

论人工智能时代量刑规范化的实体进路

王熠珏^{*}

内容提要：人工智能辅助量刑是当代中国数字司法建设的重要环节，量刑智能化程度的不断提高是量刑精准化需求与信息技术发展的双重合力使然。然而，精准的量刑预测未必得出公正的量刑结论。在向决策精准化迈进时，智能式量刑易削弱经验式量刑曾有的灵活性与适应性，并存在技术“辅助”异化为技术“主导”的隐忧。克服现有困难的重点在于廓清量刑精准化的可能与限度，从实体法角度澄清智能式量刑的理念形塑、方法提炼和规则解释。为此，有必要重新审视我国现行量刑规则体系，以责任刑与预防刑的区分为契机，分别探索面向自然人和组织体的裁量方法，从而使新时期的量刑规范化发挥出公平与效率相统一的制度功能。

关键词：人工智能 量刑规范化 量刑精准化 算法 类案同判

一、问题的提出

我国的量刑规范化改革始于 21 世纪初，既受到世界量刑改革运动的影响，也源于我国解决“类案不同判”的内在需要。以制度建构为核心的“1.0 时期”大致从 2003 年开始至 2013 年结束，其间伴随着一系列量刑指导意见的起草、制定和修订，该阶段以实现纸质案卷的电子转换、线下办案的线上转变为目标，故数字技术与量刑领域的融合并不显著。自 2014 年最高人民法院《关于常见犯罪的量刑指导意见》全面施行，我国进入了制度践行的“2.0 时期”，尔后量刑指导意见的适用范围不断拓宽。案件数量与司法资源之间的紧张关系日渐加剧，无形中增加了“类案不同判”的风险，认罪认罚从宽制度改革在此背景下应运而生，量刑规范化改革前期所形成的智识资源和量刑规则成为规范量刑建议的重要指南。然而，这一路径本质上属于经验式量刑的范

^{*} 王熠珏，北京交通大学法学院讲师。

本文为中央高校基本科研业务费专项资金项目“刑事一体化语境下的单位犯罪量刑问题研究”（2022RCW004）的阶段性成果。

畴，不仅量刑建议的提出、采纳、调整仍有争议，而且认罪认罚在常见量刑情节中的体系定位和从宽幅度尚在探索，难以圆满发挥出公正与效率相统一的制度功能。为确保量刑质效，助力“让人民群众在每一个司法案件中感受到公平正义”目标的达成，传统司法的数字化转型成为一种必然选择。智能式量刑作为数字司法建设的关键节点，通过算法模型对刑事案件进行数字“画像”、数据分析和类案对比，能够为个案量刑提供精准预测，对克服量刑失衡和提高司法效率有所裨益。

围绕智能式量刑的争议，包括但不限于以下方面：一是算法缺陷导致预测结果出现系统性偏差，二是回顾性数据（backward-looking data）中蕴含了直接或间接歧视，三是算法黑箱引发的透明度和问责制争议。^{〔1〕}对此，极端审慎的态度是，即使承认人类决策有缺陷也拒斥以量刑算法为代表的技术革新，这种保守立场体现了对数字司法的先验怀疑，如法国“第2019—222号”法案第33条禁止使用法官和书记官的个人数据来评估、分析、比较或预测其实际或将要作出的专业行为。^{〔2〕}与之相较，适度审慎的立场更为可取，如我国最高人民法院在2022年发布的《关于规范和加强人工智能司法应用的意见》中确立了人工智能辅助审判原则，以防量刑算法的风险外溢。学界通过一系列的程序设计来应对智能式量刑的潜在风险，如提高类案检索、推送的精准度，明晰人工智能辅助量刑的应用场景，构建算法解释制度体系，设立智能量刑系统的监督、问责机制等。^{〔3〕}然而，单从程序角度进行规制，既难以从根本上提升智能量刑结论的合理性与科学性，亦无法说明，以精准的量刑预测来提高量刑一致性是否契合量刑规范化改革的初衷，能否经受量刑基本原理的检验。质言之，智能式量刑的规范化进路，既是一个程序问题，更是一个实体问题。由于相同的事实可能因考虑不同的刑罚目的而产生迥异的后果，如果不能妥善地解决量刑算法的法理问题，即便相关的程序机制建立起来，仍难以准确识别变量和设置权重。

基于此，本文首先将智能式量刑蕴含的“量刑精准化”理念置于量刑规范化的演进脉络中反思，这些分析会导向人工智能时代量刑规范化的理论重塑；继而以此为契机，为量刑智能化发展提供更充分的实体根据和可行路径。

二、悖论与难题：人工智能驱动下的量刑规范化变革

囿于人工智能领域的技术迭代迅速，依循本体论思维易将复杂世界简单化和静态化，难以揭示事物的复杂性与多样性，致使规制目的落空。^{〔4〕}因此，较之界定智能式量刑的“事物本质”，

〔1〕 See Mirko Bagaric & Dan Hunter, *Enhancing the Integrity of the Sentencing Process through the Use of Artificial Intelligence*, in Jesper Ryberg & Julian V. Roberts eds., *Sentencing and Artificial Intelligence*, Oxford University Press, 2022, p. 123.

〔2〕 参见施鹏鹏：《法国缘何禁止人工智能指引裁判》，载 https://www.spp.gov.cn/spp/llyj/201910/t20191030_436678.shtml，最后访问时间：2025年3月6日。

〔3〕 参见邓矜婷：《同案同判中同案认定标准的可计算化研究》，载《法制与社会发展》2024年第5期；丰怡凯：《人工智能辅助量刑场景下的程序正义反思与重塑》，载《现代法学》2023年第6期；甄航：《人工智能介入量刑机制：困境、定位与解构》，载《重庆大学学报（社会科学版）》2023年第4期；张玉洁：《人工智能辅助量刑的法律风险与制度建构》，载《学术交流》2023年第3期；等等。

〔4〕 参见林涓民：《论人工智能立法的基本路径》，载《中国法学》2024年第5期。

更重要的是明晰“量刑精准化”与“量刑规范化”之间的意义关联。正确把握这对范畴的理论基础和实践逻辑，既要从量刑规范化的演进中爬梳经验式量刑出场的时空情境与实践难题，也要从正在进行的智能式量刑实践中觉察并提炼其在下一阶段的规范指向与意义所在。

（一）经验式量刑的规范化困境

“量刑规范化”的本义为量刑“标准化”，通过建构一套合理可行的量刑标准或规则，使法官在量刑时有据可循。^{〔5〕}一套“稳定的、可知的、确定的规则”是实现法治的基本要求，量刑法治意味着司法裁决不仅应公开，而且须按照预先公布的标准作出。^{〔6〕}从价值取向看，我国量刑规范化改革旨在实现量刑公正，既要促进量刑过程的公开透明，更要在实体上追求类案同判，并兼顾合理的个案量刑差异。从比较法上看，实现量刑结果的一致性（consistency）并非追求量刑结果的统一性（uniformity），而是通过探寻量刑方法的一致性来实现“类案同判”。^{〔7〕}在此意义上，探求规范自由裁量权的理想方案就成为经验式量刑背景下的重要议题。

就经验式量刑而言，英美法官在 20 世纪中叶前拥有较大的量刑裁量权。美国法官弗兰克尔指出，量刑标准的缺失导致法官自由裁量权不受约束，引发司法实践中的歧视与偏差，他称此为“量刑中的法制阙如”。^{〔8〕}鉴于此，英美等国自 20 世纪 70 年代开启了量刑规范化改革，法定量刑原则、司法自我规制（judicial self-regulation）、强制性量刑法和量刑指南由此成为克服量刑失衡的典型范式。^{〔9〕}其中，以法定原则指导量刑的代表性国家是芬兰和瑞典，该方式对于减少量刑偏差的作用有限，难以防止不同量刑者对法定原则的差异化释读。司法的自我规制则是通过上诉审查制度的自我调节来指导量刑，但该方式不仅过程缓慢而且只能提供零星指导，无法形成统摄全局的改革引擎。而强制性量刑法是针对某些犯罪设置法定最低刑，尽管能提高惩罚确定性，却严重制约法官的自由裁量权，消耗本已处于紧张状态的刑事司法资源，譬如美国加州的“三振出局”法案（Three Strikes and You're Out）。相比之下，量刑指南通过一系列步骤来规范法官的自由裁量权行使，侧重论证方法和过程的一致性，而非刻意追求统一化的量刑结果，既契合法治国要求，通过适用预先规定的原则和标准来促成清晰、一致、透明的法庭裁判，亦符合量刑个别化考虑，为了公正、平等而兼顾个案差异，我国此前聚焦量刑指导意见的改革思路也与之相谙。

理论上，量刑指南能清晰呈现量刑决策的考量因素及其权重，量刑规范化的关键难题——量刑失衡的风险能在很大程度上得以化解。但实际上，围绕量刑指南有无实效的质疑从未停歇。以我国量刑指导意见为例，有论者指出，量刑步骤的设计旨在改变传统“估堆式”量刑，但增加步骤数量可能导致结论的科学性被削弱，反而不及一次性估算更为合理。^{〔10〕}另有论者批判了“印

〔5〕 参见王熠珏：《量刑教义学研究》，中国政法大学出版社 2024 年版，第 11 页。

〔6〕 参见〔英〕安德鲁·阿什沃斯：《量刑与刑事司法》（第 6 版），彭海青、吕泽华译，中国社会科学出版社 2019 年版，第 106 页。

〔7〕 See Jose Pina-Sánchez, *Defining and Measuring Consistency in Sentencing*, in Julian V. Roberts ed., *Exploring Sentencing Practice in England and Wales*, Palgrave Macmillan, 2015, p. 79.

〔8〕 See Marvin E. Frankel, *Lawlessness in Sentencing*, 41 University of Cincinnati Law Review 1 (1972).

〔9〕 See Andrew Ashworth, *Techniques for Reducing Sentence Disparity*, in Andrew Ashworth, Andrew Von Hirsch & Julian V. Roberts eds., *Principled Sentencing: Readings on Theory and Policy*, Hart Publishing, 2009, pp. 243–253.

〔10〕 参见闫平超：《量刑规范化体系下的量刑方法改革刍议》，载《法律适用》2020 年第 22 期。

证式规范化”现象，即部分司法者仍沿用“估堆式”方法预设宣告刑，再逆推各量刑步骤及情节的刑量，致使量刑规范化流于形式。^{〔11〕}但上述论断未切中要害。一则，尽管法官因遵循量刑步骤而花费更多时间剖析案情，但是可能加剧个案之间的量刑差异，但英美的主流观点仍主张量刑指南优于高度自由裁量的量刑模式，与其让量刑活动湮没于未知状态，将量刑理论转化为可操作步骤更为可取。^{〔12〕}再则，定罪和量刑中的逆向型思维客观、真实地存在于刑事司法场域中。从知识论角度看，显性知识虽能凭借话语优势压制隐性知识的传播，但当其无法切实解决问题时，再强势的话语权也终将难以遏制隐性知识的生长。^{〔13〕}就存在论而言，黑格尔曾言“凡是合乎理性的东西都是现实的，凡是现实的东西都是合乎理性的”^{〔14〕}。据此，即使司法者基于实践经验先行预判结果，只要最终结论有法可依并通过量刑步骤校验，便不宜以“印证式规范化”否定量刑指南的价值。

真正的桎梏在于，经验式量刑的规范化程度取决于人的有限理性。在量刑过程中，法官受精力和时间所限，既难以从历史错判中汲取经验，对整个司法系统的量刑范式亦所知有限，大多仅能依据个人经验作推断。这一问题在我国“案多人少”的背景下更加突出，单靠投入更多人力来保证量刑质效难以为继。随着认罪认罚制度的大规模适用，司法实践对于量刑建议精准化的要求与日俱增，经验式量刑的可预期性明显不足，而为祛除这种不确定性，人工智能辅助量刑便成为量刑改革的新趋向。

（二）智能式量刑的精准化反思

关于量刑精准化的探索始于1953年美国学者莫里斯的倡议，他提出维护和共享量刑数据来帮助法官对比同行标准。^{〔15〕}囿于技术所限，莫里斯未能明确提出建立一个现代意义的量刑信息系统。随后，学者霍加斯于1971年首次提出一项相对具体的计算机辅助量刑方案，根据历史判决和再犯风险评估而构建算法模型，为法庭提供量刑建议以促进类案同判。但该提案侧重于预防性惩罚，忽视了报应、复归（restoration）等考量，远不足以规范和指引多个刑罚目的的关系处理。及至20世纪80年代后，杜布等学者在加拿大多个地区开展量刑信息系统试验，澳大利亚、苏格兰等地紧随其后，这些系统以克服量刑失衡为旨归，不仅为法官提供既往的量刑数据，也增进公众对量刑实践的了解。然而，当时的技术难以有效反映刑罚哲学的立场，相比量刑辅助工具，司法者更青睐于上诉法院的指导，加之多种因素导致上述国家或地区的量刑信息系统被搁浅。^{〔16〕}

就我国而言，20世纪80年代兴起的“电脑辅助量刑”研究意在解决实践中的量刑畸轻畸重

〔11〕 参见严海杰：《量刑情节数量化之困境与出路》，载石经海主编：《量刑研究》（第5辑），社会科学文献出版社2020年版，第123-124页。

〔12〕 See Julian V. Roberts, *The Evolution of Sentencing Guidelines in Minnesota and England and Wales*, 48 Crime and Justice 187, 237 (2019).

〔13〕 参见周建达：《以刑定罪的知识生产——过程叙事、权力逻辑与制约瓶颈》，载《法制与社会发展》2015年第1期。

〔14〕 〔德〕黑格尔：《小逻辑》，贺麟译，上海人民出版社2009年版，第60页。

〔15〕 See Tom O' Malley, *Living without Guidelines*, in Andrew Ashworth & Julian V. Roberts eds., *Sentencing Guidelines: Exploring the English Model*, Oxford University Press, 2013, p. 232.

〔16〕 See Jesper Ryberg & Julian V. Roberts, *Sentencing and Artificial Intelligence: Setting the Stage*, in Jesper Ryberg & Julian V. Roberts eds., *Sentencing and Artificial Intelligence*, Oxford University Press, 2022, pp. 3-4.

现象。^{〔17〕} 21 世纪初，江苏、山东等地的基层法院开展了电脑量刑实践，这项“自下而上”的量刑规范化探索经由最高人民法院的介入而日渐转变为“自上而下”，并由试点区域辐射至全国。随着人工智能司法应用场景的拓宽，依托强大算力和海量数据的智能式量刑不断迭代升级，今后基于 DeepSeek-R1 等先进大模型的量刑辅助工具可通过混合专家模型（mixture of experts）和强化学习（reinforcement learning）等技术深入挖掘司法数据的多维相关性，实现自动分析、推送和预警，使案件办理从个体经验判断向智能辅助预测迈进。此时，关键问题不是 21 世纪后人工智能如何提升了量刑精准化水平，而是早期智能式量刑所面临的理论供给不足问题如今是否依然存在。尽管 DeepSeek-R1 等大模型实现了高精度推理，但其底层算法仍依赖对历史裁判数据的统计归纳，智能预测的“精准性”尚不足以替代量刑学理的正当性论证。当前围绕量刑精准化而产生的算法和数据争议，主要源于一些基础性理论问题未能得到妥善解决。

其一，厘清刑罚目的位序是应对算法缺陷的基本前提。若不解决刑罚论的前置性问题，将出现假阴性（预测没有再犯风险，但实际发生了犯罪）和假阳性（预测具有再犯风险，但实际不会发生）两类算法缺陷，前者的危害是高风险罪犯再次实施了犯罪，后者则是低风险罪犯因被判处不应得的惩罚而承受了额外伤害。^{〔18〕} 例如，美国许多州在裁定保释、量刑和假释时依赖风险评估工具来预测个体的再犯可能性，这种做法之所以备受争议，主要是量刑决策不能仅考虑剥夺高风险犯罪人的行动能力，还应权衡其他刑罚目的。^{〔19〕} 多重刑罚目的间缺乏明确位阶与多元影响因素交织，致使量刑决策难以被公式化。但这并不意味着要走向“算法厌恶”的极端，相反，恰恰因为刑罚目的更多地以变量或权重的形式隐藏于算法决策之后，在可能严重影响公民基本权利时，其研究必要性反而更高。

其二，理性的法官群体经验是提高数据质量的重要保障。数据导致量刑预测偏差的原因包括两方面：一是直接歧视，若使用带有偏见的训练数据，算法不仅会重现相同的偏见，甚至会加剧歧视和不公平。二是间接歧视，即使未将特定群体作为显性的预测因子，算法仍能识别出就业、经济收入等因素与再犯之间的相关性，相当于将其他因素作为特定群体的间接代理而得出歧视性结论。^{〔20〕} 对此，不能“将价值问题转化为事实问题，进而把事实问题转化为技术应用问题”^{〔21〕}，以免陷入技治主义的谬误而不自知。智能量刑本身难以实现量刑理论的重构，只有提高法官裁量方法的一致性，才能形成更为理性的法官群体经验，继而以之为基础得出合目的、合逻辑的智能预测结论。

其三，强化量刑说理是解决算法黑箱等争议的根本途径。人工智能辅助量刑的算法黑箱关涉

〔17〕 参见苏惠渔、张国全、史建三：《未来的“电脑法官”——电脑辅助量刑专家系统论》，载《中国法学》1988 年第 4 期。

〔18〕 See Jesper Ryberg & Thomas S. Petersen, *Sentencing and the Conflict between Algorithmic Accuracy and Transparency*, in Jesper Ryberg & Julian V. Roberts eds., *Sentencing and Artificial Intelligence*, Oxford University Press, 2022, p. 64.

〔19〕 See Mirko Bagaric & Dan Hunter, *Enhancing the Integrity of the Sentencing Process through the Use of Artificial Intelligence*, in Jesper Ryberg & Julian V. Roberts eds., *Sentencing and Artificial Intelligence*, Oxford University Press, 2022, p. 128.

〔20〕 See Benjamin Davies & Thomas Douglas, *Learning to Discriminate: The Perfect Proxy Problem in Artificially Intelligent Sentencing*, in Jesper Ryberg & Julian V. Roberts eds., *Sentencing and Artificial Intelligence*, Oxford University Press, 2022, pp. 99–100.

〔21〕 邬晓燕：《数字治理中的技治主义：困境、根源与突破》，载《云南社会科学》2024 年第 6 期，第 41 页。

程序正义，算法问责是基于前者产生的附随问题。算法公开等程序机制固然对解决算法黑箱有所裨益，但当事人对正当程序的感知并非简单取决于算法工具的透明度。一则，算法源代码或输入数据的可访问性与其量刑决策质量之间并无直接关联。再则，算法透明度不会使量刑决策更具反思性。有研究表明，若法官为量刑决策提供解释，即使该决策是基于一个内部工作机制不甚透明的算法建议，当事人的公正性体验亦能得到满足。^{〔22〕}也就是说，在力求以智能式量刑实现量刑精准化的当下，法官之于量刑的亲历性和主体性仍不可或缺，只有加强量刑说理才能真正提高当事人对程序公正的可感知性。

前述困境表明，技术原理与刑罚理论的隔阂易使量刑算法的建构偏离量刑规范化的初衷，而理论供给不足又增添了训练数据中的非理性成分；同时，量刑说理的不清晰以及算法过高的抽象程度都会导致量刑结论无法得到细致的检验与审查，不仅有违科学的“有条理的怀疑”原则，也使问题解决最终向“抽象”遁逃。^{〔23〕}是以，智能式量刑的规范化探索首先应是实体层面的问题，无法仅凭算法的程序规制加以解决。由于刑罚目的位序、理性裁判经验和量刑决策说理对应于量刑的理念形塑、方法提炼和规则解释，相关争议必须回归量刑的基础理论予以澄清。

三、从经验裁量到精准裁量：量刑智能化发展的学理重塑

从经验裁量到精准裁量，契合了我国量刑规范化进程不断动态革新的现实。不过，经验式量刑与智能式量刑之间不是非此即彼、互相替代的零和关系，而是在理论上有所借鉴或赋新的可能。量刑智能化发展的本质是科学与人文的跨界融合，而成功的科际融合至少具备三个前提条件，即“反思本学科的专业标准、对于真正研究的问题尽早达成共识以及坚持清晰的语言表达”^{〔24〕}。若将前文归纳和提炼的量刑规范化难题视为“反思本学科的专业标准”，接下来针对量刑智能化发展的学理重塑则是“对于真正研究的问题尽早达成共识”的尝试，力求在当前算法缺陷、算法歧视、算法黑箱等研究之外，将思考的基点重置于量刑学理以寻求“清晰的语言表达”。

（一）经验式量刑与智能式量刑：互益与背反

当视域从程序规制转向实体反思时，智能量刑算法、数据争论背后的刑罚目的位序、理性裁判经验和量刑决策说理难题，本质上是量刑的理念形塑、方法提炼和规则解释层面的疑问。若仅对量刑理念、量刑方法、量刑规则作清单式罗列，无异于重蹈“论题学”的覆辙，而法学方法论关于“体系”的理论模型具有启发意义。晚近西方国家的量刑研究特点是，不受规制的自由裁量权逐渐受到了体系思考的指引。德国“幅的理论”“点的理论”等量刑学说自不待言，英美学界

〔22〕 See Jesper Ryberg, *Sentencing and Algorithmic Transparency*, in Jesper Ryberg & Julian V. Roberts eds., *Sentencing and Artificial Intelligence*, Oxford University Press, 2022, pp. 21–26.

〔23〕 美国社会学家罗伯特·默顿在其1942年的经典论文《科学与民主的社会结构》中将“有条理的怀疑”视为一项科学精神，即“在科学领域，所有的观点都可以随时被质疑”。据此，希尔根多夫进一步指出法学跨学科研究的风险，即“语言运用的不清晰，以及过高的抽象程度都会导致观点与论据无法得到细致的检验与审查，而这是违反‘有条理的怀疑’原则的”。参见〔德〕埃里克·希尔根多夫：《数字化、人工智能和刑法》，江溯、刘畅等译，北京大学出版社2023年版，第451页。

〔24〕 〔德〕埃里克·希尔根多夫：《数字化、人工智能和刑法》，江溯、刘畅等译，北京大学出版社2023年版，第424页。

亦呼吁采取“更体系性的”方法来组织量刑活动，以此克服量刑失衡。^{〔25〕}按照当代的“体系”概念，外部体系注重逻辑上的融贯性，以确保高度的概观性和法安定性；内部体系意在价值判断上的无矛盾，将整个法律秩序理解并解释为内部无矛盾的统一体或“意义整体”。^{〔26〕}智能式量刑的前述规范化要求可在这一体系构造中展开：一方面，外部体系以量刑规则为基石，无论制定规则或适用规则，均蕴含着一定方法。立足算法和数据的智能式量刑，实质上是量刑规则和方法的新型载体。另一方面，内部体系对应量刑理念维度，理念作为“前”或“超”量刑规则的抽象理论，代表了量刑领域的综合性“通理”，构成融贯现行量刑规则的基础，具有相对稳定性。由是，经验式量刑与智能式量刑首先可在内部体系中作融贯性思考。

1. 经验式量刑与智能式量刑的理论互益

第一，在量刑精准化导向下，智能式量刑比经验式量刑更注重“积极比例原则”。量刑语境中的比例原则侧重报应主义，强调刑罚严厉性与犯罪严重性相适应。20 世纪 70 年代，西方量刑理念从后果主义（consequentialism）向报应主义转变，此前量刑主要取决于未然之罪而非已然之罪，存在诸多不公平、不一致、不透明的问题。唯有报应主义的回归，才能使“比例性”从外围角色转化为核心决策。自 1980 年后，“罪有应得”（just deserts）模式颇具影响力，不仅英美国家的量刑指南蕴含比例观念，加拿大、芬兰、瑞典等国也在其量刑法制中强调比例原则的重要性。^{〔27〕}根据比例性原则的精确程度，可将其分为“积极”与“消极”两种类型。前者较为严格，致力探寻特定的、相称的刑量，对犯罪人施加相称的惩罚；后者相对宽松，仅反对向犯罪人施加不相称的刑罚。^{〔28〕}对经验式量刑而言，奉行积极比例原则意味着尽可能准确地对犯罪人科处其应得的惩罚，或至少在较窄的范围内作出量刑决策。在人工智能时代，这一原则因量刑精准化需求与数字技术的协同发展而愈发凸显其重要性。

第二，立足积极比例原则，智能式量刑对“基数比例”有所助益，但可能有损“序数比例”。为探寻量刑合比例的判断标准，赫希、阿什沃斯等学者对基数比例（cardinal proportionality）和序数比例（ordinal proportionality）作出区分，前者关注特定罪行的刑量是否合比例，后者着眼不同罪行的刑量之相对排序。^{〔29〕}换言之，基数比例旨在为量刑提供基准，类似于确定抽象个罪或具体个罪的责任刑，确保同罪同罚；而序数比例注重不同罪行之间的体系思考，强调重罪重罚、轻罪轻罚。若某罪行的理想刑量为 Ω ，适用量刑指南得出的实际刑量可形成一个相对宽泛的 μ 值区间，引入算法能够使围绕 μ 的刑罚分布范围更窄，有助于减少量刑偏差和实现类案同判，更契合基数比例的要求。然而，算法的训练数据来自历史判决，除非实际的量刑水平与理想的刑罚理论相吻合，才能证成更窄的 μ 值分布；否则，即使预测结论更精确，也可能与符合序数比例

〔25〕 See Julian V. Roberts & Andrew Ashworth, *The Evolution of Sentencing Policy and Practice in England and Wales, 2003–2015*, 45 *Crime and Justice* 307, 308 (2016).

〔26〕 参见〔德〕伯恩·魏德士：《法理学》，丁晓春、吴越译，法律出版社 2013 年版，第 318 页。

〔27〕 See Andrew Von Hirsch & Andrew Ashworth, *Proportionate Sentencing: Exploring the Principles*, Oxford University Press, 2005, p. 1.

〔28〕 See Richard S. Frase, *Theories of Proportionality and Desert*, in Joan Petersilia & Kevin R. Reitz, eds., *The Oxford Handbook of Sentencing and Corrections*, Oxford University Press, 2012, p. 131.

〔29〕 参见〔英〕杰瑞米·侯德：《阿什沃斯刑法原理》，时延安、史蔚译，中国法制出版社 2019 年版，第 23 页。

的理想量刑 Ω 相距更远（见图 1）。^{〔30〕} 因此，如何既符合基数比例又兼顾序数比例，是智能式量刑必须突破的难点。



图 1 人工智能算法对量刑中“基数比例”与“序数比例”的影响

第三，为了使智能式量刑符合序数比例的要求，有必要以“经验性应得”重释现代报应主义。序数比例是犯罪与刑罚的体系性表征，特定惩罚是否合比例取决于它与其他惩罚的关系，故首先要对所有的罪行和刑罚分别作轻重排序，继而在端点处将两个量表相互锚定（最严厉的刑罚惩罚最严重的罪行，最轻缓的刑罚对应最轻微的罪行），从而形成一个完整的刑罚分配方案。既然要实现罪刑轻重的相对排序，那如何确定合理的锚点？在阿什沃斯等学者看来，刑罚的总体严厉程度取决于所处司法辖区现有的惩罚水平，其“终端”即司法辖区所能接受的最严厉的惩罚。^{〔31〕} 罗宾逊进言之，一旦刑罚连续统一体的终端被设置（这是所有社会都必须完成的），那么应得的量刑要求便是具体的，因为犯罪人相对于他人的可谴责性，决定其被置于整个刑罚连续统一体的某个特定点之上，并得出相应的具体量刑。^{〔32〕} 问题是在推行智能式量刑的过程中如何把握犯罪人的可谴责性。这就关涉到报应主义的现代性转向。传统报应主义（报复性应得，vengeful desert）与现代报应主义（道义性/经验性应得，deontological/empirical desert）在“应得”惩罚的内涵上存在实质分野，前者强调以害制害的等效报应，后者侧重犯罪人的应受谴责程度。^{〔33〕} 道义性应得的理论基础是道德哲学，但受限于不同正义观的分歧，其难以直接指导智能式量刑。较之模糊的道义性应得，经验性应得立足于实践理性，试图以实证研究提炼社群的正义共识，更适宜作为量刑内部体系与外部环境互动的桥梁。经由这一桥梁，法官群体的理性经验、社会公众的正义诉求得以反馈至量刑的内部体系，为体系所知悉；这也要求智能量刑系统动态更新量刑规则与训练数据，重点关注样本适配性而非全样本覆盖。

2. 经验式量刑与智能式量刑的功能互竞

智能式量刑对量刑理念的赋新，体现了量刑内外体系之间的互动关联：在外部体系层面，智能量刑系统是量刑规则和方法的新型载体，有助于补足量刑指南的不确定性，尽可能精确地匹配罪刑比例，使犯罪人承担其应得的惩罚。在内部体系面向，智能量刑的精准性又反向作用于量刑理念的调整，譬如凸显积极比例原则的意义比重。然而，这种互动不总是有益的，囿于智能量刑系统的预测是基于历史判决的平均值，可能迫使法官向既往量刑水平趋同。从应然角度看，算法

〔30〕 See Jesper Ryberg, *Sentencing Disparity and Artificial Intelligence*, 57 *The Journal of Value Inquiry* 447 (2023).

〔31〕 See Andrew Ashworth, *Sentencing and Criminal Justice*, Cambridge University Press, 2015, p. 142.

〔32〕 See Paul H. Robinson, *Distributive Principles of Criminal Law: Who Should be Punished How Much?*, Oxford University Press, 2008, p. 12.

〔33〕 See Paul H. Robinson, *Distributive Principles of Criminal Law: Who Should be Punished How Much?*, Oxford University Press, 2008, pp. 172–173.

决策的意见仅为参考依据，但现实中司法者易受其左右，如哥伦比亚、印度已出现法官将 ChatGPT 等工具的意见直接作为案件最终结论的实例。^{〔34〕}可见，向决策精准化迈进时技术“辅助”可能异化为技术“主导”，智能式量刑不仅会助长司法人员的技术依赖，过度向量刑平均数倾斜，也削弱了经验式量刑曾有的灵活性与适应性，导致二者产生功能竞争。

就现阶段而言，智能量刑系统受制于量刑规则的完善以及法官群体经验的理性提升，重申其辅助定位是必要的，且不宜对法官偏离预测结论设置过多的条件约束，以确保法官在刑种选择和刑量确定方面的自由裁量权。此外，还应通过算法的规范性转向来提高量刑决策质量。“描述性”算法是将所有因素汇合，得出以往量刑实践的平均值，然后将之用于未来案件的预测，这样的刑罚分配方式无疑会吸收样本数据中的“沉疴”。例如，英国自 2011 年经历了治安危机后，法院偏离了比例原则而侧重一般威慑，此举得到了上诉法院的支持，理由是“必须施以严厉的刑罚来实现惩罚和威慑”。若未能事先将刑罚理论嵌入算法，径行使用上述倚重一般威慑的数据，其预测结论势必扭曲“标准的”合比例量刑观，并削弱刑罚的沟通功能。与之相对，“规范性”算法并非简单复制或学习历史判决中的偏见或不一致，而是按照一定的刑罚目的位序来辨明量刑中的适当因素及其相应权重。^{〔35〕}由是，量刑的内部体系不能被动地因应外部体系变化作调整，如何以积极比例原则指引规范性算法的建构，关涉智能式量刑精准化能在何种程度上得以实现。

（二）智能式量刑的精准化实现：可能与限度

1. 基本立场：规范性算法的价值二阶考量

首先要说明的是，刑罚目的位序具有抽象不可定性与具体可定性的特征，尽管难以完全脱离具体情境来预先设定各刑罚目的间的固定位序，但至少能明确，智能式量刑奉积极比例原则为主臬，所持的并合主义刑罚论不能滑向后果主义而置报应主义于不顾。这种侧重报应的刑罚价值位序可能招致刑罚目的一元论的质疑，后者认为预防才是唯一的刑罚目的，故需要对报应主义作如下澄清：一方面，一元论的批判对象存在偏差。现代的经验性应得有别于传统的报复性应得，前者的正义观源于社群的正义共识，主张合比例的量刑旨在进行刑事谴责，而非进行“以牙还牙”的等害回报。另一方面，一元论的原因溯源存在偏误。尽管 20 世纪 70 年代以来美国的大规模监禁与报应主义复兴的时间线存在重合，但报应主义提倡的是适度制裁而非严厉惩罚，不必然导致惩罚水平的提高。诸如“三振出局”法案、无假释的终身监禁、强制性最低量刑等导致美国“惩罚性转向”的制度设计不是源于报应主义，而是过度倚重威慑、剥夺犯罪能力等预防目的所致。

探求规范性算法的建构，应遵循一种“附条件优先”的设计思路，尽可能消除并合主义刑罚论中各目的之间的紧张关系。罗宾逊曾就刑罚价值位序的体系性建构提出两种方案：（1）“简单优先”（simple priority），事先按照“最高优先一次高优先……”进行刑罚目的排序，由较高优先权者发挥决定作用。（2）“附条件优先”（contingent priority），在设定优先权时应附加一定条件。例如，当危险性预测的可靠性达到一定程度时，才着重考虑“剥夺犯罪能力”；当威慑的先

〔34〕 参见郑曦：《刑事司法领域自动化决策的风险与规制》，载《上海师范大学学报（哲学社会科学版）》2024 年第 3 期。

〔35〕 See Richard L. Lippke, *Plea Bargaining, Principled Sentencing, and Artificial Intelligence*, in Jesper Ryberg & Julian V. Roberts eds., *Sentencing and Artificial Intelligence*, Oxford University Press, 2022, pp. 187–189.

决条件被满足，即行为人同时克服了法律认知、理性选择和成本分析的三重障碍，才能优先考虑威慑等。^{〔36〕}就智能式量刑而言，方案二可为量刑算法的变量选择和权重设置提供更清晰的指引，在衡平刑罚目的时包含二阶考量：一是以报应作为量刑的首要目的，无须刻意考虑一般威慑，因为报应的实现通常已伴随了一般威慑效果；二是对于特殊威慑、复归、剥夺犯罪能力等特殊预防目的，不追求去情景化的一般性标准，而是根据具体的犯罪类型和犯罪人的可改造性强弱来分别设计威慑模型、复归模型或剥夺犯罪能力模型。这种“以报应为先的预防区分”理念本质上亦遵循了从责任刑到预防刑的判断顺序，使智能式量刑不单是在基数比例与序数比例层面追求量刑“结果”的精准化，同时兼顾了量刑“方法”的精准化。

2. 智能式量刑精准化的实体边界

量刑算法在效率、一致性和持久性上超越人类，尤其在司法资源制约量刑质效的情况下，确定性和客观性是法治的至上“美德”。这种实用主义逻辑投射在法理领域，使智能式量刑呈现一种工具属性，但工具主义法律观因过度强调法安定性与裁判可预见性，可能削弱个案正义的实质考量。尽管比例原则可为量刑精准化提供正当性支撑，但为进一步消解潜在的量刑公正隐忧，应划定技术在量刑领域的实体边界：

其一，作用领域的阶段性。较之全面实现宣告刑的精准预测，研发针对责任刑（如确定量刑基准或预测不法所对应的刑量）的智能式量刑宜作为规范性算法的优先探索，而人身危险性的判定可暂属于经验式量刑的范畴。之所以渐进性落实规范性算法的价值二阶考量，一是受实然技术所限，量刑算法主要是“向后看”的，通过回顾个体所犯罪行的性质而分配相应的刑罚。这并不是说量刑算法不应包含任何“向前看”的因素，而是责任刑数据的采集范围和时间跨度相对固定，涉及的主观臆断因素相对较少，其推理过程较之预防刑的判断逻辑更容易被机器学习。是以，当前技术的主要任务是辅助司法者就所涉罪行作出“适当”的量刑决策，聚焦于犯罪严重性以及犯罪人在此过程中的罪责程度。^{〔37〕}二是与应然取向相符，最大化地发挥经验性应得在实现正义与预防犯罪之间的调和作用。一方面，社会公众容易就核心罪行的序数等级排序达成共识，而法定犯的不断增多更需要以技术手段提炼司法大数据所蕴含的经验性规律和正义共识。另一方面，经验性应得契合社会公众的正义认知，使量刑结论更具公信力，由此产生的规范内化力量具有一定的犯罪抑制效果。

其二，适用对象的差异性。我国理论和实务普遍将认罪认罚案件作为智能式量刑的主要应用场景，因其类型化特征显著，可实现高效精准的量刑预测。^{〔38〕}但认罪认罚情节的刑法意义仍待厘清，且需构建针对组织体的规范性算法模型，避免技术与量刑的一般化结合。一方面，单位承担刑事责任的哲学根据源于其内部治理缺陷而非相对意志自由，^{〔39〕}这种罪质差异决定了单位犯罪与自然人犯罪的序数比例不可作等同考察。例如，单位犯罪的涉案数额往往较高，据此得出单

〔36〕 See Paul H. Robinson, *Distributive Principles of Criminal Law: Who Should be Punished How Much?*, Oxford University Press, 2008, pp. 233–234.

〔37〕 See Richard L. Lippke, *Plea Bargaining, Principled Sentencing, and Artificial Intelligence*, in Jesper Ryberg & Julian V. Roberts eds., *Sentencing and Artificial Intelligence*, Oxford University Press, 2022, p. 188.

〔38〕 参见孙道萃：《人工智能辅助量刑的实践回视与理论供给》，载《学术界》2023年第3期。

〔39〕 参见时延安：《单位犯罪教义学的两个问题》，载《北京社会科学》2024年第2期。

位成员的自由刑也较重，若将单位罚金与自然人自由刑相捆绑，以此作为单位罪行轻重的衡量标准，无疑会将单位犯罪都归为序数位阶较高的犯罪类型，导致单位的从宽处罚空间有限。另一方面，单位罚金与自然人罚金难以通约，二者的基数比例也须分而视之。单位的偿付能力与自然人的经济状况无关，作关联考察并不妥当，故有必要提炼、识别反映组织体特性的量刑情节。此外，我国量刑规范化改革迄今只针对自然人颁行了量刑指导意见，组织体量刑的方法和规则存在明显缺位，若无规则的遵守，智能式量刑的践行终归是不可能的。有学者指出，量刑指南蕴含了一定的演绎算法，算法设计者（立法机关、量刑委员会或法院）根据其倡导的量刑理论、刑事政策和实践考量而将相关的量刑原则编码至算法中，旨在准确反映量刑因素及其权重；而归纳算法则是揭示事实因素与刑量之间的相关性，其预测准确性受制于数据质量。^{〔40〕} 促进经验式量刑（量刑指南）与智能式量刑的有效衔接，可以缓和归纳型算法固有的歧视性缺陷。对此，单从程序法切入难言实益，必须诉诸量刑实体规则的优化完善，才能形成理性的裁量经验作为训练数据，同时为量刑说理提供更合理的实体依据。

四、量刑精准化导向下智能式量刑的规范化路径

量刑精准化导向下智能式量刑的算法、数据困境，除程序规制不足之外，实体法理不清、适用标准不明亦不容忽视。从量刑学理的体系构造中把握传统量刑与智能量刑的关系，可洞悉人工智能时代量刑内部体系的理念特点，理念中的刑罚价值位序不仅是规范性算法构建的核心基准，还应融贯于量刑外部体系的规则形塑和方法提炼，因为一个公正、透明、可问责的量刑体系终归要落脚于量刑规则，不论这些规则是以制定法、判例法、量刑指南抑或算法形式呈现，规则承载着一定的刑罚裁量方法，而量刑方法又进一步促进规则的正确适用。如此理解，既是“量刑规范化”的原意，亦是技术发展的现实使然。若现行法仍以日常语言写就（即便其被转译为计算机语言），解释任务只会转移而非消失：一者，有限的符号语言难以编码案件事实的所有变量，有必要对现有的规范解释反复进行审查、修正和续造；二者，由于不能预设所有情形，规则的解释可能产生新的疑义，故需要对计算机发出新指令。^{〔41〕}

下文将结合“价值二阶考量”反思现行量刑规则体系，力求达致如下目的：一是逻辑的融贯，以规则作为经验式量刑与智能式量刑的沟通渠道，从而获得理性裁判经验来有效解决数据质量问题；二是路径的融贯，使量刑方法内部不存在明显冲突，为优化量刑说理提供合理注解。

（一）智能式量刑规范化的基本构想

在价值融贯的基础上反思现行量刑规则，有关规范性算法的探讨才能以一种合乎量刑学理要求的方式展开。

1. 从“价值二阶考量”角度反思现行量刑规则

若要将法律语言转为算法语言，首先要识别并消除现行法制中制约量刑智能化发展的因素。

〔40〕 See Netanel Dagan & Shmuel Baron, *The Compassionate Computer: Algorithms, Sentencing, and Mercy*, in Jesper Ryberg & Julian V. Roberts eds., *Sentencing and Artificial Intelligence*, Oxford University Press, 2022, pp. 150–152.

〔41〕 参见〔德〕卡尔·拉伦茨：《法学方法论》（第6版），黄家镇译，商务印书馆2020年版，第305–306页。

2021年最高人民法院、最高人民检察院《关于常见犯罪的量刑指导意见（试行）》〔以下简称2021年《量刑指导意见（试行）》〕规定的“三阶式”量刑步骤（量刑起点→基准刑→宣告刑），对大部分情节采取“同向相加、逆向相减”，对特殊主体、防卫过当、犯罪停止形态、共犯等特殊情节进行“部分连乘、部分相加减”。〔42〕如此规定的问题是：其一，由于2021年《量刑指导意见（试行）》未特意区分责任刑情节和预防刑情节，部分情节的无序化、零散化无疑会影响规范性算法的准度和精度，导致类案检索在识别量刑情节与提取争议焦点时仍有较大掣肘。其二，量刑情节适用的“同向相加、逆向相减（部分连乘）”演绎算法不符合从责任评价到预防判断的学界共识，且难以回答究竟是何种刑罚目的在影响量刑结论，故作为智能式量刑基础的训练数据本身就存在纰漏。

尽管该指导意见未言明从责任刑到预防刑的裁量思路，“三阶式”量刑步骤的部分环节已蕴含一定的区分意识：一是选择量刑起点相当于根据基本犯罪构成事实确定责任刑起点；二是基准刑环节是以其他影响犯罪构成的犯罪事实（手段、次数、数额、结果等）在量刑起点的基础上增加刑量，属于责任刑裁量；三是根据量刑情节调整基准刑，并综合全案确定宣告刑。对此，若要贯彻刑罚价值的二阶考量，首先须删去“同向相加、逆向相减”这一机械适用的算法规定，继而在第三步中强化责任刑与预防刑的区分意识，先考虑责任刑情节，之后根据特殊威慑、复归、剥夺犯罪能力的不同需求而设置预防刑情节权重，使其满足“以报应为先的预防区分”要求。若不能妥善解决实体规则中的情节分类裁量问题，即使智能量刑系统建立起来，仍难以清晰界定哪些因素为责任刑或预防刑变量，如何对其进行类型化区分并采取相应的裁量顺序。

2. 量刑规则与算法的责任刑面向

立足责任刑与预防刑的区分框架，规范性算法的建构应明晰责任刑情节的范畴。积极比例原则要求刑事制裁反映犯罪人因其行为而获得的谴责，但应受谴责程度的责任刑情节范围可能因“功能主义”刑法观的兴起而变化。所谓功能主义，是以刑法在社会系统中所承担的功能出发，试图在方法论上整合体系思考与问题思考，使教义学的封闭性与开放性得以兼容。功能主义关涉刑事政策与刑法的关系解读，进而影响应罚性与需罚性的体系定位，并在犯罪论体系构建和量刑环节中表现出来。〔43〕德国古典犯罪论体系与目的主义犯罪论体系将刑事政策排除在刑法体系之外，以应罚性作为犯罪成立与否的判断依据，关注行为的社会危害性评价，而涉及预防必要性的需罚性则归属于刑罚论。据此，行为人的罪责评价并不包含预防必要性考量，责任刑情节中不含有任何预防因素。然而，上述分离模式体现了形式逻辑的自洽性，却可能背离刑事政策上的目标设定，使犯罪论体系趋于僵化。于是，将刑事政策的合目的性考虑整合入刑法体系的融贯模式应运而生，以罗克辛和雅各布斯教授为代表的犯罪论体系内部均交织着应罚性判断和需罚性考量，通过在罪责阶层中融入预防理论，使刑事政策诉求得以进入刑法体系内部。如此一来，应受谴责程度的判断除了受有责的不法影响外，还要兼容预防必要性考虑，致使责任刑情节范围也随之变

〔42〕 2024年最高人民法院、最高人民检察院《关于常见犯罪的量刑指导意见（二）（试行）》沿用“三阶式”量刑步骤。

〔43〕 参见劳东燕：《刑事政策与功能主义的刑法体系》，载《中国法学》2020年第1期。

化。近年来,我国学界围绕“事后行为对于犯罪成立的影响”的论争即为典型例证。一般而言,行为人事后的有益行为不可能对其犯罪行为性质产生回溯性影响。但有学者敏锐发现,我国刑法中存在一系列具有“补正性”的实体构成要件要素,未经这些事后要素对不法状态的最终确认,仅凭前行为所造成的初始结果不足以认定犯罪成立。^{〔44〕}囿于罪责是对不法的评价,兼之我国刑法因罪量要素而具有量刑法属性,事后行为对不法程度的影响必然作用于行为人的应受谴责程度。若据此,意味着量刑指导意见在进行情节类型化区分时,需要将一些恢复法益的事后行为纳入责任刑评价,规范性算法在识别责任刑变量时也应作相应调整。

至此,大陆法系刑法的功能主义思潮能否成为量刑规则的应然取向?事实上,普通法系近来探讨的动态报应主义观(dynamic conceptions of retributivism)与之相谙,即罪行轻重随着时间推演而被重新界定、解释和评估。从宏观上看,应得惩罚的文化和观念可能因时而变;就微观而言,时间流逝可能会削弱罪犯过去和现在身份之间的联系,故不应将罪犯视为“自身过去的囚徒”,而是应动态地看待其罪责变化。若罪犯的回应性行为(如真诚悔罪)表明其已接受并内化了刑罚的谴责信息,则可能减轻对其最初谴责的必要性,这是沟通性惩罚观念的体现。^{〔45〕}进言之,若从回顾性的静态报应主义到前瞻性的动态报应主义,规范性算法在编码量刑规则时需增加预防性变量来衡量罪行轻重,但其弊端在于:一是难以被经验性应得证成,从理论上得出罪行轻重的动态评价,可能背离社会公众的正义直觉。二是加剧智能式量刑的序数比例缺陷,致使积极比例原则受损。一旦应得惩罚的决定因素不仅是原初犯罪的损害和罪责,还包括行为人的事后行为,基数比例的确定就会非常困难;而序数比例亦面临挑战,因为只有明晰特定罪行的轻重,才能比较不同罪行之间刑罚分配的相对排序。^{〔46〕}总之,前述理论的创新性不言而喻,但规范性算法的责任刑变量暂且不宜包括事后行为等预防因素。

3. 量刑规则与算法的预防刑面向

如前所述,现阶段的智能式量刑应注重责任刑探索,不宜期待规范性算法提供终局性量刑结论,但正视现行量刑规则中的预防刑问题,才能避免其阻碍量刑精准化的全面实现。认罪认罚从宽制度改革以量刑精准化为旨归,从2019年《关于适用认罪认罚从宽制度的指导意见》第33条所倚重的“确定刑”量刑建议可见一斑。作为智能式量刑的典型应用场景,明晰认罪认罚与其他从宽情节的相对独立关系,事关规范性算法如何设置预防刑权重。其中,独立性表现为认罪认罚情节始终对新问题保持开放性,从实践中发现并提炼经验性规律,将之上升为新的从宽事由,比如使狭义认罚的实体从宽价值凸显、将预交罚金视为“认罚”等;相对性则体现在认罪认罚与自首、坦白、赔偿损失等常见量刑情节的成立条件存在部分重合,在确定从宽幅度时不得对相同的成立条件所对应的事实作重复评价。^{〔47〕}正因如此,2021年《量刑指导意见(试行)》规定“对于被告人认罪认罚的,……具有自首、重大坦白、退赃退赔、赔偿谅解、刑事和解等情节的,可

〔44〕 参见江溯:《刑法教义学的现状与展望》,载《中国法律评论》2024年第6期。

〔45〕 See Netanel Dagan, *Overcoming Penal Boundaries: Exploring The Evolution of Retributive Time Through Parole Decision-Making*, 62 *The British Journal of Criminology* 37, 50 (2021).

〔46〕 See Jesper Ryberg, *Retributivism and the Dynamic Desert Model: Three Challenges to Dagan and Roberts*, 40 *Criminal Justice Ethics* 56, 58-60 (2021).

〔47〕 参见王熠珏:《认罪认罚从宽制度完善中的实体法向度》,载《社会科学辑刊》2025年第1期。

以减少基准刑的 60% 以下，犯罪较轻的，可以减少基准刑的 60% 以上或者依法免除处罚”。可见，认罪认罚情节虽能突出预防刑的影响效果，但不足以达到“免除处罚”的程度，除非原本的责任刑较轻（即犯罪较轻）且同时存在其他预防刑情节，才可能减少基准刑直至免除处罚。因此，认罪认罚属于可以从轻或减轻处罚的情节。

总之，认罪认罚的相对独立性，决定了将其视为一种综合性而非独立的情节更妥当。若在规范性算法中不考虑认罪认罚与其他从宽情节的竞合关系，重复设置从宽权重，既会造成预防刑情节之间的体系失衡和结构混乱，亦会得出有违积极比例原则的量刑趋轻结论。自“价值二阶考量”角度观之，无论涉及轻罪抑或重罪，对报应主义（即责任刑）的坚守是认罪认罚案件不能逾越的底线；继得出责任刑后，预防刑的权重设置不仅要考虑究竟要侧重威慑、复归还是剥夺犯罪能力，也须避免认罪认罚权重的重复设置。

（二）智能式量刑规范化的路径展开

较之自然人犯罪，单位犯罪不仅量刑规则缺位，其量刑规范化程度也亟待提高。在此背景下，智能式量刑若要实现精准预测，须探寻规则背后更深层次的方法论，将组织体同步置于量刑学理的体系框架中考察，从更普遍意义上发掘人工智能时代量刑规范化的路径选择。

1. 调整、优化自然人的现有量刑方法

将量刑规则与规范性算法建立实质联系会带来一些疑问。既然“三阶式”量刑步骤应遵循责任刑到预防刑的裁量顺序，那么如何将该区分理念内嵌于各步骤中？进言之，若要发挥责任刑的刑罚限定机能，它与智能量刑系统的偏离预警机制之间如何协调？

第一，完善自然人主体的量刑方法并非“增”而是“减”。首要的是，应否在“三阶式”量刑步骤（量刑起点→基准刑→宣告刑）中设置独立的认罪认罚量刑环节。有学者认为，认罪认罚是非固定的弹性情节，实务中被追诉者可能在认罪认罚后反悔，若在第四步骤中用认罪认罚情节再次调节基准刑，有助于促进量刑建议规范化，增强被追诉者认罪认罚的自愿性和明智性，提高被追诉者对反悔后果的预见可能性，规范检察院的量刑建议权等。^{〔48〕}然而，若将认罪认罚视为一个独立情节，会与其他从宽情节陷入重复评价，导致预防刑的权重被过度强化。我国“三阶式”量刑步骤的前两步相当于责任刑裁量，只是在第三步中混合了责任刑和预防刑裁量。次优方案是对第三步所涉情节进行分类，继得出责任刑后，若被追诉者具有认罪认罚情节，在确定各预防刑情节时再扣除重合部分的影响因素。更理想的是，对“三阶式”量刑步骤的前两步作合并考虑，并对原第三步进行情节分类，实质上形成量刑上的“二阶层论”，其优势包括：其一，合并考虑有助于缓和罪量因素之于边际刑量的影响。以普通盗窃 8000 元为例，依据“三阶式”量刑步骤，第一步用基本犯罪构成事实（3000 元为数额较大的起点）选择量刑起点，第二步用其他影响犯罪构成的事实（剩余的 5000 元）调节基准刑，最后在第三步中确定宣告刑。若据此，则量刑起点数额（前 3000 元）的边际价值明显高于基准刑调整数额（后 5000 元），该差异在同一犯罪中并无充分根据，只有合并考虑，才能使每 1 元数额之于刑罚量的边际价值一致。^{〔49〕}其二，

〔48〕 参见田立文：《认罪认罚案件量刑的四个核心问题》，载《中国应用法学》2022 年第 6 期。

〔49〕 参见彭雅丽：《量刑指导意见的司法实践与重构——以盗窃罪为切入点》，载《法学研究》2021 年第 4 期。

“二阶层论”亦满足先责任刑后预防刑的判断顺序，有利于通过提高量刑“方法”的精准化来促进类案同判，进而形成高质量的训练数据。

第二，智能量刑系统的偏离预警应兼顾我国刑事制裁体系特点。精准预测与偏离预警是克服“类案不同判”的一体两面，后者体现了智能式量刑的自主监督和自觉纠错，一旦司法者的裁判尺度与预测结论之间出现重大偏离，系统就会发出预警提示。但该方式只能对比终局性量刑结论，难以诠释是什么在影响偏离度，责任刑与预防刑的作用机理仍像一个“黑箱”，无法提供进一步改善规范性算法的重要信息。若要发挥责任刑的刑罚限定机能，就需要明确预防刑偏离责任刑的条件和限度。值得注意的是，刑事制裁体系对于量刑制度设计具有实质性影响，并体现于责任刑与预防刑的关系处理：其一，双轨制国家以保安处分作为刑罚的必要补充，对再犯风险高的罪犯可适用剥夺自由的保安处分。保安处分为刑罚分担了保护公众免受再犯侵害之重任，故可在量刑学中要求责任刑从根本上制约最终刑量，预防刑只能在责任刑之下发挥有限的作用。其二，单轨制国家则需要通过对累犯、惯犯等判处严厉惩罚来预防犯罪。譬如，普通法系即使存在一些“保安处分”性质的措施，但累犯等高风险犯罪人长期以来都被认为是加重处罚的合法基础。^{〔50〕}英美量刑学理允许基于预防需要而对应得的惩罚进行有限的偏离，其责任刑对于预防刑只有相对的限制意义。^{〔51〕}我国的保安处分体系尚待完善，因此也就不难理解，现行量刑指导意见允许基于累犯、前科等预防刑情节而不断向上累加刑期。就智能量刑系统的偏离预警机制而言，责任刑对量刑的制约可从“相对限制”逐步过渡到“绝对限制”，及至保安处分体系的完善。

2. 对组织体量刑方法加以针对性改进

我国针对组织体的刑事制裁以无限额罚金为主，其明确性困境是不争的事实。若单位罚金的裁量与自然人“解绑”，就面临如何围绕单位自身来建构清晰、可操作的裁量标准，在确保类案同判的同时，适应量刑的智能化发展。

首先，应注意罚金刑与损害赔偿、刑事没收之间的排序，据此构建组织体量刑算法。从经验式量刑的例证看，美国《联邦量刑指南》确立了损害赔偿在顺序上优先于罚金的一般量刑原则；^{〔52〕}英国《组织体欺诈、贿赂、洗钱指南》更是将损害赔偿和刑事没收作为裁量基准罚金的前置步骤。^{〔53〕}可见，罚金刑的执行受制于组织体的偿付能力，决定了算法逻辑须契合以提高执行为目的的现实逻辑，否则智能式量刑将脱离量刑实践。

其次，探求无限额罚金的裁量方法，不能径直采取预防优先方案。美国《联邦量刑指南》第8章创设了“乘数型罚金”，组织体罚金是由“基准罚金额（行为轻重）”的倍数（罪责大小）所决定，由此也影响了英国组织体量刑指南的形塑。尽管乘数型罚金的算法不以自然人量刑为参照，但侧重于威慑目的，美国联邦量刑委员会最初旨在提高法人犯罪的惩罚水平，以消弭法人所

〔50〕 参见江溯主编：《美国判例刑法》，北京大学出版社2021年版，第22页。

〔51〕 参见王熠珏：《刑事制裁体系视角下的量刑规范化路径》，载《广西大学学报（哲学社会科学版）》2024年第6期。

〔52〕 See U. S. Sentencing Guidelines and Policy Statements, § 5E. 4. 1. (b).

〔53〕 See Sentencing Council of England and Wales, *Fraud, Bribery and Money Laundering Offences (Corporate Guideline 2014)*, available at <https://www.sentencingcouncil.org.uk/publications/item/fraud-bribery-and-money-laundering-offences-definitive-guideline/>, last visited on Mar. 13, 2025.

判刑罚比街头犯罪更轻的问题。^{〔54〕}英国《组织体欺诈、贿赂、洗钱指南》亦采取基于威慑的犯罪控制模式，要求罚金必须足以对法人产生实质性的经济影响，使管理层和股东明晰守法的必要性。即便英美指南也要求法院基于犯罪严重性与组织体规模和财务状况处以合比例的罚金，但此种合比例性追求易湮灭在以相乘形式计算罚金额的倍数结构中。从量刑学理来看，合比例的量刑在报应实现时便能发挥一般威慑效果，组织体的自我约束及其对法律的尊重会得到强化，故无须在量刑决策中特别强调一般威慑。因此，乘数型罚金实则威慑目的张扬而报应属性隐退，以其算法作为我国组织体量刑方法的完善思路并不可取。

最后，无限额罚金的确切性困境应诉诸“以报应优先的预防区分”理念。其一，与责任相当的罚金数额，可以涉案数额为参照，同时考虑组织体结构性缺陷所呈现的违法程度以及其他可非难因素。例如，DeepSeek的多模态技术（如视觉语言模型 DeepSeek-VL2）展示了跨领域数据关联分析的潜力。可以想见，该类技术若延伸至单位犯罪场景，可解析企业财务报告、内部治理文件等多源数据，构建组织体结构性缺陷与犯罪行为的关联模型，为责任刑评估提供客观化参考。此外，一些有别于既有罚金裁量范式的参考因素亦值得关注，如英国反重大欺诈办公室和司法特别委员会曾建议将“法人的营业额或企业财务价值”作为罚金计算的备选方案，其组织体非预谋杀人、健康与公共安全、食品安全等量刑指南已将营业额作为法人基准罚金的确定根据。^{〔55〕}这些考量因素为单位罚金的责任刑算法提供了多元参考。其二，在报应优先的前提下，单位罚金的预防刑算法可依循不同的预防目的而建构改善型、剥夺型、威慑型罚金模型。对于经营记录良好的单位，因其具备存续必要性和整改能力，更多是考虑其改善可能性；对于采取伪造文件等方式隐瞒违法活动、刻意规避相关部门审查的单位，应强调特殊威慑以促使其回到依法经营轨道；对于以违法犯罪活动（诸如涉黑涉恶涉恐）为常业的单位，应侧重剥夺其犯罪能力，罚金须足以对单位产生经济性影响。

五、结 语

量刑精准化不是从量刑规范化这个时代命题中任意剥离的子概念，而是有着科学的理论意涵与深厚的实践基础。从经验裁量到精准裁量的演进，是我国解决“类案不同判”和应对“案多人少”的必然选择，但现有研究一直未能正面回应精准裁量的正当性与实质合理性问题。只有将视域从程序规制转向实体反思，解决智能量刑算法、数据之争背后的刑罚目的位序、理性裁判经验和量刑决策说理难题，回应量刑理念形塑、方法提炼和规则解释层面的疑义，才能走出先前的死胡同，避免以量刑精准化为导向的智能式量刑沉溺“技治主义”而走向规则之治的反面，真正对人工智能时代的量刑规范化进路有所助益。

〔54〕 See Laurie J. Rodriguez, *Regulatory Decision Making: Structure and Agency in the Development of the Organizational Sentencing Guidelines*, 51 Crime Law Soc Change 109, 123 (2009).

〔55〕 See House of Commons Justice Committee, *Fraud, Bribery and Money Laundering Offences Guideline: Consultation* (2013), available at <https://publications.parliament.uk/pa/cm201314/cmselect/cmjust/804/804.pdf>, last visited on Mar. 7, 2025.

法学方法论对内外体系的经典区分，有助于明晰量刑规范化改革在人工智能时代的定位，促进经验式量刑与智能式量刑在理论与制度上的衔接。就外部体系而言，立足算法和数据的智能量刑系统是量刑规则和方法的新型载体，将经验上得到确证的、逻辑上一贯的智识反哺于量刑规则的适用与完善；从内部体系来看，智能量刑的精准化又反向作用于量刑理念的调整，这一整体结构服务于最上位目标——促进量刑公正的达成。在人工智能模型不断涌现的当下，量刑规则构筑了一道不可逾越的藩篱，以预防算法自动化决策蕴含的“黑箱”、歧视等风险。为了给量刑活动提供清晰的、一以贯之的、可检验的规则，既要从“价值二阶考量”角度检视现行量刑规则，消除其中制约量刑智能化发展的因素，也要关注量刑规则的未来发展之可能，尽早为自然人或组织体储备量刑规则的优化方案。这不仅是人工智能时代的法治命题，更是量刑规范化研究无从回避的使命。

Abstract: AI-assisted sentencing is pivotal to China's digital judiciary development, driven by demands for precision and technological progress. However, precise sentencing predictions may not ensure equitable outcomes. While enhancing decision accuracy, AI risks undermining the flexibility of traditional sentencing practices and transforming technological “assistance” into “domination”. Addressing these challenges requires clarifying the scope and limits of AI-driven precision through substantive legal frameworks, including exploring the sentencing philosophy, refining sentencing methodologies, and reinterpreting sentencing rules. Reforming China's sentencing system should differentiate responsibility-based and prevention-based penalties, developing tailored approaches for individuals and organizations to balance fairness and efficiency.

Key Words: artificial intelligence, sentencing standardization, sentencing precision, algorithms, consistent case adjudication

(责任编辑：简 爱)