

企业合规咨询智能体性能评估与优化策略探究

林雪治

(厦门大学嘉庚学院, 2684875786@qq.com)

摘要: 在新一代认知智能大模型日趋成熟的背景下, 我们可以依托其来创建企业合规咨询智能体从而帮助数字经济企业合规管理人员快速获取相关法律知识、涉法案例甚至处理建议等参考信息, 从而提升合规管理水平。为了验证该假设, 作者运用四个不同开发者提供的大模型分别创建“中国大陆 AI 技术应用合规顾问”智能体并开展模拟实验, 最终发现尽管智能体刚创建时的查准率、查全率、F1 值等指标都不够高, 但是通过不断添加经过系统梳理的相关知识, 其性能可以得到显著提升并达到优异水平, 具有较高的应用价值。

关键词: 人工智能; 企业; 合规管理; 智能体

引言

近年来, 数字经济作为一种新型经济形态, 正在全球范围内蓬勃兴起, 逐渐成为推动经济发展的新引擎[1]。以中国大陆为例, 2023 年数字经济规模约 56.1 万亿元, 占 GDP 比重超过 40%[2]。截至 2023 年 10 月, 数字经济企业达到 1.7 万余家, 其中万亿级企业 17 家、千亿级企业 162 家[3]。许多企业都通过综合运用各种新兴技术推行了数字化管理、建立了数据驱动的决策机制, 不仅有效促进了企业转型升级, 而且也大大提升了工作效率, 甚至还推进了跨界合作与创新。

然而, 部分数字经济企业并没有因为智能化而走上普惠大众的正轨, 相反却不断挣脱法律法规和政策的监管, 陷入恶意套利的误区[4], 在数据与隐私保护、数字鸿沟与公平、数字税收等方面尝试跨越鸿沟, 甚至出现泄露或违法使用数据、侵犯他人知识产权、网络不正当竞争、网络金融犯罪等违法犯罪现象。以中国大陆为例, 近年来有关部门累计查处了多起数字经济企业违法案例, 据统计, 2018-2023 年, 指导网络交易平台删除违法商品信息 182.97 万个次, 关闭网站 2.39 万个, 责令停止网络平台服务 10.5 万个次, 查处涉网类案件 11.97 万件; 2019-2023 年, 通报或下架违规移动互联网应用程序近 3000 款; 2020-2021 年查处电商、网络餐饮外卖等领域“二选一”垄断案件 2 件; 2020-2022 年上半年平台企业未依法申报违法实施经营者集中案件 159 件; 北京冬奥会和冬残奥会期间, 推动网络平台删除涉冬奥侵权链接 11 万余个[5]。在此背景下, 如何采取更加有效的措施尤其是运用 AI 技术来更好地帮助数字经济企业提升合规管理水平, 值得我们深入研究。

1 企业合规咨询智能体创建的必要性

传统上企业合规管理模式主要可以分为管理导向型和风险导向型两种类型, 前者侧重于事前防范, 后者侧重于事后补救[6]。在这两种模式下, 违规预警通常都具有滞后性。这是因为, 在管理导向型模式下, 由于企业的合规管理制度都是依据现有的法律法规而制定的, 当法律法规出现更新时, 在企业的管理制度没有及时更新或完善的情况下, 企业的违规预警常常会滞后出现。在风险导向型模式下, 企业通常要等到企业或其工作人员出现违规行为后才会其管理制度进行纠错或补漏, 这种表现明显是滞后的。在人工智能等新兴技术飞速发展, 新的违规现象日新月异的背景下, 如果数字经济企业合规管理工作人员不能很好地运用新兴技术辅助管理的话, 企业及其工作人员违规的概率和严重程度可能会进一步提升。

当前，许多数字经济企业都已经开始尝试运用 AI 技术来辅助提升合规检测与预警水平并取得了较好的成效。例如，制造类数字经济企业一方面运用生物识别技术如指纹识别和面部识别等验证员工的身份，确保他们有权进入相应的场所、访问敏感信息或执行特定任务；另一方面运用增强现实和虚拟现实技术模拟生产过程，确保所有步骤都符合制造的安全标准；有的还运用区块链技术和物联网设备跟踪原材料的采购、存储、领用以及产品的生产和流通过程，防范盗窃、不安全操作、违法排放等行为发生。又如，商业贸易类数字经济企业不仅运用人工智能技术对交易数据进行实时监测，识别出可疑的交易模式，减少欺诈和洗钱等违法行为的发生；而且也运用智能合同技术自动执行合同条款，确保按时交付货物，并自动处理付款和退货等事务并留下记录，从而降低违约风险；还通过监测客户反馈，及时发现企业涉嫌违法的线索。再如，服务类数字经济企业，有的运用数字水印技术保护数字内容如文档、图像和视频等的知识产权，以防止未经授权的复制和分发；也有的运用预测性分析工具分析历史数据和发展趋势，帮助企业预测潜在的法律风险，并提前采取措施避免违法行为，还有的依据大数据技术对企业的财务报表进行深入分析，识别出潜在的税收风险。

在新一代认知智能大模型不断涌现并日趋成熟的大背景下，我们完全可以基于这类模型精心打造产品设计、企业竞争、知识产权、AI 技术运用等不同主题企业合规咨询智能体来引导数字经济企业不断提升其合规管理水平。

2 企业合规咨询智能体创建及应用实验

为了验证上述提法，我们进行了如下实验：

2.1 实验假设

我们可以尝试聚焦“中国大陆 AI 技术应用合规”这一主题来创建智能体，角色统一设定为“中国大陆 AI 技术应用合规顾问”，能力设定为“熟悉中国大陆 AI 技术应用相关法律法规和政策，能够为用户提供相关法律知识、涉法案例和法律问题处理建议等参考信息”，输出要求统一设定为“能根据用户提出的个性化问题清晰而又精准地输出答案。”当智能体创建后，我们分别向其提出如下两个问题：第一题（以下简称为 Q1）：当前在深圳运用 AI 技术辅助开展营销需要遵守哪些法律法规？第二题（以下简称为 Q2）：截至目前，中国大陆曾出现过哪些因违法运用 AI 技术而被受到法律制裁的案例？在此基础上，根据智能体生成的答案来分析上述思路的可行性与科学性。为了取得更加有说服力的效果，我们选取了 4 个由中国大陆企业开发的 AI 大模型创建了“中国大陆 AI 技术应用合规”智能体并对其开展了性能评估。具体情况如下：

2.2 实验步骤

步骤 1：统一按照假设输入名称、角色、能力、任务等相关内容，分别创建了 1 号、2 号、3 号和 4 号智能体（简称为 AG1、AG2、AG3、AG4）；

步骤 2：分别向 4 个智能体提出上述两个问题并获得相应的答案；

步骤 3：分析步骤 2 所获得的答案。

2.3 数据统计

为了便于数据统计，我们首先对样本做了如下定义，Q1 以答案里提到每个法律名称作为一个样本，Q2 以答案里提到每个案例作为一个样本。在此基础上，我们分别用 TP 代表被智能体预测为正样本的正确样本，FP 代表被智能体预测为正样本的错误样本，FN 代表被智能体预测为负样本的正确样本，TN 代表被智能体预测为负样本的错误样本。通过深入分析我们发现，针对 Q1 生成的答案中所提到的法律名称（TP+FP）数量分别为 6、6、6 和 1；针对 Q2 生成的答案中所提到的案例（TP+FP）数量分别为 1、5、6 和 5，详见表 1。

表 1 不同智能体的 TP+FP 数量

智能体	Q1	Q2
AG1	6	1
AG2	6	5
AG3	6	6
AG4	1	5

通过对其中的法律名称和案例进行研究，我们发现，针对 Q1 生成的答案中所提到的法律名称全部与题目正相关，属于正确样本，而针对 Q2 生成的答案中所提到的案例，前三个智能体都是正相关的，属于正确样本，第 4 个智能体里有 2 个案例与题目是负相关的，属于错误样本。具体数量详见表 2。

表 2 不同智能体的 TP 数量

智能体	Q1	Q2
AG1	6	1
AG2	6	5
AG3	6	6
AG4	1	3

3 企业合规咨询智能体性能评估与优化

准确率、查准率、查全率、F1 值等都是评估模型性能的关键指标。准确率是指模型正确预测的样本数量与总样本数量之间的比例，其计算公式为 $(TP+TN)/(TP+TN+FP+FN)$ ；查准率是指被正确预测为正样本的样本数占有所有预测为正样本的样本数的比例，其计算公式为 $(TP)/(TP+FP)$ ；查全率是指在原始样本中所有真正样本中，被正确预测为正样本的样本数占有所有真正样本数的比例，其计算公式为 $(TP)/(TP+FN)$ ；F1 值是指查准率和查全率的加权平均值，取值范围在 0 到 1 之间，值越接近 1 表示模型性能越好，其计算公式为 $2 * \text{查全率} * \text{查准率} / (\text{查全率} + \text{查准率})$ [7]。根据上述公式，我们对所创建的智能体进行了性能评估与优化。

3.1 性能评估

鉴于我们创建的智能体仅对正样本进行预测，我们可以从被 AG1 智能体预测为正样本的答案中分析出正确样本与错误样本即 TP 和 FP；也可以从 AG2、AG3、AG4 智能体提供的答案中分析出尚未被 AG1 智能体预测为正样本的正确样本，在此我们将其认定为 AG1 智能体的 FN，依次类推；但却无法分析出被预测为负样本的错误样本即 TN，因此，本实验仅选取查准率、查全率和 F1 值作为评估指标。

在评估过程中，针对 Q1，我们以 4 个智能体所生成的答案里所有正样本法律名称（重复的仅计算一次）作为该题所有真正样本（TP+FN）；针对 Q2，我们以 4 个智能体所生成的答案里所有正样本案例（重复的仅计算一次）作为该题所有真正样本（TP+FN）。经过合并归类，我们发现 Q1 和 Q2 的所有真正样本总数量分别为 15 和 10。计算结果详见表 3 和表 4。

表 3 不同智能体的查准率

智能体	Q1	Q2
AG1	100%	100%
AG2	100%	100%
AG3	100%	100%
AG4	100%	60%

表 4 不同智能体的查全率

智能体	Q1	Q2
AG1	40%	10%
AG2	40%	50%
AG3	40%	60%
AG4	6.67%	30%

基于以上数据，我们进一步计算出了不同智能体的 F1 值，详见表 5。

表 5 不同智能体的 F1 值

智能体	Q1	Q2
AG1	0.5714	0.1818
AG2	0.5714	0.6667
AG3	0.5714	0.75
AG4	0.1251	0.4615

通过研究上述数据，我们发现，基于新一代认知智能大模型创建的“中国大陆 AI 技术应用合规”智能体，其生成的答案查准率普遍较高，具有一定的参考价值；但查全率普遍较低，最终的 F1 值也相对较低，仍然有待于进一步强化训练；从总体上来看，AG3 的各项评估值均较高，综合性能较好。

3.2 性能优化

为了进一步提升“中国大陆 AI 技术应用合规”智能体的性能，针对上述两个问题，我们依托智能体生成的现有答案，通过广泛搜集相关资料进行核实并进一步拓展相关内容，形成了完善的参考知识，分别命名为“中国大陆 AI 技术应用相关法律法规综述”和“中国大陆已定案的违法运用 AI 技术案例汇编”，上传至支持添加知识的智能体 AG3 中，然后再次向其提问并获得新的答案。通过对新答案进行分析，我们可以发现针对 Q1 生成的新答案中正样本法律名称数量分别为 13；针对 Q2 生成的答案中所提到的案例数量分别为 10，为此我们得出了新的查准率、查全率和 F1 值，详见表 6。

表 6 智能体 AG3 性能优化后的指标

智能体	Q1	Q2
查准率	100%	100%
查全率	86.67%	100%
F1 值	0.9286	1

由此可见，为 AG3 智能体添加知识后的实验数据与添加知识前相比，各项评估指标均有显著提升，智能体性能达到了优秀水平。

4 结论

综上，基于新一代认知智能大模型创建的企业合规咨询智能体具有较高的应用价值，可以在数字经济企业中加大力度推广。当然，随着各种新兴技术与数字经济的不断发展，企业将会面临越来越多的新问题与新挑战，国家也会持续制定并加快完善相关法律法规。为此，我们还需要不断加强对数字经济企业相关法律法规以及人

工智能技术的研究，尽量收集更多有参考价值的素材，形成更丰富更完善的知识体系并对企业合规管理咨询智能体开展强化训练，从而不断提升其性能，帮助数字经济企业工作人员更好地提升合规管理水平。

基金项目

本文受 2023 年度福建省教育系统哲学社会科学研究项目《数字经济领域创业新型法律风险及防范策略》（项目编号：JAS23204）资助。

参考文献

- [1] 唐鸣, 吕培亮. 促进数字经济向更高水平发展[EB/OL]. (2021-08-02) https://theory.gmw.cn/2021-08/02/content_35046025.htm [2024-05-15].
- [2] 李彬. 加快数字经济发展, 推动网络强国建设[EB/OL]. (2024-04-23) https://theory.gmw.cn/2024-04/23/content_37280638.htm [2024-05-16].
- [3] 刘洁, 刘亮. 我国数字经济企业整体发展水平稳中有升 呈现“双高”特征[EB/OL]. (2023-10-27) <https://news.cctv.com/2023/10/27/ARTIY077Vpx4MB38qLWEkBNc231027.shtml> [2024-05-18].
- [4] 唐林垚. 人工智能时代的算法规制:责任分层与义务合规[J]. 现代法学, 2020, 42(01): 194-209.
- [5] 中国国务院新闻办公室. 《新时代的中国网络法治建设》白皮书. 2023 年 3 月版.
- [6] 陈瑞华. 企业合规管理体系建设的两种模式[J]. 法学评论, 2024, 42(01): 15-28.
- [7] SOKOLOVA M, LAPALME G. A systematic analysis of performance measures for classification tasks[J]. Information Processing & Management, 2009, 45(4): 427-437.