



中国翻译协会标准

T/TAC X—XXXX

机器翻译伦理规范

Code of Ethics for Machine Translation

在提交反馈意见时，请将您知道的相关专利与支持性文件一并附上。

(征求意见稿)

2023-11-6

XXXX - XX - XX 发布

XXXX - XX - XX 实施

中国翻译协会 发布

目 次

前 言	III
引 言	IV
机器翻译伦理规范	1
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
3.1 机器翻译相关概念	1
3.2 算法相关概念	2
3.3 伦理规范相关概念	2
3.4 机器翻译系统性能相关概念	3
3.5 利益相关者相关概念	4
4 基本原则	4
4.1 技术向善	4
4.2 公平公正	4
4.3 可靠可控	4
4.4 遵守法律法规	4
4.5 维护国家安全与公共利益	5
5 设计伦理	5
5.1 数据安全与隐私保护	5
5.2 公平性与无偏性	5
5.3 包容性与多样性	5
5.4 无障碍性与适应性设计	5
5.5 用户利益与责任	5
5.6 可持续性与环境友好	5
6 研发伦理	5
6.1 算法伦理	5
6.2 数据伦理	6
7 推广伦理	7
7.1 公平竞争	7
7.2 宣传真实情况	7
7.3 遵循行业规范	7
7.4 保护知识产权	7
7.5 加强用户教育	7
7.6 社会责任与可持续性	7
8 应用伦理	8
8.1 合规性与法律遵从	8
8.2 合理收费	8

8.3 人机协作与责任界定8

8.4 缩小数据鸿沟和技术鸿沟8

8.5 持续改进与用户反馈8

参考文献9

征求意见专用

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。

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、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）网络技术有限公司、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、中国标准化研究院、全国科技名词委、中科院自动化所、上海交通大学、上海海事大学、北京语智云帆科技有限公司、北京博硕星睿教育科技有限公司、上海一者信息科技有限公司、成都优译信息技术有限公司、新译信息技术有限公司、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、四川语言桥信息技术有限公司。

本文件主要起草人：暂略。

征求意见专用

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，为全球交流和合作创造了条件。然而，机器翻译的创新技术也带来了伦理挑战，包括数据隐私、知识产权、信息安全、公平竞争等方面的问题。我国虽然已制定了多个机器翻译相关的标准，但尚无相关伦理规范或标准。在大力研发、普及应用机器翻译的同时，加强机器翻译的人工智能伦理规范有助于规避研发、推广和应用过程中的伦理问题，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。为此，有必要建立一套完善的机器翻译伦理规范体系，指导各方共同遵循道德准则，塑造一个公平、公正、透明的行业生态环境。

机器翻译系统的生命周期通常可以划分为以下四个阶段：（1）设计阶段 (Design Phase)：在设计阶段，需要确定产品的目标、功能和需求，基于应用场景和目标人群确定治理策略，以便为后续研发阶段奠定基础；（2）研发阶段 (Research and Development Phase)：研发阶段主要关注机器翻译系统的开发、训练和优化。在这个阶段，需要处理大量的数据，选择合适的模型和算法，以实现高质量的翻译性能。同时，需要关注伦理、安全和隐私等问题，确保产品在技术和道德层面上负责任地开发；（3）推广阶段 (Promotion Phase)：推广阶段的目标是将机器翻译产品推向市场，吸引潜在用户。在这个阶段，需要制定有效的市场推广策略，包括宣传、广告、商业合作等。同时，应确保推广活动遵循相关法规、行业标准和道德原则，维护市场公平竞争；（4）应用阶段 (Application Phase)：应用阶段关注机器翻译产品在实际场景中的使用效果。在这个阶段，厂商需要关注用户体验和需求，收集反馈，持续改进和优化产品，确保产品质量和服务水平满足用户的期望。在这四个阶段中，伦理问题需要贯穿始终，需要确保机器翻译产品在全流程中都能符合伦理和规范。

机器翻译伦理规范

1 范围

本标准规范了机器翻译在设计、研发、推广及应用等四个阶段的相关伦理准则，旨在供机器翻译技术利益相关者使用，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）、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。

通过实施本标准，机器翻译提供方可以判断其机器翻译技术在设计、研发、推广与应用等方面是否符合机器翻译伦理规范要求；机器翻译需求方可以判断机器翻译是否能够满足伦理规范，安全用于某特定用途。同时，本标准有利于促进行业领域内具体机器翻译伦理问题的解决，促进国家层面相关伦理规范与标准的建设。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅注日期的版本适用于本文件。凡是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 40036—2021 翻译服务 机器翻译结果的译后编辑 要求

ISO 18587: 2017 Translation services — Post-editing of machine translation output — Requirements

GB/T 37988—2019 信息安全技术 数据安全能力成熟度模型

GB/T 5271.1—2000 信息技术 词汇 第1部分：基本术语

3 术语和定义

GB/T 40036—2021、ISO 18587: 2017、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/ISO 18587: 2017，3.1.1，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

用于在机器翻译系统训练和评测等过程的数据。

3.1.5

模型 model

用于描述或预测现象或过程的抽象或简化表示。

3.1.6

训练 training

用于优化**模型** (3.1.5) 参数的过程，通常基于大量的**数据** (3.1.3)。

3.2 算法相关概念

3.2.1

算法 algorithm

为解决问题严格定义的有限的有序规则集。

[来源：GB/T 5271.1—2000, 01.05.05]

3.2.2

算法歧视 algorithmic discrimination

描述机器翻译系统对某些语言、文化或群体进行不公正或不平等对待的现象。

3.2.3

算法偏见 algorithmic bias

描述机器翻译系统对某些语言、文化或群体进行有利或不利的倾向的现象。

3.2.4

算法滥用 algorithmic abuse

描述机器翻译系统被恶意使用或滥用，导致损害他人或社会利益的现象。

3.3 伦理规范相关概念

3.3.1

伦理规范 code of ethics

在特定领域中所遵循的道德准则和规范。

3.3.2

机器翻译伦理 machine translation ethics

机器翻译系统设计、研发、推广及应用过程中所遵循的道德准则。

3.3.3

设计伦理 design ethics

机器翻译系统在设计时遵循的**伦理规范** (3.3.1) 。

3.3.4

算法伦理 algorithm ethics

利用**算法** (3.2.1) 研发机器翻译系统时遵循的**伦理规范** (3.3.1) 。

3.3.5

数据伦理 data ethics

机器翻译系统在**数据** (3.1.3) 采集、处理和应用时遵循的**伦理规范** (3.3.1) 。

3.3.6

应用伦理 application ethics

机器翻译系统在设计、应用、宣传和普及时遵循的**伦理规范** (3.3.1) 。

3.4 机器翻译系统性能相关概念

3.4.1

可解释性 explainability

机器翻译系统的内部运作逻辑可以被理解和解释的特性。

3.4.2

可追溯性 traceability

机器翻译系统在处理过程中的透明度和可审查性。

3.4.3

公平性 fairness

机器翻译系统对不同语言、文化和群体进行公平对待的特性。

3.4.4

包容性 inclusiveness

机器翻译系统具备对全球各种语言（含方言）的理解和翻译能力，尊重并考虑到各种民族、文化、习俗和社会背景的多样性，满足不同用户群体翻译需求的特性。

3.4.5

可靠性 reliability

机器翻译系统提供准确和稳定的机器翻译输出，满足用户特定的服务需求的特性。

3.4.6

安全性 security

机器翻译系统防止恶意攻击和**数据** (3.1.3) 泄露，确保**数据** (3.1.3) 和信息安全的特性。

3.4.7

可审计性 auditability

机器翻译系统能够被有效地监督和审查，以确保其符合道德准则和规范的特性。

3.4.8

无障碍性 accessibility

机器翻译系统能够确保所有人都能方便地访问和使用的特性。

3.5 利益相关者相关概念

3.5.1

利益相关者 stakeholders

数据制造者、数据服务提供者、数据消费者等在**机器翻译数据** (3.1.4) 处理、应用及监管过程中的参与方。

3.5.2

数据安全 data security

通过管理和技术措施，确保**利益相关者** (3.5.1) 的**机器翻译数据** (3.1.4) 被有效保护和合规使用的一系列管理和技术措施。

[来源：GB/T 37988—2019, 3.1, 有修改——将“数据”修改为“利益相关者的机器翻译数据”，将“有效保护和合规使用的状态”修改为“被有效保护和合规使用的一些列管理和技术措施”]

3.5.3

数据权利 data rights

利益相关者 (3.5.1) 在**机器翻译数据** (3.1.4) 处理过程中享有的权益，如数据采集权、数据使用权、数据收益权、数据可携权、数据知情同意权、数据修改权、数据被遗忘权等。

4 基本原则

4.1 技术向善

机器翻译技术应服务于提升人类福祉，促进社会进步与和平，推动文化交流与理解，促进全球合作与发展。

4.2 公平公正

机器翻译技术应关注数据与算法的公平性，避免歧视与偏见，确保对不同语言和文化的平等对待。

4.3 可靠可控

机器翻译技术应确保整个翻译系统的准确性与一致性，同时允许用户在必要时对系统进行干预和调整，实现系统的可控性。

4.4 遵守法律法规

机器翻译技术应严格遵守国家法律法规，尊重社会公德、公序良俗，尊重并维护国际法律准则。

4.5 维护国家安全与公共利益

机器翻译技术应在尊重国家安全与公共利益的前提下开展研究与应用，防范潜在安全风险。

5 设计伦理

设计伦理指机器翻译系统在设计时需要遵循的伦理规范。机器翻译系统设计须秉承向善的科技原则，增强安全透明，服务于跨文化交流和对外话语体系建设。同时在设计开发与开发时应充分考虑到语言、区域、文化、用户群体特征，避免引发偏见。

5.1 数据安全与隐私保护

机器翻译系统设计应遵循相关法律法规，确保数据的安全存储、处理和传输，同时保护用户隐私，防止数据泄露和滥用。

5.2 公平性与无偏性

机器翻译系统设计应注重平等对待各种语言、文化和群体，防止潜在的歧视和偏见。

5.3 包容性与多样性

机器翻译系统的设计应关注不同用户的包容性和多样性需求，提高对多元用户需求的适应性，服务于更广泛的用户群。

5.4 无障碍性与适应性设计

机器翻译系统设计应充分考虑不同用户群体的需求，包括视觉、听力、认知等方面的障碍，提供无障碍性设计和多样化的交互方式，确保广泛的用户能够方便地使用和享受机器翻译服务。

5.5 用户利益与责任

机器翻译系统设计应关注不同用户群体的需求和利益，按照不同的等级服务标准提供相应的服务，满足用户在不同场景下的翻译需求。同时，在设计过程中明确机器和人的责任划分，出现问题时可追溯责任归属，维护用户的权益。

5.6 可持续性与环境友好

机器翻译系统设计应充分考虑资源和能源的有效利用，降低环境负担，支持可持续发展，为绿色科技的推广和应用贡献力量。

6 研发伦理

在研发阶段，机器翻译伦理主要关注如何确保机器翻译技术的限定开发和评测，遵循相关法规、行业标准和道德规范。

6.1 算法伦理

算法伦理指研发机器翻译技术时引发的算法安全、算法歧视、算法偏见、算法滥用等伦理风险。机器翻译技术在全过程中须避免因算法引发的相关问题，确保算法的安全性，避免因算法泄露给技术研发方、技术拥有方、技术使用方带来财产和人身安全问题，并引发一系列法律追责困境。

6.1.1 算法安全性

机器翻译算法应确保算法不会被恶意攻击、篡改或滥用，从而避免对技术研发方、技术拥有方和技术使用方造成财产损失或人身安全风险。

6.1.2 算法可靠性

机器翻译算法应确保系统能够以稳定的响应速度和质量持续地提供符合预期的服务或实现预定的算法目标，保证系统性能和输出结果的准确性与稳定性。

6.1.3 算法可解释性

机器翻译算法应力求逻辑清晰和易于理解，增强用户对算法和系统的信任度。

6.1.4 算法可审计性

机器翻译算法应可被委托方检查、追溯和验证，以评估其合规性，监督其符合道德标准和规范的程度。

6.1.5 算法歧视与偏见

机器翻译算法开发过程中应关注数据集的多样性和平衡性，避免因数据不均衡导致歧视、偏见和滥用现象发生，确保公平对待各种语言和文化背景。

6.1.6 算法责任和问责

机器翻译算法开发者应明确算法的责任归属，确保在算法出现问题时能够追溯责任并采取相应措施。同时，应建立相应的问责机制，确保算法的合规性和道德性得到有效监管。

6.2 数据伦理

数据伦理指在机器翻译系统研发、训练、测试及应用过程中产生的一系列不规范数据行为，涉及到利益相关者（数据制造者、数据服务提供者、数据消费者等）的数据安全、数据隐私、数据权利、数据滥用等。机器翻译研发方在数据采集、数据训练、数据测试等过程中须遵循国际通用伦理规范，确保数据安全和互操作性，以避免产生数据垄断、数据霸权现象，影响数据提供方或机器翻译用户方的数据产权及数据安全。同时，机器翻译技术的数据伦理还涉及到所采集和训练所用数据内容的真实性、可靠性、易用性、权威性等多方面。此外，机器翻译数据接口、数据储存、数据传输设备技术的规范性也是机器翻译技术的数据伦理需要考虑的范围。

6.2.1 数据隐私与保护

在机器翻译的数据采集、训练、评测和应用过程中，应重视利益相关者的数据隐私和保护，防止未经授权的访问、泄露或滥用。同时，应遵循相关法律法规和行业标准，明确告知用户数据收集、使用和共享的范围、目的和方式，尊重用户的数据采集权、数据使用权、数据收益权、数据可携权、数据知情同意权、数据修改权、数据被遗忘权等。

6.2.2 数据来源合规性

在机器翻译的数据采集、训练、评测和应用过程中，应确保数据来源合法合规、处理方法妥当、使用范围明确，充分尊重数据制造者的权益，避免使用侵权或非授权数据。

6.2.3 数据质量与可靠性

在机器翻译的数据采集、训练、评测和应用过程中，应确保用于翻译引擎训练的数据具有较高的质量，以提高系统的准确性和可信度。

6.2.4 数据多样性与代表性

在机器翻译的数据采集、训练、评测和应用过程中，应确保数据具有广泛的领域覆盖和丰富的样本种类，以确保机器翻译系统的公平、公正和包容。针对低资源语言，应综合运用多种方法和策略，包括但不限于数据增强、预训练模型以及数据共享和合作等，确保低资源语言数据的多样性和代表性，从而改善低资源语言的机器翻译性能。

6.2.5 数据存储、使用与传输

机器翻译技术应遵循数据接口、数据储存、数据使用和数据传输设备的技术规范，确保数据的安全存储、合规使用以及安全传输，防止数据泄露、篡改或滥用，保障数据安全和完整性。

6.2.6 数据共享与开放

机器翻译的数据采集和使用应遵循开放和共享原则，在保障隐私和安全并符合法律法规要求的前提下，与相关利益方共享数据，促进数据的可获取性和互操作性。

7 推广伦理

在机器翻译的推广阶段，主要关注如何确保产品被负责任地推广和应用，遵循相关法规、行业标准和道德原则。

7.1 公平竞争

机器翻译产品在推广过程中，厂商应秉持公平竞争的原则，遵循市场规则，不进行恶意诋毁竞争对手、低价竞争等不正当竞争行为。

7.2 宣传真实情况

机器翻译产品在推广过程中，厂商应确保宣传内容的真实性，避免过度夸大产品性能、误导用户，确保消费者能够充分了解产品的优点和局限性，对技术能力和服务有正确的认知。

7.3 遵循行业规范

机器翻译产品在推广过程中，应遵循相关行业规范和标准，确保产品在市场推广、广告宣传和商业合作等方面的合法合规性。

7.4 保护知识产权

机器翻译产品在推广过程中，厂商应尊重和保护相关知识产权，确保不侵犯他人的专利权、著作权、商标权、发明权等知识产权。

7.5 加强用户教育

厂商应承担用户教育的责任，提供易于理解的产品使用教程和案例，帮助用户了解机器翻译系统的正确使用方法，以及如何避免因误用而产生的潜在风险。

7.6 社会责任与可持续性

厂商在推广机器翻译产品时，应关注其对社会、环境和经济的影响，承担社会责任，确保产品在推广过程中符合社会伦理和可持续发展目标。

8 应用伦理

应用伦理指机器翻译技术在应用过程中产生的伦理规范问题，涵盖了诸多方面，包括但不限于技术的合规性、资费合理性、安全性以及用户权益等。

8.1 合规性与法律遵从

机器翻译系统在各行业应用中应遵守国家法律法规和行业标准，确保合规性和道德性，防范法律风险。

8.2 合理收费

厂商应设定合理的收费标准，维护市场秩序。

8.3 人机协作与责任界定

在复杂应用场景中，厂商应明确机器翻译系统与人的责任边界，强调人机协作，确保翻译质量和遵守道德规范。用户应对机器翻译结果进行进一步的质量检查，确保人在AI使用过程中的主导作用。

8.4 缩小数据鸿沟和技术鸿沟

厂商应关注数据鸿沟和技术鸿沟问题，推动技术普及和知识传播，为更多用户在数据和技术获取及应用方面提供平等的机会和资源。

8.5 持续改进与用户反馈

厂商应鼓励用户提供反馈，关注系统在实际应用中的问题和挑战，持续改进技术性能，满足不断变化的行业需求。

参 考 文 献

- [1] ISO 18587: 2017 Translation services — Post-editing of machine translation output — Requirements
- [2] 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
- [3] GB/T 40036—2021 翻译服务 机器翻译结果的译后编辑 要求
- [4] GB/T 25000.10—2016 系统与软件工程 系统与软件质量要求和评价（SQuaRE）第10部分：系统与软件质量模型
- [5] GB/T 35273—2020 信息安全技术 个人信息安全规范
- [6] GB/T 37988—2019 信息安全技术 数据安全能力成熟度模型
- [7] GB/T 5271.1—2000 信息技术 词汇 第1部分：基本术语
- [8] GF 2006 机器翻译系统评测规范
- [9] 任文. 机器翻译伦理的挑战与导向[J]. 上海翻译, 2019(5): 46-52+95
- [10] 王华树, 刘世界. 大数据时代翻译数据伦理研究：概念、问题与建议[J]. 上海翻译, 2022(2): 12-17
- [11] 肖冬梅, 文禹衡. 数据权谱系论纲[J]. 湘潭大学学报（哲学社会科学版）, 2015(6): 69 – 75
- [12] 中华人民共和国科学技术部 新一代人工智能伦理规范