HuMEL Emerging Technologies in Humanitarian Settings Series Session 1:

Humanitarian Data Insights Project

June 2024

datakind.org | @DataKind

Learn more & join our community of practice

HuMEL Peer Network
COLLABORATE, EVALUATE, EMPOWER
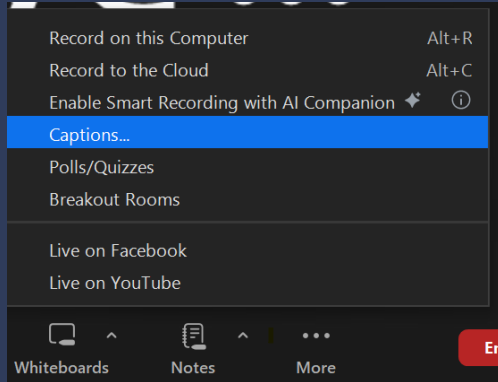
House Keeping

- Activate your video when you are speaking, if possible.
- Mute your microphone when you are not speaking.
- During Q&A, raise your hand if you'd like to speak out loud.
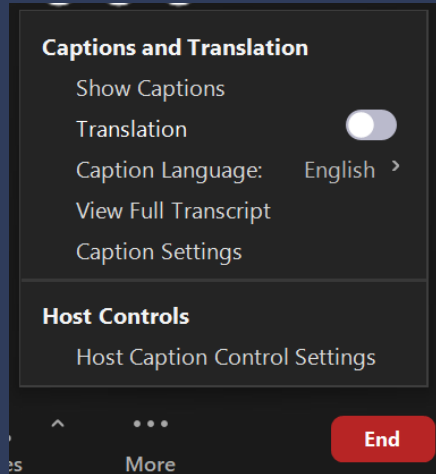- Use the chat box if you have any questions or comments!

HuMEL Peer Network
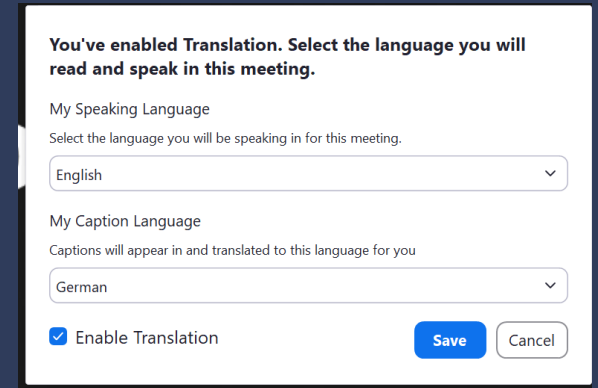COLLABORATE, EVALUATE, EMPOWER

Translated Captions

Step 1: Click Captions under "More"

Step 2: Switch translation on

Step 3: Select language

Record on this Computer — Alt+R
Record to the Cloud — Alt+C
Enable Smart Recording with AI Companion
Captions...
Polls/Quizzes
Breakout Rooms
Live on Facebook
Live on YouTube

Whiteboards   Notes   More   En

Captions and Translation
Show Captions
Translation
Caption Language   English
View Full Transcript
Caption Settings

Host Controls
Host Caption Control Settings

More   End

You've enabled Translation. Select the language you will read and speak in this meeting.

My Speaking Language
Select the language you will be speaking in for this meeting.
English

My Caption Language
Captions will appear in and translated to this language for you
German

☑ Enable Translation   Save   Cancel

HuMEL Peer Network
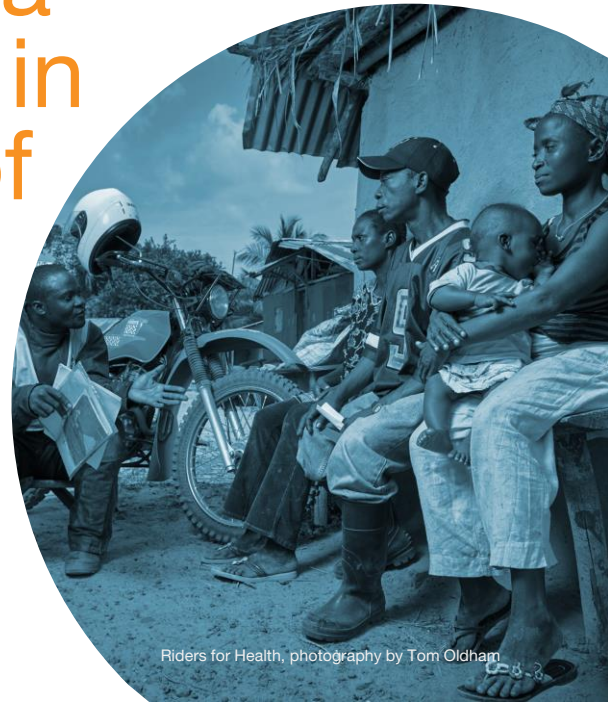COLLABORATE, EVALUATE, EMPOWER

Presentation overview

1. DataKind Overview

2. Humanitarian Data Insights Project: overview

3. User and solutions research: summary

4. Humanitarian AI Assistant: a sector-facing product

5. Discussion

DataKind

MF

DataKind®

Harnessing the power of data science + AI in the service of humanity

Riders for Health, photography by Tom Oldham

Sustainability

Success is when DataKind and its partners have created data science + AI solutions that are adopted widely and contribute positively to sector-level problems.

Accessible
Open source where possible and offering best in class documentation and training opportunities for end users

Flexible
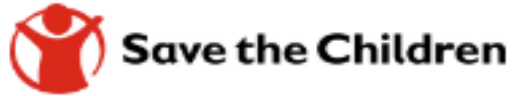Allow for user-defined customization for local context

Reusable
Allow for transferability between locations and contexts

DataKind®

Timely and accurate data and data analysis is a key component of

- planning a humanitarian response
- coordinating with other agencies
- planning and acting upon the mitigation and preparedness phases of disaster risk reduction.

The problem of **"combining meaningful data meaningfully"** and empowering the technical teams that are deployed to crises to more rapidly conduct data analysis to inform program response are consistently high pain points felt across humanitarian teams.
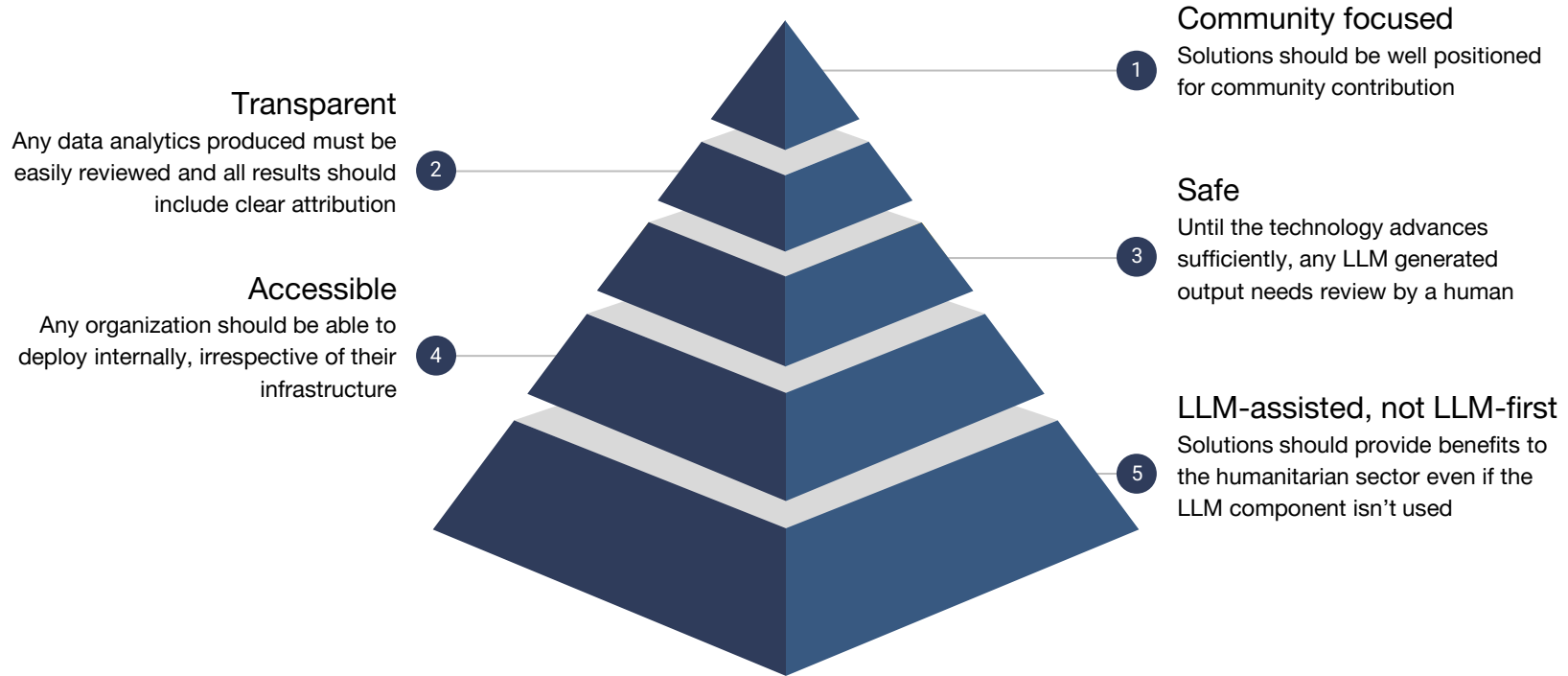
Humanitarian Data Insights Project

DataKind, Save the Children and Microsoft joined forces to design and build a set of self-contained, AI-powered tooling to enable humanitarian actors overcome persistent and pervasive data pain points to prioritize action, coordinate more effectively, and make more timely data-informed decisions.

Objective: Building on a history of data analytics projects, create community-centric, transparently designed, cost-conscious solutions that can be operationalized by any humanitarian organization

DataKind

Using GenAI safely

- DataKind has a track record of working with nonprofit and humanitarian organizations to harness the power of data for social good and places a strong emphasis on ethical data use.

- As a trusted partner for such initiatives making use of the best-in-class tools and techniques, in a safe, responsible way is core to our approach

- We explore new techniques, such as Gen AI, but not in place of safer and more powerful existing techniques

- We contributed to, and significantly aligned with, the UN-OCHA report on the guardrails for safe and ethical use of Generative AI for Humanitarians

Design principles

Community focused
Solutions should be well positioned for community contribution

1

Transparent
Any data analytics produced must be easily reviewed and all results should include clear attribution

2

Safe
Until the technology advances sufficiently, any LLM generated output needs review by a human

3

Accessible
Any organization should be able to deploy internally, irrespective of their infrastructure

4

LLM-assisted, not LLM-first
Solutions should provide benefits to the humanitarian sector even if the LLM component isn't used
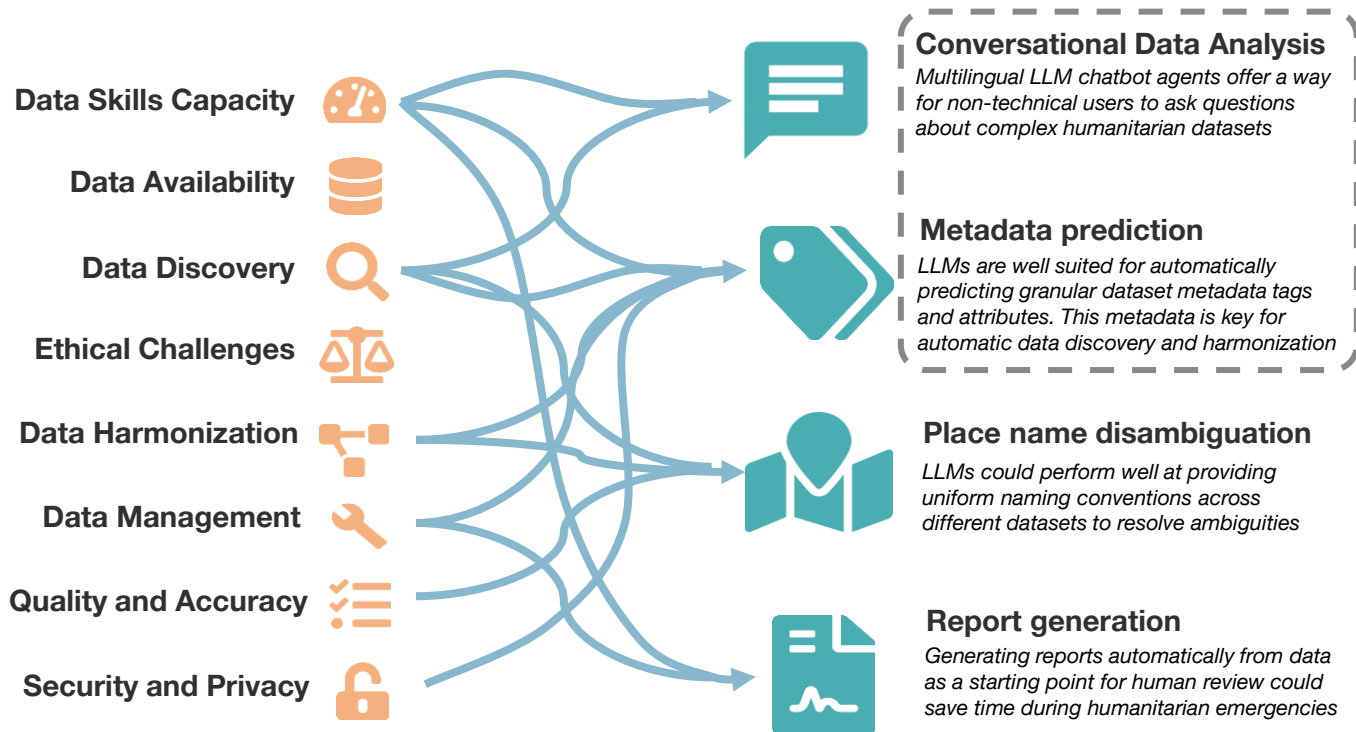
5

Research

A taxonomy of pain points mapped to potential GenAI solutions

DataKind®

Community informed priorities

Actors across the humanitarian community raised similar data challenges hindering effective response & coordination - for which a taxonomy was developed

Taxonomy	Data Pain Points
Data Harmonization	Standardization, interoperability, GIS linkage, data analysis
Data Discovery	Sharing, overwhelming amount of data, search, relevance, silos, delivery
Data Availability	Completeness, frequency, timeliness, localization & inclusiveness, disaggregation and mapping resolution
Ethical Challenges	Standards & regulations, informed consent, misuse and distortion, bias, retention
Data Skills Capacity	Literacy & awareness; capacity building & training, technical infrastructure
Security and Privacy	Data damage, security & permissions, privacy, regulations & standards, governance, risk analysis
Quality and Accuracy	Consistency, bias, accountability, verifiability
Data Management	Collection, sustainability, stress testing, consultation, evaluation and improvement, ownership, integration, scalability, technical infrastructure

DataKind

Mapping pain points to GenAI solution applications

Data Skills Capacity

Data Availability

Data Discovery

Ethical Challenges

Data Harmonization

Data Management

Quality and Accuracy

Security and Privacy

Conversational Data Analysis
Multilingual LLM chatbot agents offer a way for non-technical users to ask questions about complex humanitarian datasets

Metadata prediction
LLMs are well suited for automatically predicting granular dataset metadata tags and attributes. This metadata is key for automatic data discovery and harmonization

Place name disambiguation
LLMs could perform well at providing uniform naming conventions across different datasets to resolve ambiguities

Report generation
Generating reports automatically from data as a starting point for human review could save time during humanitarian emergencies

Humanitarian Data Pain Point Category

Generative AI Solution

DataKind

From individual solutions to sector-wide products

3+ community workshops and deep-dives into prototypes

4+ external convenings

**10+** organization interviews with stakeholders/potential users

- **Humanitarian AI Assistant:** a LLM powered conversational tool to facilitate efficient and accurate data analysis of publicly available humanitarian datasets

- **HXL Metadata Prediction Analysis:** a LLM model capable of predicting granular metadata tags on thousands of datasets on HDX demonstrating efficiency gains over current, manual processes + enabling humanitarians to automate data discovery and harmonization.

DataKind

Humanitarian AI Assistant

An LLM powered conversational tool

Product development: Humanitarian AI Assistant

Goal: making usable and useful data analysis possible, with lower barriers to access for non-technical users to discover, explore and ask questions of humanitarian datasets, so that humanitarians and front-line responders can act with data-informed responses and organizations can steward data resources more effectively to producing timely and actionable information.
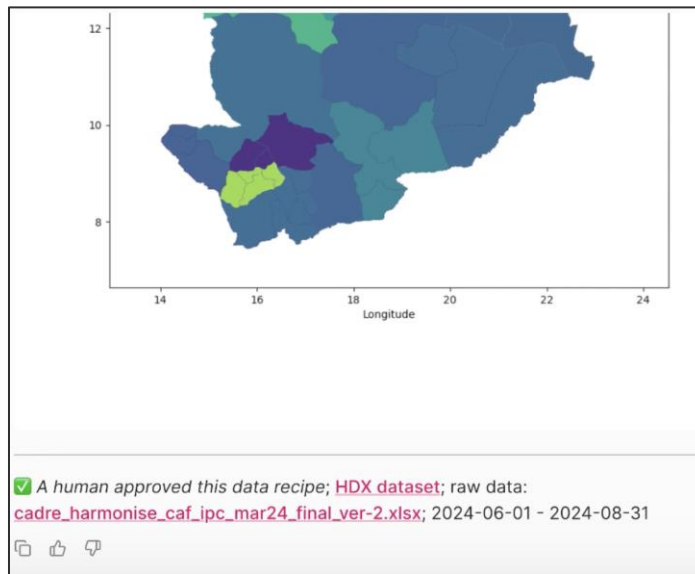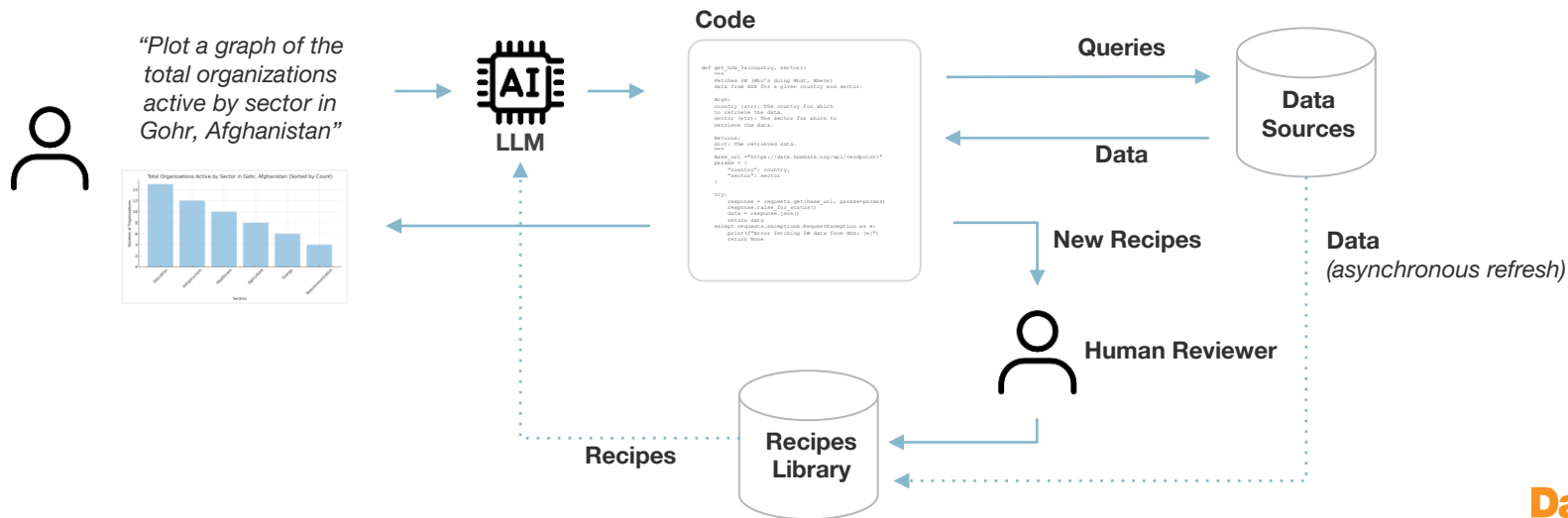

DataKind

Hi. I'm your humanitarian AI assistant.

Ensuring a human is in the loop through data recipes
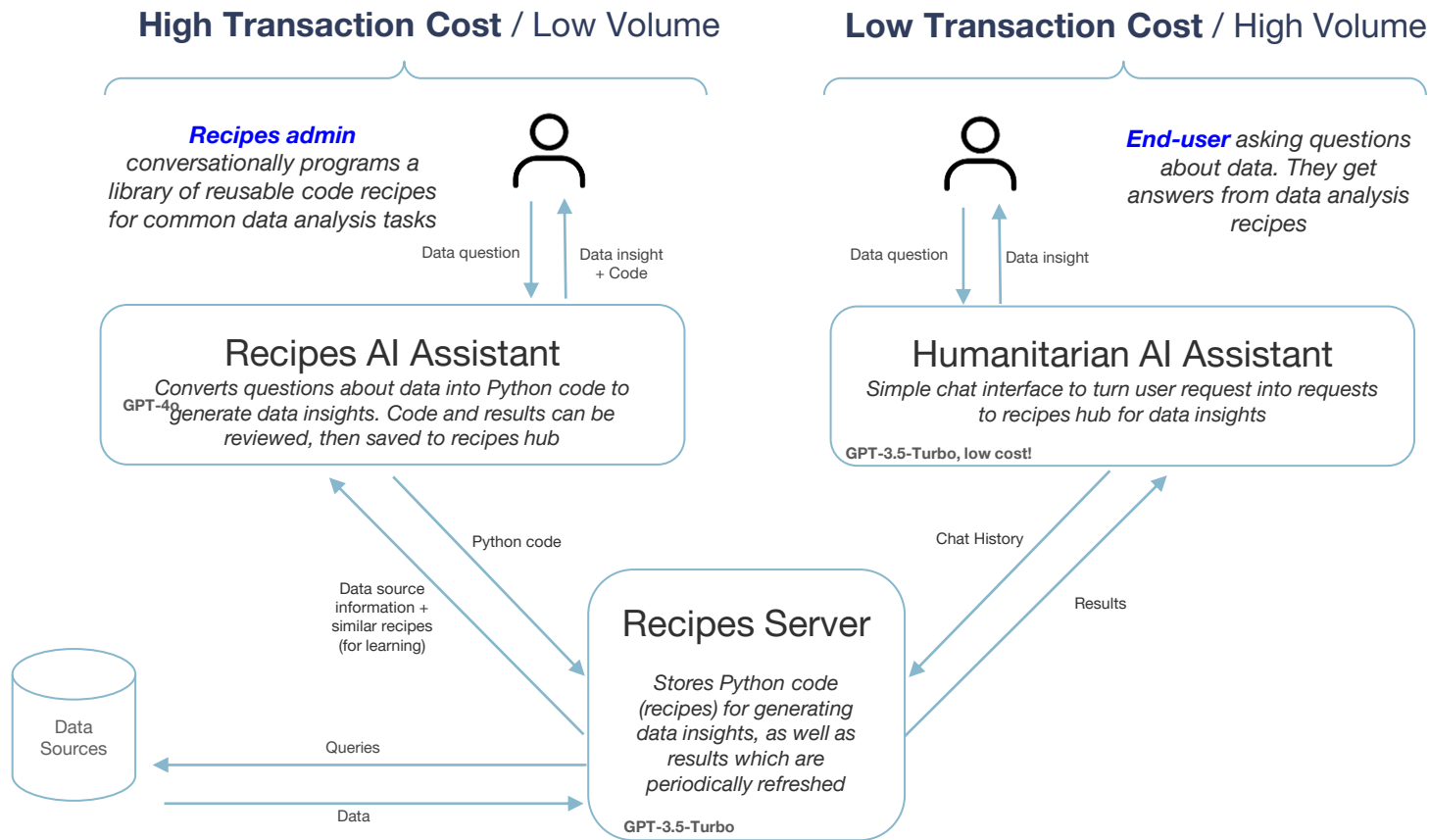
Don't ask the LLM to analyze data directly …
   ask the LLM to generate *code* recipes to analyze data …
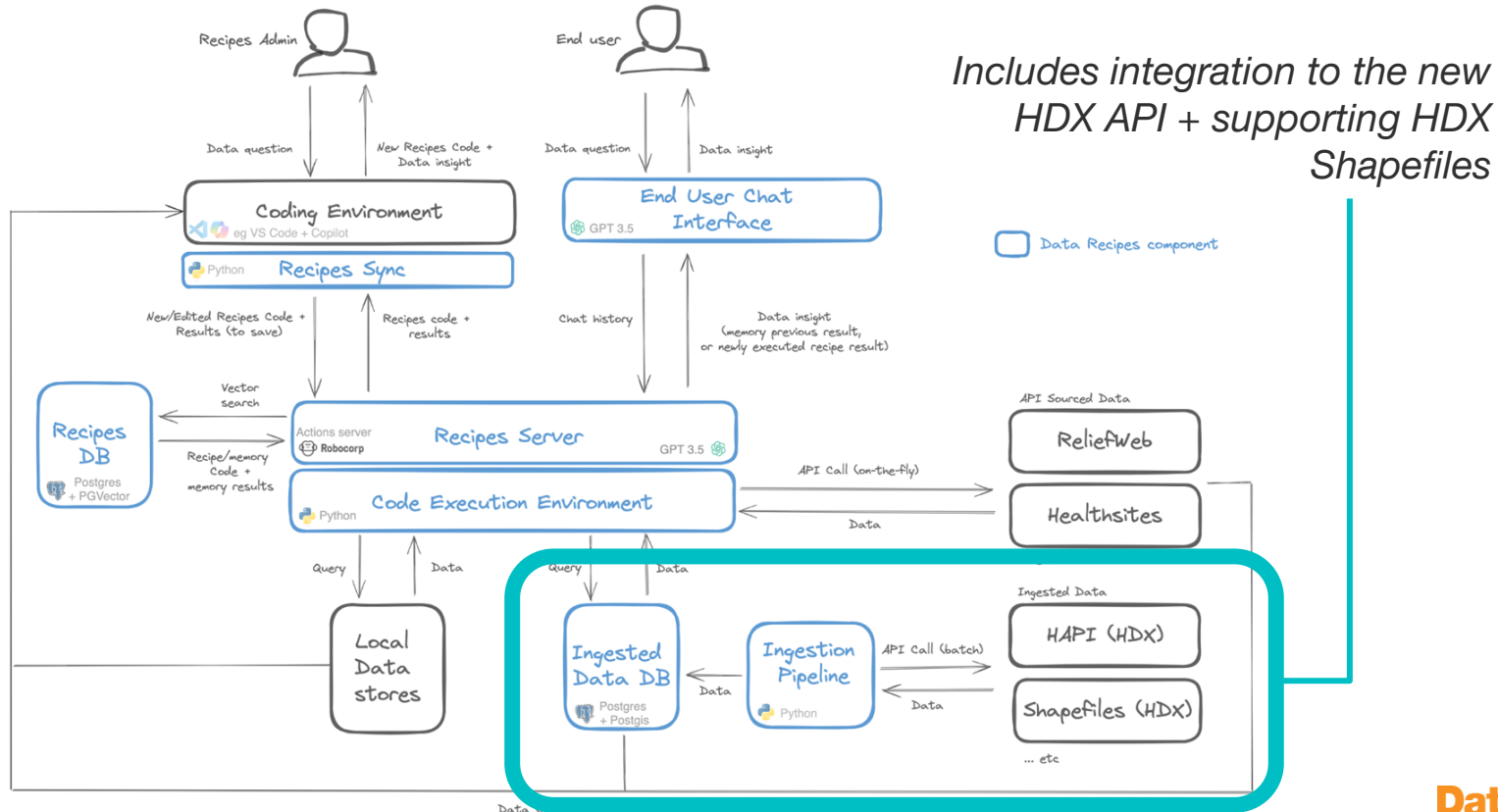   have a human review, then save this code into a **reusable library** …
         rerun recipes behind-the-scenes to keep results **up-to-date**

Deliberately managing computing costs

High Transaction Cost / Low Volume

Low Transaction Cost / High Volume

Recipes admin conversationally programs a library of reusable code recipes for common data analysis tasks

End-user asking questions about data. They get answers from data analysis recipes

Data question

Data insight + Code

Data question

Data insight

Recipes AI Assistant

Converts questions about data into Python code to generate data insights. Code and results can be reviewed, then saved to recipes hub

GPT-4o

Humanitarian AI Assistant

Simple chat interface to turn user request into requests to recipes hub for data insights

GPT-3.5-Turbo, low cost!

Python code

Chat History

Data source information + similar recipes (for learning)

Results

Recipes Server

Stores Python code (recipes) for generating data insights, as well as results which are periodically refreshed

GPT-3.5-Turbo

Data Sources

Queries

Data

DataKind

Humanitarian AI Assistant - System flows

Includes integration to the new HDX API + supporting HDX Shapefiles

Deployment - Docker

We are including a docker environment to run all components which organizations can deploy on their own infrastructure

Humanitarian AI Assistant Component	Docker container
Recipes Server	Robocorp AI Actions Server
End user chat interface	Chainlit
Recipes configuration database	Postgres SQL + PGVector
Recipes datasets database	Postgres SQL + Postgis
Ingestion pipeline	Python
Recipes management (command line interface)	Python

We're looking for feedback!

We see the immediate use cases as supporting the development of sitreps and needs assessments - both vital reports in the humanitarian space, but resource-intensive to construct.

- We think this platform is extensible - what other use cases might you see?

- What datasets / data platforms should we prioritize including in our beta release of the tool apart from HDX API?

- What are some potential risks and concerns you identify which should be mitigated/addressed?

DataKind
Join our community of practice

Publicly published report images

Join our community of practice

Humanitarian Data Insights
Generation: Join our Community

We're looking for:

- Humanitarian aid workers
- Advisors
- Researchers
- Data professionals
- Admin professionals
- Funders
- Advocates
- Others in the humanitarian sector

https://tinyurl.com/3r9avf2t

DataKind

Humanitarian Data Insights Project

June 2024

datakind.org | @DataKind

Learn more & join our community of practice

Upcoming Webinars in the HuMEL Emerging Tech in Humanitarian Settings

16th July 2pm CET

DEEP - Data Entry and Exploration Platform:

is an intelligent web-based platform offering a suite of collaborative tools tailored for qualitative and secondary data review