

论智慧税务中的法技价值对齐

许多奇 董家杰

摘要 在发展新质生产力的背景下,人工智能赋能智慧税务需要防范对税法价值取向和基本原则的偏离。人工智能对齐问题源自技术范式演进带来的自主性跃升,本质上是一种基于目标意图的价值对齐,在税收征管智能化进程中表现为人工智能技术与税收征管理法之间的张力调和。从智慧税务的具体应用场景来看,税务人工智能的权力寻求倾向、代理不适格以及内外部归责缺位冲击征管理法原则,泛化不足和偏见循环则将给征管公平造成挑战,征管效率的实现也可能因人工智能幻觉的存在而受到阻滞。对此,既应立足数据优化和“人在回路”强化技术应对,也要探求权利保障和义务前置的法治路径,并在包容审慎理念下规避“技术解决主义”陷阱。

关键词 智慧税务;税收征管;人工智能;税法原则;价值对齐;纳税人权利;算法决策;算法治理

中图分类号 D912.29 **文献标识码** A **文章编号** 1672-7320(2025)03-0028-11

基金项目 国家社会科学基金重大项目(22&ZD203)

“发展新质生产力是推动高质量发展的内在要求和重要着力点”^[1](P6),深化税收征管改革亦是高质量发展的应有之义,两者辩证统一。新质生产力的发展离不开税收体系的引导助力,所带来的生产关系变革亟须税收征管作出回应,同时也将为新一轮税收征管改革提供强有力的技术赋能。中共中央、国务院印发的《关于进一步深化税收征管改革的意见》(以下简称《意见》)指出,应充分运用大数据、云计算、人工智能、移动互联网等现代信息技术加快建设智慧税务,全面推进税收征管数字化升级和智能化改造。作为促进新质生产力形成的关键驱动力,人工智能在智慧税务建设中发挥着智能引擎的作用,可立足数据要素有效提高税收执法的规范性、税收服务的便捷性、税收监管的精准性。但往越是前沿的技术,在带来前所未有之机遇的同时,所蕴含的风险也越高。对于智能化技术而言,除潜在的故意滥用风险外,更大的挑战在于如何避免系统行为违背人类意图的控制,这便是人工智能的价值对齐问题。

人工智能的价值对齐不单是一个技术性问题,还是一个规范性问题。唯有先行确立所应遵循之价值共识,方得以在系统的开发调试过程中进行编码输入,而法律恰可被视为一套包罗社会价值观的知识库。相对模糊的人类意图在法律的制定和解释过程中被转化为清晰的指令,使得“法律塑造代码”成为引导人工智能实现人类目标的重要方法^[2](P314),这在法律人工智能领域尤为突出。由是观之,人工智能技术理性与税法实践形式理性在税务人工智能构建上的交叉融合,无疑是按照依法治税原则推进税收征管技术革新的应有之义。“人工智能+税收征管”场域中的法技价值对齐问题,具体就表现为如何驾驭人工智能技术为税务部门所用而使其不至于偏离税法的价值取向与基本原则,最终建成安全、以人为本、值得信赖和负责任的智慧税务。

一、智慧税务中法技价值对齐的问题生成

人工智能的价值对齐问题伴随其范式转换相生相长,在税收征管领域的应用则引发了广泛的法技张力问题。智慧税务的发展标志着税收征管的深刻转型,其中税务人工智能扮演着核心角色。然而,新技术与原有法治体系之间尚需磨合。面对税法规范与税收实践的复杂性,为确保算法系统的介入能够与税收征管的价值目标保持协调一致,税务人工智能不仅应遵循具体的税收规则,更须深入理解和运用抽象的税法原则,从而在提升效率的同时平衡法治与公平,实现技术理性与价值理性的有机结合。

(一) 人工智能范式演进背景下的价值对齐

范式是共有的范例,从中提供了产生特定而连贯之科学研究传统的模式。成熟科学的发展道路往往体现为一种范式通过革命向另一种范式的接续过渡,且一旦发生范式转换,用以确定问题和解答正当性的标准通常也会发生重大变化^[3](P8-10, 92-93, 156-157)。早在20世纪50年代,科幻作家阿西莫夫为机器人制定了不伤害人类、服从人类命令、保护自己这三项优先性依次递减的法则。然而,随着人工智能技术从“符号主义”到“联结主义”的范式变迁,完全顺从与高度智能之间不可调和的矛盾日益突出。人工智能不再仅仅被动地遵循为其设定的规则,作出基于逻辑推理的知识表达,而是在算法和数据的驱动下开展“自下而上”的主动学习。以ChatGPT为代表的生成式人工智能,更是实现了从执行简单的判别式任务到完成复杂的创造性任务的能力跨越。但自主性的跃升也带来了偏见、欺骗、迎合乃至操纵等挑战,使得人工智能的应用可能产生事与愿违的结果。这便是所谓“机器的诈术——为了统治,它们服务”^[4](P201),并逐渐在被利用的过程中悄无声息地反客为主。为此,正如控制论之父维纳所言:“如果我们期望使用机器来达成特定目的,而一旦启动运作后又无法有效干预,那么我们最好确保输入机器的目的是我们真正想要的而非其斑斓的赝品。”^[5](P1358)

任何智力活动都是目标驱动的。人工智能的定位是帮助人类,其所执行任务的目标只能通过人类用户的意图间接获得,即需要经由对齐来创建按照用户意图行事的代理。换言之,作为人工智能安全的核心组成部分,对齐问题建立在人工智能系统已拥有强大能力的基础之上,旨在克服动机错位的潜在危害,本质上是一种价值对齐。当然,能力不足同样会引发对目标价值的偏离,错误的动机亦外显于涌现的不良行为之中。值得注意的是,价值对齐不等于用户满意。虽然价值判断带有主观性,但在一定的社会环境下往往存在一套可资客观参照的普遍价值观念,从而防止人工智能的滥用。

尽管长久以来拥护技术中立的观点都主张价值无涉,但根据技术价值论,技术本身是有价值负荷的。一件技术产品的工具价值不仅仅等于协助政治、经济、社会等领域普遍目标实现的外在价值,其价值的基础乃是建立在自身内在特性之上的,只不过该内在价值在道德上的适当地位要归功于更广泛的外在终极价值^[6](P133-137)。相较于其他技术,人工智能系统的内在价值体系带有更高的自主性,强化价值管理的必要性尤为突出。

(二) 税收征管智能化进程中的法技张力

鉴于税务部门有限的人力物力,我国的税收征管改革始终沿着提升征管效率的主线前进,而技术赋能无疑是主要手段。金税工程的建设就充分体现出技术发展进步与税收征管改革的同步演进趋势。金税一期和金税二期聚焦增值税领域的“以票控税”,金税三期则通过统一技术基础平台的搭建强化“信息管税”。到了金税四期,仅仅依靠税收征纳的流程化、标准化和规范化,在面对数字经济背景下的生产交易复杂化以及相应的税制结构转变时日益捉襟见肘,建设智慧税务以实现与实体税制变革相适配的征管流程再造势在必行。智慧税务以数字化、智能化、场景化为特点,兼顾税务机关和纳税缴费人两端,借助大数据智能化应用实现对税务执法、服务、监管的全过程强智控,从而提升税收治理的整体效能。质言之,智慧税务受到“以数治税”和“算法治税”的双轮驱动,海量税务数据背后的关联规律经由智能算法的挖掘解析得以呈现,人工智能技术在智慧税务建设中的关键性不言而喻。

人工智能对税收征管体系的重塑,本质上是用类人化的智能算法接管人工。一方面将有限的人力资源从高频重复操作的事务性工作中解放出来,另一方面凭借其强大的交互能力、推理能力、生成能力,助推精确执法、精细服务、精准监管、精诚共治。但人工智能与税收征管并非天然耦合,在法治轨道上正确处理繁复严谨的税收法规、灵活应对复杂多样的征管情景,对于人工智能而言极具挑战。经济合作与发展组织(OECD)发布的《税收征管3.0:税收征管的数字化转型》报告则指出,成功的关键在于将人力技能与高级分析决策支持工具相结合,构建兼具人性化和高科技适应性的组织形态。必须清楚地认识到,税务人工智能之所以被人类创造出来,归根结底是为了服务于人类社会的税收治理,因而理应坚持其工具性定位,遵循“以人为本”的价值原则来实现对其行为的最终控制,促成智慧税务由工具理性向价值理性的回归。这不仅是一种改革理念转变,更是技术方案更新的客观要求。

一言以蔽之,人工智能应用中的人机价值对齐问题,在智慧税务场景下表现为人工智能技术与税收征管法治之间的张力调和。在任何社会,规范相较于价值都更易于被明显察觉,因为在日常交往中的频繁应用使它们更容易提醒行为者^[7](P65-66)。税法规范的作用即在于根据社会价值共识协调多元主体的税收利益。税收征管中税务机关和纳税人之间可能发生的、已经类型化的利益冲突,被立法者依据一定价值标准作出优位判断,并最终落实到所制定的相应税收规范中,作为后续争议的推论基础^[8](P159-160)。因此,从通用人工智能迈至税务人工智能的关键一步,便是内嵌入具体的税收规则,以奠定税收征管行动逻辑的一致性与合法性基础。但对于复杂税务的执行而言,原则的掌握和运用更为关键。在整个税法体系中,税法原则在立法目的和具体规则之间发挥着承上启下的作用,系统化地归纳呈现税收立法的精神追求。较之于规则,原则具有一般性和普遍性,通常指向某种意向、目标、权利或价值。许多个别存在的税收规则可被视为同一税收原则的体现,在税收规则的适用过程中原则往往扮演着说明性及证立性的角色^[9](P260)。这种抽象,正是应对复杂系统的基本策略。在税收征管实践中,规则总是以或有或无的二元方式适用,一旦适用于某种情况就限定了它的价值;原则却是有“分量”的,互相冲突的原则之间必须进行衡量^[10](P11)。当税务人工智能面对未曾预料的征纳情形,简单依据规则作出“如果一那么”的条件式判断往往难以奏效。此时,税法原则的介入恰如往价值天平上配置砝码,进而能够对不确定状态下的行为策略赋予一个期望值“估价”,为税法规则的解释适用与漏洞填补提供价值支撑,税法的合理性也由此在执行过程中获得具象化并整体呈现扩大趋势。综上,税收征管法的基本原则既是税法原则的程序延伸,也是行政法原则的特殊体现,构成了税务人工智能需要对齐的价值目标。

二、智慧税务中法技价值对齐的场景透视

在智慧税务浪潮中,税收征管的权力结构与执行框架正经历着一场史无前例的大变革,产生了多样化的法技价值错位场景。自动化算法成了税收征管权力的新分享者,甚至表现出主动攫取额外权力的可怖倾向。面对税务机关与智能算法之间趋于离散化的权力界分,加之“算法黑箱”和主体性质带来的双重问责困境,征管权力与纳税人权利的失衡风险急剧增加,形式法治所要求的依法治税原则岌岌可危。与此同时,实际税收征管情形的分布偏移与多元税收征管目标的权衡难题自始至终考验着税务人工智能的泛化能力,与多来源的系统偏见共同冲击着实质法治所强调的税收公平。即便是作为智慧税务最直观追求的征管效率,也难以逃脱人工智能幻觉的羁縻。

(一) 征管法治的权力主体诘问

征管法治原则是税收法定主义的执行面向,税收课征的构成要件与效力后果皆应由法律明确规定。就其中的主体性要素而言,一般情况下默认由税务机关依法治税,即税务机关及其工作人员依照法定职权和法定程序履行税收征管职责。《中华人民共和国税收征收管理法》(以下简称《税收征收管理法》)第29条亦明定,除税务机关、税务人员以及经税务机关依照法律、行政法规委托的单位和人员外,任何单位和个人不得进行税款征收活动。但在智慧税务时代,算法在税收征管中被广泛用于事实认定和法律适

用,在充当税务人员的辅助之余,甚至直接替代税务机关作出具有行政效力的征管决定。因触及技术赋能下征管权力的委托或创制,权力主体及其权限问题便值得重新考量。

一个合法行政行为的成立,首先要求其作出主体满足资格要件,具有为法律所承认的行政权能,且行政权的实际行使不得超出法定权限。智慧税务构建中的“算法系统体现了一组策略,其中权力内在于具体的行动和情境”^[11](P3)。当面对复杂的税收任务时,一个受困的超级智能可能用一些我们现在无法想象的方法来战胜人类“狱卒”^[12](P188),具有战略意识的代理规划使其有动机在这一过程中逾越既定的税收法治框架获取和维护权力。这种危险的“权力寻求”行为势必直接冲击征管法治。即便在“确认式申报”等应用场景下,通过规则的代码转译可保证算法最大程度地按照税收法律运行,部署机关得以对算法保有较大的控制力。但从实际的自动化运行机制来看,税务机关和智能算法之间仍不可否认地构成了“委托—代理”关系,智能算法事实上成了税收征管权力的分享者。而无论按照行政委托的一般法理,还是根据《委托代征管理办法》(国家税务总局公告2013年第24号)的具体规定,受委托方都应具备独立的权利能力、行为能力和责任能力,即拥有法律人格。

“若先进技术在发展中被认为是值得享有权利的,也只能在相应的责任和义务范围内获得。”^[13](P317)在外部可问责性方面,算法自动决策的介入将切断传统法律中的“行为—责任”逻辑链条。虽然税务机关作为委托方责无旁贷,但算法的不透明度和低可解释性将给举证蒙上一层阴影,现实的司法能力与算法监督的技术要求之间亦存在显著错配。由于存在算法黑箱,连税务人员都可能受缚于一个“数字牢笼”。它通过信息架构和算法而非韦伯式规则和程序来合理化任务执行,导致无法控制或很难影响其动态和结果^[14](P181-182)。处于相对弱勢的纳税人要夺回被算法吞噬之实体性和程序性权利,难度自不待言。在近三年时间里错误地要求约40万名受害者偿还福利的澳大利亚“机器人追债计划”(Robodebt计划),即为典型的反面案例。即便是作为该国最强大公共调查机构的皇家委员会都花费了11个月才发布最终报告,遑论受影响的普通人。该案集体诉讼首席原告就坦言,她曾向社会福利联络中心确认是如何计算得出其债务数额的,却被告知所基于的算法对她来说难以理解,未偿付金额则从其退税中直接扣除。内部问责则更是有如天方夜谭。在传统税务行政中,税务人员代表税务机关实际行使征管权。他们作为国家公务人员通过严格的选拔任用,并可在考核评价的基础上对其实现问责追责。虽有“生命3.0”之称,人工智能系统的硅基属性还是从本质上决定了它无法像人一样承担义务和责任。算法权力与责任的不对称,无疑将打破原先税收法治体系中税务机关权力与纳税人权利之间的微妙平衡,引发规则程序在价值内涵和拘束功能上的变异与耗损^[15](P23)。

(二) 征管公正的泛化与偏见挑战

宪法上的平等原则投射至税法领域,不仅在实体上要求税收债务的平等负担,亦强调征管程序中的平等执行。税收的公平在横向上要求条件相同的人应负担相同税收,在纵向上则要求条件不同的人应负担不同税收。税收征管的公正原则,一方面是对实体层面税收公平原则的延续,税务机关经由依法行政严格落实税制安排,有效防杜纳税义务人违法逃税或脱法避税;另一方面也具有独特的程序价值,税务机关在自由裁量范围内须综合考量各种因素采取合理措施,同时还应平等地保障征管对象的程序性权利^[16](P63-67)。在智慧税务场景下,人工智能的脆弱性和偏见构成了公正征管的双重挑战。

人工智能的稳健性是一种在多样化情境下保持可靠运行且对意外干扰具有韧性的能力,尤其强调能力可靠之外的目标正确。量能负担原则确立了税法从征税者之法到纳税人之法的变革方向,而新质生产力的发展不仅将导致“产消者”大量涌现,催生海量的自然人纳税行为,也使得收入来源更加多元、形式更加复杂,加大税收征管难度^[17]。对此,《意见》提出了实现法人税费信息“一户式”、自然人税费信息“一人式”智能归集的创新思路,各地利用税收大数据进行纳税人画像以精准防控税收风险的实践也如火如荼。但当前的人工智能系统在面对现实世界的复杂情形和未知因素时通常很脆弱。受限于资格问题和分支问题,人工智能系统对一切进行建模既不现实也不可取,因而必然产生分布外泛化能力不足

的问题,导致税务人工智能在与训练数据存在分布偏移的测试场景下易出现性能的显著下降,无法应对税收征管实践中因异常情况引发的“长尾风险”。例如,在实质课税的裁量上,如何将目光往返于事实与规范之间,结合法律实质主义和经济实质主义来公平地确定每个纳税人的真实应纳税额,就是对税务人工智能稳健性的重大考验。同时,无论是依法征税还是实质课税的过程中,整体法治和个案公平之间难免会发生冲突。但面对多个原则之间的价值衡量,人工智能系统的目标错误泛化风险不容忽视。即便已经提供了正确的规范并在训练情境中表现良好,学习模型仍可能表现出对非预期目标的追求优化。税法原则的抽象性和复合性,使得不同价值目标的位阶和量级难以被绝对化地表征。公平原则要求的结果公正不等于平均且缺乏统一评价标准,技术上所实现的所谓公平不一定能为社会价值所完全接受。

如果说稳健性对于征管公正之影响还带有间接性的话,模型偏见则属于税收公平的直接威胁。偏见意味着对标准的偏离,尽管有时对于识别所使用之数据或语言中某些统计模式是必要的,但根据相关自动化系统所部署的背景及其预期执行的任务,一旦基于某些特征对个人或群体作出了不平等的对待,便产生了歧视^[18](P72-73)。在轰动一时的荷兰育儿津贴丑闻中,税务管理当局从2013年起开始借助算法创建更有可能实施欺诈的居民风险档案,使用几乎没有监督的自动化系统对福利申请者进行扫描并标记出潜在欺诈者。但由于该算法本身因考虑了申请者是否拥有第二国籍等变量而带有歧视性,导致26000多个无辜家庭被错误地指控社会福利欺诈而被迫偿还并未欠过的钱,许多家庭因此濒临经济崩溃甚至支离破碎,其中有色人种受到的影响尤为严重。类似事件给智慧税务中的自动化决策应用敲响了警钟,有关部门务必时刻提防陷入偏见生成扩大的“数据—算法—用户”循环^[19](P3-4)。

(三) 征管效率的幻觉阻滞

征管效率原则要求税务机关“以符合成本效益的方式稽征税捐”^[20](P125),成本端强调税务机关征管成本和纳税人遵从成本的双重降低,效益端则注重量与质的协同提升。效率原则体现出典型的技术理性,但同时也为价值理性的实现提供了治理工具,乃至在许多时候被视为一项单独的制度价值。

“任何理性系统中的不确定性、不可预测性和低效性大都来源于人。”^[21](P115)意在辅助税务人员和纳税人的决策支持、精准服务等功能,由此成了人工智能在智慧税务中的重要应用场景,具体表现为结合税收规范和事实提供智能化的咨询问答、决策建议等。而税收领域的特殊性对人工智能的此类内容输出提出了更高的准确性要求,一旦虚假或错误的输出内容被税务机关或者纳税人所采纳,不但无法达成提升效率的目的,还将严重扰乱税收征管秩序,并给相应主体带来严重的法律后果。但现实不容乐观,生成式人工智能的幻觉问题可谓司空见惯,即经常生成与提供的源内容不一致或无意义的文本,这些输出看似契合真实语境,实际上却存在矛盾或无法验证。更可怕的是,人工智能系统能够从训练中学会通过操纵、迎合和作弊等手段欺骗人类。华盛顿邮报对两个业界领先报税软件中引入的智能聊天机器人进行了测试,结果发现在16个测试问题中,一个答错了50%以上且大多数回答极其不相关,另一个给出不够有帮助之回答的比例也超过30%。用户若使用这些智能软件进行报税,且不论浪费时间,更可能招致审计。可以说,克服人工智能幻觉对税收征纳效率的阻滞还任重道远。

三、智慧税务中法技价值对齐的路径探析

在技术支持、补充乃至替代规则的“法律3.0”时代,我们固然应当从技术方案本身着手来实现税务人工智能与税收征管原则的对齐,特别是需要在人类与机器深度交互基础上发展出知识与数据协同驱动的税务人工智能系统,从而兼顾适应性与精确性,并妥善把握其不同场景中的介入限度。然而,一方面,唯有在税收征管法层面对智慧税务作出反身性回应,确立正确的价值取向且据之构建完善的规范体系,税务人工智能的对齐方才能有的放矢、有迹可循;另一方面,“对齐税”^①的存在使税务机关以及相

① 对齐税是指为对齐模型而需付出的额外成本。任何对齐税较高的对齐技术都可能无法被采用,因而应尽量予以避免。

关技术团队缺乏十足的对齐动机,需要法律的激励与约束加以补强。因此,人工智能赋能智慧税务的落脚点还在于法治,关键是实现技术与法律的深度融合,在保护纳税人权利的同时前置风险管理义务。

(一) 技术路径:数据奠基与“人在回路”交互并行

人工智能从1.0发展到2.0,知识驱动转变为数据驱动。尽管人工智能系统脱离了对专家的依赖,转面可以一种适应现实世界混杂性的流动方式从经验中学习,但在一定程度上丧失精确性也在所难免^[22](P22-24),而这对于税收征管这样的法治场域而言显然是不可接受的。唯一的办法便是融合人类智力与数据要素,走向算力支持下知识与数据并行驱动算法的四要素式税务人工智能。

1. 数据处理优化来源数量与准入质量

数据的收集与处理是驱动智慧税务的关键,唯有获取完整准确的税务信息,方得实现税款的精准确定和及时征收。具体而言,需要从量和质上对税务人工智能所使用的涉税数据进行优化。

构建可靠的税务人工智能系统,在很大程度上依赖于其适应多样化涉税数据的能力。拓展数据来源以构建接近真实部署场景的数据集,不仅对于稳健性的提升至关重要,亦能为测试评估的有效开展奠定情境基础。虽然历经前三期金税工程,税收管理信息系统已然积累了大量税收数据,《意见》也进一步强调“着力推进内外部涉税数据汇聚联通”以“深化税收大数据共享应用”,但仍然突出的数据孤岛现象还是使数据的全面性得不到保障。对此,首选之策当然是在切实履行相应数据安保义务的前提下强化面向社会的必要涉税数据采集,并推动涉税数据的跨部门、跨层级、跨地域协同共享。而面对真实涉税数据不足的现实困境,人工数据增强技术或可在一定程度上解燃眉之急。

智慧税务建设过程中税务机关的数据清洗与管理能力也面临着重大考验,涉税数据的冗余、分散、低质问题亟待克服。由于模型的知识几乎都是在预训练阶段学到的,不同类型和表现形式数据中的敏感特征及其因果影响、某些群体的过度或不足代表性都可能将偏见引入税务系统,海量数据中含有的虚构、过时或偏颇的噪声信息则是造成模型幻觉的重要来源。故在税务人工智能的训练过程中,必须贯彻数据准入理念,加强涉税数据的预处理。经由精心地采样、挑选和配比实现涉税数据的平衡分布,并尽可能减少不可验证或不可靠的涉税数据。

2. “人在回路”把控输入反馈与应用界限

“人之所以为人者,非特以二足而无毛也,以其有辨也。”(《荀子·非相》)人工智能系统自始至终都无法脱离人类智能而单独存在,从其开发形成到测试运用的每个环节都深深根植于人类的行为模式和经验体系。可以说,最优模型是通过机器与人类智能之间建设性、持续性的交互关系创建的,即务必贯彻“人在回路”^①策略。

在训练阶段,“人在回路”首先表现为参与知识输入和数据处理。智慧税务的本质是将税法规范和税收事实置换为代码,实现从数据输入到信息输出^[23](P50)。一方面,直接将先验的税收知识注入税务人工智能模型是一种减少所需训练数据量的原则性方法,也可极大降低模型从数据中自行归纳知识的不确定性;另一方面,监督学习模型几乎总是随着标注数据的增加变得更加准确,有必要通过对原始涉税数据的标注为税务人工智能创建高质量的训练数据。囿于法律语言的模糊性和法律事实的复杂性,为克服代码转译偏差,不可避免地需要技术专家与税务专家进行大量直接互动以实现领域知识的交换。

除了前端输入,智能模型性能的提升也离不开后端的反馈干预。对此,由Anthropic提出的“宪法型人工智能”改进并部分替代了基于人类反馈的强化学习,其中的人类监督完全来源于一套规范人工智能行为的原则,仅辅之以少量示例用于小样本提示。这一思路生动地诠释了“代码即法律”,形成了人类向人工智能提供宪法式的简短原则或指令清单、再由人工智能对人工智能进行扩展监督的“元治理”模式,

^① “人在回路”意即人始终保持处于算法回路之中,强调在人工智能的开发应用过程中充分发挥人类智能对机器智能的控制、辅助与干预作用,从而提升智能系统的精准性、适应性与合规性。

从而得以将税收征管的基本原则真正内化为税务人工智能所根本遵循的总章程。

此外,人类自主性是数字正义的究极防线。如果能够找到实现任何既定价值目标之可计算和可预测的有效方法,那么构建公平、透明和可问责的税务人工智能与税收征管的理性化追求完全一致^[24](P1372)。但理想终究输给了现实,客观上的技术瓶颈和主观上的归责困境决定了,即便经训练得到了高性能的税务人工智能,在实际部署应用中仍应遵循比例原则谨慎把握其嵌入税收征管的边界与限度。税务机关对人工智能的使用须以不逾越现有税收征管权义配置为前提,细分场景考量技术方案与征管目标之间的适当性与必要性,防止对其他税收价值抑或纳税人权利造成不相称的损害。对此,总体上可从以下三项标准加以把握:一是服务型应用宽于管理型应用。服务型政府建设重在创新政府管理方式,寓管理于服务之中。税务机关在对各类税收开展征收管理之余,亦负有宣传税收法律法规、普及纳税知识和提供纳税咨询服务等职责。此类服务行为具有明显的增益性,人工智能的赋能更是能够显著提升服务的便捷性、个性化与精准度。即便出错,经过事前的提醒与事后的补救,风险的实害化几率也能被降到最低。二是介入终局性行为较之过程性行为更为审慎。一项直接影响纳税人权益的征管决定在被正式作出之前,通常需要经历多个程序阶段,例如基于风险预警调查核实后方可下达处罚。税务人工智能无论是否主导过程性行为,其输出结果均仅作为正面信息来源或者反向预警参考而不产生最终效力,亦不排斥人工监督与矫正;终局性行政决定的作出则必须坚持人工智能的辅助性工具定位,确保由税务人员掌握对外决策权,以维系具体行政行为的合法正当。在域外立法例中,《德国税收通则》第155条第4款允许税务机关使用全自动流程作出执行、更正、撤回、撤销、取消或修改等操作的事项就限于纳税评估以及预缴预扣税款两类过程性行为,且前提是没有理由需要由税务人员亲自处理个案。三是与量化情景的契合度高于质化情景。由于“人工智能不可能超越理性计算范畴”^[25](P196),在税额核算等可以充分量化的情景中固能通过对偶然性和差异性因素的排除发挥更佳效用,然而,一旦触及像反避税中的合理性认定、税务处罚中的酌情裁量等需要进行实质判断与价值衡量的场域,不免黔驴技穷。

(二) 法治路径:权利保障与义务前置双管齐下

税收法治是税收治理现代化的精髓。鉴于法律运行的全部过程均围绕着权利和义务而展开,智慧税务的法治化路径既应秉持融贯主义将对税务人工智能的矫正纳入改造后的纳税人权利保护框架,又须立足工具主义通过风险管理义务的赋予强化对齐风险的事前预防^[26](P31-36)。

1. 权利保障夯实价值根基

在法治轨道上推动征纳双方的税权平衡是依法治税的核心。兹因税务机关与纳税人之间的不平等地位,为防范征管行为对纳税人基本权的过度干预,税收征管法必须在正当程序中贯彻纳税人权利保护之精神,从着眼于征收管理的“权力之法”转向以纳税人为中心的“权利之法”。《意见》也强调大力推行优质高效智能税费服务,切实维护纳税人缴费人合法权益。现实却是,随着税务人工智能在税收征管改革中的嵌入程度不断拓展深化,现有纳税人权利保护框架与智慧税务实践的错配与脱节日臻明显。税收法治原则构筑起来的限制税务机关权力的藩篱被机器决策带来的权力扩张和责任规避所突破,量能平等负担的理想面对智能系统的泛化局限和偏见歧视沦为泡沫,自动化征管的瞬时性、不可控性、不透明性更是打破了传统人工参与下税收征管程序的渐进性、可控性和公开性,正当程序原则所要求的纳税人知情权、陈述申辩权、听证权、救济权等在一定程度上被架空^[27](P81-82)。究其根本,主要是对征管效率的片面追求凌驾于纳税人权利之上所致。

虽然说征纳双方就税收正义的追寻组成共同体,但基于程序上的优势地位,最终责任仍归征管机关^[28](P44)。税务机关在税法价值发生冲突时作出的优先性抉择,很可能通过反馈被具有迎合属性的税务人工智能作为一种偏好在输出。因此,在智慧税务建设过程中,税务机关纳税人权利保护意识的重申和强化至关重要,从而在多目标对齐过程中确保纳税人基本权利处于更高位阶,不致使现有纳税人权利体系因税务人工智能的应用而陷入“技术利维坦”的泥淖。当然,传统的纳税人权利框架已

不足以回应智慧税务进程中数据和算法加持下明显向税务机关倾斜的税收征管格局,需要在征管规范层面针对性保障纳税人的信息隐私权益,并成比例地赋予纳税人抵御税务算法的新权能^[29](P86)。欧盟制定的《通用数据保护条例》序言第71条就言明,数据控制者虽被允许根据法律授权在欺诈和逃税预防目的项下开展基于自动化处理的决策,且为实现此类重要公共利益可对数据主体之数据权利进行必要而相称的限制克减,但须以尊重基本权利和自由的核心内容为前提,并受到一系列适当保障措施约束,包括向数据主体提供特定信息并赋予其获取人工干预、表达观点、获得解释以及提出质疑的权利。

涉税数据是税务人工智能正常运行的“原材料”,特别是高质量涉税数据的广泛收集利用对于提升税务人工智能的稳健性和精确性意义重大,但纳税人的信息安全在此过程中也存在着一系列违规处理的明患隐忧。以过度收集为例,斯洛伐克金融管理局为打击税务欺诈,在2019年推出了将所有收银机在线连入其系统的eKasa项目,但随后受到议员团体的反对并被诉至宪法法院。原告方的核心主张之一便是,eKasa项目要求为买方开具唯一标识符从而使税务机关获得某一消费者的完整购买历史记录,但其中一些数据与税务合规目的无关,故违反了数据最小化原则。对此,法院裁定为买方开具唯一标识符不符合数据保护要求,回应了原告方关于在处理个人消费习惯数据时可能产生偏见和歧视风险的担忧(Constitutional Court of the Slovak Republic, PL. ÚS 25/2019-117)。回到我国现有规范,《税收征收管理法》第6条聚焦涉税信息的利用而忽视信息权利之保护,较早的《纳税人涉税保密信息管理暂行办法》(国税发[2008]93号)则仅关注涉及纳税人商业秘密和个人隐私的信息,无法覆盖对一般纳税人数据的保护。随着《个人信息保护法》的颁布实施,《税收征收管理法》亦应同步跟进,对纳税人信息处理作出适应征管实践的系统性规定,尤其需要强化对纳税人信息权的保障,克服2015年版修订草案征求意见稿虽设“信息披露”专章却仍片面强调增强税务机关涉税信息获取能力的不足。

回应税务人工智能自动化决策的算法权利构建,则可经由对原有纳税人权利权能的扩充改造来实现,其中陈述申辩权无疑堪当一个有效的程序抓手。作为程序正当原则赋予纳税人的一项基本权利,陈述申辩权的行使横亘于完成基本流程手续后形成初步处理意见和正式下达税务决定之间的关键节点,在纳税人据理力争维护自身合法权益的过程中以程序正义铸牢了结果非正义的滤网。智能算法的介入非但不构成适用陈述申辩程序的例外事由,反而因价值偏离可能引发的不合法、不公平、不准确风险更凸显出强化必要性。在前端,税务机关的告知是纳税人针对性提出陈述申辩的前提。智利《税法典》在其“纳税人权利”一节就新增要求,不论自动化征管算法使用与否,纳税人均有权要求税务机关明确说明采取相应行为的理由,包括表述事实、法律依据以及得出结论的推理逻辑。一旦税务人工智能参与了决策,税务机关除了原有的告知内容,亦需在部署阶段的系统性说明基础上就算法在个案中的适用履行额外说明义务^[30](P30),从而保障纳税人的知情权与解释权。在中端,传统的纳税人陈述申辩权带有明显的消极防御色彩,表现为受到税务机关告知“刺激”后对不利意见的竭力辩驳。面对税务人工智能带来的强烈价值冲击,应允许纳税人采取更加主动的姿态,在质疑结果之余还能够基于自身权益考量选择拒绝自动化决策,从而免受算法能力局限以及随之而来的救济难题之扰。英国《数据保护法》第14条对经依法授权之自动化决策所确立的一系列保障措施中,就明确数据主体可以在收到数据控制者关于重大决定通知后的一个月内,要求数据控制者重新考虑该决定或者作出一个不完全基于自动化处理的新决定。在后端,税务机关根据纳税人的陈述申辩对税务人工智能所作决定的审查复核,恰与前述“人在回路”的技术要求不谋而合,体现出法技秩序的和谐统一。

纳税人权利保护的观念强化与框架更新,最终得以为税务人工智能的“价值导向设计”提供前提。通过将抽象的税法原则和纳税人权利经由规范化转换为依赖情境的具体设计需求^[31](P2-3),实现从单纯关注客观技术风险向落实主体权利的跨越,进而弥合征管技术进步与纳税人主张之间的隔阂。纳税人的诚实协作意愿亦会因权利保护的周全而显著提升,表现为愿意提供更多精确数据和详细内容,促进征管效率与实质平等^[32](P18)。

2. 义务前置平衡创新风险

由于天然势差的存在,仅仅赋予纳税人权利不足以对抗人工智能赋能下日益强大的征管权力。完全寄希望于税务人员的监督控制,也受限于其能力不足和趋避动机而并不现实。是故,以制度化方式强化智慧税务的内部治理与外部约束不可或缺。特别是税收征管智能化改造给纳税人权利带来的不再是单一的个体性风险,而是快速变化和高度不确定的系统性风险,具有波及范围广、侵害形式新、利益衡量难、救济难度大等特征。“基于权利的方法”旨在通过合法性的二元判断为所有个体提供不可谈判的最低保护水平,在风险数字社会背景下难免力绌难支,促使其转向“基于风险的方法”,即遵循一种精细粒度的可扩展性逻辑,采取预防性、前瞻性、主动性的态度和措施来管理风险^[33](P2-3)。

作为指导和控制一个组织应对风险的协调活动,风险管理的目的是创造和保护价值,在智慧税务场景中要求智能系统的开发者与部署者将对纳税人权利的关切及其与课税目标之间的权衡转化为组织层面的应对策略,使用风险级别作为履行特定前置性义务的阈值,在确保比例性的前提下落实对税法价值风险的事先识别、评估与因应^[34](P247-249)。然而,就我国立法实践而言,虽然从国务院印发的《新一代人工智能发展规划》(国发[2017]35号)中已可管窥这一思路,强调“增强风险意识,重视风险评估和防控,强化前瞻预防和约束引导”,但统一的人工智能监管规则至今阙如。相较来看,欧盟最新颁布的《人工智能法》延续了《通用数据保护条例》的风险规制进路,明确要求建立、实施、记录和维护与高风险人工智能系统有关的风险管理系统,并有区别地规定了提供者、部署者以及其他各方的义务,可资镜鉴。尽管欧盟《人工智能法》在其第59条序言中指出专门用于税务和海关部门行政程序的人工智能系统不应被归类为执法当局用于预防、检测、调查和起诉刑事犯罪目的的高风险人工智能系统,且在附件三中亦未予以列明。但一方面税收征管的行政程序本就与涉税犯罪的刑事程序紧密关联,另一方面依赖生物识别来辨识纳税人、直接或间接地以税收抵免或各类扣减的形式发放福利在现实中也十分常见。考虑到引发基本权利问题的可能性,即使是出于纯粹行政目的而使用税务人工智能,有关部门也理应履行前述风险管理的积极义务^[35](P56-57)。通过施加意识增强和风险量化相结合的全流程治理义务,不仅可以督促税务部门以真正值得信赖和以人为本的方式开发和部署智能化系统,还能够显著促进问责制^[36](P18),在算法黑箱之外重新建立起征管机关的行为表征,从而为行政争议的违法性判断提供直接依据,以此使征纳双方的举证责任回复至平衡状态,可以有效防止税务部门遁入向技术推卸责任的代理模式。

当然,内部治理离不开外部约束,有效手段之一便是建立独立抑或复合的人工智能监管机构,用权力制约权力。以法国为例,自2014年采用“欺诈识别与请求评估系统”以来,其数据保护监管机构国家信息与自由委员会就一直在积极监测税务部门使用人工智能算法的进展,对每个由财政部推动的新工具都进行了数据保护影响评估,并就防范所识别的风险推荐了措施。西班牙则在2023年建立了欧盟首个独立的人工智能监管局,将自行通过质量和责任认证来监督人工智能标准的合规性。

值得注意的是,前置义务的赋予及其问责必须平衡缓解价值风险的效益与引发监管执法创新寒蝉效应的成本。税务领域的监管科技同样需要在实验性治理理念下借助沙盒机制等被赋予一定的试错和容错空间,使兼具被监督者身份的税务机关在开发部署税务人工智能过程中,得经由对实施经验充分的递归审查进行持续的修正迭代。

就像歌德笔下的魔法师学徒,税务部门在智慧税务建设进程中精神奋发地召唤出智能精灵的神奇力量,希望它自主而又完全顺从地按照旨令行事,作出一桩现代化税收治理的奇迹,然而一旦意识到税收征管的法治、公正和效率诸原则被如同脱缰之马的数据与算法洪流冲击漫漶,只能求助于技术与法治两位“师傅”以重新获得对这个“仆人”的掌控,引导其实现与税法价值的对齐,在既定的规范和程序轨道

上不偏不倚地前进^①。但价值对齐要求不等于因噎废食,法律系统的包容审慎在技术发展初期应被视为一种适当策略,重要的是秉持风险控制与鼓励创新相平衡的理念,防止陷入税务人工智能开发应用的“科林格里奇困境”^②。“技术解决主义”则是另一个需要提防的陷阱,“只要正确的算法到位,所有复杂的税收征管情形就可以被重构为具有明确可计算解决方案之清晰问题或是可以轻松优化之透明直观过程”的想法并不现实^[37](P5-6)。在税务执法和合规中使用人工智能无法弥补税法本身的缺陷,务必要认识到税收政策与税务管理之间更广泛的动态共生关系,避免本质上属于政治决策的事项向技术方案逃逸,方可最大化税务人工智能应用的整体效益^[38](P119-120)。

参考文献

- [1] 习近平. 发展新质生产力是推动高质量发展的内在要求和重要着力点. 求是, 2024, (11).
- [2] John J. Nay. Law Informs Code: A Legal Informatics Approach to Aligning Artificial Intelligence with Humans. *Northwestern Journal of Technology and Intellectual Property*, 2023, 20(3).
- [3] 托马斯·库恩. 科学革命的结构. 金吾伦、胡新和译. 北京: 北京大学出版社, 2012.
- [4] Samuel Butler. *Erewhon or Over the Range*. London: Trübner & Co., 1872.
- [5] Norbert Wiener. Some Moral and Technical Consequences of Automation. *Science*, 1960, 131(3410).
- [6] Ibo van de Poel. Werthaltigkeit der Technik// Armin Grunwald. *Handbuch Technikethik*. Stuttgart: J. B. Metzler, 2013.
- [7] James B. McKee. *Introduction to Sociology*. New York: Holt, Rinehart and Winston, 1974.
- [8] 卡尔·拉伦茨. 法学方法论. 黄家镇译. 北京: 商务印书馆, 2020.
- [9] H. L. A. Hart. *The Concept of Law*. Oxford: Oxford University Press, 2012.
- [10] Michael D. Bayles. *Principles of Law: A Normative Analysis*. Dordrecht: D. Reidel Publishing Company, 1987.
- [11] Taina Bucher. *If... Then: Algorithmic Power and Politics*. New York: Oxford University Press, 2018.
- [12] 迈克斯·泰格马克. 生命3.0. 汪婕舒译. 杭州: 浙江教育出版社, 2018.
- [13] Hutan Ashrafian. Artificial Intelligence and Robot Responsibilities: Innovating beyond Rights. *Science and Engineering Ethics*, 2015, 21(2).
- [14] Rik Peeters, Arjan Widlak. The Digital Cage: Administrative Exclusion through Information Architecture——The Case of the Dutch Civil Registry's Master Data Management System. *Government Information Quarterly*, 2018, 35(2).
- [15] 马长山. 数字法治政府的机制再造. 政治与法律, 2022, (11).
- [16] 刘剑文. 税收征管法. 武汉: 武汉大学出版社, 2003.
- [17] 李平. 建立与新质生产力相适应的税收征管体系. 中国税务报, 2024-04-03.
- [18] Xavier Ferrer, Tom van Nuenen, Jose M. Such et al. Bias and Discrimination in AI: A Cross-disciplinary Perspective. *IEEE Technology and Society Magazine*, 2021, 40(2).
- [19] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena et al. A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*, 2021, 54(6).
- [20] 黄茂荣, 葛克昌, 陈清秀. 税捐稽征法. 台北: 元照出版公司, 2020.
- [21] George Ritzer. *The McDonaldization of Society*. Los Angeles: Pine Forge Press, 2008.
- [22] Stuart Russell, Peter Norvig. *Artificial Intelligence: A Modern Approach*. Hoboken: Pearson, 2021.
- [23] 黎江虹, 李思思. 人工智能嵌入税务执法的应用风险及法律限度. 税务研究, 2024, (6).
- [24] Jakob Mökander, Ralph Schroeder. Artificial Intelligence, Rationalization, and the Limits of Control in the Public Sector: The Case of Tax Policy Optimization. *Social Science Computer Review*, 2024, 42(6).
- [25] 陈敏光. 司法人工智能的理论极限研究. 社会科学战线, 2020, (11).
- [26] Roger Brownsword. *Law 3.0: Rules, Regulation, and Technology*. Abingdon: Routledge, 2021.

① 叙事诗《魔法师的学徒》由歌德创作于1797年,讲述了一个魔法师学徒趁师傅外出擅自施展咒语将笤帚变成仆人替他汲水,却不知道如何终止,最终导致房子被水淹没,不得不向师傅求助才得以复原的故事。

② 这一控制困境可概括为:试图控制一项技术是困难的,甚至往往是不可能的。因为在技术尚可控的早期阶段,人们对它可能造成的社会危害知之甚少,从而不足以支持对其发展的控制;而当这些危害后果变得明显时,控制却变得代价高昂且实现缓慢。

- [27] 黎江虹,李思思. 重塑纳税人权利:“以数治税”时代的底层逻辑. 华中科技大学学报(社会科学版),2022,(6).
- [28] Rolf Wittmann. Mitwirkungspflicht und Aufklärungspflicht in der AO: Reduktion der Mitwirkungspflicht durch finanzbehördliche Verhalten. *Steuer und Wirtschaft*, 1987, 64(1).
- [29] 茅孝军. 人工智能时代税收征管的功能转型与法治因应——以欧盟税务人工智能实践为切入点. 税收经济研究, 2024,(3).
- [30] 赵鹏,张硕. 论自动化行政决定的说明义务. 山西大学学报(哲学社会科学版),2024,(2).
- [31] Evgeni Aizenberg, Jeroen van den Hoven. Designing for Human Rights in AI. *Big Data & Society*, 2020, 7(2).
- [32] 迪特尔·比尔克. 德国税法教科书. 徐妍译. 北京:北京大学出版社,2018.
- [33] Raphaël Gellert. *The Risk-based Approach to Data Protection*. Oxford: Oxford University Press, 2020.
- [34] Tobias Mahler. Between Risk Management and Proportionality: The Risk-based Approach in the EU's Artificial Intelligence Act Proposal//Liane Colonna, Stanley Greenstein. *Nordic Yearbook of Law and Informatics 2020-2021: Law in the Era of Artificial Intelligence*. Stockholm: IRI, 2022.
- [35] Bruno Peeters. European Law Restrictions on Tax Authorities' Use of Artificial Intelligence Systems: Reflections on Some Recent Developments. *EC Tax Review*, 2024, 33(2).
- [36] Alessandro Mantelero. The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, Legal Obligations and Key Elements for a Model Template. *Computer Law & Security Review*, 2024, 54.
- [37] Evgeny Morozov. *To Save Everything, Click Here: The Folly of Technological Solutionism*. New York: PublicAffairs, 2013.
- [38] Rita de la Feria, María Amparo Grau Ruiz. The Robotisation of Tax Administration//María Amparo Grau Ruiz. *Interactive Robotics: Legal, Ethical, Social and Economic Aspects——Selected Contributions to the INBOTS Conference 2021, 18-20 May, 2021*. Cham: Springer, 2022.

On the Value Alignment between Law and Technology In Smart Taxation

Xu Duoqi, Dong Jiajie (Fudan University)

Abstract In the context of developing new quality productive forces, AI-enabled smart taxation needs to prevent deviations from the value orientations and basic principles of tax law. The issue of AI alignment, stemming from the enhanced AI autonomy brought about by the evolution of technological paradigms, is essentially a kind of value alignment based on objectives and intentions, which is manifested as the tension reconciliation between AI technology and the rule of law in the smart transformation of tax collection and administration. From the perspective of specific application scenarios of smart taxation, the power-seeking tendency, agency ineligibility and absence of internal and external accountability of AI will impact the principle of the rule of law in tax administration, while the insufficiency of generalization ability and the feedback loop of bias will challenge fairness and the realization of the efficiency may be hindered due to hallucinations about AI. In this regard, it is necessary to not only strengthen the technical response based on data optimization and human-in-the-loop approaches, but also explore the path of the rule of law that protects rights and prioritizes obligations, and avoid the trap of technosolutionism with an inclusive and prudent approach.

Key words smart taxation; tax collection and administration; artificial intelligence; principles of tax law; value alignment; taxpayers' rights; algorithmic decision-making; algorithmic governance

-
- 作者简介 许多奇,复旦大学法学院教授,上海 200438;
董家杰,复旦大学法学院博士研究生。
- 责任编辑 李媛