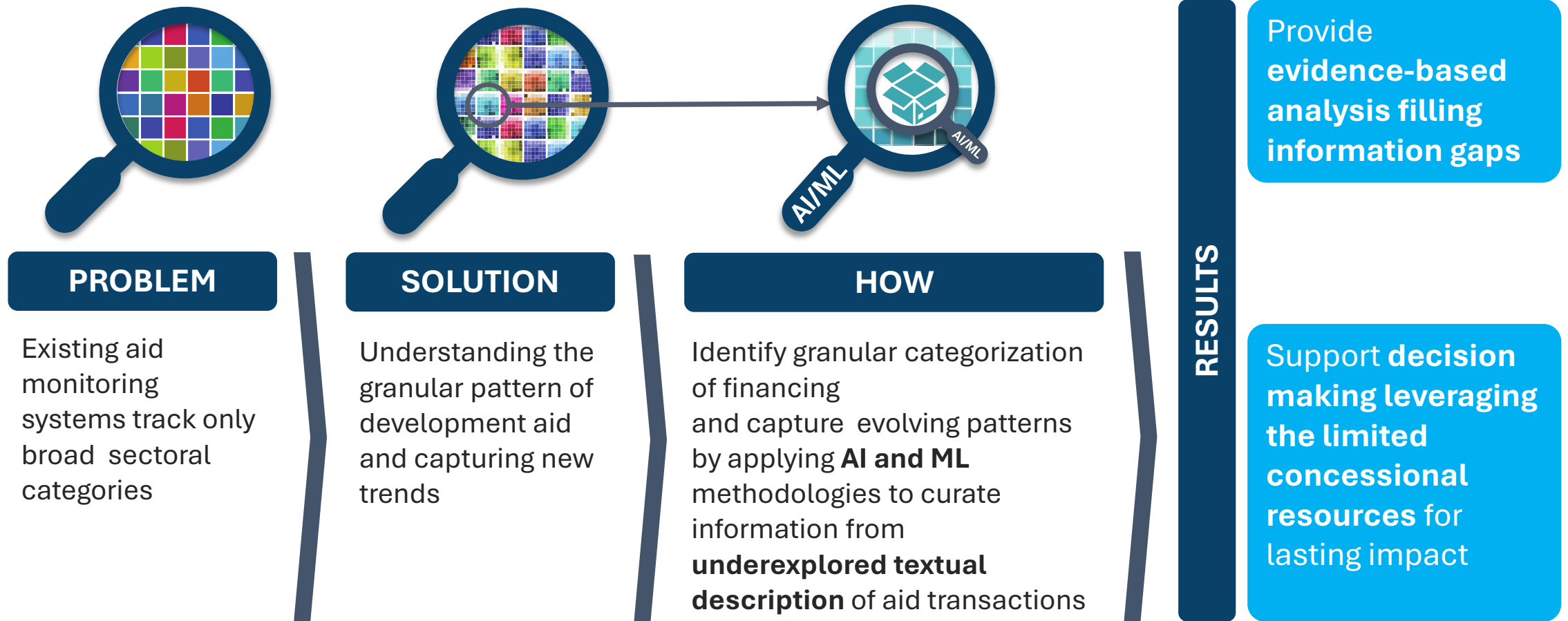

Monitoring Global Aid Flows: An AI & Machine Learning Approach

Xubei Luo

**Senior. Fellow FERDI
The World Bank**

The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the World Bank.

AI and ML for monitoring aid flows



**Emerging application
of AI/ML methods to
obtain richer insights
into trends in the
global aid
architecture**

Opportunity



**Bottom-up,
unsupervised algorithm
identification of
emerging trends in aid
activities**

*Overcome the limits of
existing monitoring
frameworks*



**Recategorize CRS aid
activities into granular
activity clusters**

*Drill down to analyze
trends at a more detailed
level.*



**Estimation of financial
support to various
cross-cutting and
thematic issues**

*Address weakness of existing
monitoring frameworks of tracking
thematic or multi-sector issues.
For instance, a health and
education project might be
categorized into the Health,
Education, or Multi-Sector sectors*

I Value

This project applies AI/ML methods to help overcome the limitations of existing monitoring frameworks of global aid flows



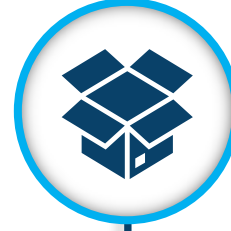
Our machine learning algorithm can “read” the rich, textual descriptions which accompany each aid activity in the CRS data



Using this text, our algorithm **re-organized** all **395,306** aid records in the 2021 CRS data into **841** highly granular activity clusters



Identify newly emerging aid activity trends



Unpack multi-sector aid

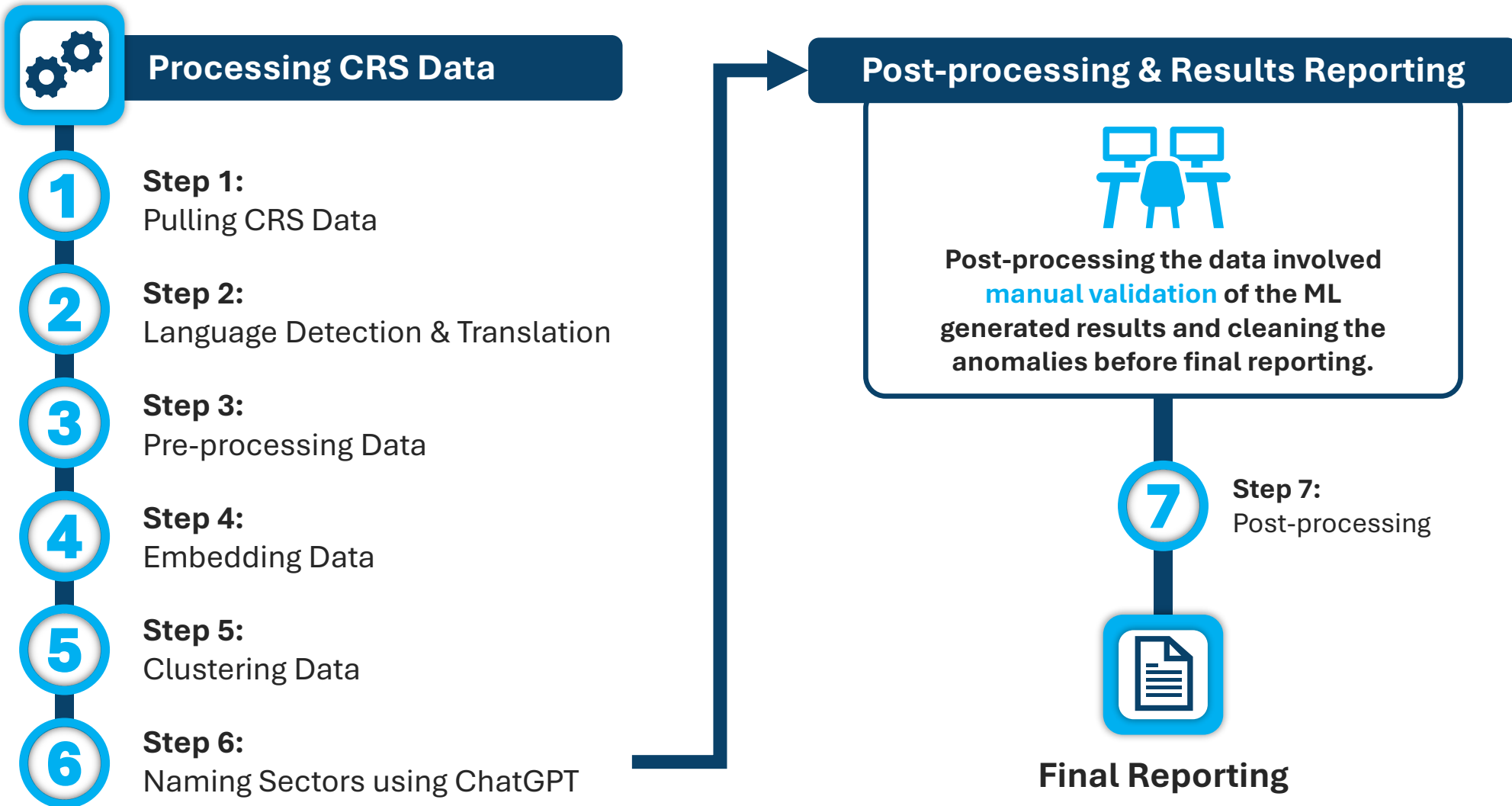


Illuminate “sector not specified” CRS group



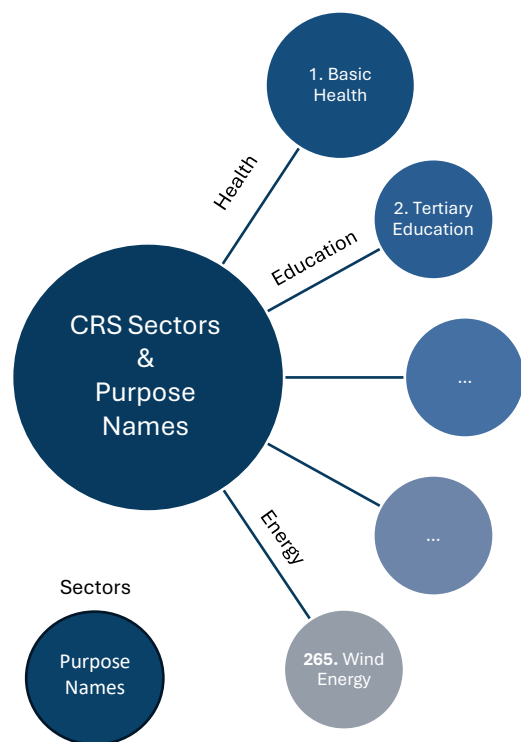
Estimate aid flows to new key thematic areas

| Processing and Results



Our machine learning model produced 841 highly granular activity clusters compared to 265 in original CRS for 2021

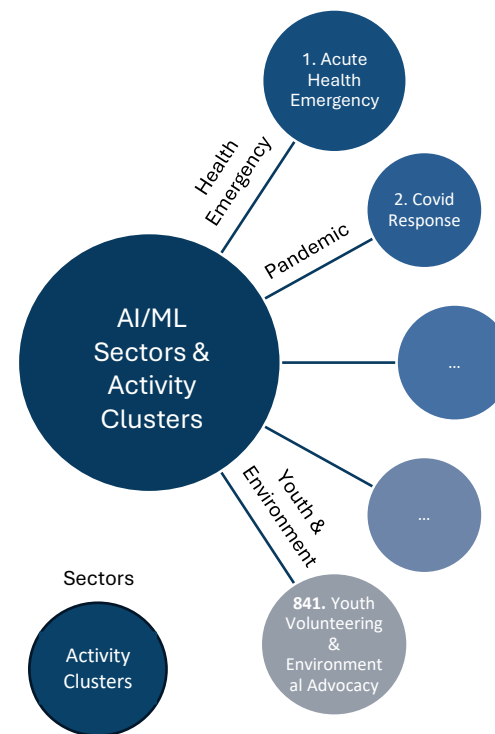
CRS Sectors/Purpose Names Before ML



The default OECD dataset has two levels of identifying a transaction- Sectors as level 1 and Purpose Name which is Level 2 under sectors.

In 2021, there were **265** purpose names

CRS Sectors/Purpose Names After ML



The ML model helped cluster the aid transactions in highly granular activity clusters, which enable us to create new and emerging sector names as needed.

In 2021, there were **841** Activity Clusters

I Main results



Identified 36 broad emerging activity trends not currently captured by existing CRS sectors.

E.g. "Food Security", "Private Sector Mobilization", "Digitalization", "Pandemics", "Employment Creation".



Unpacked multi-sectoral aid at a greater level of detail, identifying flows to specific multi-sectoral areas.

E.g. Intersections of "Women and Youth" "Agriculture and Climate", "Health and Climate", and "Education and Health".



Illuminated the previously opaque "sector not specified" CRS sector group.



Estimated the support to cross-cutting themes.

I Caveats



Due to the statistical nature of the ML algorithm, it is **impossible to perfectly reproduce the same activity clusters when re-running the algorithm.**

To mitigate this, we have re-run our algorithm on the 2021 CRS data and found **that 99% of new clusters had cosine similarity scores of over 80% with clusters from the previous round.**



As the ML-generated clusters and therefore the manually grouped sub-sectors are input data specific, **cluster and sub-sector results will change as additional data are included.**

The value of this organic, bottom-up methodology is its **ability to adaptively capture unexpected patterns in the data**



The estimates created for thematic issues such as climate, pandemics, etc. are high-level and **likely under-estimating flows to these areas.**

For future research, we may **individually tag the record-level data for these thematic issues** by directly categorizing data according to predefined thematic issues

Monitoring Global Aid Flows

A Novel Approach Using Large Language Models

*Xubei Luo
Arvind Balaji Rajasekaran
Andrew Conner Scruggs*



WORLD BANK GROUP

Development Finance Vice Presidency
November 2025

Thank You!