ITUPublications
Recommendations

International Telecommunication Union
Standardization Sector

Recommendation

ITU-T F.748.26 (02/2024)

SERIES F: Non-telephone telecommunication services

Multimedia services

Technical specification for artificial intelligence cloud platforms: Performance evaluation

ITU-T F-SERIES RECOMMENDATIONS

Non-telephone telecommunication services

TELEGRAPH SERVICE	F.1-F.109
Operating methods for the international public telegram service	F.1-F.19
The gentex network	F.20-F.29
Message switching	F.30-F.39
The international telemessage service	F.40-F.58
The international telex service	F.59-F.89
Statistics and publications on international telegraph services	F.90-F.99
Scheduled and leased communication services	F.100-F.104
Phototelegraph service	F.105-F.109
MOBILE SERVICE	F.110-F.159
Mobile services and multidestination satellite services	F.110-F.159
TELEMATIC SERVICES	F.160-F.399
Public facsimile service	F.160-F.199
Teletex service	F.200-F.299
Videotex service	F.300-F.349
General provisions for telematic services	F.350-F.399
MESSAGE HANDLING SERVICES	F.400-F.499
DIRECTORY SERVICES	F.500-F.549
DOCUMENT COMMUNICATION	F.550-F.599
Document communication	F.550-F.579
Programming communication interfaces	F.580-F.599
DATA TRANSMISSION SERVICES	F.600-F.699
MULTIMEDIA SERVICES	**F.700-F.799**
ISDN SERVICES	F.800-F.849
UNIVERSAL PERSONAL TELECOMMUNICATION	F.850-F.899
ACCESSIBILITY AND HUMAN FACTORS	F.900-F.999

For further details, please refer to the list of ITU-T Recommendations.

Recommendation ITU-T F.748.26

Technical specification for artificial intelligence cloud platforms: Performance evaluation

Summary

Recommendation ITU-T F.748.26 provides a comprehensive performance evaluation framework for artificial intelligence (AI) cloud platforms. Recommendation ITU-T F.748.26 gives an overview of the evaluation framework, configuration specification, workloads, metrics, requirements on evaluation results and evaluation suggestions. Recommendation ITU-T F.748.26 can be a unified guideline for developers, users, third party test agencies and researchers to analyse and access the performance of AI cloud platforms.

History *

Edition	Recommendation	Approval	Study Group	Unique ID
1.0	ITU-T F.748.26	2024-02-13	16	11.1002/1000/15847

Keywords

Artificial intelligence, benchmark, metrics, performance, workload.

* To access the Recommendation, type the URL https://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents/software copyrights, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the appropriate ITU-T databases available via the ITU-T website at http://www.itu.int/ITU-T/ipr/.

Table of Contents

		Page
1	Scope	1
2	References	1
3	Definitions	1
	3.1 Terms defined elsewhere	1
	3.2 Terms defined in this Recommendation	2
4	Abbreviations and acronyms	2
5	Conventions	2
6	Overview of AI cloud platform performance evaluation framework	2
	6.1 Evaluation object	2
	6.2 Evaluation principle	2
	6.3 Workflow of performance evaluation framework	2
7	Configuration specification	3
	7.1 Computing resource cluster configuration	3
	7.2 Node configuration	3
	7.3 Software configuration	4
	7.4 Physical environment configuration	4
8	Evaluation workloads and metrics	4
	8.1 Operation level	4
	8.2 Model level	6
	8.3 Platform level	7
9	Requirements on evaluation results	8
	9.1 Benchmark report	8
	9.2 Benchmark materials	8
Appendix I – Evaluation suggestions		9
	I.1 Evaluation program	9
	I.2 Dataset preparation	9
	I.3 Evaluation workloads	9
Bibliography		10

Recommendation ITU-T F.748.26

Technical specification for artificial intelligence cloud platforms: Performance evaluation

1 Scope

This Recommendation provides a comprehensive performance evaluation framework for artificial intelligence (AI) cloud platforms.

In particular, this Recommendation includes:

– an overview of the performance evaluation framework;

– a configuration specification for an AI cloud platform for performance evaluation;

– workloads and metrics for AI cloud platform performance evaluation; and

– requirements on evaluation results.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

[ITU-T F.748.11] Recommendation ITU-T F.748.11 (2020), *Metrics and evaluation methods for a deep neural network processor benchmark.*

[ITU-T F.748.17] Recommendation ITU-T F.748.17 (2022), *Technical specification for artificial intelligence cloud platform – Artificial intelligence model development.*

[ITU-T F.748.18] Recommendation ITU-T F.748.18 (2022), *Metric and evaluation methods for AI-enabled multimedia application computing power benchmark.*

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 artificial intelligence [b-ISO/IEC 2382]: Branch of computer science devoted to developing data processing systems that perform functions normally associated with human intelligence, such as reasoning. learning, and self-improvement.

3.1.2 benchmark [ITU-T F.748.11]: Benchmark is an evaluation method with a long-term application in the entire computer field. Example: As computer architecture advanced, it became more difficult to compare the performance of various computer systems simply by looking at their specifications. Therefore, tests were developed that allowed comparison of different architectures (i.e., providing benchmarks).

3.1.3 model [b-ITU-T F.748.16]: An output created by an algorithm running on training data that can generate an inference or prediction, based on input data.

3.2 Terms defined in this Recommendation

None.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

AI Artificial Intelligence

CPU Central Processing Unit

FP Floating Point

GPU Graphics Processing Unit

IO Input/Output

IOPS Input/output Operations Per Second

5 Conventions

The following conventions are used in this Recommendation:

– The phrase "**is required**" indicates a requirement that must be strictly followed and from which no deviation is permitted if conformance to this Recommendation is to be claimed.

– The phrase "**is recommended**" indicates a requirement that is recommended but which is not absolutely required. Thus, this requirement need not be present to claim conformance.

6 Overview of AI cloud platform performance evaluation framework

6.1 Evaluation object

The evaluation object is the AI cloud platform. It manages the full lifecycle of AI model development and deployment. It usually consists of hardware resources and software stacks. It is recommended that heterogeneous AI computing power, mainstream open-source deep learning framework and multiple development environment for developers to build AI models efficiently be supported. For requirements for an AI cloud platform, see [ITU-T F.748.17].

6.2 Evaluation principle

– Reproducibility: The environment configuration, workloads, evaluation metrics and the implementation of the benchmark is required to be clearly defined and avoid ambiguity. Hence, the developers of the AI cloud platform or the testing agencies can reproduce the evaluation on demand.

– Scalability: The evaluation is recommended to be able to test the ability of the platform to handle heavy workloads and a large number of users.

– Practicality: The evaluation is recommended to be applicable to mainstream AI cloud platforms. Moreover, it is recommended to provide results practical for the detection of performance bottlenecks and optimization of computing efficiency, resource usage and power consumption.

6.3 Workflow of performance evaluation framework

The workflow of the performance evaluation framework for an AI cloud platform consists of five parts as shown in Figure 6-1. The evaluation object is the AI cloud platform with specified configuration. The input workloads are classified by level as operation, mode and platform. Each workload is associated with specific metrics. Moreover, the output results include an evaluation report and evaluation material.

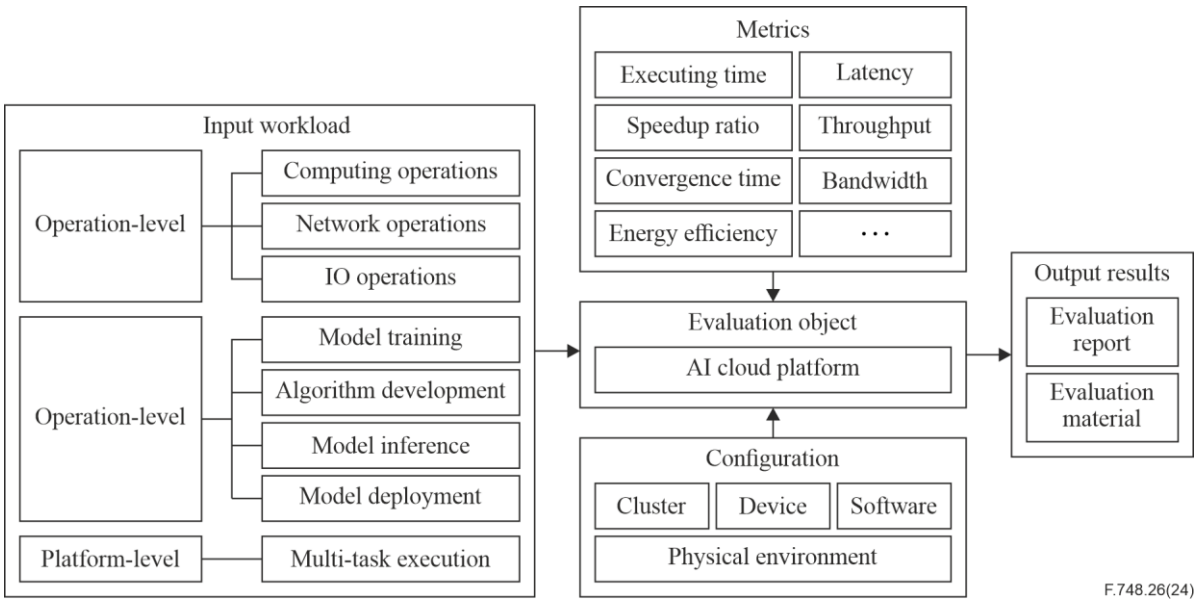

Figure 6-1 – A recommended workflow for a performance evaluation framework

7 Configuration specification

The performance of an AI cloud platform is affected by a large number of factors, such as hardware, libraries, schedulers, deep learning frameworks, etc. For performance evaluation, the configuration about AI cloud platform is recommended to be clearly specified.

7.1 Computing resource cluster configuration

An AI cloud platform is deployed on a computing resource cluster. The architecture and composition of a computing resource cluster are crucial to the performance of the platform. Therefore, the following information about a computing resource cluster is recommended to be specified.

– Nodes: A computing resource cluster comprises various types of node, which include those for computing, storage and management. The number of each type of node is required to be specified.

– Topology: Topology determines how nodes are connected to each other and how data is transmitted in a computing resource cluster. The topology of the computing resource cluster is recommended to be specified.

– Network: The network refers to the actual communication infrastructure between the nodes. The information such as the network protocol and the bandwidth are recommended to be specified.

7.2 Node configuration

Each node in a computing resource cluster comprises heterogeneous devices with different computational capabilities. The following information is recommended to be specified.

– Computing devices: A node may comprise heterogeneous computing devices, which include but are not limited to a central processing unit (CPU), graphics processing unit (GPU), neural network processing unit, field programmable gate array and application specific integrated circuit. The number of computing devices (all types) is recommended to be specified.

– Computing power: The parameters that contribute to the computing power of each device are recommended to be specified. For example, a CPU device is specified by the following indicators: the micro-architecture; the number of cores; the number of threads; CPU clock speed; cache; word size; manufacturing process of CPU; etc. A GPU may include additional indicators such as GPU memory capacity and the number or type of streaming processors.

– Memory: The memory contributes to the amount of data the node can process at once. The type, the bandwidth, speed and the capacity of memory are recommended to be specified.
– Disk: The disk contributes to the storage capacity of the node. The disk type, bandwidth, speed and storage capacity are recommended to be specified.
– System bus: The system bus determines the speed of inter-device transmission of data, address and control instructions. The generation, speed and bandwidth are recommended to be specified.
– The network adapter: The network adapter determines the speed of intra-device data transmission. The bandwidth, interface and network protocol are recommended to be specified.

7.3 Software configuration

Each node in a computing resource cluster has its own software stack. The following information is recommended to be specified.

– Operating system: The operating system is software that manages the hardware and software resources of the node. The type, version and release number of the operating system are recommended to be specified.
– Hardware driver: The hardware driver is software that allows the operating system to access hardware devices. The version of the hardware driver is recommended to be specified.
– Runtime environment: The runtime environment is software that provides a running environment for tasks. Common runtime environments used in computing resource clusters include but are not limited to the following: bare metal server; virtual machine; and container. The type of runtime environment is recommended to be specified.

7.4 Physical environment configuration

A computing resource cluster is designed to operate within a specific physical environment to ensure its optimal performance and longevity. The following information is recommended to be specified.

– Temperature: The computing resource cluster is required to run within a certain range of temperature to prevent overheating and damage to hardware devices. The temperature is recommended to be specified.
– Humidity: The computing resource cluster is required to run within a certain range of humidity to prevent corrosion and damage to electronic components. The humidity is recommended to be specified.
– Power supply: The computing resource cluster is required to be supported by a certain level of electrical power. The voltage and the power of the power adapter are recommended to be specified.

8 Evaluation workloads and metrics

The performance of an AI cloud platform can be indicated by the execution time and the resource consumption of a number of computing tasks. The tasks are classified into three levels: operation; model; and platform.

8.1 Operation level

In the context of an AI cloud platform, the notion of operations refers to atomic building blocks of complex AI tasks. They are mathematical calculations or data transformations that are used to process input data and generate output data. The performance of an AI computing task is influenced by the performance of each operation that comprises the entire task. The basic operations are classified into those for: computing; network; and input/output (IO).

8.1.1 Computing operations

8.1.1.1 Workloads for computing operations

The recommended computing operations used for evaluation include but are not limited to general matrix multiply, element-wise matrix calculation, convolution, pooling, rectified linear units, batch normalization, residual addition and self-attention.

8.1.1.2 Metrics for computing operations

Execution time (T_{co}): the time it takes for an operation to be executed from the start time point T_{co}^{s} to the end time point $T_{\mathrm{co}}^{\mathrm{e}}$.

$$T_{\mathrm{co}} = T_{\mathrm{co}}^{\mathrm{e}} - T_{\mathrm{co}}^{s} \tag{8-1}$$

Throughput (TP_{co}): the number of operations N_{co} that can be executed per time T_{u}.

$$TP_{\mathrm{co}} = \frac{N_{\mathrm{co}}}{T_{\mathrm{u}}} \tag{8-2}$$

Energy efficiency (EE_{co}): The ratio of computing power to electric power consumption. The computing power (CP_{co}) is measured in floating point of operations per second or operations per second. The electric power (EP_{co}) consumption can be measured with the power meter during the execution time [ITU-T F.748.18].

$$EE_{\mathrm{co}} = \frac{CP_{\mathrm{co}}}{EP_{\mathrm{co}}} \tag{8-3}$$

8.1.2 Network operations

8.1.2.1 Workloads for network operations

Network operations are used to evaluate communication efficiency. The typical workloads include but are not limited to scatter, all-reduce and all-to-all.

8.1.2.2 Metrics for network operations

Execution time (T_{no}): The time it takes for an operation to be executed from the start time point T_{no}^{s} to the end time point $T_{\mathrm{no}}^{\mathrm{e}}$.

$$T_{\mathrm{no}} = T_{\mathrm{no}}^{\mathrm{e}} - T_{\mathrm{no}}^{s} \tag{8-4}$$

Bandwidth (*B*): The transmission speed of data and model parameters between devices, measured in tera-. giga-, mega or kilobytes per second. For each data transmission between two nodes, record the volume of transmitted data (D_{no}) and the time duration T_{no} of transmission.

$$B = \frac{D_{\mathrm{no}}}{T_{no}} \tag{8-5}$$

8.1.3 IO operations

8.1.3.1 Workloads for IO operations

The IO operations used for evaluation include but are not limited to parallel IO and data replication.

8.1.3.2 Metrics for IO operations

IOPS: the number of input/output operations per second that can be processed.

Response time (T_{io}): The time it takes for the disk to respond to an IO request. Record the start time point $T_{\mathrm{io}}^{\mathrm{s}}$ of receiving an IO request, and the end time point $T_{\mathrm{io}}^{\mathrm{e}}$ of the disk completing the response.

$$T_{\mathrm{io}} = T_{\mathrm{io}}^{\mathrm{e}} - T_{\mathrm{io}}^{\mathrm{s}} \tag{8-6}$$

8.2 Model level

Based on the lifecycle of AI development, the model level tasks for performance evaluation include: model training; algorithm development; model inference; and model deployment.

8.2.1 Workloads for model level benchmark

The model level tasks are recommended to be representative AI applications, such as image classification, machine translation or recommendation. The recommended workloads include but are not limited to the following: residual network; single shot detector; bidirectional encoder representations from transformers; deep learning recommendation model; mask region-based convolutional neural network; and transformer.

For recommended workloads and test information, see Table 10-1 of [ITU-T F.748.11]. Further, the following workload-related information is recommended to be specified.

– Model definition: The model structure, model optimizer, loss function, initializer, etc.

– Dataset: The dataset, data enhancement approaches and data arrival modes for inference and deployment.

– Computing precision: The computing precision, which can be INT8 (where INT is integer), FP16, FP32, FP64 (where FP is floating point), mixed, etc.

– Hyperparameters: The batch size, learning rate, momentum, etc.

– Accuracy: The target accuracy for each workload.

– Development tools: The AI framework, third party libraries, integrated development environment, etc.

– Computing resource: The type and number of computing devices.

8.2.2 Metrics for model training task

Training time (T_{mt}): From the start time point $T_{\mathrm{mt}}^{\mathrm{s}}$ of loading and training data until the time point $T_{\mathrm{mt}}^{\mathrm{e}}$ that the training process ends.

$$T_{\mathrm{mt}} = T_{\mathrm{mt}}^{\mathrm{e}} - T_{\mathrm{mt}}^{\mathrm{s}} \tag{8-7}$$

Convergence time T_{ct}: From the time point of starting the first iteration $T_{\mathrm{ct}}^{\mathrm{s}}$ until the time point $T_{\mathrm{ct}}^{\mathrm{e}}$ of reaching the target accuracy.

$$T_{\mathrm{ct}} = T_{\mathrm{ct}}^{\mathrm{e}} - T_{\mathrm{ct}}^{\mathrm{s}} \tag{8-8}$$

Speedup ratio for distributed training SR_p: The ratio between the computing time $T_{\mathrm{mt}-1}$ needed by one processor and the computing time $T_{\mathrm{mt}-p}$ needed by p processors.

$$SR_p = \frac{T_{\mathrm{mt}-1}}{T_{\mathrm{mt}-p}} \tag{8-9}$$

Throughput: The amount of data that is processed per time, for the calculation method, see Equation (8-2).

Energy efficiency: For the calculation method, see Equation (8-3).

8.2.3 Metrics for algorithm development task

Development environment launch time T_{ad}: From the time point $T_{\mathrm{ad}}^{\mathrm{s}}$ of calling for launch a development environment until the time point T_{ad}^{s} where the development environment gets ready.

$$T_{\mathrm{ad}} = T_{\mathrm{ad}}^{\mathrm{e}} - T_{\mathrm{ad}}^{s} \tag{8-10}$$

Throughput: The amount of data that is processed per time, for the calculation method, see Equation (8-2).

Energy efficiency: For the calculation method, see Equation (8-3).

8.2.4 Metrics for model inference task

Inference latency (T_{mi}): from the time point $T_{\mathrm{mi}}^{\mathrm{s}}$ of start sending the first sample until the time point $T_{\mathrm{mi}}^{\mathrm{e}}$ of receiving the inference result of the last sample.

$$T_{\mathrm{mi}} = T_{\mathrm{mi}}^{\mathrm{e}} - T_{\mathrm{mi}}^{\mathrm{s}} \tag{8-11}$$

Throughput: The amount of data that is processed per time, for the calculation method, see Equation (8-2).

Energy efficiency: For the calculation method, see Equation (8-3).

8.2.5 Metrics for model deployment task

Response latency (T_{md}): The response time duration consumed by the complete inference request for a sample. Record the start time point $T_{\mathrm{md}}^{\mathrm{s}}$ of inference request for a sample, and the time point $T_{\mathrm{md}}^{\mathrm{e}}$ of receiving the inference result of the sample.

$$T_{\mathrm{md}} = T_{\mathrm{md}}^{\mathrm{e}} - T_{\mathrm{md}}^{\mathrm{s}} \tag{8-12}$$

Speedup ratio for batched data (SR^{b}): The ratio between the computing time $\sum_{i=1}^{p} T_{\mathrm{md}}^{i}$ needed by processing p requests in order and the computing time T_{bmd}^{p} needed by processing a batch of p requests.

$$SR^{\mathrm{b}} = \frac{\sum_{i=1}^{p} T_{\mathrm{md}}^{i}}{T_{\mathrm{bmd}}^{p}} \tag{8-13}$$

Request throughput: The amount of inference requests processed per time, for the calculation method, see Equation (8-2).

Energy efficiency: For the calculation method, see Equation (8-3).

8.3 Platform level

AI cloud platforms can support multiple tasks simultaneously; these tasks can be homogeneous or heterogeneous. The efficiency of processing multi-tasks is an important indicator of the AI cloud platform performance. The workloads for platform level tasks are a combination of model level tasks as specified in clause 8.2.1.

8.3.1 Platform level task information

In addition to model level task information, the following information is recommended to be specified.

– Type of tasks: The type of tasks that are being processed simultaneously by the platform.

– Number of tasks (*n*): The number of each type task that are being processed simultaneously by the platform.

– Order of task requesting: The order in which tasks are requested from the platform.

8.3.2 Metrics for platform level performance

It is recommended to evaluate the platform level performance according to the following metrics:

In this clause, it is assumed that a queue of *n* tasks is requested by users. The request time, the start execution time and the end point of execution time of each task are denoted by $T_{\mathrm{r}}^{i}, T_{\mathrm{s}}^{i}$ and T_{e}^{i} (i = 1, 2, ... *n*), respectively, where i is the task number.

Total execution time (T_{tt}): The time duration to complete all tasks.

$$T_{\mathrm{tt}} = \max(T_{\mathrm{e}}^{i}) - \min(T_{\mathrm{r}}^{i}) \tag{8-14}$$

Maximum waiting time ($T_{\mathrm{max-wt}}$): The maximum waiting time for tasks to be executed.

$$T_{\text{max-wt}} = \max(T_{\text{s}}^{i} - T_{\text{r}}^{i}) \quad (8\text{-}15)$$

Average waiting time ($T_{\text{avg-wt}}$): The average waiting time of all tasks.

$$T_{\text{avg-wt}} = \frac{\sum_{i=1}^{n}(T_{\text{s}}^{i} - T_{\text{r}}^{i})}{n} \quad (8\text{-}16)$$

Energy efficiency: For the calculation method, see Equation (8-3).

9 Requirements on evaluation results

The evaluation results are recommended for formulation into a benchmark report for the test agency and for maintenance of necessary benchmark materials.

9.1 Benchmark report

It is recommended that all test information and results be arranged in a benchmark report. A benchmark report comprises the following components.

– Metadata information: It is recommended that metadata information of the test be recorded, including the test agency, the test time and the test object.

– Configuration information: It is recommended that the test information specified in clause 7 be included.

– Workloads descriptions: It is recommended that the descriptions of workloads specified in clause 8 be included.

– Metrics: It is recommended that the metrics specified in clause 8 be included.

9.2 Benchmark materials

It is recommended that all benchmark materials be submitted to the test agency. The benchmark materials include but are not limited to the following.

– Source code: It is recommended that all source code implemented for the benchmark be included.

– Logs: It is recommended that the test logs, which include but are not limited to the timestamp, the number of runs, the number of epochs, the number of queries or samples processed, the accuracy, the throughput or other necessary information, be included.

– Model file: It is recommended that the model file that records the weights and structure information be included.

– Metrics calculation: It is recommended that the scripts that calculate and record the metrics be included.

Appendix I

Evaluation suggestions

(This appendix does not form an integral part of this Recommendation.)

The performance of the AI cloud platform is recommended to be evaluated by implementing a certain benchmark. The choice, implementation and execution of the benchmark follow the principles of the performance evaluation specified in clause 6. In this appendix, some suggestions for benchmark implementation are provided.

I.1 Evaluation program

For an evaluation program, recommendations follow:

– implementation of interfaces to collect performance information;

– implementation of functions to compute performance metrics;

– implementation of functions to record testing logs;

– maintaining the immutability of the test program during the test.

I.2 Dataset preparation

For data preparation, recommendations follow:

– application of necessary data pre-processing procedures;

– utilization of a high-performance distributed file system for data storage;

– maintenance of the integrity and orderliness of the dataset and the immutability of the dataset samples;

– avoidance of unnecessary observation of training or validation dataset by extracting features from the samples.

I.3 Evaluation workloads

For evaluation workloads, recommendations follow:

– avoidance of the application of additional optimization approaches, such as weight deletion, model pruning, knowledge distillation and re-training in model level task;

– maintenance of the model architecture and hyperparameters unchanged during the test;

– Avoidance of the execution of any process that differs from the test program during the test.

Bibliography

[b-ITU-T F.748.16] Recommendation ITU-T F.748.16 (2022), *Requirements for applications and services in smart manufacturing based on machine vision*.

[b-ISO/IEC 2382] International Standard ISO/IEC 2382:2015, *Information technology – Vocabulary*.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	Tariff and accounting principles and international telecommunication/ICT economic and policy issues
Series E	Overall network operation, telephone service, service operation and human factors
Series F	**Non-telephone telecommunication services**
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Environment and ICTs, climate change, e-waste, energy efficiency; construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
Series Q	Switching and signalling, and associated measurements and tests
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	Global information infrastructure, Internet protocol aspects, next-generation networks, Internet of Things and smart cities
Series Z	Languages and general software aspects for telecommunication systems

Published in Switzerland
Geneva, 2024