IBM x Columbia Weekly Meeting
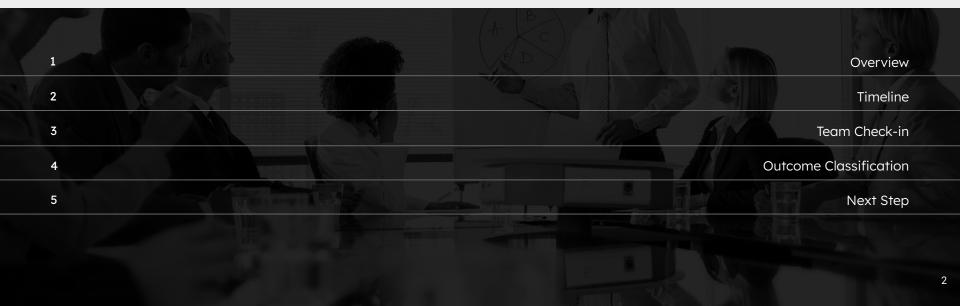
IBM x Columbia Core Team

10/16/2025

AGENDA

Why we're here today

Overview

This weekly meeting will be focusing on establishing a clear roadmap and reviewing initial progress for the IBM x Columbia project. Including the timeline, team check-in, and some Q&A at the end.

Timeline

Phase	Start Week	End Week	W1 09/23-09/29	W2 09/30-10/06	W3 10/07-10/13	W4 10/14-10/20	W5 10/21-10/27	W6 10/28-11/03	W7 11/04-11/10	W8 11/11-11/17	W9 11/18-11/24	W10 11/25-12/01
Initiation & Planning	1	2										
Objective Clarification			▓									
Literature Review			▓	▓								
Timeline & scope			▓	▓								
Data Exploration & Requirements	3	5										
Data Collecting & Preprocessing					▓	▓	▓					
Data Cleaning & Structuring					▓	▓	▓					
RAG Pipeline Prototyping					▓	▓	▓					
Prototype Development (POC)	3	7										
MCP Workflow Diagrams					▓	▓						
Wireframes						▓	▓					
Mockups						▓	▓	▓	▓			
Prototype Implementation							▓	▓	▓			
Full Solution Development & Testing	8	9										
Performance Test										▓	▓	
Usability feedback										▓	▓	
Final Presentation & Reporting	9	10										
Project Powerpoint											▓	▓
Project Report											▓	▓
Demo Presentation												▓

Team Check-in

1. Domain & Requirements Team

2. Data & Knowledge Engineering Team

3. Generative AI & Model Development Team

4. User Experience & Application Development Team

5. Evaluation & Compliance Team

Evaluation & Compliance Team

Category	Metric	Description
Text-based Metrics	BLEU	Measures overlap between generated and reference text (translation).
Text-based Metrics	ROUGE	Compares n-gram overlap in text summarization.
Text-based Metrics	Perplexity	Evaluates fluency by measuring probability distribution in language models.
Image-based Metrics	FID	Quantifies realism of generated images by comparing distributions.
Image-based Metrics	Inception Score (IS)	Measures image quality and diversity using a pre-trained classifier.
Diversity Metrics	Embedding Space Analysis	Evaluates variability and originality of generated outputs.
Computational Metrics	Latency	Analyzes time taken to generate outputs.
Computational Metrics	Throughput	Measures number of outputs generated per unit time.

reference:https://www.ijfmr.com/papers/2025/1/37182.pdf
Key words:Evaluation and testing process for generative ai

Identified and Quantified Pain Points: Completed deep dive into Jack's document, successfully framing core operational pain points.
Establishing Baselines: Currently defining critical baseline metrics for success.
Low-Cost Tool Strategy: External research is focused on low-cost/community tools to track metrics, avoiding significant new licensing costs.
Goal: Deliver a rigorous, data-backed metric system with maximum analytical value and minimal new investment.

Domain & Requirement Team

- Reviewed Development Phase templates to get required data

- In progress: Summarized required input fields and structure for each template — repetitive tasks

- List for sections with information gaps

Version History / Approval Info	Training Plan, Test Case, SLA/MOU, O&M Manual
Roles and Responsibilities	Training Plan, SLA/MOU, O&M Manual
Contact Information / Key Contacts	Training Plan, SLA/MOU, O&M Manual
Requirements Section	Training Plan, SLA/MOU, Test Case
Service or Operational Info	SLA/MOU, O&M Manual, Training Plan (under Training Resources)

Data & Knowledge Engineering Team

- Completed preprocessing and text extraction for three CDC UP templates, converting them into a structured format for the knowledge base.

- Helped subteam members set up GitHub collaboration workflow, including guidance on cloning the repo, pulling before editing, committing, and pushing updates.

Before

After

Table data was imported and can be adjusted.

phase	deliverable	section_title	related_eplc_reference	source	summary
Development	Training Plan	1.0 Introduction	EPLC Framework – Development Phase	CDC UP Training Plan Template	[This section of the Training
Development	Training Plan	1.1 Background and Scope	EPLC Framework – Development Phase	CDC UP Training Plan Template	[This subsection of the Tra
Development	Training Plan	1.2 Points of Contact	EPLC Framework – Development Phase	CDC UP Training Plan Template	[This subsection of the Tra
Development	Training Plan	1.3 Document Organization	EPLC Framework – Development Phase	CDC UP Training Plan Template	[This subsection of the Tra
Development	Training Plan	1.4 Security and the Privacy Act	EPLC Framework – Development Phase	CDC UP Training Plan Template	[This subsection of the Tra
Development	Training Plan	1.5 Glossary	EPLC Framework – Development Phase	CDC UP Training Plan Template	[This subsection of the Tra
Development	Training Plan	2.0 Instructional Analysis	EPLC Framework – Development Phase	CDC UP Training Plan Template	
Development	Training Plan	2.1 Needs and Skills Analysis	EPLC Framework – Development Phase	CDC UP Training Plan Template	[This subsection of the Tra
Development	Training Plan	2.2 Development Approach	EPLC Framework – Development Phase	CDC UP Training Plan Template	[This subsection of the Tra
Development	Training Plan	2.3 Issues and Recommendations	EPLC Framework – Development Phase	CDC UP Training Plan Template	[This subsection of the Tra
Development	Training Plan	3.0 Instructional Methods	EPLC Framework – Development Phase	CDC UP Training Plan Template	
Development	Training Plan	3.1 Training Methodology	EPLC Framework – Development Phase	CDC UP Training Plan Template	[This subsection of the Tra
Development	Training Plan	3.2 Training Database	EPLC Framework – Development Phase	CDC UP Training Plan Template	[This subsection of the Tra
Development	Training Plan	3.3 Testing and Evaluation	EPLC Framework – Development Phase	CDC UP Training Plan Template	[This subsection of the Tra
Development	Training Plan	4.0 Training Resources	EPLC Framework – Development Phase	CDC UP Training Plan Template	
Development	Training Plan	4.1 Course Administration	EPLC Framework – Development Phase	CDC UP Training Plan Template	[This subsection of the Tra
Development	Training Plan	4.2 Resources and Facilities	EPLC Framework – Development Phase	CDC UP Training Plan Template	[This subsection of the Tra
Development	Training Plan	4.3 Schedules	EPLC Framework – Development Phase	CDC UP Training Plan Template	[This subsection of the Tra
Development	Training Plan	4.4 Future Training	EPLC Framework – Development Phase	CDC UP Training Plan Template	[This subsection of the Sof
Development	Training Plan	5.0 Training Materials List	EPLC Framework – Development Phase	CDC UP Training Plan Template	
Development	Training Plan	5.1 Purpose and Scope	EPLC Framework – Development Phase	CDC UP Training Plan Template	The purpose of this docum
Development	Training Plan	5.2 Training Materials List	EPLC Framework – Development Phase	CDC UP Training Plan Template	[Populate the table below v
Development	Training Plan	6.0 Training Curriculum	EPLC Framework – Development Phase	CDC UP Training Plan Template	[This section of the Training

GenAI & Model Development

Model	Core Features	Applicable Scenarios	Overall Evaluation
GPT - 5 Nano	• Equipped with reasoning tokens • Supports large context (million-level tokens)	• EPLC document validation • Q&A and content generation	• Most economical • Best-performing
GPT - 4.1 Nano	• "Balanced" mid-tier model • Reliable but without reasoning tokens	• Lightweight Q&A • Moderately complex tasks	• Stable • More expensive
GPT - 4o Mini	• Fastest speed	• Automation scripts • UI Q&A	• High cost • Low performance

Comparison of GPT API Options (Budget & Efficiency)

Most Efficient Option: **GPT-5 Nano** → Optimal balance = reasoning × speed × cost

- Reasoning tokens → better for
  ▫ Policy & compliance logic
  ▫ Document Q&A (RAG)
  ▫ Structured text generation

- Most cost-efficient among all models
- Balanced speed

GenAI & Model Development

- Optimal Range of Budget: $150-$200

API Output Testing In Progress:

- Building API Testing pipeline using python +Json input
- Standardized Prompts and outputs
- Building a small EPLC database for testing (5-10 chunks)

User Experience & Application Development Team

- In progress:
  - Human in progress workflow (UX workflow)
  - Conversation script
    - Identify 6 key stages:
      - Onboarding & greeting
      - Source documents intake
      - Users' familiarity identification
      - Content creation
      - Review & output
      - Feedback
- Next Week:
  - Wireframe drawing

Outcome Clarification

NEXT STEPS

Keep up with the timeline, communicate with the sub-team for better support. Diving into the topics we've touched on today to deepen our understanding on the project, and starting the data preparation procedure.

THANK YOU!

Columbia Practicum Core Team