Project Scoping - AI Copilot for Deskless Workers

Team Members:

- Harshitkumar Brahmbhatt
- Krishna Venkatesh
- Raghav Gali
- Rishi Raj Kuleri
- Sujitha Godishala
- Swathi Baba Eswarappa

1. Introduction

Despite making up **70‑80 % of the workforce**, deskless workers remain underserved by traditional HR and knowledge systems, largely because of structural and communication barriers.

HR professionals face persistent challenges in effectively reaching and engaging this population:

- **Limited access to computers during work hours (62 %)**, **irregular schedules (56 %)**, and **lack of face‑to‑face communication (55 %)** are top barriers to communication.
- Many organizations still use standard engagement strategies (e.g. printed materials, presentations, online tools) that appear in data as common but *ineffective* for deskless workers. For instance, presentations at onboarding and one‑on‑one HR consultations are among the most effective, but many less‑direct strategies (online materials, email, internal messaging) score low in perceived effectiveness.

These gaps result in serious downstream problems: **lower benefits enrollment/utilization, weaker training adoption, lower retention, and reduced engagement**.

To address the persistent engagement and training challenges faced by deskless workers, we propose an **AI Copilot powered by retrieval-augmented generation (RAG)**. The system is designed to provide **accurate, grounded responses** to policy and training-related queries by retrieving relevant information from internal documents such as employee handbooks, safety protocols, and HR guidelines.

Built on top of this RAG foundation, the copilot also incorporates an **agentic layer** that enables it to go beyond answering questions — by initiating follow-up actions like scheduling training, verifying compliance deadlines, and escalating unresolved queries to HR. This layered approach ensures reliability, actionability, and adaptability, making the AI Copilot a practical and scalable tool for supporting frontline teams in real-world settings.

As a future enhancement, we also plan to integrate **voice interaction capabilities**, enabling even greater accessibility for workers in hands-busy, on-the-move environments.

2. Dataset Information

1. **Dataset Introduction**: The dataset consists of **employee handbook Q&A pairs** generated from a diverse set of employee handbooks across industries (healthcare, retail, logistics, construction, finance, etc.).

- **Purpose**: To train and evaluate a retrieval-augmented generation (RAG) system that can answer HR and policy questions for deskless workers.

- **Relevance**: Deskless employees face limited access to HR resources due to irregular schedules, lack of computer access, and fragmented communication. By converting handbook policies into structured Q&A, the dataset ensures that HR policies are accessible via natural language queries.

2. **Data Card**:

Name: Deskless Worker Handbook Q&A Dataset
Size: ~200–20,000 Q&A pairs (depending on expansion & synthesis)
Sources: Employee handbooks across multiple industries (see Data Sources section)
Format:

- **CSV**: for exploration
- **JSONL**: for fine-tuning/instruction-tuning models
- **PDF/TXT**: original handbooks for retrieval embedding
- **Data Types**:

  Natural language questions (employee queries)
  Concise answers (policy excerpts/handbook guidance)
  Metadata (source handbook, industry, section)

3. **Data Sources**:

Datasets are compiled from publicly available **employee handbooks** across multiple industries. Examples include:

Healthcare: Crouse Medical Handbook (2019)

Cleaning & Maintenance: CleanSpace Employee Handbook (2024)

Retail: Lunds & Byerlys Handbook (2019)

Hospitality: Alta Peruvian Lodge Handbook (2016)

Finance: Old National Bank Handbook

Automobile: Lowe Auto Handbook (2023)

4. **Data Rights and Privacy**:

- **Source Material**: All handbooks are **publicly available** PDFs hosted on official company or nonprofit websites.
- **Usage Rights**: The dataset is intended strictly for **research and educational purposes**. Redistribution of proprietary content without permission is not allowed.
- **Privacy**: No personal or employee-identifiable data is included. Only general policy text (holiday schedules, leave rules, conduct policies) is used.
- **Compliance**: Dataset preparation respects **GDPR/CCPA principles** by excluding personal data. Policies extracted are organizational, not individual.

3. Data Planning and Splits

Preprocessing Steps

1. **Loading**:
   - Extract raw text from PDFs (using tools like PyMuPDF or PDFMiner).
   - Store metadata: document title, source URL, industry.
2. **Preprocessing**:
   - **Chunking**: Split handbook text into manageable policy sections (e.g., paragraphs, bullet points).
   - **Cleaning**: Remove headers, footers, duplicates, and formatting artifacts.
   - **Q&A Generation**: Convert each policy chunk into multiple natural questions + concise answers (manual + synthetic generation).
   - **Normalization**: Ensure consistent format (JSONL with `{"instruction": ..., "output": ...}` schema).
3. **Managing Data**:
   - Store handbooks + processed Q&A in a structured repository.
   - Tag entries with **industry**, **policy type** (leave, benefits, conduct, etc.), and **source**.

   - Version control via Git/GitHub.

Splitting Strategy

- **Training (70%)**: Main set of Q&A pairs for fine-tuning models.
- **Validation (15%)**: Held-out Q&A pairs for tuning hyperparameters and preventing overfitting.
- **Test (15%)**: Used for final evaluation.

Additional considerations:

- **Stratified splits** by industry (ensures retail, healthcare, etc. are represented in all splits).
- **Deduplication checks** to avoid leakage (no near-identical Q&A in both train/test).
- **Synthetic vs. Human-curated**: Keep some synthetic data in validation/test to evaluate robustness.

4. GitHub Repository

https://github.com/Raghavgali/MLOps-Project-

5. Project Scope

1. **Problems:**
   - **Limited access to digital tools during work hours:**
     Deskless workers often lack consistent access to computers, email, or HR portals while on the job, making it difficult to consume time-sensitive or policy-related information.
   - **Training is hard to access and retain:**
     Most training materials are delivered in formats that are not optimized for mobile or real-time usage. Workers struggle to recall or revisit training content during critical tasks.
   - **Low engagement with HR programs and benefits:**
     Due to communication gaps, deskless employees often miss out on wellness programs, compliance deadlines, and benefits enrollment windows.
   - **Fragmented systems lead to poor user experience:**
     Training modules, HR data, schedules, and policy documents are siloed across LMS platforms, HRIS tools, and internal document stores, with no unified interface.
   - **Inaccuracy from generic AI assistants:**
     Without grounding in internal policy documents, standard LLMs can hallucinate responses, leading to confusion and compliance risks.
   - **Lack of accessibility in field environments:**
     Workers in mobile, hands-busy roles need alternatives to traditional point-and-click interfaces, especially when operating machinery or moving frequently.
2. **Current Solutions:**

Enterprise HCM Suites (e.g., Oracle, SAP SuccessFactors, Workday):

These platforms offer end-to-end HR functionality, including scheduling, benefits, performance tracking, and sometimes embedded digital assistants.

However, they are:

- Closed-source and vendor-locked, making customization and integration with external tools difficult
- Optimized for HR administrators, not frontline employees
- Lacking in dynamic, conversational interfaces for in-the-moment knowledge retrieval

Digital Assistants in Enterprise Systems (e.g., Oracle Digital Assistant):

Prebuilt bots embedded into HCM suites support common workflows like leave requests or benefits FAQs.

Limitations include:

- Rigid conversational flows tied to backend schema
- Limited flexibility to support custom policy logic or domain-specific workflows
- Complex tooling is required for expansion

Learning Management Systems (LMS) with Microlearning:

Platforms like Cornerstone, Docebo, or TalentCards deliver bite-sized training content to mobile users.

Challenges:

- Content is static and not retrievable on demand by query
- Lacks personalization or dialogue-based reinforcement
- Often disconnected from the real-time operational context

Generic HR Chatbots (e.g., Leena AI, Talla):

These bots offer conversational interfaces for common HR questions.

They are often:

- Limited to pre-scripted FAQ flows without deeper document grounding
- Not capable of tool orchestration or decision-making
- Inflexible in adapting to new policy content or emerging queries

Self-Service Portals and HR Intranets:

Web-based systems allow employees to log in and view HR content.

Limitations:

- Desktop-centric; not mobile-optimized for real-world deskless scenarios
- Require proactive effort from the user to locate information
- Provide no conversational interface or natural query support

Internal Communication Tools (e.g., Slack, Teams):

Many HR teams use these channels to push announcements or policy updates.

Issues:

- Messages are transient and easy to miss
- No intelligent retrieval, summarization, or context adaptation
- Users still have to “know what to ask” and “where to look”

3. Proposed Solution:

Retrieval-Augmented Generation (RAG) for Grounded Answers

- The AI Copilot retrieves relevant chunks from company-specific documents (e.g., safety manuals, benefits guides, onboarding material).
- Ensures that answers are accurate, contextually relevant, and cited from authoritative sources.

Agentic Layer for Orchestration and Actions

- Beyond answering questions, the system can schedule trainings, verify completion, escalate unclear issues, or update HR systems.
- This is achieved through integration with internal APIs, tools, and workflow logic.

Contextual Personalization and Memory

- Remembers user history (e.g., training completed, prior queries) to provide relevant follow-ups.
- Supports multi-turn conversations with persistent context.

System Integration and Interoperability

- Connects to LMS, HRIS, calendar tools, and document stores to provide a unified interface.
- Reduces friction by pulling live data and avoiding duplicate interactions.

Safe Fallback and Compliance

- When uncertain, the system provides source references or offers to escalate to a human HR representative.
- All interactions are logged for audit and review.

Voice Interaction for Real-Time, Hands-Free Use

- Allows employees to ask and receive answers via speech, reducing friction in hands-busy environments (e.g., warehouses, hospital floors, field service sites).
- Uses proven speech-to-text (STT) and text-to-speech (TTS) systems to enable accessibility.

6. Current Approach Flow Chart and Bottleneck Detection

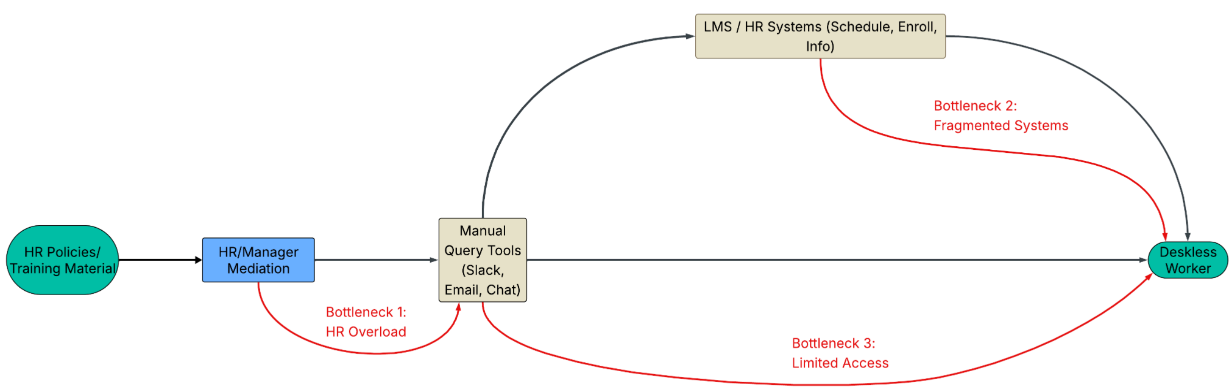

Figure 1: Traditional Flow of Deskless workers

The diagram above illustrates the traditional flow through which deskless workers seek HR or training-related information. Typically, employees rely on **HR managers** to mediate access to resources like policy documents or training schedules. These requests are often routed through **manual communication tools** such as Slack, email, or informal chat, before reaching systems like LMS platforms or HR portals.

However, this setup introduces three critical bottlenecks:

1. **HR Overload:** All queries funnel through HR or management, creating delays and inconsistency in responses.

2. **Fragmented Systems:** The LMS, HRIS, and document repositories are siloed, forcing workers to navigate multiple disconnected tools.

3. **Limited Access for Workers:** Deskless employees often lack convenient access to these systems during working hours, making it difficult to retrieve information in real-time.

These bottlenecks result in delays, missed training deadlines, policy confusion, and low engagement with HR programs — all of which reduce workforce effectiveness and compliance.

Improvements:

7. Metrics, Objectives, and Business Goals

Project Objectives:

- **Develop a RAG-based AI Copilot grounded in company policies**
  Build a reliable, retrieval-augmented system that can accurately answer HR, training, and policy-related queries by pulling directly from internal documents. Ensure answers are grounded, traceable, and explainable via citations.

- **Enable action-oriented task execution via an agentic layer**
  Extend the copilot beyond information retrieval by integrating agent-based orchestration that can perform tasks such as scheduling training, checking compliance status, or escalating unresolved issues to HR systems.

- **Integrate natural language interfaces accessible to deskless workers**
  Provide seamless access through a conversational interface with optional voice interaction, enabling workers in mobile, hands-busy, or offline environments to retrieve information or complete tasks without needing desktop access.

- **Ensure safety, auditability, and fallback mechanisms**
  Design the system to handle uncertainty gracefully by including confidence thresholds, falling back to HR escalation, and full interaction logging for transparency and compliance review.

- **Evaluate usability, trust, and system performance**
  Measure system accuracy (RAG), task completion rate (agentic), latency, and speech clarity (voice), and user feedback on helpfulness and satisfaction. Use metrics to inform continuous improvements.

Business Goals Alignment:

Business Goal	How the Project Supports It
Improve HR efficiency	Reduces the HR team's workload by automating common policy queries, training requests, and follow-ups.
Increase training compliance	Enables timely scheduling and reminders for mandatory training, reducing the risk of missed deadlines.
Boost employee engagement	Offers an accessible, responsive support experience tailored to the needs of deskless employees.
Enhance policy compliance and reduce risk	Minimizes misinformation by grounding answers in source documents and logging all interactions for audit.
Enable scalable, cost-effective support	Avoids the need for hiring additional HR staff as the organization scales; the AI copilot can handle growing query volumes.
Foster innovation and AI adoption in HR workflows	Demonstrates how AI and LLMs can be responsibly applied to real business processes, laying the foundation for broader digital transformation.

8. Key Metrics

RAG:

Metric	Description	Goal/Target
Retrieval Recall@k	% of gold/reference answers where the correct supporting document chunk was retrieved in top-k	>90% Recall@5
Answer Factuality/ Citation Match Rate	% of answers where cited content exits in retrieved documents	> 85%
Exact Match (EM)	% of generated answers that exactly match the gold answer (in QA setting)	> 70%
F1 Score(Token - level)	Measures overlap between predicted and	> 80%

	ground-truth answers	
BLEU/ROUGE Scores	Evaluates semantic similarity between generated and reference answers (for open Q&A)	ROUGE-L >0.6
Hallucination Rate	% of answers that contain unsupported or invented facts	< 5%
User-rated Helpfulness	Manual or crowd feedback on helpfulness of answers	> 4/5 avg rating

Agentic Layer (Tool Use, Action-Oriented AI):

Metric	Description	Goal/Target
Tool Accuracy	% of times the correct tool was selected and invoked with the correct parameters	>90%
End-to-End Task Success Rate	% of multi-step task (e.g. “schedule training”, “check compliance”) completed correctly	> 85%
Fallback/Recovery Rate	% of low-confidence cases where agent safely falls back or escalates rather than guessing	> 95%
Average Task Completion Time	Time taken from user input to final action (including tool calls)	< 5 seconds
User Confirmation Accuracy	% of cases where the agent confirms with the user before taking irreversible action (e.g., scheduling)	100%
Action Audit Log Completeness	% of tool calls and agent actions properly logged for traceability	100%

Voice Interaction Feature :

Metric	Description	Goal/Target
Word Error Rate(WER)	Error rate of speech-to-text transcription(lower is better)	< 10%
Response Latency	Total time from voice input to audio reply	< 3 seconds
TTS Clarity Score	Subjective user rating of text-to-speech naturalness and clarity	> 4/5
End-to-End Voice Task Completion	% of queries correctly handled fully via voice	> 80%
Fallback to Text Mode	% of voice failures that gracefully revert to text	> 95%
Microphone Accessibility / Compatibility	Number of supported input modes(desktop mic, mobile mic, browser)	Broad Support (> 2 platforms)

FORMATTED TILL HERE

8. Failure Analysis
➔ Discuss potential risks, including what could go wrong during the project and after deployment, and provide an analysis of pipeline failures and mitigation strategies.

9. Deployment Infrastructure
➔ Provide detailed information about the infrastructure required to deploy your project, along with a list of supported platforms. Be sure to include necessary flowchart Diagrams.

10. Monitoring Plan
➔ Provide a broad description of your monitoring plan, including what you intend to monitor and why. Prepare for detailed documentation.

11. Success and Acceptance Criteria
➔ Define the criteria for success and acceptance of the project.

12. Timeline Planning
➔ Create a preliminary project timeline, which can be modified based on given deadlines and constraints.

13. Additional Information
➔ Include any other relevant information you believe is necessary for a comprehensive project scoping submission.