



团 体 标 准

T/TAC 10—2024

机器翻译伦理 要求

Ethics for machine translation—Requirements

2024-03-30 发布

2024-04-30 实施

中 国 翻 译 协 会 发 布
中 国 标 准 出 版 社 出 版

目 次

前言 III

引言 IV

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 基本原则 4

5 设计伦理 4

6 研发伦理 5

7 推广伦理 6

8 应用伦理 7

参考文献..... 8

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。

本文件由中国翻译协会提出并归口。

本文件起草单位：北京外国语大学、中国外文局翻译院、中国标准化研究院、教育部语言文字应用研究所、哈尔滨工业大学、全国科学技术名词审定委员会事务中心、鹏城实验室、北京百度网讯科技有限公司、华为技术有限公司、阿里巴巴(中国)有限公司、科大讯飞股份有限公司、中译语通科技股份有限公司、传神语联网网络科技股份有限公司、中国科学院自动化研究所、上海交通大学、上海海事大学、西安外国语大学、天津外国语大学、北京语智云帆科技有限公司、上海一者信息科技有限公司、成都优译信息技术有限公司、北京外研在线数字科技有限公司、北京博硕星睿教育科技有限公司、北京悦尔信息技术有限公司、新译信息科技(深圳)有限公司、深圳云译科技有限公司、甲骨易(北京)语言科技股份有限公司、上海创凌翻译服务有限公司、四川语言桥信息技术有限公司。

本文件主要起草人：王华树、任文、王飞、邢玉堂、杜楠、梁倩如、王海涛、冯志伟、赵铁军、代晓明、相洋、吴华、赵燕青、杨宝嵩、刘俊华、程国良、何征宇、张家俊、管新潮、刘世界、曹达钦、朱华、魏勇鹏、张井、张马成、商其坤、周东莉、蒙永业、田亮、丁丽、李光华、杨颖波、朱宪超。

引 言

在以大语言模型为代表的人工智能驱动之下,机器翻译技术迅速发展,已在众多领域广泛应用,为深化全球交流与合作创造了有利条件。然而,机器翻译的创新技术也带来了数据隐私、知识产权、信息安全、公平竞争等方面的伦理挑战。我国虽然已制定了多部机器翻译相关的标准,但尚无伦理相关规范或标准。在大力研发、普及、应用机器翻译的同时,加强机器翻译伦理标准体系建设有助于规避设计、研发、推广和应用过程中潜在的伦理风险,规范相关利益者的关系,对推动语言服务行业良性发展具有重要意义。因此,有必要建立一套完善的机器翻译伦理标准体系,指导各方共同遵循,塑造公平、公正、透明的行业生态环境。

本文件的制定按照机器翻译系统的生命周期划分为设计阶段、研发阶段、推广阶段及应用阶段,机器翻译伦理意识需要贯穿始终,旨在确保机器翻译产品在全流程符合相应伦理要求。通过实施本要求,机器翻译产品提供方能够判断其机器翻译技术在设计、研发、推广与应用方面是否符合机器翻译伦理要求;机器翻译需求方能够判断机器翻译产品或服务是否满足伦理要求,安全用于某特定用途。同时,本文件有利于促进行业内具体机器翻译伦理问题的解决,促进国家层面相关伦理标准体系的建设。

机器翻译伦理 要求

1 范围

本文件规定了机器翻译系统在设计、研发、推广及应用4个阶段的相关伦理准则。

本文件旨在供机器翻译系统利益相关者使用,包括但不限于机器翻译研发方(含工程师、设计师)、供应方(含产品/技术推广企业或厂商)、用户方(使用者、产品消费者)、语言数据企业,政策制定者和科研机构。

本文件适用于机器翻译系统的开发、维护和管理过程,旨在确保机器翻译技术的发展和应用能够遵守高标准的伦理准则,并促进其在全球范围内的公平、透明和可持续发展。

2 规范性引用文件

GB/T 5271.1—2000 信息技术 词汇 第1部分:基本术语

GB/T 37988—2019 信息安全技术 数据安全能力成熟度模型

GB/T 40036—2021 翻译服务 机器翻译结果的译后编辑 要求

3 术语和定义

GB/T 40036—2021、GB/T 37988—2019 和 GB/T 5271.1—2000 界定的以及下列术语和定义适用于本文件。

3.1 机器翻译相关概念

3.1.1

自然语言 **natural language**

一种其规则是基于当前的用法且无需特别规定的语言。

[来源:GB/T 5271.1—2000,01.05.08]

3.1.2

机器翻译 **machine translation; MT**

一种使用计算机程序将文本从一种自然语言(3.1.1)自动翻译成另一种自然语言(3.1.1)的技术。

[来源:GB/T 40036—2021,3.1.1,有修改——补充“一种”和“的技术”]

3.1.3

数据 **data**

信息的可再解释的形式化表示,以适用于通信、解释或处理。

[来源:GB/T 5271.1—2000,01.01.02]

3.1.4

机器翻译数据 **machine translation data**

用于机器翻译系统训练、验证和测试等过程的数据(3.1.3)。