

团 体 标 准

T/CCF 0003—2024

人工智能模型训练碳排放评估方法

Assessment method for carbon emission measurement of AI model training

2024-12-31 发布

2024-12-31 实施

目 次

前 言.....	II
1 范围.....	1
2 规范性引用文件.....	1
3 术语和定义.....	1
4 缩略语.....	1
5 评估要求.....	2
5.1 核算范围.....	2
5.2 系统边界.....	2
5.3 AI 模型碳排放计算涉及参数.....	2
5.4 数据质量要求.....	4
5.5 数据取舍准则.....	4
6 评估方法.....	4
6.1 AI 模型训练碳排放.....	4
6.2 AI 模型训练电力消耗量计算.....	5
附 录 A（资料性）某 AI 模型训练碳排放评估简要示例.....	6
参考文献.....	10

前 言

本文件按照 GB/T 1.1—2020《标准化工作导则 第 1 部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由中国计算机学会提出并归口。

本文件起草单位：浪潮电子信息产业股份有限公司。

本文件主要起草人：范宝余，刘璐，王丽，郭振华，王永霞，卢冶，安小米，李长升，方芳，宋桂香，邸贺亮。

人工智能模型训练碳排放评估方法

1 范围

本文件规定了人工智能模型（以下简称 AI 模型）训练碳排放评估的核算范围、系统边界、参数数据，描述了 AI 模型训练碳排放的评估方法。

本文件适用于 AI 模型训练碳排放核算工作。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 24040 环境管理 生命周期评价 原则与框架

GB/T 24044 环境管理 生命周期评价 要求与指南

GB/T 32150 工业企业温室气体排放核算和报告通则

3 术语和定义

GB/T 24040、GB/T 24044、GB/T 32150界定的以及下列术语和定义适用于本文件。

3.1

模型训练 model training

利用训练数据，基于机器学习算法，确定或改进机器学习模型参数的过程。

[来源：GB/T 41867-2022，3.2.18]

3.2

AI 模型训练碳排放 carbon emission of AI model training (CEAIT)

基于生命周期评价，以CO₂e表示的AI模型系统训练产生的温室气体排放之和。

3.3

功能单位 functional unit

用来作为基准单位量化的产品系统性能。

[来源：GB/T 24044-2008，3.20]

3.4

单元过程 unit process

进行生命周期清单分析时为量化输入和输出数据而确定的最基本部分。

[来源：GB/T 24044-2008，3.34]

4 缩略语

以下缩略语适用于本文件

AI 人工智能 (Artificial Intelligence)

CPU 中央处理器 (Central Processing Unit)

FPGA 现场可编程门阵列 (Field Programmable Gate Array)

GPU 图形处理器 (Graphics Processing Unit)

RAM 随机存取存储器 (Random Access Memory)

TPU 张量处理器 (Tensor Processing Unit)

GHG 温室气体 (Greenhouse Gas)

5 评估要求

5.1 核算范围

在确定AI模型训练碳排放核算范围过程中，应考虑并描述包括但不限于下列各项：

- a) 模型（系统）范围：明确模型名称、结构、功能、功能单位和系统边界；
- b) 时间范围：选择核算碳排放训练相关的时间段；
注：与产品生命周期中具体单元过程相关的温室气体排放和清除随时间变化，选择的时间范围应可以确定产品生命周期中温室气体排放和清除的平均值，如：季节性生产的产品应覆盖产品生产的整个时间周期，不能仅使用部分时间段的数据进行核算。
- c) 温室气体范围：二氧化碳（CO₂）、甲烷（CH₄）、氧化亚氮（N₂O）、氢氟碳化合物（HFCs）、全氟碳化合物（PFCs）、六氟化硫（SF₆）和三氟化氮（NF₃）。

5.2 系统边界

核算AI模型训练碳排放应核算产品在模型训练阶段（离线训练、在线训练）的温室气体排放。

5.2.1 AI 模型离线训练阶段

AI模型离线训练阶段应收集与以下单元过程相关的数据：

- a) CPU、GPU运行功率；
- b) AI模型训练线程在CPU、GPU的占用率；
- c) AI模型训练的内存开销；
- d) AI模型训练的CPU、GPU、内存的使用状态、使用时间；
- e) AI模型训练其它相关的过程。

5.2.2 AI 模型在线训练阶段

AI模型在线训练阶段应收集以下数据：

- a) AI模型在云平台训练时各种设备的使用状态的能效；
- b) AI模型训练在云平台各种设备的使用时间；
- c) 云平台与AI模型训练使用电力相关的 GHG 排放量。

5.3 AI 模型碳排放计算涉及参数

5.3.1 模型相关参数

模型相关参数指与设备无关的算法层、软件层、编译层相关参数、模型算力需求和模型探索准备等相关参数。

- a) 与设备无关参数
 - 1) 模型参数数目
 - 2) 模型结构（深度、layer 特征、操作）
 - 3) 模型实现
 - 4) 模型编译方法
 - 5) 模型优化方法
 - 6) 模型底层 kernel 库
 - 7) 模型运行时间
 - 8) 数据规模
 - 9) 数据类型
 - 10) 网络通信开销
 - 11) 学习方法及优化器

- b) 模型算力需求
 - 1) 模型浮点数计算总量（训练）
 - 2) 模型浮点数计算总量（1次前向传播）
 - 3) 模型训练时间
 - 4) 模型推理时间（1次前向传播）
 - 5) 模型训练步数（steps）
 - 6) 模型每小时训练步数（steps/hours）
- c) 模型探索相关参数
 - 1) 模型数据获取工作
 - 2) 模型数据转移工作
 - 3) 模型探索时间
 - 4) 模型探索规模
 - 5) 模型结构探索技术（NAS）
 - 6) 模型实验阶段工作

5.3.2 设备相关参数

设备相关参数指与模型无关参数（model-agnostic factors），具体可分为设备统计信息和设备运行信息两类参数。

- a) 设备统计信息
 - 1) 设备名称：CPU、GPU、主存、网络、风扇、主板等
 - 2) 设备架构
 - 3) 设备数目
 - 4) 设备额定功率
 - 5) 设备标准差
 - 6) 设备理论算力（每秒钟可执行的浮点数计算总量，TFLOPS/s）
 - 7) 设备最大寿命
- b) 设备运行信息
 - 1) 设备最大功率
 - 2) 设备平均功率
 - 3) 设备运行功率（当前时刻设备功率，相关硬件驱动提供结果）
 - 4) 设备运行时间
 - 5) 设备占用率
 - 6) 设备运行算力（每秒钟真实执行的浮点数计算总量，TFLOPS/s）

5.3.3 数据中心相关参数

数据中心相关参数指超过单机硬件设备规模的相关参数（beyond one machine）；依照相关参数的控制主体，可分为数据中心内部参数、数据中心外部参数两类。

- a) 数据中心内部参数
 - 1) 数据中心能效比（PUE）
 - 2) 数据中心制冷系统
 - 3) 数据中心供能系统
 - 4) 数据中心外购电量
 - 5) 数据中心外购热量
 - 6) 设备最大寿命
 - 7) 数据中心总消耗能量当量（电能、热能、制冷、排水等）
 - 8) 云计算开销

- 9) 云计算电力消耗当量
- b) 数据中心外部参数
 - 1) 数据中心碳排放强度(carbon intensity)单位为千克二氧化碳当量每千瓦时(kgCO₂e/KWh)
 - 2) 数据中心绿色能源当量(carbon free energy, beyond scope 2), 超出世界资源研究所(World Resource Institute)定义的绿色温室气体排放协议GHG Protocol所定义的第二范围(scope 2)能源消耗。提供精确实时测量、地理位置信息、本地能源购买等优化。
 - 3) 数据中心绿色能源比(绿色能源当量与全部能源消耗当量比)
 - 4) 数据中心碳抵消当量(carbon offset, under scope 2)
 - 5) 数据中心碳抵消比(使用碳抵消技术的耗电量与全部电力消耗量之比)

5.4 数据质量要求

数据收集与处理过程中, 相关数据应满足以下数据质量要求:

- a) 技术代表性: 数据反映实际技术情况, 即体现实际模型训练流程、技术和设备类型、原料与能耗类型、训练规模等因素的影响;
- b) 时间代表性: 数据反应训练各单元过程的实际时间;
- c) 地理代表性: 排放因子等相关参数的选择考虑单元过程所处的地理位置;
- d) 数据完整性: 按照数据取舍准则, 判断是否已收集各训练过程的主要消耗和排放数据, 尽可能避免数据缺失, 缺失的数据需在报告中说明;
- e) 数据准确性: 所有数据均有相关的数据来源和数据处理算法; 估算或引用文献的数据需在报告中说明;
- f) 数据收集原则: 活动水平数据优先采用直接计量、测量获得的原始数据, 其次采用通过原始数据折算获得的二次数据, 以上数据均不可获得时可采用来自相似单元过程的替代数据。使用阶段可使用统计数据、设计数据或估算数据。

5.5 数据取舍准则

在模型训练碳排放核算过程中, 需明确数据取舍准则, 舍弃碳排放影响较小的因素, 简化数据收集过程。例如, 模型训练过程中人员产生的温室气体排放可舍弃。

6 评估方法

6.1 AI 模型训练碳排放

AI模型训练碳排放的核算采用排放因子法, 应包括离线训练阶段和在线训练阶段涉及的所有单元过程, 计算见公式(1)。

$$CEAIT = (E_{\text{离线训练}} + E_{\text{在线训练}}) \times PUE_{\text{数据中心}} \times IC_{\text{碳排放强度}} \dots\dots\dots (1)$$

式中:

CEAIT——AI 模型训练碳排放, 单位为千克二氧化碳当量 (kgCO₂e);

E_{离线训练}——AI 模型使用本地机器进行离线训练阶段电力消耗量, 单位为千瓦时 (kWh);

E_{在线训练}——AI 模型用云计算进行在线训练阶段电力消耗量, 单位为千瓦时 (kWh);

PUE_{数据中心}——数据中心电源使用效率, 为大于 1 的常数, 无单位;

IC_{碳排放强度}——单位产品产量(产值)或服务量的碳排放量, 单位为千克二氧化碳当量每千瓦时 (kgCO₂e/kWh)。

注: AI模型的全生命周期可能会包含多个单元过程, 如数据预处理、模型探索、模型训练、模型使用推理、模型量化压缩等生命周期各阶段的温室气体排放。其中, 模型训练阶段是深度学习模型研发阶段的主要温室气体排放时间, 本文件确定的AI模型训练碳排放核算的系统边界仅包含上述阶段。

6.2 AI 模型训练电力消耗量计算

AI模型训练阶段电力消耗量应包括模型离线训练和模型在线训练阶段涉及的所有单元过程，计算见公式（2）。

$$E_{\text{训练}} = E_{\text{离线训练}} + E_{\text{在线训练}} \dots\dots\dots (2)$$

式中：

$E_{\text{训练}}$ ——模型训练阶段电力消耗量，单位为千瓦时（kWh）；

$E_{\text{离线训练}}$ ——模型使用本地机器离线训练阶段电力消耗量，单位为千瓦时（kWh）；

$E_{\text{在线训练}}$ ——模型使用云计算、分布式计算等算力提供商进行在线训练阶段电力消耗量，单位为千瓦时（kWh）。

AI 模型离线训练阶段电力消耗量的计算应包括模型离线训练过程中所有显著硬件设备活动单元电力消耗量。AI 模型离线训练包含多个多种硬件设备时，如 CPU、AI 芯片、GPU、FPGA、TPU、智能网卡、主板、硬盘等，深度学习模型离线训练阶段碳排放核算工作主要涉及 CPU、GPU、RAM 三类设备电力消耗，计算见公式（3）与公式（4）。

$$E_{\text{离线训练}} = E_{\text{CPU}} + E_{\text{GPU}} + E_{\text{RAM}} \dots\dots\dots (3)$$

$$E_{\text{离线训练}} = \sum_{i=\text{设备类型}} (n_i \times p_{\text{线程占用率}} \times w_{\text{设备运行功率}} \times t_{\text{设备时间}}) \div 1000 \div 3600 \dots\dots\dots (4)$$

式中：

E_{CPU} ——AI模型离线训练阶段CPU电力消耗量，单位为千瓦时（kWh）；

E_{GPU} ——AI模型离线训练阶段GPU电力消耗量，单位为千瓦时（kWh）；

E_{RAM} ——AI模型离线训练阶段RAM电力消耗量，单位为千瓦时（kWh）；

$p_{\text{线程占用率}}$ ——AI模型离线训练阶段线程p的设备占用率；

$w_{\text{设备运行功率}}$ ——设备运行时的功率，不同于设备额定功率，单位为瓦特（W）；

$t_{\text{设备时间}}$ ——设备运行时间，单位为秒（s）；

n_i ——设备*i*的个数；

i ——设别类型。

AI模型在线训练阶段电力消耗量的计算应考虑云服务提供商的系统功率、云服务时间与云服务系数，计算见公式（5）。

$$E_{\text{在线训练}} = w_{\text{云服务系统功率}} \times t_{\text{云服务时间}} \times r_{\text{云服务系数}} \div 1000 \div 3600 \dots\dots\dots (5)$$

$w_{\text{云服务系统功率}}$ ——AI模型在线训练阶段云服务提供商计算系统平均功率，单位为瓦特（W）；

$t_{\text{云服务时间}}$ ——AI模型在线训练阶段云服务提供商计算系统服务时间，单位为秒（s）；

$r_{\text{云服务系数}}$ ——AI模型在线训练阶段占用云服务计算系统比率。

AI模型训练碳排放评估示例可参考附录A。

附录 A

(资料性)

某 AI 模型训练碳排放评估简要示例

A.1 基本信息

A.1.1 AI模型基本信息

模型名称：面向图片分类的卷积神经网络模型 ResNet

模型结构：卷积神经网络、池化层、归一化层、全连接层.....

模型名称：XX-XX-XX

模型功能：图片分类、特征提取等。

A.1.2 硬件设备基本信息

CPU：Intel i7 Gen10 CPU4.7GHz×8

GPU：NVidia GTX3060×2

内存：16GB

其他：硬盘、主板、电源等

A.1.3 联系人基本信息

联系人：XXX

电话：XXX-XXXXXXXX

Email：xxxx@xxx.com

A.2 概述

A.2.1 核算范围

产品范围：本报告的核算对象为 XX 有限公司生产的模型名称为 XX-XX-XX 的 AI 模型（训练后）。

时间范围：本报告选取 20XX 年作为产品碳排放核算的核算期。

温室气体范围：二氧化碳（CO₂）、甲烷（CH₄）、氧化亚氮（N₂O）、氢氟碳化合物（HFCs）、全氟碳化合物（PFCs）、六氟化硫（SF₆）和三氟化氮（NF₃）。

核算依据：依据《人工智能模型训练碳排评估方法》进行产品碳排放核算和核算报告编制。

A.2.2 功能单位

本报告以 1 个模型名称为 XX-XX-XX 的 AI 模型（训练后）为功能单位。

A.2.3 系统边界

本报告核算 AI 模型产品在训练阶段的温室气体排放。系统边界示意图略。

A.3 数据收集与处理

A.3.1 离线训练阶段数据收集与处理

某 AI 模型产品训练阶段始于模型参数初始化，经过一系列梯度计算和梯度反向传递过程，到模型参数与训练数据拟合完毕，AI 模型训练结束。训练阶段的清单数据主要包含训练过程中的能源消耗。某 AI 模型产品训练过程中相关硬件设备使用信息如表 A.1 所示。

表A.1 某 AI 模型产品训练过程中相关硬件设备使用信息

编号	部件名称	数量	平均设备运行功率	平均设备运行时长	平均设备占用率
1	CPU	8	100 瓦	120 小时	45%
2	GPU	2	60 瓦	100 小时	95%
3	内存	1	40 瓦	120 小时	20%

A.3.2 在线训练数据收集与处理

经调研（略），某 AI 模型产品训练阶段未使用在线训练信息如表 A.2 所示。

表A.2 某 AI 模型产品训练阶段未使用在线训练信息

编号	服务商	运行功率	运行时长	云服务系数
1	XXX	200 瓦	120 小时	90%
2	XXX	240 瓦	50 小时	95%

A.4 核算过程和结果

A.4.1 离线训练阶段温室气体核算

某 AI 离线模型训练阶段温室气体排放主要为电力消耗产生的温室气体排放，包括典型能源消耗（typical energy consumption, TEC，单位为 kWh）。计算方法见公式（A.1），公式（A.2）

$$E_{\text{离线训练}} = E_{\text{CPU}} + E_{\text{GPU}} + E_{\text{RAM}} \dots\dots\dots (A.1)$$

$$E_{\text{离线训练}} = \sum_{i=\text{设备类型}} (n_i \times p_{\text{线程占用率}} \times w_{\text{设备运行功率}} \times t_{\text{设备时间}}) \div 1000 \dots\dots\dots (A.2)$$

式中：

- E_{CPU} : 离线训练阶段 CPU 电力消耗量, 单位为千瓦时 (kWh);
- E_{GPU} : 离线训练阶段 GPU 电力消耗量, 单位为千瓦时 (kWh);
- E_{RAM} : 离线训练阶段 RAM 电力消耗量, 单位为千瓦时 (kWh);
- p_{CPU} : 用于 AI 模型训练的 CPU 设备占用率, 依上表取 45%;
- p_{GPU} : 用于 AI 模型训练的 GPU 设备占用率, 依上表取 95%;
- p_{RAM} : 用于 AI 模型训练的内存设备占用率, 依上表取 20%;
- w_{CPU} : CPU 运行时平均功率, 单位为瓦 (W), 依上表取 100W;
- w_{GPU} : GPU 运行时平均功率, 单位为瓦 (W), 依上表取 60W;
- w_{RAM} : 内存运行时平均功率, 单位为瓦 (W), 依上表取 40W;
- t_{CPU} : CPU 运行时长, 依上表取 120h;
- t_{GPU} : GPU 运行时长, 依上表取 100h;
- t_{RAM} : 内存运行时长, 依上表取 120h。

则本模型的离线训练能源消耗为:

$$E_{\text{离线训练}} = (8 \times 0.45 \times 100W \times 120h) + (2 \times 0.95 \times 60W \times 100h) + (1 \times 0.2 \times 40W \times 120h) / 1000$$

$$= 444.36kWh$$

A. 4. 2 在线训练阶段温室气体核算

某AI依托云服务提供商, 进行在线模型训练, 其温室气体排放主要为电力消耗产生的温室气体排放。计算方法见公式 (A.3)

$$E_{\text{在线训练}} = w_{\text{云服务系统功率}} \times t_{\text{云服务时间}} \times r_{\text{云服务系数}} \div 1000 \dots\dots\dots (A.3)$$

式中:

- $w_{\text{云服务系统功率}}$: AI 模型在线训练阶段云服务提供商计算系统平均功率, 单位为瓦特 (W);
- $t_{\text{云服务时间}}$: AI 模型在线训练阶段云服务提供商计算系统服务时间;
- $r_{\text{云服务系数}}$: AI 模型在线训练阶段占用云服务计算系统比率。

则本模型的在线训练能源消耗为:

$$E_{\text{在线训练}} = (200W \times 120h \times 0.9) + (240W \times 50h \times 0.95) / 1000 = 33kWh$$

A. 4. 3 AI模型训练的碳排放

本 AI 模型训练的碳排放为:

$$FAIT = (E_{\text{离线训练}} + E_{\text{在线训练}}) \times PUE_{\text{数据中心}} \times IC_{\text{碳排放强度}}$$

$$= (444.36kWh + 33kWh) * 0.85 * 0.8843tCO_2e/MWh = 358.812kgCO_2e$$

A.5 结论和不确定性说明

本部分包括碳排放核算结果、对模型训练技术优化管理等方面的建议、以及不确定性说明等，具体内容略。

参考文献

- [1] GB 40879—2021 数据中心能效限定值及能效等级
 - [2] GB/T 41867—2022 信息技术 人工智能 术语
 - [3] GB 43630—2023 塔式和机架式服务器能效限定值及能效等级
 - [4] Patterson D, Gonzalez J, Le Q, et al. Carbon emissions and large neural network training[J]. arXiv preprint arXiv:2104.10350, 2021.
 - [5] Wu C J, Raghavendra R, Gupta U, et al. Sustainable ai: Environmental implications, challenges and opportunities[J]. Proceedings of Machine Learning and Systems, 2022, 4: 795-813.
 - [6] Anthony L F W, Kanding B, Selvan R. Carbontracker: Tracking and predicting the carbon footprint of training deep learning models[J]. arXiv preprint arXiv:2007.03051, 2020.
 - [7] Henderson P, Hu J, Romoff J, et al. Towards the systematic reporting of the energy and carbon footprints of machine learning[J]. Journal of Machine Learning Research, 2020, 21(248): 1-43.
 - [8] Lannelongue L, Grealey J, Inouye M. Green algorithms: quantifying the carbon footprint of computation[J]. Advanced science, 2021, 8(12): 2100707.
-