

Meet
MALENA
IFC's Machine Learning ESG Analyst



MINISTRY OF FOREIGN AFFAIRS
OF DENMARK
Danida

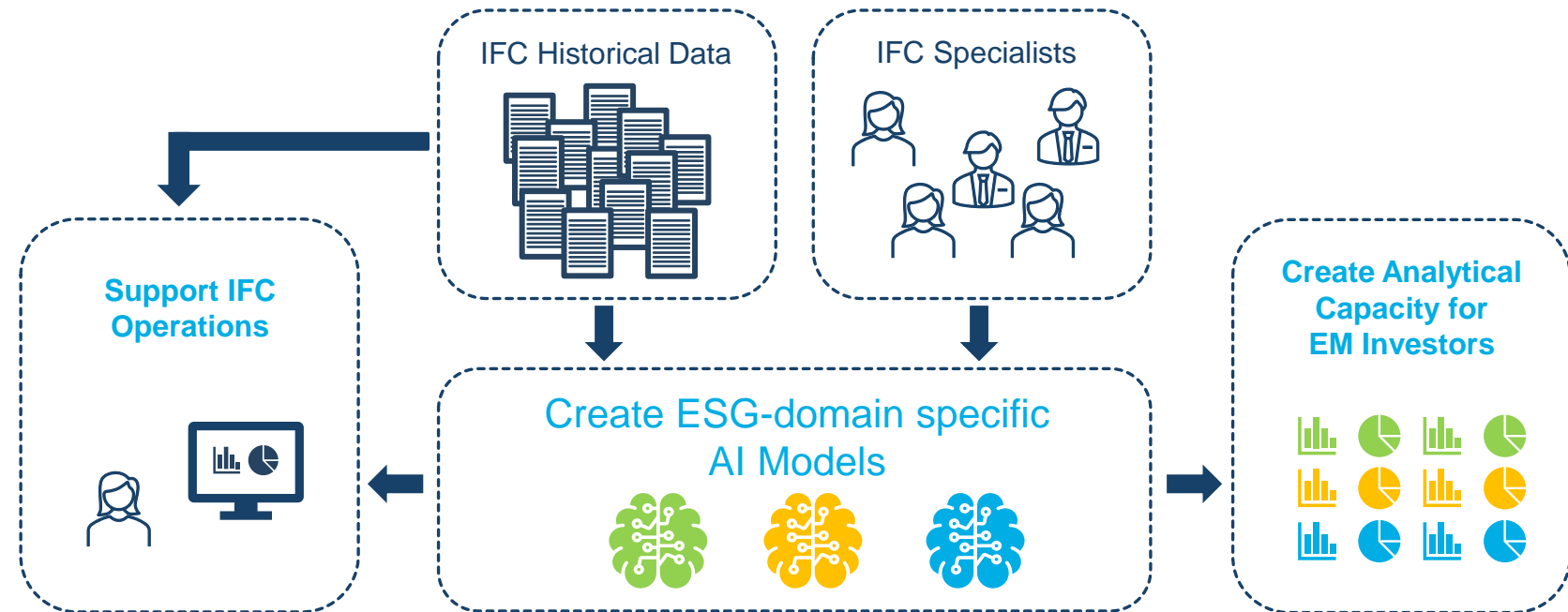
IN PARTNERSHIP WITH



Leveraging Artificial Intelligence for Development

Goal: Develop innovative ESG solutions leveraging disruptive technologies to drive sustainable investment in emerging markets

- Use **institutional knowledge** and **historical data** to create ESG-domain artificial intelligence models
- Benefit Operations: **Enhance** ESG due diligence efficiency
- Create **analytical capacity at scale** for emerging market investors



Driving Investments in Emerging Markets

Institutional Investors are instrumental in addressing the \$2.5 trillion investment shortfall needed to address the UN SDGs



ESG Integration



- **lower risk,**
less portfolio **volatility,**
and **higher returns**

Strong Demand for EMs, but:



- **Lack of uniform reporting standards**
- **Poor EM ESG disclosures**
- Lack of **EM focused** ESG indicators
- **Diverging ESG rankings**

Market Gap

- Unstructured ESG data is valuable but underused

Natural Language Processing to address ESG data and capacity needs

Recent developments in cloud computing and NLP techniques have led to innovations in the analysis of unstructured text data on a massive scale

Common NLP techniques

Named Entity Recognition

Identifies entities such as locations, companies, and organization names in unstructured text

Topic Modelling

Extracts key concepts from text to summarize and map information by topics

Sentiment Analysis

Classifies content by positive, negative, and neutral sentiment to detect risks and context

What does MALENA do: NLP to unlock ESG text for emerging markets



MALENA can read documents to identify risks and find insights

MALENA will predict a negative sentiment for risk terms that occur in an ESG risk or performance gap context

The screenshot shows the MALENA interface with a search bar and filters. The sentiment bar at the top is red, indicating negative sentiment. The text below contains several words highlighted in red boxes, indicating negative sentiment: "community", "noise", "fugitive emissions", "emissions", "wastewater", and "hazards".

Negative Sentiment Positive Sentiment Neutral Sentiment

Search Refine by PS Refine by Topics Refine by Risk Terms

The main concerns amongst other issues raised included the effect of the project on the proposed location of the site which the appellants claimed was a residential estate, a rich marine environment and citing concerns that the terminal would endanger the lives of the **community**.

On 13 January 2021 NEMA extended the EIA License validity period by an additional 24 months. The sponsors will ensure that the ESIA and associated licenses are renewed, all conditions of the licenses fulfilled, the ESIA addendum for the STT and truck park are approved prior to start of construction and maintained thereafter.

The ESIA identified the following key issues; (i) construction related impacts such as dust, **noise** and **fugitive emissions** emissions **emissions**; (ii) **wastewater** management; (iii) operational **hazards**; (iv) potential increase in traffic during construction and operation from trucks conveying LPG to rail heads and customers.

Goto: Showing 3 of 13

There are no “negative” words. MALENA makes predictions based on context, for example: “no **fatalities**” = positive sentiment

The screenshot shows the MALENA interface with a search bar and filters. The sentiment bar at the top is green, indicating positive sentiment. The text below contains several words highlighted in green boxes, indicating positive sentiment: "penalties", "fines", "non-compliance", "compliance", and "fatalities".

Negative Sentiment Positive Sentiment Neutral Sentiment

Search Refine by PS Refine by Topics Refine by Risk Terms

Based on information provided during appraisal, Robust did not incur over the last three years any material regulatory **penalties**, **fines** or sanctions for contraventions or **non-compliance** compliance **compliance** with statutory obligations, as well as not reported any **fatalities** on sites at any of its operations. Supply Chain RiskRiskRisk Assessment & Management.

The two main commodities (sesame seeds and RCN) are purchased either at farm gate from smallholder (of more than 1 ha) growers or through a network of Licensed Buying Agents (LBAs) or aggregators who cover the main production areas.

Micro-sized smallholder farmers (those with <1ha of land) are often encouraged to aggregate with other farmers to sell their stock as a cooperative.

However, the company is yet to establish the database and supplier mapping of its sourcing for its two main commodities.

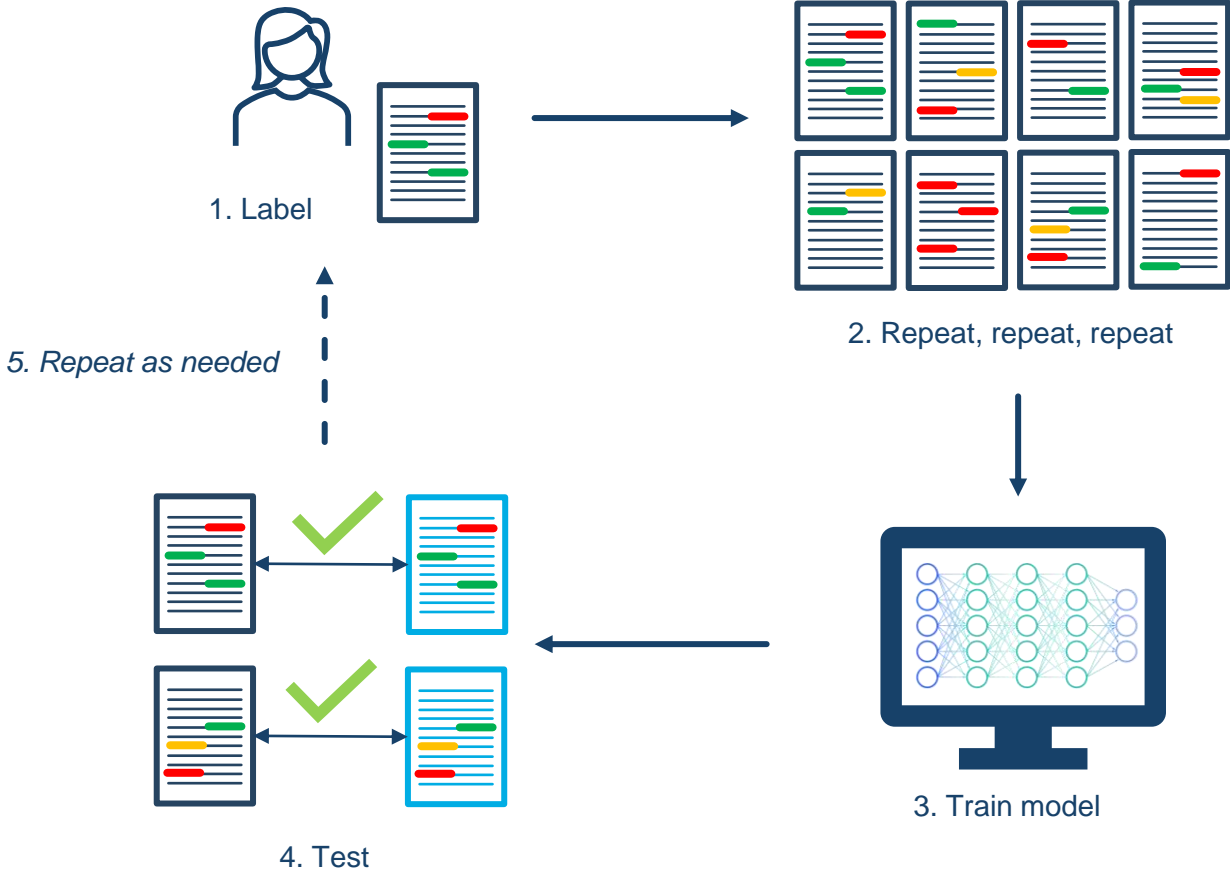
The company is committed to have 100% traceability by 2025. Robust has HP policies and **Supplier**

Goto: Showing 4 of 13

How to Train a Model

How to train a ... (supervised) model

- Clearly define the **purpose** and **goals** for the model
- **High quality** training data in **large quantities** is a key success factor for supervised machine learning models
- Follow **consistent** rules
- Establish **quality controls** for labeling



Training MALENA and Model Performance

Training MALENA

MALENA is based on a pretrained, open-source model from Meta AI: **RoBERTa** (*Robustly optimized Bidirectional Encoder Representations from Transformers approach*)

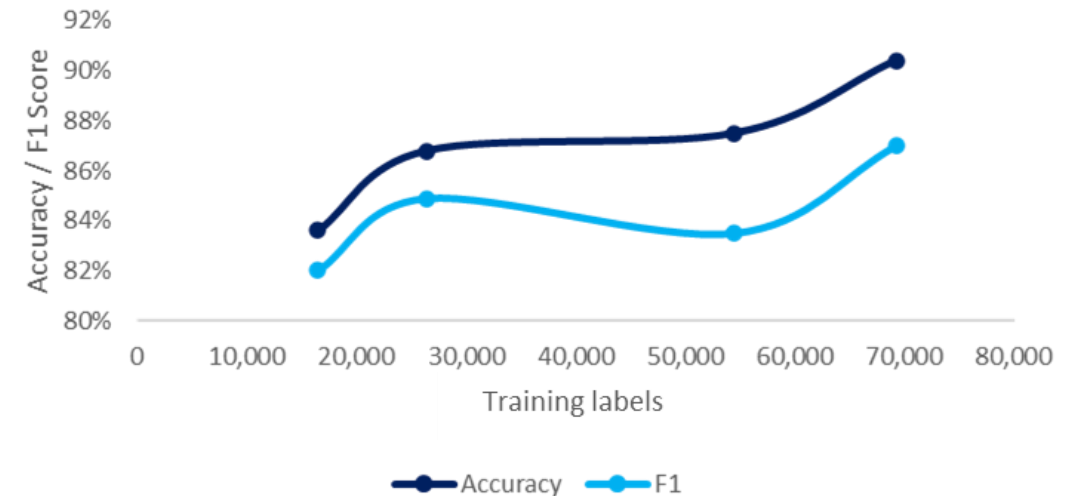
- **Manually labeled training data** is used to teach MALENA through transfer learning
- **Quality controls** for labeling consistency
- **Active Learning** provides feedback for model refinement
- Data **quality checks** to address **bias** in training data
- Model performance: **90% Accuracy / 87% F1 Score**

MALENA

Training Data

- **125,000+ labels** created by ESG analysts
- **1,200+ ESG risk terms** used for labeling
- Inventory includes E&S, CG, climate, and gender

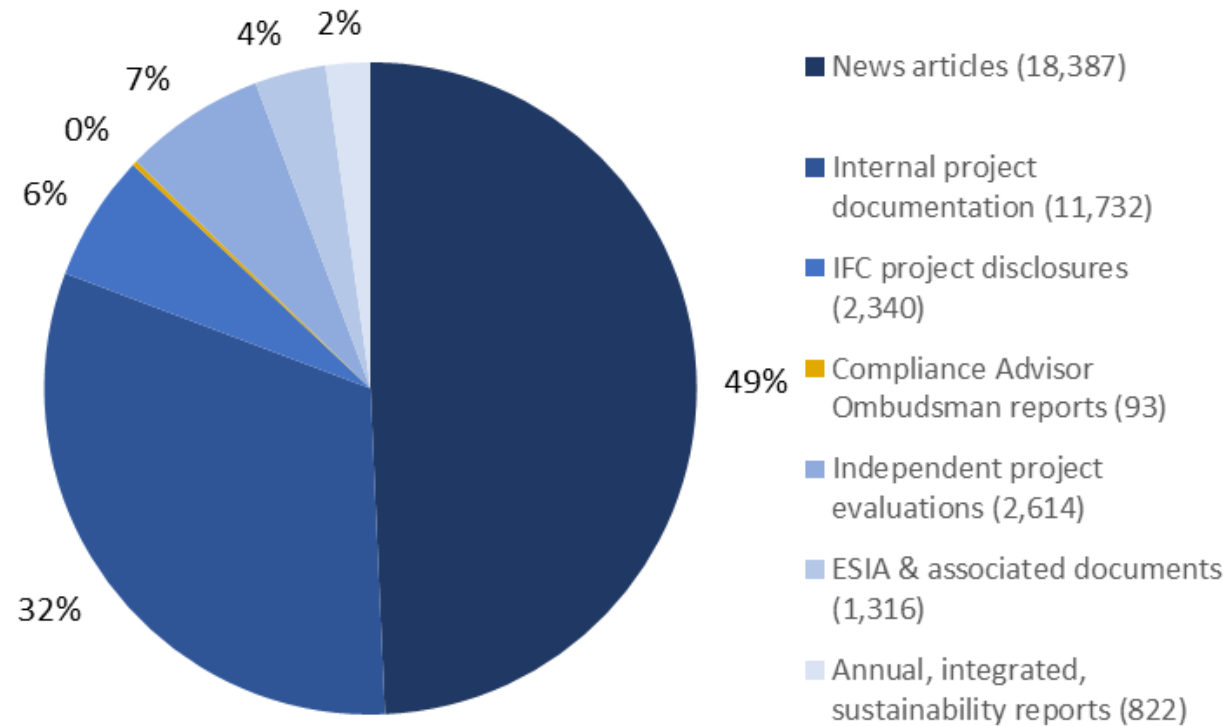
MALENA Model Performance



Documents Analyzed

Documents analyzed & predictions made

To date, MALENA has been tested on more than 37,000 documents, resulting in over **14 million ESG signals**



Ethical Artificial Intelligence

Importance of ethics in AI recognized by industry and governing bodies
UN, EU, OECD, NIST



IFC developed draft Technology Code of Conduct (TCoC) for the sustainable development of AI



MALENA assessed using draft TCoC	Data Bias	Available data for the model training & inference is not representative
	Model Drift	Degradation of model performance due to changes in data and relationships between input and output variables
	XAI	Solutions allowing humans to interpret why an AI model arrived at the results it produced



MALENA Data and Model Governance Framework

Q&A

Case Study #1: Emerging Market Issuer ESG Scoring

Using NLP to unlock ESG data for emerging market Financial Institution issuers

Collaboration with AMUNDI Asset Management to compare MALENA results with AMUNDI's ESG scores

Sample:



804
EM FI issuers of
hard currency debt



441
Data identified for 441
FIs (55%)



402 Corporate reports



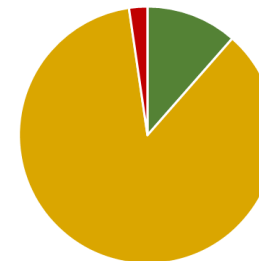
415 Bond prospectuses



428 ESG-related news reports

1,260
Documents

1.6M
ESG Signals



Neutral 1.3M
Positive 189K
Negative 38K

■ POS ■ NEU ■ NEG

Case Study #1: Results

441 Company ESG Profiles



Correlation with asset manager ESG scores

MALENA validated ESG scores for 205 FIs

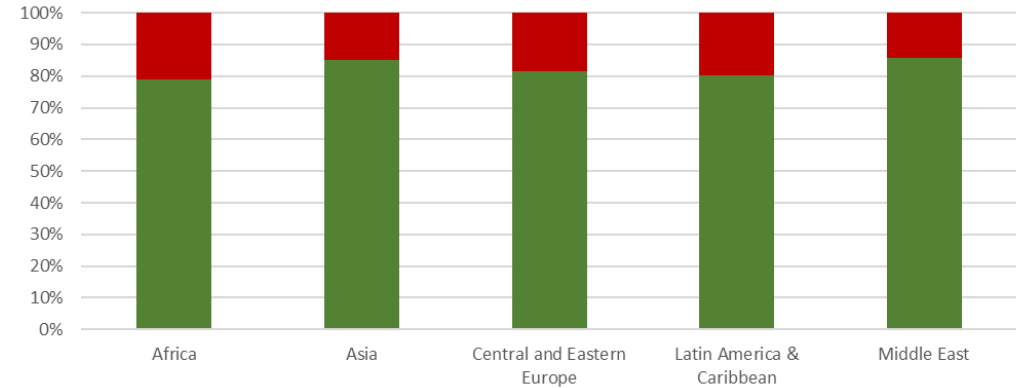


Additional ESG profiles

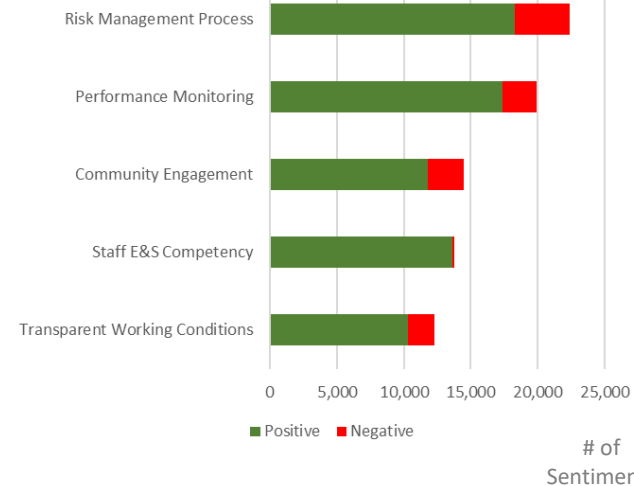
MALENA provided an additional 236 ESG scores doubling asset manager coverage

EM FI Industry Profile

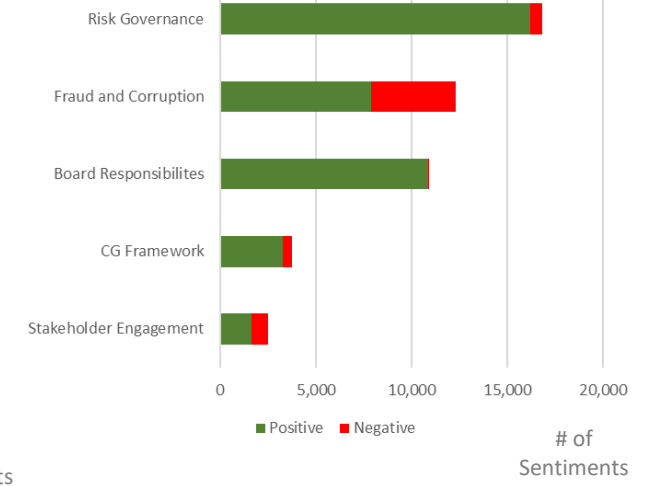
Proportion of Negative Sentiment by Region



Top 5 E&S Topics



Top 5 CG Topics



Case Study #2: ESIA as ESG Performance Predictor*

NLP to predict ESG performance of projects based on early-stage due diligence

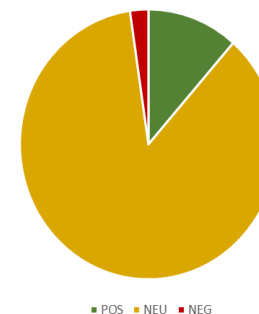
Sample:

 530 IFC projects

1,316
ESIAs & related
Documents



1.4M
ESG Signals



Neutral 1.1M
Positive 210K
Negative 112K

Results:



Correlation between
ESIA Sentiment Score
and project E&S
performance



ESIA
documents
the best
predictor in
sample



Predictions most
accurate for riskier
projects

MALENA for Emerging Market Investors

MALENA Value Proposition

- **Enable ESG integrated investing** in EMs
- Analytical capacity to **rapidly screen ESG and impact data**, conduct **ESG risk assessment and management**
- **Time and cost savings**
- Scalable **Model as a Service** solution
- **Secure service** to analyze confidential documents
- **Beta testing** underway



Investor Use Case	
Due Diligence Analyze investees: <ul style="list-style-type: none">- Investment proposals- Bond prospectus'- Corporate reports- Impact Assessments- Regulatory Documents	Portfolio Management Analyze performance <ul style="list-style-type: none">- Reporting and compliance documents- Corporate reports- Green Bond Impact Reports- TCFD/Paris Alignment Disclosures

MALENA for Climate Finance
180+ climate risk terms covering **physical climate risk; factors impacting climate change; impact of climate change to communities and livelihoods**

Looking ahead

- **Expanded MALENA taxonomy** - additional **SDG, climate, gender, and biodiversity impact** terms
- Continued improvements to model performance through **Active Learning**
- Data and Model Governance Framework to manage data **bias**, model **drift**, and **explainability** features
- Training MALENA NLP model to understand **additional languages**
- Publications in pipeline

HOW CAN YOU ACCESS MALENA?

- MALENA is currently in beta testing
- Investors can securely access the MALENA sentiment analysis model via API
- Receive sentiment predictions for ESG risk terms
- Contact us if you are an investor in emerging markets interested in beta testing MALENA

Thank you!



Atiyah Curmally

[linkedin.com/in/atiyah-curmally/](https://www.linkedin.com/in/atiyah-curmally/)



Florian Skene

[linkedin.com/in/florian-skene/](https://www.linkedin.com/in/florian-skene/)

For more information, visit: <https://www.ifc.org/sustainability/malena>