Use of AI Technology
to Support Data Collection for Project Preparation and Implementation:
A 'Learning-by-doing' Process

THE WORLD BANK
IBRD • IDA | WORLD BANK GROUP

GFDRR
Global Facility for Disaster Reduction and Recovery

Use of AI Technology
to Support Data Collection for Project Preparation and Implementation:
A 'Learning-by-doing' Process

GPSS Global Program for Safer Schools

Contact information
gpss@worldbank.org

THE WORLD BANK
IBRD • IDA | WORLD BANK GROUP

GFDRR
Global Facility for Disaster Reduction and Recovery

Table of contents

Acknowledgments

1

Objective of the Note

This is a guidance note for World Bank task teams seeking to use artificial intelligence (AI) technology to support field data collection activities for project preparation and implementation. The note stems from an ongoing school infrastructure-related project under the Global Program for Safer Schools (GPSS). Rather than being a comprehensive AI technical document, the note describes the key factors to be considered by task teams during project preparation and implementation, with potential opportunities and benefits as well as challenges and risks. The note aims to guide task teams in support of clients with field data collection activities, in either Bank-executed trust funds (BETF) or client-executed trust funds (CETF) projects.

Innovations are needed in field data collection methods—not only to improve project efficiency but also to enable continuous operations under mobility restrictions, such as in post-disaster or fragility, conflict, and violence (FCV) contexts and during the ongoing COVID-19 pandemic. In most cases, field data collection is the necessary initial step in identifying the baseline for project preparation and implementation because of the lack of data in developing countries. Such data collection is usually to be conducted by specialists with necessary technical understanding to ensure the completeness and quality of the data collected. In developing countries with limited technical capacity, this leads to the need for procuring, training, and mobilizing qualified specialists on the field including remote areas. This can be costly and time-consuming and sometimes result in data quality issues when training is not sufficient. Such field data collection activities can be even more challenging when it comes to post-disaster or FCV contexts due to mobility restrictions. Given the current health risks and travel limitations in most countries posed by COVID-19, the need to change the traditional methods of field data collection has become even more pressing.

The recent advances in AI have provided the opportunity to address such needs, through augmenting expert knowledge and widening the spectrum of data format to inform project preparation and implementation. Since the early AI research initiated in the 1950s, the use of AI benefited from the growing data volumes, algorithms, computing power, and storage and has been blooming in various areas, from the tools used in our daily lives such as personal assistants on our smartphones to smart operating systems such as self-driving cars. While AI is continuously evolving and aiming toward simulating human intelligence for a wide range of human-like cognitive tasks, AI today can be designed to perform a specifically defined narrow task with the possibility of outperforming humans. However, different from robotic automation, AI is not a direct automation of manual tasks but an automation of repetitive and progressive learning through data. This allows expert knowledge to be transferred into AI models through existing data, complementing expert resources in developing countries for data collection and processing. From various machine learning (ML) algorithms to today's deep learning algorithms, AI is becoming increasingly capable of analyzing larger and deeper data, including numbers, languages, images, audios, videos, and so on. The flexibility in the format of data that AI can handle further opens up the possibilities to conduct data collection remotely or with the help of the local community without a technical background.

While the power and flexibility of AI could provide strong support in field data collection, it is important for task teams to define necessary technical requirements and identify the most suitable and sustainable form of data collection. As AI learns directly from data, the role of data has become more important than ever before. With more and better-quality data, AI becomes more accurate and reliable in performing particular tasks. Also, AI solutions may not be suitable for all types of problems and conditions. To efficiently harvest the power of AI, several elements need to be taken into account to shape a reliable and efficient data collection approach in the long term. In this note, some key considerations for task teams are discussed.

2

Key Considerations during Inception: A Working Backwards Approach

This section discusses some key considerations in a working backwards approach that task teams may have to adopt for the inception stage of the project, to utilize AI technology in facilitating field data collection for project preparation and implementation.

What is the scope of the problem and is AI a good candidate to solve the problem?

At the inception stage, task teams have to fully understand the scope of the problem, address the challenge, and identify the type of solution required. A good approach is to 'work backwards' from the needs of targeted users while ensuring that any realized benefits flow down to the end beneficiary. Clearly outlining the problem that needs to be solved, the users/beneficiaries, and the impact sought is critical for identifying the type of solution required and defining the appropriate use case. Based on the scope of the problem, task teams should further evaluate if AI would provide a suitable and efficient solution and what AI capabilities are called for. Major characteristics of the problem that would successfully benefit from AI solutions include the following:

- The difficulties associated with large-scale data-related operational efforts. For instance, if the problem is to collect detailed information about limited cases (for example, 20 cases), the development of AI solutions would not be beneficial.

- The information needed from the data is not on a case-by-case basis, meaning there are standardizable patterns for scalable use cases. This also means any labeling effort should not be a case-by-case work. For instance, if the problem is to collect site-specific details that cannot be grouped/clustered, it may not be a good AI use case.

- The size of the data available or accessible is sufficient for the training of AI models. Although the amount of data needed to train a good AI model depends on multiple factors such as the complexity of the problem and the model, a general rule of thumb is to have at least 10 times the amount of training data compared to the number of parameters in the model.

- There is an enabling data regulatory environment for the local government to adopt AI solutions, considering data security and privacy concerns.

What information is needed and what could be the format of the data collected?

Once the scope of the problem is defined and identified to be a suitable AI use case, task teams must clearly define the scope of the information

and level of details to be captured in the data to ensure that the data collected will meet the needs of the project. Also, for AI to drive impact, the identified human benefit and user experience must be translated into a concrete pattern represented in the existing or future collected data. This important prework will guide the selection of suitable data sources and formats, efficient AI-based methods, and data collection methods. The data formats that can be handled by AI-based methods include number, text, time series, photo, satellite imagery, video, and audio, among others. The selection of data formats highly depends on the information to be extracted. For example, understanding the baseline in infrastructure projects usually requires engineers to collect visual characteristics of structures on the ground with relatively detailed resolutions, which leads to potential data formats such as photos or videos with the necessary level of spatial resolution.

What would be the size of the data needed?

With the format of the data identified, task teams need to further understand the data speed and size. By doing so, task teams will have an idea about how quickly records can be created, how much data have to be stored for identifying suitable cloud solutions, and how much data preparation effort is required (for example, labeling of data). For instance, images and videos will require

much more cloud storage capacity than text and numbers. In addition, task teams must understand that depending on how much data are available and accessible in the short or long term, the data collection for developing AI solutions with satisfactory performance can be a rather iterative and continuous process. This note identifies it as a 'learning-by-doing' process.

How can AI support the workflow?

Knowing the format of the data to be collected and the information needed from the data, task teams will have to determine the step/s that can be supported by AI for improved efficiency. This requires teams to identify the part of the workflow between inputs (data collected) and outputs (information needed) that is usually done by experts. Also, task teams need to find a value proposition and an ideal output. Listing value propositions can help task teams and experts to have a clear idea on objectives. For example, AI could support key parts of the workflow to limit the need for mobilizing experts, thus freeing up their time for higher-value tasks and reducing logistic costs of projects. However, AI solutions help augment, but do not replace, human expertise. An AI model may need to adapt to unknown scenarios which could be caused by changes in time, location, and so on. As a result, it will always be necessary to have human experts in the loop to supervise AI's operation.

Each AI use case is unique, and several factors will drive the required expertise and methodology. Those factors include the format and resolution (for example, number of different classes) of the inputs and outputs, suitable type of AI-based methods, algorithms, and the data preparation needs. In general, AI-based methods can be categorized into supervised learning, semi-supervised learning, and unsupervised learning, with decreasing needs in data preparation and decreasing maximum resolution in the output. If the use case requires the classification of only simple and broad clusters with obvious and distinct differences to be identified from data, both supervised and unsupervised learning can be considered depending on the project and data context. However, unlike supervised learning, unsupervised learning leverages unlabeled data to extract features and patterns, so that no labeling, and less data preparation effort is required. On the other hand, when more complicated and detailed classifications or predictions are needed, supervised learning would be necessary for preparing completely labeled training data. Semi-supervised learning is a mix of both that is suitable for handling tasks with complexity, output resolution, and data preparation needs in between. Table 1 briefly identifies these main categories of AI-based methods, data preparation needs, general types of models, and required expertise. In this note, more focus is given to supervised learning, aiming to share the lessons learned from the GPSS experience.

TABLE 1 **Identification of main categories of AI-based methods and algorithm, data preparation needs, and expertise**

AI-based methods	Data preparation needs	Types of models	Example of use case	Expertise needed
Supervised learning	Need full training data: completely labeled (inputs and labeled outputs)	• Classification • Regression	Real-time wildfire prediction	• Data science • Data engineering • Cloud computing • Python • Tensorflow and Keras • Convolutional Neural Network (CNN) Depending on the format of input data, the following may also be needed: • Image processing • Geospatial analysis • Time series analysis
Semi-supervised learning	Need partial training data: partly labeled with rest unlabeled	• Clustering • Association • Dimensionality reduction • Classification • Regression	Building footprint extraction	
Unsupervised learning	Do not need training data: completely unlabeled	• Clustering • Association • Dimensionality reduction	Using geospatial data to cluster land cover (for example, water, forest)	

How should the AI solution be evaluated?

As the AI solution development is usually an experimenting and iterative process, it is important for task teams to define a single-number evaluation metric for the AI model. This will help them compare the performance of different models based on this metric and choose the best model. With various metrics to evaluate the AI model performance, it is essential that task teams understand and clearly communicate with stakeholders on how the model is evaluated and selected. Also, task teams need to consider the size of the testing dataset for performance evaluation. A larger testing dataset will provide more reliable performance evaluation. Some common metrics to evaluate AI models for classification and regression are discussed.

For classification

1. Accuracy

This is perhaps the most straightforward metric to use and explain – the ratio of number of correct predictions to the total number of predictions made. However, it works well only if the dataset is balanced with the same amount of samples per class. Otherwise, it may provide a false sense of high accuracy when it is just predicting the dominant class (most frequent class) in the dataset. Such cases create bias against the underrepresented classes.

2. Precision

For imbalanced datasets (one class is much more frequent than other classes), class-specific metrics are needed. Precision looks at the ratio of true positive prediction cases to the total of true positive and false positive predictions of each class. It evaluates the precision of the model.

3. Recall

Another important metric for imbalanced datasets is recall, which is the ratio of true positive prediction cases to the total of true positive and false negative prediction cases of each class. It evaluates the robustness of the model.

4. F1 score

F1 score is the harmonic mean between precision and recall, with a range between 0 and 1. It evaluates both the precision and robustness of the model. A high precision but low recall model is accurate when it predicts true positive cases, but it misses a large number of cases that are difficult to classify. F1 score provides a good overall performance metric.

5. Area under curve (AUC) or average precision (AP)

AUC, sometimes called AP, is specifically used in binary classification problems. It evaluates the probability that the model ranks a random positive case higher than a random negative case based on the confidence level of prediction. AUC/AP has a range between 0 and 1. Higher AUC/AP indicates better model performance. When there are multiple classes, the performance can be evaluated through

mean average precision (mAP), which is basically the average of the AUC/AP of each class.

For regression

1. Mean squared error (MSE) and root mean squared error (RSME)

MSE is perhaps the most commonly used metric for regression problems. It evaluates the average squared error between predicted and real values. It is sometimes used in terms of RSME, which is the square root of MSE. The lower the MSE or RSME, the better the performance of the model.

2. Mean absolute error (MAE)

MAE evaluates the average absolute difference between predicted and real values. It is more robust to outliers than MSE.

Can any existing data be used?

As discussed in Table 1, different levels of effort would be needed to prepare the data for the AI model to learn from. Thus, before collecting any new data, it is important to review the existing data, make use of the data to prepare the AI component in the data collection workflow, and standardize any future new data collection. This use of existing data for AI model development includes the following:

- To review and define a systematic structure for the input and output for the AI algorithm

- To identify the needs of data preparation (for example, labeling of training data) or collection of new data and prepare the data in the defined systematic structure

- To conduct quality control of the data prepared, which is fundamental to the performance of the AI model

- To make sure existing datasets have no data privacy and bias concerns

- To check the data governance and security components of the existing datasets

- To develop, train, and test the AI model using the prepared data from the existing data

- To define consistency in the existing data

and inform the standardization of future data collection compatible with the AI model

- To monitor the AI model and input data for drift once deployed.

For instance, if task teams have previously collected field data or existing data received from relevant stakeholders (for example, infrastructure inventory held by the government) and identified the use of field photos together with satellite imagery as input for a classification output, it needs to prepare the existing data following the same structure and format defined for the target AI component in the workflow. Such preparation process could include cleaning of the data, quality assurance review, and necessary preprocessing of the data, which usually has to be conducted by data engineers/scientists and/or geospatial specialists with the support of other relevant specialists/engineers. In case no existing data are available for the training of the AI model, task teams should consider the preparation of training data in new data collection activities or the generation of synthetic data. Such data collection activities typically require intensive

manual labor to label the data, for which tools for capturing and annotating the data need to be prepared. Once the AI model is trained and tested for satisfactory performance, task teams can deploy the AI model in the workflow with standardized and compatible data collection activities in the future.

How should the data be collected?

Based on the type of data needed, task teams can identify a suitable and efficient individual data collection approach or combination compatible with the AI model for new data collection activities. Table 2 suggests the overall considerations to identify the data collection approach. The pros and cons of some commonly considered data collection approaches that could be paired with AI-based support are briefly discussed below. Note that the possible sources of data include but are not limited to those mentioned below. Technical support on sources of data and data-related expertise is available within the World Bank, such as from the ITS Technology and Innovation Lab (ITSTI) and Geospatial Operations Support Team (GOST).

TABLE 2 Overall consideration to identify suitable data collection approaches compatible with AI-based methods

Considered data collection approach	Type of data	Technical skills required for user adoption	Resolution of the captured information	Cost	Potential limitations
Satellite imagery	Image	High	High ↓ Low	High ↓ Free	Coverage availability Extractable details Time-consuming ground truthing
UAV	Image, video, sensor data	Medium	High ↓ Low	High ↓ Low	In-country regulations Weather limitations Risks to project operation Low communication level
Mobile app	Image, video, number, text	Low	High	Low	Availability of smartphone Availability of cell phone signal Potential gender and community segmentation
Social media	Image, video, number, text	Low	Uncertain	Low	Availability of internet Reliability and completeness of information Potential gender and community segmentation
Sensors and IoT	Image, video, audio, and other digital signals	High	High	Depends	Platform fragmentation Data privacy and security

- Satellite imagery

Satellite imagery can be either cost- and time-efficient to enable remote data collection or highly expensive depending on the required spatial resolution of the image and the scale of the data collection. The spatial resolution of satellite images can range from finer than 1 m per pixel to coarser than 30 m per pixel. If the needed level of detailed information from the image is relatively coarse – to understand the land cover (for example, water, forest, vegetation, urbanization) – free satellite imagery sources can provide adequate spatial resolution. However, if more detailed information is needed (for example, plan and elevation dimensions of buildings and other characteristics such as health conditions of structures at the national level) and the coverage scale is large, collecting data using satellite imagery can be fairly expensive and extracting the detailed information could be challenging. In addition, coverage limitations may apply to areas without high-resolution imagery or in post-disaster scenarios when imagery is needed quickly but is blocked by clouds. Task teams should also be aware that ground truthing is necessary for using satellite imagery, which can be a time-consuming process. Another thing to note is that collecting data using satellite imagery requires high technical skills in geospatial analysis and the engagement of geospatial specialists.

- Unmanned aerial vehicles (UAVs)

UAVs, or drones, can provide relatively remote data collection as they are flexible to operate and less affected by time. The data resolution can be as fine as 5 cm per pixel, depending on the flying altitude. This can be quite helpful in a post-disaster context to survey hard-to-reach sites and other types of dangerous jobs and reduce the risk of a manned solution. UAVs can also collect videos or other types of sensor data (for example, Light Detection and Ranging, known as LiDAR). The cost of data collection using UAVs depends on the required spatial resolution of the data. If higher spatial resolution is desired, to support the visual inspection of infrastructure conditions for instance, the flying altitude will be limited and the drone pilot will still have to travel to areas near the site to ensure the drone stays within the visual line of sight, which leads to considerable total cost and time. This is to ensure the safety of the people and property along the flight path as well as the UAV device. A recent UAV experience of the GPSS on collecting visual inspection aerial photos of school infrastructure indicates an average cost of US$890 per school, with a flying altitude of 50 m. On the other hand, if the required spatial resolution is lower, such as for mapping surveys, the cost-efficiency can be improved with less mobilization of pilots needed and faster coverage of large areas. Further, it is worth noting that the potential limitations of using UAVs include the regulations in different countries on the operation of UAVs (for example, regulation on flying altitude and areas), non-flying weather conditions, and risks to the project operation (for example, damages to UAV devices delaying the survey, falling hazards for the people and properties along the flight path). The operation of UAVs requires skilled pilots or training of inexperienced individuals (for example, government staff, university students).

- Mobile applications (apps)

Mobile apps have been widely used in various types of data collection. They make a versatile and valuable option due to (a) ease of implementation with a low requirement on technical skills for user adoption; (b) controllable format of the data to be collected; (c) flexibility on the information that can be collected through a smartphone (from detailed specifications to something as simple as taking photos); (d) reduced time lag between collecting data and analyzing and sharing the data; (e) more efficient two-way communication; (f) easier quality control than paper solutions; and (g) provision of accessibility for illiterate people to participate (for example, icon based). The cost of having a mobile app can be relatively low or even free using existing survey tools on the market. The resolution of the data collected through a mobile device can be as detailed as high-resolution photos of cracks on a wall. However, a relatively remote data collection initiative using a mobile app requires good participation of local communities in the site area, local availability of smartphones, and near-site cell phone signal availability. Otherwise, surveyors will

still need to be mobilized to collect data on site. Another thing to note is that the expertise required to manage the digital technology may invite greater reliance on external facilitators. Yet lack of uniform smartphone access may lead to potential gender and community bias which could skew the data. Other than smartphones, SMS mobile phones should not be discounted as a potential method of data collection as that technology may represent the only available communication infrastructure in a given community. For example, SMS data sources might include structured, transactional text messages between parties. These could be of significant value (for example, SMS-based microfinance) and may represent a viable data collection or intervention point well suited for AI application depending on the preidentified use case. There are existing application program interface (API) services to enable two-way SMS data feed integration which can help bridge the gap between SMS devices, smartphone applications, and cloud-based applications.

- Social media

Social media is another source from which different types of information can be collected free of charge and accessed remotely. However, such data availability requires internet connection, and the resolution, completeness, and reliability of the data are uncertain. Note that lack of uniform social media access may lead to potential gender and community bias which could skew the data. Also, the process of collecting data from this source will require skilled data engineers/scientists to gather, compile, clean, and structure data in various formats.

- Sensors and internet of things (IoT)

With the technology advances in recent decades, there has been a significant increase in the use of sensors, from monitoring weather (for example, humidity, air quality sensors) or infrastructure health conditions (for example, deformation, vibration sensors) to supporting public security, transportation, or smart homes (for example, motion, video, and audio sensors). With all these differently purposed devices embedded with sensors and connected over the internet to exchange data, a network of physical objects

is formed, which is known as IoT. The IoT could provide ample and various types of data, supporting a wide range of data analytics. However, due to the variety of devices that make up the IoT, the hardware and software variations lead to a lack of interoperability and common technical standards in the data. This makes it challenging for the data to work consistently between inconsistent technology ecosystems, which is known as platform fragmentation. In addition, privacy and security considerations should be included to adopt IoT solutions, as the data collected in the IoT may contain details about individual end users of the device which can be vulnerable to privacy violation.

In addition to the abovementioned data collection methods, another consideration and positive factor is the trend toward global deployment of cheap and ubiquitous broadband internet via satellite connections staged in low earth orbit. For instance, SpaceX's Starlink has launched 895 Starlink satellites out of a planned 12,000 to date and is currently advertising 100 Mbps at US$90 per month and US$600 start-up fees for equipment. Another example is Amazon's Project Kuiper, which aims to follow suit with a network of an estimated 3,000 broadband satellites in the coming years. These networks can transmit data to and from cloud infrastructure where many AI/ML services and workloads currently exist. While the affordability and adoption rates in the developing world may still be unclear, this is a promising trend for opening new data collection and communication horizons with client countries.

What are the tools and technical support needed?

Depending on different priorities and available resources, task teams may need different levels of effort in developing the AI component with associated tools for different types of users and scale. Table 3 suggests an example of phased approach to achieve AI support in the data collection work considering the different priorities and needs by use case and user. The note also includes annexes on terms of references (ToRs) for individual consultants and firms for reference.

TABLE 3 **Example of phased approach to achieve AI support in data collection work**

Phase	Aim of AI component development	Supported users - roles	User interface	Technical support needed
Initial phase: feasibility study	The initial base development can focus on having a prototype AI model in place to support any time-consuming and labor-intensive post-processing of data (for example, review of large amount of information and/or photos for classification). This is also the initial phase to validate the performance and feasibility of aimed AI component in the workflow.	Task teams or government stakeholders – own and manage the tool; Relevant specialists or engineers – end users.	During this phase, a complete product may not be necessary, and the user interface can be a simple local executable file or existing tools with AI capabilities.	During this phase, technical support from individual consultants in relevant fields, including computer science, AI engineering, data science, and so on, may be adequate. In addition to consulting support, cloud resources are needed for initial implementation. Note that data security considerations should be referred to World Bank ITS.
Scale-up phase: preparation for implementation	Once the feasibility is confirmed in the initial phase, the following phase can focus on having in place necessary tools for deployment of the AI model.	Task team or government stakeholders – own and manage the tool; Relevant specialists or engineers – end users for quality control; Non-technical end users for data collection (for example, community, government stakeholders) – end users for data collection.	User-friendly interfaces are needed for deployment. Depending on the use case, it can be developed from existing tools with AI capabilities.	To develop a complete deployable product, task teams may need to hire a firm to conduct the development and deployment of the tool/s in the project's cloud environment. Note that data security considerations should be referred to World Bank ITS and government stakeholders. In this phase, more emphasis should be on ensuring the security of the data, performance of the model, and integration with existing technology. During implementation, new data collected can be aggregated as new training data for model updating and validation.

How to validate and ensure data quality?

As AI learns from data, any mistakes or bias in the training data could be introduced in the AI component and the results from the AI support may not be fully reliable. Therefore, it is important for task teams to have a robust quality control process for both the data prepared to train the AI model and the data processed by it, in which field validations may be necessary. This means the AI model will by no means replace the important role of engineers/specialists in relevant fields; instead, human experts and AI need to work together for AI to provide successful support and augmentation of expert knowledge. Specifically, task teams need certain end users/clients/experts to help test the output of the AI solutions. In addition, the quality control process for the AI component can be an opportunity to contribute to the technical capacity building for the local communities (for example, involvement of local universities and students in relevant fields), for which task teams need to plan and adjust the capacity-building activities accordingly.

How to set up a sustainable framework in the long term?

As discussed earlier, in data preparation, collection, processing, and quality control, the development and deployment of the AI model to support the workflow require multiple considerations from task teams. Also, since AI learns from the data it is trained on, there will always be the need for it to learn and validate new knowledge from new data when the scope or context changes. Therefore, the development of such AI models is unlikely to be a one-time effort in the long term. However, a sustainable framework can be established to adapt to the changing contexts for task teams to improve and benefit from AI support in every data collection activity in the long term. This entails having user-friendly tools (for example, apps, dashboards) and reliable workflows in place to streamline those associated activities to allow the AI model to learn and update progressively. As an example, Figure 1 shows a conceptual learning-by-doing framework with experts in the loop, setting a progressive learning cycle for the AI component to improve its performance and adapt to different contexts. This type of long-term framework will (a) integrate expert knowledge in the initial training of the AI component, for the design and quality assurance of the key input data and (b) have quality control conducted by human experts to identify and correct potential mistakes, and allow the AI model to learn and improve from mistakes in new field data collection activities.

FIGURE 1 **Example of 'learning-by-doing' framework for long-term AI component development in data collection workflow**
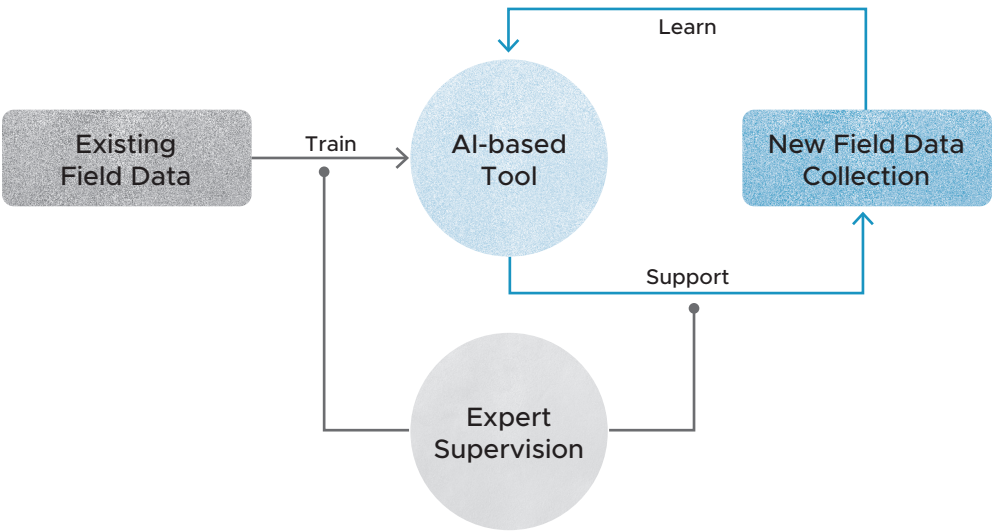
3
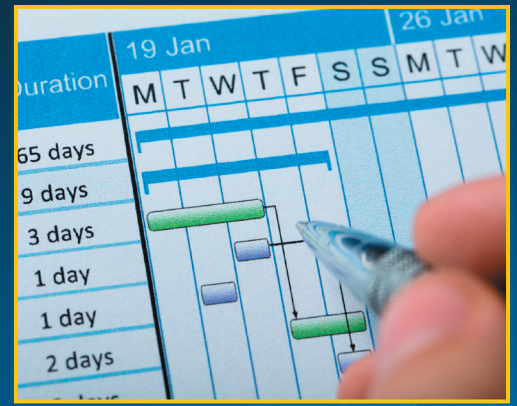
Timeline and Cost Implications

The timeline and costs for developing AI support for data collection activities depend greatly on the technical requirements; the desired resolution (for example, number of different classes) of the output information; and the existing resources including training data and a user interface. Therefore, the timeline and cost are not a one-size-fits-all scale but rather a case-by-case estimation depending on the specifics of a particular project. However, it is possible to understand the timeline and cost implication through major influencing factors, which are illustrated in Figure 2 and discussed under this section.

FIGURE 2 **Overview of timeline and cost implications for developing AI support for data collection**
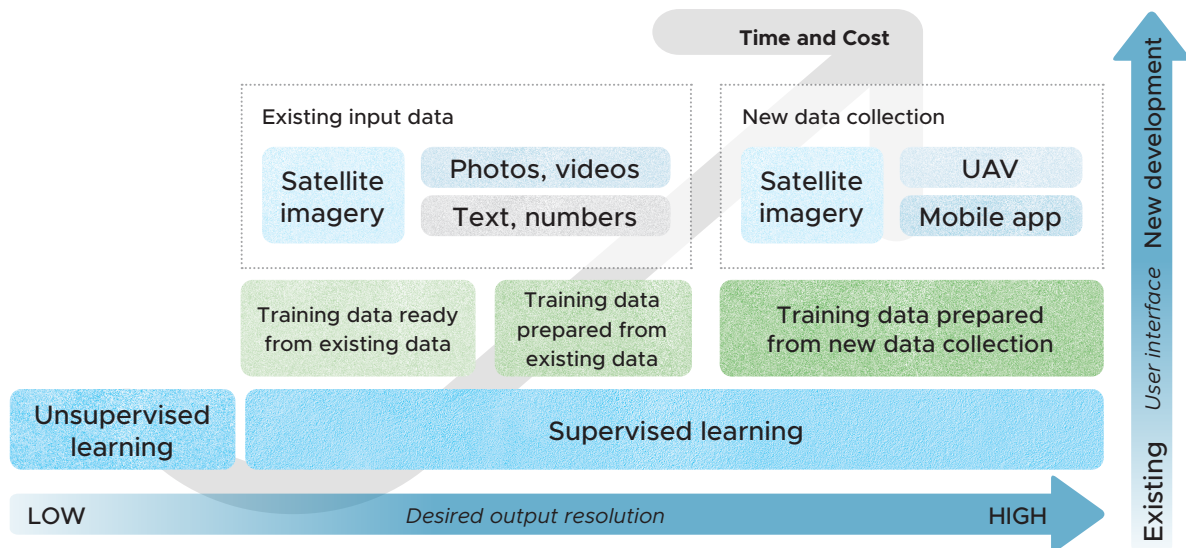
In general, the higher the resolution needed in the output information (for example, finer classes in a classification problem), the more the training data and computing power requirement (for example, memory, Graphics Processing Unit (GPU), cloud computing), leading to longer time and higher cost. In the case of low-resolution requirement for the output information, such as a coarse classification of land cover (for example, water/land) from satellite imagery, unsupervised learning of AI models can be applied without the need to prepare training data, which could indicate a shorter time frame (in terms of weeks). On the other hand, when high-resolution output information is required, such as detailed categories of building structural systems or geometric measures of building objects, supervised learning of AI models is needed, and training data must be prepared in terms of annotation or measurement. For preparing training data, the time and cost further depend on the availability of annotation/measurement data from existing datasets. More efforts required in the preparation of training data (for example, labeling) imply longer time and higher cost. Especially when there is a need to collect new data and conduct new data labeling for training data, the timeline required can be months, including the training data collection and preparation and AI component development, with cost ranging from thousands to hundreds of thousands of dollars. When considering the additional design and development of user interfaces, such as a mobile app or an online dashboard for the deployment of AI support, the cost can be even higher. Based on the specific needs of the project and feasibility, it is suggested that task teams first explore existing interfaces with AI capabilities and possible pretrained models, instead of developing the whole tool with a user interface from scratch. Table 4 lists some indicative costs and time ranges under a couple of assumed scenarios with high-resolution needs for output information (for example, classification of building structural types), to give task teams an idea of the magnitude of timeline and cost.

TABLE 4 **Assumed example scenarios with rough timeline and cost indications**

Assumed Scenario	Training data	User interface	Team needed	Timeline	Cost (US$)
A	Ready in existing data	Able to use/integrate with existing tools	Individual consultants	10–12 weeks	Several thousands
B	New data collection	Able to use/integrate with existing tools	Individual consultants; Firm	3–6 months	Several hundreds of thousands
C	New data collection	Develop tools from scratch	Individual consultants; Firm	6 months to over a year	Several hundreds of thousands and more

4

Risks and Challenges

Although success is seen in many AI applications, the use of AI models is still case specific and contains uncertainties. This means that task teams have to themselves identify the most suitable application approach tailored for their project needs. This section shares some common risks to be mitigated and challenges to overcome in the use of AI to support data collection.

The main risks task teams need to be aware of while using AI to support data collection are as follows:

- **Potential liability issue induced by the technical rigor and quality of the AI-based methods as well as the outputs from the AI models.** As the AI task to be performed is specified by task teams, it is important that those clearly define the technical rigor and resolution of the desired outputs and plan necessary input data accordingly. In general, the output results from AI-based tools should be used as evidence to inform decisions rather than as direct recommendations. For example, the use of AI to support the identification of building vulnerability should be a classification task on building structural types to inform the vulnerability assessment instead of directly expressing a measure of vulnerability. Further, as AI learns from the data it has been trained on, it is critical for task teams to have adequate supervision in place to ensure the quality of the training data. Also, it is worth noting that the results from AI-based tools always come with potential uncertainties due to the nature of AI methods,

especially when applied under a new context it has not been trained on. Therefore, task teams must have certain quality controls over the outputs from AI support as well as the ability to monitor the performance characteristics of the AI model on a continuous basis.

- **Data management and security measures should comply with the relevant management information policy.**

- **Adoption by government stakeholders.** As the task and scope an AI model can handle with reliable performance depend on the data it has been trained on, clarifications should be given to the government stakeholders on the valid scope and period for the use of the AI-based tool. Also, capacity-building activities or technical assistance is needed to support the government stakeholders on the use of the tool and to inform them on potential updates required (for example, retrain the AI model with new data covering a new scope, context, or period).

The main challenge and opportunity presented by AI-based tools is that they entail more than one-time development: clear communications to

the stakeholders will facilitate the setting up of a long-term framework to benefit from AI solutions. Since the AI model learns from the data it has been exposed to, its reliability will grow over time as more data become available, for which AI integration into all relevant field data collection is worthwhile. On the other hand, its valid applicable scope and period are governed by the training data and needs to be updated when applied under a new local context. For example, in the case of infrastructure, the types and visual characteristics can vary from one country to another, or from decades ago to now, which makes it necessary to have the learning cycle proposed in this note to allow the AI model to adapt accordingly. In all cases, it will be essential that the AI-enabled solution has the ability to seamlessly integrate new labeled datapoints into the training set, allowing for the training and improvement of the model.

5

Case Study

This section shares a case study with relatively high-resolution requirements, to illustrate the journey of a task team to develop AI support for data collection activities. It originates from an ongoing GPSS school infrastructure project 'Enhancing Resilience in Kyrgyzstan' (ERIK) with the objective of improving the seismic safety and functional conditions of school facilities in the Kyrgyz Republic.

To inform an efficient seismic risk reduction intervention strategy and investment plan at scale, field inspection campaigns at the national level covering over 2,000 schools and preschools were conducted to gather necessary school infrastructure baseline and inventory information. Some lessons learned during the process as well as past large-scale field inspection activities like the ones discussed below motivated the task team to seek AI support in similar field data collection activities:

- The assurance of data quality with limited local technical capacity requires considerable time and resources. In such cases, the training of local teams over a limited time frame is not sufficient and data quality control by local teams would require multiple revisions. This is especially true with large-scale data collection at the national level. In the ERIK field inspection campaign, over 30 percent of the resources were allocated toward such quality assurance efforts. These additional quality assurance activities caused the project timeline to double.

- Weather and transportation difficulties pose further time delay and additional costs in

field data collection activities, for example, mountainous remote areas and unforeseen extreme snow. This also applies to the current travel restrictions induced by the global pandemic, which has caused an early halt of the field inspection campaign and left more than 40 percent of the national public schools and preschools uninspected.

To address these challenges, the task team started developing AI support in similar field data collection activities to meet the following objectives:

- Set up a sustainable framework to develop AI-based tool learning from field photos to enable growing support from the AI-based tool over time.

- Allow the AI-based tool to transfer the knowledge from experts to non-technical users, compensating shortage of experts with desired technical capacity.

- Use AI-supported field data collection to facilitate community engagement and capacity building. An AI-based tool could empower non-technical school communities to be part of the effort to

make their schools safer and more resilient with a better understanding of the government's decision-making process. Also, it could provide open access for engineering students to learn and practice with the expert knowledge carried by the AI-based tool on the structural typologies of school buildings and understand the implication for associated vulnerabilities.

- Standardize and streamline data collection, mitigating data quality issues. Since no judgment needs to be made by the user who collects data (that is, photos), the data collection can be easily

standardized and streamlined, with the quality of the assessment through AI support and remote supervision and quality control from experts.

- Reduce the need for mobilization and travel of engineers, improving project efficiency and continuity under any travel limitations, especially considering the ongoing global pandemic.

With the objectives discussed above, the specific considerations of working backwards are summarized in Table 5, followed by conceptual and proof-of-concept results, and lessons learned.

TABLE 5 **Case study: Specific considerations of working backwards**

Key considerations of working backwards	Specifications in the case study
What is the scope of the problem and is AI a good candidate to solve the problem?	The current practice of school infrastructure baseline data collection requires the mobilization of a large number of local engineers in the field and considerable effort in quality control work, which is susceptible to travel restrictions due to conditions such as snow constraints during winter and the ongoing global pandemic. Since the school infrastructure baseline data are collected for building types following a systematic visual classification methodology that is scalable, AI solutions are identified to have high potential to address this challenge by reducing the need for field mobilization and time delay risks.
What information is needed and what could be the format of the data collected?	As the evaluation of the school infrastructure baseline at scale is a classification problem based on visual characteristics of the school building to identify structural system typologies (that is, structural taxonomy), visual structural characteristics need to be collected and images can be the format of the data.
What would be the size of the data needed?	Since the targeted AI solution will be a classifier based on photos and there are already photos collected in the recent field inspections, the task team starts with a total size of about 18 GB of photos. The collected photos are stored in cloud services with accessible API for constructing a data pipeline for the AI model. The cloud service is directly linked with the field data collection mobile app, which would facilitate the iterative and long-term improvement of the AI model with new data being collected.
How can AI support the workflow?	Desired AI support should enable the end data collection to be deployed by local non-technical communities (for example, school staff, students) to reduce the need for mobilization of experts and improve efficiency.
How should the AI solution be evaluated?	To understand how the AI model performs on classifying the school building type parameters, the initial metric selected by the task team was accuracy. However, after experimenting with existing data and the model, the task team learned that the dataset is imbalanced especially for some of the structural taxonomy parameters, which led to consider other metrics such as precision, recall, and AUC/AP.
Can any existing data be used?	Photos collected in past field inspections of school infrastructure covering the similar structural types can be used to initiate the training of the AI support. Proof-of-concept AI model results are discussed below and in Table 6.
How should the data be collected?	As visual details are needed for both external and internal features of school buildings, photos collected through a mobile app provide the best flexibility and cost-efficiency to meet the needs. As a result, satellite imagery and drone data collection approaches are not considered here.

Key considerations of working backwards	Specifications in the case study
What are the tools and technical support needed?	To allow the AI-based tool to be deployable by non-technical communities, both the AI model and the tool interface must be developed. This requires technical support from AI engineers and IT front-end developers as well as guidance from structural engineers on the classification of structural types for the preparation of training data and quality control of the outputs from the AI model. A phased approach can be planned to prioritize the model development and integration with existing tool interfaces, followed by tailored user interface at a later stage. A proof-of-concept user interface design is illustrated in Figures 3-4 and discussed below.
How to validate and ensure data quality?	Robust quality assurance and quality control processes conducted by experts should be included in the data collection workflow. If needed, additional field validation may apply. An online platform can be set up for remote quality assurance and quality control in the workflow, as shown in Figure 5.
How to set up a sustainable framework in the long term?	A conceptual 'learning-by-doing' framework for the data collection workflow is illustrated in Figure 5.

Proof-of-concept: Image-based GLOSI structural taxonomy classification

Based on the structural taxonomy classification defined under the Global Library of School Infrastructure (GLOSI)[1] and photos collected and annotated in past field inspections in Nepal after the 2015 earthquake, an initial prototype AI model was trained and evaluated, as specified in Table 6, with output covering the classification of three main structural taxonomy parameters. Further, with the recent photos collected in the field inspection campaign in the Kyrgyz Republic, the model was adjusted and retrained to adapt to the local context of building type classifications. This initial and adjusted prototype model was developed under the collaboration of the GPSS, Digital Transformation Hub (DxHub), and California Polytechnic State University (CalPoly).

It can be seen from Table 6 that the initial prototype model provides relatively good classification accuracy for the taxonomy parameters with relatively coarse classes, while the parameter with much finer classes (that is, main structural system) has the lowest model prediction accuracy over the testing data. This is because (a) only a single façade

photo is available per school building from the existing data, lacking necessary features captured in the photo to support the classification of the main structural system, and (b) finer classifications lead to higher demand on the amount and captured features of the input data. This clearly indicates the need to further train the model with more and better photos and to tailor future field data collection to gather photos capturing more structural features (for example, both external and internal photos of the building). However, the prototype model sets an initial base for the integration of AI support in the data collection workflow for school infrastructure.

With the recent data available from the field inspection campaign conducted in the Kyrgyz Republic, the prototype model has been adjusted to a multi-input model instead of the initial single-input model, and to adapt to the new local context. The aim was to include important external and internal features of the building for the AI model to learn from. The input for the adjusted model includes four building side images and four internal diaphragm images. Table 6 demonstrates an improvement of the adjusted model in the prediction accuracy of the building category from 81 to 86 percent.

---

[1] <u>Global Library of School Infrastructure</u>

TABLE 6 Case study: Proof-of-concept prototype AI model

Prototype model	Existing data used to train/ retrain the prototype AI model	AI method and base model		Building category	Main structural system	Height range
Initial single input model[2]	About 18,000 school building façade photos, collected in Nepal after the 2015 earthquake. Training data: 85% Testing data: 15%	Convolutional neural networks; transfer learning.	**Model output**	4 classes	16 classes	6 classes
			Accuracy (testing)	81%	67%	95%
Adjusted multi-input model[3]	Over 16,000 photos capturing both external and internal building features, collected in the Kyrgyz Republic in early 2020. Training data: 85% Testing data: 15%	Xception/ InceptionResnetV2 via Python Tensorflow Keras	**Model output**	5 classes	12 classes	4 classes
			Accuracy (testing)	86%	43%	72%

[2] "School Safety Project" Documentation
[3] "Making Schools Safer and Resilient at Scale in Kyrgyzstan" Project Documentation

However, the prediction accuracy of the main structural system and height range decreased. During the model experiments and evaluation, the task team identified that the Kyrgyz dataset is heavily imbalanced, with some underrepresented classes as low as 3 percent of the dataset. To address this, data augmentation and weighted losses during training were applied to those underrepresented classes. However, in this case, the accuracy would not be a suitable evaluation metric and the model versions are not comparable using the accuracy metric. As the AI solution development is an iterative and long-term process, the task team identified future work to improve and evaluate the model using metrics more suitable for an imbalanced dataset, such as AUC/AP.

Proof-of-concept: user interface design

To enable the engagement of non-technical local communities in data collection and reduce the need for mobilizing engineers, an intuitive and user-friendly interface is required to ensure key input data are captured in the photos collected, following desired standards. Figure 3 shows the conceptual user interface of a mobile app designed under the collaboration of the GPSS, DxHub, and CalPoly to collect necessary photos for structural classification purposes. It includes an easy-to-follow tutorial with clear examples of the type of photos to be taken, an AI autodetection function to support the identification of key features to be captured through camera, and a concise report of the AI-identified structural typology suggestion. This will facilitate the engagement of local non-technical communities in the data collection process through an easy-to-use interface, potentially including students from those schools. Note that the targeted non-technical users will only take relevant photos following the guidance but will not label the data. The labeling through the interface is designed to be provided by the AI model in the back end, and it is still necessary to have experts in the loop for remote quality control of the labeling. Additionally, a simple interface will not compromise the technical rigor of the AI model, which can also be used to support local capacity building for engineers and students. Figure 4 presents the results from a survey conducted among CalPoly students to evaluate the suitability of the interface design for the age groups of 7–10 years, 11–14 years, and 15–18 years, who would be able to use such a data collection interface. The results indicate the potential for the school community to engage.

FIGURE 3 Conceptual user interface design[4]

Camera tutorial 1.0 → Camera tutorial 1.1 → Camera tutorial 1.2 → Camera tutorial 1.3

Camera in Use → Photo Capture → A.I Auto Suggestions → Generated Report
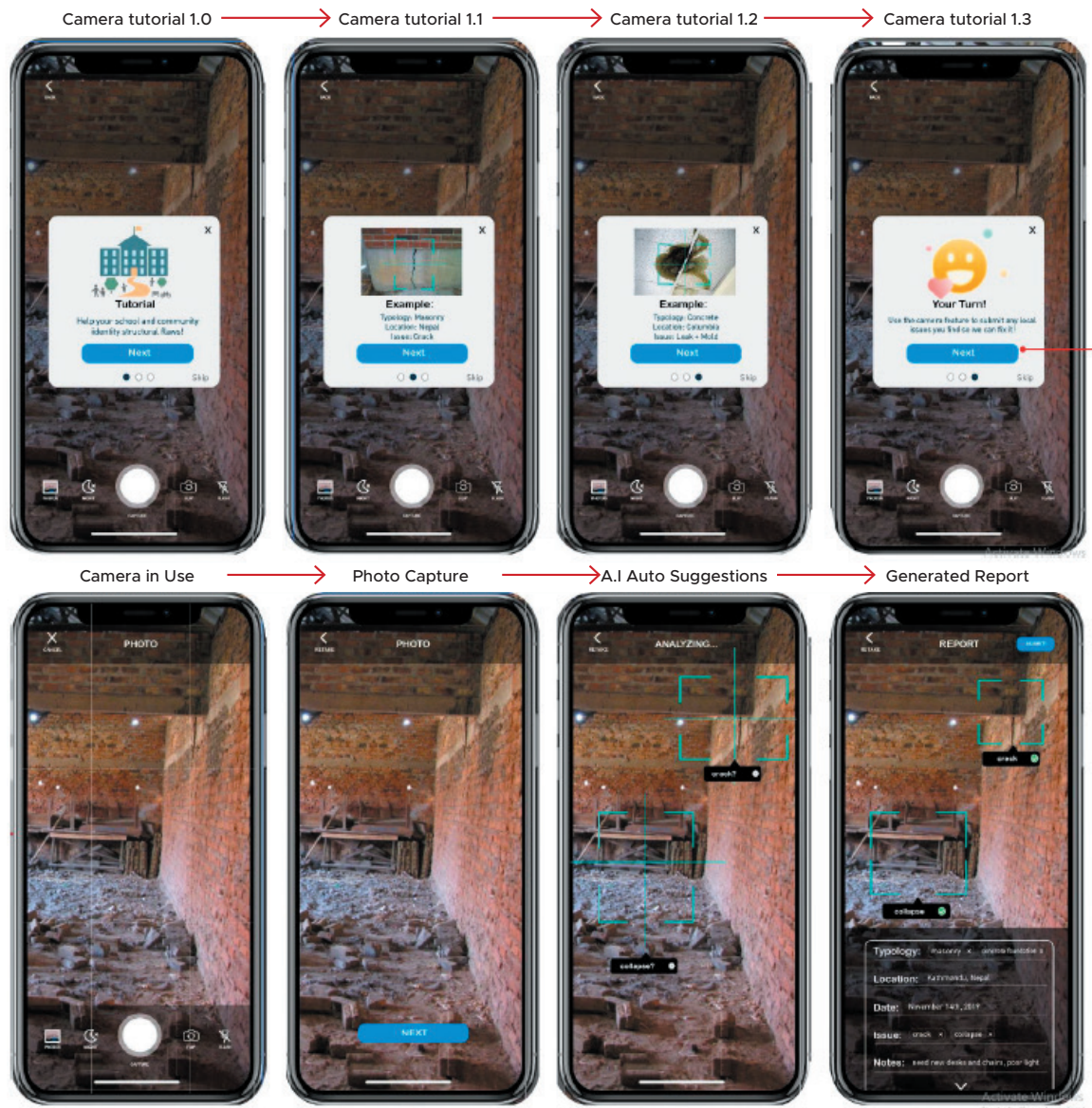
FIGURE 4 Survey results on the suitable age groups for the user interface design[4]

- 3-6 years old
- 7-10 years old
- 11-14 years old
- 15-18 years old

30%
40%
30%

---
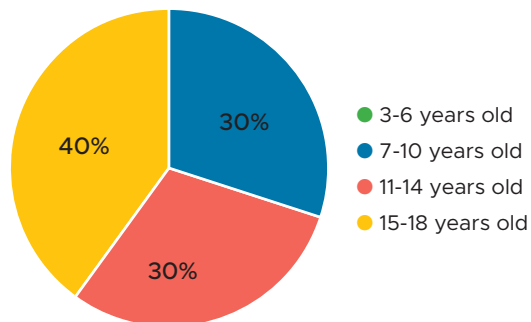
[4] Infrastructural Scanning Guide and Data Collection Tool

Conceptual 'learning-by-doing' AI-supported data collection workflow

A conceptual data collection workflow is proposed, as shown in Figure 5, to enable a progressive learning cycle in the long term for the AI-based tool with remote interventions of experts. It proposes that end data collection be functional both offline and online (that is, data can be recorded offline and synced once online), to allow simultaneous data inputs from school communities in different areas through the mobile app. With input data synced online, the AI model will be able to provide structural taxonomy classification suggestions. Further, with an online quality control and quality assurance (QA/QC) platform, experts will be able to sample and review the classification results from the AI model remotely and provide corrections as needed. If deemed necessary, further field validation may be considered. Such corrected classification will then be used to further train the AI model to improve its performance. This expert-in-the-loop approach will particularly be necessary in the context to which the AI model has not yet been exposed. After this QA/QC process, the structural classification results can be finalized to support

FIGURE 5 Conceptual learning-by-doing AI-supported data collection workflow for school infrastructure baseline data collection
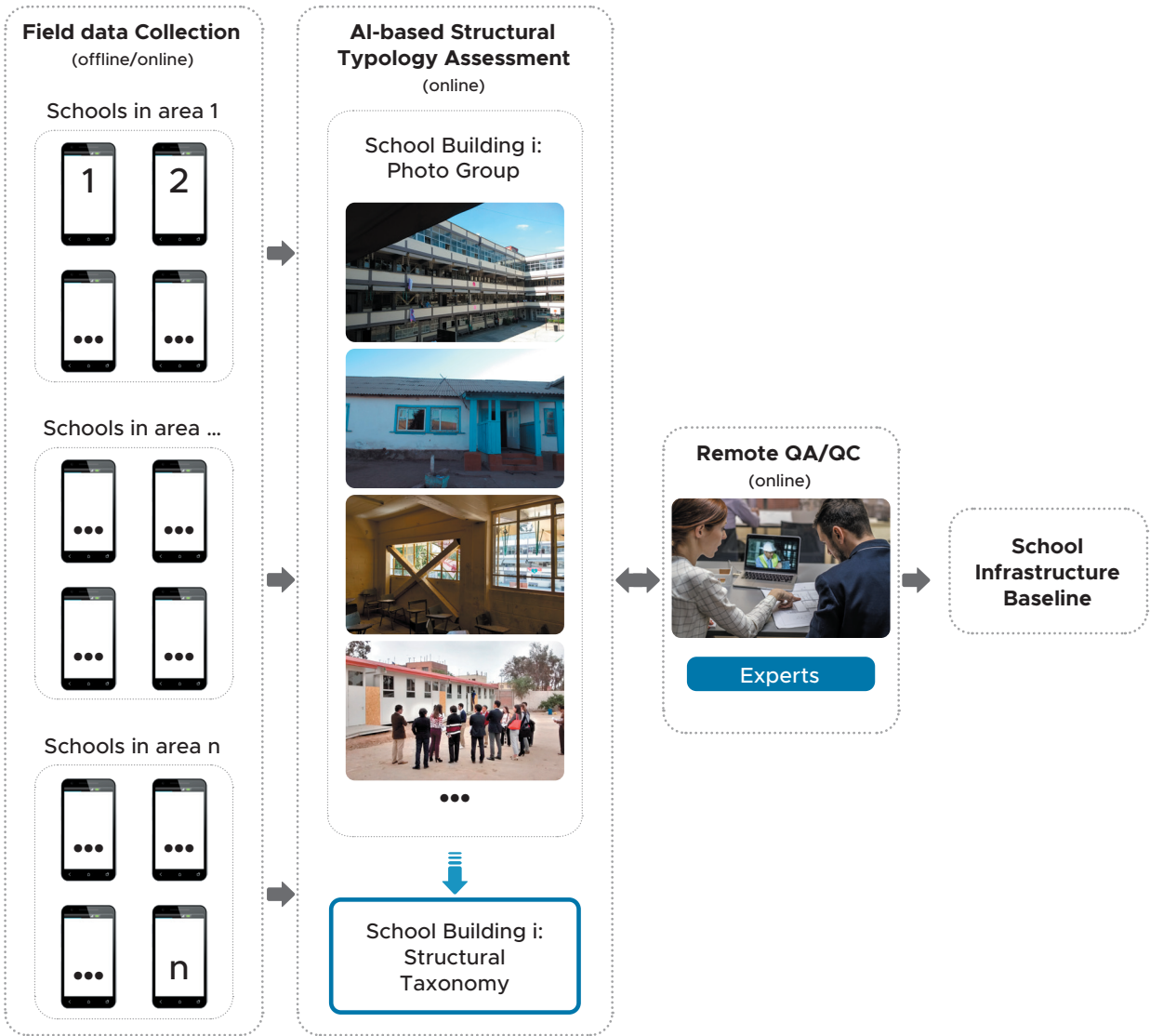
the further vulnerability assessment of school infrastructure portfolio and inform large-scale risk reduction strategies. With the proposed workflow followed in similar data collection activities, it is expected that the AI-based tool will be improved over time through a learning-by-doing process.

With this proposed AI-supported data collection workflow, based on the GPSS's recent field data collection experience in the Kyrgyz Republic, a potential reduction of more than 50 percent of the assessment time and the overall cost is projected, assuming school communities are successfully engaged in the data collection. Time and cost efficiency could vary depending on how smooth the engagement of the school community is in the specific local context.

Lessons learned

With the ongoing iterative development of the GPSS's AI solution for the structural taxonomy classification of school buildings, some key lessons learned from the case study can be shared with task teams seeking AI solutions to address challenges in data collection for project preparation and implementation:

- Development of AI solutions supporting data collection is more an iterative learning-by-doing process than a one-time effort.

- Setting up a 'learning-by-doing' framework for iterative long-term development and improvement with experts in the loop to monitor the performance of the AI solution is essential for AI solutions to be adaptive and reliable.

- Consistency in the format and structure of relevant data collections is necessary to avoid fragmentation in the accumulation of training data and to allow an iterative and long-term improvement of the AI solution.

- It is important to identify any imbalance or bias in the dataset that needs to be addressed.

- Setting up a suitable evaluation metric is critical to guide the selection and improvement of the AI model.

- Accurately communicating the model's performance (in terms of capabilities and limitations) with stakeholders will facilitate the adoption of the solution.

Sample STC ToR for Developing an AI-based Tool for Infrastructure Baseline Data Collection

1. Background

With an estimated **875 million** children exposed to earthquakes and typhoons worldwide, there is an urgent need to provide safer and resilient school infrastructure for children to develop their full potential. To address this issue, the World Bank's Global Program for Safer Schools (GPSS) is supporting developing countries in the design and implementation of large-scale investments to improve the safety and resilience of school infrastructure and the quality of learning environments.

One of the biggest challenges of this program in developing countries is the lack of high-quality data about school building inventory and the absence of efficient mechanisms to update and manage this information. The data collection process to assess school infrastructure is commonly done through field inspections conducted by engineers, which are usually costly and time-consuming. Therefore, innovative and more efficient approaches to collect baseline data are essential to strengthen the capacity of developing countries in scaling up safer schools activities.

To facilitate the development of high-quality school infrastructure baselines at scale, the GPSS has developed a strong underlying tool—the Global Library of School Infrastructure (GLOSI). It provides a global catalog of classified structural typologies of school buildings with vulnerability curves and scalable vulnerability reduction solutions. This classification is performed at a building level and consists of parameters like the structural system, number of stories, irregularities of the building, among other structural parameters, which influence the performance of the building during an earthquake.

Along this line, building upon GLOSI, the GPSS will develop AI-based tools to facilitate more cost-efficient school infrastructure baseline data collection, with the aim of leveraging efforts and engagement from school communities contributing to the planning of risk reduction interventions.

2. Scope of tasks

The selected candidate will participate and contribute to activities such as:

- Development and improvement of ML/AI algorithm for classification of structural typology parameters for school buildings based on photos from existing data

- Preparation of school infrastructure baseline data for ML/AI

- Development and deployment of end-to-end data pipelines

- Integration of ML/AI model with existing ArcGIS dashboards

- Online research on image-based ML/AI methods/ tools as needed

- Familiarization with basic concepts such as risk, hazard, exposure, vulnerability, and disaster risk management

- Learning from technical discussions within task team and with partner teams on ongoing safer schools projects

- Support for the preparation of reports and presentations.

3. Activities and timeline

Activity		Week									
		1	2	3	4	5	6	7	8	9	10
1	Review current AI classifier prototype model and existing country data	▨	▨								
2	Set up data pipeline and cloud computing		▨	▨							
3	Prepare training data			▨							
4	Improve current AI model meeting latest data structure and train with latest data				▨	▨					
5	Review results and identify improvement needs with the task team					▨	▨				
4	Conduct necessary model improvements						▨	▨			
5	Integrate model with existing ArcGIS dashboard							▨	▨		
6	Deliver preliminary results									▨	
7	Task team final review and feedback									▨	
8	Deliver final AI model and integrated online dashboard										▨

4. Key competencies

- Bachelor's degree or Master's degree (preferred) in data science, computer science, or related field

- Proven experience with ML/AI algorithms and techniques, specifically CNNs using frameworks such as Tensorflow and Keras

- Minimum of one year experience with programming languages such as Python using data science libraries (NumPy, SciPy, Pandas, and so on) and analytics suites such as Jupyter Notebook

- Minimum of one year Extract Transform Load (ETL) or data pipeline engineering experience

- Minimum of one year experience in image processing

- Minimum of one year cloud computing experience

- Professional interest in utilizing disruptive technology to support international development

- Excellent analytical skills and a passion for learning

- Excellent communication skills in written and spoken English.

ANNEX 2

Sample Firm ToR for Developing an AI-based Tool for Infrastructure Baseline Data Collection

1. Background

With an estimated **875 million** children exposed to earthquakes and typhoons worldwide, there is an urgent need to provide safer and resilient school infrastructure for children to develop their full potential. To address this issue, the World Bank's Global Program for Safer Schools (GPSS) is supporting developing countries in the design and implementation of large-scale investments to improve the safety and resilience of school infrastructure and the quality of learning environments.

One of the biggest challenges of this program in developing countries is the lack of high-quality data about school building inventory and the absence of efficient mechanisms to update and manage this information. The data collection process to assess school infrastructure is commonly done through field inspections conducted by engineers, which are usually costly and time-consuming. Therefore, innovative and more efficient approaches to collect baseline data are essential to strengthen the capacity of developing countries in scaling up safer schools activities.

To facilitate the development of high-quality school infrastructure baselines at scale, the GPSS has developed a strong underlying tool—the Global Library of School Infrastructure (GLOSI). It provides a global catalog of classified structural typologies of school buildings with vulnerability curves and scalable vulnerability reduction solutions. This classification is performed at a building level and consists of parameters like the structural system, number of stories, irregularities of the building, among other structural parameters, which influence the performance of the building during an earthquake.

Along this line, building upon GLOSI, the GPSS will develop AI-based tools to facilitate more cost-efficient school infrastructure baseline data collection, with the aim of leveraging efforts and engagement from school communities contributing to the planning of risk reduction interventions.

2. Objective

The specific objectives of this work include the following:

- Develop an ML/AI model for the classification of two to five major structural typology parameters for school buildings based on multiple photos per building, with a satisfactory performance in discussion with the World Bank task team.

- Integrate the ML/AI model with the Survey123 mobile app and ArcGIS online dashboard or web application for deployment.

- Establish retraining and model updating mechanism for the ML/AI model from new photo data collected through the Survey123 mobile app.

3. Scope of work

The development of the ML/AI model should be based on the structural classification system defined by the GLOSI, compatible with existing GPSS data collection strategies and tools, and based on discussions with the GPSS team. Expected activities are discussed below.

3.1 Inception meeting

The meeting will be held by video conference with team members and key members of the World Bank task team, and they will discuss the project background, requirements, workplan, and deliverables.

3.2 Desktop review

- Review the existing prototype ML/AI model, existing photo data provided by the World Bank task team, and existing GPSS data collection mobile app and online dashboards.

- Identify data preparation needs in discussion with the World Bank task team.

- Identify security, compliance, and reporting requirements for implementation of AI/ML model in a production setting.

3.3 Data preparation

- Establish and implement data pipeline.

- Establish cloud computing resources for the data pipeline, AI/ML model, model monitoring, and necessary security controls.

- Develop any necessary databases or data repositories for model inputs and outputs and underlying geospatial data.

- Conduct necessary data cleaning, compiling, and structuring to prepare training data.

3.4 Model development

- Develop the ML/AI classification model based on the GLOSI structural classification system and photo data provided by the World Bank task team.

- Define model performance characteristics (for example, expectations for precision, accuracy, and recall) and process for monitoring, measuring, and reporting.

- Review model performance and identify improvement needs with the World Bank task team.

- Improve the ML/AI model performance to meet a satisfactory level for two to five key GLOSI classification parameters in discussion with the World Bank task team.

3.5 Model deployment

- Integrate the developed ML/AI model with the existing GPSS mobile app for data collection.

- Integrate the developed ML/AI model with the existing GPSS quality assurance and quality control online dashboard.

- Set up a feedback loop between the ML/AI model and the GPSS mobile app and dashboard for model retraining/updating upon new photos collected through the app.

- Support the World Bank task team on issues related to the deployment of the developed tools.

4. Deliverables

The deliverables for this project will be the following:

- An ML/AI classification model for two to five key GLOSI structural typology parameters for school buildings with satisfactory performance in discussion with the World Bank task team

- An improved GPSS mobile app for data collection with the ML/AI model integrated, with a feedback loop for model retraining/updating

3.6 Training and dissemination

- Deliver a training and dissemination workshop with the World Bank task team and relevant stakeholders on the use of the developed tools.

- An improved GPSS quality assurance and quality control dashboard with the ML/AI model integrated, with a feedback loop for model retraining/updating

- A report on the development of the model and tools, including but not limited to methodology, results, limitations, and recommendations.

5. Tentative time frame

Activity		Month						Payment terms
		1	2	3	4	5	6	
1	Inception meeting	X						10% upon commencement of the services
2	Desktop review	X						
3	Data preparation	X	X					
4	Model development		X	X				30% upon submission and the World Bank acceptance of the model
5	Model deployment				X	X		30% upon submission and the World Bank acceptance of the integrated tools
6	Training and dissemination						X	
7	Final report						X	30% upon submission and the World Bank acceptance of the final version of the report

6. Qualifications

General

The firm should be able to:

- Demonstrate experience in data science and software development, including the development of AI-based tools.

Specific qualifications

Team Leader

- Master's and/or doctorate degree in data science, computer science, or related field

- Minimum of five years of professional experience in data science and/or ML/AI development

- Experience with international organizations and/or governments

- Experience with Python or similar programming language

- Experience with ArcGIS tools (preferred)

- Experience with deploying ML/AI solutions, data pipelines, and web applications in cloud environments.

Project Manager

- Master's degree in data science, computer science, or related field

- Minimum of three years of professional experience in data science and/or ML/AI development

- Proven ability to work with international organizations, national governments, or nongovernmental organizations

- Experience in managing software development projects

- Experience with ArcGIS tools (preferred).

GPSS Global Program for Safer Schools

Contact information
gpss@worldbank.org

April 2021

For the Working Group on Remote Supervision established by SD VP