## Default Question Block

Thank you for completing this survey in support of our research study to identify mismatches in Machine Learning (ML)-enabled systems. The goal of this survey is to understand the types of ML mismatches that create barriers when deploying and sustaining ML-enabled systems. The target audience for the survey is practitioners who participate in the development, deployment, and operations of ML-enabled systems.

We define an **ML mismatch** as a problem that occurs in the development, deployment, and operation of an ML-enabled system due to **incorrect assumptions** that results in a negative consequence. ML mismatch can be traced back to information that could have been shared between stakeholders that would have avoided the problem.

In this survey, you will find two sections of questions

- The Demographics section contains generic information
- The ML Mismatch section contains sections for each mismatch category, and several sub-categories for each. The categories and subcategories were derived from the collective set of interviews we conducted with practitioners. Please rate how important you consider this type of information to be available in order to avoid mismatch.

This survey is voluntary. Feel free to stop at any time. All data collected will remain anonymous and stored in an access controlled area.

If you have read the information above and consent to taking this survey, please answer "Yes" below. If not, please answer "No" (this will exit the survey).

- ○ Yes
- ○ No

# Demographics Questions

Please fill in the demographics-related questions below.

Please select the type of organization you currently work in

- ○ Industry
- ○ Government
- ○ Academia/Research
- ○ Other

Please select your primary role/perspective on machine learning projects

- ○ Data Scientist
- ○ Software Engineer
- ○ Operations
- ○ Other

[Optional] Please select your secondary role/perspective on machine learning projects

- ○ Data Scientist
- ○ Software Engineer
- ○ Operations
- ○ Other

Please select your total years of professional work experience

- ○ 1-3
- ○ 4-7
- ○ 8-11
- ○ 12 or more

Please select your total years of machine learning-related experience

- ○ 1-3
- ○ 4-7
- ○ 8-11
- ○ 12 or more

# ML Mismatch Questions

In this section, please indicate how important it is to share (i.e., provide/receive) information related to each of the following categories in order to avoid mismatches during ML system development, deployment and operations.

### Development Environment

*Computing environment for model integration and testing.*

Please indicate how important it is to share (i.e., provide/receive) information about the Development Environment in order to avoid mismatches during ML system development, deployment and operations.

| | Not Important | Somewhat Important | Important | Very Important |
|---|---|---|---|---|
| Programming Language: Programming Language/ML Framework/Tools used in the development environment | ○ | ○ | ○ | ○ |
| Upstream and Downstream System Components: Specifications/APIs for how data comes in from upstream components and is fed to downstream components | ○ | ○ | ○ | ○ |

| | Not Important | Somewhat Important | Important | Very Important |
|---|---|---|---|---|
| Computing Resources: Computing resources available in the development environment, such as CPU, GPU, memory, and storage | ○ | ○ | ○ | ○ |
| Development and Integration Timelines: Development and integration timelines for integration of trained models into the larger system | ○ | ○ | ○ | ○ |

If we missed any topics that you consider important related to Development Environment above, please add them below.

**Operational Data**

*This category refers to data that is input to the trained model at serving time.*

Please indicate how important it is to share (i.e., provide/receive) information about Operational Data in order to avoid mismatches during ML system development, deployment and operations.

| | Not Important | Somewhat Important | Important | Very Important |
|---|---|---|---|---|
| Data Statistics: Operational data statistics, such as distribution and other metrics | ○ | ○ | ○ | ○ |
| Data Pipelines: Details on the implementation of data pipelines for the operational model | ○ | ○ | ○ | ○ |

| | Not Important | Somewhat Important | Important | Very Important |
|---|---|---|---|---|
| Data Sources: Sources for operational data for the operational model | ○ | ○ | ○ | ○ |
| Data Syntax and Semantics: Syntax and semantics of the data that constitutes the input for the operational model | ○ | ○ | ○ | ○ |
| Data Rates: Rates at which operational data feeds into the operational model | ○ | ○ | ○ | ○ |

If we missed any topics that you consider important related to Operational Data above, please add them below.

**Operational Environment**

*This category refers to the computing environment in which the model will be served.*

Please indicate how important it is to share (i.e., provide/receive) information about the Operational Environment in order to avoid mismatches during ML system development, deployment and operations.

| | Not Important | Somewhat Important | Important | Very Important |
|---|---|---|---|---|
| Runtime Metrics and Data: Runtime metrics, logs, model version, data, user feedback, etc. collected for troubleshooting, debugging, or retraining | ○ | ○ | ○ | ○ |

| | Not Important | Somewhat Important | Important | Very Important |
|---|---|---|---|---|
| Computing Resources: Computing resources available in the operational environment (e.g., GPU, memory, storage) | ○ | ○ | ○ | ○ |
| Required Model Inference Time: Required model inference time (e.g., time for the model to produce a result) | ○ | ○ | ○ | ○ |

If we missed any topics that you consider important related to Operational Environment above, please add them below.

**Raw Data**

*This category refers to data available for model training that has not been cleansed or pre-processed.*

Please indicate how important it is to share (i.e., provide/receive) information about Raw Data in order to avoid mismatches during ML system development, deployment and operations.

| | Not Important | Somewhat Important | Important | Very Important |
|---|---|---|---|---|
| Metadata: Metadata about raw data (e.g., collection details, distribution, geographic location, timeframes) | ○ | ○ | ○ | ○ |
| Data Dictionary: Description of data elements (e.g., field names, description, values, meaning of missing values) | ○ | ○ | ○ | ○ |

| | Not Important | Somewhat Important | Important | Very Important |
|---|---|---|---|---|
| Proxy Data: Process used to generate or acquire proxy data due to sensitivities, legal, or policy reasons | ○ | ○ | ○ | ○ |
| Restrictions: Data sensitivities that would result in prohibiting certain actions (e.g. upload to public cloud environments) | ○ | ○ | ○ | ○ |
| Anonymization: Process used to anonymize data due to PII or other constraints | ○ | ○ | ○ | ○ |

If we missed any topics that you consider important related to Raw Data above, please add them below.

**Task and Purpose**

*This category refers to high-level requirements and constraints for the model.*

Please indicate how important it is to share (i.e., provide/receive) information about Task and Purpose in order to avoid mismatches during ML system development, deployment and operations.

| | Not Important | Somewhat Important | Important | Very Important |
|---|---|---|---|---|
| Business Goals: Business goals or objectives that the model is going to help satisfy | ○ | ○ | ○ | ○ |
| Success criteria: Information used to determine if the model is performing correctly (e.g. success criteria, client expectations, validation scenarios, acceptance criteria) | ○ | ○ | ○ | ○ |

| | Not Important | Somewhat Important | Important | Very Important |
|---|---|---|---|---|
| Task: Task that model is expected to perform | ○ | ○ | ○ | ○ |
| Usage Context: How results of the model will be used by end users or in the context of a larger system | ○ | ○ | ○ | ○ |
| Data Rights and Policies: Known data rights, legal, privacy, and other policies that need to be met by model and data | ○ | ○ | ○ | ○ |

If we missed any topics that you consider important related to Task and Purpose above, please add them below.

**Trained Models**

*This category refers to models trained and ready for integration into a larger system.*

Please indicate how important it is to share (i.e., provide/receive) information about Trained Models in order to avoid mismatches during ML system development, deployment and operations.

| | Not Important | Somewhat Important | Important | Very Important |
|---|---|---|---|---|
| Test Cases and Data: Test Cases plus Test Data that can be used for integration testing | ○ | ○ | ○ | ○ |
| API/Specifications: Model APIs and specifications that provide greater detail into inputs, outputs, and internals | ○ | ○ | ○ | ○ |

| | Not Important | Somewhat Important | Important | Very Important |
|---|---|---|---|---|
| Decisions: Decisions, assumptions, limitations and constraints that have an effect on model integration and deployment | ○ | ○ | ○ | ○ |
| Model Output Interpretation: Information necessary to interpret model output, results or inferences | ○ | ○ | ○ | ○ |
| Programming Language: Programming Language, ML Framework, Tools and/or Libraries used to develop and train the model | ○ | ○ | ○ | ○ |
| Evaluation Metrics: Evaluation metrics and results of trained model evaluation (e.g., false positive rate, accuracy) | ○ | ○ | ○ | ○ |
| Versioning: Version information for trained model | ○ | ○ | ○ | ○ |
| System Configuration Requirements: System configuration requirements for trained model to execute (e.g., number of GPUs, libraries, tools, and dependencies) | ○ | ○ | ○ | ○ |
| Data Buffering/Window Requirements: Data buffering or time window requirements that would indicate that data has to be delivered in "chunks" instead of streamed | ○ | ○ | ○ | ○ |

If we missed any topics that you consider important related to Trained Models above, please add them below.

### Training Data

*This category refers to data that is pre-processed and ready for input into a model for training purposes.*

Please indicate how important it is to share (i.e., provide/receive) information about Training Data in order to avoid mismatches during ML system development, deployment and operations.

| | Not Important | Somewhat Important | Important | Very Important |
|---|---|---|---|---|
| Data Preparation Pipelines: Details of data preparation pipelines to derive training data from raw data | ○ | ○ | ○ | ○ |
| Data Statistics: Training data statistics, such as distribution and other metrics | ○ | ○ | ○ | ○ |
| Versioning: Version information for training data | ○ | ○ | ○ | ○ |

If we missed any topics that you consider important related to Trained Data above, please add them below.

## Additional Questions

## General Feedback:

Please share an example of an issue you faced due to one of these ML-mismatch categories:

If you have any other comments please let us know: