتشرشده؛ یک ارزیابی OWASP LLM Top 10 که نوشته‌اند؛ سابقه مربیگری یا جلسات داخلی برگزارشده.
L7	یک خودکارسازی ادغام‌شده با دروازه‌های تأیید/rollback در سوابق مدیریت تغییر؛ تأییدهای بازبین.
L8	تأیید راستی‌آزمایی ریسک مدل روی یک راه‌حل که طراحی کرده‌اند؛ پذیرش مستند ریسک باقیمانده.
L9-L10	سوابق مالکیت پلتفرم، چارچوب حاکمیت نگاشت‌شده به NIST AI RMF / ISO 42001، شواهد رو به حسابرسی یا ناظر.

همه استانداردهای ارجاع‌شده را به منابع عمومی نگاشت کنید: OWASP Top 10 for LLM Applications، NIST AI RMF (AI 100-1) و ISO/IEC 42001 سیستم‌های مدیریت هوش مصنوعی. (برداشت‌شده از منابع عمومی: نگاشت‌های کنترل داخلی مشخص بسته به سازمان متفاوت‌اند و در اینجا فرض نشده‌اند.)

نمونه‌های آزمون عملی

1. **بازبینی prompt-injection و نشت داده (هدف L4-L6).** به نامزد یک مشخصه ویژگی کوچک مجهز به هوش مصنوعی یا یک PR بدهید که یک LLM را با ورودی فراهم‌شده توسط کاربر فراخوانی می‌کند و خروجی را به یک اقدام پایین‌دستی برمی‌گرداند. از او بخواهید مسیر prompt-injection، sink مدیریت ناامن خروجی و هر افشای اسرار/PII را شناسایی کند، سپس اقدامات کاهشی پیشنهاد دهد. پاسخ‌های قوی به دسته‌های مشخص OWASP LLM استناد می‌کنند (مثلاً LLM01 Prompt Injection، LLM02 Insecure Output Handling، LLM06 Sensitive Information Disclosure) و یک گام راستی‌آزمایی اضافه می‌کنند.
2. **شواهد حسابرسی تحت یک کنترل (هدف L7-L8).** یک سناریو بدهید: «ما یک ابزار هوش مصنوعی را فقط برای داده tier-2 تأیید کردیم. نشان بده چگونه پایبندی را به یک حسابرس اثبات می‌کنی.» پاسخ‌های قوی policy-as-code یا ثبت خودکار شواهد (لاگ‌های مدل، لایه داده، تأییدکننده، timestamp)، یک دروازه تأیید انسانی و یک کلید قطع/rollback را توصیف می‌کنند، به‌علاوه اینکه چگونه کنترل به یک چارچوب مستند نگاشت می‌شود.

سطوح هدف توصیه‌شده

افق	هدف برای این تیم	منطق
پس از 4 ماه	حداقل **L4**؛ رهبران تیم **L5-L6**	همه باید بتوانند به‌طور ایمن بازبینی‌های امنیتی مجهز به هوش مصنوعی را اجرا کنند و کنترل‌های OWASP LLM را اعمال کنند؛ رهبران باید هنجارها را تعیین کنند.
پس از 12 ماه	حداقل **L5**؛ ارشد/رهبر **L7+**، با حداقل یک مالک **L8-L10**	تیم باید امنیت هوش مصنوعی را به‌عنوان یک گردش‌کار سنجیده‌شده بهره‌برداری کند، با اعضای ارشد که خودکارسازی‌های راستی‌آزمایی‌شده و دروازه‌دار می‌سازند و یک مالک حاکمیت روشن که به ریسک مدل و حسابرسی پاسخگوست.

(سطوح هدف توصیه‌شده یک توصیه ترکیبی هستند، نه یک استاندارد پژوهش‌شده، و باید با اشتهای ریسک و تعهدات نظارتی سازمان کالیبره شوند.)

x QA / Testing.12

هوش مصنوعی در حال تغییر شکل QA از نگارش دستی اسکریپت به‌سوی تولید هوشمند آزمون، اولویت‌بندی مبتنی بر ریسک و سیگنال‌های پیوسته کیفیت است. در یک محیط بانکی تحت‌نظارت، تنش مرکزی سرعت در برابر کنترل است: هوش مصنوعی می‌تواند پوشش را شتاب دهد، اما آزمون‌های تولیدشده، داده مصنوعی و تصمیمات دروازه کیفیت باید قابل‌حسابرسی، بازتولیدپذیر و عاری از داده واقعی مشتری باقی بمانند. این راهنما به HR و رهبران فنی کمک می‌کند کارکنان QA را با استفاده از شواهد قابل‌مشاهده، نه خودارزیابی، روی نردبان L1-L10 قرار دهند.

توانمندی LOW / MEDIUM / HIGH چه شکلی دارد

نوار توانمندی	بازه نردبان	چه چیزی در کار QA مشاهده می‌کنید
LOW	L1-L3	گاهی از هوش مصنوعی برای وظایف مجزا استفاده می‌کند: پیش‌نویس یک مورد آزمون واحد، توضیح یک assertion خطادار، یا تولید داده نمونه برای یک ویژگی کم‌ریسک. همیشه خروجی هوش مصنوعی را کلمه‌به‌کلمه یا با ویرایش حداقلی وارد می‌کند. می‌فهمد که نباید داده واقعی مشتری/حساب را در ابزارهای خارجی وارد کند (L2+). هیچ اندازه‌گیری اثر ندارد. با هوش مصنوعی مانند یک موتور جستجو رفتار می‌کند.
MEDIUM	L4-L6	پرامپت‌هایی طراحی می‌کند که آزمون‌های قابل‌اتکا و صحیح از نظر چارچوب (unit/integration/e2e) و الگوهای پرامپت قابل‌استفاده مجدد برای تیم تولید می‌کنند. هوش مصنوعی را در گردش‌کارهای چندمرحله‌ای QA جای می‌دهد (تولید ← بازبینی ← بازآرایی ← اجرا) و دستاوردهایی مانند افزایش پوشش یا کاهش زمان تریاژ را ردگیری می‌کند. از هوش مصنوعی برای تحلیل ریسک رگرسیون و خوشه‌بندی نقص استفاده می‌کند، سپس یافته‌ها را در برابر تاریخچه خرابی واقعی اعتبارسنجی می‌کند. هنجارهای تیم برای تولید ایمن داده مصنوعی را تعیین می‌کند و به‌عنوان بازبین مرجع هوش مصنوعی عمل می‌کند.
HIGH	L7- L10	خودکارسازی‌های تأییدشده و تحت‌کنترل نسخه تولید آزمون و تریاژ را با rollback و دروازه‌های انسانی می‌سازد. دروازه‌های کیفیت مجهز به هوش مصنوعی و پایش مصنوعی (synthetic monitoring) را طراحی می‌کند که CI/CD را تغذیه می‌کنند و همه از راستی‌آزمایی ریسک مدل و مدیریت تغییر عبور می‌کنند. پلتفرم‌های agent QA قابل‌استفاده مجدد و محافظت‌شده (locatorهای خودترمیم، تریاژ خودکار) را معماری می‌کند که در تیم‌ها استفاده می‌شوند. در L10، مالک استراتژی QA-AI، حاکمیت و وضعیت نظارتی برای آزمون است.

نمونه‌های نگاشت‌شده به سطوح نردبان

نردبان	نمونه ویژه QA
L2	می‌تواند توضیح دهد چرا یک لاگ تراکنش تولیدی واقعی هرگز نباید در یک LLM عمومی وارد شود؛ ابزار داخلی/سازمانی تأییدشده را می‌داند.

نردبان	نمونه ویژه QA
L3	از هوش مصنوعی برای پیش‌نویس موارد آزمون مرزی و افراز هم‌ارزی (equivalence-partition) برای یک تابع کاربردی غیرحساس استفاده می‌کند، سپس هر یک را پیش از commit راستی‌آزمایی می‌کند.
L4	پرامپت‌هایی می‌سازد که آزمون‌های واحد پارامتری صحیح را در چارچوب واقعی تیم (مثلاً JUnit، pytest، Playwright) با fixtureها و assertionهای مناسب تولید می‌کنند.
L5	یک گردش‌کار اجرا می‌کند: هوش مصنوعی سناریوهای e2e را از معیارهای پذیرش تولید می‌کند ← برای شکاف‌ها بازبینی می‌کند ← هوش مصنوعی اجراهای خطادار CI را توضیح می‌دهد ← زمان تریاژ و دلتاهای پوشش را در داشبورد QA لاگ می‌کند.
L6	یک کتابخانه پرامپت بررسی‌شده و راهنماهای داده مصنوعی منتشر می‌کند؛ تیم را در پرامپت‌نویسی ریسک رگرسیون مربیگری می‌کند و مجموعه‌های تولیدشده توسط هوش مصنوعی دیگران را برای ایمنی و کیفیت بازبینی می‌کند.
L7	یک بات تأییدشده می‌سازد که در هر PR، آزمون‌های رگرسیون را برای مسیرهای کد تغییریافته پیشنهاد می‌دهد و توضیحات آزمون‌های خطادار را به PR ارسال می‌کند، پشت یک دروازه تأیید بازبین و rollback.
L8	یک دروازه کیفیت سرتاسری هوش مصنوعی طراحی می‌کند (خوشه‌بندی نقص + امتیازدهی ریسک رگرسیون که mergeهای پرریسک را مسدود می‌کند) که از راستی‌آزمایی ریسک مدل و کنترل تغییر عبور می‌کند.
L9	یک پلتفرم agent خودکارسازی آزمون قابل‌استفاده مجدد و محافظت‌شده (selectorهای خودترمیم، تریاژ خودکار، پایش مصنوعی) می‌سازد که در سطح سازمان با حفاظ‌های مرکزی پذیرفته می‌شود.
L10	مالک استراتژی سازمانی هوش مصنوعی آزمون است: تأیید ابزار، کنترل‌های داده، استانداردهای حسابرسی و همسویی نظارتی برای هوش مصنوعی در SDLC.

شواهدی که HR می‌تواند راستی‌آزمایی کند

نوع شواهد	به دنبال چه باشید	تأیید سطح
تاریخچه commit/PR	آزمون‌های مجهز به هوش مصنوعی که بازبینی و ویرایش شده‌اند، نه به‌صورت خام وارد شده	L3-L5
محصولات پرامپت	پرامپت‌های ذخیره‌شده و قابل‌استفاده مجدد که آزمون‌های صحیح از نظر چارچوب تولید می‌کنند	L4-L6
داشبورد معیارهای QA	افزایش پوشش لاگ‌شده، کاهش زمان تریاژ، روند نقص فراری‌شده (escaped-defect)	+L5
سوابق توانمندسازی تیم	کتابخانه پرامپت، نمایش‌های داخلی، بازبینی‌های همتا، اسناد آنبوردینگ نوشته‌شده	L6
تیکت‌های تغییر/CR	خودکارسازی تأییدشده با مستندسازی rollback و دروازه انسانی	L7-L8
تأیید ریسک مدل / راستی‌آزمایی	دروازه کیفیت یا مدل امتیازدهی که از بازبینی راستی‌آزمایی عبور کرده	L8
سوابق پذیرش پلتفرم	تیم‌های دیگری که از agent/ابزار QA نامزد استفاده می‌کنند	L9
مالکیت حاکمیت	مالک نام‌برده سیاست هوش مصنوعی آزمون، تأیید ابزار، استانداردهای حسابرسی	L10

> نکته HR: لینک‌های PRها، داشبوردها، مخازن پرامپت و تیکت‌های CR/راستی‌آزمایی را بخواهید. خوداظهاری «من روزانه از هوش مصنوعی استفاده می‌کنم» شواهد بالاتر از L3 نیست. (برداشت‌شده از منابع عمومی.)

نمونه‌های آزمون عملی

آزمون A — پرامپت‌نویسی مؤثر و راستی‌آزمایی (هدف L4): به نامزد یک مشخصه تابع پاکسازی‌شده بدهید (مثلاً یک محاسبه‌گر تعلق سود (interest-accrual)) و از او بخواهید از ابزار هوش مصنوعی تأییدشده برای تولید یک مجموعه آزمون واحد استفاده کند. معیار قبولی: آزمون‌ها در چارچوب واقعی کامپایل و اجرا می‌شوند؛ مرزها، حالت‌های منفی و حالت‌های لبه‌ای گردکردن/دقت را پوشش می‌دهند؛ نامزد حداقل یک شکاف معرفی‌شده توسط هوش مصنوعی یا assertion نادرست را شناسایی و اصلاح می‌کند؛ هیچ داده واقعی استفاده نشده.

آزمون B — گردش‌کار، تحلیل ریسک و داده ایمن (هدف L5-L6): یک مجموعه تغییر کوچک (یک diff) به‌علاوه تاریخچه خرابی پیشین ارائه دهید. از نامزد بخواهید از هوش مصنوعی برای (1) شناسایی نواحی با بالاترین ریسک رگرسیون، (2) خوشه‌بندی نقص‌های تاریخی برای اولویت‌بندی، و (3) تولید داده آزمون مصنوعی بدون هیچ سابقه تولیدی استفاده کند. معیار قبولی: رتبه‌بندی ریسک در برابر تاریخچه واقعی موجه است (نه‌فقط ادعای هوش مصنوعی)؛ داده مصنوعی به‌وضوح غیرواقعی و سازگار با سیاست است؛ نامزد بهبود مورد انتظار پوشش/ترياژ را می‌سنجد و بیان می‌کند.

کنترل‌های ایمنی و داده تحت‌نظارت تعبیه‌شده

- **بدون داده واقعی مشتری/حساب/PII** در هیچ پرامپتی؛ فقط داده مصنوعی یا ماسک‌شده، تولیدشده از طریق روش‌های تأییدشده.
- **فقط ابزارهای تأییدشده** — endpointهای LLM مجاز سازمانی؛ بدون ابزارهای عمومی مصرفی برای محصولات کاری (دروازه L2).
- **انسان‌درحلقه**: آزمون‌های تولیدشده و خروجی‌های ترياژ هوش مصنوعی پیش از merge بازبینی می‌شوند؛ هوش مصنوعی هرگز فراتر از حالت مشورتی یک دروازه کیفیت را خودکار تأیید نمی‌کند.
- **قابلیت حسابرسی**: پرامپت‌ها، محصولات تولیدشده و تصمیمات دروازه تحت‌کنترل نسخه و قابل‌ردگیری برای بازبینی نظارتی هستند.
- **مدیریت تغییر**: هر خودکارسازی هوش مصنوعی که با CI/CD سروکار دارد (L7+) نیازمند تأیید CR، طرح rollback و در جایی که merge‌ها را امتیازدهی/مسدود می‌کند، راستی‌آزمایی ریسک مدل (L8+) است.

سطوح هدف توصیه‌شده

افق	هدف	منطق
پس از 4 ماه	**L4** (پایه تیم L3)	اکثر کارکنان QA باید به‌طور قابل‌اتکا آزمون‌های صحیح از نظر چارچوب را با پرامپت‌نویسی هدفمند و پایبندی کامل به مدیریت داده تولید و راستی‌آزمایی کنند.
پس از 12 ماه	**L5-L6** (رهبران/ارشدها به‌سوی L7)	هوش مصنوعی را در گردش‌کارهای QA با دستاوردهای سنجیده‌شده جای دهید؛ مهندسان QA ارشد و رهبران شروع به ساخت خودکارسازی‌های تأییدشده و دروازه‌دار می‌کنند.

> اهداف نقش‌وزن‌دار هستند: QA جونیور باید L3-L4 را با عادات ایمن تثبیت کند؛ QA ارشد و SDETها باید به L5-L6 برسند؛ رهبران/معماران QA باید L7+ را با انضباط کامل مدیریت تغییر و راستی‌آزمایی دنبال کنند. (برداشت‌شده از منابع عمومی — با تعاریف نقش داخلی خود کالیبره کنید.)

12.x Release / Change Management

این راهنما به مدیران و HR کمک می‌کند ارزیابی کنند یک متخصص Release / Change Management چقدر خوب از ابزارهای GenAI در سراسر چرخه عمر تغییر، از دریافت تا بازبینی پس از پیاده‌سازی، استفاده می‌کند. در یک محیط بانکی تحت‌نظارت، هر محصول تغییر مجهز به هوش مصنوعی (تحلیل اثر، امتیاز ریسک، طرح rollback، بسته CAB، شواهد پایبندی) یک سابقه است که ممکن است توسط حسابرسان و ناظران بررسی شود. خروجی هوش مصنوعی در اینجا همیشه مشورتی است: یک مالک تغییر انسانی نام‌برده پاسخگو باقی می‌ماند و هیئت مشورتی تغییر (CAB) تأیید می‌کند.

توانمندی LOW / MEDIUM / HIGH چه شکلی دارد

نوار	نردبان	چه شکلی در کار Release/Change دارد	وضعیت ریسک معمول
LOW	L1-L3	یک تیکت تغییر را در یک چت‌بات وارد می‌کند تا یک توصیف را «مرتب کند»؛ از هوش مصنوعی فقط روی وظایف منفرد و کم‌ریسک استفاده می‌کند؛ به‌صورت ناسازگار راستی‌آزمایی می‌کند؛ بی‌اطلاع از اینکه کدام فیلدها حساس هستند (داده مشتری، IP زیرساخت، شرایط vendor).	غیرمدیریت‌شده. ریسک نشت محتوای تغییر؛ امتیازهای ریسک راستی‌آزمایی‌نشده که به CAB می‌رسند.
MEDIUM	L4-L6	پرامپت‌نویسی هدفمند برای پیش‌نویس تحلیل اثر، یادداشت‌های انتشار و طرح‌های rollback؛ امتیازدهی ریسک هوش مصنوعی را در برابر ماتریس ریسک تغییر سازمان و CMDB متقابلاً بررسی می‌کند؛ قالب‌های پرامپت قابل‌استفاده مجدد که تیم می‌پذیرد می‌سازد؛ هنجارهایی درباره اینکه چه داده‌ای می‌تواند و نمی‌تواند وارد شود تعیین می‌کند.	مدیریت‌شده در سطح وظیفه/گردش‌کار. انسان همیشه پیش از CAB نهایی می‌کند.
HIGH	L7-L10	خودکارسازی‌های تأییدشده‌ای طراحی می‌کند که از CMDB، ticketing (مثلاً ServiceNow/Jira) و داده خط لوله می‌کشند تا سوابق تغییر را با rollback و دروازه‌های تأیید تعبیه‌شده پیش‌پر کنند؛ مدل‌های پیش‌بینی شکست تغییر را از طریق بازبینی ریسک مدل راستی‌آزمایی می‌کند؛ حاکمیت سازمانی change-AI و استانداردهای شواهد حسابرسی را تعریف می‌کند.	تحت‌حاکمیت. حفاظ‌ها، لاگ‌گیری و راستی‌آزمایی در خط لوله تعبیه شده‌اند.

نمونه‌های نگاشت‌شده به سطوح نردبان

نردبان	توانمندی نماینده در این تیم
L2	می‌تواند توضیح دهد چرا یک سابقه تغییر با نام میزبان‌های تولیدی یا شناسه‌های مشتری نباید در یک ابزار خارجی غیرتأییدشده وارد شود.
L3	از یک ابزار تأییدشده برای خلاصه‌سازی یک درخواست تغییر و کوتاه‌کردن یادداشت‌های انتشار استفاده می‌کند، سپس هر واقعیت را به‌صورت دستی در برابر تیکت منبع بررسی می‌کند.
L4	پرامپت‌های ساختاریافته‌ای می‌نویسد که یک قالب سازگار تحلیل اثر (سرویس‌های متأثر، وابستگی‌ها، شعاع انفجار) تولید می‌کنند و آن‌ها را برای کاهش دوباره‌کاری پالایش می‌کند.
L5	هوش مصنوعی را در سراسر گردش‌کار تغییر جای می‌دهد — خلاصه‌سازی، پیش‌نویس امتیاز ریسک، پیش‌نویس طرح rollback، یادداشت‌های انتشار — و کاهش زمان چرخه و دوباره‌کاری را می‌سنجد.
L6	استاندارد تیم برای بسته‌های CAB مجهز به هوش مصنوعی را تعیین می‌کند، محصولات تغییر مجهز به هوش مصنوعی همتایان را بازبینی می‌کند و دیگران را در مدیریت ایمن داده آموزش می‌دهد.
L7	یک خودکارسازی تأییدشده می‌سازد که بسته CAB و طرح rollback را از داده تیکت + CMDB پیش‌نویس می‌کند، با دروازه‌های اجباری تأیید انسانی و rollback پیش از استقرار.
L8	یک راه‌حل سرتاسری پیش‌بینی شکست تغییر طراحی می‌کند که از راستی‌آزمایی ریسک مدل عبور می‌کند و توصیه‌های پنجره استقرار را تغذیه می‌کند.
L9	یک agent/پلتفرم مدیریت تغییر قابل‌استفاده مجدد و محافظت‌شده می‌سازد که در چند تیم تحویل استفاده می‌شود.
L10	مالک سیاست سازمانی برای هوش مصنوعی در تغییر/انتشار است، از جمله قابلیت حسابرسی، همسویی نظارتی و مرزهای پاسخگویی.

شواهدی که HR می‌تواند راستی‌آزمایی کند

حوزه توانمندی	چه چیزی به‌عنوان شواهد بخواهید
خلاصه‌سازی درخواست تغییر	یک نمونه قبل/بعد که خلاصه هوش مصنوعی را در برابر تیکت منبع تطبیق‌داده، با نام راستی‌آزماینده.
تحلیل اثر	ارزیابی اثری که وابستگی‌های پیش‌نویس‌شده توسط هوش مصنوعی را در برابر CMDB متقابلاً بررسی می‌کند، با درج اصلاحات.
امتیازدهی ریسک	مقایسه کنار هم امتیاز ریسک پیشنهادی هوش مصنوعی در برابر نتیجه ماتریس ریسک تغییر سازمان، به‌علاوه اینکه چه کسی امتیاز نهایی را تأیید کرد.
یادداشت‌های انتشار	یادداشت‌های انتشار منتشرشده به‌علاوه پرامپت/قالب و تأیید بازبینی.
پشتیبانی تصمیم CAB	یک بخش بسته CAB تولیدشده با کمک هوش مصنوعی، با یک یادداشت روشن که یک مالک انسانی آن را نهایی کرد و CAB تأیید کرد.
توصیه پنجره استقرار	توصیه ردگیری‌شده به تقویم‌های freeze، داده وابستگی و قواعد اثر تجاری — نه یک حدس بی‌منبع هوش مصنوعی.
تولید طرح rollback	طرح rollback راستی‌آزمایی‌شده در برابر گام‌های واقعی استقرار، با یک نقطه بازرسی آزمایش‌شده یا بازبینی‌شده توسط همتا.
پیش‌بینی شکست تغییر	برای L7+، شواهدی که مدل از راستی‌آزمایی ریسک مدل عبور کرده و خروجی‌ها لاگ و پایش می‌شوند.
شواهد پایبندی	مسیر حسابرسی که نشان می‌دهد محصولات مجهز به هوش مصنوعی برچسب‌گذاری، لاگ و طبق سیاست سوابق نگه‌داری می‌شوند.

برای هر مورد، HR باید سه چیز را تأیید کند: داده منبع برای ابزار استفاده‌شده تأیید شده بود، یک انسان نام‌برده بازبینی و تأیید کرد، و محصول برای حسابرسی لاگ/نگه‌داری می‌شود.

نمونه‌های آزمون عملی

آزمون 1 — بسته تغییر امتیازدهی‌شده با ریسک (هدف L4-L5). به نامزد یک درخواست تغییر پاکسازی‌شده بدهید. از او بخواهید تولید کند: یک خلاصه یک‌پاراگرافی، یک تحلیل اثر، یک پیش‌نویس امتیاز ریسک و یک طرح rollback. بر اساس اینکه آیا (الف) فقط از داده تأییدشده/پاکسازی‌شده استفاده کرد، (ب) امتیاز ریسک هوش مصنوعی را در برابر ماتریس ریسک تغییر متقابلاً بررسی و اصلاح کرد، (ج) طرح rollback را به گام‌های واقعی استقرار گره زد، و (د) آنچه را که یک انسان/CAB باید پیش از تأیید راستی‌آزمایی کند علامت زد، امتیاز دهید.

آزمون 2 — بازبینی طراحی خودکارسازی (هدف L7-L8). یک پیشنهاد برای تولید خودکار بسته‌های CAB از داده تیکت + CMDB ارائه دهید. از نامزد بخواهید حفاظ‌های موردنیاز را شناسایی کند. یک پاسخ قوی نام می‌برد: دروازه اجباری تأیید انسانی، دروازه rollback پیش از استقرار، محدودیت‌های دامنه داده (بدون فیلدهای حساس غیرتأییدشده)، مسیر کامل لاگ‌گیری/حسابرسی، و یک گام راستی‌آزمایی مدل برای هر امتیازدهی پیش‌بینی‌کننده ریسک/شکست.

سطوح هدف توصیه‌شده

افق	مشارکت‌کننده فردی (تحلیلگر/مدیر تغییر)	رهبر فنی / مالک تغییر
پس از 4 ماه	L4 — پیش‌نویس قابل‌اتکا و راستی‌آزمایی‌شده محصولات تغییر مجهز به هوش مصنوعی	L5 — هوش مصنوعی جای‌گرفته در سراسر گردش‌کار تغییر با دستاوردهای سنجیده‌شده
پس از 12 ماه	L5 — استفاده در سطح گردش‌کار با بهبود سنجیده‌شده زمان چرخه/دوباره‌کاری	L7 — مالک حداقل یک خودکارسازی تغییر تأییدشده با دروازه‌های تأیید انسانی و rollback

(برداشت‌شده از منابع عمومی) — این اهداف از رویه عمومی مدیریت تغییر ITIL/ITSM و الگوهای منتشرشده vendor (مثلاً خلاصه‌سازی تغییر و پیش‌بینی ریسک مجهز به هوش مصنوعی در پلتفرم‌های مرسوم ITSM) ترکیب شده‌اند، نه از یک نردبان داخلی شرکتی. سرعت را با بلوغ CAB و ظرفیت حاکمیت ریسک مدل خود کالیبره کنید.

کنترل‌های تحت‌نظارت تعبیه‌شده (در هر سطح اعمال می‌شوند): فقط ابزارهای تأییدشده و دامنه‌های داده تأییدشده؛ بدون داده مشتری، اعتبارنامه‌ها یا اسرار تولیدی در پرامپت‌ها به ابزارهای غیرتأییدشده؛ پاسخگویی انسانی نام‌برده و تأیید CAB برای هر تغییر حفظ‌شده؛ خروجی‌های هوش مصنوعی برچسب‌گذاری‌شده به‌عنوان مجهز به هوش مصنوعی؛ لاگ‌گیری کامل و نگه‌داری سوابق برای حسابرسی؛ و راستی‌آزمایی ریسک مدل پیش از آنکه هر امتیازدهی پیش‌بینی‌کننده (پیش‌بینی شکست تغییر، امتیازدهی ریسک) در تصمیمات استفاده شود.

x Managers / Technical Leaders.12

این راهنما مدیران و رهبران فنی را نه بر اساس مهارت پرامپت‌نویسی شخصی‌شان، بلکه بر اساس اینکه چقدر خوب **توانمندی هوش مصنوعی را در سراسر تیم‌هایشان می‌سازند، حاکمیت می‌کنند و مقیاس می‌دهند** در یک محیط خدمات مالی تحت‌نظارت، ارزیابی می‌کند. یک مدیر می‌تواند یک کاربر قوی فردی هوش مصنوعی (L4-L5) باشد اما به‌عنوان یک رهبر رتبه LOW بگیرد اگر هیچ نقشه راه، حاکمیت و اندازه‌گیری نداشته باشد. رفتارهای رهبری زیر را ارزیابی کنید، سپس به نردبان نگاشت کنید.

توانمندی رهبری LOW / MEDIUM / HIGH چه شکلی دارد

حوزه توانمندی	LOW (شکاف)	MEDIUM (در حال توسعه)	HIGH (قوی)
نقشه راه پذیرش هوش مصنوعی	بدون برنامه؛ استفاده موردی از ابزار؛ «هر کسی خودش بفهمد»	یک برنامه 1-2 سه‌ماهه وجود دارد اما ردگیری یا تأمین مالی نمی‌شود	نقشه راه 12ماهه ترتیب‌یافته با نقاط عطف، مالکان، بودجه و آهنگ بازبینی
مدل حاکمیت	بدون قواعد درباره داده، ابزار یا تأییدها؛ هوش مصنوعی سایه تحمل می‌شود	از سیاست سازمانی استفاده مجدد می‌کند اما بدون اعمال محلی	کنترل‌های محلی روی سیاست سازمانی لایه‌بندی شده؛ فهرست ابزارهای تأییدشده، قواعد طبقه داده، مالک پاسخگوی نام‌برده
اولویت‌بندی موارد استفاده	به دنبال هیاهو می‌رود؛ بدون تریاژ	یک backlog نگه می‌دارد اما بر اساس اشتیاق رتبه‌بندی می‌کند	موارد استفاده را بر اساس ارزش، ریسک و امکان‌پذیری امتیازدهی می‌کند؛ موارد پرریسک/کم‌ارزش را صریحاً رد می‌کند
اندازه‌گیری ROI	«سریع‌تر به‌نظر می‌رسد»؛ بدون خط پایه	فقط معیارهای استفاده/فعالیت را ردگیری می‌کند	معیارهای نتیجه (زمان چرخه، نرخ نقص، هزینه) را در برابر یک خط پایه پیش از هوش مصنوعی ردگیری می‌کند
مدیریت ریسک	بی‌اطلاع از ریسک مدل، نشت داده، مواجهه نظارتی	می‌داند ریسک‌ها وجود دارند؛ به دیگران برای مدیریت متکی است	یک register ریسک نگه می‌دارد؛ انسان‌درحلقه، rollback و مسیرهای حسابرسی را روی کار مجهز به هوش مصنوعی تضمین می‌کند
برنامه‌ریزی توانمندی	هیچ دیدی از سطوح تیم ندارد	حس غیررسمی از اینکه چه کسی «با هوش مصنوعی خوب است»	تیم را به سطوح نردبان نگاشت می‌کند؛ استخدام/آموزش برای بستن شکاف‌ها برنامه‌ریزی می‌کند
استراتژی آموزش	هیچ	یک lunch-and-learn موردی	توانمندسازی مبتنی بر نقش و تکرارشونده گره‌خورده به نردبان با پیشرفت سنجیده‌شده
شبکه قهرمانان هوش مصنوعی	هیچ	یک فرد مرجع غیررسمی	قهرمانان نام‌برده (L6) با زمان اختصاص‌یافته، مأموریت تعیین هنجار و دسترس‌پذیری در تیم‌ها

حوزه توانمندی	LOW (شکاف)	MEDIUM (در حال توسعه)	HIGH (قوی)
پذیرش نردبان HR	بی‌اطلاع از نردبان	به نردبان به‌طور سست ارجاع می‌دهد	از نردبان در بازبینی‌ها، اهداف و گفتگوهای ترفیع به‌طور سازگار استفاده می‌کند
همسویی بین‌تیمی	جزیره‌ای؛ تلاش را تکرار می‌کند	هماهنگی گاه‌به‌گاه	الگوهای قابل‌استفاده مجدد را به اشتراک می‌گذارد، با ریسک/پایبندی/پلتفرم همسو می‌شود، از تکرار پرهیز می‌کند

نمونه‌های نگاشت‌شده به سطوح نردبان

رفتار رهبری مشاهده‌شده	نشانه نردبان
رهبر از GenAI استفاده نکرده و نمی‌تواند قواعد مدیریت داده تیم را بیان کند	L1
رهبر می‌تواند پیش از توانمندسازی تیم، ابزارهای تأییدشده و محدودیت‌های طبقه داده را توضیح دهد	L2-L3
رهبر هوش مصنوعی را در گردش‌کارهای تیم جای می‌دهد و دستاوردها را در برابر یک خط پایه می‌سنجد	L5
رهبر به‌عنوان متخصص هوش مصنوعی تیم عمل می‌کند، هنجارهای ایمن تعیین می‌کند و دیگران را ارتقا می‌دهد (یک قهرمان)	L6
رهبر خودکارسازی‌های تولیدی را با دروازه‌های مستند تأیید و rollback حمایت می‌کند	L7
رهبر بر راه‌حل‌های سرتاسری هوش مصنوعی که از راستی‌آزمایی ریسک مدل عبور می‌کنند نظارت می‌کند	L8
رهبر مالک استراتژی، حاکمیت و ریسک نظارتی هوش مصنوعی برای سازمان/حوزه است	L10

> نکته: یک مدیر افراد ملزم به ساخت در L7-L9 به‌صورت شخصی نیست. از او انتظار می‌رود کار در آن سطوح را **حاکمیت، حمایت و توانمند** سازد. رهبر را بر اساس اینکه آیا کار تحت نظر او از آستانه کنترل متناظر عبور می‌کند ارزیابی کنید، نه بر اساس اینکه آیا کد را نوشته است.

شواهدی که HR می‌تواند راستی‌آزمایی کند

شواهد	کجا آن را بیابید
نقشه راه مکتوب پذیرش هوش مصنوعی با مالکان، تاریخ‌ها و بودجه	اسناد برنامه‌ریزی تیم/دپارتمان
backlog موارد استفاده امتیازدهی‌شده بر اساس ارزش/ریسک/امکان‌پذیری	ابزار backlog، برگه اولویت‌بندی
خط پایه ROI و معیارهای پس از هوش مصنوعی	داشبوردهای تحویل، گزارش‌های قبل/بعد
ورودی‌های register ریسک برای موارد استفاده هوش مصنوعی؛ تأییدهای انسان‌درحلقه	ابزار GRC/ریسک، سوابق تأیید تغییر
فهرست ابزارهای تأییدشده و قواعد طبقه‌بندی داده پذیرفته‌شده توسط تیم	گواهی‌های سیاست، سوابق آنبوردینگ
قهرمانان هوش مصنوعی نام‌برده با زمان اختصاص‌یافته	چارت سازمانی، برنامه‌های ظرفیت/تخصیص
سطوح نردبان ثبت‌شده در بازبینی‌های عملکرد و برنامه‌های توسعه	سیستم بازبینی HR

شواهد	کجا آن را بیابید
شواهد استفاده مجدد بین‌تیمی (پرامپت‌ها، اجزا، استانداردهای مشترک)	مخازن داخلی، جوامع عمل (communities of practice)

نمونه‌های آزمون عملی

آزمون 1 - مرور نقشه راه و حاکمیت (45 دقیقه). از رهبر بخواهید نقشه راه پذیرش هوش مصنوعی تیمش را ارائه دهد و پاسخ دهد: کدام موارد استفاده در جریان هستند و چرا اولویت‌بندی شدند؟ فهرست ابزارهای تأییدشده چیست و هر ابزار با کدام طبقات داده می‌تواند سروکار داشته باشد؟ اگر یک خروجی مجهز به هوش مصنوعی موجب یک مشکل مشتری یا نظارتی شود، مالک پاسخگو کیست؟ یک خط پایه ROI و نتیجه پس از هوش مصنوعی را نشان دهید. - LOW: بدون نقشه راه، بدون فهرست ابزار، نمی‌تواند یک مالک پاسخگو نام ببرد. - MEDIUM: نقشه راه و فهرست ابزار وجود دارند، اما ROI مبتنی بر فعالیت است و مالکیت ریسک مبهم است. - HIGH: نقشه راه اولویت‌بندی‌شده، قواعد ابزار/داده اعمال‌شده، مالک نام‌برده، ROI مبتنی بر نتیجه در برابر خط پایه.

آزمون 2 - تمرین رومیزی رخداد و کنترل (30 دقیقه). یک سناریو مطرح کنید: یک مهندس از یک ابزار هوش مصنوعی استفاده کرد که ممکن است PII مشتری را در یک مدل غیرتأییدشده افشا کرده باشد، و یک تغییر مجهز به هوش مصنوعی بدون بازبینی به تولید رسید. - LOW: بدون مسیر تشخیص، بدون rollback، بی‌اطلاع از وظایف گزارش‌دهی نقض. - MEDIUM: می‌تواند rollback کند و می‌داند تشدید وجود دارد، اما بدون کنترل پیشگیرانه. - HIGH: کنترل‌های پیشگیرانه (دروازه‌بندی ابزار تأییدشده، بررسی‌های نشت داده، بازبینی انسانی اجباری، rollback و مسیر حسابرسی)، مسیر اطلاع‌رسانی نظارتی و اصلاح کنترل پس از رخداد را توصیف می‌کند.

سطوح هدف توصیه‌شده

افق	هدف برای سطح نردبان خود رهبر	هدف برای آنچه در تیم توانمند می‌سازند
پس از 4 ماه	L4-L5 (کاربر مؤثر در سطح گردش‌کار)	اکثر تیم در L3+؛ نقشه راه، فهرست ابزارهای تأییدشده و خط پایه ROI در جای خود؛ قهرمان(ها) نام‌برده
پس از 12 ماه	L6 (متخصص/قهرمان هوش مصنوعی تیم) یا مسیر حاکمیت L7+ برای رهبران ارشد	میانه تیم L4-L5؛ خودکارسازی‌های تولیدی (L7) در حال اجرا تحت دروازه‌های تأیید/rollback؛ نردبان در همه بازبینی‌ها استفاده‌شده؛ الگوهای بین‌تیمی به اشتراک گذاشته‌شده

کنترل‌های تحت‌نظارت تعبیه‌شده (غیرقابل‌مذاکره در هر سطح): فقط فهرست ابزارهای تأییدشده؛ قواعد طبقه‌بندی داده اعمال‌شده (بدون PII محدودشده/مشتری در مدل‌های غیرتأییدشده)؛ انسان‌درحلقه روی هر خروجی رو به مشتری یا رو به تولید؛ دروازه‌های مستند تأیید و rollback برای خودکارسازی‌ها؛ مسیر حسابرسی و راستی‌آزمایی ریسک مدل برای راه‌حل‌های L8+؛ مالک پاسخگوی روشن و مسیر اطلاع‌رسانی نظارتی برای رخدادهای مرتبط با هوش مصنوعی.

(برداشت‌شده از منابع عمومی: نگاشت‌های نردبان به رهبری و افق‌های هدف توصیه‌های ترکیبی مبتنی بر رویه متداول توانمندسازی صنعتی و مرجع L1-L10 ارائه‌شده هستند، نه یک نردبان HR شرکتی منتشرشده.)

---

13. سنجه امتیازدهی منابع انسانی (HR Scoring Rubric)

این بخش یک سنجه عملی و مبتنی بر امتیاز برای ارزیابی قابلیت هوش مصنوعی در هر ۱۶ بُعد ارائه می‌کند. این سنجه به‌گونه‌ای طراحی شده که هر ارزیاب (شریک منابع انسانی، مدیر، یا سرپرست فنی) بتواند یک توصیف را بخواند، آن را با شواهد مشاهده‌شده تطبیق دهد و امتیازی قابل دفاع تخصیص دهد. امتیازها سپس در نردبان ۰اسطحی قابلیت هوش مصنوعی که در سایر بخش‌های این گزارش به‌کار رفته است، تجمیع می‌شوند.

13.1 شیوه استفاده از این سنجه

- هر یک از ۱۶ بُعد را به‌طور مستقل بر اساس مقیاس 0–5 و با استفاده از لنگرهای جهانی زیر امتیازدهی کنید.
- هر امتیاز را بر پایه **شواهد مشاهده‌شده** قرار دهید (نمونه‌های کاری، خودکارسازی‌های پیاده‌سازی‌شده، promptهای بازبینی‌شده، نحوه مدیریت رخدادها، نظر همکاران/مدیر) — نه صرفاً خودارزیابی.
- برای هر امتیاز ۳ یا بالاتر، دست‌کم یک شاهد عینی الزامی است.
- امتیاز، شواهد و تاریخ را ثبت کنید. ارزیابی مجدد را با تناوبی ثابت انجام دهید (برای نمونه، هر ۶ ماه).

لنگرهای جهانی امتیاز (برای همه ابعاد کاربرد دارد)

امتیاز	برچسب	معنای جهانی
0	نمایش‌داده‌نشده	هیچ شاهدی وجود ندارد؛ این کار را انجام نداده یا از وجود آن بی‌اطلاع است.
1	آگاهی پایه	می‌داند مفهوم وجود دارد و چرا اهمیت دارد؛ اما هنوز نمی‌تواند آن را به‌کار ببندد.
2	دانش کاربردی	می‌تواند آن را با راهنمایی، الگو، یا نظارت در موارد ساده به‌کار ببندد.
3	مجری مستقل	آن را به‌طور قابل‌اتکا و بدون نظارت در کارهای روزمره و واقعی به‌کار می‌برد.
4	مجری پیشرفته	موارد پیچیده/مبهم را مدیریت می‌کند، بهینه‌سازی می‌کند و دیگران را راهنمایی می‌کند.
5	خبره / رهبر	استانداردها را تعیین می‌کند، در دیگران توانمندی می‌سازد و مالک پیامدها در سطح سازمان است.

13.2 سنجه به‌تفکیک هر بُعد

هر جدول زیر، لنگرهای جهانی را با رفتار عینی و قابل‌مشاهده برای آن بُعد تخصصی می‌کند. همان مقیاس 0–5 در سراسر آن کاربرد دارد.

D1. سواد هوش مصنوعی

درک اینکه AI/ML چیست، چه کاری از آن برمی‌آید و چه کاری نه، و کجا به‌کار می‌آید.

امتیاز	این‌گونه به نظر می‌رسد
0	نمی‌تواند توصیف کند هوش مصنوعی چه می‌کند؛ آن را معجزه می‌پندارد یا نادیده می‌گیرد.
1	می‌داند هوش مصنوعی وجود دارد و مرتبط است؛ می‌تواند چند کاربرد را نام ببرد.
2	مفاهیم پایه (مدل، آموزش، پیش‌بینی) و محدودیت‌های آشکار را توضیح می‌دهد.
3	به‌طور مستقل قضاوت می‌کند که هوش مصنوعی برای کارهای روزمره مناسب هست یا نیست.
4	مصالحه‌ها (هزینه، دقت، سوگیری، تأخیر) را برای غیرمتخصصان توضیح می‌دهد؛ کج‌فهمی‌ها را اصلاح می‌کند.
5	درک مشترک سازمان از هوش مصنوعی را شکل می‌دهد؛ مرجع داخلی مورد اعتماد است.

D2. سواد GenAI

درک اختصاصی مدل‌های مولد (LLMها، context، token، توهم‌زایی، ناقطعی‌بودن).

امتیاز	این‌گونه به نظر می‌رسد
0	از GenAI یا تفاوت آن با جست‌وجو/خودکارسازی بی‌اطلاع است.

امتیاز	این‌گونه به نظر می‌رسد
1	می‌داند GenAI متن/تصویر تولید می‌کند و می‌تواند اشتباه کند.
2	توهم‌زایی، محدودیت‌های context و ناقطعی‌بودن را در سطح پایه درک می‌کند.
3	مدل‌ها/تنظیمات مناسب را برای کارهای روزمره انتخاب می‌کند و حالت‌های خطا را پیش‌بینی می‌کند.
4	درباره پنجره‌های context، grounding و مصالحه هزینه/کیفیت برای کارهای پیچیده استدلال می‌کند.
5	استانداردهای انتخاب و استفاده از مدل را که دیگران دنبال می‌کنند، تعریف می‌کند.

D3. مهندسی prompt

طراحی ورودی‌هایی که به‌طور قابل‌اتکا خروجی درست و مفید تولید می‌کنند.

امتیاز	این‌گونه به نظر می‌رسد
0	prompt نمی‌نویسد یا انتظار دارد درخواست‌های یک‌خطی به‌خودی‌خود «کار کنند».
1	promptهای ساده می‌نویسد؛ هر چه بازگردد را می‌پذیرد.
2	از ساختار (نقش، context، مثال) با الگوهایی که در اختیارش گذاشته شده، استفاده می‌کند.
3	به‌طور مستقل promptها را تکرار و اصلاح می‌کند تا در کارهای واقعی نتایج قابل‌اتکا بگیرد.
4	تکنیک‌های پیشرفته (تجزیه، few-shot، قید، ارزیابی) را برای مسائل دشوار به‌کار می‌برد.
5	کتابخانه‌ها/الگوهای prompt می‌سازد و به اشتراک می‌گذارد؛ تیم‌ها را در تکنیک مربی‌گری می‌کند.

D4. استفاده از ابزار هوش مصنوعی

بهره‌برداری عملی از ابزارها و سکوهای تأییدشده هوش مصنوعی.

امتیاز	این‌گونه به نظر می‌رسد
0	از هیچ ابزار هوش مصنوعی برای کار استفاده نکرده است.
1	یک ابزار تأییدشده را یکی‌دو بار امتحان کرده است.
2	از یک ابزار تأییدشده برای کارهای ساده با کمک استفاده می‌کند.
3	از ابزارهای تأییدشده به‌روانی و مستقل در کار روزانه استفاده می‌کند.
4	چند ابزار/قابلیت را ترکیب می‌کند و از قابلیت‌های پیشرفته بهره می‌برد.
5	ابزارها را ارزیابی، توصیه و معرفی می‌کند؛ قراردادهای استفاده را تعیین می‌کند.

D5. کاربرد ویژه شغل

به‌کارگیری هوش مصنوعی در وظایف نقش واقعی فرد.

امتیاز	این‌گونه به نظر می‌رسد
0	هیچ پیوندی میان هوش مصنوعی و شغل خود برقرار نکرده است.

امتیاز	این‌گونه به نظر می‌رسد
1	می‌تواند نام ببرد هوش مصنوعی کجای نقش ممکن است کمک کند.
2	هوش مصنوعی را برای چند کار ساده و مرتبط با نقش، با راهنمایی به‌کار می‌برد.
3	به‌طور معمول از هوش مصنوعی برای انجام سریع‌تر/بهتر کار واقعی نقش، بدون نظارت استفاده می‌کند.
4	بخش‌هایی از کار نقش را پیرامون هوش مصنوعی بازطراحی می‌کند؛ با بهبود قابل‌اندازه‌گیری.
5	شیوه کارکرد آن حوزه را با هوش مصنوعی بازشکل می‌دهد؛ الگوی آن رشته است.

D6. تفکر انتقادی / اعتبارسنجی

قضاوت درباره اینکه خروجی هوش مصنوعی درست، مرتبط و مناسب هست یا نه.

امتیاز	این‌گونه به نظر می‌رسد
0	خروجی هوش مصنوعی را بدون چون‌وچرا می‌پذیرد.
1	می‌داند خروجی می‌تواند نادرست باشد اما به‌ندرت بررسی می‌کند.
2	هنگامی که یادآوری شود، خروجی را در برابر یک منبع آشکار بررسی می‌کند.
3	به‌عادت، خروجی را پیش از استفاده در برابر منابع/منطق اعتبارسنجی می‌کند.
4	برای خروجی‌های مبهم یا پرمخاطره، رویکردهای اعتبارسنجی طراحی می‌کند.
5	استانداردهای اعتبارسنجی را تعیین می‌کند و شک‌گرایی سالم نسبت به هوش مصنوعی را آموزش می‌دهد.

D7. بازبینی / راستی‌آزمایی انسانی

اطمینان از حضور مناسب انسان در چرخه پیش از استفاده یا انتشار خروجی.

امتیاز	این‌گونه به نظر می‌رسد
0	خروجی هوش مصنوعی را مستقیم و بدون بازبینی منتشر می‌کند.
1	از لزوم بازبینی آگاه است اما ناپایدار عمل می‌کند.
2	خروجی هوش مصنوعی خود را پیش از استفاده کم‌مخاطره بازبینی می‌کند.
3	به‌طور پایدار، عمق بازبینی مناسب را بر اساس مخاطره کار اعمال می‌کند.
4	جریان‌های کاری بازبینی/تأیید و دروازه‌های تصویب را برای تیم‌ها طراحی می‌کند.
5	مالک سیاست انسان‌درچرخه است؛ پاسخگویی را در سراسر سازمان تضمین می‌کند.

D8. توان خودکارسازی

ساخت خودکارسازی‌های تکرارپذیر مبتنی بر هوش مصنوعی.

امتیاز	این‌گونه به نظر می‌رسد
0	هیچ‌چیز را با هوش مصنوعی خودکار نکرده است.

امتیاز	این‌گونه به نظر می‌رسد
1	می‌فهمد خودکارسازی ممکن است؛ اما چیزی نساخته است.
2	یک خودکارسازی ساده و تک‌مرحله‌ای را با کمک می‌سازد.
3	خودکارسازی‌های چندمرحله‌ای قابل‌اتکا را برای کار خود/تیم، مستقل می‌سازد.
4	خودکارسازی‌های عملیاتی با مدیریت خطا، تأیید و بازگشت (rollback) می‌سازد.
5	الگوها/سکوهای خودکارسازی را طراحی می‌کند که در میان تیم‌ها بازاستفاده می‌شوند.

D9. درک داده و RAG

استوار کردن هوش مصنوعی بر داده‌های مورد اعتماد (بازیابی، embedding، منابع، تازگی).

امتیاز	این‌گونه به نظر می‌رسد
0	بی‌اطلاع از اینکه هوش مصنوعی می‌تواند بر داده‌های مشخص استوار شود.
1	در سطح کلی می‌داند هوش مصنوعی می‌تواند از «اسناد شما» استفاده کند.
2	مفهوم بازیابی/grounding و اینکه چرا خطاها را کاهش می‌دهد درک می‌کند.
3	ابزارهای سبک RAG را درست به‌کار می‌برد و کیفیت/تازگی منبع را قضاوت می‌کند.
4	خط‌لوله‌های بازیابی، chunking و کنترل‌های کیفیت داده را طراحی می‌کند.
5	معماری و استانداردهای استوارسازی بر داده را در سطح سازمان تعیین می‌کند.

D10. درک هوش مصنوعی عامل‌محور (Agentic)

درک عامل‌های هوش مصنوعی خودمختار و چندمرحله‌ای که اقدام انجام می‌دهند.

امتیاز	این‌گونه به نظر می‌رسد
0	از عامل‌های هوش مصنوعی بی‌اطلاع است.
1	در سطح کلی می‌داند عامل‌ها می‌توانند اقدام کنند/از ابزار استفاده کنند.
2	حلقه‌های عامل، ابزارها و مخاطره افزوده ناشی از خودمختاری را درک می‌کند.
3	از ابزارهای عامل‌محور به‌صورت ایمن و در مرزهای تعیین‌شده استفاده می‌کند.
4	عامل‌هایی با مجوزهای محدود، حفاظ‌ها و پایش طراحی می‌کند.
5	سکوهای عامل بازاستفاده‌پذیر را با حفاظ‌های سراسر سازمانی معماری می‌کند.

D11. ایمنی و امنیت هوش مصنوعی

محافظت در برابر سوءاستفاده، نشت داده، prompt injection و خروجی ناایمن.

امتیاز	این‌گونه به نظر می‌رسد
0	از مخاطرات ایمنی/امنیت هوش مصنوعی بی‌اطلاع است.

امتیاز	این‌گونه به نظر می‌رسد
1	از وجود مخاطراتی مانند نشت داده و «اسرار را نچسبانید» آگاه است.
2	قواعد پایه استفاده ایمن را به‌طور پایدار رعایت می‌کند.
3	به‌طور مستقل از مخاطرات رایج (injection، نشت، اشتراک‌گذاری بیش از حد) پرهیز می‌کند و آن‌ها را گزارش می‌دهد.
4	تدابیر حفاظتی و جریان‌های کاری هوش مصنوعی آگاه به تهدید را طراحی می‌کند.
5	مالک وضعیت امنیتی هوش مصنوعی است؛ کنترل‌ها و واکنش را تعریف می‌کند.

D12. حاکمیت و انطباق

کار در چارچوب سیاست، تأییدها، ممیزی و الزامات مستندسازی.

امتیاز	این‌گونه به نظر می‌رسد
0	از هیچ سیاست یا الزام تأیید هوش مصنوعی بی‌اطلاع است.
1	می‌داند سیاست‌ها وجود دارند؛ از جزئیات مطمئن نیست.
2	سیاست تأیید/استفاده را با راهنمایی رعایت می‌کند.
3	به‌طور مستقل در چارچوب سیاست کار می‌کند و استفاده از هوش مصنوعی را مناسب مستند می‌کند.
4	جریان‌های کاری منطبق می‌سازد؛ شواهد را برای ممیزی/بازبینی مخاطره مدل آماده می‌کند.
5	حاکمیت، کنترل‌ها و سیاست هوش مصنوعی را برای سازمان شکل می‌دهد.

D13. آگاهی از مخاطره بانکی / صنعت تحت‌مقررات

درک تعهدات خاص یک بافت خدمات مالی تحت‌مقررات.

امتیاز	این‌گونه به نظر می‌رسد
0	هیچ آگاهی از تعهدات مقرراتی/داده‌مشتری پیرامون هوش مصنوعی ندارد.
1	می‌داند بانکداری قواعد سخت‌گیرانه داده/مقرراتی مرتبط با هوش مصنوعی دارد.
2	مخاطرات آشکار داده تحت‌مقررات (PII، داده مشتری) را پیش از اقدام شناسایی می‌کند.
3	به‌طور مستقل احتیاط صنعت تحت‌مقررات را در استفاده روزمره از هوش مصنوعی اعمال می‌کند.
4	مواجهه مقرراتی (انصاف، توضیح‌پذیری، سوابق) را در طراحی هوش مصنوعی پیش‌بینی می‌کند.
5	مالک راهبرد مخاطره مقرراتی برای هوش مصنوعی است؛ با ریسک/انطباق/تنظیم‌گران تعامل می‌کند.

D14. همکاری / اشتراک دانش

گسترش رویه ایمن و مؤثر هوش مصنوعی به دیگران.

امتیاز	این‌گونه به نظر می‌رسد
0	منزوی کار می‌کند؛ چیزی به اشتراک نمی‌گذارد.

امتیاز	این‌گونه به نظر می‌رسد
1	گاه‌به‌گاه نکات هوش مصنوعی را به‌صورت غیررسمی مطرح می‌کند.
2	هنگامی که از او خواسته شود، یک prompt/تکنیک مفید را به اشتراک می‌گذارد.
3	به‌طور منظم رویه‌های کارآمد را به اشتراک می‌گذارد و به همکاران کمک می‌کند.
4	برنامه‌های توانمندسازی (مستندات، جلسات، هنجارها) را اجرا می‌کند که یک تیم را ارتقا می‌دهد.
5	جامعه عمل (community of practice) هوش مصنوعی را در سراسر سازمان می‌سازد.

D15. اثر کسب‌وکاری

پیوند دادن استفاده از هوش مصنوعی به ارزش قابل‌اندازه‌گیری (زمان، هزینه، کیفیت، درآمد، کاهش مخاطره).

امتیاز	این‌گونه به نظر می‌رسد
0	هیچ پیوندی میان استفاده از هوش مصنوعی و پیامدها وجود ندارد.
1	منفعت را به‌صورت حکایتی ادعا می‌کند؛ بدون اندازه‌گیری.
2	دستاوردهای زمان/کیفیت را به‌صورت غیررسمی یادداشت می‌کند.
3	دستاوردهای عینی از استفاده خود از هوش مصنوعی را اندازه‌گیری و گزارش می‌کند.
4	اثر قابل‌اندازه‌گیری در سطح تیم را با سنجه‌های روشن پیش می‌برد.
5	ارزش هوش مصنوعی در سطح سازمانی را ارائه می‌دهد و مالک آن است؛ به راهبرد گره می‌زند.

D16. رهبری و راهبرد

تعیین جهت، اولویت‌بندی و پاسخگویی برای هوش مصنوعی.

امتیاز	این‌گونه به نظر می‌رسد
0	هیچ نقشی در جهت‌دهی هوش مصنوعی ندارد.
1	دیدگاه‌هایی درباره جهت هوش مصنوعی را به‌صورت غیررسمی بیان می‌کند.
2	هنگامی که مشارکت داده شود، در تصمیم‌های هوش مصنوعی سهم می‌گذارد.
3	پذیرش هوش مصنوعی را در تیم/حیطه خود رهبری می‌کند.
4	اولویت‌ها و استانداردهای هوش مصنوعی را در چند تیم تعیین می‌کند.
5	مالک راهبرد، حاکمیت و پاسخگویی هوش مصنوعی در سطح سازمان است.

13.3 ماتریس اصلی (مرجع سریع)

#	بُعد	1 آگاهی پایه	3 مجری مستقل	5 خبره / رهبر
D1	سواد هوش مصنوعی	می‌داند هوش مصنوعی مرتبط است	تناسب هوش مصنوعی را مستقل قضاوت می‌کند	مرجع سازمانی در هوش مصنوعی

#	بُعد	1 آگاهی پایه	3 مجری مستقل	5 خبره / رهبر
D2	سواد GenAI	می‌داند GenAI می‌تواند اشتباه کند	مدل/تنظیمات را برمی‌گزیند، خطا را پیش‌بینی می‌کند	استانداردهای استفاده از مدل را تعیین می‌کند
D3	مهندسی prompt	prompt ساده می‌نویسد	تا نتیجه قابل‌اتکا تکرار می‌کند	کتابخانه prompt می‌سازد، مربی‌گری می‌کند
D4	استفاده از ابزار هوش مصنوعی	یک ابزار را امتحان کرده	از ابزارها روزانه و روان استفاده می‌کند	ابزارها را ارزیابی/معرفی می‌کند
D5	کاربرد ویژه شغل	نام می‌برد هوش مصنوعی کجا کمک می‌کند	از هوش مصنوعی برای کار واقعی نقش استفاده می‌کند	آن حوزه را بازشکل می‌دهد
D6	تفکر انتقادی / اعتبارسنجی	می‌داند خروجی می‌تواند نادرست باشد	به‌عادت اعتبارسنجی می‌کند	استانداردهای اعتبارسنجی را تعیین می‌کند
D7	بازبینی / راستی‌آزمایی انسانی	از لزوم بازبینی آگاه است	بازبینی مبتنی بر مخاطره را به‌طور پایدار انجام می‌دهد	مالک سیاست انسان‌درچرخه است
D8	توان خودکارسازی	می‌داند خودکارسازی ممکن است	خودکارسازی‌های چندمرحله‌ای می‌سازد	سکوهای خودکارسازی طراحی می‌کند
D9	درک داده و RAG	می‌داند هوش مصنوعی می‌تواند از اسناد استفاده کند	از RAG استفاده می‌کند، منابع را قضاوت می‌کند	معماری grounding را تعیین می‌کند
D10	درک هوش مصنوعی عامل‌محور	می‌داند عامل‌ها می‌توانند اقدام کنند	عامل‌ها را در مرزها به‌کار می‌برد	سکوهای عامل را معماری می‌کند
D11	ایمنی و امنیت هوش مصنوعی	از مخاطره نشت آگاه است	از مخاطرات رایج پرهیز/گزارش می‌کند	مالک وضعیت امنیتی است
D12	حاکمیت و انطباق	می‌داند سیاست وجود دارد	در چارچوب سیاست کار می‌کند، مستند می‌کند	حاکمیت/سیاست را شکل می‌دهد
D13	مخاطره بانکی / تحت‌مقررات	می‌داند بانکداری سخت‌گیرانه است	احتیاط تحت‌مقررات را اعمال می‌کند	مالک راهبرد مقرراتی هوش مصنوعی است
D14	همکاری / اشتراک	نکات را غیررسمی مطرح می‌کند	رویه‌ها را به اشتراک می‌گذارد، به همکاران کمک می‌کند	جامعه عمل می‌سازد
D15	اثر کسب‌وکاری	منفعت حکایتی	دستاوردها را اندازه‌گیری/گزارش می‌کند	مالک ارزش سازمانی است
D16	رهبری و راهبرد	دیدگاه‌های غیررسمی	پذیرش تیمی را رهبری می‌کند	مالک راهبرد سازمانی است

13.4 شیوه تجمیع امتیاز ابعاد در نردبان ۱۰سطحی

۱۶ بُعد، قابلیت‌ها را توصیف می‌کنند؛ نردبان (L0–L10) سطح کلی فرد را توصیف می‌کند. از این روش دوقاعده‌ای برای تبدیل امتیازها به سطح نردبان استفاده کنید.

قاعده ۱ — ابعاد دروازه‌ای (کف ایمنی الزامی). از آنجا که این یک بافت خدمات مالی تحت‌مقررات است، چهار بُعد به‌عنوان دروازه عمل می‌کنند. هیچ فردی را نمی‌توان بالاتر از سطحی که امتیازهای دروازه‌ای‌اش اجازه می‌دهد قرار داد، صرف‌نظر از اینکه سایر امتیازهایش چقدر قوی باشند. ابعاد دروازه‌ای عبارت‌اند از:

- D6 تفکر انتقادی / اعتبارسنجی

- D7 بازبینی / راستی‌آزمایی انسانی
- D11 ایمنی و امنیت هوش مصنوعی
- D12 حاکمیت و انطباق
- D13 آگاهی از مخاطره بانکی / صنعت تحت‌مقررات

قاعده ۲ — ابعاد محرک (آنچه سطح را بالا می‌کشد). در محدوده سقفی که دروازه‌ها تعیین می‌کنند، سطح بر اساس عالی‌ترین قابلیتی تعیین می‌شود که فرد به‌طور پایدار و با شواهد نشان می‌دهد (prompting ← جریان‌های کاری ← توانمندسازی ← خودکارسازی ← طراحی راه‌حل ← ساخت سکو ← راهبرد).

جدول نگاشت

سطح نردبان	نام	پروفایل معمول ابعاد (با شواهد)	الزام دروازه
L0 / L1	بدون مواجهه	همه ابعاد در 0؛ بی‌اطلاع از ابزارها و مخاطرات داده.	هیچ.
L2	آگاهی نسبت به هوش مصنوعی	D1، D2، D12، D13 در 1؛ قواعد مدیریت داده را پیش از استفاده از ابزار درک می‌کند.	دروازه‌ها ≤ 1.
L3	کاربر پایه هوش مصنوعی	D4، D5 در 2–3؛ D6، D7 در 3 (در کارهای کم‌مخاطره راستی‌آزمایی می‌کند).	دروازه‌ها ≤ 2؛ D6، D7، D11 ≥ 2.
L4	کاربر مؤثر prompt	D3 در 4؛ D2، D5 در 3؛ نتایج قابل‌اتکا از طریق تکنیک سنجیده.	دروازه‌ها ≤ 2؛ D6، D7 ≥ 3.
L5	کاربر جریان‌کاری هوش مصنوعی	D5 در 4؛ D8 در 3؛ D15 در 3 (هوش مصنوعی را در کار چندمرحله‌ای جای می‌دهد و دستاوردها را اندازه می‌گیرد).	دروازه‌ها ≤ 3؛ D6، D7 ≥ 3.
L6	کاربر قدرتمند هوش مصنوعی	D14 در 4؛ D3، D6 در 4؛ هنجارهای ایمن را تعیین و دیگران را ارتقا می‌دهد.	دروازه‌ها ≤ 3؛ D11، D12، D13 ≥ 3.
L7	مشارکت‌کننده خودکارسازی هوش مصنوعی	D8 در 4 (خودکارسازی‌های عملیاتی با تأیید و بازگشت)؛ D11 در 4.	دروازه‌ها ≤ 3؛ D7، D11، D12 ≥ 4.
L8	طراح راه‌حل هوش مصنوعی	D9 در 4؛ D8 در 4؛ D12 در 4 (راه‌حل‌های سرتاسری که از اعتبارسنجی مخاطره مدل عبور می‌کنند).	دروازه‌ها ≤ 4؛ D12، D13 ≥ 4.
L9	سازنده عامل / سکوی هوش مصنوعی	D10 در 5؛ D8، D9 در 5؛ D11 در 4–5 (سکوهای عامل بازاستفاده‌پذیر با حفاظ‌های سراسر سازمانی).	دروازه‌ها ≤ 4؛ D11، D12 ≥ 4.
L10	رهبر راهبرد / حاکمیت هوش مصنوعی	D16 در 5؛ D12، D13 در 5؛ D15 در 5 (مالک راهبرد سازمانی، حاکمیت، مخاطره مقرراتی).	دروازه‌ها ≤ 4؛ D12، D13 = 5.

مثال‌های حل‌شده

- **سازنده قوی، انطباق ضعیف:** D3=4، D8=4، D10=4 اما D12=2 و D13=1. قابلیت محرک حاکی از L7–L8 است، اما دروازه‌های حاکمیت/مخاطره تحت‌مقررات، جایگاه را به **L3** محدود می‌کنند. اقدام: پیش از پیشروی، شکاف دروازه‌ای را پر کنید.
- **کاربر متوازن جریان‌کاری:** D3=4، D5=4، D8=3، D15=3، همه دروازه‌ها ≤ 3. به‌روشنی در **L5** قرار می‌گیرد.
- **رهبر توانمندسازی:** D12=4، D11=4، D6=4، D14=5، دروازه‌ها ≤ 4، اما بدون خودکارسازی عملیاتی (D8=3). تا زمانی که شواهد خودکارسازی (D8 ≥ 4) پشتیبان L7 شود، در سقف **L6** می‌ماند.

یادداشت‌هایی درباره نگاشت

- (استنباط‌شده از منابع عمومی) مدل دروازه‌ومحرک، رویه گسترده و منتشرشده چارچوب‌های شایستگی (مقیاس‌های مهارت، معیارهای ایمنی الزامی) را با تأکید بر کنترل صنعت تحت‌مقررات که در جای دیگر این گزارش توصیف شده، ترکیب می‌کند. این یک روش پیشنهادی است، نه استانداردی منتشرشده توسط یک فروشنده.
- آستانه‌های ابعاد در جدول نگاشت، **نقاط شروع پیشنهادی** هستند. آن‌ها را در برابر ارزیابی‌های واقعی سازمان خود واسنجی کنید و پیش از استفاده رسمی در تصمیم‌های ارتقا یا حقوق، تنظیم نمایید.
- فهرست دروازه‌ها را کوتاه و غیرقابل‌مذاکره نگه دارید. در یک بافت بانکی‌گونه، هزینه ارتقای کسی که خروجی هوش مصنوعی راستی‌آزمایی‌نشده یا نامنطبق منتشر می‌کند، به‌مراتب بالاتر از هزینه نگه‌داشتن او یک سطح پایین‌تر است.

---

14. فرایند ارزیابی مبتنی بر شواهد

ادعای شایستگی دقیقاً به اندازه شواهد پشتیبان آن اعتبار دارد. در یک محیط خدمات مالی تحت نظارت، عبارت «من با AI خوب کار می‌کنم» یک رتبه نیست؛ بلکه یک فرضیه آغازین است که باید در برابر رفتار قابل‌مشاهده، مصنوعات (artefacts) و نتایج آزموده شود. این بخش هفت نوع ارزیابی را تعریف می‌کند: هر کدام برای چه هدفی به کار می‌رود، چگونه امتیازدهی می‌شود، چه شواهدی باید مطالبه شود و چگونه می‌توان سوگیری را از نتیجه دور نگه داشت.

هر نوع ارزیابی به نردبان شایستگی AI (از L0/L1 تا L10) نگاشت می‌شود. هیچ روش واحدی به‌تنهایی کافی نیست: خودارزیابی ارزان اما غیرقابل‌اتکاست، وظایف عملی قابل‌اتکا اما محدودند و پنل‌های کالیبراسیون منصفانه اما پرهزینه‌اند. طراحی پیشنهادی از روش‌های ارزان‌تر برای غربالگری سطوح پایین‌تر (L0/L1–L3) استفاده می‌کند و روش‌های سنگین‌تر (وظیفه عملی، نمونه‌کار/portfolio و پنل) را برای سطوح با ریسک بالاتر (L5 به بالا) نگه می‌دارد؛ جایی که هزینه یک رتبه‌بندی نادرست — مثلاً فردی فاقد صلاحیت که یک اتوماسیون تولیدی (production) را منتشر می‌کند یا راهکاری مرتبط با ریسک مدل طراحی می‌کند — بیشترین مقدار است.

اصل طراحی ارزیابی (استنتاج‌شده از منابع عمومی)

> این اصل کلی که هیچ منبع رتبه‌بندی واحدی قابل‌اتکا نیست، که رتبه‌ها باید میان ارزیابان کالیبره شوند و اینکه خودارزیابی معمولاً متورم است، از پژوهش‌ها و تجربه‌های پرشمار منتشرشده درباره ارزیابی عملکرد و سنجش شایستگی برگرفته شده است (برای نمونه، Kruger & Dunning درباره تورم خودارزیابی؛ ادبیات مربوط به اعتبار مصاحبه ساختاریافته و نمونه‌کار عملی؛ و رویه جلسات کالیبراسیون که در HR رایج است). نگاشت مشخص به نردبان AI که در ادامه می‌آید، یک توصیه ترکیب‌شده (synthesized) است، نه یک رویه تثبیت‌شده شرکتی.

A. خودارزیابی (Self-Assessment)

حوزه	جزئیات
هدف	غربالگری سریع و کم‌هزینه؛ آشکارسازی دیدگاه خودِ فرد نسبت به سطحش و شناسایی شکاف‌ها برای بررسی‌های بعدی. برای تفکیک اولیه (triage) سطوح L0/L1 تا L3 و برای برانگیختن تأمل پیش از گفت‌وگو با مدیر سودمند است.
روش امتیازدهی	فرد خود را بر اساس توصیفگرهای نردبان (L0 تا L10) در هر شایستگی ارزیابی می‌کند. این ارزیابی به‌منزله‌ی یک ادعا تلقی می‌شود، نه یک امتیاز. شکاف = (خودارزیابی − ارزیابی تأییدشده)؛ شکاف‌های مثبت بزرگ نشانه‌ی اعتمادبه‌نفس بیش از حد هستند و باید بررسی شوند. هرگز به‌تنهایی بالاتر از L3 استفاده نمی‌شود.
شواهد لازم	روبریک نردبانِ تکمیل‌شده همراه با یک مثال عینی برای هر سطح ادعاشده (اینکه چه کاری انجام داده‌اند، با کدام ابزار، با چه داده‌ای، و چه گام تأییدی را طی کرده‌اند). ادعاهای بدون مثال به‌عنوان فاقد شواهد امتیازدهی می‌شوند.

حوزه	جزئیات
خطرات سوگیری	اعتمادبه‌نفس بیش از حد (Dunning-Kruger)؛ کم‌بینی در کارکنان محتاط یا تازه‌کار؛ سوگیری مطلوبیت اجتماعی (ادعای «استفاده‌ی ایمن» چون این انتظار وجود دارد)؛ دستکاری به‌منظور ترفیع.
اقدامات کاهنده	الزام به ارائه‌ی یک مثال عینی برای هر ادعا؛ مقایسه‌ی متقابل با ارزیابی مدیر و دست‌کم یک روش عینی؛ کالیبره کردن شکاف خودارزیابی در برابر ارزیابی تأییدشده در سطح کل تیم، به‌جای تفسیر هر امتیاز به‌صورت منفرد و مجزا.

نمونه پرسش‌ها - «یک وظیفه در ماه گذشته نام ببرید که در آن از یک ابزار GenAI استفاده کرده‌اید. چه داده‌ای را وارد کردید و پیش از استفاده، خروجی را چگونه بررسی کردید؟» - «کدام سطح از نردبان به‌بهترین شکل استفاده‌ی معمول شما را توصیف می‌کند و چرا نه سطح بالاتر از آن؟» - «زمانی را توصیف کنید که تصمیم گرفتید برای یک وظیفه از هوش مصنوعی استفاده نکنید. خطر آن چه بود؟»

B. ارزیابی مدیر

حوزه	جزئیات
هدف	ارائه دیدگاه سرپرست بر پایه مشاهده روزمره؛ منبع اصلی راستی‌آزمایی برای L3 تا L5؛ ورودی (و نه رأی تعیین‌کننده) برای L6 به بالا.
روش امتیازدهی	مدیر فرد را بر اساس همان معیار سنجش نردبان امتیاز می‌دهد و به نمونه‌های مشخصِ مشاهده‌شده استناد می‌کند. امتیازدهی باید به نمونه‌های تاریخ‌دار ارجاع دهد، نه برداشت‌های کلی. اختلاف‌نظر با خودارزیابی ثبت می‌شود، نه آنکه با میانگین‌گیری از میان برداشته شود.
شواهد لازم	دست‌کم دو مشاهده مشخص و تاریخ‌دار برای هر سطح ادعاشده (برای نمونه: «خروجی یک مدل را پیش از ارسال به مشتری در برابر منبع راستی‌آزمایی کرد، ۱۲ مه»)؛ برای L4 به بالا، شواهدی از تکنیک تکرارپذیر، نه یک رخداد یک‌باره.
خطرهای سوگیری	اثر هاله/شاخ (halo/horns)؛ سوگیری تازگی (recency bias)؛ تبعیض و طرفداری؛ سوگیری اغماض یا گرایش به میانه؛ مدیرانی که مهارت AIای را امتیاز می‌دهند که خود توان داوری درباره آن را ندارند.
اقدامات کاهنده	الزام به ارائه نمونه‌های تاریخ‌دار؛ آموزش مدیران درباره نردبان پیش از امتیازدهی؛ ارجاع هر سطحی که مدیر شخصاً توان داوری درباره آن را ندارد (معمولاً L7 به بالا) به بازبینی فنی؛ نمایش توزیع امتیازهای مدیر در فرایند کالیبراسیون برای شناسایی اغماض/سختگیری.

نمونه پرسش‌ها (برای مدیر) - «پیچیده‌ترین وظیفه با کمک AI که شخصاً دیده‌اید این فرد به‌درستی انجام دهد چه بوده است؟» - «آیا دیده‌اید که خروجی‌ها را راستی‌آزمایی و داده‌ها را تحت فشار زمانی به‌شکلی ایمن مدیریت کند، یا تنها در شرایط ایده‌آل؟» - «در کجا توانمندی دیگران را ارتقا داده است، در برابر آنکه صرفاً کار خود را تحویل دهد؟»

C. تکلیف عملی (نمونه‌کار)

حوزه	جزئیات
هدف	مشاهده مستقیم توانمندی در یک تکلیف نمونه و کنترل‌شده. بالاترین اعتبار را برای سطوح L3 تا L5 دارد (استفاده ایمن، prompting، گردش‌کار)؛ روش اصلی برای تأیید مهارت عملی به جای صرفِ ادعای کلامی.
روش امتیازدهی	امتیازدهی بر اساس یک rubric از پیش تدوین‌شده با معیارهای قبول/نسبی/ردود برای هر رفتار (مثلاً مدیریت داده، راستی‌آزمایی، تکرار و پالایش prompt، کیفیت نتیجه). تکلیف و rubric یکسان برای همه داوطلبان یک سطح. امتیازدهی کور (blind) در صورت امکان.

حوزه	جزئیات
شواهد لازم	خودِ بروندادِ تکلیف (promptهای به‌کاررفته، خروجی‌ها، رد پایِ راستی‌آزمایی) به‌علاوه rubric تکمیل‌شده. برای سطوح مربوط به گردش‌کار، یک نمونهٔ عملیِ چندمرحله‌ای با مقایسهٔ سنجیده‌شدهٔ قبل/بعد.
خطرهای سوگیری	نماینده نبودن تکلیف از کار واقعی؛ نوسان دشواری میان داوطلبان؛ سخاوتمندی داور؛ برتری داوطلبانی که با ابزار خاصِ به‌کاررفته در آزمون آشنا هستند.
اقدام کاهنده	استفاده از سناریوهای واقع‌گرایانه و مرتبط با نقش با دادهٔ مصنوعی (هرگز دادهٔ زندهٔ مشتری)؛ تثبیت یک rubric برای هر سطح؛ به‌کارگیری دو داورِ مستقل و آشتی‌دادن نتایج آن‌ها؛ مجاز دانستن انتخاب ابزار در جایی که موضوعِ سطح، توانمندی است نه یک محصول خاص.

نمونه تکالیف - L3: «این خط‌مشی داخلی را با یک ابزار GenAI خلاصه کنید. نشان دهید چه چیزی را وارد کرده‌اید و چگونه از دقت خلاصه اطمینان حاصل کرده‌اید. هر چیزی را که وارد نمی‌کردید مشخص کنید.» (آزمونِ استفادهٔ روتین و ایمن + راستی‌آزمایی.) - L4: «برای این سناریو یک پیش‌نویس اولیهٔ قابل‌استفاده از یک ایمیلِ مشتری تهیه کنید. تکرارهای prompt خود و دلیل کارآمدبودنِ prompt نهایی را نشان دهید.» (آزمونِ تکنیک هدفمندِ prompting.) - L5: «یک گردش‌کارِ مبتنی بر کمکِ هوش مصنوعی برای این تکلیف گزارش‌دهیِ تکرارشونده طراحی کنید و زمان صرفه‌جویی‌شده نسبت به امروز را برآورد کنید.» (آزمونِ تعبیهٔ چندمرحله‌ای + اندازه‌گیری.)

D. شواهد مبتنی بر نمونه‌کار (Portfolio)

میدان	جزئیات
هدف	ارزیابی توانمندی پایدار و واقعی در طول زمان با استفاده از محصولات کاری واقعی. روش اصلی برای L5 تا L9، جایی که توانمندی در گردش‌کارهای عرضه‌شده، اتوماسیون‌ها، پلتفرم‌ها و پذیرش آن‌ها توسط دیگران نمایان می‌شود.
روش امتیازدهی	بررسی مصنوعات (artefacts) ارائه‌شده در برابر توصیفگرهای نردبان، با بازبینی اصالت تألیف، دروازه‌های تأیید/بازگشت (approval/rollback) و اثر اندازه‌گیری‌شده. هر مورد بر اساس آنچه واقعاً نشان می‌دهد امتیاز می‌گیرد، نه بر پایه بلندپروازی. اصالت تألیف و سهم مشخص فرد راستی‌آزمایی می‌شود.
شواهد لازم	پیوندها/خروجی‌های مصنوعات واقعی: کتابخانه‌های prompt، گردش‌کارهای مستندشده همراه با سنجه‌ها، اتوماسیون‌های تولیدی (production) با کنترل‌های تأیید و بازگشت (L7)، طرح‌های راهکار که اعتبارسنجی ریسک مدل (model-risk) را پشت سر گذاشته‌اند (L8)، پلتفرم‌های agent قابل‌استفاده مجدد همراه با حفاظ‌ها (guardrails) (L9)، به‌علاوه شواهد استفاده دیگران از آن‌ها.
ریسک‌های سوگیری	خطای انتساب (به نام خود زدن کار تیمی)؛ گزینش گزینشی بهترین کارها؛ ارائه صیقل‌خورده‌ای که ضعف محتوا را پنهان می‌کند؛ سوگیری بازماندگی (پنهان ماندن شکست‌ها).
اقدامات کاهنده	راستی‌آزمایی سهم فرد با همکاران نام‌برده و تاریخچه نسخه‌ها؛ مطالبه نمونه‌ای نماینده و نه صرفاً نقاط برجسته؛ درخواست دست‌کم یک مورد که شکست خورده و درس‌های آموخته‌شده از آن؛ بررسی وجود مصنوعات حاکمیتی (تأییدها، بازگشت، تأییدیه اعتبارسنجی) و نه فقط خودِ ساخت.

E. مصاحبه / گفت‌وگو (ساختاریافته)

حوزه	جزئیات
هدف	کندوکاو در استدلال، قضاوت و آگاهی از ریسک پشت کار انجام‌شده — «چرایی»، نه فقط «چه چیزی». این مرحله برای L2 (درک نحوه مدیریت داده)، L6 (هنجارسازی) و L8 تا L10 (طراحی، حاکمیت و قضاوت تنظیم‌گرانه) حیاتی است.

حوزه	جزئیات
روش امتیازدهی	مصاحبه ساختاریافته: پرسش‌های اصلی یکسان برای هر سطح، با مقیاس امتیازدهی لنگرگذاری‌شده برای هر پرسش (تعریف‌شده بر مبنای رفتار، ۱ تا ۵)، که در برابر شاخص‌های پاسخ از پیش توافق‌شده امتیازدهی می‌شود. این یک گفت‌وگوی آزاد نیست.
شواهد لازم	امتیازهای ثبت‌شده برای هر پرسش همراه با یادداشت درباره آنچه نامزد واقعاً گفته است؛ سناریوهای مشخصی که از او خواسته شده درباره آن‌ها استدلال کند.
ریسک‌های سوگیری	سوگیری شباهت/قرابت؛ اثر «سخنور روان» (پاداش‌دادن به روانی بیان به‌جای محتوا)؛ پرسش‌های جهت‌دار؛ لنگرگذاری بر اساس برداشت اولیه.
اقدامات کاهنده	استفاده از مجموعه پرسش ثابت و امتیازدهی لنگرگذاری‌شده؛ امتیازدهی مستقل دو مصاحبه‌گر پیش از گفت‌وگو با یکدیگر؛ تمرکز پرسش‌ها بر رفتار واقعی گذشته و استدلال سناریومحور، نه فرضیات درباره عملکرد ایده‌آل؛ تفکیک «خوب توضیح می‌دهد» از «خوب عمل کرده است».

نمونه پرسش‌ها - L2: «چه نوع داده‌هایی هرگز نباید در اینجا وارد یک ابزار GenAI بیرونی شود، و چرا؟ اگر مطمئن نبودید چه می‌کنید؟» - L6: «چگونه برای تیمی که برخی اعضای آن در سطح L3 و برخی در سطح L1 هستند، هنجارهای ایمن استفاده از AI تعیین می‌کنید؟» - L8: «گام‌به‌گام توضیح دهید چگونه یک راهکار AI را طراحی می‌کنید تا اعتبارسنجی ریسک مدل (model-risk validation) را پشت سر بگذارد. از همان ابتدا چه کنترل‌هایی در آن تعبیه می‌کنید؟» - L10: «چگونه میان پذیرش AI در سطح سازمان و ریسک تنظیم‌گرانه و اعتباری تعادل برقرار می‌کنید؟»

F. بازبینی همتا / فنی (Peer / Technical Review)

میدان	جزئیات
هدف	به‌کارگیری داوری خبرگان و همتایان در جایی که مدیران توانایی ارزیابی عمق فنی را ندارند. این روش برای سطوح فنی (L7–L9: خودکارسازی، طراحی راهکار، ساخت پلتفرم) و برای راستی‌آزمایی تأثیرگذاری ادعاشده بر همتایان (L6) ضروری است.
روش امتیازدهی	بازبین یا بازبینان واجد صلاحیت، مصنوعات فنی و انتخاب‌های طراحی را در برابر توصیف‌گرهای نردبان و استانداردهای مهندسی/حاکمیتی ارزیابی می‌کنند؛ همتایان نیز هنجارسازی و ارتقای سطح را گواهی می‌دهند. امتیازدهی در برابر معیارهای فنی تعریف‌شده انجام می‌شود و صلاحیت بازبین ثبت می‌گردد.
شواهد موردنیاز	تأییدیه‌ی بازبین با ارجاع به مصنوعات مشخص (کد، دروازه‌های خودکارسازی، معماری، نرده‌های محافظ)؛ گواهی همتایان مبنی بر آموزش‌دیدن، رفع انسداد، یا پذیرش هنجارها/ابزارهای آن فرد.
خطرهای سوگیری	بده‌بستان متقابل/زدوبند («مرا بالا ببر، من هم تو را بالا می‌برم»)؛ سوگیری درون‌گروهی؛ بازبینی که در واقع برای سطح موردبازبینی صلاحیت ندارد؛ تعارض منافع درون یک تیم واحد.
اقدامات کاهنده	الزام به اینکه بازبینان در سطحی برابر یا بالاتر از سطح موردارزیابی صلاحیت داشته باشند؛ استفاده از دست‌کم یک بازبین خارج از تیم نزدیک نامزد؛ ثبت استدلال بازبین؛ وزن‌دهی به گواهی همتا به‌عنوان شواهد پشتیبان و نه شواهد یگانه.

نمونه پرسش‌ها (برای بازبین) - «آیا دروازه‌های تأیید و بازگشت (rollback) این خودکارسازی، استاندارد تولیدی ما را برآورده می‌کنند؟ به محل آن استناد کنید.» - «آیا این پلتفرم agent به‌راستی با نرده‌های محافظ قابل استفاده‌ی مجدد است، یا یک ساخت یک‌بارمصرف وابسته به سازنده‌اش؟» - «فارغ از آنچه نامزد ادعا می‌کند، این مصنوع چه چیزی را درباره‌ی توانایی او اثبات می‌کند؟»

G. پنل کالیبراسیون (Calibration Panel)

میدان	جزئیات
هدف	بررسی نهایی انصاف و انسجام. تجمیع شواهد چندمنبعی در یک رتبه‌بندی واحد و قابل دفاع و اطمینان از اینکه یک سطح مشخص در همه تیم‌ها معنای یکسانی دارد. برای همه رتبه‌بندی‌های L6 به بالا و برای رفع هر گونه اختلاف میان خودارزیابی و ارزیابی مدیر الزامی است.
روش امتیازدهی	پنلی میان‌وظیفه‌ای (cross-functional) بسته شواهد گردآوری‌شده (خودارزیابی، مدیر، وظیفه، نمونه‌کار، مصاحبه، بازبینی فنی) را بررسی می‌کند و بر سر یک سطح نهایی به توافق می‌رسد. تصمیم‌ها و نظرات مخالف مستند می‌شوند. پنل بررسی می‌کند که رتبه‌بندی‌ها به یک استاندارد ثابت پایبند باشند، نه به رتبه‌بندی نسبی درون یک تیم.
شواهد لازم	بسته شواهد کامل برای هر نامزد؛ توجیه مکتوب برای سطح مورد توافق؛ ثبت هر گونه تعدیل نسبت به رتبه‌بندی پیشنهادی مدیر و دلیل آن.
ریسک‌های سوگیری	تفکر گروهی/اثر صدای مسلط؛ فشار توزیع رتبه‌بندی (منحنی‌های اجباری که سطوح واقعی را تحریف می‌کنند)؛ بده‌بستان (horse-trading)؛ سوگیری منزلتی به نفع نامزدهای ارشد.
اقدامات کاهشی	پیش‌مطالعه‌های مستقل پیش از بحث؛ چرخش/محدودسازی اینکه چه کسی نخست صحبت می‌کند؛ رئیس جلسه چالش مبتنی بر شواهد را اعمال می‌کند؛ ارزیابی بر اساس توصیفگرهای مطلق نردبان (ladder)، نه سهمیه؛ مستندسازی نظرات مخالف؛ ممیزی تصمیم‌ها برای الگوهای اثر نامطلوب در میان گروه‌های جمعیتی.

انواع رایج شواهد

نوع شواهد	چه چیزی را نشان می‌دهد	قوی‌ترین برای
نمونهٔ عملی مشخص (prompt + خروجی + رد راستی‌آزمایی)	رفتار عملی واقعی و عادت‌های ایمنی	L3–L5
مشاهدهٔ تاریخ‌دار مدیر/همکار	عملکرد واقعی در شرایط عادی	L3–L6
نمونه‌کار همراه با نمرهٔ rubric	توانمندی کنترل‌شده و قابل‌مقایسه	L3–L5
گردش‌کار مستندشده با معیارهای پیش/پس	اثر سنجش‌شده بر بهره‌وری	L5
اتوماسیون عملیاتی با دروازه‌های تأیید/بازگشت	توانایی ساخت ایمن تحت حاکمیت	L7
طراحی راهکار با تأییدیهٔ اعتبارسنجی ریسک مدل	طراحی سرتاسری که آستانهٔ مقرراتی را برآورده می‌کند	L8
پلتفرم عامل قابل‌استفادهٔ مجدد همراه با guardrail و شواهد پذیرش	توانمندی مهندسی در سطح سازمان	L9
مصنوعات راهبرد/حاکمیت سازمانی (سیاست، چارچوب ریسک)	مالکیت راهبردی و مقرراتی	L10
گواهی همکاران دربارهٔ آموزش/ارتقا	تأثیر بر توانمندی دیگران	L6
تأییدیهٔ بازبین با ارجاع به مصنوعات مشخص	عمق فنی تأییدشده به‌صورت مستقل	L7–L9

روش‌های ارزیابی در یک نگاه

روش	هزینه	پایایی به‌تنهایی	بهترین تناسب با نردبان	کاربرد به‌عنوان
A. خودارزیابی	پایین	پایین	غربالگری L0/L1–L3	فقط غربالگری / ادعا
B. ارزیابی مدیر	پایین	متوسط	L3–L5	راستی‌آزمایی اصلی (سطوح پایین‌تر)
C. تکلیف عملی	متوسط	بالا	L3–L5	تأیید مهارت عملی
D. کارنامه (Portfolio)	متوسط تا بالا	بالا	L5–L9	تأیید توانمندی واقعی و پایدار
E. مصاحبهٔ ساختاریافته	متوسط	متوسط تا بالا	L2، L6، L8–L10	کاوش قضاوت و استدلال
F. بازبینی همکار/فنی	متوسط	بالا (فنی)	L6–L9	راستی‌آزمایی عمق و تأثیر
G. پنل کالیبراسیون	بالا	بالا (انصاف)	L6 به بالا و همهٔ اختلاف‌نظرها	تطبیق و جمع‌بندی نهایی

خلاصهٔ توصیه‌شدهٔ شواهد بر اساس سطح (توصیهٔ استنتاجی / ترکیب‌شده)

سطح نردبان	حداقل شواهد برای تأیید
L0/L1 بدون مواجهه	خودارزیابی + تأیید مدیر
L2 آگاهی از AI	خودارزیابی + مصاحبهٔ ساختاریافتهٔ کوتاه دربارهٔ قواعد مدیریت داده
L3 کاربر پایهٔ AI	تکلیف عملی (استفادهٔ روزمرهٔ ایمن + راستی‌آزمایی) + مشاهدهٔ مدیر
L4 کاربر اثربخش prompt	تکلیف عملی که promptنویسی هدفمند را نشان دهد + نمونهٔ عملی
L5 کاربر AI در سطح گردش‌کار	گردش‌کار در کارنامه با سنجش پیش/پس + تأییدیهٔ مدیر
L6 کاربر حرفه‌ای AI	گواهی همکاران دربارهٔ ارتقا + ورودی مدیر + پنل کالیبراسیون
L7 مشارکت‌کنندهٔ اتوماسیون AI	اتوماسیون در کارنامه با دروازه‌های تأیید/بازگشت + بازبینی فنی
L8 طراح راهکار AI	طراحی راهکار با تأییدیهٔ اعتبارسنجی ریسک مدل + بازبینی فنی + پنل
L9 سازندهٔ عامل/پلتفرم AI	پلتفرم قابل‌استفادهٔ مجدد همراه با guardrail + شواهد پذیرش + بازبینی فنی + پنل
L10 رهبر راهبرد/حاکمیت AI	مصنوعات راهبرد/حاکمیت سازمانی + بازبینی اجرایی/پنل

15. آزمون‌های عملی بر حسب نقش

x.15 تحلیل‌گر میز خدمت (Service Desk Analyst)

لنگرگاه نردبان: این آزمون سطوح L3 تا L4 (کاربر پایه‌ی هوش مصنوعی تا کاربر مؤثر در طراحی پرامپت) را هدف می‌گیرد و دارای نشانگرهای کششی برای L5 (کاربر هوش مصنوعی در سطح جریان کاری) است. فرض بر این است که داوطلب سطح L2 را گذرانده باشد (مبانی GenAI و قواعد مدیریت داده را می‌داند). تحلیل‌گران میز خدمت در یک بانک تحت نظارت با تیکت‌های زنده‌ی رخداد سروکار دارند که حاوی داده‌های مشتریان و کارکنان است؛ بنابراین قضاوت ایمن در مدیریت داده به همان اندازه‌ی دقت فنی وزن‌دهی می‌شود.

سناریو

شما یک تحلیل‌گر میز خدمت در میز خدمت IT یک بانک خرد (retail) هستید. یک تیکت رخداد با اولویت P2 به شما ارجاع داده شده است. یکی از کارکنان شعبه گزارش می‌دهد که برنامه‌ی دسکتاپ داخلی «خدمات حساب» (Account Servicing) به‌طور مکرر در بارگذاری صفحه‌ی جست‌وجوی مشتری ناموفق است و از امروز صبح خطای عمومی «Connection timed out» را نمایش می‌دهد. حدود ۴۰ کاربر شعبه در سه مکان مختلف تحت تأثیر قرار گرفته‌اند، بنابراین تحویلداران (tellers) ناچار به استفاده از یک فرایند دستی کندتر شده‌اند.

تیکتی که به شما تحویل داده شده شامل موارد زیر است:

- شرح کاربر نهایی و تصاویر صفحه (screenshots).
- یک قطعه از لاگ خطای کپی‌شده‌ی برنامه.
- نام کامل کارمند گزارش‌دهنده، شناسه‌ی کارمندی، کد شعبه و شماره‌ی تلفن مستقیم او.
- یک تصویر صفحه که نام یک مشتری واقعی، شماره‌ی کامل حساب و بخشی از نشانی او را در پنجره‌ی برنامه نمایان می‌کند.
- قطعه‌ای از connection string موجود در لاگ که شامل نام میزبان (hostname) یک پایگاه‌داده‌ی داخلی و نام کاربری یک حساب سرویس (service-account) است.

شما به یک دستیار GenAI تأییدشده و میزبانی‌شده در سطح سازمانی دسترسی دارید (ابزار مجاز بانک شما که تحت توافق‌نامه‌ی پردازش داده‌ی سازمانی پوشش داده شده است). شما به هیچ ابزار هوش مصنوعی مصرفی/عمومی دسترسی ندارید. وظیفه‌ی شما استفاده از هوش مصنوعی برای تسریع غربالگری اولیه (triage) است، ضمن آن‌که داده‌های حساس را از هر پرامپتی که اکیداً ضروری نیست و در چارچوب سیاست‌ها قرار ندارد، دور نگه دارید.

وظیفه

1. تصمیم بگیرید چه محتوایی از تیکت را می‌توان و چه محتوایی را نمی‌توان در دستیار هوش مصنوعی تأییدشده قرار داد، و هر چیزی را که نباید ارسال شود حذف (redact) یا جایگزین کنید.
2. از هوش مصنوعی برای خلاصه‌سازی نقص، پیشنهاد محتمل‌ترین علت(های) ریشه‌ای، و پیشنهاد یک اقدام اصلاحی ایمن به‌عنوان پاسخ نخست و یک مسیر روشن ارجاع (escalation) به تیم پشتیبانی برنامه استفاده کنید.
3. یک به‌روزرسانی ایمن برای مشتری و یک به‌روزرسانی داخلی برای تیکت تهیه کنید.
4. به‌صراحت بیان کنید کدام خروجی‌های هوش مصنوعی را پیش از اقدام راستی‌آزمایی کرده‌اید یا باید به یک انسان/تیم متخصص واگذار کنید.

کاربرد مجاز هوش مصنوعی

مجاز	غیرمجاز
استفاده از دستیار GenAI سازمانی تأییدشده برای خلاصه‌سازی، پیش‌نویس به‌روزرسانی‌ها و طوفان فکری درباره‌ی علل محتمل.	درج PII خام مشتری (نام، شماره‌ی کامل حساب، نشانی) در هر پرامپتی.
درج متن لاگ پاک‌سازی‌شده/حذف‌شده و یک شرح عمومی از مسئله.	درج اعتبارنامه‌های حساب سرویس، نام‌های میزبان داخلی کامل، یا connection string.
درخواست از هوش مصنوعی برای پیشنهاد پرسش‌های تشخیصی و یک چک‌لیست غربالگری.	استفاده از هر ابزار هوش مصنوعی مصرفی/تأییدنشده، حتی برای بررسی‌های «سریع».
استفاده از هوش مصنوعی برای پیش‌نویس به‌روزرسانی‌های رو به مشتری و داخلی جهت بازبینی انسانی.	اجازه‌دادن به هوش مصنوعی برای اجرای خودکار اقدام اصلاحی، بستن تیکت، یا انجام هرگونه اقدام سیستمی.

خروجی مورد انتظار

یک بسته‌ی غربالگری کوتاه شامل:

1. **تصمیم مدیریت داده** — فهرستی از مواردی که پیش از هرگونه استفاده از هوش مصنوعی حذف، پنهان (mask) یا جایگزین شده‌اند، همراه با یک خط استدلال برای هر مورد (مثلاً شماره‌ی حساب به ۴ رقم آخر ماسک شد، نام میزبان با `<DB_HOST>` جایگزین شد، کاربر حساب سرویس با `<SVC_USER>` جایگزین شد، شماره‌ی تلفن شخصی کارمند کنار گذاشته شد).
2. **خلاصه‌ی نقص با کمک هوش مصنوعی** — ۳ تا ۵ جمله به زبان ساده.
3. **علت(های) ریشه‌ای محتمل** — رتبه‌بندی‌شده، همراه با یک یادداشت اطمینان (مثلاً «شبکه/اتصال در برابر در دسترس‌بودن سرویس DB در برابر اعتبارنامه/انقضا») و این‌که چه شواهدی هر مورد را تأیید می‌کند.
4. **اقدام اصلاحی ایمن به‌عنوان پاسخ نخست** — تنها گام‌های کم‌خطر و برگشت‌پذیری که میز خدمت مجاز به انجام آن‌هاست (مثلاً تأیید دامنه‌ی تأثیر، بررسی‌های پایه‌ای اتصال، راه‌اندازی مجدد برنامه‌ی کلاینت روی یک دستگاه)، به‌علاوه‌ی ارجاع/واگذاری صحیح به پشتیبانی برنامه یا پایگاه‌داده برای هر چیز دارای دسترسی ویژه.
5. **دو به‌روزرسانی تیکت** — یکی مناسب مشتری/شعبه، و دیگری یک یادداشت فنی داخلی.
6. **فهرست نیازمند راستی‌آزمایی انسانی** — مواردی که نباید صرفاً بر اساس خروجی هوش مصنوعی روی آن‌ها اقدام شود (مثلاً هر تغییری که پایگاه‌داده، حساب سرویس یا دیوار آتش (firewall) را لمس کند).

معیارهای ارزیابی

#	معیار	شاخص عملکرد مطلوب
1	غربالگری صحیح داده	همه‌ی عناصر حساس را شناسایی می‌کند و پیش از هرگونه استفاده از هوش مصنوعی آن‌ها را حذف/جایگزین می‌کند؛ هیچ مورد ممنوعی ارسال نمی‌شود.
2	انتخاب ابزار	تنها از ابزار سازمانی تأییدشده استفاده می‌کند؛ هرگز «صرفاً برای آزمایش» به سراغ ابزار مصرفی نمی‌رود.
3	کیفیت خلاصه	دقیق، به زبان ساده، بدون جزئیات لاگ ساختگی یا کدهای خطای جعلی.
4	استدلال علت ریشه‌ای	فرضیه‌های رتبه‌بندی‌شده همراه با اطمینان و شواهد لازم برای تأیید ارائه می‌دهد — نه یک حدس واحد که به‌عنوان واقعیت بیان شود.
5	دامنه‌ی اقدام اصلاحی ایمن	تنها اقدامات مجاز و برگشت‌پذیر میز خدمت را پیشنهاد می‌دهد؛ اقدامات دارای دسترسی ویژه را به‌درستی ارجاع می‌دهد.
6	راستی‌آزمایی انسانی	پیشنهادهای هوش مصنوعی را به‌روشنی از اقداماتی که نیازمند تأیید انسان/متخصص هستند جدا می‌کند.
7	ارتباطات	به‌روزرسانی مشتری اطمینان‌بخش و بدون اصطلاحات فنی است؛ یادداشت داخلی دقیق و قابل‌اقدام است.
8	تکنیک پرامپت (کششی، L4)	از پرامپت‌نویسی ساختاریافته (نقش، زمینه، محدودیت‌ها، قالب خروجی) استفاده می‌کند و به‌جای پذیرفتن نخستین خروجی، آن را تکرار و اصلاح می‌کند.

سنجه‌ی امتیازدهی (0-5)

امتیاز	سطح	توضیح
0	ناایمن / بدون توانایی	PII یا اعتبارنامه‌ی خام را در هوش مصنوعی درج می‌کند، از ابزار تأییدنشده استفاده می‌کند، یا اصلاً نمی‌تواند از ابزار استفاده کند. متناظر با L0/L1.

امتیاز	سطح	توضیح
1	تنها آگاه	قواعد داده را در تئوری می‌داند اما به‌طور ناهمسان به کار می‌بندد؛ خروجی عمدتاً بدون بازکاری غیرقابل‌استفاده است. متناظر با L2.
2	پایه، ناقص	بیشتر داده‌های حساس را حذف می‌کند و یک خلاصه‌ی قابل‌استفاده تولید می‌کند، اما یک عنصر حساس را از قلم می‌اندازد یا خروجی هوش مصنوعی را بدون راستی‌آزمایی می‌پذیرد. در آستانه‌ی L3.
3	کاربر روزمره‌ی ایمن	غربالگری داده‌ی کاملاً صحیح، خلاصه‌ی دقیق، اقدام اصلاحی منطقی، و یک فهرست روشن نیازمند راستی‌آزمایی انسانی. L3 محکم.
4	کاربر مؤثر در پرامپت	تمام موارد بالا به‌علاوه‌ی پرامپت‌نویسی هدفمند (ساختاریافته، تکرارشده)، علل ریشه‌ای رتبه‌بندی‌شده همراه با اطمینان، و ارجاع دقیق و هدف‌گذاری‌شده. L4 محکم.
5	جریان‌کاری‌محور	تمام L4 به‌علاوه‌ی مصنوعات قابل‌استفاده‌ی مجدد (یک چک‌لیست حذف داده یا قالب پرامپت غربالگری قابل‌استفاده‌ی مجدد) و سنجه‌ای از بهره‌ی زمانی/کیفی، که نشانگر آمادگی به‌سوی L5 است.

پرچم‌های قرمز

- درج هرگونه PII مشتری، شماره‌ی حساب، نشانی، اعتبارنامه، یا نام میزبان داخلی در یک پرامپت هوش مصنوعی — سقف خودکار امتیاز در 0 صرف‌نظر از سایر کیفیت‌ها.
- استفاده از یک ابزار هوش مصنوعی مصرفی/تأییدنشده برای هر بخشی از وظیفه.
- تلقی یک اقدام اصلاحی پیشنهادی هوش مصنوعی (به‌ویژه تغییرات DB، حساب سرویس، یا دیوار آتش) به‌عنوان تأییدشده و اقدام بر روی آن بدون راستی‌آزمایی انسانی.
- ارائه‌ی کدهای خطا، خطوط لاگ، یا علل ریشه‌ای ساخته‌ی هوش مصنوعی به‌عنوان واقعیت تأییدشده (عبور توهم (hallucination) به درون تیکت یا به مشتری).
- بستن تیکت یا علامت‌گذاری آن به‌عنوان حل‌شده تنها بر اساس خروجی هوش مصنوعی.
- قراردادن شماره‌ی تلفن شخصی یا شناسه‌های کامل کارمند گزارش‌دهنده در یک پرامپت، در حالی که نیازی به آن نیست.

بررسی‌های ایمنی/انطباق

بررسی	شرط قبولی
کمینه‌سازی داده	تنها کمینه‌ی متن پاک‌سازی‌شده‌ی لازم برای غربالگری به هوش مصنوعی ارسال می‌شود؛ PII و اسرار ماسک یا جایگزین شده‌اند.
ابزار تأییدشده	تنها از دستیار GenAI مجاز و تحت پوشش قراردادی بانک استفاده می‌شود.
مدیریت اسرار	هیچ اعتبارنامه، connection string، یا نام میزبان داخلی در معرض هوش مصنوعی قرار نمی‌گیرد؛ هرگونه افشای مشکوک اعتبارنامه در تیکت طبق سیاست برای چرخش (rotation) علامت‌گذاری می‌شود.
انسان در حلقه (Human-in-the-loop)	هیچ اقدام دارای دسترسی ویژه یا تغییردهنده‌ی سیستم بر اساس توصیه‌ی هوش مصنوعی بدون راستی‌آزمایی و تأیید متخصص انجام نمی‌شود.
قابلیت ممیزی	پرامپت(ها) و خروجی‌های هوش مصنوعی به‌کاررفته قابل بازبینی هستند (در نشست لاگ‌شده‌ی ابزار تأییدشده نگه داشته می‌شوند)؛ هیچ ابزار سایه‌ای (shadow tooling) وجود ندارد.
محرمانگی مشتری	به‌روزرسانی رو به مشتری هیچ نام سیستم داخلی، داده‌ی حساب، یا گمانه‌زنی علت ریشه‌ای که بتواند گمراه‌کننده باشد را فاش نمی‌کند.

بررسی	شرط قبولی
یکپارچگی رخداد	تغییرات وضعیت تیکت و حل‌وفصل‌ها یک تصمیم انسانی باقی می‌مانند؛ هوش مصنوعی صرفاً پشتیبان تصمیم است.

> **یادداشت (استنباط‌شده از منابع عمومی):** سطوح نردبان و انتظارات حذف داده/ارجاع در این‌جا از بهترین‌رویه‌های پرکاربرد و منتشرشده‌ی استفاده از GenAI در میز خدمت و صنایع تحت نظارت ترکیب و استخراج شده‌اند (کمینه‌سازی داده، استفاده‌ی صرف از ابزار تأییدشده، انسان در حلقه برای اقدامات دارای دسترسی ویژه). این موارد بازتاب هیچ نردبان HR داخلی خاص یا سیاست اختصاصی هیچ شرکتی نیستند.

x.15 مهندس DevOps

سناریو

یک خط لوله‌ی شبانه‌ی CI/CD برای **سرویس مجوزدهی پرداخت‌ها** (Payments Authorization Service) بانک، طی مرحله‌ی استقرار (deploy) در محیط پیش‌ازتولید (pre-production) با شکست مواجه شده است. پنجره‌ی انتشار (release) دو ساعت دیگر بسته می‌شود، و فردا یک انجماد تغییرات (change freeze) الزام‌شده توسط نهاد ناظر آغاز می‌گردد. یک لاگ کنسول Jenkins با ۱٬۲۰۰ خط به شما تحویل داده شده است. لاگ شامل ترکیبی از خطاهای ساخت (build)، یک مهاجرت پایگاه‌داده‌ی ناموفق، و یک قطع‌شدن اتصال (timeout) به یک سرویس پایین‌دستی است.

نکته‌ی حیاتی این‌که لاگ خام حاوی موارد حساس جاسازی‌شده است: یک connection string پایگاه‌داده همراه با اعتبارنامه‌ها، یک طرح نام میزبان/IP داخلی، یک PAN مشتری (شماره‌ی کارت) که در یک دستور SQL در سطح debug ظاهر شده، یک OAuth client secret که توسط یک افزونه‌ی به‌اشتباه‌پیکربندی‌شده چاپ شده، و توپولوژی سرویس داخلی. شما می‌توانید از یک دستیار GenAI سازمانی تأییدشده برای تسریع غربالگری استفاده کنید، اما سیاست مدیریت داده‌ی بانک، ارسال اسرار تولیدی (production secrets)، داده‌ی مشتری (در دامنه‌ی PII/PAN/PCI)، یا جزئیات زیرساخت داخلی به ابزارهای هوش مصنوعی خارجی یا تأییدنشده را ممنوع می‌کند.

وظیفه

1. تمام داده‌های حساس در لاگ را **پیش از** هرگونه تعامل با هوش مصنوعی شناسایی و حذف کنید.
2. از دستیار هوش مصنوعی تأییدشده برای خلاصه‌سازی نقص، پیشنهاد یک فرضیه‌ی علت ریشه‌ای رتبه‌بندی‌شده، و پیشنهاد گام‌های اصلاحی ایمن استفاده کنید.
3. به‌صراحت پیشنهادهای تولیدشده توسط هوش مصنوعی را از واقعیت‌هایی که خودتان راستی‌آزمایی کرده‌اید جدا کنید.
4. بیان کنید پیش از اعمال هر تغییری در پیش‌ازتولید چه چیزی باید توسط انسان راستی‌آزمایی شود.
5. با توجه به انجماد تغییرات قریب‌الوقوع، یک توصیه‌ی برو/نرو (go/no-go) ارائه دهید.

کاربرد مجاز هوش مصنوعی

مجاز	غیرمجاز
خلاصه‌سازی یک لاگ **پیش‌ازحذف‌شده** (pre-redacted)	درج لاگ‌های خام حاوی اسرار، PAN/PII، یا توپولوژی داخلی
تولید فرضیه‌های نامزد علت ریشه‌ای	اجازه‌دادن به هوش مصنوعی برای اعمال خودکار اصلاحات در هر محیطی
پیش‌نویس گام‌های اصلاحی برای بازبینی انسانی	تلقی خروجی هوش مصنوعی به‌عنوان علت ریشه‌ای راستی‌آزمایی‌شده
پیش‌نویس الگوهای حذف/regex	استفاده از یک ابزار هوش مصنوعی تأییدنشده/مصرفی برای داده‌ی بانک

مجاز	غیرمجاز
توضیح امضاهای خطای ناآشنا	ارسال اعتبارنامه‌ها یا connection string برای دریافت یک «اصلاح سریع»

خروجی مورد انتظار

- یک **فهرست حذف (redaction list)**: هر مورد حساس یافت‌شده، دسته‌ی آن (اعتبارنامه / PII / PAN-PCI / توپولوژی داخلی)، و نحوه‌ی ماسک‌کردن آن.
- یک **خلاصه‌ی نقص** مختصر (چه چیزی شکست خورد، در کجای خط لوله، نشانه‌های مشاهده‌شده).
- یک **فرضیه‌ی علت ریشه‌ای رتبه‌بندی‌شده** (از محتمل‌ترین تا کم‌احتمال‌ترین) همراه با شواهد پشت هر یک.
- **گام‌های اصلاحی** که به‌روشنی به‌عنوان پیشنهاد هوش مصنوعی در برابر راستی‌آزمایی‌شده توسط خودتان برچسب خورده‌اند.
- یک **چک‌لیست راستی‌آزمایی انسانی** (مثلاً تأیید این‌که مهاجرت خودتوان (idempotent)/برگشت‌پذیر است، اعتبارسنجی سلامت سرویس پایین‌دستی، بررسی مسیر بازگشت (rollback)).
- یک تصمیم **برو/نرو** همراه با استدلالی که انجماد تغییرات را در نظر می‌گیرد.

معیارهای ارزیابی

حوزه	شاخص عملکرد مطلوب
مدیریت داده	هر مورد حساس را پیش از استفاده از هوش مصنوعی می‌یابد و حذف می‌کند؛ PAN/PII/اسرار/توپولوژی را به‌درستی دسته‌بندی می‌کند
انتخاب ابزار	تنها از ابزار هوش مصنوعی سازمانی تأییدشده استفاده می‌کند؛ این را به‌صراحت بیان می‌کند
کیفیت غربالگری	خلاصه‌ی نقص دقیق است و به مرحله‌ی صحیح خط لوله نگاشت می‌شود
استدلال	فرضیه‌های علت ریشه‌ای رتبه‌بندی‌شده و مبتنی بر شواهدند، نه حدس‌زده‌شده
انضباط راستی‌آزمایی	خروجی هوش مصنوعی را به‌روشنی از واقعیت راستی‌آزمایی‌شده جدا می‌کند؛ بررسی‌های انسانی مشخص را فهرست می‌کند
اقدام اصلاحی ایمن	تغییرات برگشت‌پذیر و دروازه‌دار (gated) همراه با مسیر بازگشت پیشنهاد می‌دهد؛ بدون ویرایش مستقیم تولید
قضاوت	استدلال برو/نرو ریسک را در برابر انجماد و پنجره‌ی انتشار صادقانه می‌سنجد

سنجه‌ی امتیازدهی (0-5)

امتیاز	توضیح
0	لاگ خام (شامل اسرار/PAN) را در هوش مصنوعی درج می‌کند، یا از ابزار تأییدنشده استفاده می‌کند؛ خروجی هوش مصنوعی را به‌عنوان واقعیت تلقی می‌کند.
1	مقداری استفاده از هوش مصنوعی، اما بیشتر داده‌های حساس را از قلم می‌اندازد؛ بدون حذف؛ علت ریشه‌ای یک حدس است.
2	برخی موارد را حذف می‌کند اما دست‌کم یک سرّ/PII را نشت می‌دهد؛ خلاصه تا حدی صحیح است؛ بدون درنظرگرفتن بازگشت.
3	همه‌ی اسرار آشکار را حذف می‌کند؛ خلاصه‌ی نقص معقول؛ پیشنهادهای هوش مصنوعی به‌کار می‌رود اما راستی‌آزمایی ضعیف است.
4	حذف کامل با دسته‌بندی صحیح؛ فرضیه‌های رتبه‌بندی‌شده‌ی مبتنی بر شواهد؛ چک‌لیست روشن راستی‌آزمایی انسانی؛ مسیر بازگشت یادداشت شده.

امتیاز	توضیح
5	تمام L4 به‌علاوه‌ی: تفکیک پیشنهاد هوش مصنوعی در برابر راستی‌آزمایی‌شده در سراسر کار؛ برو/نروی منطقی گره‌خورده به انجماد/ریسک؛ پیشنهاد یک گام حذف قابل‌استفاده‌ی مجدد برای خط لوله‌ی تیم (سطح انتظار را برای دیگران بالا می‌برد).

پرچم‌های قرمز

- ارسال هرگونه سرّ، اعتبارنامه، PII، PAN، یا نام میزبان/IP داخلی به هوش مصنوعی بدون حذف.
- استفاده از یک حساب هوش مصنوعی شخصی یا مصرفی برای داده‌ی بانک.
- پذیرفتن علت ریشه‌ای هوش مصنوعی بدون راستی‌آزمایی مستقل.
- اجازه‌دادن به هوش مصنوعی برای تولیدواعمال مستقیم تغییرات در پیش‌ازتولید یا تولید.
- غیرفعال‌کردن مهاجرت یا بررسی پایین‌دستی «برای آن‌که خط لوله قبول شود».
- نادیده‌گرفتن انجماد تغییرات یا تحریف ریسک برای تحمیل یک انتشار.
- سخت‌کدنویسی (hardcoding) یا بازتاب سرّ بازیابی‌شده در پیکربندی به‌جای چرخش آن.

بررسی‌های ایمنی/انطباق

- [ ] تمام داده‌های حساس پیش از هرگونه تعامل با هوش مصنوعی حذف و دسته‌بندی شده‌اند.
- [ ] تنها ابزار هوش مصنوعی سازمانی تأییدشده‌ی بانک استفاده شده؛ طبق سیاست لاگ شده است.
- [ ] اعتبارنامه/OAuth secret افشاشده برای **چرخش (rotation)** علامت‌گذاری شده، نه استفاده‌ی مجدد.
- [ ] افشای PAN در لاگ‌ها به‌عنوان یک **نقض احتمالی لاگ‌نویسی PCI-DSS** به مالک مربوطه گزارش شده است.
- [ ] هیچ تغییری بدون بازبینی انسانی و یک مسیر بازگشت تأییدشده در پیش‌ازتولید/تولید اعمال نشده است.
- [ ] سیاست انجماد تغییرات رعایت شده است؛ هر استثنا از فرایند مستندشده‌ی تغییر اضطراری پیروی می‌کند.
- [ ] تصمیمات اتخاذشده با کمک هوش مصنوعی برای قابلیت ممیزی ثبت شده‌اند (چه چیزی پرسیده شد، چه چیزی راستی‌آزمایی شد).

(استنباط‌شده از منابع عمومی: به سطوح نردبان L4-L5 — کاربر مؤثر در پرامپت تا کاربر هوش مصنوعی در سطح جریان کاری — نگاشت می‌شود، با نشانه‌گذاری رفتار L6 هنگامی که داوطلب خط لوله‌ی تیم را با یک گام حذف قابل‌استفاده‌ی مجدد مستحکم می‌کند.)

x.15 مهندس SRE

جایگاه نردبان (باند هدف): از L3 (کاربر پایه‌ی هوش مصنوعی) تا L5 (کاربر هوش مصنوعی در سطح جریان کاری). بیشتر کارهای SRE رو به تولید در این تمرین در سطح L5 سقف می‌خورد؛ رفتارهای L7+ (اصلاح خودکار، پلتفرم‌های عامل (agent)) به‌عنوان نشانگرهای کششی یادداشت شده‌اند اما برای قبولی **الزامی نیستند**.

سناریو

شما SRE کشیک پلتفرم **درگاه پرداخت** (Payments Gateway) بانک هستید. ساعت ۰۲:۱۴، PagerDuty هشدار می‌دهد: pipeline شبانه‌ی `payments-settlement` در Jenkins در مرحله‌ی `deploy-to-prod` با شکست مواجه شده و اکنون فرآیند تسویه‌ی دسته‌ای پایین‌دستی متوقف شده است. SLA تعیین‌شده برای تکمیل تسویه ساعت 06:00 است. شما به این موارد دسترسی دارید:

- یک log کنسول Jenkins با ۴،۰۰۰ خط (build #1183).
- دسترسی به runbook استقرار و یک دستیار GenAI غیرتولیدی که از سوی بانک تأیید شده است (tenant سازمانی، با شرط قراردادی "no-training"، و فقط در شبکه‌ی داخلی).
- یک چت‌بات GenAI **عمومی** در سطح مصرف‌کننده روی گوشی شخصی‌تان (که برای داده‌های بانکی تأیید نشده است).

این log در میان خروجی عادی، شامل این موارد است: یک stack trace، یک خطای rollout در Kubernetes، یک connection string پایگاه‌داده با اعتبارنامه‌های جاسازی‌شده، سه شماره PAN مشتری (شماره‌های اصلی حساب) که توسط یک debug logger با پیکربندی نادرست چاپ شده‌اند، یک طرح نام‌گذاری hostname/IP داخلی، و ایمیل یک کارمند در فیلد نویسنده‌ی commit.

شما تصمیم می‌گیرید برای تسریع triage از AI استفاده کنید. باید تصمیم بگیرید **چه چیزی** را paste کنید، **کجا**، و **چه چیزی را نمی‌توانید برای تأیید به AI اعتماد کنید**.

وظیفه

1. بخش موردنظر از لاگ را برای استفاده از کمک AI آماده کنید: داده‌های حساس را **پیش از** خروج از ترمینال خود شناسایی و پنهان (redact) کنید.
2. از دستیار AI **تأییدشده** برای موارد زیر استفاده کنید: خلاصه‌سازی خطا، ارائه یک فرضیه‌ی رتبه‌بندی‌شده درباره‌ی ریشه‌ی مشکل، و پیشنهاد گام‌های ایمن برای رفع آن.
3. گام‌های **پیشنهادشده توسط AI** را به‌صورت صریح از گام‌هایی که **حتماً باید توسط انسان راستی‌آزمایی شوند** جدا کنید.
4. پیش از هرگونه تغییر در محیط production، برنامه‌ی بازگشت/توقف (rollback/abort) و گیت تأیید انسانی را مشخص کنید.
5. تصمیم بگیرید که آیا اساساً می‌توان از چت‌بات عمومی استفاده کرد یا خیر، و دلیل آن را توجیه کنید.

کاربردهای مجاز AI

اقدام	مجاز است؟	شرط
استفاده از دستیار سازمانی تأییدشده برای خلاصه‌سازی/تحلیل ریشه‌ای	بله	تنها پس از پاک‌سازی داده‌ها (redaction)؛ خروجی باید با runbook راستی‌آزمایی شود
چسباندن لاگ خام در یک chatbot مصرفی عمومی	خیر	حاوی داده‌های تحت مقررات و داخلی است؛ پردازنده‌ای تأییدشده نیست
چسباندن لاگ **پاک‌سازی‌شده** در دستیار تأییدشده	بله	پاک‌سازی ابتدا توسط انسان تأیید شود
درخواست از AI برای پیش‌نویس دستورهای ترمیم (remediation)	بله	به‌عنوان پیش‌نویس تلقی می‌شود؛ هرگز بدون بازبینی روی محیط prod اجرا نمی‌شود
اجازه به AI برای اجرای مستقیم اقدامات ترمیمی	خیر (برای این تمرین)	نیازمند کنترل تغییر (change-control) و دروازهٔ تأیید انسانی است

خروجی مورد انتظار

داوطلب باید موارد زیر را تولید کند:

1. **فهرست پاکسازی (Redaction list)** — هر عنصر حساسِ شناسایی‌شده و نحوه برخورد با آن (حذف / ماسک‌گذاری / توکن‌سازی).
2. **prompt(های) هوش مصنوعی** به‌کاررفته، که تنها محتوای پاکسازی‌شده را نمایش می‌دهند.
3. **خلاصه خرابی** (۲ تا ۴ جمله، به زبان ساده).
4. **فرضیه‌های ریشه‌ای رتبه‌بندی‌شده** (از محتمل‌ترین تا کم‌احتمال‌ترین، همراه با شواهد لاگ برای هر کدام).

5. **برنامه رفع مشکل** که به دو ستون تقسیم شده است: پیشنهادشده توسط هوش مصنوعی در برابر نیازمند تأیید انسانی.
6. **برنامه بازگشت/توقف (Rollback/abort)** و **دروازه تأیید انسانیِ** مشخص‌شده (چه کسی پیش از تغییر در محیط production تأیید نهایی می‌دهد).
7. **تصمیم درباره ابزار** — اینکه کدام ابزار هوش مصنوعی استفاده شد و چرا چت‌بات عمومی رد شد.

معیارهای ارزیابی

#	معیار	«کیفیت مطلوب» چگونه است
1	مدیریت داده‌های حساس	هر هفت قلم داده حساس شناسایی شده‌اند؛ شماره کارت‌ها (PAN) و اطلاعات احراز هویت حذف شده‌اند (نه صرفاً masking)، و انتخاب ابزار توجیه شده است
2	انتخاب درست ابزار	از tenant تأییدشده استفاده می‌کند؛ به‌صراحت chatbot عمومی را با ذکر دلیل مرتبط با مدیریت داده رد می‌کند
3	کیفیت تحلیل ریشه‌ای	فرضیه‌ها اولویت‌بندی شده و به شواهد مشخص از لاگ‌ها متصل‌اند، نه حدس‌های کلی
4	انضباط در تأیید انسانی	به‌روشنی مشخص می‌کند چه مواردی را نمی‌توان به تأیید AI سپرد (مثلاً وضعیت prod، علت ریشه‌ای واقعی)
5	اصلاح ایمن	هیچ اجرای مستقیمی روی prod انجام نمی‌شود؛ دستورها به‌عنوان پیش‌نویس و پشت دروازه تغییر/تأیید (change/approval gate) در نظر گرفته می‌شوند
6	آمادگی برای rollback	پیش از هر تغییر، مسیر مشخصی برای توقف/بازگردانی (abort/rollback) بیان شده است
7	ارتباطات	خروجی برای رویداد آماده است: مختصر، قابل مرور سریع، و قابل تحویل تمیز به مسئول بعدی

سنجه نمره‌دهی (0-5)

نمره	سطح	توضیح
0	ناایمن	لاگ خام (شامل PAN یا گواهی‌های دسترسی) را در هر AI کپی می‌کند، یا دستورهای AI را مستقیماً روی محیط prod اجرا می‌کند. مردودی خودکار.
1	پایین‌تر از L3	از AI استفاده می‌کند اما بیشتر داده‌های حساس را نادیده می‌گیرد؛ خروجی AI را به‌عنوان حقیقت قطعی می‌پذیرد؛ هیچ برنامه‌ای برای بازگشت (rollback) ندارد.
2	در حال نزدیک‌شدن به L3	برخی موارد را پنهان‌سازی (redact) می‌کند و از ابزار تأییدشده استفاده می‌کند، اما پنهان‌سازی ناقص است (مثلاً PANها یا گواهی‌های دسترسی را از قلم می‌اندازد) یا تفکیک بازبینیِ انسانی را انجام نمی‌دهد.
3	L3 — کاربر پایه AI	پنهان‌سازی کامل، ابزار تأییدشده، خلاصه‌ای منطقی، تحلیل ریشه‌ای پایه، و اعلام وجود مسیر rollback. ایمن اما فاقد تحلیل عمیق.
4	L4 — کاربر اثربخش پرامپت	پرامپت‌نویسی هدفمند (ساختاریافته، با تعیین نقش، درخواست فرضیه‌های اولویت‌بندی‌شده به‌همراه شواهد)؛ تفکیک شفاف میان کار AI و کار انسانی؛ rollback مشخص و دروازه تأیید (approval gate).

نمره	سطح	توضیح
5	L5 — کاربر گردش‌کار AI	تمام موارد L4، به‌علاوه ادغام AI در گردش‌کار رخداد (مثلاً گام پاک‌سازیِ قابل‌استفاده مجدد، قالب پرامپت، و سنجش زمان صرفه‌جویی‌شده در triage) و پیشنهاد یک حفاظ (guardrail) برای جلوگیری از تکرار ثبت PANها در لاگ‌های دیباگ.

نشانه‌های هشدار

- لاگ (یا هر بخشی که حاوی PAN، اطلاعات احراز هویت یا IPهای داخلی است) را در چت‌بات **عمومی** یا در هر AI **پیش از** پاک‌سازی داده‌های حساس می‌چسباند.
- یک PAN را به‌صورت ناقص پنهان می‌کند (مثلاً ۴ رقم آخر به‌علاوه BIN را باقی می‌گذارد) و آن را امن می‌داند، بدون اینکه تأیید کند با استاندارد توکن‌سازی بانک مطابقت دارد.
- حدس AI درباره علت ریشه‌ای را تأییدشده تلقی می‌کند و بدون بررسی محیط prod یا runbook به سراغ رفع مشکل می‌رود.
- اجازه می‌دهد AI دستورها را علیه محیط production **هم تولید و هم اجرا** کند، بدون وجود یک دروازه تأیید انسانی.
- هیچ برنامه‌ای برای rollback یا توقف ندارد؛ رویکرد «همین‌طور به جلو رفع می‌کنیم» بدون هیچ گزینه جایگزین.
- اساساً متوجه نمی‌شود که PANهای مشتریان در لاگ ثبت شده‌اند — تنها روی stack trace تمرکز می‌کند.
- در یک نکته نظارتی/مقرراتی بیش از حد به AI اعتماد می‌کند (مثلاً از AI می‌پرسد «آیا این مطابق با PCI است؟» و پاسخ را معتبر و قطعی می‌پندارد).

بررسی‌های ایمنی/انطباق

بررسی	الزام
پاک‌سازی داده‌های تحت نظارت	شماره‌های کارت (PAN) و اعتبارنامه‌ها پیش از هرگونه استفاده از AI حذف شوند؛ بدون استثنا. مدیریت PAN باید با اصول PCI-DSS هم‌راستا باشد — داوطلب باید تصمیم‌گیری درباره انطباق را به یک انسان/مالک کنترل واگذار کند، نه به AI. (استنتاج‌شده از منابع عمومی: PCI-DSS ذخیره/انتقال PAN را محدود می‌کند.)
فقط پردازشگر تأییدشده	محتوای مجاورِ حساس (پاک‌سازی‌شده) تنها به محیط سازمانی تأییدشده بانک با شرایط عدم‌آموزش (no-training) ارسال شود؛ هرگز به چت‌بات عمومی.
کنترل تغییر	هر تغییر تولیدی (production) از دروازه مدیریت تغییر/تأیید موجود عبور کند؛ خروجی AI صرفاً مشورتی است.
پاسخگویی انسانی	یک انسانِ مشخص اقدام تولیدی را تأیید می‌کند و مالک پیامد آن است؛ AI تأییدکننده نیست.
قابلیت ممیزی	پرامپت، ورودی پاک‌سازی‌شده و خروجی AI همراه با تیکت رخداد ثبت شود تا بعداً قابل بازبینی باشد.
بهداشت رخداد	نقص ثبت PAN در لاگ به‌عنوان یک پیگیری جداگانه مطرح شود (بهداشت لاگ‌برداری/ترمیم نشت داده)، نه آنکه بی‌صدا نادیده گرفته شود.

یادداشت (تفکیک بهترین‌روش از توصیه): گردش‌کار «پاک‌سازی پیش از AI» و دروازه تأیید انسانی، بهترین‌روش‌های جاافتاده مدیریت رخداد و مدیریت داده هستند. لایه‌بندی ابزارهای خاص (محیط تأییدشده در برابر چت‌بات عمومی) و دروازه تأیید نام‌بُرده، **توصیه‌های ترکیب‌شده** برای این تمرین‌اند؛ پیش از استفاده، آن‌ها را به مدیریت تغییر واقعی و فهرست ابزارهای تأییدشده سازمان خود نگاشت کنید.

x.15 مهندس پلتفرم (Platform Engineer)

سناریو

شما یک مهندس پلتفرم (Platform Engineer) در تیم Shared Services یک بانک خرده‌فروشی تحت نظارت هستید. خطلوله شبانه CI/CD برای سرویس اصلی مجوزدهی پرداخت (`pay-auth-svc`) در مرحله استقرار (deploy) با شکست مواجه شد. تیم شما برای تسریع ترياژ به دستیارهای AI (یک LLM سازمانی تأییدشده) تکیه می‌کند، اما لاگ کنسول Jenkins بزرگ است و ترکیبی از خروجی ساخت (build)، جزئیات زیرساخت و داده‌های عملیاتی را در بر دارد. این شکست پنجره انتشار صبحگاهی را مسدود کرده است و یک SRE در ۳۰ دقیقه آینده خلاصه‌ای از علت ریشه‌ای می‌خواهد.

لاگ خام ساخت Jenkins به شما داده شده است. این لاگ در میان خروجی عادی ساخت، شامل موارد زیر است: - مقدار یک secret در Kubernetes که توسط یک گام دیباگ `kubectl` با پیکربندی نادرست بازتاب (echo) شده است (یک رمز عبور پایگاه‌داده کدگذاری‌شده با base64). - یک رشته اتصال (connection string) پایگاه‌داده تولیدی شامل host، port و نام کاربری حساب سرویس (service account). - نام میزبان‌های داخلی و بازه‌های IP خصوصی برای VPC پرداخت. - یک PAN مشتری (شماره حساب اصلی) که از یک تست یکپارچه‌سازی ناموفق با استفاده از داده تست توکنیزه‌نشده به یک stack trace نشت کرده است. - stack traceهای استاندارد، خطاهای ناسازگاری نسخه Helm chart و یک readiness probe ناموفق.

وظیفه

1. از دستیار AI سازمانی تأییدشده برای کمک به ترياژ شکست استفاده کنید.
2. پیش از ارسال هرگونه محتوای لاگ به AI، تمام داده‌های حساس را شناسایی و پاک‌سازی کنید و مستند کنید چه چیزی را و چرا حذف کرده‌اید.
3. یک خلاصه علت ریشه‌ای، فهرست رتبه‌بندی‌شده‌ای از علل محتمل و یک طرح ترمیم ایمن تولید کنید.
4. نتیجه‌گیری‌های پیشنهادی AI را به‌صراحت از آنچه انسان باید پیش از اقدام تأیید کند، تفکیک کنید.
5. مشخص کنید کدام گام‌های ترمیم به تأیید تغییر و دروازه‌های بازگشت (rollback) نیاز دارند.

استفاده مجاز از AI

مجاز	غیرمجاز
استفاده از LLM سازمانی تأییدشده برای خلاصه‌سازی گزیده‌های پاک‌سازی‌شده لاگ، فرضیه‌سازی علل ریشه‌ای و پیش‌نویس گام‌های ترمیم.	چسباندن لاگ‌های خام و پاک‌سازی‌نشده در هر ابزار AI.
استفاده از AI برای تولید الگوها/regexهای پاک‌سازی و یک چک‌لیست بهداشت‌سازی.	ارسال secretها، PANها، رشته‌های اتصال یا جزئیات شبکه داخلی به AI.
استفاده از AI برای پیش‌نویس (نه اجرای) دستورات `kubectl` /Helm جهت بازبینی انسانی.	استفاده از یک ابزار AI تأییدنشده/مصرف‌کننده برای هر داده بانکی.
استفاده از AI برای پیش‌نویس تیکت تغییر و طرح بازگشت.	اجازه دادن به AI برای اعمال خودکار تغییرات روی تولید بدون تأیید و دروازه‌های بازگشت.

خروجی مورد انتظار

1. یک سابقه بهداشت‌سازی: جدولی از هر مورد حساس یافت‌شده، دسته آن و اقدام پاک‌سازی.
2. یک خلاصه علت ریشه‌ای یک‌پاراگرافی به زبان ساده.
3. فهرست رتبه‌بندی‌شده‌ای از علل محتمل (از محتمل‌ترین به کم‌احتمال‌ترین) همراه با شواهد پشتیبان از لاگ پاک‌سازی‌شده.
4. یک طرح ترمیم تفکیک‌شده به: (الف) گام‌های ایمن برای انجام، (ب) گام‌های نیازمند تأیید تغییر، (ج) طرح بازگشت.
5. یک بخش به‌وضوح برچسب‌خورده «باید توسط انسان تأیید شود».

6. یک یادداشت درباره اینکه کدام ابزار AI استفاده شد و تأیید اینکه تنها محتوای پاک‌سازی‌شده به اشتراک گذاشته شده است.

معیارهای ارزیابی

معیار	نمونه خوب چگونه است
تشخیص داده‌های حساس	هر پنج دسته را می‌یابد (secret، رشته اتصال، IP، PAN/نام میزبان داخلی، حساب سرویس) پیش از هرگونه استفاده از AI.
انضباط پاک‌سازی	secretها را با جایگزین (placeholder) عوض می‌کند؛ هرگز مقادیر خام را نمی‌چسباند؛ هر حذف را مستند می‌کند.
کیفیت علت ریشه‌ای	ناسازگاری نسخه Helm / شکست readiness probe را به‌درستی بر نویز اولویت می‌دهد.
انسان‌درحلقه (Human-in-the-loop)	فرضیه‌های AI را به‌وضوح از حقایق تأییدشده تفکیک می‌کند؛ آنچه نیازمند تأیید انسانی است را علامت‌گذاری می‌کند.
کنترل تغییر	مشخص می‌کند کدام گام‌ها به تأیید نیاز دارند و یک بازگشت عینی تعریف می‌کند.
حاکمیت ابزار	تأیید می‌کند که تنها LLM سازمانی تأییدشده، آن هم فقط روی داده پاک‌سازی‌شده، استفاده شده است.

سنجه نمره‌دهی (0-5)

نمره	توضیح
0	لاگ خام (یا هر secret/PAN) را در AI می‌چسباند، یا از ابزار تأییدنشده استفاده می‌کند. مردودی خودکار بر اساس ایمنی.
1	مقداری تریاژ انجام شده، اما داده حساس را از قلم می‌اندازد یا محتوای پاک‌سازی‌نشده جزئی را به AI می‌فرستد.
2	بیشتر موارد حساس را پاک‌سازی می‌کند اما یک دسته را از قلم می‌اندازد (مثلاً PAN یا حساب سرویس)؛ علت ریشه‌ای ضعیف است یا خروجی AI به‌عنوان حقیقت پذیرفته می‌شود.
3	تمام داده‌های حساس را پاک‌سازی می‌کند؛ یک خلاصه علت ریشه‌ای قابل‌استفاده تولید می‌کند؛ مقداری تفکیک میان AI و تأییدشده توسط انسان دارد، اما کنترل تغییر یا بازگشت مبهم است.
4	تمام داده‌های حساس پاک‌سازی و مستند شده؛ علت ریشه‌ای درست؛ تفکیک شفاف AI در برابر انسان؛ دروازه‌های تأیید و بازگشت را با خلأهای جزئی شناسایی می‌کند.
5	تمام موارد بالا، به‌علاوه: یک رفع پایدار پیشنهاد می‌دهد (مثلاً توقف بازتاب secretها در CI، الزام داده تست توکنیزه‌شده، افزودن پاک‌سازی لاگ در خط‌لوله)؛ بهبود قابل‌اندازه‌گیری در زمان تریاژ؛ یک تیکت تغییر آماده ثبت همراه با بازگشت.

پرچم‌های قرمز

- ارسال هرگونه secret خام، PAN، رشته اتصال یا نام میزبان/IP داخلی به یک ابزار AI.
- استفاده از یک حساب AI شخصی یا مصرف‌کننده برای داده بانکی.
- تلقی علت ریشه‌ای یا دستورات تولیدشده توسط AI به‌عنوان حقیقت قطعی بدون راستی‌آزمایی.
- پیش‌نویس دستوراتی که تولید را بدون دروازه تأیید یا بازگشت تغییر می‌دهند.
- نبود مستندسازی درباره آنچه پاک‌سازی شد یا کدام ابزار استفاده شد.
- پاک‌سازی بیش‌ازحد تا جایی که AI نمی‌تواند کمک کند، و سپس حدس زدن به‌جای ارجاع به بالادست.

بررسی‌های ایمنی/انطباق

بررسی	الزام
مدیریت داده	هیچ PAN، secret، اعتبارنامه یا رشته اتصالی از مرز بانک خارج نمی‌شود؛ مدیریت PAN با انتظارات PCI-DSS هم‌راستاست.
تأیید ابزار	تنها LLM سازمانی تأییدشده بانک استفاده می‌شود؛ ابزارهای مصرف‌کننده برای هر داده داخلی ممنوع‌اند.
کمینه مواجهه	تنها کمینه گزیده پاک‌سازی‌شده موردنیاز برای تریاژ با AI به اشتراک گذاشته می‌شود.
مدیریت تغییر	گام‌های مؤثر بر تولید از تأیید تغییر همراه با یک بازگشت مستند عبور می‌کنند.
قابلیت ممیزی	سابقه بهداشت‌سازی و استفاده از ابزار AI برای بازبینی بعدی ثبت می‌شود.
بهداشت رخداد	secret نشت‌یافته به‌عنوان به‌خطرافتاده تلقی و چرخش (rotate) می‌شود؛ نشت PAN مطابق فرایند رخداد داده بانک گزارش می‌شود. (استنتاج‌شده از منابع عمومی: PCI-DSS و رویه استاندارد پاسخ به رخداد.)

جایگاه پیشنهادی در نردبان (استنتاج‌شده از منابع عمومی): عملکرد قوی در اینجا به **L4-L5** نگاشت می‌شود (پرامپت‌نویسی هدفمندِ تعبیه‌شده در یک گردش‌کار تریاژ واقعی با صرفه‌جویی زمانی اندازه‌گیری‌شده). داوطلبی که علاوه‌بر این، حفاظ‌های پایدار خط‌لوله را پیشنهاد می‌دهد (پاک‌سازی لاگ، داده تست توکنیزه‌شده، خودکارسازی چرخش secret) و آن‌ها را برای تیم مدون می‌کند، به سمت **L6-L7** گرایش دارد.

x.15 مهندس مشاهده‌پذیری (Observability Engineer)

تناسب با نردبان: این آزمون **L3-L5** را هدف می‌گیرد (کاربر پایه AI تا کاربر گردش‌کاری AI)، با نشانگرهای کششی برای **L6/L7** (کاربر قدرتمند / مشارکت‌کننده خودکارسازی). این آزمون برای یک مهندس مشاهده‌پذیری (Observability Engineer) ساخته شده که مالک متریک‌ها، لاگ‌ها، traceها، داشبوردها و هشداردهی برای سامانه‌های بانکی تولیدی است.

سناریو

ساعت 02:40 یکی از روزهای کاری است. **سرویس پرداخت بلادرنگ (real-time payments)** بانک یک هشدار نرخ مصرف بودجهٔ SLO تأخیر (latency SLO burn-rate) را فعال کرده است. مهندس Observability آماده‌به‌کار (on-call) برای بررسی و عیب‌یابی فراخوانده می‌شود. سیگنال‌های در دسترس عبارت‌اند از:

- یک داشبورد **Prometheus/Grafana** که نشان می‌دهد تأخیر p99 روی `payments-api` طی ۲۰ دقیقه از 80ms به 1.9s افزایش یافته و این موضوع با افزایش نرخ خطاهای 5xx همبستگی دارد.
- یک **ردیابی توزیع‌شده (distributed trace)** (OpenTelemetry/Jaeger) برای یک درخواست کند، که نشان می‌دهد بخش عمدهٔ زمان صرف فراخوانی یک سرویس پایین‌دستی به نام `customer-profile` شده است.
- یک **بریدهٔ لاگ برنامه (application log)** (حدود ۶۰۰ خط) شامل ردپای پشتهٔ خطا (stack trace)، هشدارهای استخر اتصال (connection-pool)، و بار دادهٔ درخواست/پاسخ (request/response payload).

این بریدهٔ لاگ حاوی دادهٔ زندهٔ مشتری است: شماره‌های کامل کارت (PAN)، شماره‌های ملی، نام مشتریان، مانده‌حساب‌ها، یک توکن احراز هویت bearer مربوط به سرویس داخلی، و یک رشتهٔ اتصال پایگاه داده (connection string) با گذرواژهٔ تعبیه‌شده در آن.

از مهندس خواسته می‌شود از یک دستیار GenAI سازمانیِ تأییدشده برای **تسریع عیب‌یابی** استفاده کند: خلاصه‌کردن خرابی، پیشنهاد یک علت ریشه‌ای، و ارائهٔ یک راهکار اصلاحی ایمن — بدون نشت دادهٔ حساس و بدون انجام هیچ اقدام تأییدنشده در محیط تولید (production).

وظیفه

1. درباره‌ی **اینکه چه داده‌هایی مجاز یا غیرمجاز برای ارسال** به دستیار هوش مصنوعی هستند تصمیم بگیرید. پیش از هرگونه paste، داده‌ها را پاک‌سازی (redact) یا ساختگی‌سازی کنید.
2. از هوش مصنوعی برای **خلاصه‌سازی خرابی** (علامت، دامنه‌ی اثر، خط زمانی) بر اساس سیگنال‌های پاک‌سازی‌شده استفاده کنید.
3. از هوش مصنوعی برای **ارائه‌ی یک فرضیه‌ی رتبه‌بندی‌شده درباره‌ی علت ریشه‌ای** استفاده کنید که جهش تأخیر (latency spike)، نقطه‌ی داغ trace و هشدارهای log را به هم پیوند دهد.
4. از هوش مصنوعی برای **پیشنهاد گزینه‌های ایمن اصلاحی** استفاده کنید که هر کدام با سطح ریسک و مسیر بازگشت (rollback) برچسب‌گذاری شده باشند.
5. به‌صراحت بیان کنید **چه چیزی باید پیش از رسیدن هر تغییر به production، توسط انسان راستی‌آزمایی شود**.
6. یک **بهبود پیگیرانه در حوزه‌ی observability** پیشنهاد دهید (برای مثال یک SLI، alert، پنل داشبورد یا ویژگی trace جدید) تا این دسته از خرابی‌ها زودتر شناسایی شوند.

کاربردهای مجاز AI

فعالیت	مجاز است؟	شرط
خلاصه‌سازی logها/traceها/metricهای **پاک‌سازی‌شده**	بله	فقط با ابزار AI سازمانی تأییدشده؛ داده‌های حساس ابتدا حذف شوند
تهیه پیش‌نویس PromQL / log query / قواعد هشدار (alert rules)	بله	خروجی پیش از استفاده در محیط non-prod بازبینی و آزمایش شود
پیشنهاد فرضیه‌های ریشه‌یابی علت (root-cause)	بله	به‌عنوان فرضیه تلقی شود نه واقعیت؛ در برابر سیگنال‌های واقعی راستی‌آزمایی شود
تهیه پیش‌نویس خلاصه رخداد (incident) / اطلاع‌رسانی‌ها	بله	پیش از انتشار برای ذینفعان توسط انسان بازبینی شود
چسباندن logهای خام حاوی PAN/شناسه/token/secret	**خیر**	تخلف جدی — به بخش پرچم‌های قرمز مراجعه کنید
اجازه دادن به AI برای اجرای دستورات روی محیط production	**خیر**	هیچ اصلاح خودکاری بدون تأیید و دروازه‌های بازگشت (rollback gates) مجاز نیست
استفاده از ابزار AI عمومی/مصرف‌کننده	**خیر**	فقط ابزارهای مجاز و دارای حاکمیت داده

خروجی مورد انتظار

یک سند کوتاه ترياژ (تقریباً یک صفحه) شامل موارد زیر:

1. **یادداشت پاکسازی (Sanitization)** — فهرستی از فیلدهای حساس شناسایی‌شده و نحوه برخورد با هر یک (حذف، ماسک‌گذاری، یا جایگزینی با یک placeholder)، به‌همراه بیانی صریح از آنچه به AI ارسال نشده است.
2. **خلاصه خرابی** — علامت/نشانه، سرویس(های) متأثر، اثر قابل‌مشاهده برای کاربر، و یک خط‌زمانی ۳ تا ۵ نقطه‌ای.
3. **فرضیه‌های ریشه‌یابی رتبه‌بندی‌شده** — دست‌کم دو فرضیه، هر یک به‌همراه شواهد پشتیبان آن و سیگنالی که برای تأیید یا رد آن لازم است.
4. **جدول گزینه‌های اصلاحی** — هر گزینه به‌همراه اقدام، سطح ریسک، اثر مورد انتظار، و مسیر بازگشت (rollback).
5. **چک‌لیست تأیید انسانی** — واقعیت‌ها/اقدامات مشخصی که باید پیش از هرگونه تغییر در محیط production توسط یک فرد تأیید شوند.
6. **بهبود مشاهده‌پذیری پیگیری‌شونده** — یک تغییر مشخص و قابل‌سنجش در سازوکار تشخیص یا سیگنال‌ها.

معیارهای ارزیابی

#	معیار	نمونه عملکرد مطلوب
1	**اولویت با مدیریت داده**	تمام فیلدهای حساس را شناسایی می‌کند (PAN، کد ملی، نام، مانده حساب، bearer token، رمز عبور/connection string پایگاه داده) و پیش از هرگونه استفاده از AI آن‌ها را پاک‌سازی می‌کند
2	**روش صحیح پاک‌سازی**	داده‌ها را ماسک یا با placeholder جایگزین می‌کند و در عین حال ارزش تشخیصی را حفظ می‌کند (مثلاً نوع خطاها، زمان‌بندی، status code و نقش هاست‌ها را نگه می‌دارد)
3	**خلاصه دقیق**	علائم، دامنه اثر (blast radius) و خط زمانی با سیگنال‌ها هم‌خوانی دارد؛ هیچ جزئیات ساختگی وجود ندارد
4	**ریشه‌یابی منطقی**	فرضیه‌ها به‌طور منطقی متریک‌ها + log + trace را به هم پیوند می‌دهند (مثلاً اشباع شدن connection-pool سرویس `customer-profile` که موجب تأخیر/خطای 5xx در سرویس‌های بالادست می‌شود)
5	**رفع ایمن مشکل**	گزینه‌ها برگشت‌پذیر، محدوده‌بندی‌شده و بر اساس ریسک رتبه‌بندی شده‌اند؛ هیچ ری‌استارت کورکورانه روی محیط prod یا تغییر schema بدون gate انجام نمی‌شود
6	**حضور انسان در حلقه**	پیشنهادهای AI را به‌روشنی از واقعیت‌های تأییدشده جدا می‌کند؛ مشخص می‌کند چه چیزی را باید انسان تأیید کند
7	**تردید نسبت به خروجی AI**	ریشه‌یابی AI را به‌عنوان یک فرضیه در نظر می‌گیرد؛ آن را با داشبوردها/traceهای واقعی راستی‌آزمایی متقابل می‌کند
8	**بهبود سازوکار تشخیص**	یک سیگنال هشدار زودهنگام قابل اندازه‌گیری پیشنهاد می‌دهد (مثلاً هشدار اشباع pool، یا SLI مربوط به وابستگی‌های پایین‌دست)

سنجه امتیازدهی (0-5)

امتیاز	سطح	توضیح
0	بدون مواجهه (L0/L1)	قادر به استفاده از AI برای triage نیست یا تمایلی به آن ندارد؛ یا لاگ‌های حساس خام را بدون آگاهی از ریسک کپی و الصاق می‌کند

امتیاز	سطح	توضیح
1	آگاه (L2)	می‌داند که داده‌های حساس نباید به اشتراک گذاشته شوند، اما قادر به انجام یک triage کاربردی نیست؛ پاک‌سازی (sanitization) مبهم یا ناقص است
2	کاربر مقدماتی (L3)	فیلدهای آشکار را پاک‌سازی می‌کند و یک خلاصه عمومی از AI دریافت می‌کند، اما برخی داده‌های حساس را از قلم می‌اندازد یا ریشه مشکل را در برابر سیگنال‌های واقعی راستی‌آزمایی نمی‌کند
3	کاربر مؤثر prompt (L4)	تمام فیلدهای حساس را به‌طور کامل پاک‌سازی می‌کند، با دقت prompt می‌نویسد، خلاصه‌ای دقیق و یک ریشه مشکل محتمل و رتبه‌بندی‌شده دریافت می‌کند و گام‌های راستی‌آزمایی انسانی را بیان می‌کند
4	کاربر workflow (L5)	تمام موارد بالا به‌علاوه گزینه‌های ترمیم برگشت‌پذیر با مسیرهای rollback، شک‌گرایی نسبت به خروجی AI و یک بهبود قابل‌اندازه‌گیری در شناسایی برای پیگیری؛ متریک‌ها + traceها + لاگ‌ها را به‌صورت منسجم یکپارچه می‌کند
5	کاربر حرفه‌ای / مشارکت‌کننده در automation (L6/L7)	یک الگوی ایمن و قابل‌استفاده مجدد تولید می‌کند: یک گام پاک‌سازی/پنهان‌سازی (redaction) که دیگران بتوانند از آن استفاده مجدد کنند، یک query/alert آزموده‌شده و پیشنهادی برای automation کنترل‌شده (تأیید + rollback) — در حالی که اقدامات production همچنان نیازمند تأیید انسانی باقی می‌مانند

آستانه قبولی برای این نقش: 3 یا بالاتر. هرگونه نشت داده حساسِ کنترل‌نشده، صرف‌نظر از کیفیت triage، امتیاز را در سقف **0** محدود می‌کند.

نشانه‌های هشدار (Red flags)

- درج لاگ‌های خام حاوی **شماره کارت (PAN)، کد ملی، نام مشتریان، مانده حساب، توکن‌های bearer یا اطلاعات احراز هویت پایگاه داده (DB credentials)** در هر ابزار AI.
- استفاده از **ابزار AI عمومی/مصرفی** به‌جای دستیار سازمانی مجاز.
- تلقی علت ریشه‌ای ارائه‌شده توسط AI به‌عنوان **واقعیت قطعی** و اقدام بر اساس آن بدون بررسی سیگنال‌های واقعی.
- پیشنهاد **تغییرات تولیدی (production) بدون مسیر بازگشت (rollback)** (برای مثال «فقط پایگاه داده production را restart کن»).
- پیشنهاد اینکه AI به‌صورت **اجرای خودکار (auto-execute)** اقدامات اصلاحی را روی محیط production بدون دروازه تأیید انجام دهد.
- پنهان‌کردن نشانه به‌جای پرداختن به علت (برای مثال بالا بردن آستانه هشدار برای خاموش‌کردن آن).
- ساختن خطوط لاگ، مقادیر متریک یا trace spanهایی که در منبع وجود ندارند («پذیرش hallucination به‌عنوان شواهد»).
- حذف بیش از حد زمینه (context) هنگام پاک‌سازی به‌گونه‌ای که تشخیص بی‌معنا می‌شود (over-redaction)، یا باقی‌گذاشتن شناسه‌های جزئی (برای مثال ۴ رقم آخر PAN به‌همراه نام و مانده حساب) که موجب بازشناسی مجدد مشتری می‌شود.

بررسی‌های ایمنی/انطباق

بررسی	الزام
کمینه‌سازی داده‌های حساس	هیچ PAN، PII (کد ملی، نام)، مانده حساب، راز (secret) یا توکنی نباید از مرز کنترل‌شده بانک به سمت AI خارج شود. این موضوع را از طریق یک چک‌لیست پیش از ارسال تأیید کنید.

بررسی	الزام
استفاده تنها از ابزارهای تأییدشده	دستیار AI باید نمونه سازمانی و تحت حاکمیت داده باشد — نه یک حساب شخصی/مصرف‌کننده.
همسویی با PCI-DSS (استنباط‌شده از منابع عمومی)	داده‌های دارنده کارت (PAN) نباید در معرض سیستم‌های تأییدنشده قرار گیرد؛ پوشاندن/کوتاه‌سازی (masking/truncation) باید از کنترل‌های PCI بانک پیروی کند.
چرخش رازها (Secret rotation)	اگر یک bearer token یا رمز عبور پایگاه‌داده در لاگ‌ها حضور داشت، آن را برای چرخش علامت‌گذاری کرده و مسیر نشت را گزارش دهید — لاگ‌کردن رازها خود یک رخداد (incident) است.
مدیریت تغییر	هرگونه اقدام اصلاحی که به محیط production می‌رسد باید از دروازه‌های تأیید و بازگردانی (rollback) عبور کند؛ هیچ اقدام مستقیم تولیدی هدایت‌شده توسط AI مجاز نیست.
رد ممیزی (Audit trail)	برای بازبینی رخداد ثبت کنید که چه چیزی به AI ارسال شد، چه چیزی بازگردانده شد و چگونه بر تصمیم اثر گذاشت.
پاسخگویی انسانی	یک مهندس مشخص مالک تصمیم نهایی اصلاحی است؛ خروجی AI صرفاً مشورتی است.

نکته: شیوه‌های PCI-DSS و لاگ‌گیری امن که در اینجا به آن‌ها اشاره شده، استانداردهای عمومی و به‌طور گسترده پذیرفته‌شده صنعتی هستند. این آزمون هیچ نردبان داخلی بانک یا کنترل خاصی فراتر از این استانداردهای عمومی را فرض نمی‌گیرد؛ پیش از استفاده، بررسی‌های بالا را به کاتالوگ کنترل واقعی سازمان خود نگاشت کنید.

x.15 مهندس DataOps

سناریو

یک خط لوله ETL دسته‌ای شبانه که داده‌های تراکنشی را از سیستم بانکداری متمرکز (core banking) به انبار داده گزارش‌گری نظارتی (regulatory-reporting data warehouse) بارگذاری می‌کند، با شکست مواجه شده است. مهندس DataOps کشیک در ساعت ۰۲:۱۰ فراخوانده می‌شود. ارکستریتور خط لوله (مثلاً Airflow/Jenkins) یک اجرای شکست‌خورده را برای کار `core_txn_to_regulatory_dwh` نشان می‌دهد. این شکست، به‌روزرسانی صبحگاهی گزارش‌گری AML (مبارزه با پولشویی) و نقدینگی را مسدود می‌کند که مهلت تحویل نظارتی آن ساعت ۰۹:۰۰ است.

مهندس می‌خواهد از یک دستیار GenAI داخلی/تأییدشده برای تریاژ سریع‌تر شکست استفاده کند. لاگ خام شکست شامل stack traceها، قطعات connection string، SQLها و نمونه‌ردیف‌هایی از رکوردهای ناموفق است که برخی از آن‌ها شامل PII مشتری (نام‌ها، کدهای ملی، شماره‌حساب‌ها، IBANها) و جزئیات تراکنش هستند.

مهندس باید یک خلاصه تریاژ تولید کند، یک اقدام اصلاحی ایمن پیشنهاد دهد و تصمیم بگیرد چه چیزی را می‌توان و چه چیزی را نمی‌توان با ابزار AI به اشتراک گذاشت.

وظیفه

1. از دستیار AI تأییدشده برای خلاصه‌کردن شکست و شناسایی محتمل‌ترین علت ریشه‌ای از لاگ (پاک‌سازی‌شده) استفاده کنید.
2. یک طرح اصلاحی با دروازه‌های ایمنی صریح پیشنهاد دهید (چه چیزی ابتدا در محیط non-prod اجرا می‌شود، چه چیزی به تأیید نیاز دارد، گام‌های بازگردانی).
3. به‌روشنی بیان کنید چه چیزی باید پیش از هرگونه تغییر در production توسط انسان تأیید شود.
4. تمام داده‌های حساس را پیش از ارسال هر محتوایی به ابزار AI شناسایی و پنهان (redact) کنید و مستند کنید چه چیزی و چرا حذف شد.
5. یک یادداشت رخداد کوتاه مناسب برای تیکت مدیریت تغییر و لاگ حاکمیت داده بنویسید.

این وظیفه سطوح نردبان **L4-L7** را هدف می‌گیرد: promptنویسی سنجیده (L4)، تعبیه AI در یک جریان کاری عملیاتی با صرفه‌جویی زمانی اندازه‌گیری‌شده (L5)، تعیین هنجارهای ایمن برای تیم (L6) و پیشنهاد اتوماسیون مبتنی بر دروازه تأیید/بازگردانی (L7).

استفاده مجاز از AI

مجاز	غیرمجاز
چسباندن گزیده‌های لاگ **پاک‌سازی‌شده** (با PII/رازهای حذف یا توکنیزه‌شده) در ابزار AI داخلی تأییدشده	چسباندن لاگ‌های خام حاوی PII، شماره‌حساب، IBAN، کد ملی یا اعتبارنامه‌های زنده (live credentials)
درخواست از AI برای توضیح stack traceها، خلاصه‌کردن خطاها و پیشنهاد علل ریشه‌ای کاندید	اجازه‌دادن به AI برای اجرای مستقیم دستورات روی production بدون تأیید انسانی
درخواست از AI برای پیش‌نویس گام‌های اصلاحی، اسکریپت‌های بازگردانی و یادداشت رخداد	پذیرفتن خروجی AI به‌عنوان مرجع معتبر بدون راستی‌آزمایی در برابر runbook، schema و دودمان داده (data lineage)
استفاده از AI برای تولید الگوهای redaction/regex جهت کمک به پاک‌سازی لاگ‌ها	ارسال connection stringها، توکن‌ها یا کلیدها به هر ابزار AI

خروجی مورد انتظار

داوطلب باید هر پنج خروجی را تولید کند:

1. **خلاصه شکست** (۳ تا ۶ خط): چه چیزی شکست خورد، در کجای DAG/کار، و دسته خطای آشکارشده.
2. **فرضیه علت ریشه‌ای** با حداقل یک گزینه جایگزین، هرکدام با یک سطح اطمینان و نحوه تأیید آن.
3. **طرح اصلاحی** با توالی non-prod-first، دروازه تأیید صریح و گام‌های بازگردانی.
4. **فهرست راستی‌آزمایی انسانی**: مواردی که باید پیش از اجرای production توسط یک فرد بررسی شوند (مثلاً اثر تغییر schema، تعداد ردیف‌ها، تطبیق پایین‌دستی، اثر بر مهلت نظارتی).
5. **سابقه مدیریت داده**: جدولی از فیلدهای حساس یافت‌شده، اقدام انجام‌شده (redact/توکنیزه/حذف) و دلیل آن، به‌علاوه یک جمله یک‌خطی که تأیید می‌کند هیچ PII/رازی به AI ارسال نشده است.

معیارهای ارزیابی

بُعد	شاخصِ کیفیت مطلوب
مدیریت داده	تمام فیلدهای حساس را شناسایی می‌کند، پیش از استفاده از AI پاک‌سازی می‌کند و آن را مستند می‌کند
کیفیت تشخیص	علت ریشه‌ای محتمل که به شواهد واقعی لاگ گره خورده، نه حدس‌های کلی
دروازه‌گذاری ایمنی	non-prod-first، دروازه تأیید و بازگردانی آزمون‌شده به‌صراحت ذکر شده‌اند
انسان‌درحلقه (Human-in-the-loop)	به‌روشنی پیشنهادهای AI را از آنچه انسان باید تأیید کند جدا می‌کند
تناسب عملیاتی	خروجی در یک تیکت تغییر واقعی و لاگ حاکمیت قابل‌استفاده است
آگاهی از انطباق	مهلت نظارتی، رد ممیزی و قواعد اقامت داده (data-residency) را در نظر می‌گیرد

سنجه امتیازدهی (0-5)

امتیاز	توضیح
0	لاگ خام (با PII/رازها) را در AI می‌چسباند، یا بدون هیچ راستی‌آزمایی به خروجی AI تکیه می‌کند. ناآگاه از ریسک داده. (رفتار L0-L1)
1	می‌داند که PII نباید به اشتراک گذاشته شود اما نمی‌تواند آن را به‌طور قابل‌اعتماد شناسایی یا پنهان کند؛ هیچ اقدام اصلاحی واقعی ندارد. (L2)
2	فیلدهای بدیهی را پاک‌سازی می‌کند و یک خلاصه کلی از AI می‌گیرد؛ اقدام اصلاحی فاقد دروازه‌های ایمنی یا بازگردانی است. (L3)
3	redaction کامل همراه با مستندسازی؛ promptنویسی سنجیده یک علت ریشه‌ای معتبر می‌دهد؛ اقدام اصلاحی بازگردانی دارد اما دروازه‌گذاری تأیید/non-prod ضعیف است. (L4)
4	هر پنج خروجی حاضر و منسجم‌اند؛ AI در جریان کاری با فهرست راستی‌آزمایی انسانی و دروازه‌های تأیید + بازگردانی تعبیه شده است؛ زمان صرفه‌جویی‌شده در مقابل تریاژ دستی را ذکر می‌کند. (L5-L6)
5	مانند بالا، به‌علاوه یک اتوماسیون پاک‌سازی-و-تریاژ قابل‌استفاده مجدد و دارای حفاظ (مثلاً یک گام redaction پیش از AI در runbook) با کنترل‌های تأیید و بازگردانی پیشنهاد می‌دهد و هنجار را با تیم به اشتراک می‌گذارد. (L7)

پرچم‌های قرمز

- لاگ خام، یک connection string یا هر اعتبارنامه‌ای را در ابزار AI می‌چسباند.
- PII مشتری (کدهای ملی، IBANها، شماره‌حساب‌ها) در هر prompt ارسالی به AI ظاهر می‌شود.
- پیشنهاد اجرای مستقیم یک اصلاح پیشنهادی AI در production بدون تأیید یا بازگردانی.
- علت ریشه‌ای AI را بدون بررسی در برابر لاگ، schema یا دودمان واقعی می‌پذیرد.
- مهلت نظارتی ۰۹:۰۰ را نادیده می‌گیرد یا هنگام درخطربودن، تشدید (escalate) نمی‌کند.
- از یک ابزار AI تأییدنشده/عمومی ("shadow AI") برای داده‌های بانک استفاده می‌کند.
- بدون رد ممیزی: نمی‌تواند نشان دهد چه چیزی پنهان شده یا چه چیزی از AI پرسیده شده است.

بررسی‌های ایمنی / انطباق

بررسی	شرط قبولی
پاک‌سازی داده‌های حساس	هیچ PII، شماره‌حساب، IBAN، کد ملی یا رازی در هیچ prompt ارسالی به AI نیست؛ redaction مستند شده است
استفاده تنها از ابزارهای تأییدشده	از ابزار AI داخلی مجاز بانک استفاده می‌کند، نه یک سرویس عمومی/مصرف‌کننده
اقامت داده (Data residency)	تأیید می‌کند که داده و پردازش AI درون حوزه قضایی/محیط مجاز باقی می‌مانند
دروازه تأیید انسانی	هیچ تغییری در production بدون تأیید انسانی مستند (سابقه مدیریت تغییر) انجام نمی‌شود
آمادگی بازگردانی	پیش از اجرای هر اقدام اصلاحی، یک مسیر بازگردانی آزمون‌شده وجود دارد
رد ممیزی	ورودی‌های prompt، خروجی‌های AI و اقدامات redaction برای بازبینی لاگ می‌شوند
مهلت / تشدید	مهلت نظارتی ۰۹:۰۰ پیگیری می‌شود؛ مسیر تشدید در صورت درخطربودن فعال می‌شود

بررسی	شرط قبولی
کمینه مواجهه (Least exposure)	تنها حداقل گزیده پاک‌سازی‌شده مورد نیاز برای ترياژ به اشتراک گذاشته می‌شود، نه کل لاگ

x.15 مهندس اتوماسیون QA

سطح هدف نردبان برای «قبولی»: L4 (کاربر مؤثر Prompt) با شاخص‌های کششی به‌سوی L5 (کاربر AI در جریان کاری). این آزمون به‌گونه‌ای طراحی شده که یک داوطلب L2-L3 بتواند ایمنی پایه را نشان دهد، درحالی‌که داوطلبان L5+ بلوغ جریان کاری و اندازه‌گیری را آشکار می‌کنند.

سناریو

شما از فرآیند تست رگرسیون خودکار برای یک اپلیکیشن وب بانکداری خرد (retail banking) پشتیبانی می‌کنید. شب گذشته، **خط لوله Jenkins به نام** `core-banking-regression` در مرحله `integration-tests` با شکست مواجه شد. دوازده تست از مجموع ۳۴۰ تست Selenium/REST API ناموفق بودند. سرپرست QA از شما خواسته است که به‌سرعت اولویت‌بندی و بررسی اولیه (triage) را انجام دهید تا مسیر انتشار نسخه (release train) متوقف نشود.

لاگ خام خط لوله حجیم است و شامل موارد زیر می‌شود:

- ردیابی پشته (stack trace) و خطاهای assertion
- یک test fixture که **PAN (شماره کارت)، نام و نام خانوادگی کامل، کد ملی، شماره حساب و یک session bearer token متعلق به یک مشتری واقعی** را که در جریان یک تست ناموفق فرآیند پرداخت ثبت شده، لاگ کرده است
- نام میزبان‌های (hostname) داخلی، رشته‌های اتصال پایگاه داده (database connection string) همراه با اعتبارنامه‌های تعبیه‌شده، و یک ارجاع به تیکت داخلی Jira

شما مجاز هستید از **دستیار GenAI داخلی و تأییدشده بانک** استفاده کنید (مستأجر سازمانی، بدون آموزش مدل روی ورودی‌ها، داده‌ها در محیط بانک باقی می‌مانند). ابزارهای GenAI عمومی/مصرفی برای هیچ‌یک از داده‌های کاری **تأیید نشده‌اند**.

وظیفه

با استفاده از دستیار هوش مصنوعی تأییدشده، یک خلاصه‌ی دسته‌بندی (triage) تهیه کنید که:

1. ۱۲ خرابی را خلاصه کند و آن‌ها را بر اساس نوع محتمل خرابی گروه‌بندی کند (نقص محصول در برابر ناپایداری آزمون/محیط در برابر آماده‌سازی داده).
2. محتمل‌ترین علت یا علل ریشه‌ای را همراه با شواهد پشتیبان از لاگ‌ها شناسایی کند.
3. گام‌های اصلاحی ایمن را پیشنهاد دهد و به‌روشنی موارد «ایمن برای اقدام فوری» را از موارد «نیازمند تأیید انسانی پیش از اقدام» جدا کند.
4. به‌صراحت بیان کند که یک انسان چه چیزی را باید تأیید کند (برای مثال، اینکه آیا یک خرابی پیش از بازگشایی یک نسخه، بازتاب‌دهنده‌ی یک نقص واقعی است یا خیر).
5. **پیش از** آنکه هرگونه محتوای لاگ به دستیار هوش مصنوعی ارسال شود، هر قطعه داده‌ی حساس در لاگ و نحوه‌ی مدیریت آن (حذف، ماسک‌گذاری یا توکن‌سازی) را شناسایی کند.

کاربردهای مجاز AI

مجاز	غیرمجاز
دستیار GenAI سازمانی تأییدشده برای خلاصه‌سازی لاگ‌های **پاک‌سازی‌شده**، تدوین فرضیه‌های ریشه‌یابی و تدوین گام‌های اصلاحی	چسباندن لاگ‌های خام و پاک‌سازی‌نشده در هر ابزاری
استفاده از AI برای تولید regex یا اسکریپت‌های ماسک‌گذاری/حذف اطلاعات (با بازبینی انسانی پس از آن)	هر ابزار AI مصرفی/عمومی (نسخه رایگان ChatGPT و مانند آن) برای داده‌های کاری
استفاده از AI برای تدوین ساختار خلاصه triage و متن کامنت Jira	اجازه دادن به AI برای ادغام خودکار، بستن خودکار تیکت‌ها یا راه‌اندازی مجدد اجرا بدون تأیید انسانی
استفاده از AI برای پیشنهاد اینکه کدام تست‌ها بر اساس الگوها احتمالاً ناپایدار (flaky) هستند	ارسال PAN، کد ملی، توکن‌ها یا اعتبارنامه‌ها به AI حتی روی tenant تأییدشده

خروجی مورد انتظار

داوطلبی که در حد قبولی عمل می‌کند، باید این موارد را ارائه دهد:

1. **دفترچه ثبت داده‌های حساس پیش از ارسال** که هر عنصر حساس یافت‌شده و اقدام انتخاب‌شده برای آن را فهرست می‌کند:

عنصر حساس	نمونه در لاگ	اقدام مدیریتی	دلیل
PAN (شماره کارت)	`4111 11** **** 1111`	حذف کامل	PCI-DSS؛ حتی روی tenant تأییدشده هرگز به AI ارسال نشود
کد ملی	`national_id=...`	حذف کامل	PII / مقرراتی
نام کامل مشتری	`name=...`	حذف یا جایگزینی با `CUSTOMER_A`	PII
شماره حساب	`acct=...`	ماسک‌گذاری (فقط ۴ رقم آخر) یا توکن‌سازی	PII / قابلیت ردیابی
توکن bearer نشست	`Authorization: Bearer ...`	حذف و چرخش/ابطال	اعتبارنامه فعال؛ رخداد امنیتی
رشته اتصال DB + اعتبارنامه‌ها	`jdbc:...user=...password=...`	حذف و چرخش	افشای اعتبارنامه
نام‌های میزبان داخلی	`host-int-...`	جایگزینی با placeholder عمومی	کاهش سطح حمله داخلی
ارجاع تیکت Jira	`BANK-1234`	می‌تواند نگه داشته شود (حساسیت پایین)	زمینه عملیاتی

1. **یک گزیده لاگ پاک‌سازی‌شده** (یا توصیفی از نحوه پاک‌سازی آن) که تأیید می‌کند هیچ چیزی بالاتر از آستانه توافق‌شده به AI ارسال نشده است.
2. **یک خلاصه دسته‌بندی‌شده از شکست‌ها** (نقص / ناپایدار / مشکل آماده‌سازی داده) به همراه تعداد هر دسته.
3. **فرضیه‌های ریشه‌یابی علت** که به شواهد لاگ مشخص (و پاک‌سازی‌شده) گره خورده‌اند.

4. **یک برنامه اصلاحی** که به دو بخش «اکنون ایمن است» در برابر «نخست نیازمند تأیید انسانی» تقسیم شده باشد.
5. **یک فهرست صریح از موارد نیازمند تأیید انسانی** و یادداشتی مبنی بر اینکه خروجی AI بازبینی شده و کورکورانه به آن اعتماد نشده است.
6. یادداشتی مبنی بر اینکه توکن bearer فعال باید به‌عنوان یک **رخداد بالقوه افشای اعتبارنامه** گزارش شود (چرخش/ابطال)، نه صرفاً پنهان‌سازی.

معیارهای ارزیابی

بُعد	نمونه عملکرد مطلوب
مدیریت داده (با بیشترین وزن)	تمام عناصر حساس، به‌ویژه توکن فعال و اطلاعات احراز هویت (credentials) را تشخیص می‌دهد؛ برای هر عنصر اقدام درست را انتخاب می‌کند؛ پیش از ارسال، داده‌ها را پاک‌سازی (sanitize) می‌کند؛ و تنها از ابزار تأییدشده استفاده می‌کند
تشخیص رخداد (incident)	توکن bearer افشاشده و اطلاعات احراز هویت پایگاه داده را به‌عنوان یک رخداد امنیتی که نیازمند چرخش (rotation) است تشخیص می‌دهد، نه صرفاً موضوعی برای پنهان‌سازی (redaction)
کیفیت تریاژ (triage)	نقص‌ها (defects) را به‌درستی از ناپایداری‌ها (flakiness) تفکیک می‌کند؛ ریشه‌یابی‌ها مبتنی بر شواهد است، نه حدس و گمان
تکنیک طراحی پرامپت	نقش، زمینه و محدودیت‌های روشنی به AI می‌دهد؛ به‌جای پذیرش اولین خروجی، فرایند را به‌صورت تکرارشونده بهبود می‌دهد
حضور انسان در فرایند (Human-in-the-loop)	به‌روشنی مشخص می‌کند چه چیزی پیشنهاد AI بوده و چه چیزی باید پیش از هر تصمیم انتشار توسط انسان تأیید شود
انتخاب ابزار و استدلال	از tenant سازمانی تأییدشده استفاده می‌کند؛ و می‌تواند توضیح دهد چرا ابزارهای عمومی ممنوع هستند
ارتباطات	خروجی به‌قدر کافی روشن است که سرپرست QA (QA lead) بتواند بر اساس آن اقدام کند

سنجه نمره‌دهی (0-5)

نمره	شرح
0	لاگ خام (شامل PAN/token/credentials) را در هر ابزار AI کپی می‌کند، یا از ابزاری تأییدنشده/عمومی استفاده می‌کند. صرف‌نظر از کیفیت دسته‌بندی اولیه، مردودی خودکار است.
1	از ابزار تأییدشده استفاده می‌کند اما در پاک‌سازی، یک یا چند عنصر پرریسک (PAN، token یا credentials) را از قلم می‌اندازد. ممکن است دسته‌بندی اولیه انجام شده باشد اما مدیریت داده ناایمن است.
2	PII آشکار را پاک‌سازی می‌کند اما bearer token / DB credentials را صرفاً به‌عنوان موردی برای «حذف» تلقی می‌کند و زاویه‌ی رخداد افشای credential را نمی‌بیند. دسته‌بندی اولیه سطحی است یا بیشتر عیناً برگرفته از AI و با راستی‌آزمایی اندک است.
3	همه‌ی عناصر حساس را به‌درستی پاک‌سازی می‌کند و از ابزار تأییدشده بهره می‌برد. یک دسته‌بندی اولیه‌ی گروه‌بندی‌شده‌ی قابل استفاده و فرضیه‌های ریشه‌ای تولید می‌کند. برخی نیازهای راستی‌آزمایی انسانی را علامت‌گذاری می‌کند. آستانه‌ی قبولی L3-L4 را برآورده می‌سازد.

نمره	شرح
4	همه‌ی موارد بالا، به‌علاوه: token/credentials را به‌عنوان رخدادی نیازمند چرخش (rotation) علامت‌گذاری می‌کند؛ به‌روشنی «اکنون ایمن» را از «نیازمند راستی‌آزمایی انسانی» تفکیک می‌کند؛ prompting سنجیده (نقش/زمینه/محدودیت‌ها) و دست‌کم یک تکرار برای بهبود خروجی را نشان می‌دهد. L4 محکم.
5	همه‌ی موارد بالا، به‌علاوه: یک گام پاک‌سازی قابل استفاده‌ی مجدد و قابل بازبینی می‌سازد (مثلاً یک اسکریپت/regex ماسک‌گذاری که دیگران بتوانند اجرا کنند)، زمان صرفه‌جویی‌شده را اندازه‌گیری یا تخمین می‌زند، و پیشنهاد می‌دهد پاک‌سازی در جریان‌کار مدیریت لاگ CI تعبیه شود تا این کار تکرارپذیر و قابل ممیزی باشد. تفکر جریان‌کاری L5 را نشان می‌دهد.

آستانه‌ی قبولی: 3. نمرات 0-1 به دلایل ایمنی، مستقل از مهارت دسته‌بندی اولیه، ردکننده هستند.

نشانه‌های هشدار

- محتوای لاگ پاک‌سازی‌نشده را در یک ابزار AI کپی می‌کند، یا برای داده‌های بانک از یک ابزار AI عمومی/مصرفی استفاده می‌کند.
- bearer token فعال یا اعتبارنامه‌های پایگاه‌داده را صرفاً موضوعی برای «حذف و پنهان‌سازی» تلقی می‌کند و نگرانی مربوط به افشای اعتبارنامه/چرخش (rotation) آن را مطرح نمی‌کند.
- ادعای AI درباره علت ریشه‌ای را به‌عنوان یک واقعیت می‌پذیرد و بدون راستی‌آزمایی مستقل انسانی، بازگشایی مجدد یا متوقف‌کردن انتشار (release) را توصیه می‌کند.
- اجازه می‌دهد AI بدون تأیید انسانی، یک pipeline را دوباره اجرا کند، تیکت‌ها را ببندد، یا تغییرات را merge کند.
- نمی‌تواند توضیح دهد چرا tenant سازمانی تأییدشده از یک ابزار عمومی ایمن‌تر است (محل اقامت داده، عدم آموزش مدل روی ورودی‌ها).
- به AI «اعتماد بیش از حد» می‌کند: گام‌های اصلاحی را بدون بازبینی، عیناً در اسکریپت‌های production کپی می‌کند.
- PAN را پنهان (mask) می‌کند اما کد ملی، شماره حساب یا connection string های داخلی را دست‌نخورده باقی می‌گذارد.
- یک «قطعه کوچک» حاوی یک secret را ارسال می‌کند، چون به‌نظرش کم‌خطر آمده است.

بررسی‌های ایمنی/انطباق

بررسی	الزام
تأیید ابزار	تنها از تننت سازمانی GenAI مورد تأیید بانک استفاده می‌شود؛ عدم استفاده از ابزارهای عمومی تأیید شده است
کمینه‌سازی داده	تنها محتوای لاگِ پاک‌سازی‌شده و حداقلِ ضروری به AI ارسال می‌شود
PCI-DSS	تحت هیچ شرایطی PAN/داده‌های دارنده کارت به AI ارسال نمی‌شود، حتی روی تننت تأییدشده
مدیریت PII	شناسه ملی، نام و شماره حساب پیش از ارسال مطابق سیاست‌ها حذف/ماسک/توکنایز می‌شوند
افشای اطلاعات هویتی	bearer token فعال و اعتبارنامه‌های پایگاه‌داده برای چرخش/ابطال از طریق فرایند رخداد امنیتی گزارش می‌شوند، نه آنکه بی‌سروصدا حذف شوند
دروازه‌های تأیید انسانی	هیچ اقدام انتشار، تیکت، merge یا pipeline صرفاً بر مبنای خروجی AI انجام نمی‌شود؛ یک فرد مشخص آن را تأیید می‌کند
قابلیت ممیزی	مرحله پاک‌سازی و آنچه به AI ارسال شده ثبت می‌شود تا ترياژ قابل بازتولید و بازبینی باشد

بررسی	الزام
راستی‌آزمایی خروجی	علل ریشه‌ای و اقدامات اصلاحیِ تولیدشده توسط AI پیش از اثرگذاری بر هر تصمیم انتشار، توسط انسان راستی‌آزمایی می‌شوند

x.15 مهندس امنیت (Security Engineer)

جایگاه در نردبان: این آزمون **L3–L5** (کاربر پایه AI ← کاربر AI در گردش‌کار) را هدف می‌گیرد، همراه با شاخص‌های کششی برای **L6**. فرض بر این است که مهندس پیش‌تر الزامات L2 (آگاهی از AI و قواعد مدیریت داده) را برآورده کرده است. نقش‌های بالاتر از L6 به‌صورت جداگانه از طریق تمرین‌های طراحی راهکار و حاکمیت ارزیابی می‌شوند.

سناریو

شما مهندس امنیتِ کشیک (on-call Security Engineer) در یک بانک خرده‌فروشی تحت نظارت هستید. ساعت 09:14، سامانه SIEM یک هشدار با شدت متوسط صادر می‌کند: یک Web Application Firewall (WAF) که رو به اینترنت عمومی قرار دارد، انبوهی از درخواست‌های مشکوک علیه API فرایند جذب مشتری (onboarding) را ثبت کرده است. بسته‌ای 600 خطی برای تریاژ سریع به شما تحویل داده می‌شود:

- لاگ‌های خام دسترسیِ WAF و reverse-proxy (خطوط کامل درخواست، هدرها، IPهای مبدأ)
- یک stack trace از سرویس onboarding
- تکه‌ای از diff مربوط به `terraform plan` برای security group این API
- بخشی از یک رکورد مشتری که در یکی از خطوط لاگ ظاهر شده است (نام، کد ملی، PAN ماسک‌شده، ایمیل)

رهبری سازمان انتظار دارد ظرف 30 دقیقه یک خلاصهٔ تریاژ دریافت کند. شما مجاز هستید از **دستیار GenAI داخلی و تأییدشدهٔ بانک** استفاده کنید (داده‌ها تا سطح "Internal" مجازند؛ این دستیار برای سطوح "Confidential/Restricted" یا PII مشتری **تأیید نشده** است). استفاده از ابزارهای AI عمومی/مصرفی برای هرگونه دادهٔ کاری ممنوع است.

وظیفه

1. با استفاده از AI، رخداد را از بستهٔ لاگ‌ها **خلاصه** کنید و یک روایت تریاژ به زبان ساده تولید نمایید.
2. **علت ریشه‌ای محتمل** را شناسایی کنید (الگوی حمله در برابر پیکربندی نادرست در برابر هشدار کاذب) به همراه مؤلفهٔ متأثر.
3. گام‌های **اصلاح ایمن** را پیشنهاد دهید که به دو دستهٔ «مهار فوری» و «اصلاح اصولی» تفکیک شده باشند و هر کدام با تأییدیه‌های لازم برچسب‌گذاری شوند (تیکت تغییر، بازبینی همتا، تأیید نهایی مسئول امنیت).
4. به‌صراحت بیان کنید که **پیش از انجام هر اقدامی، انسان چه چیزی را باید راستی‌آزمایی کند**.
5. **پیش از ارائهٔ prompt به AI**، تمام داده‌های حساسی را که نباید برای دستیار ارسال شوند شناسایی و حذف (redact) کنید و مستند نمایید که چه چیزی را و به چه دلیلی حذف کرده‌اید.

استفاده مجاز از AI

فعالیت	مجاز است؟	شرط
خلاصه‌سازی لاگ‌های پاک‌سازی‌شده	بله	PII و اسرار حذف شده باشند؛ داده ≤ سطح "Internal"

فعالیت	مجاز است؟	شرط
دسته‌بندی الگوی حمله / پیشنهاد قواعد شناسایی	بله	فقط بر روی گزیده‌های پاک‌سازی‌شده
تدوین پیش‌نویس گام‌های اصلاح و بازگشت (rollback)	بله	خروجی به‌عنوان پیش‌نویس تلقی شده و توسط انسان راستی‌آزمایی شود
درج رکورد خام مشتری (نام، کد ملی، شماره کارت/PAN)	**خیر**	PII محدودشده — هرگز به AI ارسال نشود
درج IPهای مبدأ واقعی، نام میزبان‌های داخلی، اسرار، توکن‌ها	**خیر**	ابتدا حذف یا توکن‌سازی (redact/tokenize) شود
اعمال خودکار اصلاح (تغییر firewall/SG) از خروجی AI	**خیر**	نیازمند تیکت تغییر + تأیید انسانی + برنامه بازگشت است
استفاده از یک ابزار AI عمومی/مصرفی	**خیر**	نقض سیاست، صرف‌نظر از محتوا

خروجی مورد انتظار

داوطلب باید موارد زیر را تولید کند:

1. **گزارش پاک‌سازی (Redaction log)** — جدولی کوتاه که هر قلم حساسِ یافت‌شده، طبقه‌بندی آن و اقدام انجام‌شده (حذف / توکن‌سازی / ماسک‌گذاری) را پیش از هرگونه استفاده از AI فهرست می‌کند.
2. **پرامپت پاک‌سازی‌شده** — متن/گزیدهٔ واقعی که به دستیار تأییدشده ارسال شده و نشان می‌دهد که PII، اسرار (secrets) و شناسه‌های داخلی حذف شده‌اند.
3. **خلاصهٔ تریاژ** — ۵ تا ۱۰ خط: چه اتفاقی افتاد، کدام مؤلفه متأثر شد، ارزیابی شدت و سطح اطمینان.
4. **فرضیهٔ ریشه‌یابی** — به‌صورت یک فرضیه همراه با شواهد پشتیبان بیان شود، نه به‌عنوان یک قطعیت.
5. **برنامهٔ رفع مشکل** — تفکیک‌شده به «مهار فوری» (Contain Now) در برابر «اصلاح اصولی» (Fix Properly)، که هر یک دارای مسئول، دروازهٔ تأیید (approval gate) و یادداشت بازگشت (rollback note) است.
6. **چک‌لیست راستی‌آزمایی انسانی** — فهرست صریحی از ادعاها/اقداماتی که پیش از اجرا باید توسط یک فرد تأیید شوند.

معیارهای ارزیابی

#	معیار	نمودِ «عملکرد خوب» چگونه است
1	اولویت‌دادن به مدیریت داده	داده‌های PII و اطلاعات محرمانه را **پیش از** نخستین prompt پاک‌سازی می‌کند؛ این کار را مستند می‌سازد
2	انتخاب درست ابزار	تنها از دستیار داخلی تأییدشده استفاده می‌کند؛ ابزارهای عمومی را رد می‌کند
3	کیفیت prompt	promptهای شفاف و دارای دامنهٔ مشخص که خروجی مفید و ساختاریافته تولید می‌کنند (نشانهٔ سطح L4)
4	دقت فنی	علت ریشه‌ای و راهکار اصلاحی، با توجه به شواهد، باورپذیر و درست است

#	معیار	نمودِ «عملکرد خوب» چگونه است
5	چارچوب‌بندی فرضیه	میزان اطمینان را بیان می‌کند؛ میان حمله، پیکربندی نادرست و هشدار کاذب (false positive) تمایز قائل می‌شود
6	انضباط در تأیید و بازگشت به وضعیت پیشین	هیچ تغییری بدون تیکت، بازبینی و مسیر rollback اعمال نمی‌شود
7	راستی‌آزمایی انسانی	به‌روشنی پیش‌نویس تولیدشده توسط AI را از آنچه باید توسط انسان تأیید شود جدا می‌کند
8	تفکر فرایندی (هدف کششی)	یک جریان تکرارپذیر sanitize→summarize→triage یا یک قاعدهٔ تشخیص پیشنهاد می‌دهد (سطح L5/L6)

سنجه نمره‌دهی (0–5)

نمره	سطح	توضیح
0	بدون تعامل / ناایمن	داده‌های شخصی (PII) یا اسرار خام را مستقیماً در AI جای‌گذاری می‌کند، یا از ابزاری ممنوعه استفاده می‌کند. صرف‌نظر از کیفیت فنی، مردودی خودکار.
1	آگاه اما ناایمن	می‌داند که AI می‌تواند کمک کند، اما به‌صورت ناهماهنگ داده‌ها را حذف یا پنهان می‌کند، یا نام میزبان‌ها/آدرس‌های IP داخلی را بدون تأمل ارسال می‌کند.
2	کاربر مقدماتی	داده‌های شخصی بدیهی را حذف می‌کند و خلاصه‌ای قابل‌استفاده به دست می‌آورد، اما خروجی AI را با راستی‌آزمایی اندک می‌پذیرد؛ اقدامات اصلاحی مبهم است.
3	کاربر روزمرهٔ ایمن (**معادل L3**)	حذف تمیز داده‌ها همراه با ثبت مستند، خلاصهٔ تریاژ منطقی، علت ریشه‌ای محتمل و راستی‌آزمایی ادعاهای کلیدی. پایه‌ای مستحکم.
4	کاربر مؤثر prompt (**معادل L4**)	همهٔ موارد L3 به‌علاوهٔ promptنویسی هدفمند و خوش‌ساختار؛ تفکیک اقدامات اصلاحی به مهار/رفع همراه با دروازه‌های تأیید و امکان بازگردانی؛ فهرست شفاف موارد نیازمند راستی‌آزمایی انسانی.
5	کاربر گردش‌کار (**معادل L5/L6**)	همهٔ موارد L4 به‌علاوهٔ یک گردش‌کار قابل‌استفادهٔ مجدد برای پاک‌سازی/تریاژ یا یک قانون تشخیص پیشنهادی، صرفه‌جویی زمانی قابل‌اندازه‌گیری، و هنجارهایی که دیگران بتوانند از آن‌ها پیروی کنند. کل تیم را ارتقا می‌دهد.

آستانهٔ قبولی: 3 از 5. هر نمرهٔ 0 به‌منزلهٔ مردودی خودکار در این تمرین است.

نشانه‌های هشدار (Red flags)

- چسباندن (paste) مستقیم سابقه‌ی مشتری، کد ملی یا شماره‌ی کارت (PAN) در هر ابزار AI. (مردودی خودکار.)
- ارسال IPهای مبدأ واقعی، نام میزبان‌های داخلی، اسرار (secrets)، توکن‌های API یا diff خام `terraform` بدون پاک‌سازی و حذف اطلاعات حساس.
- استفاده از یک دستیار AI عمومی/مصرفی برای داده‌های کاری بانک.
- برخورد با خروجی AI به‌مثابه‌ی حقیقت مطلق — اعمال مستقیم یک تغییر در firewall یا security-group بر اساس پیشنهاد AI.
- نداشتن طرح بازگشت (rollback)، نبود تیکت تغییر (change ticket) یا نبود تأیید انسانی پیش از پیشنهاد اقدام.
- ناتوانی در تشخیص تفاوت میان یک پاسخ قاطعانه‌ی AI و یک واقعیت راستی‌آزمایی‌شده.
- اشتراک‌گذاری بیش از حد جزئیات رخداد به‌صورت خارجی یا در کانال‌های تأییدنشده.

- توجیه حذف مرحله‌ی پاک‌سازی داده‌ها با استناد به فشار زمانی ۳۰ دقیقه‌ای («وقتی برای پاک‌سازی نیست»).

بررسی‌های ایمنی / انطباق

بررسی	الزام	شرط قبولی
طبقه‌بندی داده	هیچ داده‌ای بالاتر از سطح «داخلی» به AI ارسال نشود	لاگ پاک‌سازی (Redaction) نشان دهد که تمام اقلام Confidential/Restricted حذف شده‌اند
حفاظت از PII	نام مشتری، کد ملی، شماره کارت (PAN) و ایمیل هرگز به AI نرسد	در پرامپت پاک‌سازی‌شده تأیید شده باشد
بهداشت رازها (Secrets)	هیچ توکن، کلید، نام میزبان یا IP داخلی ارسال نشود	پیش از پرامپت‌دهی توکنیزه یا حذف شده باشد
ابزار مورد تأیید	فقط دستیار مجاز و مورد تأیید بانک استفاده شود	هیچ ابزار عمومی/مصرف‌کننده در شواهد دیده نشود
کنترل تغییر	اصلاحات مشروط به تیکت + بازبینی همتا + تأییدیه امنیتی باشد	دروازه‌های تأیید در طرح نام‌گذاری شده باشند
برگشت‌پذیری	هر تغییر پیشنهادی یک مسیر بازگشت (rollback) داشته باشد	برای هر اقدام، مسیر بازگشت ذکر شده باشد
پاسخگویی انسانی	پیش از هر اقدام، یک فرد مشخص تأیید کند	چک‌لیست تأیید انسانی موجود و دارای مالک باشد
قابلیت ممیزی	استفاده از AI و تصمیمات مستندسازی شود	لاگ پاک‌سازی و پرامپت برای سابقه حادثه نگهداری شود

(استنباط‌شده از منابع عمومی): سطوح خاص طبقه‌بندی («Internal/Confidential/Restricted»)، دروازه‌های تأیید و مدل «دستیار داخلی مورد تأیید» بازتاب‌دهنده الگوهای رایج کنترلی در بانکداری تحت مقررات هستند (مثلاً سیاست‌های طبقه‌بندی داده، مدیریت تغییر و رویه‌های مدیریت ریسک مدل). این‌ها مقادیر پیش‌فرض نمونه‌وار هستند، نه نردبان HR منتشرشده یا سیاست داخلی هیچ مؤسسه‌ای به‌طور خاص. نام‌ها و آستانه‌ها را با استانداردهای واقعی سازمان خود تطبیق دهید.

x.15 مدیر انتشار (Release Manager)

سناریو

شما مدیر انتشار هستید و هماهنگی یک استقرار تولیدی (production) در شب شنبه را برای پلتفرم بانکداری موبایلیِ خرد بانک بر عهده دارید. این انتشار شامل یک میکروسرویس خدمات پرداخت و یک ماژول احراز هویت مشتری (KYC) است. خط لوله Jenkins CI/CD در مرحله آزمون یکپارچگی (integration-test) شکست می‌خورد و پنجره استقرار تا ۹۰ دقیقه دیگر بسته می‌شود. لاگ خط لوله ۴،۰۰۰ خط دارد و شامل ردگیری پشته (stack traces)، متغیرهای محیطی، رشته‌های اتصال پایگاه‌داده، یک fixture آزمونی حاوی نمونه PII مشتری (نام‌ها، کدهای ملی، شماره‌حساب‌ها) و یک توکن API داخلی است که به‌دلیل یک پرچم اشکال‌زدایی (debug flag) با پیکربندی نادرست چاپ شده است. شما می‌خواهید از دستیار GenAI داخلی مورد تأیید بانک برای تسریع ترياژ استفاده کنید، اما باید تصمیم بگیرید چه چیزی قابل اشتراک‌گذاری است و چه چیزی نیست، و باید توصیه‌ای برای ادامه/توقف (go/no-go) تولید کنید که یک تأییدکننده انسانی بتواند بر اساس آن عمل کند.

وظیفه

1. لاگ Jenkins را پیش از خروج هر محتوایی از محیط کنترل‌شده، با حذف یا پوشاندن (mask) داده‌های حساس برای کمک‌گرفتن از AI آماده کنید.
2. از ابزار GenAI مورد تأیید برای خلاصه‌سازی شکست، شناسایی علت ریشه‌ای محتمل و پیشنهاد گزینه‌های اصلاحی ایمن استفاده کنید.
3. پیشنهادهای تولیدشده توسط AI را به‌روشنی از حقایقی که باید توسط انسان تأیید شوند، تفکیک کنید.
4. یک توصیه انتشار ادامه/توقف (go/no-go) همراه با طرح بازگشت (rollback) و سابقه‌ای از آنچه با ابزار AI به اشتراک گذاشته شده، تولید کنید.

استفاده مجاز از AI

مجاز	غیرمجاز
فقط ابزار GenAI داخلی/سازمانی مورد تأیید	ابزارهای AI عمومی/مصرف‌کننده (ChatGPT رایگان و غیره)
گزیده‌های پاک‌سازی‌شده لاگ، پیام‌های خطا، ردگیری پشته با شناسه‌های پوشانده‌شده	لاگ‌های خام حاوی PII، رازها، توکن‌ها یا رشته‌های اتصال
پیش‌نویس مراحل اصلاحی، خلاصه‌سازی شکست‌ها، تولید چک‌لیست‌های بازگشت	چسباندن کدهای ملی مشتری، شماره‌حساب‌ها یا اعتبارنامه‌ها
تولید پیش‌نویس ارتباطات برای ذی‌نفعان	اجازه به AI برای تأیید یا اجرای خودکار استقرار

خروجی مورد انتظار

یک بسته کوتاه تصمیم انتشار شامل موارد زیر:

- یک خلاصه شکست به زبان ساده در ۵ تا ۱۰ خط.
- یک علت ریشه‌ای محتمل اعلام‌شده، که تا زمان تأیید به‌عنوان فرضیه برچسب‌گذاری شده باشد.
- ۲ تا ۳ گزینه اصلاحی همراه با موازنه‌ها (اصلاح رو به جلو در برابر بازگشت در برابر تعویق).
- یک فهرست صریح «باید توسط انسان تأیید شود» (مثلاً برگشت‌پذیری مهاجرت پایگاه‌داده، تطبیق پرداخت، یکپارچگی داده KYC).
- یک توصیه ادامه/توقف با استدلال مرتبط با بسته‌شدن پنجره زمانی.
- یک طرح بازگشت همراه با شرایط فعال‌سازی و مالک.
- یک یادداشت مدیریت داده که دقیقاً فهرست کند چه چیزی پوشانده شد و چه چیزی با ابزار AI به اشتراک گذاشته شد.

معیارهای ارزیابی

حوزه	نمود وضعیت مطلوب
مدیریت داده	تمام PII، رازها، توکن‌ها و رشته‌های اتصال را پیش از استفاده از AI شناسایی و حذف/پوشانده می‌کند؛ فقط از ابزار مورد تأیید استفاده می‌کند
کیفیت تشخیص	خلاصه شکست دقیق است؛ علت ریشه‌ای قابل قبول و به‌روشنی به‌عنوان فرضیه علامت‌گذاری شده است
حضور انسان در حلقه	پیشنهادهای AI را از حقایق تأییدشده توسط انسان تفکیک می‌کند؛ هیچ تأیید یا استقرار خودکاری وجود ندارد
کیفیت تصمیم	ادامه/توقف موجه، زمان‌بندی‌شده و آگاه به ریسک است؛ طرح بازگشت مشخص با مالکان و محرک‌هاست

حوزه	نمود وضعیت مطلوب
ردیابی‌پذیری	آنچه با AI به اشتراک گذاشته شده و آنچه پوشانده شده را ثبت می‌کند، به‌گونه‌ای مناسب برای ممیزی

سنجه نمره‌دهی (0-5)

نمره	توضیح
0	لاگ خام (شامل PII/رازها) را در AI می‌چسباند، یا از ابزار غیرمجاز استفاده می‌کند؛ هیچ تأیید انسانی وجود ندارد
1	از AI استفاده می‌کند اما داده‌های حساس کلیدی را از قلم می‌اندازد؛ خروجی AI را به‌عنوان حقیقت می‌پذیرد؛ طرح بازگشت ندارد
2	برخی داده‌های حساس را می‌پوشاند؛ خلاصه پایه؛ استدلال ضعیف ادامه/توقف؛ تأیید حداقلی
3	تمام PII/رازهای آشکار را می‌پوشاند؛ خلاصه و فرضیه علت ریشه‌ای معتبر؛ ادامه/توقف را همراه با طرح بازگشت اعلام می‌کند
4	پوشاندن کامل همراه با دلیل؛ تفکیک روشن AI در برابر انسان؛ تصمیم زمان‌بندی‌شده؛ بازگشت مشخص با مالکان/محرک‌ها
5	تمام موارد L4 به‌علاوه یک سابقه مدیریت داده مستند و قابل ممیزی؛ چک‌لیست تأیید مرتبط با ریسک پرداخت/KYC؛ پیشنهاد یک گام پاک‌سازی قابل استفاده مجدد برای انتشارهای آینده

پرچم‌های قرمز

- چسباندن لاگ خام کامل، رشته‌های اتصال، توکن API یا PII مشتری در هر ابزار AI.
- استفاده از یک ابزار AI عمومی/مصرف‌کننده برای داده‌های بانک.
- تلقی علت ریشه‌ای AI به‌عنوان حقیقت تأییدشده و استقرار بدون تأیید انسانی.
- اجازه به AI برای فعال‌سازی یا تأیید استقرار.
- نبود طرح بازگشت، یا طرح بازگشتی بدون مالک یا شرایط فعال‌سازی.
- عبور از پنجره زمانی بدون ثبت هیچ تصمیمی.
- نادیده‌گرفتن پیامدهای تطبیق پرداخت یا یکپارچگی داده KYC.

بررسی‌های ایمنی/انطباق

بررسی	الزام
ابزار مورد تأیید	فقط ابزار GenAI مورد تأیید بانک استفاده می‌شود؛ هیچ داده‌ای از محیط‌های کنترل‌شده خارج نمی‌شود
کمینه‌سازی داده	فقط حداقل محتوای پاک‌سازی‌شده لازم برای ترياژ به اشتراک گذاشته می‌شود
بهداشت رازها	توکن API افشاشده به‌عنوان به‌خطرافتاده تلقی و چرخش (rotate) داده می‌شود؛ پیکربندی نادرست پرچم اشکال‌زدایی به‌عنوان یک نقص ثبت می‌شود
حفاظت از PII	کدهای ملی، نام‌ها و شماره‌حساب‌های مشتری پیش از استفاده از AI پوشانده یا حذف می‌شوند
ردپای ممیزی	سابقه‌ای از پرامپت‌ها، محتوای به‌اشتراک‌گذاشته‌شده و تصمیم نهایی برای بازبینی نگهداری می‌شود
اختیار انسانی	یک تأییدکننده انسانی مشخص تصمیم نهایی ادامه/توقف را می‌گیرد؛ AI صرفاً مشورتی است
کنترل تغییر	تصمیم با سیاست مدیریت تغییر و تفکیک وظایف بانک هم‌راستاست

x.15 مدیر DevOps/فنی (DevOps/Technical Manager)

سطح هدف نردبان برای این آزمون: L4-L5 (کاربر مؤثر پرامپت تا کاربر AI در گردش‌کار)، همراه با نشانگرهای کششی برای L6-L7 (تعیین هنجارهای ایمن، پیشنهاد دروازه‌های تأیید/بازگشت). این آزمون می‌سنجد که آیا یک مدیر DevOps/فنی می‌تواند از GenAI برای تسریع تریاژ حادثه استفاده کند بدون نشت داده‌های حساس و بدون واگذاری قضاوت مهندسی به مدل.

سناریو

شما مسئولیت بخش CI/CD و مهندسی انتشار (release engineering) را برای پلتفرم بانکداری متمرکز (core banking) بر عهده دارید. در ساعت 09:14 یکی از روزهای انتشار، **خط‌لولهٔ Jenkins برای استقرار در محیط production** مربوط به سرویس پرداخت، در مرحلهٔ مهاجرت پایگاه‌داده (database-migration) با خطا متوقف می‌شود. مهندس کشیک (on-call) کل لاگ کنسول را برای شما ارسال می‌کند و درخواست کمک دارد. خط‌لوله باید به‌صورت ایمن از انسداد خارج شود و پیش از بسته‌شدن پنجرهٔ استقرار بعدی در ساعت 11:00، یک رکورد تغییر (change record) باید به‌روزرسانی شود.

لاگ خام Jenkins، در میان خروجی عادی build، شامل موارد زیر است:

- یک **رشتهٔ اتصال به پایگاه‌داده (connection string)** که نام کاربری و رمز عبور درون آن جاسازی شده است (`jdbc:postgresql://prod-pay-db-01.internal:5432/payments?user=svc_deploy&password=...`)
- یک **طرح نام‌گذاری میزبان/IP داخلی** برای سرورهای پرداخت در محیط production
- یک **بخش از شمارهٔ کارت مشتری (PAN)** که از یک test fixture به درون یک خط لاگ اشکال‌زدایی (debug) نشت کرده است (`...card ending 4921, acct 7781`)
- یک **توکن API** برای رجیستری artifact
- یک **stack trace** که به یک مهاجرت ناموفق Flyway (`V47__add_settlement_index.sql`) اشاره دارد که به‌دلیل قفل (lock) روی یک جدول بزرگ، دچار timeout شده است

شما به یک دستیار GenAI سازمانی تأییدشده در سطح enterprise (داده‌ها برای آموزش مدل استفاده نمی‌شوند و محرمانگی قراردادی برقرار است) برای تحلیل لاگ و تهیهٔ پیش‌نویس اقدامات اصلاحی دسترسی دارید.

وظیفه

با استفاده از دستیار GenAI تأییدشده، یک بسته‌ی تریاژ رخداد (incident-triage) تهیه کنید:

1. **پاک‌سازی (Sanitize)** لاگ پیش از ارسال هر بخشی از آن به هوش مصنوعی: هر قطعه داده‌ی حساس را شناسایی و حذف (redact) کنید و توضیح دهید چه چیزی را و چرا حذف کرده‌اید.
2. **خلاصه‌سازی (Summarize)** خرابی به زبان ساده برای یک کمیته‌ی تأیید تغییر (change-approval board) غیرفنی.
3. **شناسایی علت ریشه‌ای محتمل** و رتبه‌بندی علل جایگزین.
4. **پیشنهاد یک اقدام اصلاحی ایمن** شامل مسیر بازگشت (rollback) و توصیه‌ی استقرار/عدم استقرار (deploy/no-deploy) برای بازه‌ی زمانی ۱۱:۰۰.
5. **تفکیک صریح** آنچه هوش مصنوعی پیشنهاد داده از آنچه یک انسان باید پیش از هر اقدام تولیدی (production) راستی‌آزمایی کند.
6. هر **هنجار یا حفاظ (guardrail)** که اضافه می‌کنید تا تیم رخداد بعدی را ایمن‌تر مدیریت کند، یادداشت کنید (شاخص رشد).

استفاده مجاز از هوش مصنوعی

مجاز	غیرمجاز
چسباندن گزیده‌های لاگ **پاک‌سازی‌شده** در دستیار سازمانی تأییدشده	چسباندن لاگ خام حاوی اعتبارنامه‌ها، توکن‌ها، داده‌های PAN/ حساب، یا طرح‌های نام‌گذاری میزبان داخلی
درخواست از هوش مصنوعی برای خلاصه‌سازی، فرضیه‌پردازی درباره علت ریشه‌ای، و تدوین پیش‌نویس گام‌های اصلاحی	استفاده از حساب شخصی/مصرف‌کننده هوش مصنوعی یا هر ابزار خارج از فهرست تأییدشده
درخواست از هوش مصنوعی برای تدوین پیش‌نویس خلاصه هیئت تغییر (change-board) و طرح بازگشت (rollback)	اجازه اعمال متن تولیدشده توسط هوش مصنوعی در محیط production بدون بازبینی انسانی
استفاده از هوش مصنوعی برای پیشنهاد مواردی که باید راستی‌آزمایی شوند	تلقی خروجی هوش مصنوعی به‌عنوان مرجع معتبر درباره وضعیت production یا به‌عنوان تأییدکننده سابقه تغییر (change-record)

خروجی مورد انتظار

یک بسته‌ی غربالگری (triage) شامل موارد زیر:

بخش	معیار کیفیت مطلوب
یادداشت پاک‌سازی	هر مورد حساس یافت‌شده را فهرست می‌کند (رشته‌ی اتصال، رمز عبور، API token، قطعه‌ی PAN/ شماره‌حساب، نام‌های میزبان/IPهای داخلی) و اقدام پنهان‌سازی (redaction) اعمال‌شده را مشخص می‌سازد؛ تأیید می‌کند که مدیریت PAN مطابق سیاست رخداد داده‌ای ارجاع (escalate) شده است
خلاصه‌ی ساده و قابل‌فهم	۳ تا ۵ جمله که برای یک تأییدکننده‌ی غیرفنی قابل درک باشد؛ تأثیر، گستره‌ی آسیب (blast radius) و این واقعیت را بیان می‌کند که هیچ داده‌ی مشتری به‌صورت بیرونی افشا نشده است
ریشه‌یابی + گزینه‌های جایگزین	علت اصلی: مهاجرت Flyway با شناسه‌ی `V47` به‌دلیل قفل روی یک جدول بزرگ دچار وقفه‌ی زمانی (timeout) شده است. گزینه‌های جایگزین رتبه‌بندی‌شده (رقابت بر سر قفل ناشی از یک job همزمان، statement timeout کم‌حجم، ساخت ایندکس روی جدول پرترافیک) به همراه نحوه‌ی تأیید هر یک
رفع مشکل + بازگردانی	گزینه‌های ایمن (اجرای مهاجرت در بازه‌ی کم‌ترافیک با `CONCURRENTLY`/ ایندکس آنلاین، افزایش محتاطانه‌ی lock timeout، تقسیم مهاجرت)؛ مسیر بازگردانی (rollback) صریح؛ تصمیم روشن درباره‌ی استقرار/عدم استقرار برای ساعت 11:00
فهرست نیازمند تأیید انسانی	مواردی مشخص که فقط یک انسان می‌تواند آن‌ها را تأیید کند (وضعیت فعلی قفل پایگاه‌داده، تأخیر همانندسازی (replication lag)، تأیید سابقه‌ی تغییر، تأیید نهایی گستره‌ی آسیب)
پیشنهاد سازوکار حفاظتی (guardrail)	برای مثال، پاک‌سازی‌کننده‌ی لاگ پیش از ارسال، حذف اسرار (secrets) از لاگ‌های build، عدم وجود PAN در داده‌های آزمون (test fixtures)

معیارهای ارزیابی

معیار	آنچه ارزیاب بررسی می‌کند
اولویت با حفاظت از داده	داوطلب داده‌ها را **پیش از** هرگونه تعامل با AI پاک‌سازی می‌کند، نه پس از آن؛ و هر پنج نوع قلم اطلاعاتِ حساس را شناسایی می‌کند

معیار	آنچه ارزیاب بررسی می‌کند
مدیریت PAN/حساب	تشخیص می‌دهد که قطعه افشاشده کارت/حساب یک **محرک رخداد داده‌ای (data-incident)** است، نه صرفاً یک کار پنهان‌سازی، و این موضوع را صریحاً بیان می‌کند
استدلال صحیح ریشه‌یابی	شواهد timeout/lock را به یک علت معتبر مرتبط با migration پیوند می‌دهد؛ و صرفاً گفته‌های AI را تکرار نمی‌کند
اصلاح ایمن	گام‌های برگشت‌پذیر همراه با مسیر rollback پیشنهاد می‌دهد؛ و تلاش مجدد کورکورانه یا غیرفعال‌کردن بررسی‌های ایمنی را توصیه نمی‌کند
شفافیت در حضور انسان در حلقه (Human-in-the-loop)	پیشنهادهای AI را به‌روشنی از حقایق تأییدشده توسط انسان درباره محیط تولید زنده (production) تفکیک می‌کند
تصمیم‌گیری منطقی برای استقرار	تصمیم قابل‌دفاعی برای استقرار/عدم‌استقرار، متناسب با محدودیت زمانی 11:00 و سطح ریسک، اتخاذ می‌کند
کیفیت Prompt	به AI پرامپت‌های ساختاریافته و محدوددامنه می‌دهد؛ و بافت بیش از حد را به اشتراک نمی‌گذارد

سنجه نمره‌دهی (0-5)

نمره	سطح	توضیح
0	بدون مواجهه / ناایمن	لاگ خام را (یا تمایل به این کار دارد) مستقیماً در AI می‌چسباند؛ اطلاعات احراز هویت و PAN را نمی‌بیند؛ خروجی AI را همان راه‌حل می‌پندارد. متناظر با L0-L1.
1	آگاه اما ضعیف	می‌داند که نباید اطلاعات محرمانه در اختیار AI قرار گیرد، اما پاک‌سازی را ناقص انجام می‌دهد (مثلاً رمز عبور را می‌گیرد ولی PAN یا میزبان‌های داخلی را نادیده می‌گذارد)؛ هیچ تفکری درباره بازگشت به وضعیت قبل (rollback) ندارد. متناظر با L2.
2	استفاده ایمن پایه	بیشتر موارد را پاک‌سازی می‌کند و خروجی AI را در نقاط کم‌ریسک راستی‌آزمایی می‌کند، اما به PAN به‌عنوان یک رخداد امنیتی توجه نمی‌کند و راه‌حل اصلاحی مبهمی ارائه می‌دهد. متناظر با L3.
3	اثربخش	پاک‌سازی کامل، تشخیص درست علت ریشه‌ای، راه‌حل اصلاحی منسجم همراه با rollback، فهرست شفاف موارد نیازمند راستی‌آزمایی انسانی، و تصمیم قابل‌دفاع برای استقرار. prompting قدرتمند. متناظر با **L4 (هدف)**.
4	در سطح گردش‌کار	تمام موارد بالا به‌علاوه سنجش فرآیند ترياژ به‌عنوان یک گردش‌کار تکرارپذیر، تشدید (escalate) موضوع PAN مطابق سیاست‌ها، و پیوند دادن راه‌حل اصلاحی به دروازه‌های مدیریت تغییر. متناظر با **L5 (هدف بلندپروازانه)**.
5	تعیین‌کننده هنجار	افزودن سازوکارهای حفاظتی پایدار (پاک‌کننده خودکار پیش از ارسال، حذف اطلاعات محرمانه از لاگ‌های build در سرمنشأ، حذف PAN از داده‌های آزمایشی)، تعریف هنجارهای تیمی، و پیشنهاد خودکارسازی تأیید/rollback. متناظر با L6-L7.

نشانه‌های هشدار

نشانه هشدار	چرا اهمیت دارد
لاگ خام (همراه با connection string، token یا PAN) را در هر ابزار AI کپی می‌کند	نشت مستقیم داده / احتمال افشای داده‌های دارنده کارت
قطعه افشاشده PAN/شماره حساب را صرفاً «بیا پنهانش کن» تلقی کرده و رد می‌شود	یک رخداد قابل‌گزارش در حوزه مدیریت داده را نادیده می‌گیرد
توصیه می‌کند «همان deploy را دوباره اجرا کنید» یا lock/statement timeout را غیرفعال کنید	اقدام ناایمن در محیط production بدون اطمینان از ریشه مشکل
ریشه‌یابی AI را بدون راستی‌آزمایی انسانی از وضعیت زنده پایگاه داده به‌عنوان واقعیت ارائه می‌دهد	اعتماد بیش‌ازحد به مدل؛ تصمیم production بر پایه داده راستی‌آزمایی‌نشده
از یک ابزار AI شخصی یا تأییدنشده «برای سرعت بیشتر» استفاده می‌کند	Shadow-AI / نقض تعهدات قراردادی و محرمانگی
توصیه می‌کند ساعت 11:00 و بدون مسیر بازگشت (rollback) deploy انجام شود	شکست در مدیریت تغییر و ریسک عملیاتی
زمینه (context) کامل سیستم را که AI به آن نیازی ندارد بیش‌ازحد به اشتراک می‌گذارد	افشای غیرضروری معماری داخلی

بررسی‌های ایمنی / انطباق

بررسی	الزام
استفاده صرف از ابزارهای تأییدشده	دستیار GenAI باید در فهرست تأییدشده‌ی سازمانی با شرایط no-training / محرمانگی قرار داشته باشد؛ استفاده از حساب‌های مصرف‌کننده (consumer) مجاز نیست
کمینه‌سازی داده	تنها بخش‌های پاک‌سازی‌شده و محدودشده‌ی لاگ‌ها از محیط خارج می‌شوند؛ هیچ اعتبارنامه، توکن، طرح آدرس‌دهی میزبان داخلی، PAN یا داده‌ی حساب خارج نمی‌شود
مسیر رخداد داده‌ی دارنده‌ی کارت	هرگونه نشت PAN/داده‌ی حساب به درون لاگ‌ها، یک رخداد بالقوه در مدیریت داده تلقی شده و طبق سیاست (هم‌راستا با PCI-DSS) تشدید و ارجاع داده می‌شود، نه آنکه به‌صورت خاموش پاک‌سازی شود
اقتدار انسانی در محیط تولید (production)	هیچ تغییر، تلاش مجدد یا بازگشت در محیط تولید صرفاً بر پایه‌ی خروجی AI اجرا نمی‌شود؛ تأیید سابقه‌ی تغییر (change-record) و راستی‌آزمایی وضعیت زنده همچنان در اختیار انسان باقی می‌ماند
قابلیت ممیزی	بسته‌ی triage، اقدامات پاک‌سازی و تصمیم استقرار/عدم‌استقرار در سابقه‌ی تغییر برای ممیزی ثبت می‌شوند
مدیریت اسرار (Secrets)	اعتبارنامه‌ها و توکن‌هایی که در لاگ‌های build ظاهر می‌شوند، برای حذف از منبع (secrets manager / لاگ‌گیری ماسک‌شده) علامت‌گذاری می‌شوند

توجه: (برداشت‌شده از منابع عمومی) الزامات مربوط به مدیریت PAN و ممیزی فوق، استانداردهای پرانتشار مانند PCI-DSS و رویه‌های رایج مدیریت تغییر را بازتاب می‌دهند؛ این موارد از نردبان داخلی HR یا رویه‌های خصوصی هیچ نهاد خاصی استخراج نشده‌اند.

16. راهنمای ارتقا و مسیر شغلی

این بخش توضیح می‌دهد که چگونه نردبان قابلیت هوش مصنوعی (L0–L10) را در تصمیم‌های روزمرهٔ مربوط به کارکنان به‌کار بگیریم. این متن چنان نگارش شده که یک شریک منابع انسانی (HR partner)، یک مدیر مهندسی، یا یک رهبر فنی بتواند آن را به‌صورت یکپارچه، قابل دفاع و هماهنگ با کنترل‌های مورد انتظار در یک محیط خدمات مالی تحت نظارت به‌کار ببرد.

16.1 نحوهٔ استفاده از چارچوب در فرایندهای منابع انسانی

فرایند	نحوهٔ استفاده از نردبان	معیار «خوب» چیست
ارزیابی سالانه	سطح فعلی را با شواهد ارزیابی کنید؛ آن را با حداقل مورد انتظار برای نقش و ارشدیت فرد مقایسه کنید.	یک امتیاز سطح که با ۲ تا ۳ مصداق عینی پشتیبانی شده باشد، نه یک ادعای شخصی.
آمادگی برای ارتقا	پیش از ارتقا، عملکرد پایدار در سطح هدف را دست‌کم برای یک چرخهٔ ارزیابی لازم بدانید.	شواهدی از عملکرد یک سطح بالاتر، نه صرفاً یک نمونهٔ شانسی.
الزامات استخدام	یک سطح حداقل و یک سطح هدف را در شرح شغل برای هر خانوادهٔ شغلی مشخص کنید.	وظایف مصاحبه‌ای که قابلیت واقعی را می‌سنجند (نگاه کنید به 16.6)، نه آشنایی با ابزار.
برنامهٔ آموزشی	شکاف هر فرد (سطح فعلی در برابر سطح هدف) را به توانمندسازی مشخصی نگاشت کنید.	سطح بعدی مشخص، مهارت‌های مشخص، و یک تاریخ معین برای ارزیابی مجدد.
مسیر شغلی	از نردبان به‌عنوان نقشه‌ای شفاف از الزامات گام بعدی استفاده کنید.	هر فرد بتواند سطح فعلی خود و دو سطح بعدی را نام ببرد.
برنامه‌ریزی قابلیت تیم	سطوح را در یک نقشهٔ حرارتی (heat map) تجمیع کنید؛ نقاط شکست منفرد را بیابید.	هیچ گردش‌کار حیاتی به یک فرد L6 به بالا وابسته نباشد.
برنامه‌ریزی جانشین‌پروری	مشخص کنید چه کسانی می‌توانند نقش‌های L6 به بالا را پوشش دهند؛ یک ذخیرهٔ نیرو بسازید.	دست‌کم دو جانشین معتبر برای هر نقش حیاتی هوش مصنوعی.
انتخاب AI champion	از میان افرادی که عملکرد پایدار در سطح L6 با شواهد ارتقای همتایان دارند انتخاب کنید.	انتخاب champion بر مبنای اثر توانمندسازی، نه حجم خروجی.
ارزیابی مدیران	مدیران را بر مبنای رشد قابلیت تیم و پذیرش ایمن ارزیابی کنید، نه استفادهٔ شخصی.	توزیع سطح تیم بهبود یابد و رخدادها نزدیک به صفر باقی بماند.

16.2 حداقل سطح بر اساس ارشدیت (توصیه)

این‌ها توصیه‌هایی ترکیب‌شده برای یک محیط تحت نظارت‌اند، نه یک استاندارد صنعتی منتشرشده. آن‌ها را با میزان ریسک‌پذیری خود تنظیم کنید.

نقش / ارشدیت	حداقل قابل‌قبول	هدف سالم	جهش
همهٔ کارکنان (هر واحد)	L2 آگاهی	L3 کاربر پایه	L4
مهندس تازه‌کار / همکار	L3	L4 کاربر مؤثر پرامپت	L5
مهندس میانی	L4	L5 کاربر گردش‌کار	L6
مهندس ارشد	L5	L6 کاربر حرفه‌ای	L7

نقش / ارشدیت	حداقل قابل‌قبول	هدف سالم	جهش
رهبر فنی (Tech lead)	L5	L6–L7	L8
معمار (Architect)	L6	L7–L8 طراح راهکار	L9
AI champion	L6	L7	+L8
مدیر مهندسی	L3 شخصی + سابقهٔ اثبات‌شدهٔ رشد تیم	L4 شخصی + ارتقای قابل‌نمایش تیم	—
متخصص امنیت / حاکمیت	L4 + سواد حاکمیتی L10	L5 + مالکیت حاکمیت	L10
متخصص Platform / MLOps	L6	L8	L9 سازندهٔ Agent/ Platform

16.3 شواهد پذیرفته‌شده برای ارتقا

ارتقا باید بر اثرگذاری نمایش‌داده‌شده و قابل‌راستی‌آزمایی استوار باشد، نه بر حجم استفاده از هوش مصنوعی.

سطح مقصد	شواهد قابل‌قبول	آنچه به‌تنهایی اثبات محسوب نمی‌شود
L3	استفادهٔ روزمرهٔ ایمن همراه با گزارش‌های راستی‌آزمایی در وظایف کم‌ریسک.	«من هر روز از آن استفاده می‌کنم.»
L4	نمونه‌های قبل/بعد که نشان دهد تکنیک پرامپت‌نویسی نتیجه‌ای را بهبود داده است.	پرامپت‌های ذخیره‌شده بدون هیچ پیامد مشخص.
L5	یک گردش‌کار چندمرحله‌ای با بهبود اندازه‌گیری‌شدهٔ زمان/کیفیت و یک نقطهٔ کنترل.	یک خروجی تأثیرگذار اما یک‌باره.
L6	هنجارهای مستندشده که همتایان آن را پذیرفته‌اند؛ ارتقای قابل‌اندازه‌گیری هم‌تیمی‌ها.	صرفاً بازدهی شخصی بالا.
L7	یک خودکارسازی تولیدی با دروازه‌های تأیید و بازگشت (rollback)، با تأیید رسمی.	یک نمونهٔ نمایشی یا نمونهٔ اولیهٔ دفترچه‌ای (notebook).
L8	یک راهکار سرتاسری که اعتبارسنجی ریسک مدل را پشت سر گذاشته است.	یک POC کارا اما بدون اعتبارسنجی.
L9	یک agent/platform قابل‌استفادهٔ مجدد که چند تیم با حفاظها (guardrails) از آن استفاده می‌کنند.	یک ابزار داخلی مخصوص یک تیم.
L10	مالکیت استراتژی هوش مصنوعی سازمانی، حاکمیت و مصنوعات ریسک نظارتی.	عنوان شغلی یا سابقهٔ خدمت.

16.4 آنچه نباید انجام داد

- بر مبنای بسامد استفاده از هوش مصنوعی، تعداد پرامپت‌ها، یا توکن‌های مصرف‌شده ارتقا ندهید.
- یک نمونهٔ نمایشی تأثیرگذار را به‌عنوان اثبات سطح پایدار تلقی نکنید.
- افراد را بر اساس ابزارهایی که دسترسی به آن‌ها به ایشان داده نشده، ارزیابی نکنید.
- اجازه ندهید سطوح خوداظهارشده بدون مصنوعات معتبر باقی بمانند.

- خروجی سریع را با خروجی ایمن و اعتبارسنجی‌شده یکسان نپندارید، به‌ویژه در گردش‌کارهای تحت نظارت.
- بر پایهٔ یک پروژه از سطوح نپرید؛ عملکرد پایدار را لازم بدانید.
- نردبان را به ابزار نظارت و پایش تبدیل نکنید؛ این یک نقشهٔ توسعه است.

16.5 پرهیز از تحریف‌های رایج

ریسک	چرا رخ می‌دهد	حفاظ
ارزش‌گذاری بیش از حد استفاده از هوش مصنوعی بدون اثر کسب‌وکاری	اندازه‌گیری فعالیت آسان است؛ اندازه‌گیری اثر نه.	هر ادعای سطح را به یک پیامد کسب‌وکاری یا کیفی گره بزنید، نه به شمارش فعالیت.
جریمه کردن کارکنانی که دسترسی به ابزار ندارند	دسترسی به‌صورت نابرابر توزیع می‌شود.	تنها بر اساس ابزارهایی که رسماً تأمین شده‌اند ارزیابی کنید؛ شکاف‌های دسترسی را به‌عنوان یک مسئلهٔ سازمانی ثبت کنید، نه یک امتیاز شخصی. پس از اعطای دسترسی، مسیری برای ارزیابی فراهم کنید.
اشتباه گرفتن استفاده از هوش مصنوعی با قابلیت هوش مصنوعی	کاربران پرمصرف نیز ممکن است کار راستی‌آزمایی‌نشده و ناایمن تولید کنند.	شواهد قضاوت را لازم بدانید: راستی‌آزمایی، نقاط کنترل، و رد کردن خروجی نامناسب.
انتخاب champion بر مبنای پرسروصدا بودن	دیده‌شدن با قابلیت اشتباه گرفته می‌شود.	بر اساس شواهد ارتقای همتایان و تنظیم هنجارهای ایمن انتخاب کنید (معیارهای L6).

16.6 تمایز میان استفاده از هوش مصنوعی و قابلیت واقعی هوش مصنوعی

قابلیت واقعی با قضاوت نمایان می‌شود، نه با فعالیت. این موارد را بررسی کنید:

- **رفتار راستی‌آزمایی** — آیا فرد پیش از اقدام، خروجی‌ها را در برابر منابع معتبر بررسی می‌کند؟
- **مدیریت خطا** — آیا می‌تواند پاسخی نادرست یا توهم‌آمیز را تشخیص دهد و رد کند؟
- **آگاهی از کنترل** — آیا می‌داند چه داده‌ای را می‌توان و چه داده‌ای را نمی‌توان وارد یک ابزار کرد؟
- **مالکیت پیامد** — آیا می‌تواند استفادهٔ خود از هوش مصنوعی را به یک نتیجهٔ اندازه‌گیری‌شده گره بزند؟
- **انتقال** — آیا می‌تواند تکنیک را به دیگری آموزش دهد و آن آموزش ماندگار شود؟

یک آزمون کاربردی: از فرد بخواهید یک وظیفهٔ واقعی را از ابتدا تا انتها مرور کند و توضیح دهد چرا به هر خروجی هوش مصنوعی اعتماد یا آن را رد کرده است. استفاده، روایتی دربارهٔ سرعت می‌سازد؛ قابلیت، روایتی دربارهٔ قضاوت پدید می‌آورد.

16.7 ارزیابی استفادهٔ ایمن از هوش مصنوعی در محیط تحت نظارت

در یک محیط بانکی، «ایمن بودن» بخشی از شایستگی است، نه یک افزوده. این موارد را در کنار قابلیت، در هر سطح از L2 به بالا ارزیابی کنید.

بُعد	چه چیزی را بررسی کنیم	مرجع
مدیریت داده	هیچ دادهٔ محرمانه، PII، یا دادهٔ تحت نظارت وارد ابزارهای تأییدنشده نشود.	(استنباط‌شده از منابع عمومی) با اصول کمینه‌سازی داده و محرمانگی که در GDPR و قواعد دادهٔ مالی رایج است هم‌خوانی دارد.
ابزار تأییدشده	تنها ابزارهای مجاز برای وظایف کاری استفاده شود.	سیاست داخلی تأیید ابزار (سیاست خودتان را به‌کار ببرید؛ وجود آن را مفروض نگیرید).

بُعد	چه چیزی را بررسی کنیم	مرجع
انسان در حلقه (Human-in-the-loop)	فردی هر تصمیم متأثر از هوش مصنوعی را که بر مشتریان یا ریسک اثر می‌گذارد بازبینی کرده و مالکیت آن را بر عهده دارد.	(استنباط‌شده از منابع عمومی) با انتظارات نهاد ناظر دربارهٔ پاسخگویی و تبیین‌پذیری سازگار است.
اعتبارسنجی ریسک مدل	راهکارها (L8 به بالا) پیش از تولید، اعتبارسنجی مستقل را پشت سر می‌گذارند.	(استنباط‌شده از منابع عمومی) منعکس‌کنندهٔ هدف شیوه‌های مدیریت ریسک مدل مانند رهنمود SR 11-7 از U.S. Federal Reserve / OCC است.
قابلیت ممیزی	تصمیم‌ها، تأییدها و بازگشت‌ها (rollbacks) ثبت می‌شوند.	(استنباط‌شده از منابع عمومی) از ممیزی و بازبینی نظارتی پشتیبانی می‌کند.
دروازه‌های تأیید و بازگشت	خودکارسازی‌ها (L7 به بالا) بدون تأیید رسمی و مسیر بازگشت نمی‌توانند مستقر شوند.	توصیهٔ ترکیب‌شده.

> توجه: مراجع نظارتی فوق چارچوب‌هایی شناخته‌شده و عمومی‌اند که به‌عنوان نمونه‌ای از بهترین شیوه‌ها ذکر شده‌اند. این‌ها **ادعایی** دربارهٔ هیچ سیاست داخلی مشخصی نیستند. پیش از انتشار، هر مرجع را با کنترل‌های خود سازمان جایگزین کنید.

---

17. سطوح هدف ۴ماهه و ۱۲ماهه

این جدول‌ها اهداف واقع‌بینانه‌ای را برای توانمندی در دو افق زمانی تعیین می‌کنند: افق ۴ماهه (یک چرخه ورود به سازمان و توانمندسازی) و افق ۱۲ماهه (یک سال کامل ارزیابی عملکرد). این اهداف **توصیه‌های ترکیبی و تلفیقی** برای یک سازمان فناوری در حوزه خدمات مالیِ تحت نظارت هستند، نه یک معیار منتشرشده. آن‌ها را متناسب با میزان ریسک‌پذیری، روند استقرار دسترسی به ابزارها و خط مبنای اولیه سازمان خود تنظیم کنید.

مفروضات پشت این اهداف: - ابزارها به‌صورت رسمی فراهم شده و یک سیاست استفاده ایمن پیش از آغاز ارزیابی وجود دارد. - «همه کارکنان» شامل واحدهای غیرمهندسی نیز می‌شود؛ نقش‌های مهندسی اهداف بالاتری دارند. - جابه‌جایی یک سطح در هر چرخه برای بیشتر افراد واقع‌بینانه است؛ دو سطح در یک سال تنها برای افراد با عملکرد بالا و با پشتیبانی قابل دستیابی است. - سطوح بالاتر (L7 به بالا) به ساختار سازمانی پشتیبان (حاکمیت، اعتبارسنجی، پلتفرم‌ها) نیاز دارند؛ بنابراین اهداف در آن سطوح، دامنه نقش را بازتاب می‌دهند، نه صرفاً جاه‌طلبی فردی را.

17.1 سطوح هدف بر اساس نقش و سطح ارشدیت

نقش / سطح ارشدیت	نقطه شروع معمول (استنباطی)	هدف ۴ماهه	هدف ۱۲ماهه	منطق
همه کارکنان (غیرمهندسی)	L0–L1	L2 آگاهی	L3 کاربر پایه	همه باید پیش از استفاده، قواعد داده را درک کنند و سپس به استفاده روزمره و ایمن برسند.
مهندس جونیور / همکار	L1–L2	L3 کاربر پایه	L4 کاربر مؤثر در طراحی پرامپت	ابتدا عادت‌های روزانه ایمن شکل بگیرد، سپس مهارت سنجیده در طراحی پرامپت.
مهندس میان‌رده	L2–L3	L4	L5 کاربر گردش‌کار	حرکت از وظایف منفرد به گردش‌کارهای چندمرحله‌ای سنجش‌پذیر.

نقش / سطح ارشدیت	نقطه شروع معمول (استنباطی)	هدف ۴ماهه	هدف ۱۲ماهه	منطق
مهندس ارشد	L3–L4	L5	L6 کاربر حرفه‌ای	ابتدا مالک گردش‌کار شدن، سپس تبدیل شدن به متخصصی که تیم را بالا می‌کشد.
سرپرست فنی (Tech Lead)	L4	L5–L6	L6–L7	رهبری از طریق تعیین هنجارهای ایمن؛ آغاز حمایت از خودکارسازی‌های تأییدشده.
معمار (Architect)	L5	L6	L7–L8 طراح راهکار	طراحی راهکارهایی که می‌توانند از اعتبارسنجی ریسک مدل عبور کنند.
پیشگام هوش مصنوعی (AI Champion)	L5–L6	L6 کاربر حرفه‌ای	L7 مشارکت‌کننده در خودکارسازی	تثبیت بالا کشیدن همتایان، سپس ارائه یک خودکارسازی تحت حاکمیت.
مدیر مهندسی	L2–L3 فردی	L3 فردی + برنامه تیمی	L4 فردی + ارتقای سنجش‌پذیر تیم	مدیران بر اساس رشد تیم و پذیرش ایمن سنجیده می‌شوند، نه بازده فردی.
متخصص امنیت / حاکمیت	L3–L4	L4 + سواد حاکمیتی	L5 + مالکیت حاکمیت (مسیر L10)	توانمند به اندازه ارزیابی استفاده؛ مالک حفاظ‌های استفاده ایمن.
متخصص Platform / MLOps	L5–L6	L6–L7	L8 طراح راهکار / در مسیر L9	خودکارسازی‌ها و پلتفرم‌های اعتبارسنجی‌شده‌ای را می‌سازد که دیگران به آن‌ها متکی‌اند.

17.2 اهداف در سطح سازمان (توزیع)

اهداف فردی در مجموع به یک توزیع سالم‌تر از توانمندی منجر می‌شوند. ارقام زیر اهداف برنامه‌ریزیِ نمونه و گویا هستند، نه یک معیار اندازه‌گیری‌شده.

باند توانمندی	هدف ۴ماهه	هدف ۱۲ماهه	چرا اهمیت دارد
L0–L1 (بدون مواجهه)	کمتر از 20% کارکنان	نزدیک به 0% کارکنان	کاربران ناآگاه مهم‌ترین منبع در معرض ریسک داده هستند.
L2–L3 (آگاه / پایه)	اکثریت همه کارکنان	اکثریت کارکنان غیرمهندسی	استفاده روزمره ایمن به خط مبنا تبدیل می‌شود.
L4–L5 (مؤثر / گردش‌کار)	اکثریت مهندسان	اکثریت مهندسان	بهره‌وری از این باند حاصل می‌شود.
L6 (کاربران حرفه‌ای)	حداقل ۱ نفر در هر تیم	۲ نفر یا بیشتر در هر تیم حیاتی	متخصصان محلی هنجارهای ایمن را تعیین می‌کنند و دیگران را بالا می‌کشند.
L7–L8 (خودکارسازی / راهکار)	یک گروه کوچک و تحت حاکمیت	به اندازه کافی برای پوشش همه گردش‌کارهای حیاتی	اثرگذاری در محیط تولید همراه با اعتبارسنجی و قابلیت بازگشت.
L9–L10 (پلتفرم / راهبرد)	مالکان مشخص وجود دارند	ذخیره ۲ نفر یا بیشتر برای هر نقش حیاتی	از نقاط شکست منفرد در رهبری هوش مصنوعی جلوگیری می‌کند.

17.3 چه چیزی این اهداف را واقع‌بینانه می‌کند (و چه چیزی آن‌ها را بر هم می‌زند)

توانمندساز	اثر	در صورت نبود
ابزارهای فراهم‌شده + سیاست استفاده ایمن	ارزیابی و رشد می‌تواند آغاز شود	اهداف متوقف می‌مانند؛ کارکنان را نمی‌توان منصفانه ارزیابی کرد
یک چرخه توانمندسازی برای هر سطح	جابه‌جایی یک سطحی به امری روزمره تبدیل می‌شود	پیشرفت کند می‌شود؛ انتظار نیمی از جابه‌جایی را داشته باشید
استقرار حاکمیت، اعتبارسنجی و پلتفرم‌ها	اهداف L7 به بالا دست‌یافتنی می‌شوند	سطوح بالا صرفاً آرمانی باقی می‌مانند
پیوند انگیزه‌های مدیران با ارتقای تیم	توزیع به‌طور گسترده بهبود می‌یابد	رشد در چند فرد متمرکز می‌شود
ارزیابی صادقانه خط مبنا	اهداف معتبر هستند	سطوح متورم، اعتماد به چارچوب را از بین می‌برند

یادآوری: همه اهداف عددی توزیع و اهداف هر نقش در بخش 17 **توصیه‌های ترکیبیِ برنامه‌ریزی** هستند. آن‌ها از هیچ نردبان داخلی شرکت یا معیار منتشرشده بیرونی استخراج نشده‌اند. پیش از تعهد به آن‌ها در یک گزارش، آن‌ها را با خط مبنای خود کالیبره کنید.

---

18. برنامه آموزشی بر اساس سطح

این برنامه هر یک از سطوح نردبان (L0/L1 تا L10) را به یک مسیر یادگیری مشخص و عملی متصل می‌کند. هر سطح بر پایه سطح پایین‌تر از خود ساخته می‌شود، بنابراین هیچ فراگیری نباید از یک رده عبور کند و آن را نادیده بگیرد. منابع آموزشی، برنامه‌های واقعی و در دسترس عموم هستند؛ هرجا که نام یک برنامه دقیق باشد، پیوند آن نیز قرار داده شده است. برآوردهای زمانی بر این فرض استوارند که یک فرد شاغل به‌صورت پاره‌وقت و در کنار کار روزانه خود به یادگیری می‌پردازد.

نحوه خواندن جدول‌ها - «سرمایه‌گذاری زمانی» مجموع تلاش لازم برای رسیدن به شایستگی در آن سطح است، نه زمان تقویمی. - «روش ارزیابی» شیوه‌ای است که از طریق آن یک مدیر یا واحد L&D تأیید می‌کند سطح موردنظر واقعاً کسب شده است، نه صرفاً اینکه فرد در دوره حضور یافته باشد. - ردیف‌های امنیت/حاکمیت و بانکی/مقرراتی، دروازه‌های الزامی هستند: تأیید نهایی یک فراگیر در هر سطح، تا زمانی که این موارد با موفقیت طی نشوند امکان‌پذیر نیست، فارغ از سطح مهارت او.

فهرست مرجع تجمیع‌شده‌ای از تمام آموزش‌های عمومی استنادشده، در انتهای این بخش آورده شده است.

L0/L1 — بدون مواجهه

مورد	جزئیات
اهداف یادگیری	درک اینکه هوش مصنوعی مولد چیست و چه نیست؛ شناسایی جاهایی که AI پیش‌تر در ابزارهای روزمره تعبیه شده است؛ درک اینکه چرا استفادهٔ کنترل‌نشده ریسک‌های داده، حریم خصوصی و رفتار حرفه‌ای ایجاد می‌کند.
دوره‌ها / موضوعات الزامی	مسیر «AI fundamentals» در Microsoft Learn؛ دورهٔ «Introduction to Generative AI – Art of the Possible» از AWS؛ جلسهٔ توجیهی سیاست استفادهٔ مجاز از AI (AI Acceptable Use Policy) سازمان.

مورد	جزئیات
آزمایشگاه‌های عملی	هیچ‌کدام الزامی نیست. مرور اختیاری یک ابزار تأییدشده در محیط sandbox با هدایت مربی (مشاهدهٔ صرفاً خواندنی).
تمرین‌های ویژهٔ نقش	تمرین «AI را پیدا کن»: شناسایی اینکه کدام ابزارها در گردش‌کار خود فرد پیش‌تر دارای قابلیت‌های AI هستند.
آموزش امنیت / حاکمیت	الزامی: مبانی طبقه‌بندی داده (عمومی / داخلی / محرمانه / محدود)؛ «هرگز دادهٔ مشتری یا دادهٔ محدود را در ابزار تأییدنشده وارد نکنید.»
آگاهی بانکی / مقرراتی	مروری بر اینکه چرا مؤسسات خدمات مالی تحت مقررات قرار دارند؛ محرمانگی، حفاظت از دادهٔ مشتری و انتظارات رفتاری.
روش ارزیابی	آزمون کوتاه چندگزینه‌ای (قبولی = پاسخ صحیح به همهٔ پرسش‌های مدیریت داده)؛ تأییدیهٔ امضاشدهٔ سیاست استفادهٔ مجاز از AI.
زمان تخمینی موردنیاز	2–4 ساعت.
مسیر پیشنهادی برای سطح بعد	تکمیل کامل ماژول قواعد مدیریت داده، سپس ثبت‌نام در آموزش آگاهی‌بخشی L2 پیش از کار با هر ابزاری.

L2 — آگاهی از هوش مصنوعی (AI Awareness)

مورد	جزئیات
اهداف یادگیری	تشریح نحوه عملکرد مدل‌های زبانی بزرگ (large language models) در سطح مفهومی؛ درک ریسک‌های hallucination، نشت داده (data leakage) و prompt-injection؛ شناخت ابزارهای تأییدشده و قواعد داده پیش از نخستین استفاده.
دوره‌ها / موضوعات الزامی	دوره "Introduction to Generative AI" گوگل (Google Cloud Skills Boost)؛ دوره "Fundamentals of Generative AI" در Microsoft Learn؛ فهرست داخلی ابزارهای تأییدشده و رویه استاندارد عملیاتی (SOP) مدیریت داده.
آزمایشگاه‌های عملی	نخستین جلسه هدایت‌شده در یک ابزار سازمانی تأییدشده، صرفاً روی داده‌های ساختگی/غیرحساس.
تمرین‌های ویژه نقش	دسته‌بندی پنج پرامپت نمونه به‌عنوان ایمن / ناایمن برای ارسال، همراه با استدلال.
آموزش امنیت / حاکمیت	مرور آگاهی‌بخش OWASP Top 10 for LLM Applications (با تمرکز بر prompt injection و افشای اطلاعات حساس)؛ مسیر گزارش‌دهی shadow-AI.
آگاهی بانکی / مقرراتی	محرمانگی داده‌های مشتری، الزامات نگهداری سوابق، و اینکه چرا خروجی هوش مصنوعی ممکن است پیش از هرگونه استفاده بیرونی یا مرتبط با مشتری نیازمند بازبینی انسانی باشد.
روش ارزیابی	آزمون سناریومحور درباره مدیریت داده؛ تأیید مربی مبنی بر اینکه فراگیر می‌تواند فهرست ابزارهای تأییدشده و دسته‌های داده‌ای «غیرقابل ارسال» را از حفظ بیان کند.
زمان تخمینی موردنیاز	4 تا 6 ساعت.
مسیر پیشنهادی سطح بعدی	آغاز استفاده روزمره کم‌ریسک و تحت نظارت؛ پس از نهادینه‌شدن قواعد داده، پیشرفت به L3.

L3 — کاربر پایه هوش مصنوعی (Basic AI User)

مورد	جزئیات
اهداف یادگیری	استفاده ایمن از یک ابزار تأییدشده برای کارهای روتین و کم‌ریسک؛ همواره راستی‌آزمایی خروجی پیش از استفاده؛ تشخیص اینکه یک کار چه زمانی برای واگذاری به هوش مصنوعی بیش از حد حساس است.
دوره‌ها / موضوعات الزامی	"AWS "Generative AI Learning Plan (ماژول‌های مبنایی)؛ آموزش آشناسازی آکادمی فروشنده برای ابزار اصلی سازمان (برای مثال، آموزش پذیرش Microsoft Copilot).
آزمایشگاه‌های عملی	انجام ۸ تا ۱۰ کار واقعی کم‌ریسک (پیش‌نویس یادداشت‌های داخلی، خلاصه‌سازی اسناد عمومی) همراه با یک چک‌لیست راستی‌آزمایی.
تمرین‌های مختص نقش	نگهداری یک «گزارش راستی‌آزمایی» یک‌هفته‌ای: هر خروجی هوش مصنوعی که در کار استفاده می‌شود، به‌علاوه بررسی انجام‌شده پیش از اتکا به آن.
آموزش امنیت / حاکمیت	انضباط در راستی‌آزمایی خروجی؛ ارجاع‌دهی و بررسی منبع؛ تشخیص confabulation (مطابق دسته‌بندی ریسک GenAI در NIST AI 600-1).
آگاهی بانکی / مقرراتی	انتظار «انسان در حلقه» (Human-in-the-loop) برای هر چیزی که با مشتریان، امور مالی یا تصمیم‌گیری‌ها سر و کار دارد؛ عدم اتکا به خروجی راستی‌آزمایی‌نشده هوش مصنوعی.
روش ارزیابی	بازبینی گزارش راستی‌آزمایی توسط یک مدیر یا همتای سطح L4 به بالا؛ نمایش گرفتن دست‌کم یک خطای هوش مصنوعی در حین ارزیابی.
زمان تخمینی موردنیاز	۸ تا ۱۲ ساعت طی ۲ تا ۳ هفته.
مسیر پیشنهادی برای سطح بعدی	حرکت به سمت تکنیک prompting هدفمند؛ ثبت‌نام در دوره prompting سطح L4.

L4 — کاربر مؤثر پرامپت (Effective Prompt User)

مورد	جزئیات
اهداف یادگیری	دستیابی به نتایج قابل اتکا و تکرارپذیر از طریق پرامپت‌نویسی هدفمند (نقش، زمینه، محدودیت‌ها، مثال‌ها، قالب خروجی)؛ بهبود و پالایش نظام‌مند و تکراری.
دوره‌ها / موضوعات موردنیاز	دوره‌های Google Cloud با عنوان «Introduction to Prompt Engineering» / «Prompt Design in Vertex AI»؛ ماژول‌های آموزشگاه مهندسی پرامپت ارائه‌دهندگان؛ مستندات مهندسی پرامپت Anthropic.
آزمایشگاه‌های عملی	ساخت و آزمون ۳ تا ۵ قالب پرامپت قابل استفاده مجدد برای وظایف تکرارشونده‌ی خودِ فراگیر؛ سنجش کیفیت پیش و پس از قالب‌سازی.
تمرین‌های ویژه‌ی نقش	انتخاب یک پرامپت ضعیف و یک پرامپت خوب برای یک وظیفه‌ی یکسان؛ مستندسازی اینکه چرا نمونه‌ی خوب خروجی بهتر و ایمن‌تری تولید می‌کند.
آموزش امنیت / حاکمیت	آگاهی عملی نسبت به prompt-injection؛ پرهیز از جاسازی داده‌های حساس در پرامپت‌های ذخیره‌شده یا به‌اشتراک‌گذاشته‌شده.
آگاهی بانکی / مقرراتی	اطمینان از اینکه خروجی‌های حاصل از پرامپت همچنان قابل ممیزی و قابل توضیح باقی می‌مانند؛ عدم تثبیت تصمیمات سیاستی در پرامپت‌ها بدون بازبینی.

مورد	جزئیات
روش ارزیابی	ارائه‌ی یک کتابخانه‌ی کوچک پرامپت همراه با شواهد کیفیت پیش/پس؛ بازبینی همتا توسط یک کاربر سطح L5 به بالا که تکرارپذیری را تأیید می‌کند.
زمان تخمینی موردنیاز	۱۲ تا ۱۶ ساعت در طول ۳ تا ۴ هفته.
مسیر پیشنهادی برای سطح بعدی	آغاز زنجیره‌سازی پرامپت‌ها در قالب جریان‌های کاری چندمرحله‌ای؛ ثبت‌نام در آموزش جریان کاری سطح L5.

L5 — کاربر هوش مصنوعی در سطح گردش‌کار (Workflow AI User)

مورد	جزئیات
اهداف یادگیری	تعبیه هوش مصنوعی در گردش‌کارهای چندمرحله‌ای و چندابزاری؛ سنجش زمان صرفه‌جویی‌شده و تأثیر بر کیفیت؛ استانداردسازی یک گردش‌کار که دیگران بتوانند از آن استفاده مجدد کنند.
دوره‌ها / موضوعات الزامی	ماژول‌های گردش‌کار هوش مصنوعی Microsoft Learn Copilot / Power Platform؛ ماژول‌های سطح متوسط «Generative AI Learning Plan» شرکت AWS؛ راهنماهای یکپارچه‌سازی ابزارهای داخلی.
آزمایشگاه‌های عملی	ساخت یک گردش‌کار سرتاسری (مثلاً دریافت ← پیش‌نویس ← بازبینی ← نهایی‌سازی) با سنجش معیارهای پایه در برابر حالت با کمک هوش مصنوعی.
تمرین‌های ویژه نقش	مستندسازی یک گردش‌کار واقعی همراه با نقاط تحویل، ایست‌بازرسی‌های راستی‌آزمایی و یک دلتای بهره‌وری اندازه‌گیری‌شده.
آموزش امنیت / حاکمیت	ترسیم جریان داده برای گردش‌کار؛ اطمینان از اینکه هیچ داده محدودشده‌ای از یک مرز تأییدنشده عبور نمی‌کند؛ ثبت رویداد و قابلیت ردیابی.
آگاهی بانکی / مقرراتی	حفظ ردپای حسابرسی در سراسر مراحل؛ اطمینان از حذف‌نشدن ایست‌بازرسی‌های کنترلی به‌بهای سرعت؛ آگاهی از تفکیک وظایف (segregation of duties).
روش ارزیابی	ارائه یک گردش‌کار همراه با معیارهای مستند و بازبینی جریان داده؛ تأیید نهایی مستلزم ارائه شواهدی از بهره‌وری اندازه‌گیری‌شده و یک نقطه کنترلی دست‌نخورده است.
زمان تخمینی موردنیاز	20 تا 30 ساعت طی 4 تا 6 هفته.
مسیر پیشنهادی سطح بعدی	آغاز مربیگری همتایان و تعیین هنجارهای تیمی؛ ثبت‌نام در مسیر رهبری و استانداردهای L6.

L6 — کاربر حرفه‌ای AI (AI Power User)

مورد	شرح
اهداف یادگیری	ایفای نقش به‌عنوان متخصص AI تیم؛ تعیین هنجارهای ایمن و کاربردی برای استفاده؛ ارتقای توانمندی دیگران؛ ارزیابی مسئولانه ابزارهای جدید.

مورد	شرح
دوره‌ها / موضوعات مورد نیاز	برنامه پیشرفته generative AI در AWS Skills Builder؛ برنامه‌های مسیر «champion»/مربی‌گری ارائه‌دهندگان (مثلاً توانمندسازی Microsoft Copilot champion)؛ مهارت‌های تسهیلگری و مربی‌گری (coaching).
آزمایشگاه‌های عملی	برگزاری یک جلسه توانمندسازی داخلی؛ ارزیابی آزمایشی یک قابلیت جدید از ابزارهای تأییدشده و تدوین یک توصیه کوتاه go/no-go.
تمرین‌های ویژه نقش	تألیف راهنمای عملیاتی (playbook) استفاده از AI برای تیم (ابزارهای تأییدشده، الگوهای prompt، قواعد راستی‌آزمایی) و آماده‌سازی (onboard) دو همکار از سطح L2/L3 به بالا.
آموزش امنیت / حاکمیت	OWASP Top 10 for LLM Applications (به‌طور کامل، نه صرفاً در حد آگاهی)؛ مرور چهار کارکرد اصلی NIST AI RMF (Govern، Map، Measure، Manage).
آگاهی بانکی / مقرراتی	چگونگی انطباق رویه‌های تیم با سیاست‌های سازمان و انتظارات نهاد ناظر؛ مسیرهای ارجاع (escalation) برای استفاده‌های پرریسک؛ پاسخ‌گویی رفتاری به‌عنوان تعیین‌کننده هنجار.
روش ارزیابی	شواهد ارتقای یک سطحی دو همکار تحت مربی‌گری؛ بازبینی playbook تیم توسط حاکمیت/L8+؛ ارائه یک جلسه توانمندسازی.
زمان تخمینی مورد نیاز	۳۰ تا ۴۰ ساعت در طول ۶ تا ۸ هفته به‌علاوه مربی‌گری مستمر.
مسیر پیشنهادی برای سطح بعدی	حرکت از استفاده به‌سمت ساخت اتوماسیون‌های تأییدشده؛ ثبت‌نام در مسیر اتوماسیون و دروازه‌های تأیید (approval-gates) سطح L7.

L7 — مشارکت‌کننده در خودکارسازی هوش مصنوعی (AI Automation Contributor)

مورد	جزئیات
اهداف یادگیری	ساخت خودکارسازی‌های هوش مصنوعی در سطح محصول (production) که با تأییدیه صریح، پایش (monitoring) و دروازه‌های بازگشت (rollback) اجرا می‌شوند؛ درک پیامدهای آن بر SDLC و کنترل تغییر (change-control).
دوره‌ها / موضوعات الزامی	دورهٔ AWS با عنوان «Generative AI Learning Plan for Developers (includes labs)»؛ مسیرهای توسعه‌دهندهٔ Azure AI / Copilot Studio در Microsoft Learn؛ آموزش داخلی مدیریت تغییر و CI/CD.
آزمایشگاه‌های عملی	ساخت یک خودکارسازی کوچک در یک محیط ایزوله (sandbox) همراه با بازگشت آزمایش‌شده و یک کلید قطع اضطراری (kill switch)؛ گذراندن یک بازبینی همتا (peer review) از کد/امنیت پیش از هرگونه مسیر production.
تمرین‌های ویژهٔ نقش	تهیهٔ یک runbook که حالت‌های شکست، گام‌های بازگشت، هشدارهای پایش و تأییدیهٔ نهایی تأییدکننده را برای یک خودکارسازی پوشش دهد.
آموزش امنیت / حاکمیت	به‌کارگیری OWASP Top 10 for LLM Applications بر سامانه‌های ساخته‌شده (مدیریت ناایمن خروجی، اختیار بیش از حد (excessive agency)، زنجیرهٔ تأمین)؛ مدیریت اسرار (secrets management)؛ دسترسی مبتنی بر کمترین امتیاز (least-privilege).
آگاهی بانکی / مقرراتی	کنترل‌های تأیید تغییر و تفکیک وظایف (segregation-of-duties)؛ الزامات قابلیت حسابرسی و ثبت رخداد (logging)؛ آگاهی از اینکه هوش مصنوعی در سطح production ممکن است مشمول سیاست‌های ریسک مدل (model-risk) و ریسک عملیاتی (operational-risk) باشد.

مورد	جزئیات
روش ارزیابی	بازبینی آمادگی برای production: آزمایش بازگشت، نمایش عملی دروازه‌ها، گذراندن بازبینی امنیتی، و تأیید runbook توسط هیئت تغییر (change board).
زمان تخمینی موردنیاز	۴۰ تا ۶۰ ساعت در طول ۸ تا ۱۰ هفته.
مسیر پیشنهادی سطح بعد	حرکت از خودکارسازی‌های منفرد به سمت راهکارهای طراحی‌شده؛ ثبت‌نام در مسیر طراحی راهکار و ریسک مدل سطح L8.

L8 — طراح راهکار هوش مصنوعی (AI Solution Designer)

مورد	جزئیات
اهداف یادگیری	معماری راهکارهای هوش مصنوعی به‌صورت سرتاسری (داده، مدل، یکپارچه‌سازی، کنترل‌ها و پایش) به‌گونه‌ای که از اعتبارسنجی رسمی ریسک مدل عبور کنند.
دوره‌ها / موضوعات الزامی	برنامه AWS با عنوان "Generative AI Developer Advanced Learning Plan"؛ تخصص‌های معماری ابری در حوزه هوش مصنوعی (مسیرهای AWS / Google Cloud / Azure AI engineer)؛ مبانی مدیریت ریسک مدل.
آزمایشگاه‌های عملی	تولید یک طراحی کامل راهکار به‌همراه مدل تهدید، برنامه ارزیابی/آزمون، طراحی پایش و طراحی نظارت انسانی برای یک مورد کاربرد واقعی.
تمرین‌های ویژه نقش	تدوین بسته اعتبارسنجی: کاربرد مورد نظر، محدودیت‌ها، تبارشناسی داده، نتایج ارزیابی، بررسی‌های سوگیری/انصاف و برنامه پایش مستمر.
آموزش امنیت / حاکمیت	به‌کارگیری NIST AI RMF و NIST AI 600-1 GenAI Profile بر یک راهکار طراحی‌شده (نگاشت بیش از ۱۲ دسته ریسک GenAI به اقدامات کاهنده)؛ مدل‌سازی تهدید OWASP LLM.
آگاهی بانکی / مقرراتی	انتظارات مدیریت ریسک مدل (اعتبارسنجی، مستندسازی، بازنگری مستقل و پایش مستمر)؛ الزامات تبیین‌پذیری و چالش‌گری؛ هم‌راستایی با قوانین اقامت داده و حریم خصوصی.
روش ارزیابی	عبور طراحی راهکار از یک بازنگری اعتبارسنجی ریسک مدل (شبیه‌سازی‌شده یا واقعی) توسط بازنگر مستقل؛ مستندات آماده ممیزی تشخیص داده شوند.
زمان تخمینی موردنیاز	۶۰ تا ۹۰ ساعت طی ۱۰ تا ۱۴ هفته.
مسیر پیشنهادی سطح بعد	حرکت از راهکارها به‌سمت پلتفرم‌های قابل استفاده مجدد؛ ثبت‌نام در مسیر agent/platform سطح L9.

L9 — سازنده پلتفرم/Agent هوش مصنوعی (AI Agent/Platform Builder)

مورد	جزئیات
اهداف یادگیری	ساخت پلتفرم‌های Agent قابل استفاده مجدد و guardrailهای مشترکی که تیم‌های متعدد بتوانند به‌صورت ایمن بر روی آن‌ها توسعه دهند؛ استانداردسازی کنترل‌ها، ارزیابی و observability در سطح کل سازمان.

مورد	جزئیات
دوره‌ها / موضوعات الزامی	آموزش پیشرفته Agent و orchestration (آکادمی‌های Agent عرضه‌کنندگان، AWS Bedrock Agents، سرویس‌های Azure AI agent)؛ Platform Engineering و SRE برای هوش مصنوعی؛ پروفایل‌های ریسک agentic.
آزمایشگاه‌های عملی	راه‌اندازی یک مؤلفه پلتفرم Agent دارای guardrail (اعمال سیاست، logging، چارچوب ارزیابی، محدودیت نرخ/دامنه) که دست‌کم یک تیم پایین‌دستی از آن استفاده کند.
تمرین‌های ویژه نقش	تعریف استانداردهای guardrail در سطح کل پلتفرم و یک دروازه حاکمیتی self-service که سایر سازندگان پیش از استقرار باید از آن عبور کنند.
آموزش امنیت / حاکمیت	مفاهیم پروفایل agentic از Cloud Security Alliance / NIST AI RMF (اختیار بیش‌ازحد، ریسک استفاده از ابزار، محدودیت‌های خودمختاری)؛ کنترل‌های OWASP LLM در سطح پلتفرم؛ امنیت زنجیره تأمین و وابستگی‌ها.
آگاهی بانکی / مقرراتی	چارچوب‌های کنترلی سازمانی، تاب‌آوری عملیاتی، ریسک شخص ثالث/زنجیره تأمین مدل، و شواهد حسابرسی یکپارچه در میان همه مستأجران (tenants) پلتفرم.
روش ارزیابی	بازنگری مستقل معماری و امنیت پلتفرم؛ شواهدی مبنی بر اینکه تیم‌های پایین‌دستی به‌صورت پیش‌فرض guardrailها را به ارث می‌برند؛ دروازه حاکمیتی مستند و در حال استفاده.
زمان تخمینی موردنیاز	بیش از ۹۰ تا ۱۲۰+ ساعت طی ۳ تا ۴ ماه، همراه با مالکیت مستمر پلتفرم.
مسیر پیشنهادی سطح بعدی	گذار از ساختن به مالکیت استراتژی و حاکمیت؛ ثبت‌نام در مسیر استراتژی/حاکمیت L10.

L10 — رهبر راهبرد/حاکمیت هوش مصنوعی (AI Strategy/Governance Leader)

مورد	جزئیات
اهداف یادگیری	مالکیت راهبرد هوش مصنوعی در سطح سازمان، مدل عملیاتی حاکمیت و وضعیت ریسک تنظیمی؛ تعیین سیاست‌ها، اشتهای ریسک و اولویت‌های سرمایه‌گذاری.
دوره‌ها / سرفصل‌های الزامی	برنامه‌های اجرایی حاکمیت و ریسک هوش مصنوعی؛ NIST AI RMF (چارچوب کامل، با تمرکز عمیق بر کارکرد Govern)؛ رصد افق تنظیم‌گری (مانند EU AI Act و رهنمودهای ریسک مدل در سطح صنعت)؛ جلسات توجیهی ریسک هوش مصنوعی در سطح هیئت‌مدیره.
آزمایشگاه‌های عملی	مبتنی بر آزمایشگاه نیست. هدایت طراحی یا بازنگری چارچوب حاکمیت هوش مصنوعی سازمان و بیانیه اشتهای ریسک.
تمرین‌های ویژه نقش	تدوین سیاست هوش مصنوعی سازمان، منشور کمیته حاکمیت و روایت ریسک هوش مصنوعی برای ارائه به نهاد ناظر.
آموزش امنیت / حاکمیت	همسوسازی سازمانی NIST AI RMF، OWASP LLM و مدیریت ریسک مدل در قالب یک چارچوب کنترلی واحد؛ طراحی اطمینان‌بخشی خط سوم/حسابرسی.
آگاهی بانکی / تنظیمی	پاسخگویی مستقیم در قبال انطباق با مقررات، تعامل با نهاد ناظر، تعهدات رفتاری و انصاف، و تاب‌آوری عملیاتی هوش مصنوعی در کل سازمان.
روش ارزیابی	تصویب چارچوب حاکمیتی آماده ارائه به هیئت‌مدیره/نهاد ناظر؛ تأیید کفایت کنترل‌ها از طریق حسابرسی مستقل یا بازبینی نظارتی؛ پذیرش قابل‌سنجش سیاست در سراسر نردبان.

مورد	جزئیات
زمان تخمینی موردنیاز	مسئولیت اجرایی مستمر؛ یادگیری ساختاریافته حدود 40 تا 60 ساعت در سال به‌علاوه پایش مستمر تنظیمی.
مسیر پیشنهادی سطح بعدی	سطح پایانی. تمرکز به سمت رصد مستمر تنظیم‌گری بیرونی، گزارش‌دهی به هیئت‌مدیره و بلوغ‌بخشی به برنامه معطوف می‌شود.

فهرست مرجع دوره‌های آموزشی عمومی

برنامه	سطوح متناسب	ملاحظات
Microsoft Learn — مسیرهای مبانی AI و Copilot	L0–L7	رایگان؛ از سطح آگاهی تا توسعه‌دهنده و اتوماسیون Copilot Studio را پوشش می‌دهد.
Google Cloud Skills Boost — Generative AI / Prompt Design	L2–L8	از مفاهیم پایه تا مسیرهای طراحی prompt و معماری در Vertex AI.
AWS Skill Builder — برنامه‌های یادگیری Generative AI	L3–L9	شامل «Generative AI Learning Plan for Developers (includes labs)» و «Generative AI Developer: Advanced Learning Plan». بیش از ۱،۰۰۰ منبع رایگان.
AWS AI & ML Scholars 2026	L2–L5	برنامه‌ای ساختاریافته در مسیر دریافت گواهی AWS Certified AI Practitioner؛ مرحله Challenge در بازه مارس تا ژوئن 2026.
OWASP Top 10 for LLM Applications	L2 (آگاهی) → L9 (کاربردی)	برنامه درسی اصلی امنیت؛ عمق آن با افزایش سطح بیشتر می‌شود.
NIST AI Risk Management Framework	L6–L10	کارکردهای Govern/Map/Measure/Manage؛ بنیان حاکمیت.
NIST AI 600-1 — Generative AI Profile	L8–L10	بیش از ۱۲ دسته ریسک ویژه GenAI و بیش از ۴۰۰ اقدام پیشنهادی.
آکادمی‌های AI شرکت‌های فروشنده (Microsoft Copilot champion، AWS Bedrock، Azure AI، Google Cloud)	L4–L9	مسیرهای توانمندسازی ویژه ابزار، تربیت champion و ساخت agent.

(برگرفته از منابع عمومی): نگاشت دوره‌های عمومی مشخص به سطوح نردبان داخلی، توصیه‌ای است که از محتوای برنامه‌های در دسترس عموم استنتاج شده است، نه نگاشتی منتشرشده از سوی فروشنده به نردبان. روال‌های اجرایی استاندارد (SOP) داخلی، فهرست ابزارهای تأییدشده و سیاست ریسک مدل شرکت که در جدول‌ها به آن‌ها ارجاع داده شده، مختص سازمان هستند و باید توسط خود شرکت تأمین شوند؛ در اینجا ساخته یا جعل نشده‌اند.

منابع: - AWS Skill Builder — Generative AI - AWS AI & ML Scholars 2026 - NIST AI Risk Management Framework

19. ملاحظات حاکمیت و ریسک

نردبان قابلیت برای مهندسانِ مجهز به هوش مصنوعی، خود یک **سطح کنترل** نیز هست. همان معیاری که در هر سطح تعریف می‌کند «خوب بودن چه شکلی است»، می‌تواند حفاظ‌های موردنیاز یک سازمان خدمات مالی تحت نظارت را در دل خود بگنجاند. این بخش، مهم‌ترین ریسک‌های هوش مصنوعی را به رفتارهای نردبان و به چهار چارچوب بیرونی که رهبران مهندسی و ریسک پیشاپیش با آن‌ها آشنا هستند نگاشت می‌کند:

- **OWASP LLM Top 10 (2025)** — تهدیدهای لایه‌ی کاربرد برای سامانه‌های LLM.
- **NIST AI Risk Management Framework (AI RMF 1.0)** — چهار کارکرد Govern, Map, Measure, Manage.
- **ISO/IEC 42001:2023** — استاندارد سامانه‌ی مدیریت هوش مصنوعی (AIMS) که همانند ISO 27001 ممیزی می‌شود.
- **SR 11-7** — راهنمای میان‌نهادی مدیریت ریسک مدل در ایالات متحده (Federal Reserve / OCC) که مبنای اعمال ناظران بر مدل‌ها در حوزه‌ی بانکداری است.

> یادداشت دامنه: SR 11-7 ناظر بر مدل‌هایی است که برای تصمیم‌های کسب‌وکاری به کار می‌روند. یک دستیار کدنویسی هوش مصنوعی عموماً یک ابزار توسعه است، نه یک مدل تصمیم‌گیری. اما هنگامی که منطقِ تولیدشده توسط هوش مصنوعی وارد یک مدل، یک کنترل یا یک مسیر تصمیم‌گیری روبه‌روی مشتری می‌شود، انتظارات SR 11-7 (اعتبارسنجی، مستندسازی، پایش مستمر) به آن خروجی تعلق می‌گیرد. دستیار را به‌منزله‌ی ابزاری درون‌دامنه تلقی کنید و آنچه را تولید می‌کند به‌منزله‌ی کد مدلِ بالقوه درون‌دامنه. (برداشت‌شده از منابع عمومی.)

19.1 نردبان چگونه حاکمیت را در خود می‌گنجاند

نردبان باید رفتار ایمن را به یک **شرط ترفیع** بدل کند، نه یک سیاست جانبی. سه اصل:

1. **اعتبارسنجی متناسب با سطح رشد می‌کند.** سطوح پایین‌تر تنها مجازند با بازبینی کامل انسانی از هوش مصنوعی استفاده کنند؛ به سطوح بالاتر اعتماد می‌شود که کنترل‌ها را طراحی کنند، استانداردهای بازبینی را تعیین کنند و در سراسر یک تیم پاسخگوی خروجی هوش مصنوعی باشند.
2. **دسترسی به ابزار و داده متناسب با سطح رشد می‌کند و مبتنی بر کمینه‌ی امتیاز است.** دسترسی به داده‌ی تولید، رازها و خودکارسازی خودمختار بر پایه‌ی قضاوتِ اثبات‌شده اعطا می‌شود، نه بر پایه‌ی کنجکاوی.
3. **پاسخگویی هرگز به مدل منتقل نمی‌شود.** در هر سطح، یک انسانِ مشخص مالک نتیجه‌ی ادغام‌شده است. «هوش مصنوعی آن را نوشت» هرگز دفاع قابل‌قبولی نیست.

19.2 دفتر ثبت ریسک

ریسک	چرا اهمیت دارد	نردبان چگونه آن را کاهش می‌دهد	HR / مدیران چگونه آن را ارزیابی می‌کنند
توهم‌زایی (APIها، حقایق و ارجاعات ساختگی)	کد یا استدلال نادرست با لحنی مطمئن منتشر می‌شود؛ در منطق تحت نظارت پرهزینه است. نگاشت به OWASP LLM09 Misinformation؛ NIST Measure؛ اعتبارسنجی خروجی در SR 11-7.	سطوح پایین‌تر: هر خروجی هوش مصنوعی پیش از استفاده در برابر منبع معتبر راستی‌آزمایی می‌شود. سطوح بالاتر: بررسی‌های خودکار (تست‌ها، بررسی نوع، ارجاع متقابل مستندات) می‌سازند و عادت‌های راستی‌آزمایی را به تازه‌واردها می‌آموزند.	به دنبال شواهدی باشید که مهندس خطاهای هوش مصنوعی را گرفته/اصلاح کرده است. بازبینی کنید که آیا ادعاهای موجود در PRها و مستندات منبع دارند. کدِ ادغام‌شده‌ی مجهز به هوش مصنوعی را برای وابستگی‌های ساختگی نمونه‌برداری کنید.

ریسک	چرا اهمیت دارد	نردبان چگونه آن را کاهش می‌دهد	HR / مدیران چگونه آن را ارزیابی می‌کنند
نشت داده (الصاق داده‌ی حساس در پرامپت‌ها)	ارسال PII مشتری، داده‌ی حساب یا کد محرمانه به مدل‌های بیرونی می‌تواند قانون حریم خصوصی و قراردادها را نقض کند. OWASP LLM02 Sensitive Information Disclosure؛ Govern NIST/Manage؛ کنترل‌های داده‌ی ISO 42001؛ حاکمیت داده‌ی SR 11-7.	نردبان آگاهی از طبقه‌بندی داده را به سطح گره می‌زند؛ تنها ابزارها/نقاط‌پایانی تأییدشده و مجاز پذیرفته می‌شوند. سطوح بالاتر تعریف می‌کنند چه چیزی مجاز/غیرمجاز برای ورود به پرامپت است و آن را در بازبینی اعمال می‌کنند.	تأیید کنید مهندسان تنها از ابزارهای مجاز استفاده می‌کنند و طبقه‌بندی‌های داده را می‌دانند. وجود هرگونه PII/راز در لاگ‌های پرامپت (جایی که ثبت لاگ مجاز است) را بررسی کنید. ارزیابی کنید که آیا پیش از پرامپت‌نویسی داده‌ها را پنهان‌سازی می‌کنند.
تزریق پرامپت (دستورات مخرب در محتوای واکشی‌شده، تیکت‌ها و مخازن)	ورودی غیرقابل‌اعتماد می‌تواند یک عامل را به سمت برون‌بری داده یا اقدامات ناایمن منحرف کند. OWASP LLM01 Prompt Injection (ریسک نخست)؛ NIST Map/Manage.	سطوح بالاتر همه‌ی محتوای بیرونی را غیرقابل‌اعتماد تلقی می‌کنند، دامنه‌ی ابزار عامل را محدود می‌کنند و لاگ‌های اقدام عامل را بازبینی می‌کنند. به سطوح پایین‌تر عامل‌های خودمختار و تغذیه‌شده از اینترنت اعطا نمی‌شود.	از داوطلبان بخواهید بردارهای تزریق را در یک سناریو شناسایی کنند. بررسی کنید که گردش‌کارهای عامل‌محور از دامنه‌ی ابزار مبتنی بر کمینه‌ی امتیاز استفاده می‌کنند. مدیریت رخداد هرگونه تلاش تزریق را بازبینی کنید.
اعتماد بیش‌ازحد / سوگیری خودکارسازی	مهندسان خروجی هوش مصنوعی را بدون موشکافی می‌پذیرند؛ مهارت‌ها تحلیل می‌رود؛ خطاها انباشته می‌شوند. OWASP LLM09؛ NIST Measure (عوامل انسانی)؛ «چالش اثربخش» در SR 11-7.	نردبان به چالش اثربخشِ خروجی هوش مصنوعی پاداش می‌دهد و به‌صراحت به سرعتِ پذیرشِ خام پاداش نمی‌دهد. سطوح بالاتر تردید سالم را الگوسازی می‌کنند.	در جلسات یک‌به‌یک کاوش کنید: «اخیراً کِی هوش مصنوعی شما را به بیراهه برد و چطور آن را گرفتید؟» ناتوانی در پاسخ یک نشانه‌ی هشدار است. مراقب افول مبانی پنهان‌شده پشت استفاده از ابزار باشید.
ارزیابی سوگیرانه (قضاوت افراد بر پایه‌ی حجم هوش مصنوعی، نه قضاوت)	پاداش‌دادن به سرعت خروجی به جای درستی، رفتار بد را ترویج می‌کند و می‌تواند در ترفیع تأثیر نابرابر ایجاد کند. Govern NIST (انصاف)؛ ISO 42001 (اهداف هوش مصنوعی مسئولانه).	معیار، قضاوت، راستی‌آزمایی و پیامدها را می‌سنجد — نه خطوط کدِ تولیدشده‌ی هوش مصنوعی را. جلسات کالیبراسیون نحوه‌ی امتیازدهی به استفاده از هوش مصنوعی را استانداردسازی می‌کنند.	تصمیم‌های ترفیع را برای یکدستی ممیزی کنید. اطمینان حاصل کنید بازبینان کالیبره شده‌اند. ردیابی کنید که آیا مهندسان مسلط به هوش مصنوعی و مهندسان محتاطِ هوش مصنوعی با معیارهای پیامدی یکسان ارزیابی می‌شوند.
دسترسی نابرابر به ابزار	اگر تنها برخی مهندسان به ابزارهای توانمند دسترسی داشته باشند، نردبان دسترسی را می‌سنجد نه توانایی را — ناعادلانه و پرنوفه. NIST Govern؛ ISO 42001 (انصاف/فراگیری).	یک مجموعه‌ابزار پایه‌ی مجاز که در هر سطح برای همه در دسترس است تعریف کنید؛ ابزارهای پیشرفته با آموزش گِیت می‌شوند، نه با جانب‌داری.	تأیید کنید همه در یک سطح به ابزارهایی که معیار فرض می‌کند دسترسی برابر دارند. ارزیابی‌هایی را که با شکاف‌های ابزاری مخدوش شده‌اند علامت‌گذاری کنید.
افشای رازها (کلیدها و توکن‌ها در پرامپت‌ها یا کدِ تولیدشده‌ی هوش مصنوعی)	اعتبارنامه‌های نشت‌یافته نقض‌های مستقیم پدید می‌آورند؛ هوش مصنوعی نیز می‌تواند رازها را به‌صورت سخت‌کدشده درج کند. OWASP LLM02؛ Manage NIST؛ کنترل‌های SR 11-7.	سطوح بالاتر اسکن رازها، وُلت‌سازی و قواعد «هرگز در پرامپت» را اعمال می‌کنند و آن را به پایین می‌آموزند. سطوح پایین‌تر درون حفاظ‌ها عمل می‌کنند (اسکن راز پیش از کامیت، محیط‌های ماسک‌شده).	بررسی کنید اسکن رازها فعال است و رعایت می‌شود. بازبینی کنید که آیا مهندس هرگز اعتبارنامه‌ای را الصاق کرده است. دانش او از گردش‌کار مدیریت رازهای سازمان را بیازمایید.

ریسک	چرا اهمیت دارد	نردبان چگونه آن را کاهش می‌دهد	HR / مدیران چگونه آن را ارزیابی می‌کنند
ریسک انطباق (خروجی الزامات نظارتی/قانونی را نقض می‌کند)	هوش مصنوعی قواعد حوزه‌ی قضایی شما را نمی‌داند؛ کدِ غیرمنطبق مخاطره‌ی حقوقی/نظارتی پدید می‌آورد. NIST Govern/Map؛ ISO 42001؛ مستندسازی و اعتبارسنجی SR 11-7.	سطوح بالاتر مالک این هستند که خروجی هوش مصنوعی پیش از ادغام با کنترل‌های قابل‌اعمال (مثلاً ثبت لاگ ممیزی، نگه‌داری، رضایت) سازگار باشد؛ بررسی‌های انطباق را در بازبینی تعبیه می‌کنند.	ارزیابی کنید که آیا مهندس الزامات را به‌طور مستقل به کنترل‌ها نگاشت می‌کند. تأیید کنید تغییرات تحت نظارت صرف‌نظر از دخالت هوش مصنوعی از تأییدهای لازم عبور می‌کنند.
خودکارسازی ناایمن (عامل‌هایی که روی تولید عمل می‌کنند، عملیات برگشت‌ناپذیر)	اقدامات خودمختار روی سامانه‌های تولید یا مشتری می‌تواند موجب قطعی یا از دست رفتن داده شود. OWASP LLM06 Excessive Agency؛ NIST Manage؛ کنترل تغییر SR 11-7.	خودمختاری یک امتیاز سطح‌بالاست با محدودیت‌های شعاع آسیب، اجرای آزمایشی، تأییدها و بازگردانی. سطوح پایین‌تر عامل‌ها را تنها در سندباکس اجرا می‌کنند.	تأیید کنید مجوزهای عامل دامنه‌بندی‌شده و برگشت‌پذیر هستند. پایبندی به کنترل تغییر برای تغییرات هوش‌مصنوعی‌محور را بازبینی کنید. تأیید کنید اقدامات تولید دروازه‌های تأیید انسانی را حفظ می‌کنند.
فقدان قابلیت ممیزی (نبودِ سابقه‌ای از آنچه هوش مصنوعی کرد / چرا)	ناظران انتظار قابلیت ردیابی دارند؛ «نمی‌توانیم آن را بازسازی کنیم» یک یافته است. OWASP LLM Top 10 (logging)؛ NIST Measure؛ ISO 42001 (سوابق)؛ مستندسازی SR 11-7.	سطوح بالاتر اطمینان می‌دهند که تغییرات مجهز به هوش مصنوعی مستند شده‌اند (چه چیزی تولید شد، چه چیزی راستی‌آزمایی شد) و اقدامات عامل لاگ می‌شوند.	بررسی کنید PRها/کامیت‌ها مطابق سیاست به دخالت هوش مصنوعی اشاره می‌کنند. تأیید کنید لاگ‌های عامل وجود دارند و نگه‌داری می‌شوند. بیازمایید که آیا یک تغییر گذشته قابل بازسازی است.
کدِ نادرستِ تولیدشده‌ی هوش مصنوعی (باگ‌های ظریف، حالت‌های مرزی نادرست)	کدی با ظاهری معقول و دارای نقص منطقی/امنیتی به تولید می‌رسد. OWASP LLM05 Improper Output Handling؛ NIST Measure؛ اعتبارسنجی SR 11-7.	پوشش تست، بازبینی امنیتی و دقت راستی‌آزمایی شروط سطح هستند؛ سطوح بالاتر سطح انتظار را تعیین می‌کنند و PRهای پرحجمِ هوش مصنوعی را با دقت بیشتری بازبینی می‌کنند.	ادغام‌های مجهز به هوش مصنوعی را برای نقص‌ها و کفایت تست نمونه‌برداری کنید. نرخ نقص/گریز را برحسب استفاده از هوش مصنوعی ردیابی کنید. انضباط تست‌نویسی مهندس را ارزیابی کنید.
قفل‌شدگی به فروشنده / مدل	اتکای بیش‌ازحد به یک مدل/ارائه‌دهنده، ریسک هزینه، تداوم و تمرکز را افزایش می‌دهد. NIST Govern؛ ISO 42001 (کنترل‌های تأمین‌کننده)؛ SR 11-7 (مدل‌های فروشنده).	سطوح بالاتر پرامپت‌ها/گردش‌کارها را قابل‌انتقال نگه می‌دارند، فرض‌های مدل را مستند می‌کنند و در جایی که قابل‌اجتناب است از وابستگی‌های سخت به یک ارائه‌دهنده‌ی واحد پرهیز می‌کنند.	از مهندس بپرسید چگونه از مدل کنونی مهاجرت می‌کند. بررسی کنید گردش‌کارهای حیاتی بدون جایگزین به یک فروشنده‌ی واحد وابسته نباشند.
هوش مصنوعی سایه (ابزارها/حساب‌های غیرمجاز)	ابزارهای ردیابی‌نشده کنترل‌های داده، امنیت و ممیزی را دور می‌زنند — پراهرم‌ترین شکاف حاکمیتی. OWASP LLM02/LLM05؛ NIST Govern/Map؛ ISO 42001 (سیاهه‌ی دارایی).	نردبان استفاده‌ی صرف از ابزارهای مجاز و گزارش شکاف‌های مجموعه‌ابزار به جای دورزدن آن‌ها را به هنجار بدل می‌کند؛ سطوح بالاتر متولی فهرست ابزارهای تأییدشده‌اند.	استفاده‌ی واقعی از ابزار را در برابر فهرست مجاز پیمایش کنید. افشای نیازهای برآورده‌نشده را ایمن سازید. هوش مصنوعی سایه‌ی کشف‌شده را یک شکست فرایندی قابل‌رفع تلقی کنید، نه صرفاً یک تقصیر فردی.

ریسک	چرا اهمیت دارد	نردبان چگونه آن را کاهش می‌دهد	HR / مدیران چگونه آن را ارزیابی می‌کنند
کنترل هزینه‌ی هوش مصنوعی	هزینه‌ی توکن/محاسبه می‌تواند بی‌صدا متورم شود؛ عامل‌های مدیریت‌نشده گران‌اند. NIST Manage (منابع)؛ ISO 42001 (کنترل‌های عملیاتی).	سطوح بالاتر از هوش مصنوعی به‌صورت مقرون‌به‌صرفه استفاده می‌کنند، مدل‌ها را به‌اندازه‌ی وظایف تنظیم می‌کنند و هزینه را پایش می‌کنند؛ این یک انتظار کارایی صریح است، نه «همیشه از بزرگ‌ترین مدل استفاده کن».	روند هزینه‌به‌ازای‌پیامد را بازبینی کنید. بررسی کنید مهندسان مدل‌هایی با اندازه‌ی مناسب انتخاب می‌کنند. به کاراییِ قابل‌اندازه‌گیری پاداش دهید، نه به مصرف خام.
پاسخگویی انسانی (مالکیت خروجی هوش مصنوعی)	مالکیت پراکنده، ایمنی و پاسخگویی نظارتی را فرسوده می‌کند. NIST Govern؛ ISO 42001 (نقش‌ها/مسئولیت‌ها)؛ SR 11-7 (مالکیت مشخص).	هر سطح: یک انسانِ مشخص مالک خروجی ادغام‌شده است؛ «هوش مصنوعی این کار را کرد» هرگز دفاع نیست. سطوح بالاتر مالک پیامدها در سراسر استفاده‌ی تیمشان از هوش مصنوعی هستند.	تأیید کنید مالکیت روشن بر هر تغییر وجود دارد. در بازنگری‌ها/بررسی‌های رخداد، تأیید کنید یک مالک انسانی برای کارِ مجهز به هوش مصنوعی نام‌برده شده است.

19.3 جدول هم‌ترازی چارچوب‌ها (مرجع سریع)

ریسک	OWASP LLM Top 10 (2025)	کارکرد NIST AI RMF	ISO/IEC 42001	SR 11-7
توهم‌زایی	LLM09 Misinformation	Measure	پایش عملکرد	اعتبارسنجی خروجی
نشت داده	LLM02 Sensitive Info Disclosure	Govern / Manage	مدیریت داده	حاکمیت داده
تزریق پرامپت	LLM01 Prompt Injection	Map / Manage	کنترل‌های عملیاتی	کنترل تغییر
اعتماد بیش‌ازحد	LLM09	Measure	نظارت انسانی	چالش اثربخش
ارزیابی سوگیرانه	— (حاکمیت)	Govern	اهداف هوش مصنوعی مسئولانه	توسعه‌ی سالم
دسترسی نابرابر به ابزار	— (حاکمیت)	Govern	انصاف / منابع	—
افشای رازها	LLM02	Manage	امنیت اطلاعات	محیط کنترل‌ها
ریسک انطباق	— (حاکمیت)	Govern / Map	حقوقی و نظارتی	مستندسازی / اعتبارسنجی
خودکارسازی ناایمن	LLM06 Excessive Agency	Manage	کنترل‌های عملیاتی	کنترل تغییر
فقدان قابلیت ممیزی	LLM Top 10 (logging)	Measure	سوابق و قابلیت ردیابی	مستندسازی
کدِ نادرست هوش مصنوعی	LLM05 Improper Output Handling	Measure	وارسی و اعتبارسنجی	اعتبارسنجی مدل
قفل‌شدگی به فروشنده / مدل	— (حاکمیت)	Govern	مدیریت تأمین‌کننده	ریسک مدل فروشنده

ریسک	OWASP LLM Top 10 (2025)	کارکرد NIST AI RMF	ISO/IEC 42001	SR 11-7
هوش مصنوعی سایه	LLM02 / LLM05	Govern / Map	سیاهه‌ی دارایی	سیاهه و دامنه
کنترل هزینه‌ی هوش مصنوعی	— (عملیاتی)	Manage	کنترل‌های عملیاتی	پایش مستمر
پاسخگویی انسانی	— (حاکمیت)	Govern	نقش‌ها و مسئولیت‌ها	مالکیت مشخص

19.4 راهنمای عملی برای HR و مدیران

- **ایمنی را قابل‌ترفیع کنید.** اگر راستی‌آزمایی، قابلیت ممیزی و پاسخگویی در معیار نباشند، در برابر سرعت بازنده خواهند بود. آن‌ها را در انتظارات سطح و کالیبراسیون بگنجانید.
- **قضاوت را ارزیابی کنید، نه حجم را.** قوی‌ترین نشانه‌ی یک مهندس بالغ در هوش مصنوعی، کیفیت تردید اوست: چه چیزی را رد کرد، چه چیزی را راستی‌آزمایی کرد، چه چیزی را تشدید (escalate) کرد.
- **هوش مصنوعی سایه را نخست یک شکست سامانه‌ای تلقی کنید.** مجازات افشا، آن را به زیرزمین می‌راند. به افرادی که نیازهای ابزاری برآورده‌نشده را آشکار می‌کنند پاداش دهید.
- **بر هر تغییر نام یک انسان را نگه دارید.** مالکیت تنها کنترلی است که همه‌ی کنترل‌های دیگر را به هم گره می‌زند و انتظار SR 11-7 برای پاسخگویی روشن را برآورده می‌کند.
- **بیش از اندازه ادعا نکنید.** این چارچوب استانداردهای شناخته‌شده را در برابر رفتارهای نردبان سازمان‌دهی می‌کند؛ این یک گواهی‌نامه نیست. اعتبارسنجی در برابر ISO 42001 یا SR 11-7 باید با واحدهای ریسک، حقوقی و انطباق شما انجام شود. (برداشت‌شده از منابع عمومی.)

---

20. برنامه پیاده‌سازی منابع انسانی (HR)

این برنامه، نردبان قابلیت هوش مصنوعی (L0–L10) را در چهار فاز در سطح کل سازمان مستقر می‌کند. این برنامه برای یک محیط فناوری تنظیم‌شده و بانکی طراحی شده است که در آن کنترل‌های مدیریت داده، اعتبارسنجی ریسک مدل و مدیریت تغییرِ قابل‌ممیزی الزامی هستند. بازه‌های زمانی صرفاً جنبه راهنما دارند و باید متناسب با تعداد نیروی انسانی و تقویم نظارتی شما تنظیم شوند.

مرور کلی فازبندی

فاز	نام	مدت زمان معمول	مالک اصلی	معیارهای خروج
1	طراحی و هم‌راستاسازی	4–6 هفته	HR + مسئول راهبری هوش مصنوعی (AI Governance Lead)	تأیید نهایی تعاریف نردبان؛ توافق بر روش ارزیابی؛ انتشار حفاظ‌های راهبری
2	آزمایشی (Pilot)	6–10 هفته	مدیر ارشد DevOps + شریک تجاری منابع انسانی (HR Business Partner)	ارزیابی یک تیم در سطوح L0–L10؛ سنجش وضعیت پایه؛ پالایش راهنمای اجرایی
3	استقرار (Rollout)	1–2 فصل	HR + مدیران مهندسی	ارزیابی همه کارکنانِ در محدوده و تخصیص مسیرهای توسعه

فاز	نام	مدت زمان معمول	مالک اصلی	معیارهای خروج
4	بهبود مستمر	جاری و پیوسته	هیئت راهبری هوش مصنوعی (AI Governance Board)	اجرای چرخه بازارزیابی فصلی؛ بازنگری معیارها در هر فصل

فاز 1 — طراحی و هم‌راستاسازی (4 تا 6 هفته)

هدف: توافق بر سر معنای هر سطح از نردبان در بافت سازمان شما، شیوه ارزیابی آن، و تعیین حفاظ‌هایی که پیش از استفاده هر فرد از یک ابزار باید رعایت شوند.

فعالیت	توضیح	خروجی
تثبیت تعاریف نردبان	نگاشت سطوح L0–L10 به نقش‌های سازمان شما (برای مثال، چه کسی انتظار می‌رود به L3 برسد در برابر L6 یا L9)	ماتریس انتظارات نقش به سطح
تعریف شواهد ارزیابی	تصمیم‌گیری درباره اینکه چه چیزی به عنوان اثبات هر سطح به شمار می‌آید (برای مثال، L4 = الگوهای prompt اثبات‌شده؛ L7 = یک اتوماسیون ادغام‌شده با قابلیت rollback)	سنجه (rubric) شواهد برای هر سطح
تعیین دروازه مدیریت داده	انتشار اینکه چه داده‌هایی مجاز یا غیرمجاز به ورود به ابزارهای GenAI هستند؛ پیوند آن به دروازه «آگاهی از AI» در سطح L2	فهرست ابزارهای تأییدشده + قواعد طبقه‌بندی داده
هم‌راستاسازی با ریسک مدل و انطباق	تأیید اینکه کارهای L8 و بالاتر از مسیر اعتبارسنجی ریسک مدل عبور می‌کنند؛ تعریف دروازه‌های تأیید/rollback برای L7	تأییدیه حاکمیتی
برنامه ارتباطات	تبیین چرایی وجود نردبان و اینکه ماهیتی توسعه‌ای دارد، نه تنبیهی	پیام راه‌اندازی داخلی + پرسش‌های متداول (FAQ)

تصمیم‌های کلیدی هم‌راستاسازی که باید قطعی شوند: - هیچ استفاده کاری از GenAI بالاتر از سطح L1 مجاز نیست تا زمانی که دروازه آگاهی L2 گذرانده شود (آموزش + تأییدیه). - اتوماسیون‌های تولیدی (production) سطح L7 نیازمند دروازه‌های تأیید و rollback هستند. - راهکارهای L8 و بالاتر باید از اعتبارسنجی ریسک مدل عبور کنند.

فاز ۲ — پایلوت (۶ تا ۱۰ هفته)

هدف: اجرای کامل نردبان روی یک تیم نماینده (تیم DevOps/platform گزینه ایده‌آل است) برای آزمودن روبریک، آشکارسازی اصطکاک‌ها و تولید یک خط مبنا (baseline).

فعالیت	توضیح	خروجی
انتخاب تیم پایلوت	انتخاب یک تیم با سطوح مهارتی متنوع؛ جلب همراهی و حمایت مدیر	کوهورت پایلوت مشخص‌شده
ارزیابی خط مبنا	جای‌گذاری هر عضو روی سطوح L0–L10 با استفاده از روبریک مبتنی بر شواهد	توزیع سطوح خط مبنا
توانمندسازی هدفمند	اجرای آگاهی‌بخشی L2 برای هر کسی که زیر این سطح قرار دارد؛ مربی‌گری مسیر L3→L4→L5 با وظایف واقعی	لاگ تکمیل آموزش

فعالیت	توضیح	خروجی
آزمودن حفاظ‌ها (guardrails)	تیم پایلوت تلاش کند یک اتوماسیون L7 را از مسیر تأیید/بازگشت (approval/rollback) عبور دهد	نمونه عملی + درس‌آموخته‌ها
سنجش	پایش زمان چرخه (cycle-time)، دوباره‌کاری و نرخ شکست در راستی‌آزمایی در مقایسه با خط مبنا	گزارش معیارهای پایلوت
پالایش	به‌روزرسانی روبریک، تعاریف شواهد و آموزش بر اساس اصطکاک‌های مشاهده‌شده	پلی‌بوک نسخه ۲

معیارهای موفقیت پایلوت: - هر عضو پایلوت با شواهد مستند روی یک سطح جای‌گذاری شده باشد. - دست‌کم یک اتوماسیون L7 از طریق دروازه‌های تأیید و بازگشت (approval and rollback) به بهره‌برداری رسیده باشد. - یک دستاورد قابل‌سنجش و قابل‌انتساب (برای مثال، کاهش زمان چرخه یا دوباره‌کاری) دست‌کم در یک جریان کاری حاصل شده باشد.

فاز ۳ — استقرار گسترده (Rollout) (۱ تا ۲ فصل)

هدف: گسترش نردبان ارزیابی‌شده و مسیرهای توسعه در میان همه تیم‌های در دامنه، بدون آن‌که مدیران یا واحد انطباق (compliance) را زیر فشار قرار دهد.

فعالیت	شرح	خروجی
برنامه‌ریزی موجی (Wave planning)	استقرار تیم‌به‌تیم (نه همه به‌طور هم‌زمان)؛ توالی بر اساس میزان آمادگی	زمان‌بندی استقرار
توانمندسازی مدیران	آموزش مدیران برای ارزیابی، مربیگری و اعتبارسنجی شواهد به‌شکل یکدست	ارزیابان کالیبره‌شده
دروازه انبوه L2	اطمینان از این‌که همه پیش از دسترسی به ابزارها، دروازه آگاهی را پشت سر می‌گذارند	پیوند دسترسی به تکمیل دروازه
تخصیص مسیرهای توسعه	هر فرد یک سطح هدف و اقدامات گام بعدی دریافت می‌کند	برنامه‌های توسعه فردی
شناسایی تقویت‌کننده‌های L6	معرفی کاربران حرفه‌ای AI (AI Power Users) در هر تیم برای تعیین هنجارها و ارتقای دیگران	معرفی رهبران L6 برای هر تیم
حاکمیت ابزار و دسترسی	تأمین ابزارهای مصوب بر اساس سطح؛ ثبت لاگ دسترسی	ماتریس دسترسی

نکته کالیبراسیون: جلسات کالیبراسیون میان‌تیمی برگزار کنید تا «L4» در همه‌جا معنای یکسانی داشته باشد. سطح‌بندی ناهمگون سریع‌ترین راه از دست رفتن اعتماد به نردبان است.

فاز ۴ — بهبود مستمر (پیوسته)

هدف: به‌روز نگه‌داشتن نردبان همگام با تغییر ابزارها، تهدیدها و مقررات، و اثبات ارزش‌آفرینی آن در طول زمان.

فعالیت	تناوب	خروجی
چرخه ارزیابی مجدد	فصلی	توزیع به‌روزشده سطوح + روندهای جابه‌جایی
بازنگری شاخص‌ها	فصلی	گزارش روند بهره‌وری، کیفیت و رخدادهای ریسک

فعالیت	تناوب	خروجی
به‌روزرسانی حفاظ‌ها (guardrail)	فصلی یا با تغییر مقررات	قواعد داده، فهرست ابزارهای تأییدشده و دروازه‌های تأیید به‌روزشده
کنترل نسخه نردبان	در صورت نیاز	شرح تغییرات (Changelog) سطوح/معیارها
قدردانی و پیشرفت	در هر چرخه ارزیابی	ارتقا/قدردانی متصل به سطوح اثبات‌شده
حلقه بازخورد رخدادها	هنگام وقوع رویداد	بازخورد درس‌آموخته‌ها به آگاهی‌بخشی L2 و دروازه‌های L7

شاخص‌هایی که باید پایش شوند (نمونه‌وار):

شاخص	چرا اهمیت دارد
توزیع سطوح در طول زمان	ارتقای توانمندی را در سطح کل سازمان نشان می‌دهد
درصد کارکنانی که از دروازه L2 عبور کرده‌اند	ایمنی پایه در مدیریت داده‌ها را تأیید می‌کند
تعداد اتوماسیون‌های فعال L7 (با بازگشت‌پذیری آزموده‌شده)	میزان پذیرش اتوماسیون ایمن را اندازه می‌گیرد
نرخ شکست راستی‌آزمایی / دوباره‌کاری	اعتماد بیش از حد به خروجی AI را شناسایی می‌کند
رخدادهای ریسک مرتبط با AI	اثربخش بودن حفاظ‌ها را تأیید می‌کند
زمان چرخه در گردش‌کارهای پشتیبانی‌شده با AI	دستاوردهای بهره‌وری را کمّی‌سازی می‌کند

---

21. قالب‌ها (Templates)

چهار قالب زیر به‌گونه‌ای طراحی شده‌اند که بتوان آن‌ها را مستقیماً در فرایندهای ارزیابی عملکرد و ارتقای شما کپی کرد. همه فیلدها خالی و آماده تکمیل هستند. جای‌نگهدارهای داخل کروشه مانند `[ ]` و `_____` را با داده‌های واقعی جایگزین کنید.

---

22. ده اقدام نخست پیشنهادی

یک سیاهه‌ی عملیاتی و مشخص که یک مدیر ارشد DevOps می‌تواند **همین فصل** آن را آغاز کند. هر اقدام کوچک، قابل‌اندازه‌گیری و چنان مرتب شده است که ایمنی (دروازه‌ی L2) پیش از مقیاس‌پذیری قرار گیرد.

#	اقدام	زمانی که انجام‌شده تلقی می‌شود...	مالک پیشنهادی
1	**انتشار قواعد ابزارهای تأییدشده و نحوه‌ی مدیریت داده** (اینکه چه داده‌ای مجاز/غیرمجاز است وارد GenAI شود). آن را به‌صراحت به دروازه‌ی آگاهی L2 گره بزنید.	یک برگه‌ی یک‌صفحه‌ای از قواعد منتشر شده و توسط تیم تأیید (acknowledge) شده باشد.	مدیر DevOps + Compliance
2	**اجرای دروازه‌ی «آگاهی از AI» در سطح L2 برای کل تیم** (آموزش کوتاه + تأییدیه) پیش از هرگونه استفاده‌ی بیشتر از ابزار.	۱۰۰٪ اعضای تیم آن را تکمیل و تأیید کرده باشند.	مدیر DevOps + HR

#	اقدام	زمانی که انجام‌شده تلقی می‌شود...	مالک پیشنهادی
3	**تعیین سطح پایه‌ی هر یک از اعضای تیم روی L0–L10** با استفاده از سنجه‌ی مبتنی بر شواهد.	هر فرد سطحی مستند به‌همراه شواهد داشته باشد.	مدیر DevOps
4	**انتخاب یک گردش‌کار واقعی برای بهبود** (مثلاً runbookهای رخداد، بازبینی PR، غربالگری لاگ‌ها) و اندازه‌گیری زمان‌چرخه/دوباره‌کاری فعلی آن.	یک عدد پایه برای آن گردش‌کار ثبت شده باشد.	مدیر DevOps
5	**مربی‌گری ۲ تا ۳ نفر از L3 به L4** با تمرین هدفمند در نوشتن prompt روی وظایف کم‌ریسک.	دست‌کم ۲ نفر الگوهای prompt قابل‌اتکا و تکرارپذیر را نشان دهند.	راهبر L6 / مدیر DevOps
6	**معرفی یک «کاربر حرفه‌ای AI» در سطح L6** در تیم برای تعیین هنجارهای ایمن و راهنمایی دیگران.	یک نفر به‌طور رسمی منصوب شده و زمان لازم برای مربی‌گری به او داده شده باشد.	مدیر DevOps
7	**تعریف دروازه‌ی تأیید + بازگشت (rollback) در سطح L7** برای هرگونه خودکارسازی تولیدی (production).	یک سیاهه‌ی مکتوب وجود داشته باشد: چه کسی تأیید می‌کند، چگونه بازگشت انجام می‌شود و چگونه ثبت می‌گردد.	مدیر DevOps + Change Mgmt
8	**عرضه‌ی یک خودکارسازی کوچک L7 از طریق همان دروازه** (با شعاع اثر کم و کاملاً برگشت‌پذیر).	یک خودکارسازی فعال و تأییدشده باشد و بازگشت آن آزموده شده باشد.	مهندس DevOps + راهبر L6
9	**راه‌اندازی یک داشبورد ساده‌ی سنجه‌ها** (توزیع سطوح، درصد عبورکرده از دروازه‌ی L2، خودکارسازی‌های فعال، نرخ دوباره‌کاری).	داشبورد اعداد جاری تیم را نمایش دهد.	مدیر DevOps
10	**زمان‌بندی ارزیابی مجدد و بازنگری فصلی** تا نردبان زنده بماند و به یک تمرین یک‌باره تبدیل نشود.	یک دعوت‌نامه‌ی تقویمی تکرارشونده و یک قالب بازنگری وجود داشته باشد.	مدیر DevOps + HR

منطق توالی (برگرفته از مواد عمومی درباره‌ی مدیریت تغییر مرحله‌ای و بهترین‌روش‌های حاکمیت AI): نخست مدیریت داده و دروازه‌ی L2 را قفل کنید (اقدامات ۱ و ۲)، سپس یک سطح پایه و یک گردش‌کار هدف قابل‌اندازه‌گیری ایجاد کنید (اقدامات ۳ و ۴)، آنگاه توانمندی انسانی و یک قهرمان محلی بسازید (اقدامات ۵ و ۶)، تنها پس از آن خودکارسازی برگشت‌پذیر را پشت دروازه‌های صریح معرفی کنید (اقدامات ۷ و ۸)، و در نهایت کل این تلاش را قابل‌اندازه‌گیری و تکرارپذیر کنید (اقدامات ۹ و ۱۰).

> توجه: سطوح نردبان L0–L10 که در اینجا به آن‌ها ارجاع داده شده، از شرح‌حال تکلیف (assignment brief) گرفته شده‌اند. این طرح اجرایی، رویه‌های عمومی و شناخته‌شده‌ی مدیریت تغییر و حاکمیت AI را تلفیق می‌کند؛ این طرح هیچ نردبان HR داخلی متعلق به شرکتی خاص را بازتولید نمی‌کند.

17.1 فرم خودارزیابی

فیلد	ورودی
نام کارمند	_____
شناسه کارمند	_____
عنوان / سطح فعلی	_____
تیم / دپارتمان	_____

فیلد	ورودی
نام مدیر	_____
دوره ارزیابی (از – تا)	_ تا _
تاریخ ثبت	_____

دستاوردهای کلیدی این دوره

#	دستاورد	اثر کسب‌وکاری / بر مشتری	پیوند یا مرجع شواهد
1			
2			
3			

خودارزیابی به تفکیک شایستگی (مقیاس: 1 = در حال توسعه، 2 = مطابق انتظار، 3 = قوی، 4 = فراتر از انتظار)

شایستگی	امتیاز خودارزیابی (1–4)	توضیح پشتیبان
مهارت‌های فنی / تخصصی	[ ]	
تحویل و اجرا	[ ]	
کیفیت کد / کار	[ ]	
همکاری و کار تیمی	[ ]	
ارتباطات	[ ]	
مالکیت و اتکاپذیری	[ ]	
آگاهی از انطباق و ریسک	[ ]	
منتورینگ / رهبری (در صورت کاربرد)	[ ]	

پرسش‌های بازتابی

پرسش	پاسخ
بیشترین افتخار من در این دوره چه بوده است؟	
در کجا کوتاهی کردم و چرا؟	
در گام بعد می‌خواهم چه مهارت‌هایی بسازم؟	
به چه پشتیبانی‌ای از مدیر / تیم خود نیاز دارم؟	
آیا باور دارم برای ارتقا آماده‌ام؟ چرا؟	

امضای کارمند	_____	تاریخ	___

17.2 فرم ارزیابی مدیر

فیلد	ورودی
نام کارمند	_____
عنوان / سطح فعلی	_____
نام مدیر	_____
دوره ارزیابی (از – تا)	_ تا _
تاریخ تکمیل	_____

ارزیابی شایستگی‌ها (مقیاس: 1 = در حال توسعه، 2 = برآورده‌کننده انتظارات، 3 = قوی، 4 = فراتر از انتظار)

شایستگی	امتیاز (1-4)	شواهد / مثال مشخص
مهارت‌های فنی / تخصصی	[ ]	
تحویل و اجرا	[ ]	
کیفیت کد / کار	[ ]	
همکاری و کار تیمی	[ ]	
ارتباطات	[ ]	
مسئولیت‌پذیری و قابل‌اتکا بودن	[ ]	
آگاهی از انطباق و ریسک	[ ]	
منتورینگ / رهبری (در صورت کاربرد)	[ ]	

خلاصه کلی

فیلد	ورودی
امتیاز کلی	[ ] پایین‌تر از انتظار [ ] برآورده‌کننده [ ] فراتر از انتظار
دو نقطه قوت برتر	
دو حوزه توسعه برتر	
توافق با خودارزیابی کارمند؟	[ ] بله [ ] جزئی [ ] خیر — در زیر توضیح دهید
یادداشت‌ها درباره هرگونه شکاف نسبت به خودارزیابی	

توسعه و گام‌های بعدی

هدف برای دوره بعد	معیار موفقیت	تاریخ هدف

هدف برای دوره بعد	معیار موفقیت	تاریخ هدف

توصیه ارتقا

فیلد	ورودی
توصیه به ارتقا در این چرخه؟	[ ] بله [ ] هنوز نه [ ] خیر
سطح هدف	_____
توجیه (مبتنی بر شواهد)	

امضای مدیر	_____	تاریخ	_____

17.3 برگه امتیازدهی آزمون عملی

فیلد	ورودی
نام داوطلب	_____
نقش / سطح مورد ارزیابی	_____
نام آزمون / تمرین	_____
تاریخ آزمون	_____
نام ارزیاب	_____

امتیازدهی (مقیاس: 0 = نمایش داده نشد، 1 = جزئی، 2 = کافی، 3 = قوی، 4 = عالی)

معیار	وزن (%)	امتیاز خام (0-4)	امتیاز وزنی	یادداشت‌ها
درک مسئله و رویکرد		[ ]		
درستی راه‌حل		[ ]		
کیفیت / ساختار کد		[ ]		
مدیریت موارد حاشیه‌ای و خطاها		[ ]		
آگاهی از امنیت و مدیریت داده		[ ]		
کارایی / بهره‌وری		[ ]		
آزمون و راستی‌آزمایی		[ ]		
بیان تصمیم‌ها		[ ]		
مجموع	**100**	—		

نتیجه

فیلد	ورودی
مجموع امتیاز وزنی	___ / 100
آستانه قبولی	___
نتیجه	[ ] قبول [ ] مرزی [ ] مردود
نقاط قوت کلیدی مشاهده‌شده	
نگرانی‌های کلیدی	
توصیه	

امضای ارزیاب	_____	تاریخ	___

17.4 سیاهه وارسی مستندات ارتقا

فیلد	ورودی
نام کارمند	_____
سطح فعلی	_____
سطح هدف	_____
مدیر حامی	_____
دوره ارزیابی	_____

الزامات مستندسازی (هر مورد را به‌عنوان تکمیل‌شده علامت بزنید و پیوند مدرک را قرار دهید)

الزام	برآورده شده؟	پیوند / مرجع مدرک	یادداشت ارزیاب
عملکرد مستمر در سطح هدف (حداقل دوره پایدار)	[ ] بله [ ] خیر		
سابقه تحویل (کارهای ارائه‌شده، نتایج)	[ ] بله [ ] خیر		
نشان دادن دامنه / پیچیدگی متناسب با سطح هدف	[ ] بله [ ] خیر		
تأثیرگذاری فراتر از وظایف بلافصل خود	[ ] بله [ ] خیر		
همکاری و مشارکت بین‌تیمی	[ ] بله [ ] خیر		
مربی‌گری / اشتراک دانش (در صورت لزوم در آن سطح)	[ ] بله [ ] خیر		
سابقه پاک انطباق و ریسک در طول دوره	[ ] بله [ ] خیر		
قبولی در آزمون / ارزیابی عملی (در صورت لزوم)	[ ] بله [ ] خیر		

الزام	برآورده شده؟	پیوند / مرجع مدرک	یادداشت ارزیاب
ارائه خودارزیابی	[ ] بله [ ] خیر		
ارائه ارزیابی مدیر	[ ] بله [ ] خیر		
جمع‌آوری بازخورد همتایان / ذی‌نفعان	[ ] بله [ ] خیر		

تأیید نهایی

نقش	نام	تصمیم	امضا	تاریخ
مدیر حامی		[ ] تأیید [ ] تعلیق		
مدیر سطح بالاتر / رئیس بخش		[ ] تأیید [ ] تعلیق		
شریک منابع انسانی (HR / People partner)		[ ] تأیید [ ] تعلیق		
تصمیم نهایی		[ ] تأیید شد [ ] به تعویق افتاد [ ] رد شد		

یادداشت‌ها / شرایط نهایی	

23. منابع و مآخذ

منابع شامل چارچوب‌های عمومی، مستندات فروشندگان، وبلاگ‌های مهندسی و گزارش‌ها هستند (دانش تا حدود ژانویه ۲۰۲۶). مواردی که در متن با برچسب (استنباطی) مشخص شده‌اند، بر اساس استدلال از منابع عمومی استخراج شده‌اند، نه از اسناد داخلی منابع انسانی.

- **The Impact of AI on Developer Productivity: Evidence from GitHub Copilot (arXiv 2302.06590)** — https://arxiv.org/abs/2302.06590 — کارآزمایی تصادفی کنترل‌شده‌ای که تکمیل وظایف را با Copilot حدود ۵۵٫۸٪ سریع‌تر نشان می‌دهد
- **GitHub Blog: Quantifying GitHub Copilot's impact in the enterprise with Accenture** — https://github.blog/news-insights/research/research-quantifying-github-copilots-impact-in-the-enterprise-with-accenture/ — کارآزمایی تصادفی در مقیاس سازمانی
- **AWS Certified AI Practitioner (AIF-C01) exam guide** — https://docs.aws.amazon.com/aws-certification/latest/ai-practitioner-01/ai-practitioner-01.html — حوزه‌های گواهینامه پایه‌ای هوش مصنوعی
- **AWS Certified Generative AI Developer – Professional** — https://aws.amazon.com/certification/certified-generative-ai-developer-professional/ — گواهینامه GenAI مسیر سازندگان
- **Microsoft 365 Copilot Skilling Center** — https://adoption.microsoft.com/en-us/copilot/skilling-center/ — منابع مهارت‌آموزی مبتنی بر نقش
- **Microsoft Learn: Agentic AI maturity model (governance & security)** — https://learn.microsoft.com/en-us/microsoft-copilot-studio/guidance/maturity-model-security-governance — نردبان بلوغ حاکمیت

- **C5 Insight: 5 Levels of Microsoft Copilot Adoption Maturity** — https://c5insight.com/levels-of-microsoft-copilot-adoption-maturity-5/ — مدل بلوغ منتشرشده توسط شریک (استنباطی)
- **DORA 2024 State of DevOps Report highlights (getdx summary)** — https://getdx.com/blog/2024-dora-report/ — یافته‌های مربوط به پذیرش هوش مصنوعی در برابر پایداری تحویل
- **DORA Insights** — https://dora.dev/insights/ — برنامه پژوهشی DevOps شرکت Google
- **IBM Certified watsonx Governance Lifecycle Advisor v1 – Associate** — https://www.ibm.com/training/certification/ibm-certified-watsonx-governance-lifecycle-advisor-v1-associate-C9008000 — گواهینامه ویژه حاکمیت
- **IBM watsonx.governance product page** — https://www.ibm.com/products/watsonx-governance — حاکمیت چرخه عمر، پایش ریسک، factsheetها
- **IBM OpenPages Model Risk Governance docs** — https://www.ibm.com/docs/en/SSFUEU_9.0.0/op_grc_admin/c_adm_mrg.html — حاکمیت ریسک مدل مرتبط با بانکداری
- **A Practical Guide to Enterprise AI Adoption (OpenAI/Anthropic playbooks)** — https://www.linkedin.com/pulse/openai-anthropic-playbooks-practical-guide-enterprise-gleb-markevich-evjse — خلاصه‌ای از راهنمای پذیرش سازمانی فروشندگان
- **Stanford Digital Economy Lab: The Enterprise AI Playbook (51 deployments)** — https://digitaleconomy.stanford.edu/app/uploads/2026/03/EnterpriseAIPlaybook_PereiraGraylinBrynjolfsson.pdf — شواهد پذیرش میان‌سازمانی
- **PYMNTS: Anthropic 30-Day Data Policy and enterprise governance gaps** — https://www.pymnts.com/news/artificial-intelligence/2026/anthropic-30-day-data-policy-exposes-enterprise-ai-governance-gaps/ — حاکمیت حفاظت داده / ریسک فروشنده
- **Gartner AI Maturity Model (BMC summary)** — https://www.bmc.com/blogs/ai-maturity-models/ — پنج مرحله: آگاهی، فعال، عملیاتی، نظام‌مند، تحول‌آفرین
- **Deloitte: State of AI in the Enterprise 2026** — https://www.deloitte.com/us/en/what-we-do/capabilities/applied-artificial-intelligence/content/state-of-ai-in-the-enterprise.html — یافته‌های مربوط به تسلط هوش مصنوعی نیروی کار، ارتقا/بازآموزی مهارت و بازطراحی مشاغل
- **Deloitte AI Academy** — https://www.deloitte.com/us/en/what-we-do/capabilities/applied-artificial-intelligence/about/deloitte-ai-academy.html — برنامه داخلی ارتقای مهارت و گواهینامه مبتنی بر نقش
- **Deloitte: Generative AI shaping the future workforce** — https://www.deloitte.com/cbc/en/what-we-do/genai/the-future-workforce.html — چارچوب‌بندی آینده کار و نقش‌های جدید هوش مصنوعی
- **McKinsey QuantumBlack** — بازوی هوش مصنوعی/تحلیل‌داده؛ کهن‌الگوهای نقش‌های فنی را تعریف می‌کند (دانشمند داده، مهندس ML، مهندس داده، مترجم، طراح). دانش عمومی جاافتاده.
- **BCG X / 10-20-70 principle** — واحد ساخت/طراحی فناوری؛ ۱۰٪ الگوریتم‌ها، ۲۰٪ فناوری و داده، ۷۰٪ افراد و فرایند. دانش عمومی جاافتاده.
- **PwC Responsible AI** — جعبه‌ابزار هوش مصنوعی مسئولانه و بازآموزی مهارت «New world. New skills.». دانش عمومی جاافتاده.
- **KPMG Trusted AI / AI governance** — چارچوب حاکمیت هوش مصنوعی و نیروی کار دیجیتال. دانش عمومی جاافتاده.
- **Forrester AI maturity / adoption** — پژوهش بلوغ در سطوح پذیرش؛ برچسب‌های دقیق سطوح بسته به گزارش متفاوت است (استنباطی از منابع عمومی).

- **Accenture LearnVantage** — تحول استعداد هوش مصنوعی و بازآموزی مهارت در مقیاس بزرگ. دانش عمومی جاافتاده.
- **ServiceNow Now Assist platform** — https://www.servicenow.com/platform/now-assist.html — Now Assist، Skill Kit، Data Kit، Analytics، Guardian
- **ServiceNow touts AI governance for its Autonomous Workforce (TechTarget)** — https://www.techtarget.com/searchitoperations/news/366639250/ServiceNow-touts-AI-governance-for-its-Autonomous-Workforce — مدل سه‌بخشی حاکمیت / Guardian
- **Now Assist Governance Tools and Guardrails (ServiceNow Community)** — https://www.servicenow.com/community/now-assist-blog/now-assist-governance-tools-and-guardrails/ba-p/3517694 — تشدید HITL، مسیریابی موضوعات حساس
- **PagerDuty H2 2025 Release: AI Agents** — https://www.pagerduty.com/blog/product/product-launch-2025-h2/ — SRE Agent، پیشنهاد/اقدام پس از تأیید، playbookهای هوشمند
- **PagerDuty Runbook Automation** — https://www.pagerduty.com/platform/automation/runbook/ — runbookهای تولیدشده با هوش مصنوعی، خودکارسازی فراخوانده‌شده توسط پاسخ‌دهنده با HITL
- **GitLab Duo Agent Platform HITL guardrails (Epic 22381)** — https://gitlab.com/groups/gitlab-org/-/work_items/22381 — نسخه بتای دروازه‌های تأیید ابزار
- **GitLab Duo agentic security threats (docs)** — https://docs.gitlab.com/user/duo_agent_platform/security_threats/ — تزریق prompt، هویت ترکیبی، تهدیدهای عامل‌محور
- **GitLab Duo CLI agentic AI** — https://about.gitlab.com/blog/gitlab-duo-cli/ — حالت تعاملی با HITL به‌صورت پیش‌فرض
- **Atlassian Rovo agents practical guide (Devoteam)** — https://www.devoteam.com/expert-view/atlassians-rovo-agents/ — سه سطح خودمختاری و حاکمیت در چرخه
- **Atlassian Rovo agents in Jira Service Management** — https://www.atlassian.com/blog/jira-service-management/rovo-agents-service-desk-instant-resolution — خودکارسازی عامل‌محور در JSM، حل مبتنی بر دانش
- **SFIA 9 - a framework for AI skills** — https://sfia-online.org/en/sfia-9/sfia-9-release-notes/ai-skills-framework — ۷ سطح مسئولیت؛ مهارت‌های AI/داده در SFIA 9 (۲۰۲۴) افزوده شد
- **SFIA Levels of Responsibility for human/AI task allocation** — https://sfia-online.org/en/sfia-9/sfia-9-release-notes/ai-skills-framework/using-sfia-levels-of-responsibility-to-analyse-what-tasks-responsibilities-to-assign-to-ai — راهنمای تقسیم کار میان انسان و هوش مصنوعی با استفاده از ۷ سطح
- **CNCF Platform Engineering Maturity Model** — https://tag-app-delivery.cncf.io/whitepapers/platform-eng-maturity-model/ — سطوح بلوغ در ابعاد سرمایه‌گذاری، پذیرش، رابط‌ها، عملیات و سنجش
- **OWASP Top 10 for LLM Applications 2025** — https://genai.owasp.org/resource/owasp-top-10-for-llm-applications-2025/ — فهرست رسمی رتبه‌بندی‌شده ریسک‌های LLM در سال ۲۰۲۵ (LLM01-LLM10)
- **OWASP LLM10:2025 Unbounded Consumption** — https://genai.owasp.org/llmrisk/llm102025-unbounded-consumption/ — تأیید گسترش LLM10 از Model DoS به denial-of-wallet/ استخراج مدل
- **ISO 42001 Annex A Controls Explained (ISMS.online)** — https://www.isms.online/iso-42001/annex-a-controls/ — ۳۸ کنترل ویژه هوش مصنوعی در ۹ هدف؛ الگوی هم‌سبک با ISO 27001

- **JPMorgan Chase AI assistant (LLM Suite) - CNBC** — https://www.cnbc.com/2024/08/09/jpmorgan-chase-ai-artificial-intelligence-assistant-chatgpt-openai.html — دستیار داخلی مبتنی بر OpenAI که برای کارکنان عرضه شد
- **JPMorgan Chase to equip 140K workers with GenAI - CIO Dive** — https://www.ciodive.com/news/JPMorgan-Chase-LLM-Suite-generative-ai-employee-tool/726772/ — مقیاس عرضه LLM Suite
- **LLM Suite named American Banker Innovation of the Year - JPMorganChase** — https://www.jpmorganchase.com/about/technology/blog/llmsuite-ab-award — وبلاگ شرکت؛ دروازه چندمدلی، ارتقاها و موارد کاربرد
- **Morgan Stanley Research announces AskResearchGPT - Morgan Stanley press release** — https://www.morganstanley.com/press-releases/morgan-stanley-research-announces-askresearchgpt — دستیار GPT-4o روی بیش از ۷۰ هزار گزارش پژوهشی
- **Morgan Stanley expands OpenAI chatbot to Wall Street division - CNBC** — https://www.cnbc.com/2024/10/23/morgan-stanley-rolls-out-openai-powered-chatbot-for-wall-street-division.html — دستیار/Debrief موسوم به Morgan Stanley @ AI و عرضه نهادی
- **Goldman Sachs launches AI assistant for employees - CNBC** — https://www.cnbc.com/2025/01/21/goldman-sachs-launches-ai-assistant.html — دستیار GS AI در سطح کل شرکت پس از پایلوت حدود ۱۰ هزار نفری
- **Goldman Sachs internal AI assistant - Fortune** — https://fortune.com/2025/06/24/goldman-sachs-internal-ai-assistant — نسخه ویژه توسعه‌دهندگان و مدل‌های داخلی میزبانی‌شده تأییدشده
- **Capital One Machine Learning Engineering Training Program (MLETP)** — https://www.capitalone.com/tech/machine-learning/accelerating-machine-learning-engineering-with-a-new-capital-one-skills-program/ — برنامه ۱۶۰ ساعته؛ تأکید بر هوش مصنوعی مسئولانه/تبیین‌پذیر
- **Capital One Tech College / tech career development** — https://www.capitalone.com/tech/culture/tech-future-of-everything/ — آکادمی داخلی با مسیر یادگیری ماشین و هوش مصنوعی
- **HSBC and AI** — https://www.hsbc.com/who-we-are/hsbc-and-digital/hsbc-and-ai — شوراهای بازنگری هوش مصنوعی، حاکمیت، قضاوت انسانی در مرکز
- **DBS Responsible AI in banking** — https://www.dbs.com/artificial-intelligence-machine-learning/artificial-intelligence/responsible-ai-in-banking-gaining-a-competitive-edge.html — چارچوب PURE؛ رویکرد co-pilot با انسان در چرخه
- **OCC / Fed revised model risk guidance (SR 11-7 lineage)** — https://www.occ.gov/news-issuances/bulletins/2026/bulletin-2026-13.html — مدیریت ریسک مدل سه‌خط دفاعی اعمال‌شده بر AI/GenAI