# Automating Mismatch Detection in Machine-Learning (ML) Systems Using TEC

Organizations increasingly use machine learning (ML) to implement advanced capabilities in many software systems, but they often struggle to integrate ML components into production systems. These struggles usually arise because the development of ML-enabled systems involves activities that are performed by separate teams with different workflows and perspectives. Difficulty aligning the work of teams results in **ML mismatch**—a problem that occurs in the development, deployment, and operation of an ML-enabled system when stakeholders make incorrect assumptions about system elements, resulting in a negative consequence.

For example, it is typical for data scientists or ML engineers to develop the model, for software engineers to test it and integrate it into the system, and for operations staff to move it into production and set up monitoring during operations. If these teams don't communicate well or hold different assumptions about different elements of the ML-enabled system, organizations might run into problems, including delays, rework, and failures in developing, deploying, and evolving the ML system. Even when teams communicate effectively during development, the production environment can often change, resulting in further mismatch if teams don't detect these changes in a timely manner.

The Software Engineering Institute (SEI) developed the **TEC ML Mismatch Detection Tool** to help organizations improve how they integrate ML components into software systems by better aligning the work of ML stakeholders and development teams throughout the ML model and system development process.

TEC—which comes from the word "de-TEC-tive"—can help your organization discover any sources of possible mismatch in the expectations or assumptions of system stakeholders. Discovering ML mismatch early in model and system development can help correct it when it is still relatively easy and cost-effective to do so.

## How TEC Helps Avoid Mismatch

The SEI identified eight categories of ML mismatch by conducting an empirical study, gathering feedback from practitioners, and providing organizations with ongoing assistance to engineer ML-enabled systems. The categories resulted in a formal set of machine-readable descriptors to define the system attributes that teams must specify to avoid ML mismatch; these include the following (see the figure on the next page for more detail):

- system context
- raw data
- training data
- data pipeline
- trained model
- development environment
- production environment
- production data

Using these descriptors, development teams and system stakeholders can enter information about their part of the system, or they can import or export them for easy sharing if your organization is working with a contractor. TEC then makes this information visible to all stakeholders. After stakeholders enter their information, TEC examines it to detect mismatch or missing information so that the stakeholders can resolve issues before they turn into larger problems.

## Results

TEC helps ensure that ML-enabled systems meet mission and business goals, and it brings all stakeholders involved in ML-enabled systems development onto the same page. Using TEC, system developers and operations staff better understand how new ML components will work as part of the larger system.

### Organizations that use TEC can

- identify ML mismatch
- improve communication and align workflows
- ease ML component integration
- save time and money due to avoided rework
- gather requirements more effectively
- improve collaboration among stakeholders and reach agreement more easily

### Learn More About ML Mismatch

For an overview of TEC, read the paper "Characterizing and Detecting Mismatch in Machine-Learning-Enabled Systems."

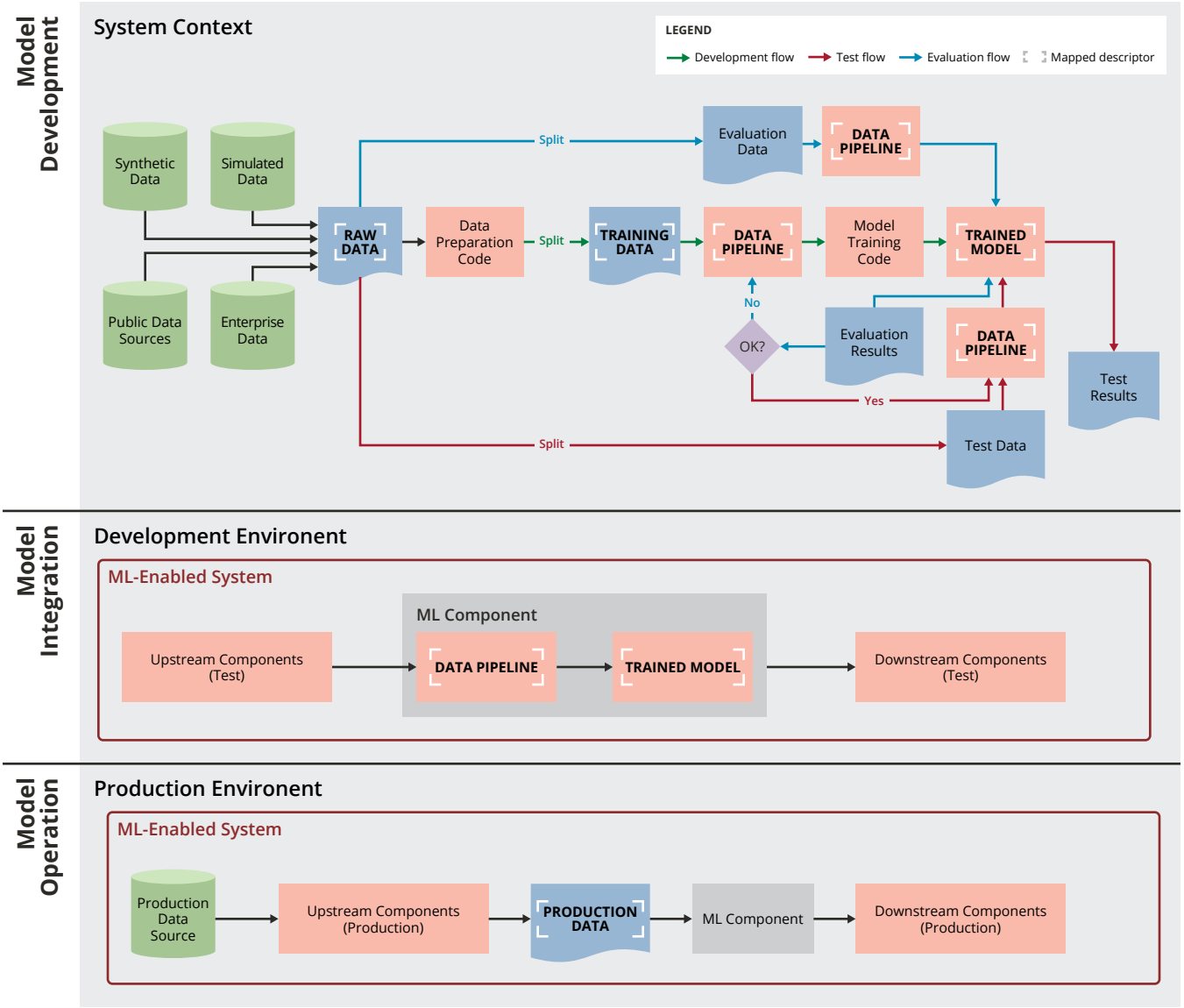
### Download TEC

The SEI can help you improve your engineering of ML-enabled systems by instantiating TEC in your organizations and integrating it into your ML-enabled systems development processes.

TEC is freely available for download at the following site: **github.com/cmu-sei/TEC**.

You can also download the ML mismatch descriptors from the SEI's GitHub site at the following location: **github.com/cmu-sei/ml-mismatch-descriptors**

*Mapping of Descriptors to Elements of the ML-Enabled Systems development workflows.*




## About the SEI

The Software Engineering Institute (SEI) advances software as a strategic advantage for national security. We serve the nation as a federally funded research anddevelopment center (FFRDC) sponsored by the U.S. Department of Defense (DoD).

## Contact Us


sei.cmu.edu
412.268.5800 | 888.201.4479
info@sei.cmu.edu