Model upgrade or data work

One card per AI project whose answers disappoint. Trace the failures before you fill in anything else; fill in what you know. Every blank line gets an owner and a date. Any blank line means: no upgrade gets priced yet.

QUESTION	ANSWER	OWNER / BY WHEN
Project, and which answers disappoint		
Sample of disappointing answers: how many, drawn from when		
What the model actually received on each (fed-in data, retrieved documents, connected systems)		
Which failures trace to data the model never received		
Which systems hold the missing answers, and who owns each		
Where our data is fragmented, poorly understood or inconsistently governed		
Sample sorted: failures sitting with the model / sitting with the data		
If most failures sit with the data, where the upgrade budget goes instead		
Which data, APIs and industry context we make AI-ready first		
Where the value sits for us: which use built on our own data		
Gold-standard eval set: who builds it, from which real questions, ready when		
Named owner of the eval set once the pilot ends		
Current model's score on the eval set, measured when and by whom		
Every replacement scored on the same set before anyone signs? (yes / no)	☐ Yes ☐ No	
Decision and date	**☐ DATA WORK FIRST ☐ UPGRADE TOO ☐ HOLD Date _____ Signed _____**	

Model upgrade or data work

Example (fictional): claims-guidelines assistant at a regional insurer

QUESTION	ANSWER	OWNER / BY WHEN
Project, and which answers disappoint	Internal assistant that answers branch staff's questions on claims-handling rules; quotes outdated rules and misses about half the questions	n/a
Sample of disappointing answers: how many, drawn from when	12 answers, drawn from last month's question log	n/a
What the model actually received on each (fed-in data, retrieved documents, connected systems)	Fed a 2023 guidelines PDF and about 400 past tickets; never connected to the underwriting wiki or the current guidelines file	n/a
Which failures trace to data the model never received	7 of 12: the answers live in the underwriting wiki or the current guidelines, neither connected	n/a
Which systems hold the missing answers, and who owns each	Underwriting wiki (underwriting lead); current claims guidelines (head of claims)	Head of claims, both connected by 30 Jan
Where our data is fragmented, poorly understood or inconsistently governed	Guidelines live in 3 places, 2 versions apart; nobody owns the definitions	Data governance lead
Sample sorted: failures sitting with the model / sitting with the data	9 with the data (7 never received, 2 from the outdated PDF); 3 with the model (misread a table)	n/a
If most failures sit with the data, where the upgrade budget goes instead	The $40k set aside for the premium model tier goes to connecting the wiki and the current guidelines	Claims operations manager
Which data, APIs and industry context we make AI-ready first	Current claims guidelines, then the underwriting wiki, then the ticket taxonomy	Data governance lead, by 28 Feb
Where the value sits for us: which use built on our own data	A claims assistant grounded in our own guidelines and past claims decisions, not a general chatbot	Head of claims
Gold-standard eval set: who builds it, from which real questions, ready when	Claims operations manager, from last quarter's 200 real questions, ready 31 Mar	n/a
Named owner of the eval set once the pilot ends	**Blank: not ready**	Head of claims, by 31 Mar
Current model's score on the eval set, measured when and by whom	**Blank: not ready**, set not built yet, so nothing has been scored	Claims operations manager, April
Every replacement scored on the same set before anyone signs? (yes / no)	☒ Yes: agreed in writing with procurement	n/a
Decision and date	**☒ DATA WORK FIRST · Date 12 Jan · "No upgrade priced until the eval set exists and the current model is scored on it."**	Head of claims