ASIAN EVALUATION WEEK 2024

Innovations for Influential Evaluation
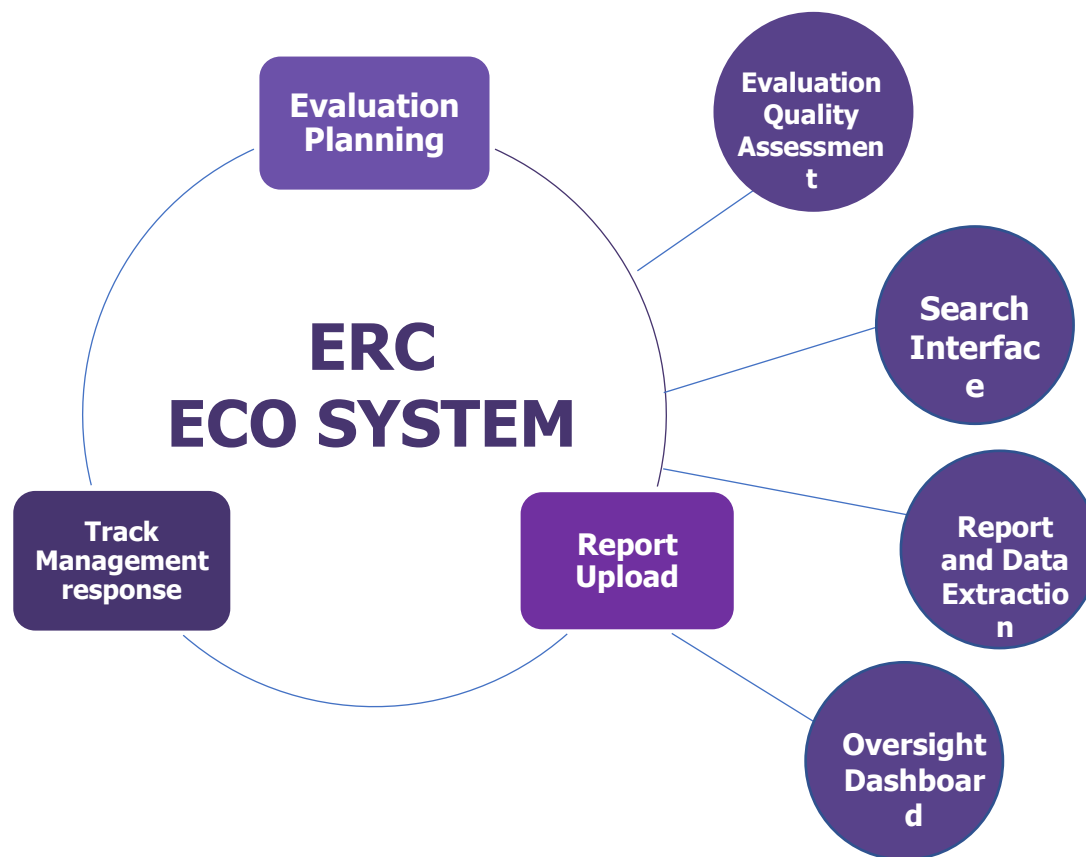2–5 September | Shanghai, People's Republic of China

UNDP — IEO Independent Evaluation Office
United Nations Development Programme

Artificial Intelligence for Development Analytics - AIDA

EVALUATION RESOURCE CENTRE

erc.undp.org

- **UNDP Country Offices**
- **Headquarter Units**
- **Funds and Programmes**

Evaluation Planning

Evaluation Quality Assessment

ERC ECO SYSTEM

Search Interface

Track Management response

Report Upload

Report and Data Extraction

Oversight Dashboard

- **Publicly & globally accessible system**
- **Connected to Corporate systems for data exchange**

Open database with more than **6,500** evaluations reports since 1997

300 to 400 new evaluation reports added annually

Key functionalities: Evaluation Planning, Report Upload and Track Management Response

AIDA's simplified pipeline

Data acquisition
- Ingestion of more than 6,500 evaluation reports from ERC

Text extraction
- Automatic extraction of key sections

Comprehension
- AI-driven labelling and categorization of extracted data
- Human-in-the-loop feedback system to retrain the machine learning model

Search engine
- ML-driven search engine
- Human-in-the-loop feedback system
- Output filtering
- Easy export to XLS

What can AIDA do for me?

01
Support project and programme formulation

02
Provide insights for synthesis

03
Provide inputs for gap map analysis

4 types of search

Results – evaluation evidence

Insight generation

https://aida.undp.org

Taxonomy

- **2 models: Thematic & non-thematic**
- **17 parent categories in total**
- **216 child labels in total**

- **Taxonomy can be updated with new labels if necessary.**
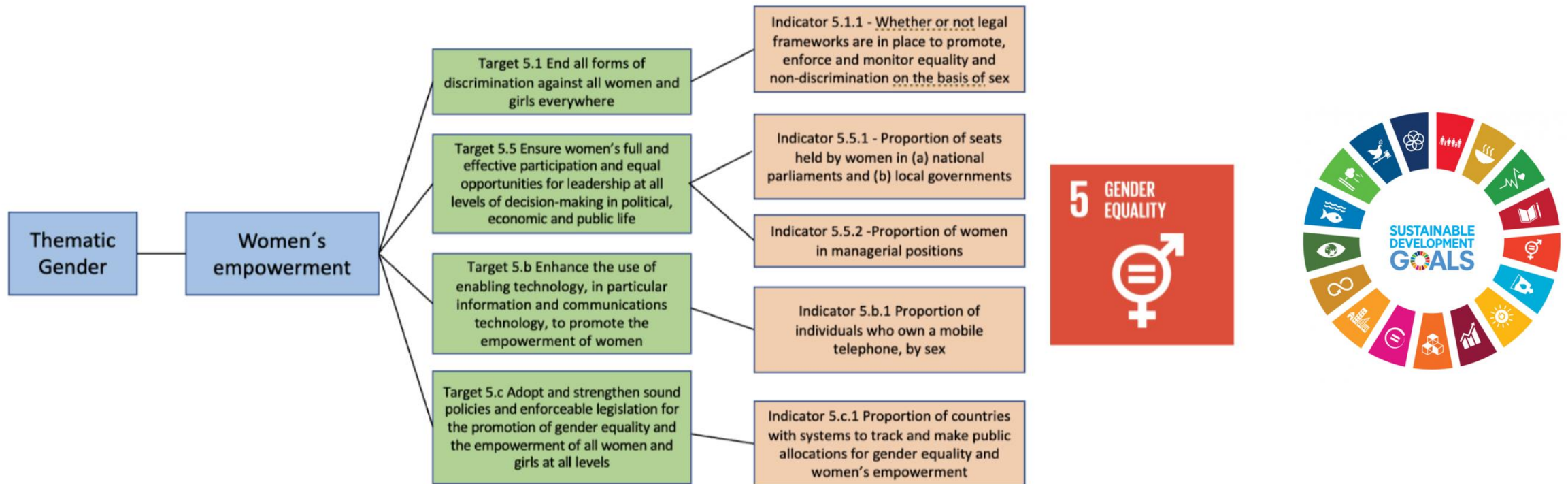- **Creation of training data set per label is required.**

Row Labels	Distinct Count of Parent_Category	Distinct Count of Child_Category
Non-thematic	8	101
Thematic	9	115
Grand Total	**17**	**216**

Model Grouping	p_i	Parent_Category	c_id	Child_Category
Thematic	12	Agriculture	130	Agriculture
Thematic	12	Agriculture	136	Agriculture Land Resources
Thematic	12	Agriculture	135	Agriculture Policy
Thematic	12	Agriculture	132	Agriculture Water Resources
Thematic	12	Agriculture	188	Aquaculture
Thematic	12	Agriculture	133	Crop Production
Thematic	12	Agriculture	131	Fishery
Thematic	12	Agriculture	25	Food Security
Thematic	12	Agriculture	58	Forestry
Thematic	12	Agriculture	146	Livestock
Thematic	12	Agriculture	137	Rural Development
Thematic	3	Climate and Disaster Resilience	78	Climate Change Adaptation
Thematic	3	Climate and Disaster Resilience	80	Climate Change Governance
Thematic	3	Climate and Disaster Resilience	255	Climate Change Mitigation
Thematic	3	Climate and Disaster Resilience	79	Climate Finance
Thematic	3	Climate and Disaster Resilience	198	Crisis Mitigation
Thematic	3	Climate and Disaster Resilience	155	Crisis Response
Thematic	3	Climate and Disaster Resilience	81	Disaster Recovery
Thematic	3	Climate and Disaster Resilience	83	Disaster Risk Assessment
Thematic	3	Climate and Disaster Resilience	151	Disaster Risk Management
Thematic	3	Climate and Disaster Resilience	82	Disaster Risk Reduction
Thematic	3	Climate and Disaster Resilience	261	Drought
Thematic	3	Climate and Disaster Resilience	263	Flooding
Thematic	3	Climate and Disaster Resilience	154	Greenhouse Gas Emission
Thematic	3	Climate and Disaster Resilience	234	Humanitarian Crisis
Thematic	3	Climate and Disaster Resilience	54	Mine Action
Thematic	3	Climate and Disaster Resilience	152	Natural Disaster
Thematic	3	Climate and Disaster Resilience	214	Recovery
Thematic	3	Climate and Disaster Resilience	92	Resilience Building

SDG linked Taxonomy

To enable extraction of evidence at SDG level, we have developed a taxonomy framework for AIDA that contains **216 labels** organized in **9 thematic and 8 non-thematic parent categories**.
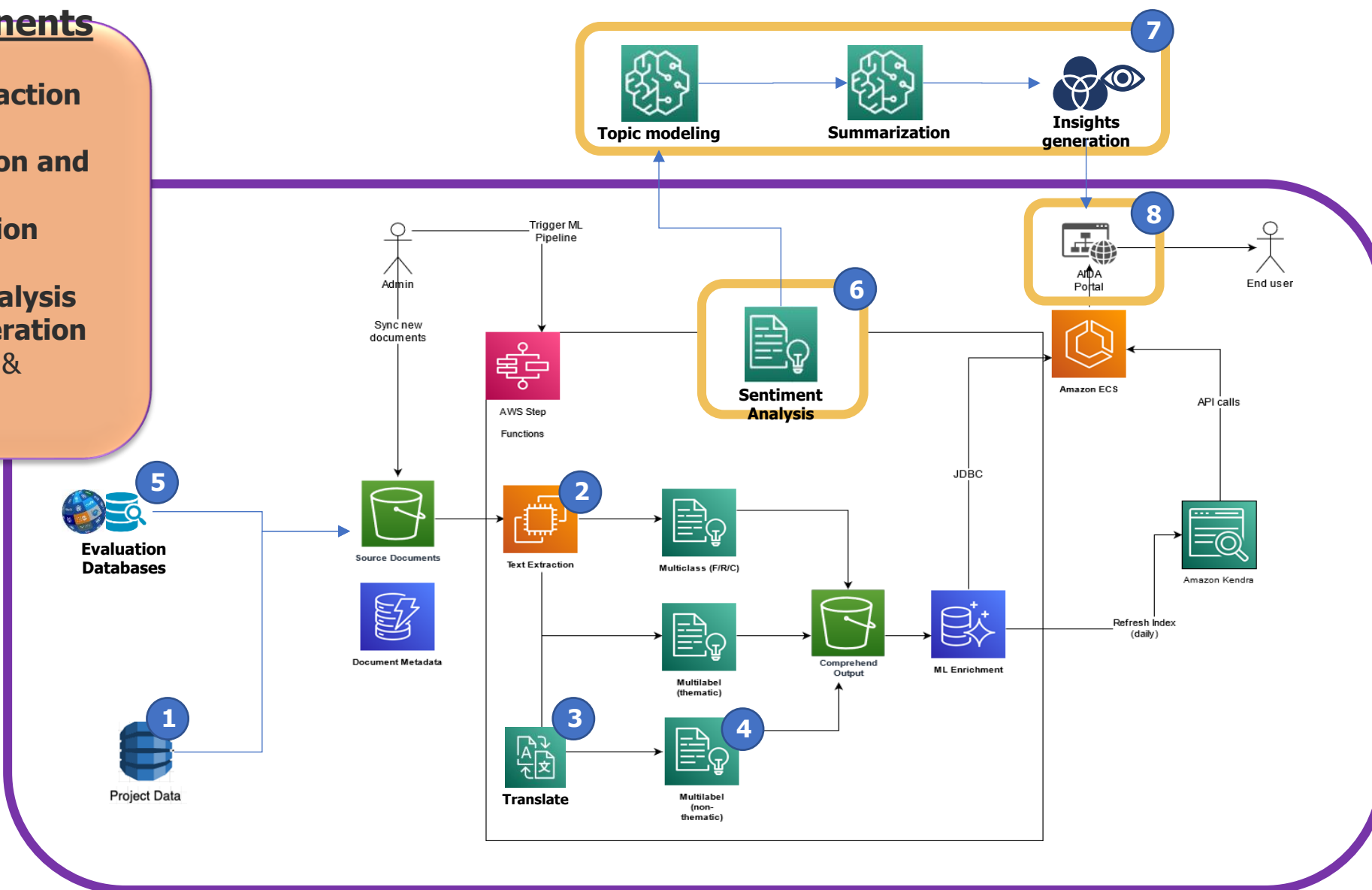
Each labels is **linked to SDG targets** and indicators.

AIDA 2.1

Main components

1. **Data pipeline**
2. **Enhance extraction**
3. **Translation**
4. **Comprehension and model training**
5. **Multi-institution design**
6. **Sentiment analysis**
7. **Insights generation**
   - Topic modelling & summarization
8. **Web portal**

Sentiment Analysis

Sentence level

Version name ▽	Creation date ▼	Accuracy ▽	Test data source ▽	F1 score ▽	Endpoints ▽	Status ▽
○ aida-zeppelin-balanced	March 22, 2023 at 14:43 (UTC)	0.99+	Customer provided	0.98	0	⊘ Trained
○ aida-zeppelin-unbalanced	March 22, 2023 at 14:42 (UTC)	0.98	Customer provided	0.97	0	⊘ Trained
○ AIDA-only	March 22, 2023 at 14:23 (UTC)	0.63	Customer provided	0.62	0	⊘ Trained

✔ **Solution:**

- **Sentence level performs best**
- This granular approach enables more precise predictions and the opportunity for specific feedback.

At the time of the Mid-Term Evaluation, a number of knowledge products produced by PSLWGEP were not finalised and could not be assessed. As part of this final evaluation, the National Policy on Gender Equality (Draft December 2019), Neary Ratanak V, Press release on launch of Media Code of Conduct, We Can Do It Video, Monitoring and Evaluation Framework report and Capacity Development Framework were reviewed to determine their technical quality. **All products reviewed as part of this evaluation were of a high standard.** This evaluation was not able to assess the quality of the Media Code of Conduct, feminist leadership materials, training and coaching materials and GMAPs as these are only available in Khmer language.

Paragraph level

Model	Accuracy	Precision	Recall	F1 Score
Old AIDA model	74%	0.74	0.74	0.73
AIDA Dataset Only	68%	0.64	0.68	0.65
AIDA + Zeppelin Data	75%	0.73	0.75	0.70
AIDA + Zeppelin Balanced	77%	0.77	0.77	0.75

Ground truth sentiment: Mixed
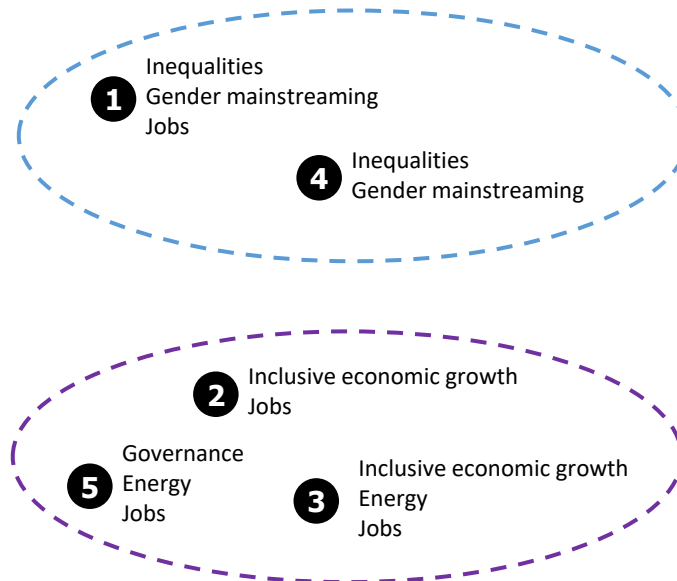**Model sentiment: Mixed**

Topic Clustering

- Leveraged machine learning algorithms to **group paragraphs into topics**

- Each topic cluster consists of thematic labels (i.e. "gender", "communication", "clean energy", etc.) that are representative of the cluster

- Steps within the process:

Data Clustering
A clustering algorithm was used to cluster paragraphs by thematic labels.

Data Summarization
For each cluster, the most represented thematic labels were assigned as representative themes of the cluster.

1 Inequalities
Gender mainstreaming
Jobs

2 Inclusive economic growth
Jobs

3 Inclusive economic growth
Energy
Jobs

4 Inequalities
Gender mainstreaming

5 Governance
Energy
Jobs

1 Inequalities
Gender mainstreaming
Jobs

4 Inequalities
Gender mainstreaming

2 Inclusive economic growth
Jobs

5 Governance
Energy
Jobs

3 Inclusive economic growth
Energy
Jobs

Topic Cluster 1
- Representative Theme 1: Inequalities
- Representative Theme 2: Gender mainstreaming

1 **4**

Topic Cluster 2
- Representative Theme 1: Jobs
- Representative Theme 2: Energy
- Representative Theme 3: Inclusive economic growth

5 **2**
3

Summarization

Objectives:
- **Breaking down lengthy text into a digestible paragraph.**
- **Extracting vital information without losing the coherence.**

Search Query: **Climate Change**
Found Topic: **Coastal Resilience**

Paragraph 1:
Project Description To seek to improve coastal resilience in Tunisia, in 2014 the UNDP embarked on the implementation of the current project entitled "Addressing climate change vulnerabilities and risks in vulnerable coastal areas of Tunisia" a GEF-financed project (US$ 5,500,000) over the period 2015-2019. The project proposes a risk-based approach to Climate Change Adaptation (CCA) by enabling flexible adaptation pathways, which will build resilience to climate change and provide maximum co-benefits. As tourism is a dominate source of revenue for the region, a set of economic instruments are proposed to be devised to signal the existing risks and drive future hotel and private residence development, including investments, away from vulnerable areas…

Paragraph 2:
From a technical monitoring perspective (not specifically project level) there is a need to start gathering real time data from the wave buoy network and re-deploy the wave buoy in Djerba which was vandalized and has not gathered any new data. This is key to show to GEF as an important contributor to the intended outcomes of the project as it also represents a key part of the Coastal Observatory work and links importantly to the argument for new data to be collated. This is even more important based on recent storm events during the winter of 2018 which had a significant impact on beach erosion rates in the pilot study areas (mainly wind and wave induced erosion). Local populations now wish to urgently see coastal interventions taking place (as opposed to more studies). Any new monitored data to support local coastal decision making now needs to be better used and conveyed to local groups to demonstrate the benefits of the SCCF project and what it has helped to provide.

Paragraph 3:
The vulnerability of SIDS to climate change is accel- erating rapidly and represents an existential threat to some. Reflecting this, the UNFCCC refers to "the specific needs and concerns of developing country Parties arising from the adverse effects of climate change especially on: (a) Small Island countries; (b) Countries with low-lying coastal areas has carried through to the Paris Agreement, which singles out island states (with LDCs) as a focus for capacity development support and scaled-up financial resources, among other things.
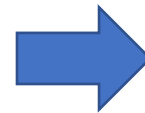
Generated Summary:

The report focuses on four specific areas: coastal flooding, livelihood/ecological resilience, water and wetland resilience, urban resilience, climate information technology and early warnings systems. It further identifies three other areas where ODA may be considered in terms of adaptation if it meets the needs of both developing countries and international actors. MedCOAST 2015 reports on the impact of rising sea levels on agricultural production and consumption; drought and severe weather

UNEG - Ethical Principles for use of AI in evaluation

The Data/AI group drafted a working paper which applies several existing UN AI ethics documents. The working paper is specifically tailored to AI applications in evaluation processes, data, methods, and analysis.   The principles include:

- Do no harm

- Transparency and Accountability

- Fairness and Mitigation

- Participation and Inclusiveness

- Privacy and Data Protection

- Accuracy and Reliability

- Human-Centered Approach

- Upholding Human Rights

Thank you

https://www.undp.org/evaluation

@UNDP_Evaluation

UNDP Independent Evaluation Office