基于大语言模型的立法文本可读性研究

胡钦谙
中国社会科学院语言研究所 / 北京
中国社会科学院语言学重点实验室 / 北京
huqa@cass.org.cn

崔玉珍†
中国政法大学人文学院 / 北京
cuiyuzhen@126.com

摘要

我国现行有效法律在内容上所呈现出的多样性及其庞大体量，使得人工方式的可读性评估难以实现全面覆盖。本研究采用大语言模型对立法文本可读性进行评估，以深度学习的端到端方式摆脱了可读性研究对传统语言特征工程的路径依赖。研究表明，大语言模型对立法文本可读性的自动化评分与人工评分具有显著相关性。本研究从部门法等维度出发，系统揭示了不同法律的显著特征差异，刻画了我国现行有效法律文本可读性的整体面貌；并通过大语言模型文本生成与人工校验，从法律适用的角度探讨了提升立法文本可读性的可能路径，为立法语言的优化提供参考。

关键词： 立法文本；大语言模型；可读性；改写

Research on Legislative Text Readability Based on Large Language Models

Qinan Hu
Institute of Linguistics/
Key Laboratory of Linguistics,
Chinese Academy of Social Sciences
Beijing
huqa@cass.org.cn

Yuzhen Cui†
School of Humanities,
China University of Political Science and Law
Beijing
cuiyuzhen@126.com

Abstract

The diversity and substantial volume of China's current effective laws render comprehensive coverage through manual readability assessments impractical. This study employs large language models (LLMs) to evaluate the readability of legislative texts, eliminating the traditional reliance on linguistic feature engineering through an end-to-end deep learning approach. Results demonstrate significant correlations between automated readability scores generated by LLMs and manual assessments. From the perspective of legal branches, this research systematically reveals notable differences in textual characteristics across various laws, delineating the comprehensive landscape

©2025 中国计算语言学大会
根据《Creative Commons Attribution 4.0 International License》许可出版
†为本文通讯作者。

第二十四届中国计算语言学大会论文集，第745页-第755页，济南，中国，2025年8月11日至14日。

(c) 2025 中国中文信息学会计算语言学专业委员会

of readability in China's current legal corpus. Through a dual mechanism combining LLM-generated text and human verification, the study explores potential pathways to enhance legislative text readability from the perspective of legal applicability, providing references for the optimization of legislative language.

Keywords: legislative text , Large Language Models , readability , adaptation

1 引言

法律与语言之间的多维互动关系一直是法学及语言学研究的重要课题。立法文本作为制度实施的重要载体，其可读性（readability）直接影响法律实施效能。立法语言的晦涩不仅会直接影响司法实践中的纠纷解决，还会制约法律解释的实质正义、阻碍法律社会政治功能的实现。对法律文本的阅读障碍会削弱公众对法律的深入理解，进而影响其遵守法律法规的意愿，最终掣肘法治建设的整体进程。因此，提升立法语言可读性不仅是立法质量的保证、司法实践的内在要求，更是实现法律公平正义终极目标的基础。

在法律语言可读性研究中，研究者普遍认为法律语言具备专业性和复杂性。有研究者采用理论建构与案例分析等方法，从影响因素、技巧方式等方面探索提升法律语言可读性[2]的有效路径 (陈佳璇, 2006; 郑金雄, 2011; 崔玉珍, 2016)。尽管这些研究对法律语言可读性提供了有益探索，但多局限于定性层面的理论探讨或零散分析，在系统性定量评估框架的构建方面仍存在明显不足，难以为立法文本的优化提供系统性解答与精准指导。

人工智能技术的突破性进展为解决这一困境提供了思路。以ChatGPT为代表的大语言模型（Large Language Models）“可以从文本的表层序列上发现其背后的层次结构，从而正确地完成基于层次性的造句和理解工作” (袁毓林, 2024, p.10-11)，“在语言运用方面表现优异、功能强大” (袁毓林, 2024, p.3)。大语言模型能够根据自然语言提示（prompt）完成多样化的任务，有望突破现有法律文本可读性研究范式的局限性。

基于上述背景，本研究构建了基于大语言模型的立法文本可读性评估框架，探讨立法语言这一专门领域文本的可读性评估方法、我国现行有效法律可读性的整体面貌与特征，以及提升立法文本可读性的可能路径等问题。

2 大语言模型在立法文本可读性评估中的可行性

2.1 可读性研究的两种范式及其分野

可读性指阅读材料被读者理解的程度 (Dale & Chall, 1949)。传统的可读性研究主要通过人工设计的语言特征，如词频（如Zipf定律）、词长（字符数/音节数）、句法结构（依存距离、嵌套层级）、词语熟悉度等，构建可读性公式 (杨孝溁, 1971; Klare, 2000; Feng et al., 2010; Sung et al., 2015; 吴思远等, 2020)，典型代表包括Flesch-Kincaid指数和SMOG公式。部分研究尝试融入语篇连贯性、修辞结构等深层特征，但受限于量化难度，实际应用仍以表层特征为主。然而，将这些可读性评估方法应用于立法语言时，面临着法律术语密集、嵌套句式复杂等特点，通用领域既有的语言特征及可读性公式的解释力显著降低。

[2]在前人研究中，并不是所有研究都使用了“法律语言可读性”这一术语，有些学者采用了“法律语言易读性”的表达，有些学者则从立法语言的简明化、通俗化或大众化的角度来探讨立法语言可读性的问题。

(c) 2025 中国中文信息学会计算语言学专业委员会

大语言模型为可读性研究开辟了新思路。Trott & Rivière (2024)通过提示工程，引导GPT-4 Turbo依据词汇复杂度、句子结构及整体清晰度等，对CLEAR (Crossley, 2023)语料库中3-12年级学生的英语读物进行可读性评估，结果显示其与人工评分的相关系数高达0.63。Farajidizaji等 (2023)进一步证实，基于零样本提示，ChatGPT与Llama-2能够根据目标可读性级别改写文本。

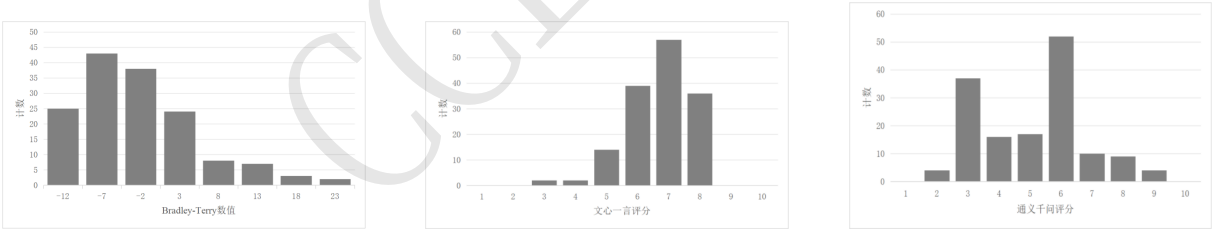
上述两种研究范式的深层次差异在于研究对象的选择。前者聚焦于文本复杂度，而后者则注重文本复杂度与读者认知之间的主客观交互。在传统研究范式下，文本复杂度被视为独立于读者的客观语言属性，其评估依赖于人工设计的语言特征组合，忽略了读者知识背景对文本理解的调节作用。而大语言模型的端到端方式，一方面摆脱了可读性研究对语言特征的路径依赖；另一方面将读者认知能力纳入评估体系。其理论基础可追溯至van Dijk & Kintsch (1983)提出的文本理解模型。该理论指出，文本理解是读者在表层结构、命题（微观/宏观结构）以及读者自身知识背景所形成的情景模型三者间交互建构的过程。大语言模型通过海量语料训练形成的神经表征，精准模拟了这一认知机制。它不仅能解析文本的表层句法结构，还能像读者一样调用、整合知识，构建起对法律术语、逻辑关系的深层情景模型，从而模拟人类在阅读过程中的注意力分配、语境推理等复杂认知机制。这使得大语言模型能够超越语言特征的孤立分析，从更接近人类认知过程的角度进行可读性评估。

2.2 实证研究：人工评估与大语言模型评估的相关性验证

为验证大语言模型在评估立法文本可读性中的可行性，本研究构建了人工-大模型对照实验，检验评估结果的相关性。实验流程如下：

1.语料构建。由法律专业人员从现行有效法律中依典型性选取150个法条句子。

2.人工评估。由2名非法律专业人员对句子进行可读性评估。标注员选择大学本科在校生，以反映公众对立法文本的理解水平。鉴于法条普遍难以理解，成对比较尤为困难。因此，在标注方式上，本研究未采用常用的成对比较方式，而选择以评分的方式标注可读性。评分为1~10分，1分最简单，10分最难。人工评分如图1(a)所示。



(a) 人工评分　(b) 文心一言评分　(c) 通义千问评分

图 1. 可读性评分

3.整合多位标注员评分。根据每位标注员的评分对法条句子依可读性进行两两比较，构建句子间可读性的难易关系，然后通过Bradley-Terry模型对多位标注员的评分结果进行整合。该模型基于概率统计原理，用于分析成对比较数据，可有效降低个体评分偏差，生成稳健的统一序关系，常用于整合多位标注员的可读性评分 (Crossley, 2023)。

整合后的评分如图1(a)所示，法条句子的可读性呈现显著分化，印证立法文本理解障碍的普遍性（Bradley-Terry数值越低，说明在易读性的比较中胜出的可能性越低，即越难理解）。

4.大语言模型评估。在选择大语言模型时，本研究综合考虑了模型性能及其可得性，一方面

(c) 2025 中国中文信息学会计算语言学专业委员会

要求大模型在基准测试中的性能表现领先，另一方面也要求易于访问。基于以上考量，我们采用文心一言(https://yiyan.baidu.com/, ERNIE-4.0-8K)及通义千问(https://tongyi.aliyun.com/, qwen-max)两种具有代表性的大语言模型对法条句子的可读性进行自动评估。通过API调用设置贪婪解码策略，使其输出确定的结果。两个大语言模型的评分如图1所示，(b)为文心一言的评分(c)为通义千问的评分，后者呈双峰分布。输入的提示语及模型输出示例如下。

（1）示例：

【提示语】

请阅读下列句子，综合考虑其所涉及的法律概念、逻辑关系、句子复杂度及相关法律知识，对句子的难易程度进行打分（1-10分，1分最简单，10分最难）。

句子：对有证据证明有犯罪事实，可能判处十年有期徒刑以上刑罚的，或者有证据证明有犯罪事实，可能判处徒刑以上刑罚，曾经故意犯罪或者身份不明的，应当予以逮捕。

【回答】

这个句子的难易程度可以评为7分。解析如下：

1. **法律概念**：句子中涉及了多个法律专业术语，如“证据”、“犯罪事实”、“判处十年有期徒刑以上刑罚”、“徒刑”、“故意犯罪”、“身份不明”及“逮捕”，这些概念需要具备一定的法律基础知识才能理解。

2. **逻辑关系**：句子通过“或者”连接了两个并列的情况，第一个是可能判处十年有期徒刑以上的罪行，第二个是可能判处徒刑以上且有故意犯罪历史或身份不明的情况，这两个条件满足其一即可适用逮捕措施。这种逻辑结构相对直接，但需要读者仔细分辨两个条件及其之间的关系。

3. **句子复杂度**：句子较长，包含多个从句和法律术语，对于非法律专业人士来说，理解起来可能会有些吃力，尤其是如何将不同的条件正确归类并理解其法律后果。

4. **相关法律知识**：理解该句子不仅需要基本的法律知识框架，还需要对刑事诉讼程序中的逮捕条件有一定的了解，这对于普通大众而言可能较为陌生。

综上所述，虽然句子结构并不极其复杂，但由于其专业术语密集且需要特定的法律背景知识，因此难度偏高。

5.对人工评分和大模型评分进行相关性分析。Spearman秩相关系数作为一种非参数统计方法，通过将原始数据转换为秩次（排序位置）并计算两组秩次间的Pearson相关系数，来评估变量间的单调关系。该方法仅依赖于数据的排序关系，不要求数据满足正态分布或等间距条件，因此特别适用于不同生成方式和尺度的数据相关性分析。在自然语言处理领域，该方法被广泛应用于判断词向量的语义相关性等任务。

Spearman秩相关系数如表1所示，人工评分与大模型评分呈现出正相关，其中文心一言一致性更高，相关系数为0.56。具体分布如图2，左侧是人工与文心一言的对比，右侧是人工与通义千问的对比。从分布来看，虽然数据点略为分散，但整体上呈现出沿拟合线上升的趋势，表明大模型评分与人工评分之间具有良好的一致性，从而验证了采用大模型模拟人工评分的可行性。

分析发现，人工评分者对罕见于日常生活的法律术语（如“用益物权”）的敏感度高于模型，而模型对含“除外条款”（如“除当事人另有约定外”）等复合句表现出评分偏高的倾向，可能低估读者的语境推理能力。这些现象值得深入探讨。

大语言模型	Spearman秩相关系数
文心一言	0.56
通义千问	0.44

表 1. 人工评分与大语言模型评分的相关系数

3 大语言模型对立法文本可读性评估的量化分析

3.1 语料库构建

本研究构建的立法文本语料库严格遵循权威性原则，数据源自中国人大网(http://www.npc.gov.cn/)及国家法律法规数据库(https://flk.npc.gov.cn/)。截至2024年9月13日十四届全国人大常委会第十一次会议闭幕，现行有效法律合计303部。本

(c) 2025 中国中文信息学会计算语言学专业委员会

图 2. 人工评分与大语言模型评分的对比

研究使用其中已生效的法律合计290部，逾230万字。以句子为单位对法律条文的可读性进行评估，以句号[3]为分隔符，共提取有效句子[4]40,343个。

依据《中国特色社会主义法律体系》白皮书确立的七分法标准，我国的法律体系框架主要包括宪法及宪法相关法、民法商法、行政法、经济法、社会法、刑法、诉讼与非诉讼程序法等七个法律部门。具体的法律数量以及句子数量见下表。可以看出，不同部门法由于法律数量的不同，句子数量存在较大差异。行政法的句子最多，共12706个句子；刑法的句子最少，共1291个句子。

法律部门	法律数量	句子数量
宪法及宪法相关法	50	4307
民法商法	24	7074
行政法	93	12706
经济法	83	9197
社会法	26	3320
刑法	4	1291
诉讼与非诉讼程序法	10	2448
总计	**290**	**40343**

表 2. 法律语料库及句子分布

3.2 法条句子的可读性特征

大语言模型根据提示语，对全部的法条句子进行可读性评分。下面为最简单评分1分及最难评分9分的示例。

（2）评分为1分的示例：

【句子】

会议简报、发言记录或者摘要可以为纸质版，也可以为电子版。

【回答】

对于这个句子“会议简报、发言记录或者摘要可以为纸质版，也可以为电子版。”，我们综合考虑其法律概念、逻辑关系、句子复杂度及相关法律知识，进行如下分析：

1. **法律概念**：句子中涉及的概念主要是“会议简报”、“发言记录”和“摘要”，以及它们的载体形式“纸质版”和“电子版”。这些概念在日常生活和工作中较为常见，不属于特别复杂或专业的法律术语。

2. **逻辑关系**：句子表达了一个并列关系，即会议简报、发言记录或摘要可以以两种形式存在：纸质版或电子版。这种逻辑关系相对简单，易于理解。

3. **句子复杂度**：从句子结构来看，该句采用了并列句式，没有嵌套复杂的从句或多重修饰成分，因此句子复杂度较低。

4. **相关法律知识**：这句话没有直接涉及具体的法律条款或复杂的法律理论，因此不需要深厚的法律知识背景来理解。

综上所述，这个句子在法律概念、逻辑关系、句子复杂度及相关法律知识方面均较为简单。因此，我给出这个句子的难易程度评分为1分（最简单）。

（3）评分为9分的示例：

[3]法律文本的表达相对规整，因此本文的“句子”直接以句号结尾为标准进行判断。

[4]当法律条文的句子含有政治、毒品等敏感内容时，大语言模型将会拒绝进行理解难度的评分，因此这部分句子将会被排除在本文的研究之外，只有能被大语言模型进行理解难度评分的句子我们才纳入评测范围，并称之为“有效句子”。

第二十四届中国计算语言学大会论文集，第745页-第755页，济南，中国，2025年8月11日至14日。

(c) 2025 中国中文信息学会计算语言学专业委员会

【句子】

本法所称间谍行为，是指下列行为：（一）间谍组织及其代理人实施或者指使、资助他人实施，或者境内外机构、组织、个人与其相勾结实施的危害中华人民共和国国家安全的活动；（二）参加间谍组织或者接受间谍组织及其代理人的任务，或者投靠间谍组织及其代理人；（三）间谍组织及其代理人以外的其他境外机构、组织、个人实施或者指使、资助他人实施，或者境内机构、组织、个人与其相勾结实施的窃取、刺探、收买、非法提供国家秘密、情报以及其他关系国家安全和利益的文件、数据、资料、物品，或者策动、引诱、胁迫、收买国家工作人员叛变的活动；（四）间谍组织及其代理人实施或者指使、资助他人实施，或者境内外机构、组织、个人与其相勾结实施针对国家机关、涉密单位或者关键信息基础设施等的网络攻击、侵入、干扰、控制、破坏等活动；（五）为敌人指示攻击目标；（六）进行其他间谍活动。

【回答】

首先，我们来分析这个句子的法律概念、逻辑关系、句子复杂度及相关法律知识。

1. **法律概念**：句子中涉及多个法律概念，如“间谍行为”、“间谍组织”、“国家秘密”、“情报”、“国家安全”等。这些概念在法律上都有其特定的定义和内涵，需要读者具备一定的法律知识背景才能准确理解。

2. **逻辑关系**：句子中列举了六种间谍行为，每种行为之间为并列关系。同时，每种行为内部又包含了多种情况，这些情况之间也存在复杂的逻辑关系，如实施、指使、资助、相勾结等。

3. **句子复杂度**：这个句子非常长，包含了多个分句和子句，结构复杂。每个分句或子句都描述了不同的间谍行为或情况，需要读者仔细阅读并理解。

4. **相关法律知识**：为了准确理解这个句子，读者需要具备一定的法律知识，特别是关于国家安全、间谍行为等方面的法律知识。

综上所述，这个句子在法律概念、逻辑关系、句子复杂度及相关法律知识等方面都较为复杂，需要读者具备较高的法律素养和阅读能力才能准确理解。因此，我给这个句子的难易程度打分为：9分（10分最难，1分最简单）。这个句子虽然复杂，但并未达到最难的10分程度，因为虽然涉及多个法律概念和复杂逻辑关系，但整体上还是围绕“间谍行为”这一主题展开，有一定的内在逻辑和连贯性。

通过对比典型高低分样本，发现大语言模型的评分机制中，术语、句法结构与逻辑结构是影响立法文本可读性的核心要素。其一，法律概念密度及熟悉度，法律概念越多，熟悉度越低，专业性越高，可读性越低；其二，句法结构复杂度，表达越长、嵌套层次越深，可读性越低；其三，法律概念之间语义关联网络的复杂度，关联越多、越复杂，可读性越低，尤其涉及多主体权利义务交织的表述更易获得高难度评分。

3.3 部门法维度下的可读性差异

按部门法对评分进行汇总后获得如表3所示的分布。结果显示，法条句子的可读性整体处于中等偏高难度水平：88.7%的句子集中在5-8分区间，构成法律条文最普遍的理解难度主体；4分及以下低难度句子仅占10.9%，而9分的高难度句子占0.4%，未出现满级难度（10分）的句子。

部门法	1分	2分	3分	4分	5分	6分	7分	8分	9分	10分
宪法及宪法相关法	15	27	413	150	713	1498	947	525	19	0
民法商法	5	60	498	191	1059	2769	1663	804	25	0
行政法	20	88	857	396	1976	4496	3145	1681	47	0
经济法	8	42	532	265	1304	3317	2410	1293	26	0
社会法	4	11	254	131	557	1255	771	328	9	0
刑法	6	1	34	15	91	405	387	338	14	0
诉讼与非诉讼程序法	2	18	276	96	397	945	481	228	5	0
总计	**60**	**247**	**2864**	**1244**	**6097**	**14685**	**9804**	**5197**	**145**	**0**

表 3. 大语言模型对法律条文中句子的可读性评分汇总

本研究以部门法为单位对可读性进行了正态分布拟合，使用正态分布曲线来近似描述大语言模型评分结果跨部门法的分布情况，如图3所示。刑法（黄色实线）的可读性特征呈现出显著异质性：刑法的分布曲线整体偏右，顶部较高，开口较窄，呈现”高均值-低离散”特征。刑法评分的均值达 μ=6.657分，在七大部门法中高居首位；而其标准差最小，为 σ=1.223，评分最为集中。这说明刑法的句子理解难度整体较高，可读性是所有部门法中最低的，这与刑法条文专业术语密集、罪刑结构复杂的特征高度契合。上文的示例（3）是刑法法条的典型例子。

相较而言，宪法及宪法相关法（μ=5.912，σ=1.448）、诉讼与非诉讼程序法（μ=5.777，σ=1.419）的曲线顶部较低、开口较宽，呈现”低均值-高离散”特征。这说明这两个部门法虽然整体难度较低，但存在部分高难度条文。

其他部门法的可读性评分相差并不是很大。经济法（μ=6.156）、民法商法（μ=6.021）、行政法（μ=6.068）的评分分布相对接近，标准差在1.3-1.4区间波动，反映其条文可读性的均

(c) 2025 中国中文信息学会计算语言学专业委员会

衡性。

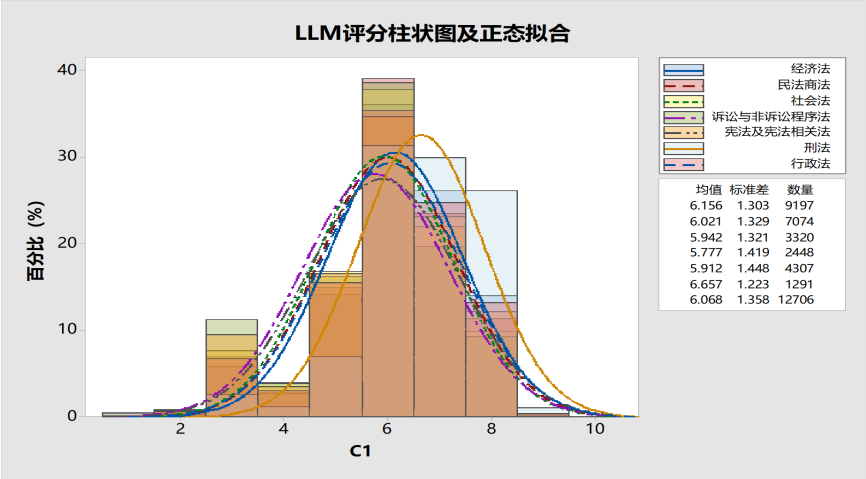


图 3. 大语言模型对法律条文的评分柱状图及正态拟合

3.4 法律维度下的可读性差异

表4是大语言模型评分中均值最高以及最低的20部法律。评分均值最高的《中华人民共和国外国中央银行财产司法强制措施豁免法》（$\mu=7.25$）与均值最低的《中华人民共和国国家通用语言文字法》（$\mu=4.57$）形成鲜明对比。

法律名称	均值	法律名称	均值
中华人民共和国外国中央银行财产司法强制措施豁免法	7.25	中华人民共和国学位条例	5.38
中华人民共和国外国国家豁免法	7.11	中华人民共和国科学技术普及法	5.37
中华人民共和国专属经济区和大陆架法	6.96	中华人民共和国村民委员会组织法	5.36
中华人民共和国乡村振兴促进法	6.80	中华人民共和国老年人权益保障法	5.36
中华人民共和国反电信网络诈骗法	6.78	中华人民共和国监狱法	5.33
中华人民共和国青藏高原生态保护法	6.75	中华人民共和国国徽法	5.32
中华人民共和国环境影响评价法	6.71	中华人民共和国户口登记条例	5.31
中华人民共和国长江保护法	6.70	中华人民共和国劳动法	5.30
中华人民共和国刑法	6.68	中华人民共和国劳动争议调解仲裁法	5.28
中华人民共和国海南自由贸易港法	6.62	中华人民共和国城市居民委员会组织法	5.26
中华人民共和国反有组织犯罪法	6.61	中华人民共和国居民身份证法	5.24
中华人民共和国黄河保护法	6.60	中华人民共和国国务院组织法	5.12
中华人民共和国网络安全法	6.59	中华人民共和国人民法院组织法	5.10
中华人民共和国银行业监督管理法	6.59	中华人民共和国仲裁法	5.07
中华人民共和国野生动物保护法	6.58	中华人民共和国全民所有制工业企业法	5.04
中华人民共和国香港特别行政区维护国家安全法	6.56	中华人民共和国烟叶税法	5.00
中华人民共和国出口管制法	6.56	中华人民共和国国歌法	4.86
中华人民共和国密码法	6.55	中华人民共和国国旗法	4.82
中华人民共和国企业国有资产法	6.54	中华人民共和国国籍法	4.63
中华人民共和国反恐怖主义法	6.53	中华人民共和国国家通用语言文字法	4.57

表 4. 大语言模型评分均值最高和最低的20部法律

本文进一步考察了不同法律的大语言模型评分在标准差方面的分布，表5是大语言模型评分中标准差最高以及最低的20部法律。不同法律在标准差方面也呈现较大差异，标准差最高的为《反分裂国家法》（$\sigma=1.82$）；最低的为《中华人民共和国城市维护建设税法》（$\sigma=0.70$）。

值得注意的是，法条句子可读性评分的均值与标准差之间存在有趣关联：评分均值最高的20部法律和标准差最小的20部法律中，有9部法律是重合的。为了能全面反映法条句子评分均值与标准差之间的关联，我们对290部法律的评分均值与标准差之间的关系进行了拟合。如图4所示，法律的评分均值与标准差之间呈负相关，即评分均值高的法律往往标准差较小，均值

(c) 2025 中国中文信息学会计算语言学专业委员会

法律名称	标准差	法律名称	标准差
反分裂国家法	1.82	中华人民共和国湿地保护法	1.08
中华人民共和国缔结条约程序法	1.81	中华人民共和国保守国家秘密法	1.06
中华人民共和国国务院组织法	1.79	中华人民共和国反电信网络诈骗法	1.06
中华人民共和国全国人民代表大会组织法	1.78	中华人民共和国促进科技成果转化法	1.05
中华人民共和国宪法修正案	1.77	中华人民共和国票据法	1.05
中华人民共和国香港特别行政区基本法	1.76	中华人民共和国长江保护法	1.05
中华人民共和国居民身份证法	1.73	中华人民共和国银行业监督管理法	1.04
中华人民共和国教育法	1.69	中华人民共和国环境影响评价法	1.02
中华人民共和国车船税法	1.68	中华人民共和国乡村振兴促进法	1.01
中华人民共和国国家通用语言文字法	1.65	中华人民共和国反垄断法	1.01
中华人民共和国全国人民代表大会议事规则	1.64	中华人民共和国反外国制裁法	1.00
中华人民共和国全民所有制工业企业法	1.63	中华人民共和国青藏高原生态保护法	0.98
中国人民解放军军官军衔条例	1.63	中华人民共和国外商投资法	0.98
中华人民共和国价格法	1.63	中华人民共和国无障碍环境建设法	0.98
中华人民共和国学位条例	1.62	中华人民共和国人民警察警衔条例	0.97
中华人民共和国护照法	1.62	中华人民共和国专属经济区和大陆架法	0.96
中华人民共和国律师法	1.61	中华人民共和国外国中央银行财产司法强制措施豁免法	0.96
中华人民共和国村民委员会组织法	1.59	中华人民共和国台湾同胞投资保护法	0.94
中华人民共和国国防法	1.58	中华人民共和国外国国家豁免法	0.90
中华人民共和国对外贸易法	1.58	中华人民共和国城市维护建设税法	0.70

表 5. 大语言模型评分标准差最高和最低的20部法律

低的法律标准差反而更大。例如，高均值法律《中华人民共和国银行业监督管理法》（$\mu=6.59$）具有低标准差（$\sigma=1.04$），而低均值法律《中华人民共和国国家通用语言文字法》（$\mu=4.57$分）则标准差较高（$\sigma=1.65$）。这表明技术性强的法律整体可读性低而且稳定，而民生类法律则存在可读性分化现象。

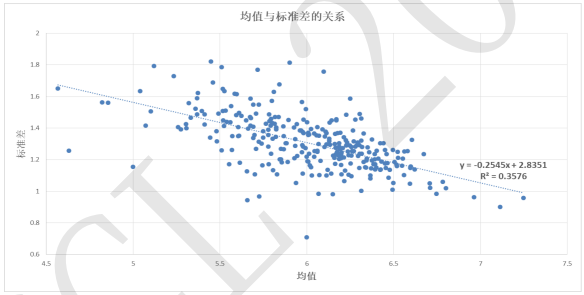


图 4. 大语言模型对法条句子评分均值与标准差的相关关系

4 基于大语言模型的立法语言可读性提升路径

提升立法语言可读性需要在法律与通用领域之间寻求语言形式的平衡，同时必须严格遵循不损害法律条文适用性的基本原则。本研究始终秉持审慎态度，并非主张使用人工智能直接修改法条，而是通过分析大语言模型在“使其变得容易理解”这一指令下的改写策略，逆向探索立法语言优化路径，为立法草案说明、法律解释、普法材料编写等提供数据驱动的参考。本研究将大模型改写的影响区分为语言和法律两个层面，进而区分出仅优化表达方式的“形式性改写”（如增强语篇衔接、拆分长句）和可能触及法律实质内容的“实质性改写”，并明确前者才是法律文本优化过程中安全、可靠的方向。

4.1 大语言模型的改写策略

本研究将可读性低（8分及以上）的法条句子作为改写对象，通过提示语“请改写下列句子，使其变得容易理解”驱动大语言模型生成改写文本，继而对比法条原句与改写后的差异。图5展

(c) 2025 中国中文信息学会计算语言学专业委员会

示了大语言模型对法律文本的改写结果。通过观察改写后的句子，我们提炼出大语言模型提高可读性的三种改写策略。

原句	改写后
工程监理单位与建设单位或~~者~~建筑施工企业串通，~~弄虚~~作假、~~降低~~工程质量~~的~~，~~责令~~改正，~~处以~~罚款，降低资质等级~~或者~~吊销资质证书；~~有违法所得的~~，~~予以~~没收；造成损失~~的~~，承担连带赔偿责任；构成犯罪~~的~~，依法追究刑事责任。	如果工程监理单位与建设单位或建筑施工企业一起欺骗、做假，导致工程质量下降，那么他们将被要求改正错误，并会被罚款。此外，他们的资质等级可能会被降低，甚至资质证书也可能被吊销。如果他们因此获得了非法收益，这些收益将被没收。如果他们的行为造成了损失，他们还需要承担连带赔偿责任。如果他们的行为构成犯罪，那么他们将依法被追究刑事责任。

图 5. 大语言模型对法条句子的改写对比（删除项用红色横线、增补项用蓝色字体）

1. “修改”是大语言模型改写过程中出现最多的手段，主要在词汇和句式两个层面展开，以实现语言简化或表达提升。

修改——词语替换：将法律专业术语替换为通用词汇。如将“履行个人信息保护职责的部门”替换为“个人信息保护部门”、“违法所得”替换为“非法收益”，又例如把“出具”改成“提供”、“责令”改成“命令”等。

修改——句式重组：通过两种方式使法条的句子结构更加清晰。第一种方式是将立法语言程式化表达改为通用表达。如立法语言中的“VP的，……”被众多学者认为是一种立法语言独特的程式化句式，但在改写中，大语言模型把所有的“VP的，……”都改成了“如果VP，（那么）……”条件句式，使逻辑关系显性化。如图5所示例子中的“构成犯罪的，……”改成了“如果他们的行为构成犯罪，……”。第二种方式将长句分解为多个短句。图5中例句的原句就是一个长句，使用分号、逗号区分层次；但改写后被拆解为五个短句。

2. “增加”主要是通过增加语篇衔接成分和补充具体信息，使得法条句子的逻辑更清晰、表达更具体。

增加——增加语篇衔接成分：通过增加连词（“此外”）、副词（“甚至”）、代词等语篇衔接成分，加强句内各成分、各层次之间的衔接连贯程度，构建显性逻辑链条。如图5改写例子中增加了复指代词“他们”来指代句子一开始提及的“工程监理单位与建设单位或建筑施工企业”，建立前后照应，消除指代模糊。

增加——信息补全：对隐含行为主体、抽象行为、后果等进行具象化补足。如在图5改写例子中，在“构成犯罪”前增加了主语“他们的行为”，明确行为主体，避免读者产生歧义。

3. “删除”主要通过删除冗余重复的表达，以实现简化。

删除——删除冗余或重复成分：使法条句子简洁明了。如把“进行反向操作或者相关操作”改成了“采取相反或其他相关操作”，这里原句中的两个“操作”在改写的时候进行了合并。

总体而言，这些改写都朝着简化、具象化和逻辑显化的方向进行，促进法律领域的表达向通用领域的表达靠近，从而提高可读性。

4.2 大模型改写策略的合理性边界

本研究进而考察了大语言模型改写在法律适用方面的表现，区分出仅优化表达方式的“形式性改写”（如增强语篇衔接、拆分长句）和可能触及法律实质内容的“实质性改写”。

1. 形式性改写

(c) 2025 中国中文信息学会计算语言学专业委员会

以图5为例，把“.....的”改写成“如果.....”、关联词“那么”“并”的添加、把“处以”改成“会被”等改写提升了可读性，且法律适用未受影响；

2. 实质性改写

为“造成损失”增加主语之后改成“他们的行为造成了损失”，增加了行为主体，让表达更加具体，似乎并不影响法律适用，但这种改写从法律专业的角度来看不无问题。法律条文的适用性本身就充满了精确性与模糊性的博弈。确定性原则、法律的可预见性、法官的权力限制等方面都要求法律具有精确性。可是，“在很多情况下，缺乏明确性和难以具体适用的法律文本——例如，宽泛、模糊、不精确、不确定的文本——是不可避免的，有时甚至是合乎情理的”。(亚历山大弗吕, 2014, p.19)因为法哲学家考夫曼提出，立法者通过“类型化”手段对生活事实进行提炼、归纳、形塑，从而形成法律规范，“立法者综合一群生活事实成为一个以概念方式加以表示的‘构成要件’，并对此规定一个法律效果” (亚图·考夫曼, 1999, p.37)。这也就意味着，法律概念需保持类型化特征以涵盖同类生活事实，过度具体化将限缩法律规范的适用范围。在该例中，如果在“造成损失”之前加上主语“他们的行为”，就表明“造成损失”的主体是“工程监理单位与建设单位或者建筑施工企业”的相关行为，但这种规定就排除了“他们的行为”间接造成损失的情况。因此从法律的原则层面来讲，这种改写会破坏法律文本的弹性空间，并非一种合理、可接受的改写方式。

类似的还有把“串通，弄虚作假、降低工程质量”改成“一起欺骗、作假，导致工程质量下降”。因为在原条文中“弄虚作假”和“降低工程质量”是并列的关系，而改写后“弄虚作假”成了“降低工程质量”的原因，“弄虚作假”和“工程质量下降”之间构成了因果关系。这种改写会带来法律适用的变化：原条文由于“弄虚作假”和“降低工程质量”是并列关系，因此“弄虚作假”的手段和“降低工程质量”的结果只要出现其中一种情况，主体都会被“责令改正，处以罚款”；但改写后，只有“弄虚作假”导致“工程质量下降”的结果出现才会被“责令改正，处以罚款”。此时法律适用性已发生实质性改变。

在确保法律适用不变的前提下，单纯针对语言形式的改写可切实提升立法文本的可读性水平。然而，如何准确判断改写行为对法律适用的潜在影响，可能需要大模型具备对法律规范内涵与外延的深度解析能力后才能实现。

5 结语

本研究表明，大语言模型对立法文本可读性的自动化评分与人工评分具有显著相关性。本研究首次实现对我国立法文本可读性的全景式考察，揭示了我国法律语言可读性的现状和特点；并尝试使用大语言模型对可读性低的法条句子进行改写。这些探索将有助于推动我国立法技术从经验主导型向数据驱动型的范式转变，最终实现法律文本专业性与可读性的平衡。

本研究尚存在以下局限性，值得在后续研究中继续深化。首先，本研究主要模拟普通公众的认知情境，未能深入对比法律专业人士与非专业人士在可读性感知上的差异。未来可引入多背景的被试群体，构建更具层次性的可读性评估模型。其次，本研究以“句子”为分析单元，虽能精准定位微观层面的语言难点，但法律文本的复杂性也体现在条文之间、章节之间的宏观结构上，未来可尝试将研究视野扩展至条文级乃至整部法律的篇章级，进行系统性的可读性评估。最后，对于大语言模型的评分机制和改写机制，本文的分析仍偏向定性归纳，未来可设计更精细的量化实验，系统探究术语密度、句法依存距离等不同语言特征对模型评分的具体影响权重，以及不同改写策略对可读性提升的效能边界，从而更深刻地揭示立法文本可读性的内在机理。

第二十四届中国计算语言学大会论文集，第745页-第755页，济南，中国，2025年8月11日至14日。

(c) 2025 中国中文信息学会计算语言学专业委员会

致谢

本研究得到国家语委“基于大数据的立法语言规范研究”(项目编号： YB145-97)、中国社会科学院学科建设“登峰战略”资助计划（项目编号： DF2023TS05）的支持。感谢审稿专家给出的专业精准的修改意见。

参考文献

陈佳璇. 2006. 法律语言的“易读性准则”与“易读性”测量[J]. 修辞学习,2006,(04).

崔玉珍. 2016. 从立法语言的连词“或者”看我国法律文本的可读性[J].当代修辞学,(02).

吴思远,于东,江新. 2020. 汉语文本可读性特征体系构建和效度验证[J].世界汉语教学,(01).

亚历山大弗吕. 2014. 法律明晰的模糊原则[C].[法]安娜瓦格纳、[爱尔兰]索菲卡西圭蒂-法伊编,法律中的晦涩与明晰,苏建华、石伟东、崔飞、黄姗译，中国政法大学出版社，2014.

亚图·考夫曼等. 1999. 类比与“事情本质”[M]. 吴从周译，台北：台湾学林文化事业有限公司.

杨孝滁. 1971. 影响中文可读性语言因素的分析[J].报学,(4).

袁毓林. 2024. ChatGPT等大型语言模型对语言学理论的挑战与警示[J].当代修辞学,(01).

郑金雄. 2011. 易读性传播:法律传播中的语言解码与理解[J].政法论坛,(06).

Dale, E., & Chall, J. S. 1949. The concept of readability[J]. *Elementary English*, (26).

Feng, Lijun, Martin Jansche, Matt Huenerfauth &. Noemie Elhadad. 2010. A comparison of features for automatic readability assessment[A]. *Proceedings of the 23rd International Conference on Computational Linguistics* (COLING 2010)，Beijing, China.

Klare, G. R. The measurement of Readability: Useful information for communicators[J]. *ACM Journal of Computer Documentation*, 2000,(24).

Sung, Yao-Ting, Ju-Ling Chen, Ji-Her Cha, Hou-Chiang Tseng, Tao-Hsing Chang & Kuo-En Chang. 2015. Constructing and validating readability models: The method of integrating multilevel linguistic features with machine learning[J]. *Behavior Research Methods*, (2).

Crossley, Scott, Aron Heintz, Joon Suh Choi, Jordan Batchelor, Mehrnoush Karimi & Agnes Malatinszky 2023. A large-scaled corpus for assessing text readability[J]. *Behavior Research Methods*, (2).

van Dijk, T. A., and Kintsch, W. 1983. *Strategies of Discourse Comprehension* [M]. New York: Academic Press.

Farajidizaji, Asma, Vatsal Raina, and Mark Gales. 2023. Is it possible to modify text to a target readability level? an initial investigation using zero-shot large language models[A]. *arXiv preprint arXiv:2309.12551.*

Trott, Sean. and Pamela D. Rivière. 2024. Measuring and Modifying the Readability of English Texts with GPT-4 [A]. *arXiv preprint arXiv:2410.14028*

(c) 2025 中国中文信息学会计算语言学专业委员会