Meet MALENA
IFC’s Machine Learning ESG Analyst
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

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.
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
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

<table>
<thead>
<tr>
<th>Technique</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Named Entity Recognition</td>
<td>Identifies entities such as locations, companies, and organization names in unstructured text</td>
</tr>
<tr>
<td>Topic Modelling</td>
<td>Extracts key concepts from text to summarize and map information by topics</td>
</tr>
<tr>
<td>Sentiment Analysis</td>
<td>Classifies content by positive, negative, and neutral sentiment to detect risks and context</td>
</tr>
</tbody>
</table>
What does MALENA do: NLP to unlock ESG text for emerging markets

AI assistant to rapidly identify over 1,200 ESG risk terms and predict sentiments based on context.
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

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

1. Label
2. Repeat, repeat, repeat
3. Train model
4. Test
5. Repeat as needed

MALENA
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.

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

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

- News articles (18,387)
- Internal project documentation (11,732)
- IFC project disclosures (2,340)
- Compliance Advisor Ombudsman reports (93)
- Independent project evaluations (2,614)
- ESIA & associated documents (1,316)
- Annual, integrated, sustainability reports (822)
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

<table>
<thead>
<tr>
<th>Data Bias</th>
<th>Available data for the model training &amp; inference is not representative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model Drift</td>
<td>Degradation of model performance due to changes in data and relationships between input and output variables</td>
</tr>
<tr>
<td>XAI</td>
<td>Solutions allowing humans to interpret why an AI model arrived at the results it produced</td>
</tr>
</tbody>
</table>

MALENA Data and Model Governance Framework
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

Pie chart:
- Neutral: 1.3M
- Positive: 189K
- Negative: 38K
Case Study #1: Results

441 Company ESG Profiles

- MALENA provided an additional 236 ESG scores doubling asset manager coverage

EM FI Industry Profile

- Proportion of Negative Sentiment by Region:
  - Africa: 20%
  - Asia: 15%
  - Central and Eastern Europe: 10%
  - Latin America & Caribbean: 25%
  - Middle East: 20%

Top 5 E&S Topics

- Risk Management Process
- Performance Monitoring
- Community Engagement
- Staff E&S Competency
- Transparent Working Conditions

Top 5 CG Topics

- Risk Governance
- Fraud and Corruption
- Board Responsibilities
- COI Framework
- Stakeholder Engagement

MALENA validated ESG scores for 205 FIs

Correlation with asset manager ESG scores

Additional ESG profiles
Case Study #2: ESIAs 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

Results:

- Correlation between ESIA Sentiment Score and project E&S performance
- ESIA documents the best predictor in sample
- Predictions most accurate for riskier projects

Neutral 1.1M
Positive 210K
Negative 112K

* These research findings are from a draft chapter. The final version will be available in Handbook of Environmental Impact Assessment edited by Alberto Fonseca, forthcoming 2022, Edward Elgar Publishing Ltd. The material cannot be used for any other purpose without further permission of the publisher and is for private use only.
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:
- Investment proposals
- Bond prospectus’
- Corporate reports
- Impact Assessments
- Regulatory Documents

Portfolio Management

Analyze performance
- 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

Florian Skene

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