1. Type of data needed and purpose of the machine learning tool

What is the final purpose of the data being gathered? What are you basically analyzing here?

The final purpose of the data being gathered and the machine learning tool is to support companies in identifying forced labor risk in their supply chain. GFEMS will then work with at-risk companies that are willing to engage to mitigate this risk and improve labor conditions in their supply chain. Potential solutions include the introduction of worker voice tool and grievance mechanisms or ethical recruitment practices. Solutions will depend on the profile and needs of companies.

“GFEMS is primarily looking for data on labor conditions that can be linked to specific and identifiable companies. (Page 2)” - Please clarify what is meant by “identifiable” here? Does anonymized data with unique ID for each enterprise count?

Our general goal is the identification of operational or supply chain features that have causal relationships or very strong correlations with forced labor. This can be accomplished with anonymized data if it also includes operational or supply chain features like number of trade partners, purchase order patterns, direct employment, etc.

What kind of variables constitute labor conditions? Or is it something we need to propose to you?

GFEMS is seeking data that is indicative of forced labor in specific companies. Variable such as number of working hours per day, number of working days per week, income per employee, etc would provide such information. Reports of forced labor cases in specific companies would also be clear evidence. Please refer to ILO 11 indicators of forced labor for a better understanding of what typical indicators would be: https://www.ilo.org/global/topics/forced-labour/publications/WCMS_203832/lang--en/index.htm

In India, firms do not keep details of labor conditions unless they are legally obliged by their investors/clients/governments. How do you plan to factor in this limitation?

This challenge is GFEMS’s main reason for seeking a data collection contractor. GFEMS is aware of the limitations. However, the Fund has come across organizations such as NGOs, companies or social auditors that have collected extensive information on labor conditions at company level. The Fund expects the contractor to identify such organizations as well as ways to access this data (Either through purchase or partnership).

What are the key indicators of data collection?

To be decided in contracting, but potential examples include number of unique cases or named entities, number of independently confirmed cases, detail (method of control, method of recruitment, presence of ILO indicators, for instance), relevance (if collected by automation).

Is there any reference or direction so that we can get in and collect the data?

The focus for the contractor will be collection of existing data (This means secondary data, no primary data or worker surveys). As one of the main components of the scope of work is to identify existing sources of data, GFEMS will not be providing reference or direction.
Is the data collected expected to be more qualitative or quantitative?
GFEMS is mostly interested in quantitative data but qualitative data can be valuable too as long as it is structured.

Is there a database of companies available to start with?
There is no predefined database of companies. In priority, the contractor should collect data on companies that have been investigated for forced labor cases or other violations of regulations.

Since the RFP expects company specific data and not aggregated data, is the idea that we select a particular company in a particular sector and collect data from all stakeholders in the company including the workers?
The goal of the project is to identify the profile of as many companies with forced labor cases as possible, to support them at a later stage (beyond this RFP) in improving their practices. Therefore, rather than collecting data on all the workers of a single company, the contractor should look for as many companies as possible with potential cases of forced labor. GFEMS is interested in a diverse set of companies rather than in in-depth information about a single company. The information can be provided either by NGOs, companies themselves, social auditors or via media reports. The contractor should think creatively of other potential sources.

What sectors would you like to cover and what size of companies would you like to include or exclude from the sample?
There is no predefined sector or size of companies. The priority is to collect as many data points as possible. Therefore, the contractor should look at a broad range of sectors and company types. Greater diversity will increase the performance of the future machine learning tool.

What if we are collecting labor conditions data at firm level? Can we use that?
If the contractor already has data on labor conditions, especially with indications of forced labor, it can be included in the scope. However, it should not exempt the contractor from identifying new sources of secondary data.

How will the geography be selected for the sample?
There is no predefined geography within India. However, the broader the number of states covered the better. This will provide more potential data points and increase the accuracy of the future machine learning tool.

How large should the dataset be?
This depends on the quality and nature of the data. If using automated scraping, we believe it should be possible to identify hundreds to thousands of unique entities with credible evidence of forced labor. If data is compiled by hand with high quality standards for timeliness, relevance, and accuracy, hundreds of cases might be more realistic. Identifying companies that are credibly free of forced labor throughout their supply chains alongside those that are not could also increase the training power of the dataset.

2. Staffing, costs, timeline and administrative questions

“Including designing and deploying surveys” - It is stated that the ideal contractor must be an expert in primary data collection but there is no mention of primary data work in the scope of work section. Please can you confirm if the primary data work is only meant to supplement the secondary data analysis whenever there is a gap identified?
The contract doesn’t need to be an expert in primary data collection. The focus of the RFP is existing data/secondary data collection. GFEMS doesn’t expect the contractor to conduct any primary data work or worker surveys.

**How many days is the field work anticipated to be completed in?**
The amount of field work has to be determined by the contractor and will be discussed with GFEMS during the contracting phase. GFEMS expects the number of field days work to be limited as the contractor should focus on collecting secondary data, not primary data through surveys.

**Will any resource person be joining during Field team training?**
The contractor should determine whether a field person is needed and include the cost in the total budget. However, as the focus is secondary data collection, GFEMS would not expect the need for field team training.

**Please specify the level of supervision from GFEMS**
GFEMS’s Project Manager and Senior Data Scientist will be reviewing the data collected at several points in time during the project and provide feedback so that the direction of the research can be corrected, if needed, as early as possible. However, GFEMS expects a limited need for supervision during the data collection phase (Component 1. of the scope of work). GFEMS expects increased collaboration between the contractor and GFEMS’s Senior Data Scientist for the analysis of the collected data (Component 2. of the scope of work).

**The RFP specifies that both qualitative and quantitative data need to be collected: Will there be a reviewer from your team for the instruments?**
GFEMS’s Project Manager and Senior Data Scientist will be reviewing the data collected at several points in time during the project and provide feedback so that the direction of the research can be corrected, if needed, as early as possible.

**Will your firm cover the fees for the proprietary data that might be required for this activity?**
The role of the contractor will be to identify sources of existing data/secondary data and how to secure it. If the quality of the data identified is satisfying enough, and based on value for money, GFEMS will cover the fees for the purchase of the data.

**Is there any government level permission letter available to show for the interviews?**
GFEMS cannot provide such letter.

**Who will design the survey tool?**
No survey tool will have to be designed. Data collected should be secondary data/existing data. GFEMS will not fund activities to conduct surveys for primary data collection.

**What will be the approximate budget?**
GFEMS expects the total cost to not exceed USD 50,000.

**Can final datasets be submitted in the month of June-July?**
Depending on the quality of the interim datasets provided, submission of final datasets could be accepted later than indicated on the RFP. This would have to be discussed between GFEMS and the selected applicant during the contracting phase.

**Is IRB required or already available?**
We envision the contractor collecting secondary data, which would not require an IRB, but we will facilitate an ethical review, if needed.