Process Mining and Intelligence Project  
Emotion Based Music Selection

Ettore Ricci

Francesco Boldrini

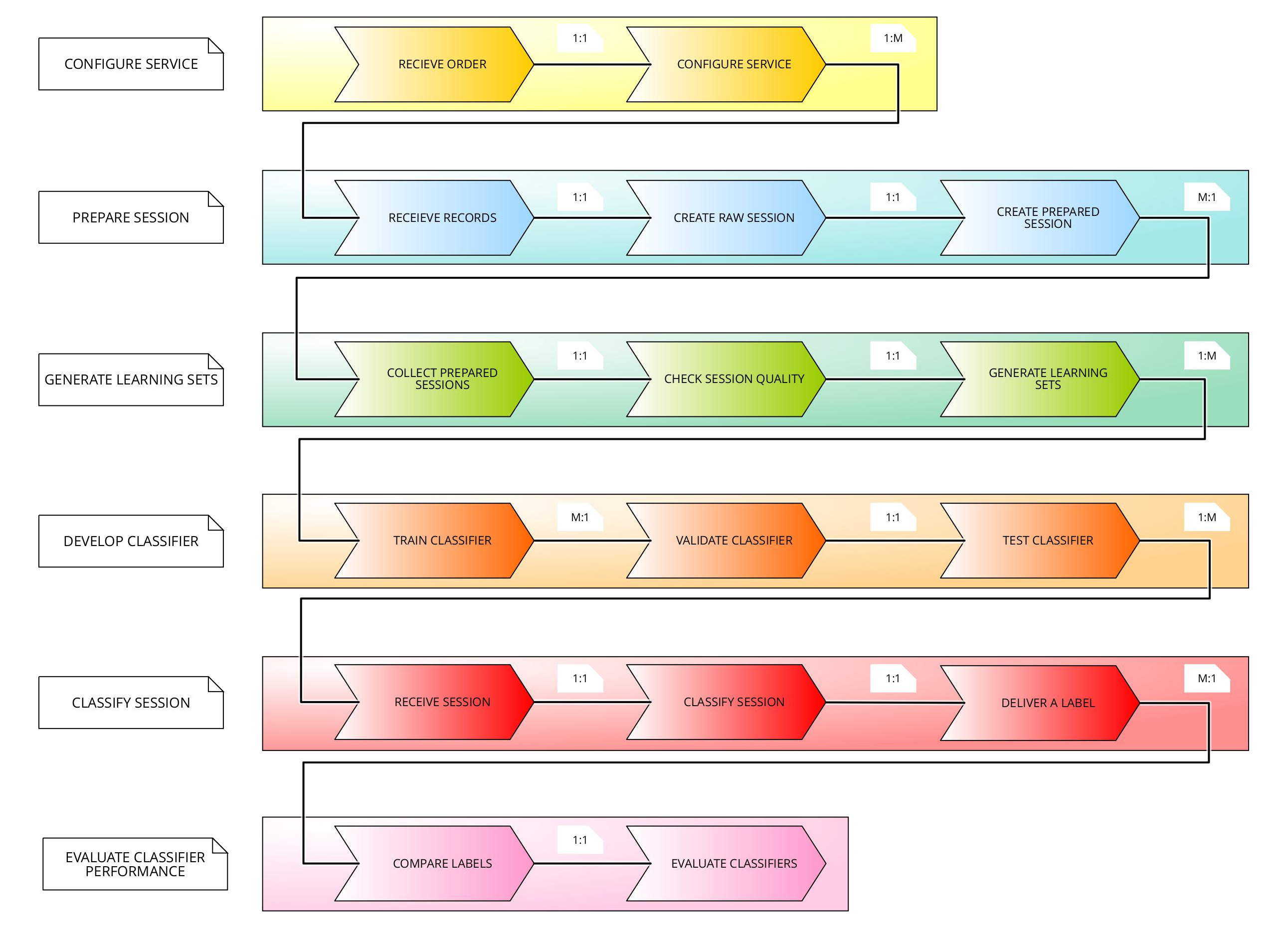
Paolo Palumbo

Zahra Omrani

# BPMN modeling

## Process landscape

[*Ettore Ricci, Paolo Palumbo, Francesco Boldrini, Zahra Omrani*]

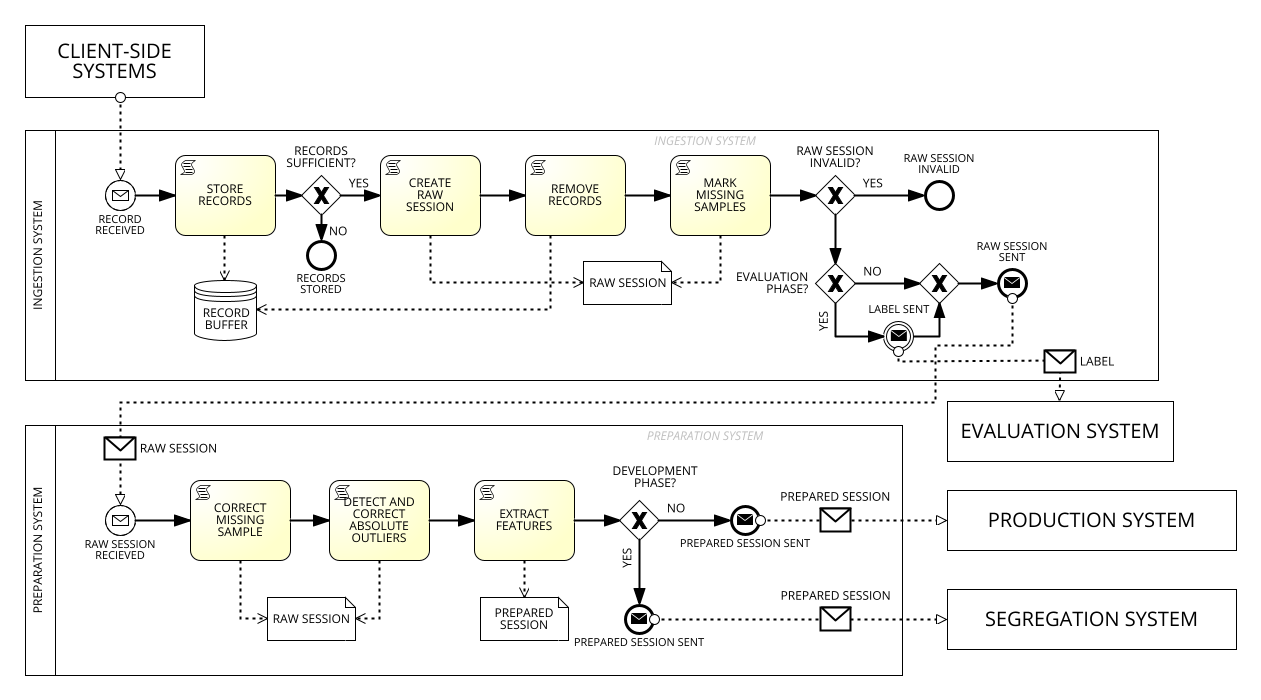


Process landscape

## Process model

### Prepare session

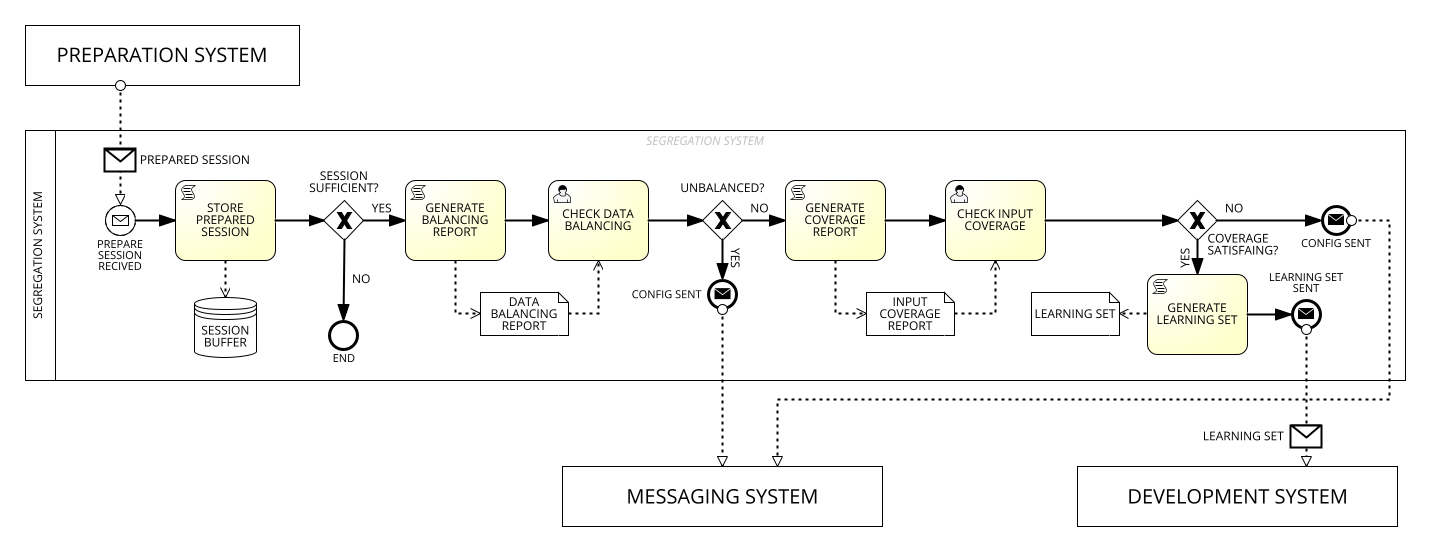
[*Ettore Ricci, Paolo Palumbo*]



Business Diagram of the "Prepare session" process

### Generate learning sets

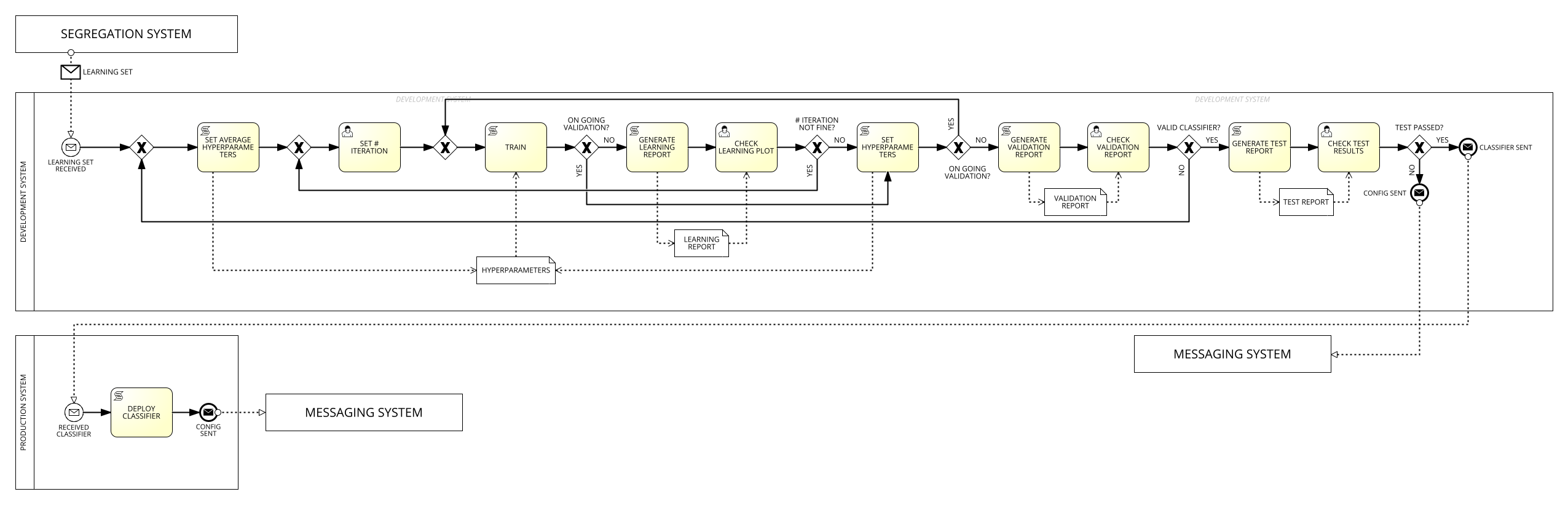
[*Ettore Ricci, Paolo Palumbo*]



Business Diagram of the "Generate learning sets" process

### Develop classifier

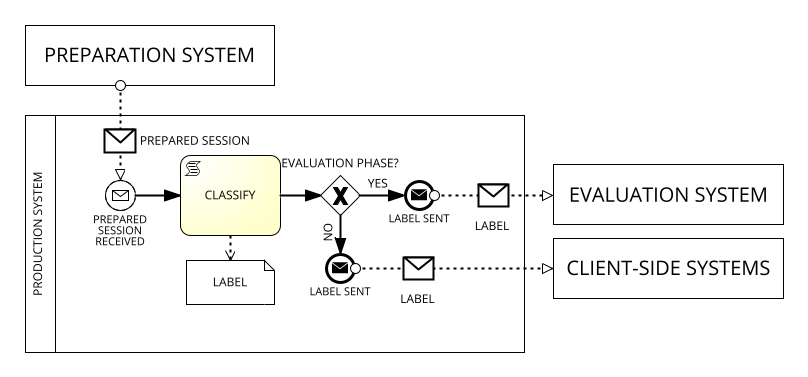
[*Ettore Ricci, Paolo Palumbo*]



Business Diagram of the "Develop classifier" process

### Classify session

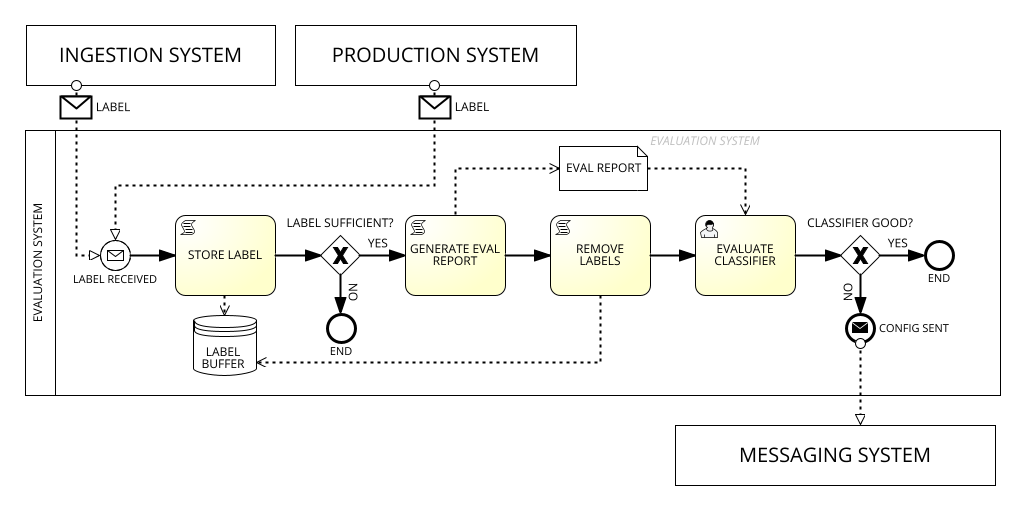
[*Ettore Ricci, Paolo Palumbo*]



Business Diagram of the "Classify session" process

### Evaluate classifier performance

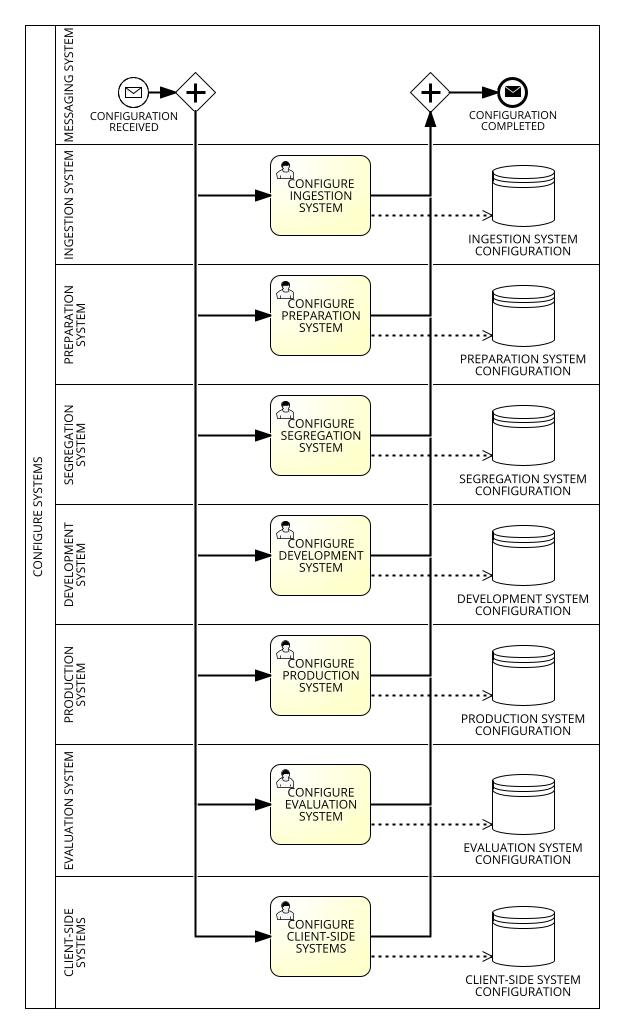
[*Ettore Ricci, Paolo Palumbo*]



Business Diagram of the "Evaluate classifier performance" process

### Configure systems

[*Ettore Ricci, Paolo Palumbo*]



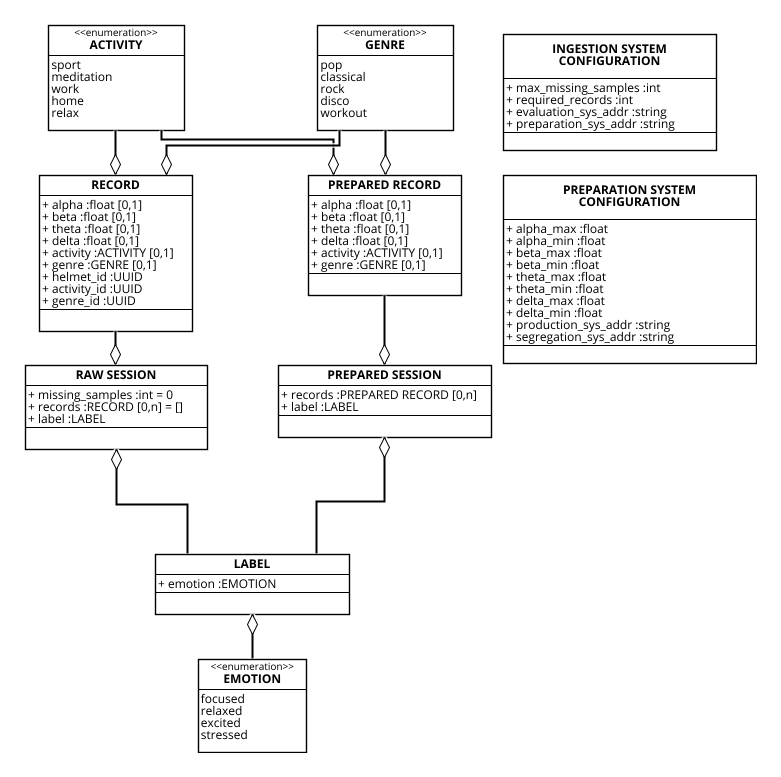
Business Diagram of the "Configure systems" process

# Data modeling

## Process model

### Prepare session

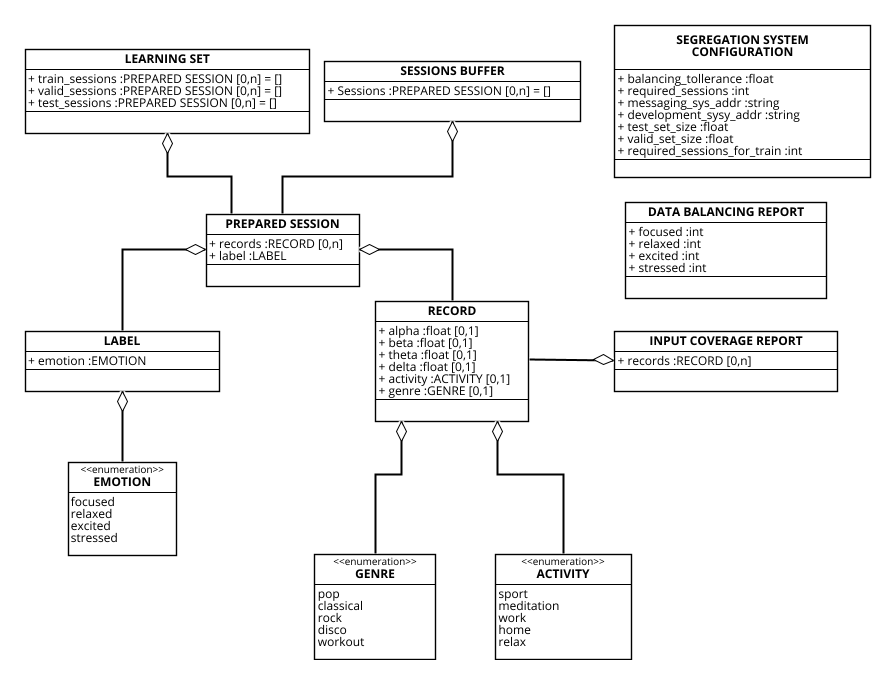
[*Ettore Ricci*]



Data Model of the "Prepare session" process

### Generate learning sets

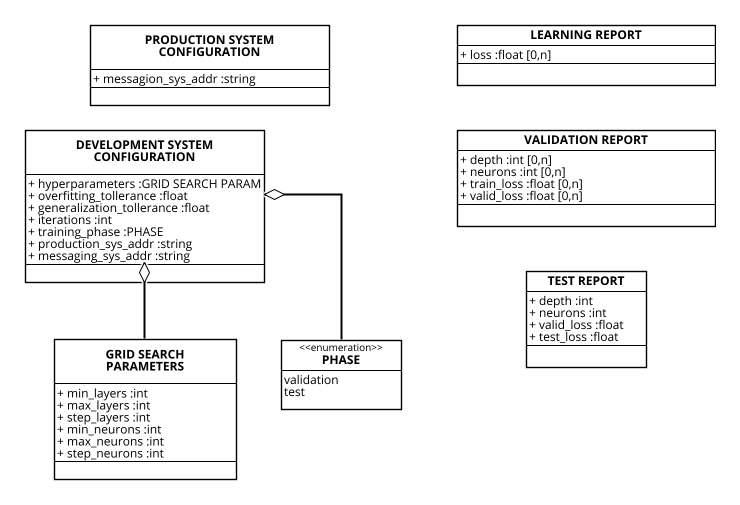
[*Paolo Palumbo*]



Data Model of the "Generate learning sets" process

### Develop classifier

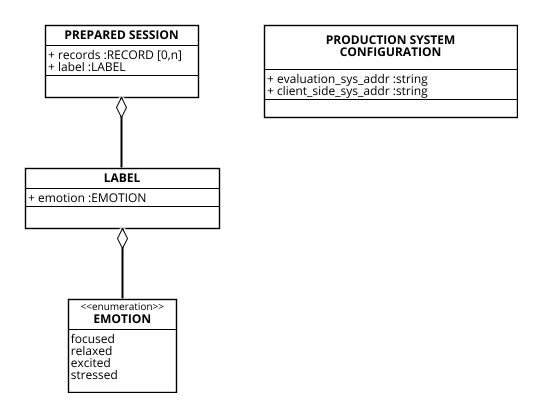
[*Paolo Palumbo*]



Data Model of the "Develop classifier" process

### Classify session

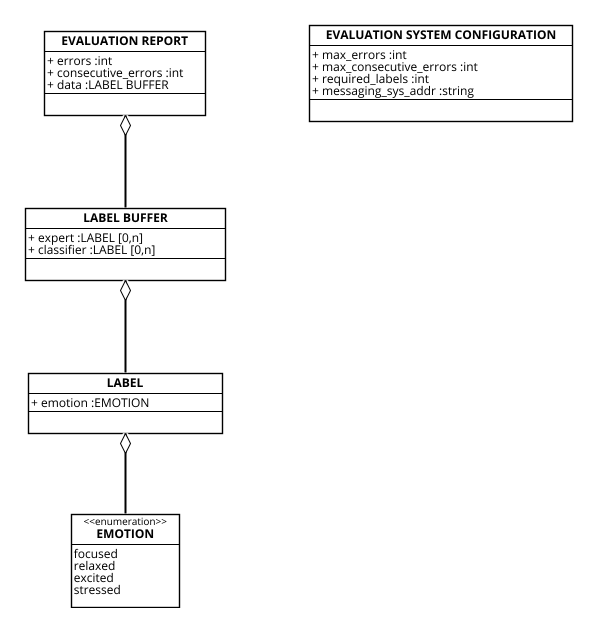
[*Francesco Boldrini*]



Data Model of the "Classify session" process

### Evaluate classifier performance

[*Zahra Omrani*]



Data Model of the "Evaluate classifier performance" process

# Task level modeling

## Roles and salaries

[*Ettore Ricci, Paolo Palumbo*]

Salary and normalized salary for each position

| **Position** | **Description** | **Salary** | **Normalized Salary** |
| --- | --- | --- | --- |
| Clerk | Handles administrative tasks, organizes documentation, and assists with data entry and labeling. Ensures smooth operations by coordinating communication and managing resources. | $52,000.00 | 1.00 |
| Data analyst | Prepares, analyzes, and visualizes data to extract insights. Collaborates on cleaning datasets, identifying trends, and supporting model validation. | $60,000.00 | 1.15 |
| ML engineer | Builds, tests, and deploys machine learning models, optimizing performance and scalability. Integrates AI solutions into production systems with a focus on efficiency. | $130,000.00 | 2.50 |
| Data scientist | Designs and experiments with AI models, applying advanced techniques to solve project challenges. Collaborates with experts to integrate domain knowledge and refine outputs. | $123,000.00 | 2.37 |
| Domain expert (Neurologist) | Provides medical expertise to guide AI development and validate results. Ensures solutions align with clinical standards and address neurological challenges. | $267,000.00 | 5.13 |
| **Minimum** | | $52,000.00 | 1.00 |

## Segregation system

### Check data balancing

[*Ettore Ricci, Paolo Palumbo*]

The task is performed by a Data Analyst.



"Check data balancing" mock-up form

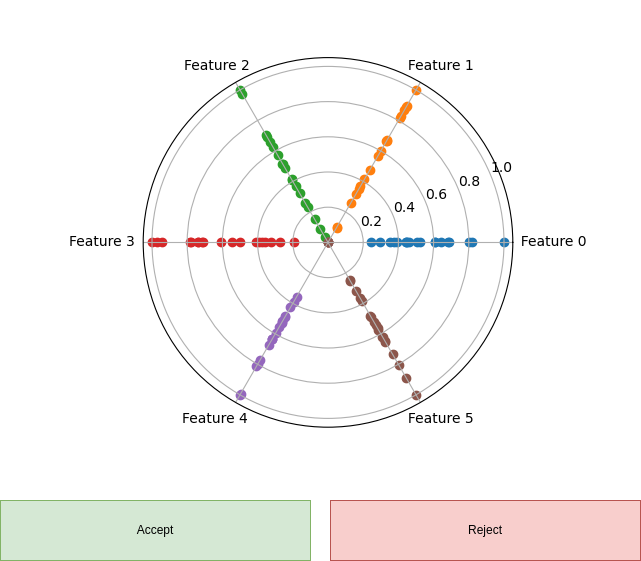
Detailed use case for "Check data balancing" task  
O - Occurrence, CL - Cognitive Level, S - Normalized Salary, SC - Step Cost

| **Step** | **O** | **CL** | **S** | **SC** |
| --- | --- | --- | --- | --- |
| **1** **ACTOR** opens "Check data balancing" form. | 1 | 1 | 1.15 | 1.15 |
| **2** **SYSTEM** shows the report. |  |  |  |  |
| **3** **SYSTEM** shows a hint whether the data is balanced or not. |  |  |  |  |
| **4** **ACTOR** checks the hint to see if the data is balanced or not. | 1 | 2 | 1.15 | 2.30 |
| **5.1** **IF** the data is balanced. | 0.2 |  |  |  |
| **5.1.1** **ACTOR** clicks "Balanced" button. | 0.2 | 1 | 1.15 | 0.23 |
| **5.2** **ELSE** | 0.8 |  |  |  |
| **5.2.1** **ACTOR** clicks "Unbalanced" button. | 0.8 | 1 | 1.15 | 0.92 |
| **7** **SYSTEM** shows a confirmation dialog. |  |  |  |  |
| **8** **ACTOR** closes the form. | 1 | 1 | 1.15 | 1.15 |
| Human task cost | | | | 5.74 |

### Check input coverage

[*Ettore Ricci, Paolo Palumbo*]

The task is performed by a Data Analyst.



"Check input coverage" mock-up form

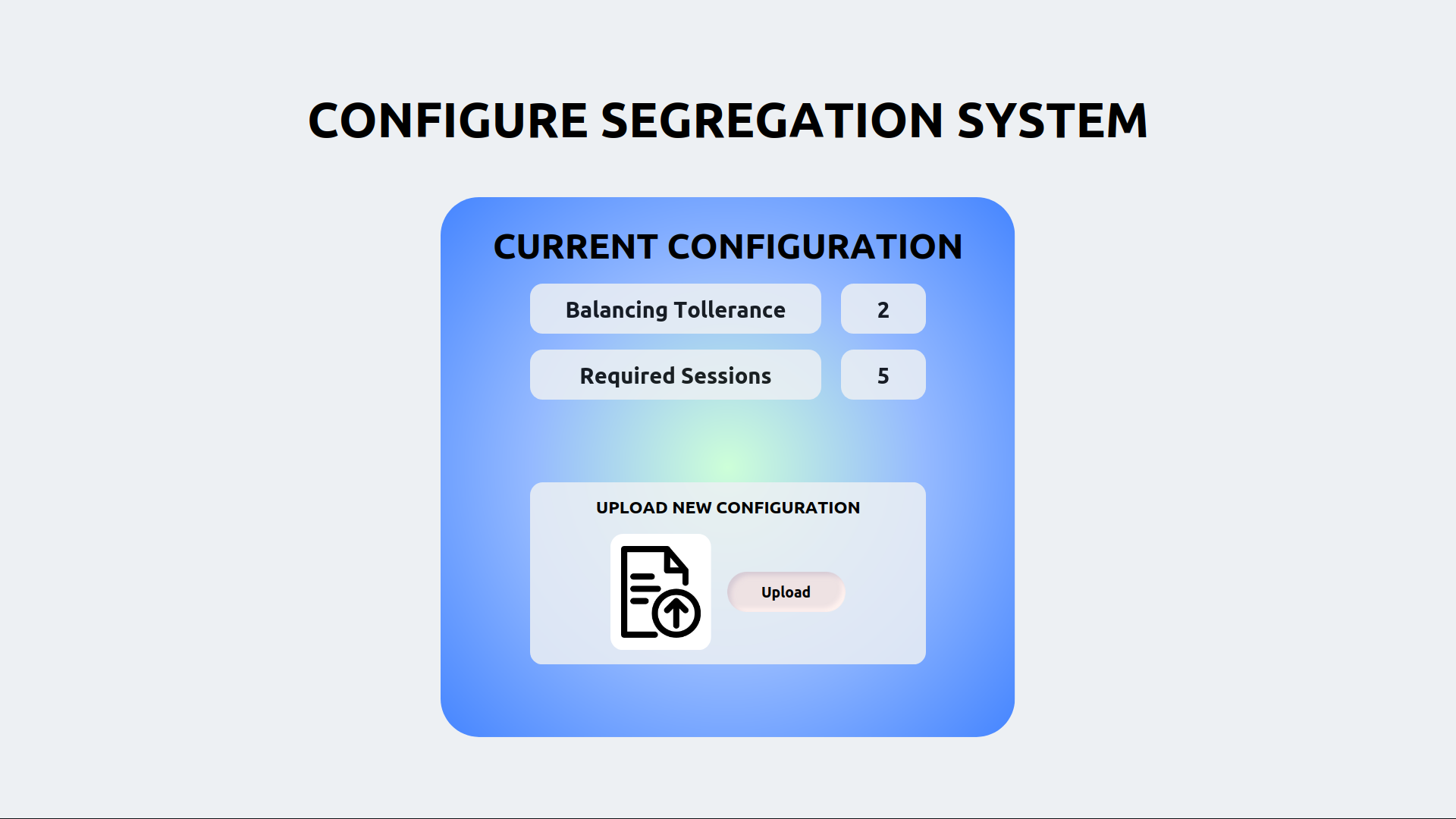
Detailed use case for "Check input coverage" task  
O - Occurrence, CL - Cognitive Level, S - Normalized Salary, SC - Step Cost

| **Step** | **O** | **CL** | **S** | **SC** |
| --- | --- | --- | --- | --- |
| **1** **ACTOR** opens "Check input coverage" form. | 1 | 1 | 1.15 | 1.15 |
| **2** **SYSTEM** shows a radar scatter plot of the input distribution. |  |  |  |  |
| **3** **FOR EACH** radius in the radar scatter plot: | 6 |  |  |  |
| **3.1** **ACTOR** checks if the distribution is uniform on the radius. | 6 | 4 | 1.15 | 27.6 |
| **3.1.1** **IF** the distribution is not uniform as expected. | 4 |  |  |  |
| **3.1.1.1** **THEN** the input coverage is not satisfied. | 4 |  |  |  |
| **4.1** **IF** the input coverage is satisfied. | 0.33 |  |  |  |
| **4.1.1** **ACTOR** clicks "Accept" button. | 0.33 | 1 | 1.15 | 0.38 |
| **4.2** **ELSE** | 0.66 |  |  |  |
| **4.2.1** **ACTOR** clicks "Reject" button. | 0.66 | 1 | 1.15 | 0.76 |
| **5** **SYSTEM** shows a confirmation dialog. |  |  |  |  |
| **6** **ACTOR** closes the form. | 1 | 1 | 1.15 | 1.15 |
| Human task cost | | | | 31.04 |

### Configure Segregation System

[*Francesco Boldrini, Zahra Omrani*]

This task is performed by a ML Engineer.



"Configure Segregation System" mock-up form

Detailed use case for "Configure Segregation" task  
O - Occurrence, CL - Cognitive Level, S - Normalized Salary, SC - Step Cost

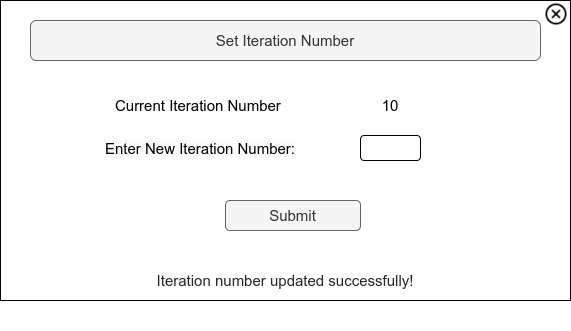
| **Step** | **O** | **CL** | **S** | **SC** |
| --- | --- | --- | --- | --- |
| **1** **ACTOR** opens the "Configure Segregation System" form. | 1 | 1 | 2.50 | 2.50 |
| **2** **SYSTEM** displays current configuration and "Upload" button. |  |  |  |  |
| **3** **ACTOR** checks parameters against previous iterations on file | 1 | 3 | 2.50 | 7.50 |
| **4** **ACTOR** adjusts file based on current parameters | 1 | 3 | 2.50 | 7.50 |
| **5** **ACTOR** pushes "Upload" button and uploads configuration file | 1 | 1 | 2.50 | 2.50 |
| **6.1** **SYSTEM** IF config is correct and correctly formatted |  |  |  |  |
| **6.1.1** **SYSTEM** shows a confirmation message. |  |  |  |  |
| **6.2** **ELSE** |  |  |  |  |
| **6.2.1** **SYSTEM** shows error message and aborts |  |  |  |  |
| **7** **ACTOR** closes the form. | 1 | 1 | 2.50 | 2.50 |
| Human task cost | | | | 22.50 |

## Development system

### Set iteration number

[*Zahra Omrani*]

The task is performed by a ML engineer.



"Set iteration number" mock-up form

Detailed use case for "Set iteration number" task  
O - Occurrence, CL - Cognitive Level, S - Normalized Salary, SC - Step Cost

| **Step** | **O** | **CL** | **S** | **SC** |
| --- | --- | --- | --- | --- |
| **1** **ACTOR** opens "Set Iteration Number" form. | 1 | 1 | 2.5 | 2.5 |
| **2** **SYSTEM** displays the current iteration number. |  |  |  |  |
| **3.1** **IF** it’s the first configuration: |  |  |  |  |
| **3.1.1** **ACTOR** inputs the desired number of iterations based on task complexity and previous experience. | 0.2 | 3 | 2.5 | 1.5 |
| **3.2** **ELSE** (subsequent configurations): |  |  |  |  |
| **3.2.1** **ACTOR** inputs the number based on the established learning curve. | 0.8 | 1 | 2.5 | 2 |
| **4** **ACTOR** clicks "Submit" button to confirm the iteration number. | 1 | 1 | 2.5 | 2.5 |
| **5** **SYSTEM** shows a confirmation dialog. |  |  |  |  |
| **6** **ACTOR** closes the form. | 1 | 1 | 2.5 | 2.5 |
| Human task cost | | | | 11 |

### Check learning report

[*Paolo Palumbo*]

The task is performed by a ML engineer.



"Check learning report" mock-up form

Detailed use case for "Check training report" task  
O - Occurrence, CL - Cognitive Level, S - Normalized Salary, SC - Step Cost

| **Step** | **O** | **CL** | **S** | **SC** |
| --- | --- | --- | --- | --- |
| **1** **ACTOR** opens "Check training report" form. | 1 | 1 | 2.50 | 2.50 |
| **2** **SYSTEM** shows the training loss curve. |  |  |  |  |
| **3** **ACTOR** checks the learning curve. | 1 | 3 | 2.50 | 7.50 |
| **3.1** **IF** the loss is flat for at least half of the iterations: | 0.4 |  |  |  |
| **3.1.1** **THEN** **ACTOR** clicks "Overfit" button. | 0.4 | 1 | 2.50 | 1.00 |
| **3.2** **IF** the loss is not flat at the end of the iterations: | 0.4 |  |  |  |
| **3.2.1** **THEN** **ACTOR** clicks "Underfit" button. | 0.4 | 1 | 2.50 | 1.00 |
| **3.3** **ELSE** | 0.2 |  |  |  |
| **3.3.1** **ACTOR** clicks "Approved" button. | 0.2 | 1 | 2.50 | 0.50 |
| **4** **SYSTEM** shows a confirmation dialog. |  |  |  |  |
| **5** **ACTOR** closes the form. | 1 | 1 | 2.50 | 2.50 |
| Human task cost | | | | 15 |

### Check validation report

[*Ettore Ricci*]

This task is performed by a ML engineer.



"Check validation report" mock-up form

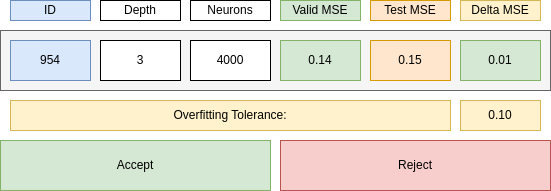
Detailed use case for "Check validation report" task  
O - Occurrence, CL - Cognitive Level, S - Normalized Salary, SC - Step Cost

| **Step** | **O** | **CL** | **S** | **SC** |
| --- | --- | --- | --- | --- |
| **1** **ACTOR** opens "Check validation report" form. | 1 | 1 | 2.5 | 2.5 |
| **2** **SYSTEM** shows the best 5 models sorted by increasing Validation Loss. |  |  |  |  |
| **3** **FOR EACH** model in the list: | 5 |  |  |  |
| **3.1** **IF** the model Validation Loss minus the Training Loss is less than the Overfitting Tolerance and the Best Model is not selected. | 1 | 2 | 2.5 | 5 |
| **3.1.1** **THEN** select the model as the Best Model. | 1 | 1 | 2.5 | 2.5 |
| **4** **FOR EACH** model in the list: | 4 |  |  |  |
| **4.1** **IF** the model is not the Best Model and the Validation Loss minus the Training Loss is less than the Overfitting Tolerance and the Second Best Model is not selected. | 1 | 2 | 2.5 | 5 |
| **4.1.1** **THEN** select the model as the Second Best Model. | 1 | 1 | 2.5 | 2.5 |
| **5.1** **IF** the Best Model is not selected. | 0.05 | 1 | 2.5 | 0.125 |
| **5.1.1** **ACTOR** clicks "Reject" button. | 0.05 | 1 | 2.5 | 0.125 |
| **5.2** **ELSE IF** the Second Best Model is not selected or the Validation Loss of the Second Best Model is one order of magnitude greater than the Validation Loss of the Best Model. | 0.3 | 3 | 2.5 | 2.25 |
| **5.2.1** **ACTOR** clicks on the Best Model. | 0.3 | 1 | 2.5 | 0.75 |
| **5.3** **ELSE** | 0.65 | 3 | 2.5 | 4.875 |
| **5.3.1** **ACTOR** clicks on the least complex model among the Best Model and the Second Best Model. | 0.65 | 3 | 2.5 | 4.875 |
| **6** **SYSTEM** shows a confirmation dialog. |  |  |  |  |
| **7** **ACTOR** closes the form. | 1 | 1 | 2.5 | 2.5 |
| Human task cost | | | | 32.91 |

### Check test results

[*Ettore Ricci*]

This task is performed by a ML engineer.



"Check test results" mock-up form

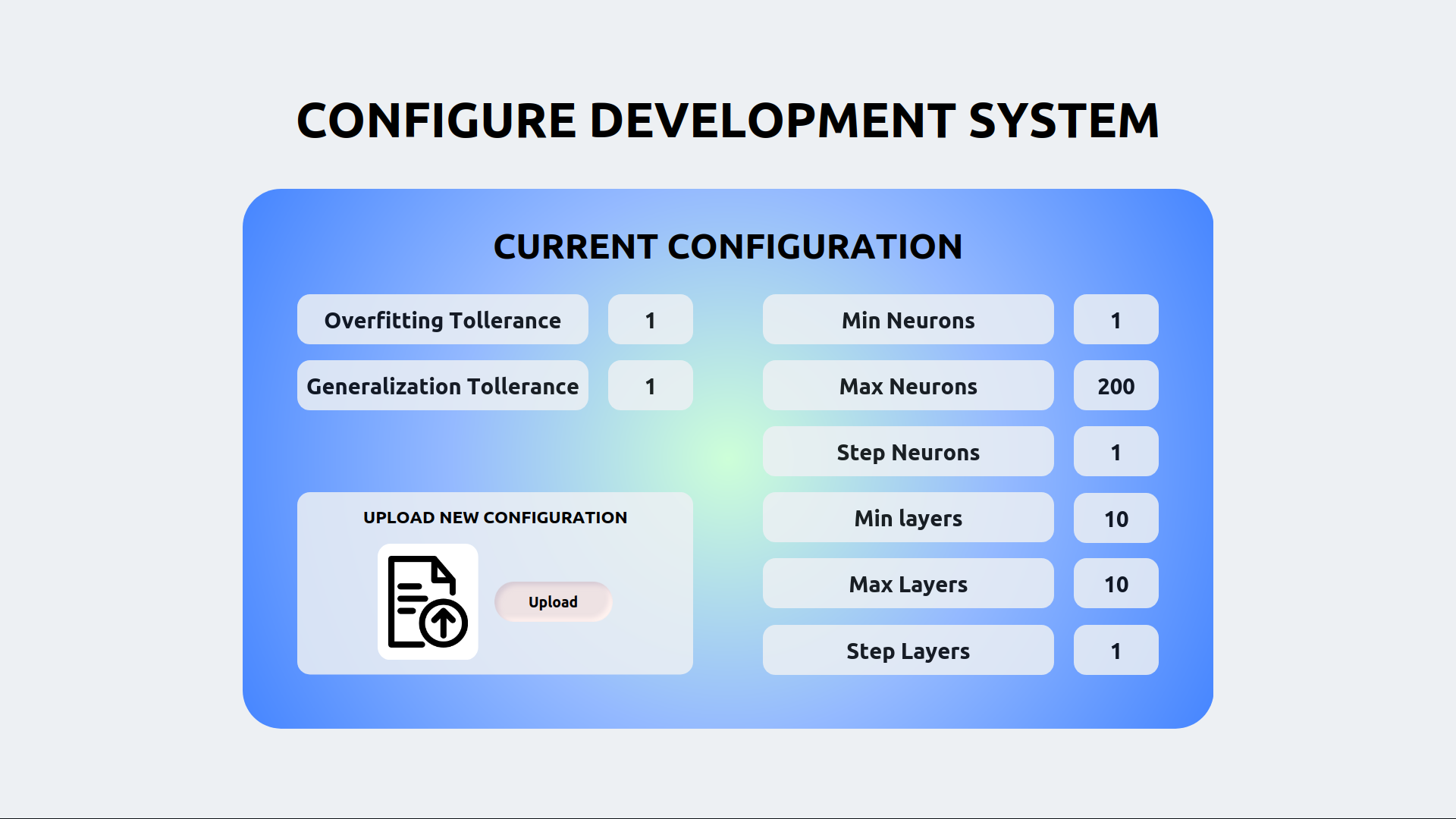
Detailed use case for "Check test results" task  
O - Occurrence, CL - Cognitive Level, S - Normalized Salary, SC - Step Cost

| **Step** | **O** | **CL** | **S** | **SC** |
| --- | --- | --- | --- | --- |
| **1** **ACTOR** opens "Check test results" form. | 1 | 1 | 2.5 | 2.5 |
| **2** **SYSTEM** shows the test results. |  |  |  |  |
| **3** **ACTOR** checks if the difference between the test results and the validation results is within overfitting tolerance. | 1 | 2 | 2.5 | 5 |
| **4.1** **IF** the test results is not satisfactory. | 0.01 |  |  |  |
| **4.1.1** **ACTOR** clicks "Reject" button. | 0.01 | 1 | 2.5 | 0.025 |
| **4.2** **ELSE** | 0.99 |  |  |  |
| **4.2.1** **ACTOR** clicks "Approve" button. | 0.99 | 1 | 2.5 | 2.475 |
| **5** **SYSTEM** shows a confirmation dialog. |  |  |  |  |
| **6** **ACTOR** closes the form. | 1 | 1 | 2.5 | 2.5 |
| Human task cost | | | | 12.5 |

### Configure Development System

[*Francesco Boldrini, Zahra Omrani*]

This task is performed by a ML Engineer.



"Configure Development System" mock-up form

Detailed use case for "Configure Development" task  
O - Occurrence, CL - Cognitive Level, S - Normalized Salary, SC - Step Cost

| **Step** | **O** | **CL** | **S** | **SC** |
| --- | --- | --- | --- | --- |
| **1** **ACTOR** opens the "Configure Development System" form. | 1 | 1 | 2.50 | 2.50 |
| **2** **SYSTEM** displays current configuration and "Upload" button. |  |  |  |  |
| **3** **ACTOR** checks parameters against previous iterations on file | 1 | 3 | 2.50 | 7.50 |
| **4** **ACTOR** adjusts file based on current parameters | 1 | 3 | 2.50 | 7.50 |
| **5** **ACTOR** pushes "Upload" button and uploads configuration file | 1 | 1 | 2.50 | 2.50 |
| **6.1** **SYSTEM** IF config is correct and correctly formatted |  |  |  |  |
| **6.1.1** **SYSTEM** shows a confirmation message. |  |  |  |  |
| **6.2** **ELSE** |  |  |  |  |
| **6.2.1** **SYSTEM** shows error message and aborts |  |  |  |  |
| **7** **ACTOR** closes the form. | 1 | 1 | 2.50 | 2.50 |
| Human task cost | | | | 22.50 |

## Evaluation system

### Evaluate classifier performance

[*Zahra Omrani*]

This task is performed by a Data Analyst.



"Evaluate Classifier Performance" mock-up form

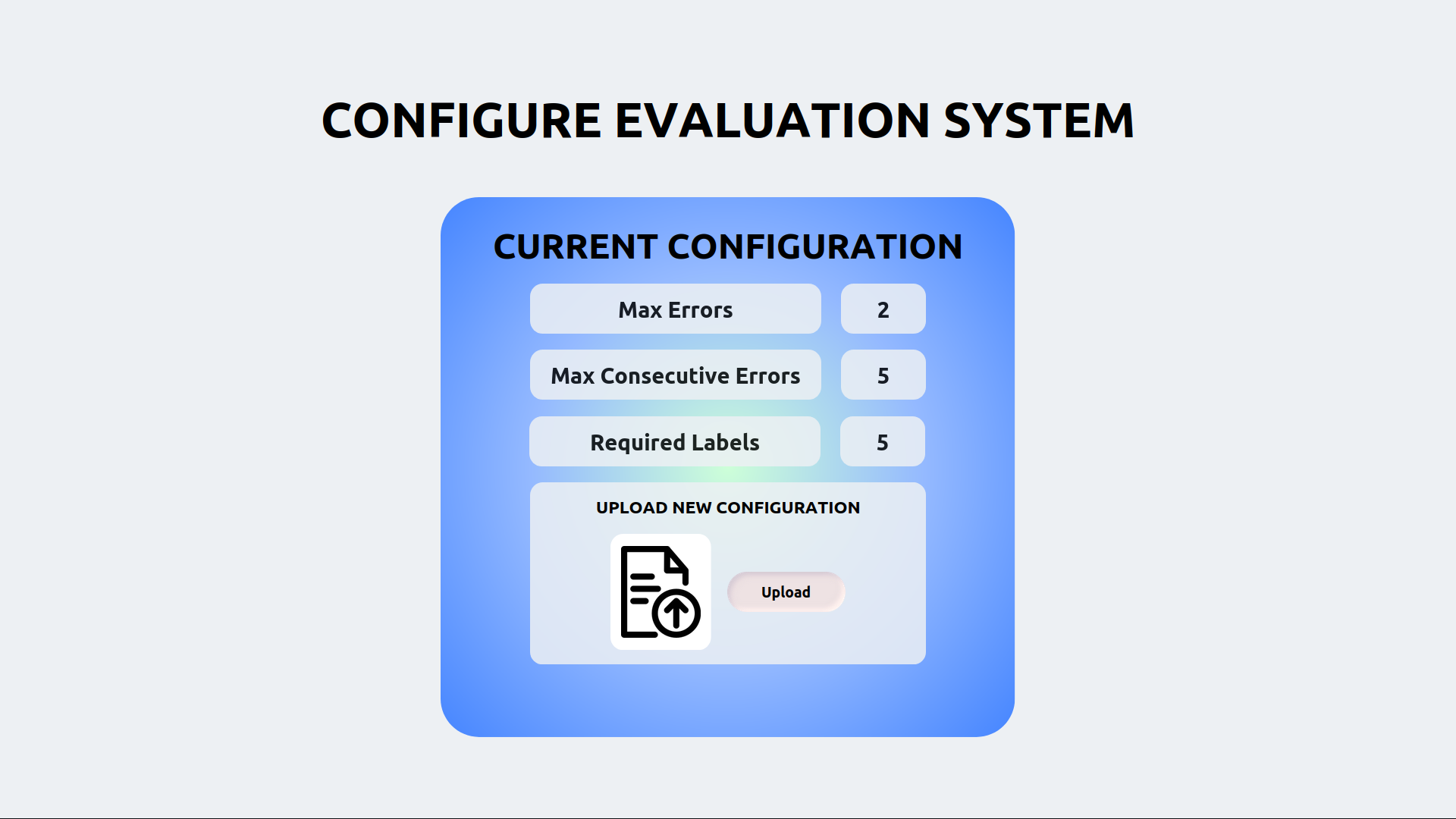
Detailed use case for "Evaluate Classifier Performance" task  
O - Occurrence, CL - Cognitive Level, S - Normalized Salary, SC - Step Cost

| **Step** | **O** | **CL** | **S** | **SC** |
| --- | --- | --- | --- | --- |
| **1** **ACTOR** opens the "Evaluate Classifier Performance" form. | 1 | 1 | 1.15 | 1.15 |
| **2** **SYSTEM** displays a table of sessions with Expert Label (ground truth) and Classifier Label (predicted label). The difference between the labels (if any) represents an error. |  |  |  |  |
| **3.1** **ACTOR** checks the total errors threshold color. | 1 | 2 | 1.15 | 2.30 |
| **3.2** **ACTOR** checks the consecutive errors threshold color | 1 | 2 | 1.15 | 2.30 |
| **3.3** **IF** at least one threshold is red |  |  |  |  |
| **3.3.1** **ACTOR** clicks the "Fail" button. | 0.4 | 1 | 1.15 | 0.46 |
| **3.4** **ELSE** |  |  |  |  |
| **3.4.1** **ACTOR** clicks the "Pass" button. | 0.6 | 1 | 1.15 | 0.65 |
| **4** **SYSTEM** shows a confirmation dialog. |  |  |  |  |
| **5** **ACTOR** closes the form. | 1 | 1 | 1.15 | 1.15 |
| Human task cost | | | | 8.05 |

### Configure Evaluation System

[*Francesco Boldrini, Zahra Omrani*]

This task is performed by a ML Engineer.



"Configure Evaluation System" mock-up form

Detailed use case for "Configure Evaluation" task  
O - Occurrence, CL - Cognitive Level, S - Normalized Salary, SC - Step Cost

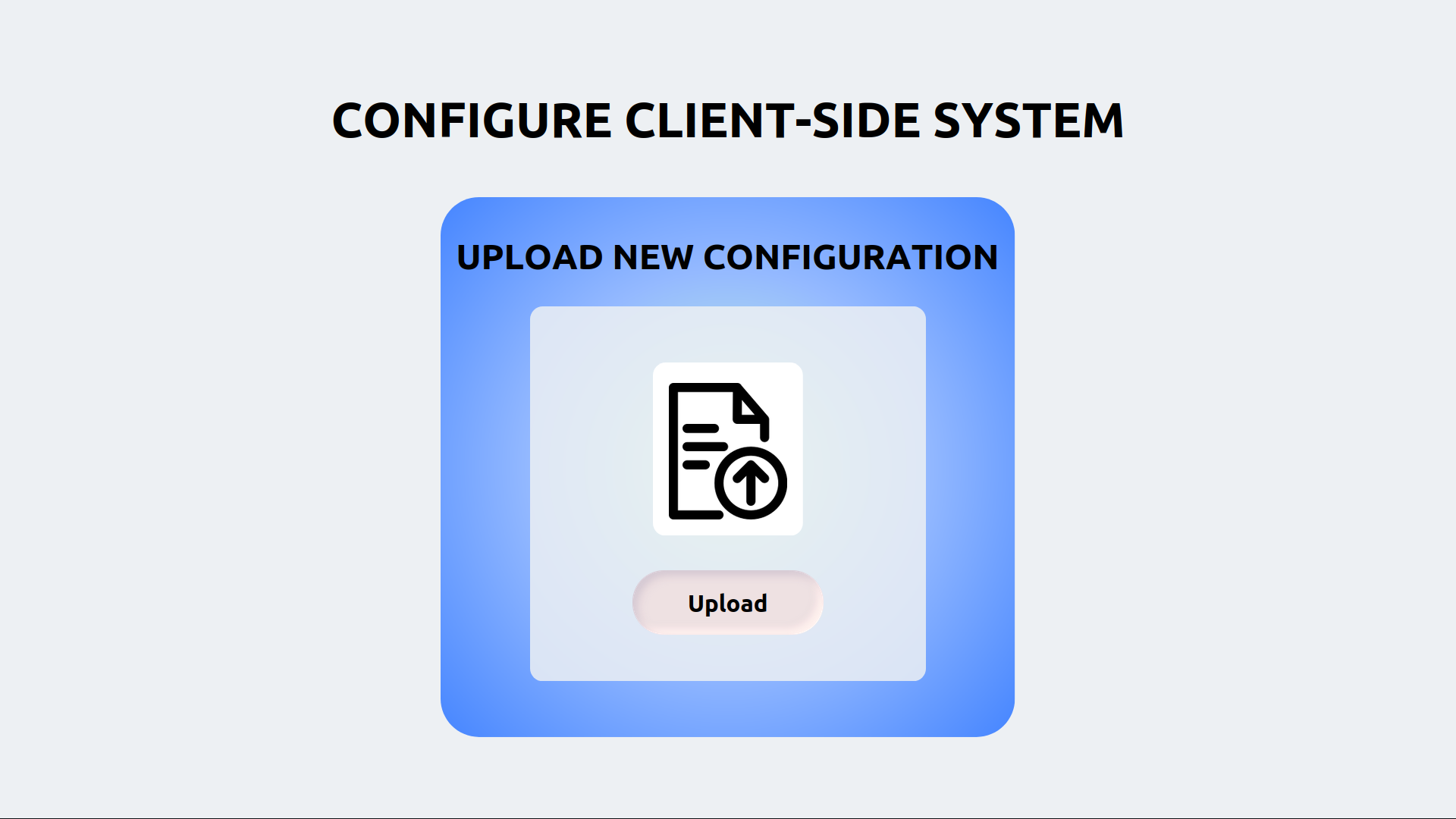
| **Step** | **O** | **CL** | **S** | **SC** |
| --- | --- | --- | --- | --- |
| **1** **ACTOR** opens the "Configure Evaluation System" form. | 1 | 1 | 2.50 | 2.50 |
| **2** **SYSTEM** displays current configuration and "Upload" button. |  |  |  |  |
| **3** **ACTOR** checks parameters against previous iterations on file | 1 | 3 | 2.50 | 7.50 |
| **4** **ACTOR** adjusts file based on current parameters | 1 | 3 | 2.50 | 7.50 |
| **5** **ACTOR** pushes "Upload" button and uploads configuration file | 1 | 1 | 2.50 | 2.50 |
| **6.1** **SYSTEM** IF config is correct and correctly formatted |  |  |  |  |
| **6.1.1** **SYSTEM** shows a confirmation message. |  |  |  |  |
| **6.2** **ELSE** |  |  |  |  |
| **6.2.1** **SYSTEM** shows error message and aborts |  |  |  |  |
| **7** **ACTOR** closes the form. | 1 | 1 | 2.50 | 2.50 |
| Human task cost | | | | 22.50 |

## Client-Side Systems

### Configure Client-Side Systems

[*Francesco Boldrini, Zahra Omrani*]

This task is performed by a ML Engineer.



"Configure Client-Side Systems" mock-up form

Detailed use case for "Configure Client-Side Systems" task  
O - Occurrence, CL - Cognitive Level, S - Normalized Salary, SC - Step Cost

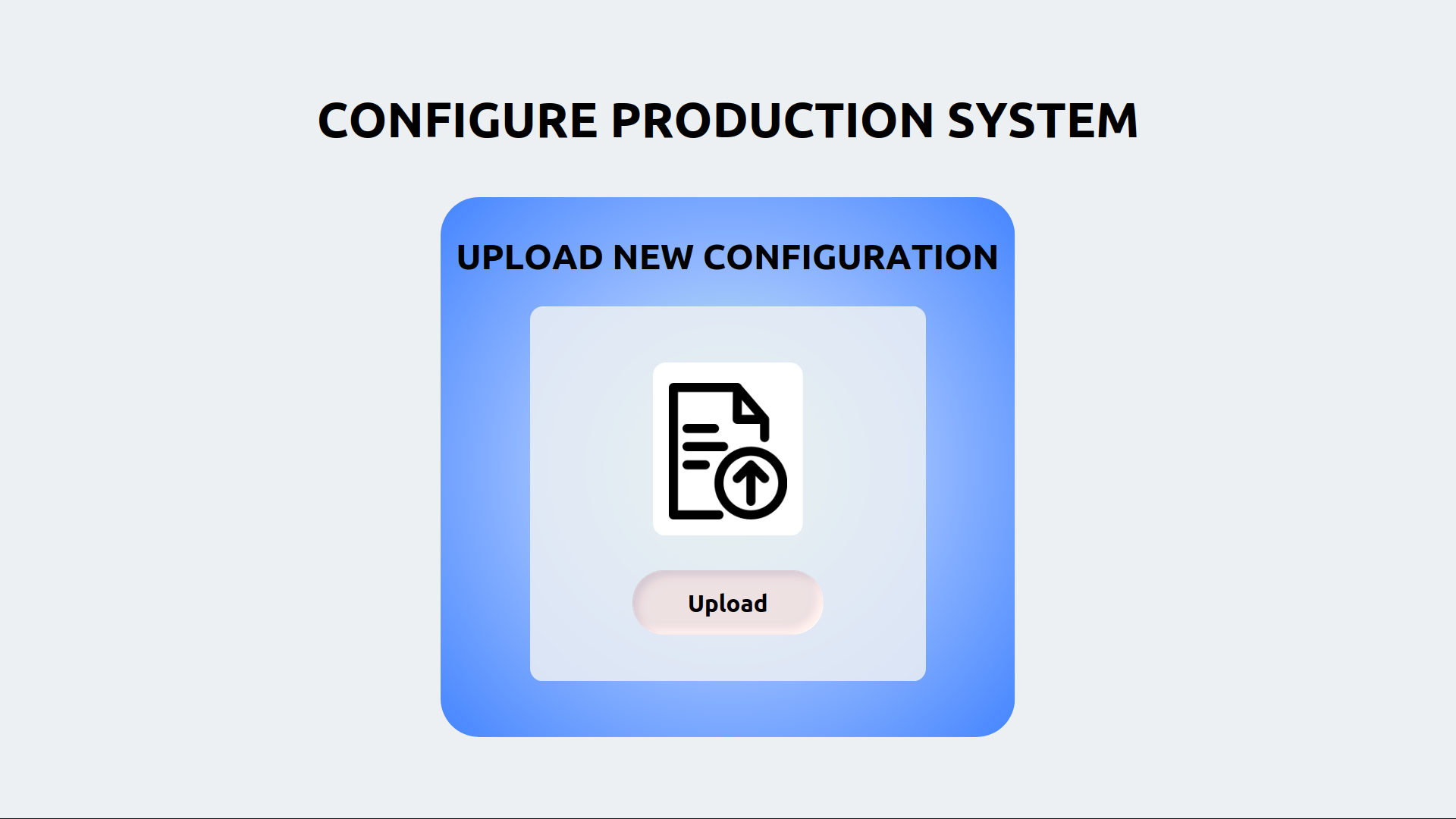
| **Step** | **O** | **CL** | **S** | **SC** |
| --- | --- | --- | --- | --- |
| **1** **ACTOR** opens the "Configure Client-Side System" form. | 1 | 1 | 2.50 |  |
| **2** **SYSTEM** displays the "Upload" button. |  |  |  |  |
| **3** **ACTOR** push the "Upload" button and upload the configuration file. | 1 | 1 | 2.50 | 2.50 |
| **4** **SYSTEM** shows a confirmation message. |  |  |  |  |
| **5** **ACTOR** closes the form. | 1 | 1 | 2.50 | 2.50 |
| Human task cost | | | | 7.50 |

## Production System

### Configure Production Systems

[*Francesco Boldrini, Zahra Omrani*]

This task is performed by a ML Engineer.



"Configure Production System" mock-up form

Detailed use case for "Configure Production" task  
O - Occurrence, CL - Cognitive Level, S - Normalized Salary, SC - Step Cost

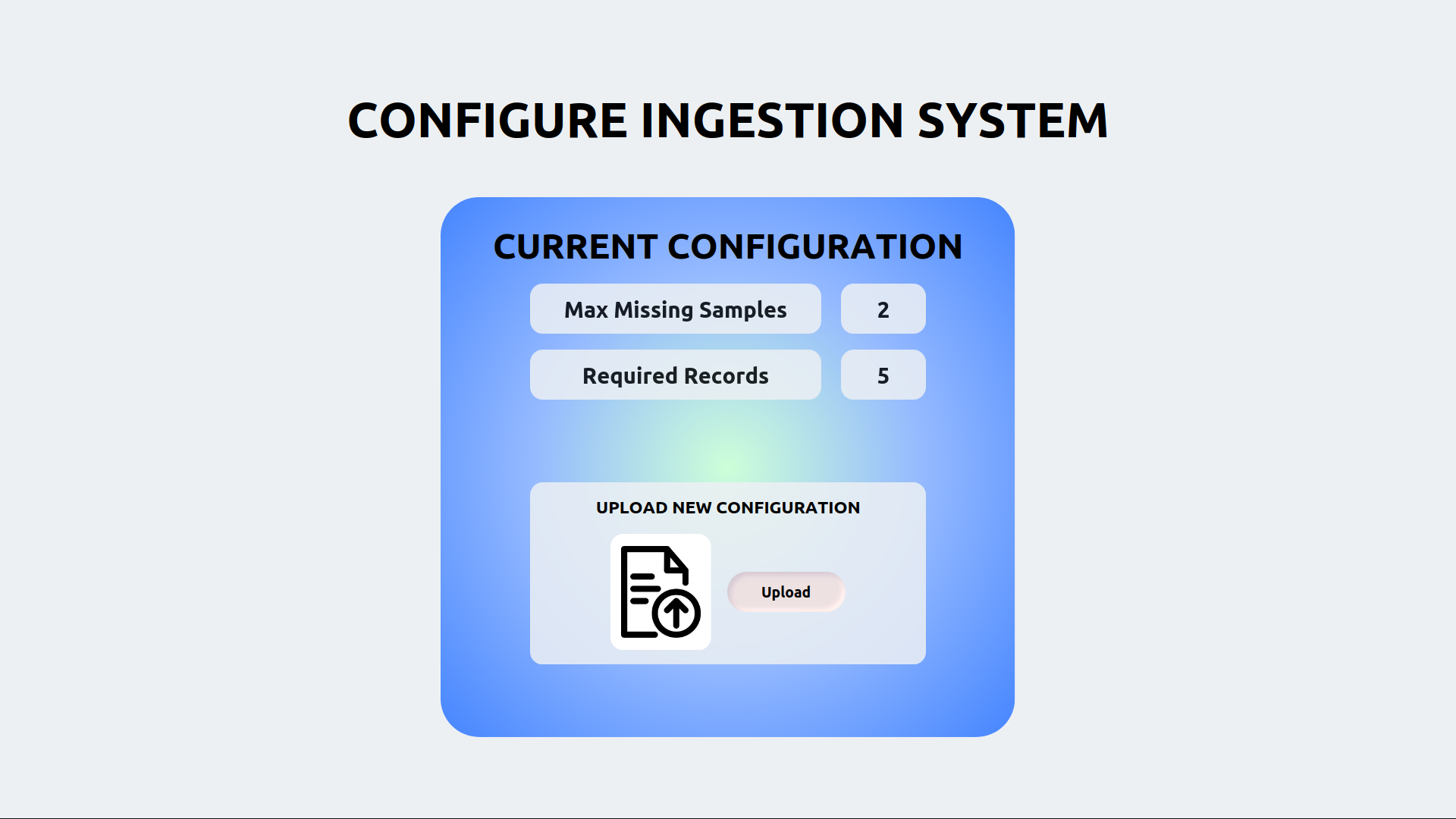
| **Step** | **O** | **CL** | **S** | **SC** |
| --- | --- | --- | --- | --- |
| **1** **ACTOR** opens the "Configure Production System" form. | 1 | 1 | 2.50 | 2.50 |
| **2** **SYSTEM** displays the "Upload" button. |  |  |  |  |
| **3** **ACTOR** push the "Upload" button and upload the configuration file. | 1 | 1 | 2.50 | 2.50 |
| **4** **SYSTEM** shows a confirmation message. |  |  |  |  |
| **5** **ACTOR** closes the form. | 1 | 1 | 2.50 | 2.50 |
| Human task cost | | | | 7.50 |

## Ingestion System

### Configure Ingestion System

[*Francesco Boldrini, Zahra Omrani*]

This task is performed by a ML Engineer.



"Configure Ingestion System" mock-up form

Detailed use case for "Configure Ingestion" task  
O - Occurrence, CL - Cognitive Level, S - Normalized Salary, SC - Step Cost

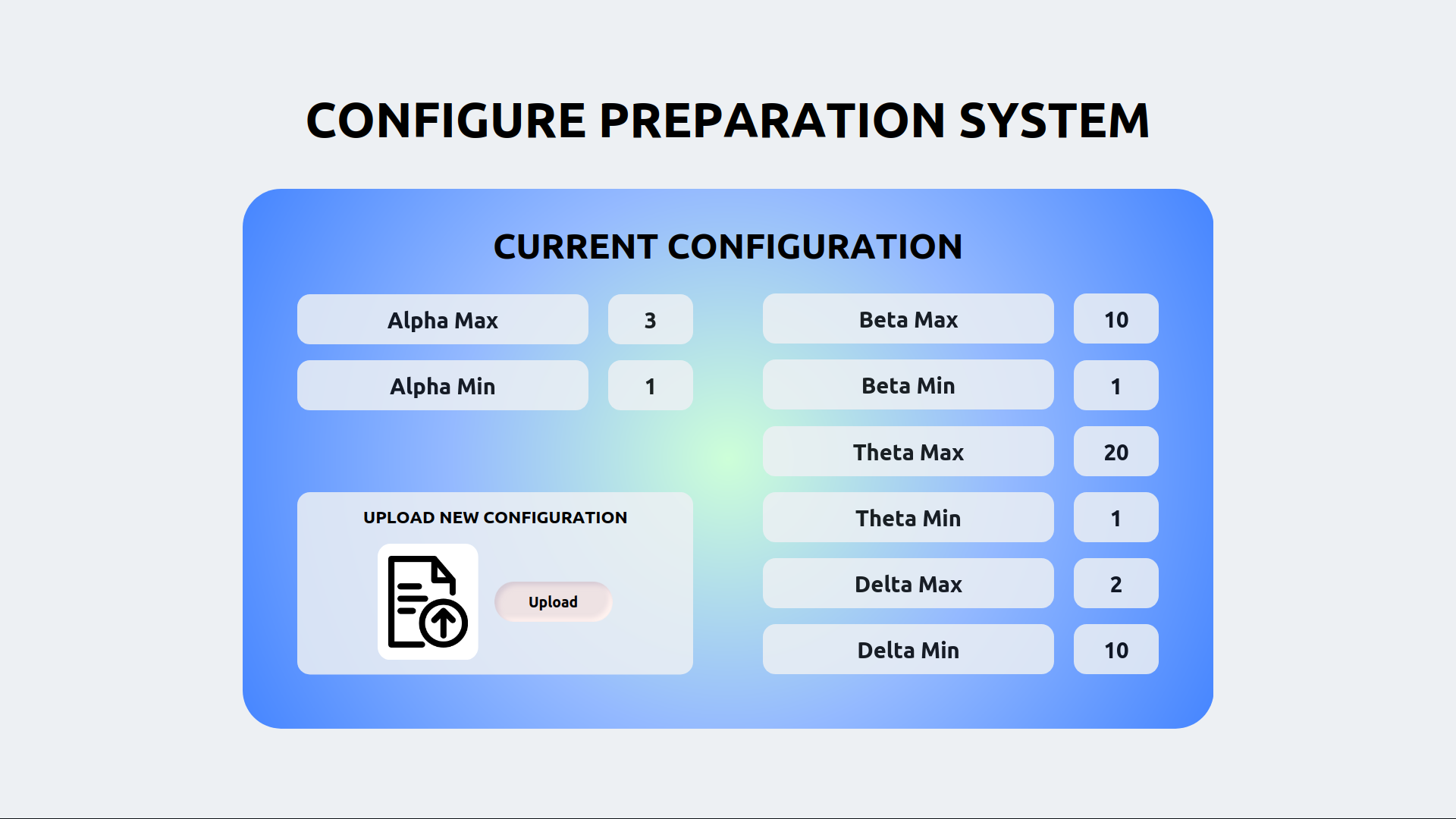
| **Step** | **O** | **CL** | **S** | **SC** |
| --- | --- | --- | --- | --- |
| **1** **ACTOR** opens the "Configure Ingestion System" form. |  |  | 2.50 |  |
| **2** **SYSTEM** displays current configuration and "Upload" button. |  |  |  |  |
| **3** **ACTOR** checks parameters against previous iterations on file | 1 | 3 | 2.50 | 7.50 |
| **4** **ACTOR** adjusts file based on current parameters | 1 | 3 | 2.50 | 7.50 |
| **5** **ACTOR** pushes "Upload" button and uploads configuration file | 1 | 1 | 2.50 | 2.50 |
| **6.1** **SYSTEM** IF config is correct and correctly formatted |  |  |  |  |
| **6.1.1** **SYSTEM** shows a confirmation message. |  |  |  |  |
| **6.2** **ELSE** |  |  |  |  |
| **6.2.1** **SYSTEM** shows error message and aborts |  |  |  |  |
| **7** **ACTOR** closes the form. | 1 | 1 | 2.50 | 2.50 |
| Human task cost | | | | 22.50 |

## Preparation System

### Configure Preparation System

[*Francesco Boldrini, Zahra Omrani*]

This task is performed by a ML Engineer.



"Configure Preparation System" mock-up form

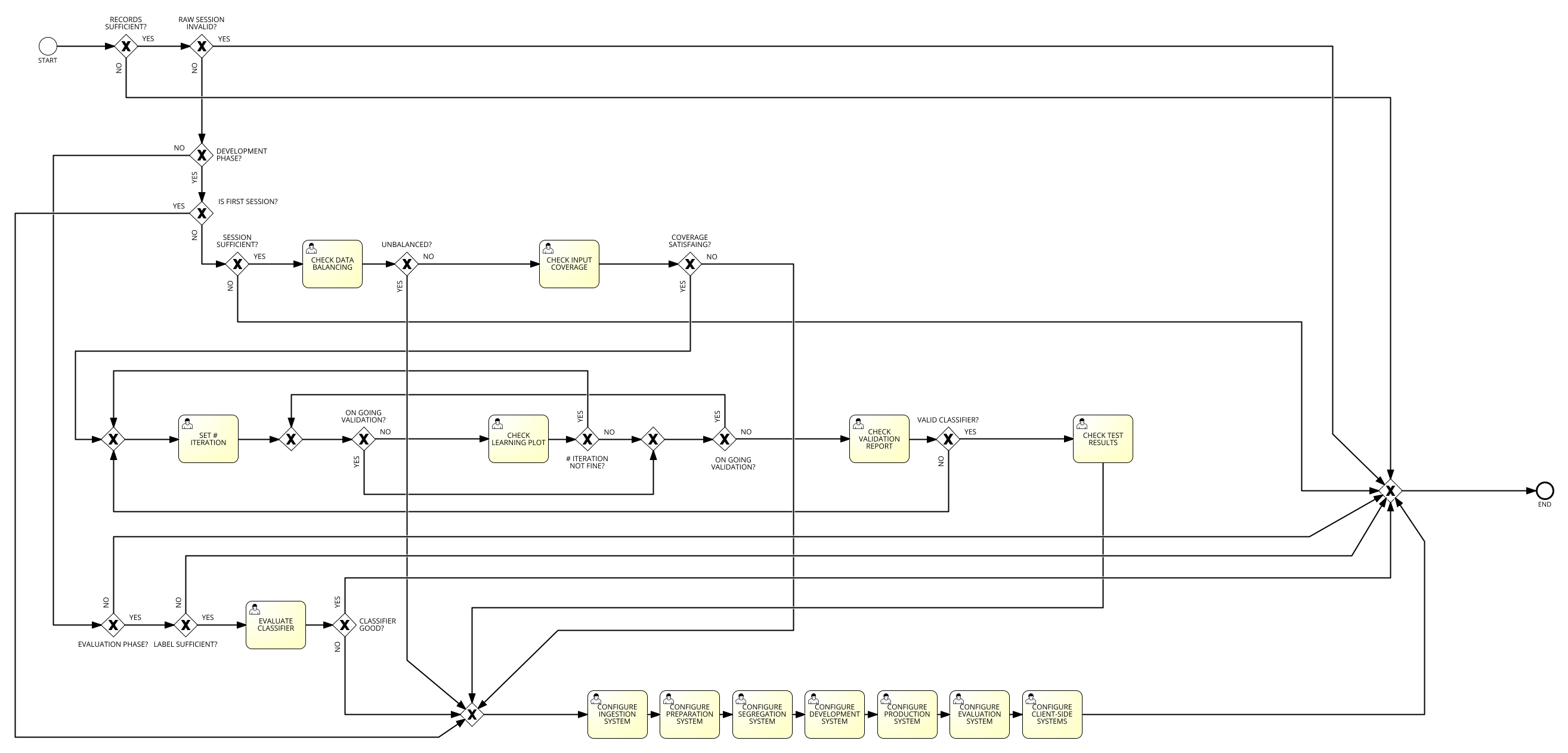
Detailed use case for "Configure Preparation" task  
O - Occurrence, CL - Cognitive Level, S - Normalized Salary, SC - Step Cost

| **Step** | **O** | **CL** | **S** | **SC** |
| --- | --- | --- | --- | --- |
| **1** **ACTOR** opens the "Configure Preparation System" form. |  |  | 2.50 |  |
| **2** **SYSTEM** displays current configuration and "Upload" button. |  |  |  |  |
| **3** **ACTOR** checks parameters against previous iterations on file | 1 | 3 | 2.50 | 7.50 |
| **4** **ACTOR** adjusts file based on current parameters | 1 | 3 | 2.50 | 7.50 |
| **5** **ACTOR** pushes "Upload" button and uploads configuration file | 1 | 1 | 2.50 | 2.50 |
| **6.1** **SYSTEM** IF config is correct and correctly formatted |  |  |  |  |
| **6.1.1** **SYSTEM** shows a confirmation message. |  |  |  |  |
| **6.2** **ELSE** |  |  |  |  |
| **6.2.1** **SYSTEM** shows error message and aborts |  |  |  |  |
| **7** **ACTOR** closes the form. | 1 | 1 | 2.50 | 2.50 |
| Human task cost | | | | 22.50 |

# Simulation

## Collapsed workflow

[*Ettore Ricci, Paolo Palumbo, Francesco Boldrini*]

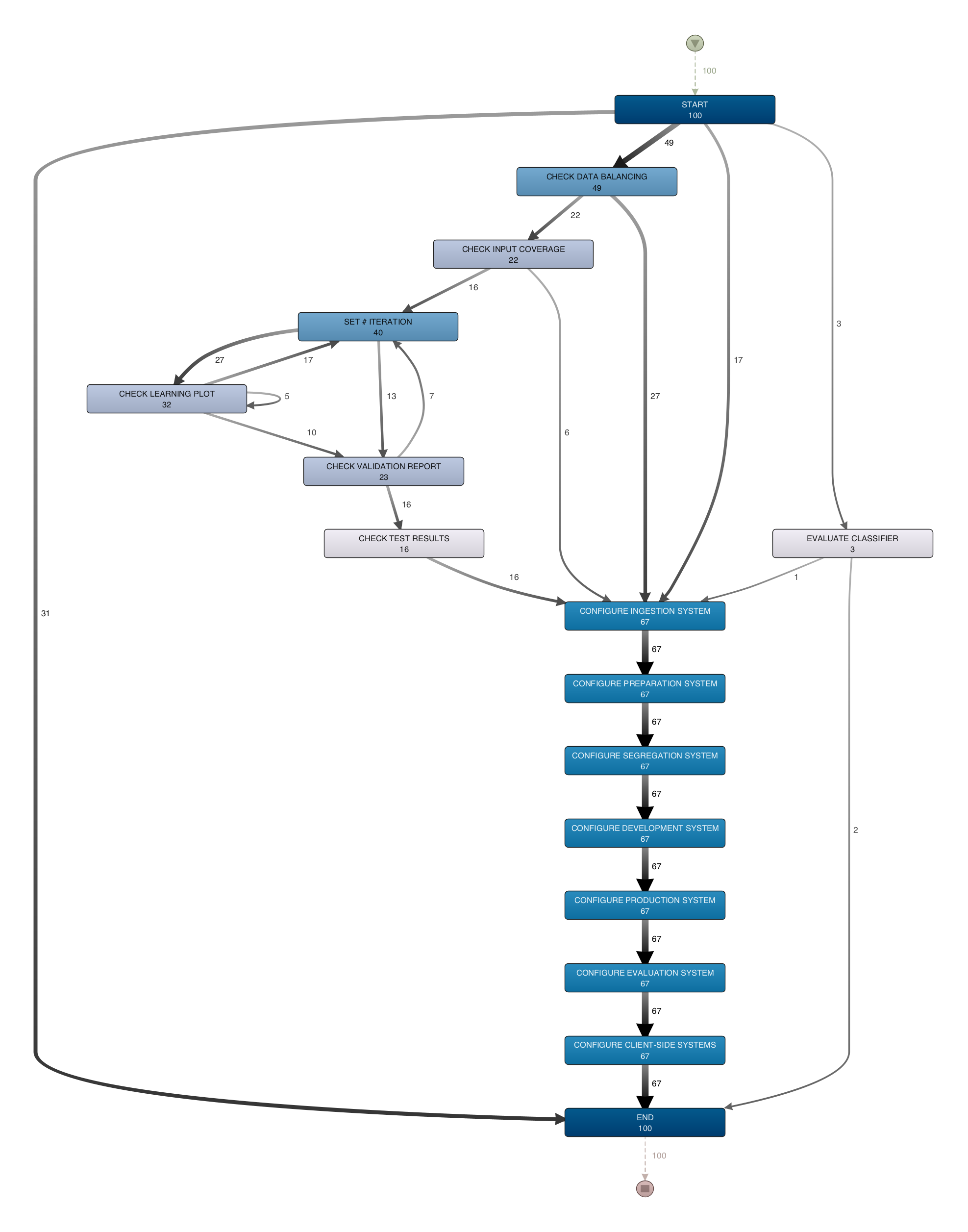


Collapsed workflow

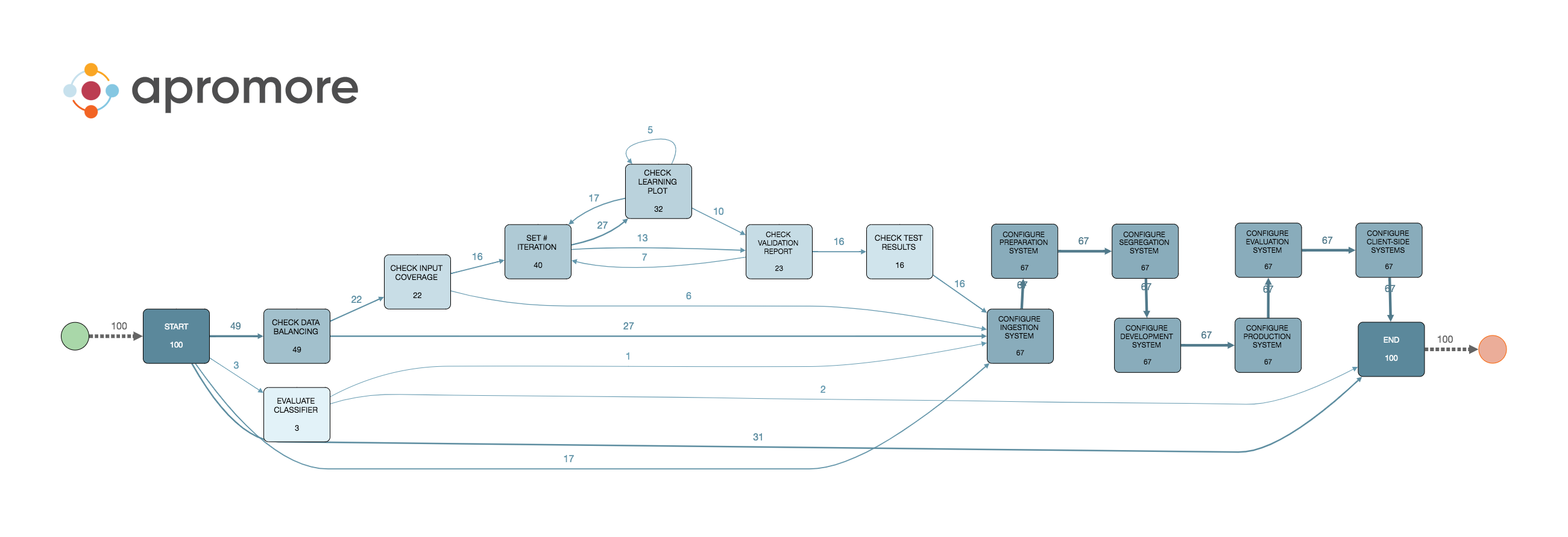
# Process mining

We mined the logs generated by the simulation of the collapsed workflow.

We modified the simulation configuration

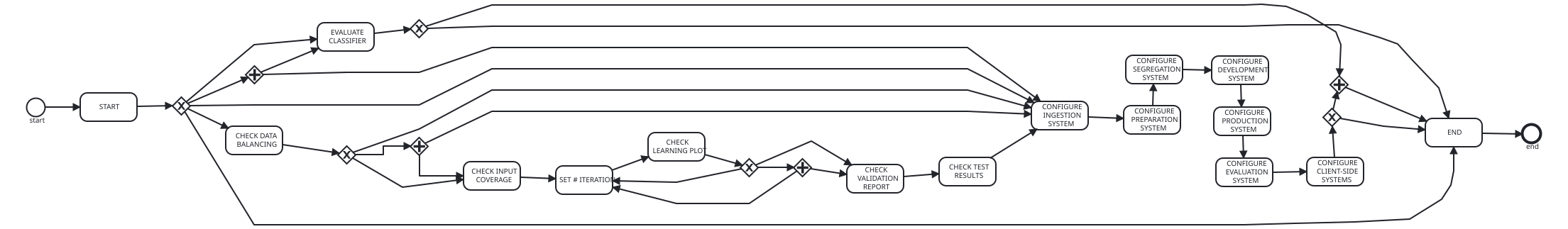


Disco analysis



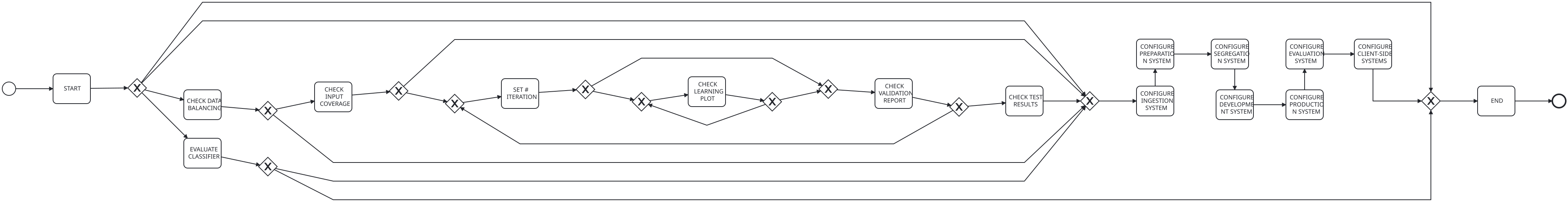
Apromore analysis

As we can see, the two transition maps mined from Disco and from Apromore are identical.



ProM mined BPMN model

The BPMN model mined from ProM is not compliant with the BPMN 2.0 specification (section 10.5). Furthermore, the model contains parallel gateways that are never present in the collapsed model.



Apromore mined BPMN model

The BPMN model mined from Apromore is much more sensible compared to the one from ProM, also it is compliant with the BPMN 2.0 specification.

Comparison of the process mining tools

| **Tool** | **Trace** | **Generalization** | **Precision** | **Simplicity** |
| --- | --- | --- | --- | --- |
| Apromore | 0.4203 | 0.9872 | 0.7566 |  |
| ProM | 0.6666 | 0 | 0 |  |