Design a Responsible AI Pilot

HR use-case planning guide & working sheets

The Problem

An organization with about 800 employees wants to make routine HR information easier to find and reduce repetitive questions to a small HR team. Employees work across locations and schedules. They currently search an intranet, email HR, or ask a manager; finding the right policy and next step can take several exchanges.

The Use Case

Leadership is considering a six-week pilot of an AI HR information assistant. It would help employees find and understand approved policies on onboarding, time off, and benefits enrollment, link to its sources, and direct individual situations to HR. It would not decide eligibility, approve leave, evaluate employees, or change personnel records.

The Data

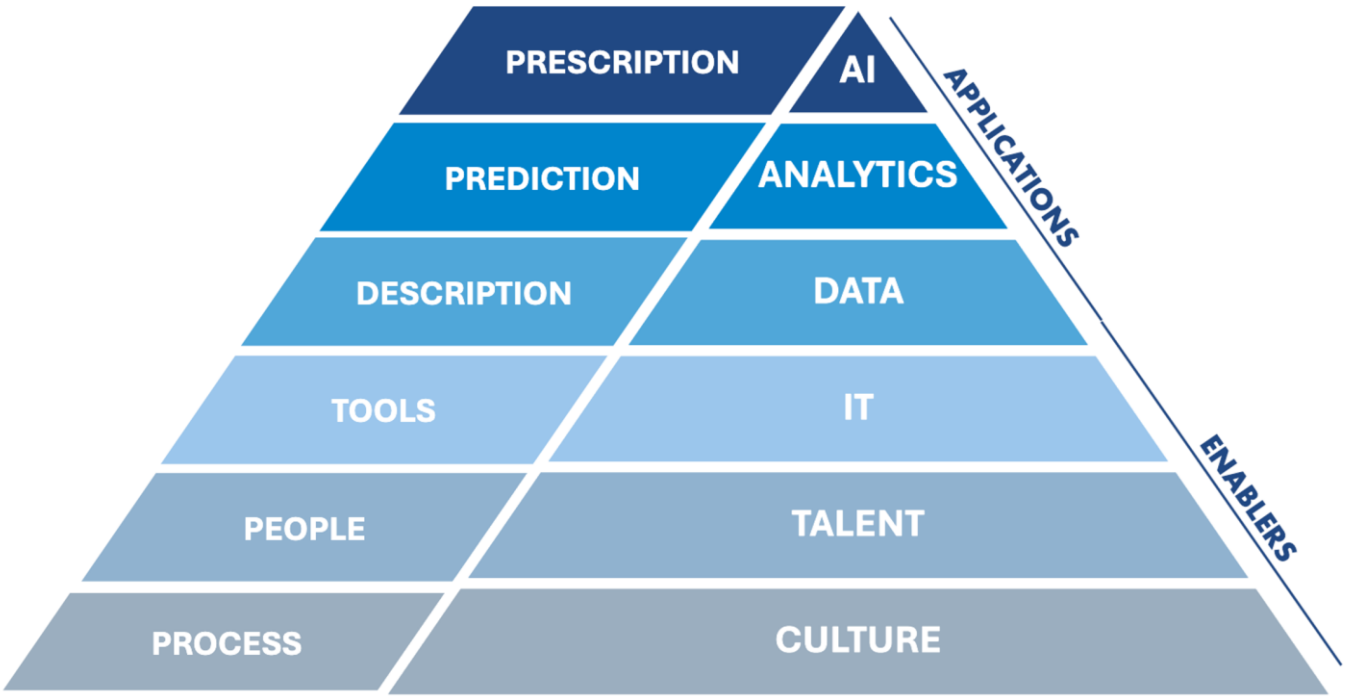
The proposed data sources include a handbook, policy pages, and general FAQs. It is unclear how consistently those materials are maintained. Individual personnel, medical, payroll, and performance records are outside the proposed scope. Employees would retain a direct route to HR.

Your Assignment

Your table is the pilot design team. Recommend the smallest, useful version worth testing, the information and people it needs, and how you would know whether it helps. Identify assumptions that must be checked. You may narrow the proposal, improve the existing process first, or recommend a simpler alternative.

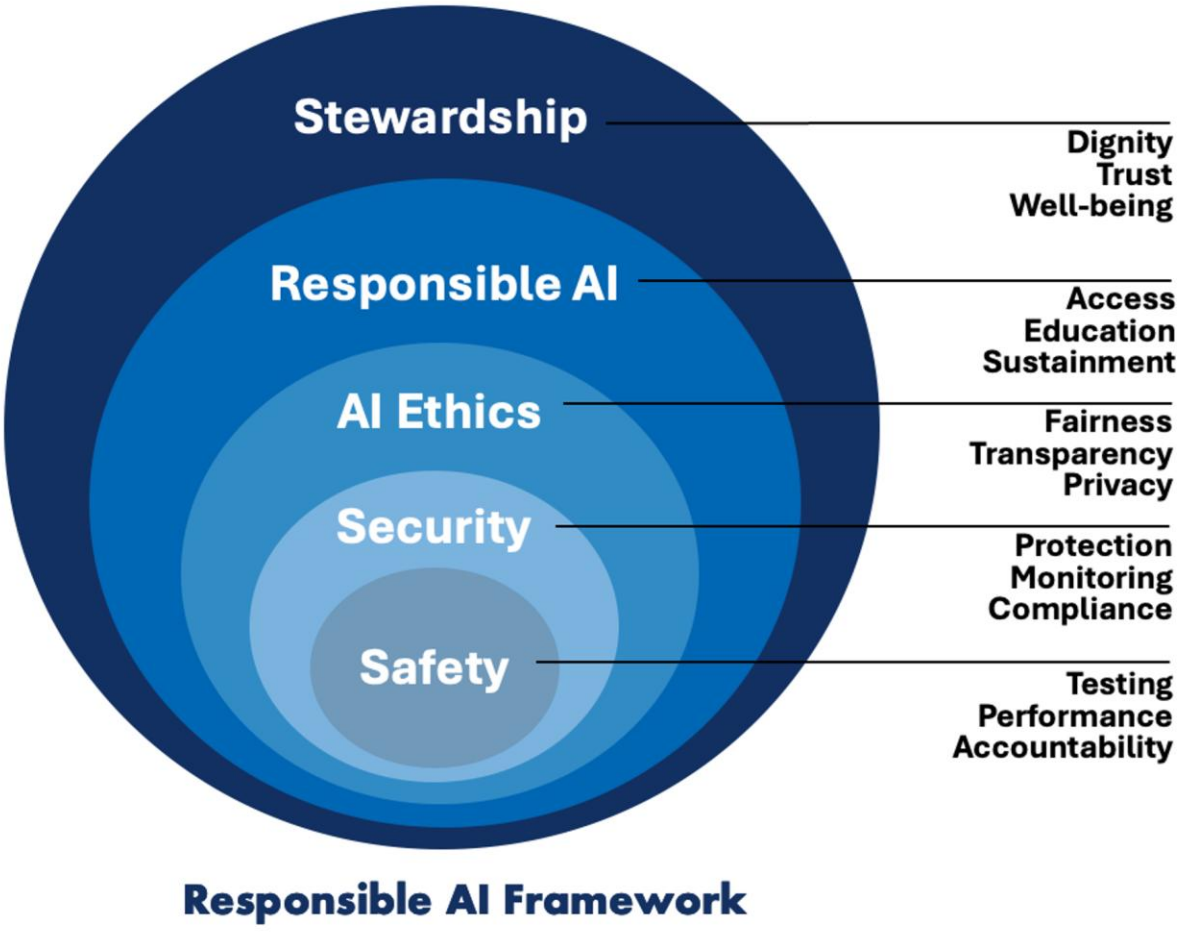
Data Readiness Pyramid: what must support the pilot?

Image layer, bottom to top	Questions to build into the case
Process / Culture	Is the HR workflow clear? Who owns it? Do employees trust the service and feel safe questioning it?
People / Talent	Who maintains content, evaluates answers, trains users, and handles exceptions? Is there capacity to do that?
Tools / IT	Are identity, permissions, approved tools, and support arrangements adequate?
Description / Data	Are policies current, authoritative, complete, and labeled for the right employee groups?
Prediction / Analytics	What evidence supports performance claims? What comparisons and subgroup checks are needed?
Prescription / AI	What may the system recommend or do? Where must it defer to a person?



Responsible AI Framework: who or what needs protection?

Image circle and terms	Suggested case-design prompt
Safety: Testing, Performance, Accountability	What harmful or wrong response must the pilot catch? Who can stop it?
Security: Protection, Monitoring, Compliance	Who can see or change information? What misuse should be detected, and what review is needed?
AI Ethics: Fairness, Transparency, Privacy	Whose experience is worse? Can users verify the source and understand data use?
Responsible AI: Access, Education, Sustainment	Can everyone use the service? Who teaches, supports, and maintains it?
Stewardship: Dignity, Trust, Well-being	Can employees seek help safely, contest an answer, and reach a person without disadvantage?



Responsible AI Framework

Participant sheet • Our responsible AI pilot

Table: _____ Complication ID: _____ Scribe: _____

Before the complication • 10–30 minutes	**Our outline**
Problem, users, and benefit	
Smallest useful pilot / excluded decisions	
Allowed data / information kept out	
Human owner, review, and escalation	
Success measure and comparison	
After the complication • 30–45 minutes	**Our revision**
What changed? Who is affected?	
Missing pyramid foundation / relevant circle	
Revised scope or approach	
Specific safeguard and accountable role	
Test, pass criterion, and evidence needed	
When to stop or involve a person	

Our recommendation: □ **Bounded test** □ **Repair first** □ **Defer / use another approach**

Report-out: "Our complication was _____. We changed _____. Our key safeguard is _____, owned by _____. Before proceeding, we need evidence that _____."

Participant sheet • Model Solution

Table: _____ Complication ID: _____ Scribe: _____

Before the complication • 10–30 minutes	**Our outline**
Problem, users, and benefit	Employees across shifts struggle to find HR guidance. Reduce repeat questions and time spent finding the right policy.
Smallest useful pilot / excluded decisions	Six weeks; 20 volunteers across shifts; benefits FAQs only. No eligibility decisions, approvals, evaluations, or record changes.
Allowed data / information kept out	HR-approved handbook, benefits pages, and general FAQs. Exclude personnel, medical, payroll, and performance records.
Human owner, review, and escalation	HR benefits lead owns content and can pause the pilot. Assigned HR reviewer checks answers daily; personal cases go to HR.
Success measure and comparison	Proposed target: 20% faster correct answers than current search. Compare matched tasks; count HR review time and repeat contacts.
After the complication • 30–45 minutes	**Our revision**
What changed? Who is affected?	Two approved-looking pages conflict; AI picks the older deadline. Employees risk missing enrollment; HR must resolve the conflict.
Missing pyramid foundation / relevant circle	Description / Data: source authority and ownership are missing. Safety: wrong deadlines; AI Ethics: misleading confidence.
Revised scope or approach	Repair first: HR verifies deadlines and removes stale sources. Use an approved FAQ and direct HR help until retesting passes.
Specific safeguard and accountable role	HR benefits lead signs off a source list with owner, version, effective date, and employee group; IT excludes all other pages.
Test, pass criterion, and evidence needed	Proposed gate: all 20 synthetic cases pass, including conflicts. Require correct cited answers or HR referral; retain test results.
When to stop or involve a person	Any wrong deadline pauses the pilot for HR review and retesting. Unclear sources or personal cases go to HR; urgent deadlines by phone.

Our recommendation: □ **Bounded test** √ **Repair first** □ **Defer / use another approach**

Report-out: "Our complication was **conflicting dates.** We changed **sources.** Our key safeguard is **verified policies,** owned by **HR.** Before proceeding, we need evidence that **conflicts reach HR."**

Take it with you

Start with one resource. The rest are there when you want them.

Tier	Resource	What it is	Example rule
Start here	NIST AI Risk Management Framework https://www.nist.gov/itl/ai-risk-management-framework	A free, voluntary guide to managing AI risk: Govern, Map, Measure, Manage.	"Every AI tool has a named owner who can pause it."
Go deeper	CISA AI security guidance https://www.cisa.gov/topics/cybersecurity-best-practices/super-intelligence	Practical steps for deploying AI securely and protecting its data.	"Chat history is restricted by default; only named roles can see it."
Go deeper	NIST Privacy Framework https://www.nist.gov/privacy-framework	Helps decide what data to collect and how long to keep it.	"Collect only what the question needs; delete logs after 30 days."
For data teams	EDM Council CDMC https://edmcouncil.org/frameworks/cdmc/	Cloud data management controls: ownership, sensitivity labels, retention.	"No data source feeds the AI until it has an owner and a sensitivity label."
For data teams	DAMA-DMBOK https://dama.org/2026/08/21/dama-dmbok-2-revised-edition-7-languages/	The common reference book for data governance, quality, and metadata.	"Every policy page shows an owner and review date; the assistant uses only pages reviewed in the last 12 months."