# ISO/IEC JTC 1 SC 42 Artificial Intelligence – Working Group 4

**Use Case Submission Form**

# The quality of use case submissions will be evaluated for inclusion in the Working Group’s Technical Report based the application area, relevant AI technologies, credible reference sources (see References section), and the following characteristics:

* Data Focus & Learning: Use cases for AI system which utilizes Machine Learning, and those that use a fixed *a priori* knowledge base.
* Level of Autonomy: Use cases demonstrating several degrees (dependent, autonomous, human/critic in the loop, etc.) of AI system autonomy.
* Verifiability & Transparency: Use cases demonstrating several types and levels of verifiability and transparency, including approaches for explainable AI, accountability, etc.
* Impact: Use cases demonstrating the impact of AI systems to society, environment, etc.
* Architecture: Use cases demonstrating several architectural paradigms for AI systems (e.g., cloud, distributed AI, crowdsourcing, swarm intelligence, etc.)

1. **General**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ID | (leave blank, for internal use) | | | |
| Use case name | Emotion-sensitive AI Customer Service | | | |
| Application domain | Retail | | | |
| Deployment  model | On-premise systems | | | |
| Status | In operation | | | |
| Scope[[1]](#footnote-1) | Extracting sentiment and its intensity from customers’ input, and responding with appropriate attitude in order to improve the quality of customers’ inquiry. | | | |
| Objective(s)[[2]](#footnote-2) | To design an efficient solution for customers’ sentiment and intensity detection, especially in the situation of limited training dataset. | | | |
| Narrative | Short description (not more than 150 words) | The emotion-sensitive AI customer service of JD.com Int., is supported by AI technology and deep learning method. It is developed for ameliorating accuracy of customer sentiment and intensity. In sentiment classification, it has achieved 74% accuracy and 90% recall score while in intensity detection, it has accomplished 85% accuracy and 85% recall. During the special sale of “618”, it has increased customer satisfaction by 57%. | | |
| Complete description | JD’s customer service representatives need to handle millions of requests on a daily basis. Regular AI customer service systems, 24/7 online, are capable of offering instant assistance, which alleviates the labor resources to a large extent. However, it is quite challenging, if not impossible, for those systems to interpret emotions from customer input and respond as friendly as human.  Under this background, based on huge data set of customer comments and rich experience of Natural Language Processing, our system can automatically detect sentiments like happy, angry, anxious, etc. Moreover, this system can also detect the intensity of customer sentiment. Furthermore, we adapt Convolutional Neural Networks, a widely used techniques in visual computing, to interpret the semantic meaning of customer’s expression. It can improve the system’s performance for sentiment classification and intensity detection. Moreover, with the adoption of transfer learning, the system can also be applied into various types of data. To overcome the difficulty of limited training data, we also use data augmentation method such as reverse translation and data noise to increase the variability of training data.  Up to now, the system has reached 90% recall and 74% accuracy rate for sentiment classification over 7 categories. The overall recall and accuracy for sentiment intensity are also around 85%，it has increased customer satisfaction by 57%. | | |
| Stakeholders[[3]](#footnote-3) | Customers targeted for the Customer Service system | | | |
| Stakeholders’  assets, values[[4]](#footnote-4) | Customer experience may be in influnced by the use of AI custemer service | | | |
| System’s threats & vulnerabilities[[5]](#footnote-5) | The low degree of humanization, and lack of semantic diversity for response;  Reducing the number of human customer service. | | | |
| Key performance indicators (KPIs) | ID | Name | Description | Reference to mentioned use case objectives |
| 1 | Customer Satisfaction | The ratio of customer satisfaction when using this system for requests. The expectation is 100% | Increasing its ratio as high as possible |
| 2 | Accuracy | Among all the predicted customer sentiment classification, the ratio of accurate prediction, current value is 76.4% | Increasing to 90% |
| 3 | Recall | Among all the customer sentiment intensity, the ratio of accurate prediction, current overall value is 90% | Increasing to 90% |
| 4 | Accuracy | Among all the predicted customer sentiment intensity, the ratio of accurate prediction, current overall value is 85% | Increasing to 90% |
| 5 | Recall | Among all the customer sentiment intensity, the ratio of accurate prediction, current overall value is 85% | Increasing to 90% |
|  |  |  |  |
| AI features | Task(s) | Natural language processing | | |
| Method(s)[[6]](#footnote-6) | Deep learning, transfer learning, data augmentation | | |
| Hardware[[7]](#footnote-7) |  | | |
| Topology[[8]](#footnote-8) |  | | |
| Terms and concepts used[[9]](#footnote-9) | Deep learning: a class of machine learning algorithms use a cascade of multiple layers of [nonlinear processing](https://en.wikipedia.org/wiki/Nonlinear_filter) units for [feature extraction](https://en.wikipedia.org/wiki/Feature_extraction) and transformation.  Transfer learning: we adopt multi-task learning method in this system. Jointly training different annotated data in same domain, this method improves the model performance for classification problems.  Data augmentation: we apply reverse translation to firstly translation Chinese into English and then translate it backward. We also use data noise to improve the data diversity. | | |
| Standardization  opportunities/ requirements | The system can be promoted to as many customer cervices companies as possible once provide with enough training data for the specific Application scenario | | | |
| Challenges and issues | Challenge: the system’s performance should be as good as the human customer server.  Issues: 1) limited training data; 2) sentiment classification among seven categories. | | | |
| Societal  Concerns[[10]](#footnote-10) | Description | Improving the corresponding efficiency of customer service, improving customer service experience；  Reducing labor costs, and reducing operating costs. | | |
| SDGs[[11]](#footnote-11) to be achieved | Industry, Innovation, and Infrastructure | | |

**Data (optional)**

|  |  |
| --- | --- |
| Data characteristics | |
| Description | For sentiment classification: conversation data from after-sales customer services. It’s annotated by professional annotators into 7 categories of sentiments.  For sentiment intensity: Only including sentiment data with “anger” and “anxious”; it’s annotated into 3 degrees of intensity: “low, medium, high”. |
| Source[[12]](#footnote-12) | Conversation data from JD.com real-time customer services. |
| Type[[13]](#footnote-13) | Text |
| Volume (size) | Around 60,000 sentences for sentiment classification and 20,000 for sentiment intensity. |
| Velocity[[14]](#footnote-14) | Batch Processing |
| Variety[[15]](#footnote-15) | Real-time data from JD.com, including various categories of products. |
| Variability  (rate of change)[[16]](#footnote-16) | Static |
| Quality[[17]](#footnote-17) | High |

**Process scenario (optional)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Scenario conditions | | | | | |
| No. | Scenario name | Scenario description | Triggering event | Pre-condition[[18]](#footnote-18) | Post-condition[[19]](#footnote-19) |
| 1 | Data Augmentation | Using reverse translation and noise processing to increase the size and diversity of data. | Annotated raw data is ready. |  | Increase the performance of model training. |
| 2 | Model Training | Based on the large training data, with deep learning method, to develop model for sentiment classification (7 categories) or sentiment intensity (3 categories). | Augmented data is ready |  |  |
| 3 | Evaluation | Evaluate data performance on open dataset and specific data. | Pretrained model is ready |  |  |
| 4 | Execution | Apply the trained model on real-time AI customer service. |  | The trained model has been evaluated as deployable |  |
| 5 | Retraining | Retraining model with new annotated data and new requirement from industry. |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

**Training (optional)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Scenario name | Training | | | | |
| Step No. | Event[[20]](#footnote-20) | Name of process/Activity[[21]](#footnote-21) | Primary actor | Description of process/activity | Requirement |
| 1 | Complete data augmentation | Design model for training | AI algorithm engineers | Using CNN for sentiment classification and intensity. |  |
| 2 | Complete model designing | Transfer learning | AI algorithm engineers | Multi-task learning with different data in same domain. |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| Specification of training data | |  | | | |

**Evaluation (optional)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Scenario name | Evaluation | | | | |
| Step No. | Event[[22]](#footnote-22) | Name of process/Activity[[23]](#footnote-23) | Primary actor | Description of process/activity | Requirement |
| 1 | Complete model training | Evaluation on open dataset | AI algorithm engineers | Evaluate different models’ performance on open dataset | Their performance shall be as good as state-of-art. |
| 2 | Complete model training | Evaluation on own dataset | AI algorithm engineers | Evaluate different models’ performance on own dataset | Their performance shall meet certain standard. |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| Input of evaluation | | Independent testing data | | | |
| Output of evaluation | | Accuracy and Recall | | | |

**Execution (optional)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Scenario name | Execution | | | | |
| Step No. | Event[[24]](#footnote-24) | Name of process/Activity[[25]](#footnote-25) | Primary actor | Description of process/activity | Requirement |
| 1 | Finish model training | Application | AI engineers | Making trained model into application of AI Customer Service system. |  |
| 2 | Given customer’s input | Data processing | AI algorithm engineers | Processing data into required format for model. |  |
| 3 | Finish data processing | Model prediction | AI algorithm engineers | Predicting sentiment or sentiment intensity. |  |
| 4 | Completion of Step3 | Making response | AI algorithm engineers | Making response according to the preidiction from previous step. |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| Input of Execution | |  | | | |
| Output of Execution | |  | | | |

**Retraining (optional)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Scenario name | Retraining | | | | |
| Step No. | Event[[26]](#footnote-26) | Name of process/Activity[[27]](#footnote-27) | Primary actor | Description of process/activity | Requirement |
| 1 | Certain period of time has passed since the last training/retrainig | Improve architecture of model | AI algorithm engineers | Collecting new requirements for model designing. |  |
| 2 | Certain period of time has passed since the last training/retrainig | Collecting new data | AI algorithm engineers | Collecting new data based on the further requirements. |  |
| 3 | Completing Step1&Step2 | Model retraining | AI algorithm engineers | Training new model on additional data. |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| Specification of retraining data | |  | | | |

**References**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| References | | | | | | |
| No. | Type | Reference | Status | Impact on use case | Originator/organization | Link |
| 1 | IT company | XiaoIce | In operation |  | Microsoft Asia |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

# Acceptable Reference Sources of Use Cases

* Peer-reviewed scientific/technical publications on AI applications (e.g. [1]).
* Patent documents describing AI solutions (e.g. [2], [3]).
* Technical reports or presentations by renowned AI experts (e.g. [4])
* High quality company whitepapers and presentations
* Publicly accessible sources with sufficient detail

***This list is not exhaustive. Other credible sources may be acceptable as well.***

## Examples of credible sources:

* [1] B. Du Boulay. "Artificial Intelligence as an Effective Classroom Assistant". IEEE Intelligent Systems, V 31, p.76–81. 2016.
* [2] S. Hong. "Artificial intelligence audio apparatus and operation method thereof". N US 9,948,764, Available at: [https://patents.google.com/patent/US20150120618A1/en. 2018](https://patents.google.com/patent/US20150120618A1/en.%202018).
* [3] M.R. Sumner, B.J. Newendorp and R.M. Orr. "Structured dictation using intelligent automated assistants". N US 9,865,280, 2018.
* [4] J. Hendler, S. Ellis, K. McGuire, N. Negedley, A. Weinstock, M. Klawonn and D. Burns. "WATSON@RPI, Technical Project Review".

URL: [https://www.slideshare.net/jahendler/watson-summer-review82013final. 2013](https://www.slideshare.net/jahendler/watson-summer-review82013final.%202013).

1. The scope defines the intended area of applicability, limits, and audience. [↑](#footnote-ref-1)
2. The intention of the system; what is to be accomplished?; who/what will benefit?. [↑](#footnote-ref-2)
3. Stakeholder are those that can affect or be affected by the AI system in the scenario; e.g., organizations, customers, 3rd parties, end users, community, environment, negative influencers, bad actors, etc. [↑](#footnote-ref-3)
4. Stakeholders’ assets and values that are at stake with potential risk of being compromised by the AI system deployment – e.g., competitiveness, reputation, trustworthiness, fair treatment, safety, privacy, stability, etc. [↑](#footnote-ref-4)
5. Threats and vulnerabilities can compromise the assets and values above - e.g., different sources of bias, incorrect AI system use, new security threats, challenges to accountability, new privacy threats (hidden patterns), etc. [↑](#footnote-ref-5)
6. AI method(s)/framework(s) used in development. [↑](#footnote-ref-6)
7. Hardware system used in development and deployment. [↑](#footnote-ref-7)
8. Topology of the deployment network architecture. [↑](#footnote-ref-8)
9. Terms and concepts used here should be consistent with those defined by Working Group 1 (AWI 22989 and AWI 23053) or to be recommended for inclusion. [↑](#footnote-ref-9)
10. To be inserted. [↑](#footnote-ref-10)
11. The Sustainable Development Goals (SDGs), also known as the Global Goals, are a collection of 17 global goals set by the United Nations General Assembly. SDGs are a universal call to action to end poverty, protect the planet and ensure that all people enjoy peace and prosperity.

    URL: <http://www.undp.org/content/undp/en/home/sustainable-development-goals.html> [↑](#footnote-ref-11)
12. Origin of data, which could be from customers, instruments, IoT, web, surveys, commercial activity, simulations, etc. [↑](#footnote-ref-12)
13. Structured/unstructured text, images, voices, gene sequences, numbers, composite: time-series, graph-structures, etc. [↑](#footnote-ref-13)
14. The rate of flow at which the data is created, stored, analysed, or visualized. Could be in real time. [↑](#footnote-ref-14)
15. Domains and types of data employed including formats, logical models, timescales, and semantics. Could be from multiple databases. [↑](#footnote-ref-15)
16. Changes in data rate, format/structure, semantics, and/or quality. [↑](#footnote-ref-16)
17. Completeness and accuracy of the data with respect to semantic content as well as syntax of the data (such as presence of missing fields or incorrect values). [↑](#footnote-ref-17)
18. Describes which condition(s) should have been met before this scenario happens. [↑](#footnote-ref-18)
19. Describes which condition(s) should prevail after this scenario happens. The post-condition may also define "success" or "failure" conditions [↑](#footnote-ref-19)
20. The event that triggers the step. This might be completion of the previous event. [↑](#footnote-ref-20)
21. Action verbs should be used when naming activity. [↑](#footnote-ref-21)
22. The event that triggers the step. This might be completion of the previous event. [↑](#footnote-ref-22)
23. Action verbs should be used when naming activity. [↑](#footnote-ref-23)
24. The event that triggers the step. This might be completion of the previous event. [↑](#footnote-ref-24)
25. Action verbs should be used when naming activity. [↑](#footnote-ref-25)
26. The event that triggers the step. This might be completion of the previous event. [↑](#footnote-ref-26)
27. Action verbs should be used when naming activity. [↑](#footnote-ref-27)