第四次工业革命中的司法转型与数字正义

站在第四次工业革命的十字路口，人工智能、大数据和云计算等技术正带来一场司法文明的范式变革。一方面，新技术通过提升司法的智能化水平，增强了司法在社会中的治理效能，有利于满足公众对司法功能的期待。另一方面，新技术的介入也使得司法运作的逻辑和流程发生了改变，并带来系列法律、技术、伦理以及司法功能异化的问题，尤其是在深层次对司法领域的“人机关系”形成冲击。面对司法智能化转型的大趋势，我们需要在价值层面坚守法治的底线，防止技术理性侵蚀人文关怀的根基，共同探索数字正义的实现路径。有鉴于此，《探索与争鸣》编辑部与上海对外经贸大学法学院共同举办圆桌会议，邀请相关领域专家围绕数字司法图景、智能化实践探索、司法价值坚守等核心议题进行深入讨论。

傅郁林教授指出，AI 法官是人工智能技术运用于司法活动的终极状态，而当下 AI 的发展程度距离担当起核心审判职能所需的法律专业化水平还有较大差距，但其在技术上仍具有可行性。而若要在制度上赋予 AI 法官以审判主体权限，由于司法所蕴含的人性特质与 AI 法官所欲实现的替代功能存在冲突，所以很难实现。**王福华**教授认为，数字技术给人类带来便捷和效率的同时，也衍生出诸多“数字不正义”现象。要通过纠正数字不公、修复数字侵害、制衡技术权力、实现程序正义和促进数字包容等途径，实现数字的矫正正义。**唐力**教授强调，人工智能的司法应用可能侵蚀人类作为司法主体的“主体性”与“创造性”，需要在人工智能司法应用的伦理准则、“人机关系”定位、应用范围，以及技术安全保障等维度采取具体措施，明确人工智能司法应用的限度。**许多奇**教授以智能化赋能金融案件“四大检察”为分析场景，认为技术层面的数据壁垒、算法局限，机制层面的协同不畅、权责不清，以及伦理和规范层面的价值冲突与法律滞后等问题，构成制约智能化赋能检察的核心。需要在破除数据壁垒、精准适配算法、筑牢安全机制等核心举措的基础上，同步消除金融案件办理过程中跨域协作、权责划分、规范适用等制度性障碍。**张兴美**教授以民事司法为例，指出当下的人工智能正作为智能的“第四方”在纠纷解决层面发挥了实质性作用，但也应警惕技术至上主义对民事司法的异化。应该秉持司法人类主体性理念、注重法官参与、坚持司法公开、加强系统监督，推动人工智能时代司法的公平正义。**黄泽敏**副教授构建了司法领域智能科技的介入的强式、中等和弱式三种标准。其中，中等和弱式应让位于强式标准，强式标准强调凡涉及法律判断的，智能科技不可介入。强式标准之所以成立，是因为除了形式要素之外，法律判断还复合了价值、道德、情感和权力等要素，而人工智能不能也不应该介入这些要素。**季平平**研究员指出，人工智能技术的发展为裁判文书的结构优化提供了可能路径。然而，由于法律推理内含事实认定与价值判断的双重属性，人工智能仍难以胜任具有规范创设与理由展开功能的文书自动生成任务。人工智能的合理定位，应是技术辅助工具而非替代者。在促进裁判文书结构变革的同时，应尊重法官对法律解释的最终权限。

——主持人　**孙冠豪**

AI 法官的底层逻辑与法律边界

傅郁林，北京大学法学院教授

法官在司法活动中的核心角色，意味着“AI 法官”注定是人工智能技术运用于司法活动的最高境界和终极状态。目前关于 AI 法官角色相对于人类法官的辅助性抑或替代性（独立性）的学术争论，主要集中在技术维度的可行性，也涉及伦理维度的可容性，但极少关注情感维度的可欲性。然而，关于人类司法的本质与 AI 法官底层逻辑的追问，却是探讨 AI 法官的“合法性”及其法律边界的基本前提。本文将从三个层面探讨这一问题：司法职能的基本定位与评价体系、司法职能的可分性及其实现途径的可替代性、人类法官实现司法职能的特质及其可替代性。

司法功能评价体系的人性基底与 AI 法官的角色悖论

司法职能的基本定位，是将具有普遍约束力的法律规范在具体案件中应用于经正当程序查明（认定）的事实，并据此作出裁判（或在自愿的基础上达成调解协议），以此保护受侵害者的权利，修复被纠纷破坏的秩序，实现社会公平正义。质言之，司法职能的制度定位是认定事实、适用法律、作出裁判，其直接功能是解决个案纠纷、实现个体正义，附带功能是维护或修复社会秩序、确立社会行为规则，间接功能是巩固政治合法性、实现社会团结与和谐。在上述所有层面上，裁判的合法性或正当性都构成司法职能评价体系的基准，无论从裁判结果的个体（当事人）可接受性还是社会效果而言，均无二致。

仅从裁判技术而言，尽管孟德斯鸠的司法权机械论和韦伯关于形式理性的“自动售货机”比喻受到广泛批评，但现代司法追求高度系统化、逻辑化、去人格化的特征已成为普遍趋势。司法职能及其实现方式的逻辑基础，存在一套已经由成文法或判例法加以确定的统一先验性标准。司法过程则在特定的裁判逻辑框架内，通过特定的程序架构，适用特定的证据规则，查明个案事实，运用该法定标准进行评价，从而获得符合社会正义的裁判结论。就此而言，AI 司法功能的阈值主要取决于 AI 技术条件的发展

状况：AI 法官替代人类法官逐步承接其承担的部分乃至全部审判职责，其技术上的可行性只是时间早晚的问题。诚然，法律、法律论据以及判决理由都是用语言表达的，AI 司法的瓶颈在于，作为大数据基础技术的自然语言处理（NLP）技术的原理和机制还不够成熟。此外，通用的知识图谱构建技术在进入司法场景后仍固守原有逻辑，AI 司法所依赖的算法与法律专业知识无法实现深度融合。然而，据保守估计，AI 完全掌握自然语言处理（NLP）技术只需要几十年的时间，通过构建模拟神经元（即大脑）的人工神经网络的深度学习过程还可能明显缩短这一时间；而 AI 针对法学专业知识进行量身定做的升级迭代也在日新月异。通过算法的发展与运用，无论是获取、分享或审核证据、事实信息的能力，还是海量检索并精准适配法律规范和司法先例的能力，AI 法官能够替代甚或超越人类法官，在技术上的可行性受到社会各界的普遍认同。

然而，从司法的性质、功能及其实现方式而言，无论是法官进行事实判断所依赖的自由心证原则，还是法律赋予法官进行实体处理时所享有的自由裁量权，抑或临时措施、紧急救济机制等无处不在的司法衡平，法官作出符合真相的事实判断和合乎正义（社会正义乃至自然正义）的决定，都依赖其作为人类的心智和良心，都要借助人类的经验和对同类的共情。在司法的合法性评价体系中，“同类审判”（trial by one’s peers）恰恰是陪审团审判的正当性基础，陪审团裁判无需说理，并超越于司法理性，免予上诉审查，却可以获得难以被推翻的合法性。

同样，法律解释作为法官适用法律的职责的内在权限，也不可能超越人类的共情和价值共识。无论是在职业法官独享裁判权、以成文法为主要渊源的欧陆国家，抑或是在职业法官与陪审团共享裁判权、以判例法为主要渊源的普通法国家，法官在司法职能中都普遍担当着“法律的嘴巴”的角色。法官由于智力、能力、阅历、信仰、文化背景、道德倾向、政治观点等因素的影响，无论其法律职业训练如何成功、职业伦理境界怎样超脱，都无法脱离人性底色，因而其可能在审判中加入个人偏见，导致审判结果的不确定性。这当然是法治主义者通过精妙的程序设计力图避免和减少的缺陷，甚至也因此为算法在确保裁判一致性和司法公平性方面的优势提供了富有说服力的支撑。然而，姑且不论算法与人类一样也可能存在偏见，并且算法缺陷一旦产生，通过溯源来矫正更加困难，更为关键的问题在于，人类法官是人类的一部分，无论其裁判是否存在偏见，以及偏见是如何发生的，人类法官作出的裁判与接受裁判的人类（当事人）及评价其裁判的人类（社会公众）之间总会形成共鸣或张力，这是同属于人类的心智体验和以此为基础而产生的社会反馈，而这种反馈恰恰是司法的合法性或正当性的主观评价体系的内在机理。一旦丧失这种基于人类心智和情绪的“同类”评价机制，即使算法可以将主流价值观、伦理规范植入司法过程中，并且能够根据社会反馈随时调整算法，也会因超越人性的底层逻辑和确定且一致的司法结果而形成僵化的司法方法。其司法功能即使得以如期实现因而证明是可行的，在情感上也会是荒谬的、不可欲的。

综上，讨论 AI 法官局部或全面替代人类法官的可能性，其逻辑前提不仅包括技术维度的考量，更应包括价值、伦理、心智、情感维度的考量，不仅要评估人类法官承担的司法职能是否“可能”被 AI 所替代，而且要评估其在何种限度内“可以”被替代，尤其需要评估哪些司法职能的履行和司法功能的实现在性质上依赖人类特质和人性力量，因而即使技术上可能，制度上也不可以，即使技术上可行，制度上也不可欲。特别值得关注的是，随着司法系统应对专业化需求、效率化压力、技术化革新的适应性调整，司法职能在悄然无声的动态变革中仅剩下“处理案件”，而不再是实现社会正义、修复社会秩序、提升政治凝聚力的重要途径。倘如此，则 AI 法官相对于人类法官可能更具优势，其全面替代人类法官也为时不远了。因此，为了应对社会发展的频率、高强度的冲击，司法功能及其实现方式不得不努力回应和变革时，很有必要追问这种动态发展的边界和底线究竟在哪里。

司法职能的可分性与 AI 法官角色替代的逻辑前提

司法功能评价体系的人性基底与 AI 法官角色定位的逻辑冲突，意味着 AI 法官全面替代人类法官因受到来自生命基础和伦理维度的阻碍而尚不可能，但这并不影响司法 AI 化或 AI 法官替代人类法官目前承担的部分司法职能。实际上，“人案矛盾”的全球司法困局与传统审判效能的结构性桎梏已成为 AI 法官替代人类的内在动因，并且中国的超职权主义诉讼模式对 AI 法官的发展空间而言，一方面更具有现实迫切性，

另一方面也具有更大的可行性。因为无论是普通法国家的当事人主义诉讼模式，还是德日以辩论主义为基础的职权主义诉讼模式，抑或以当事人进行主义为特色的法意诉讼模式，其流程化、一元化、去人性化的特征受制于当事人程序参与权乃至程序控制权，在 AI 化进程中可能遇到更多结构性的障碍。

然而，AI 法官替代人类法官承担部分司法职能并在司法评价体系中获得正当性的逻辑前提，是确立司法职能的可分性及其可替代性的界定标准。目前关于 AI 法官的角色，形成了“辅助说”与“替代说”两种主要观点，但争议双方在方法论上都是以司法职能和法官职权的完整性为逻辑前提的。这一固化思维，在我国进行审判人员分类改革中已经成为建立司法“辅助”职能由司法辅助人员“独立”行使的审判权分享机制的逻辑障碍，当下探索建立 AI 法官与人类法官之间制度化的审判权限分享模式，必须突破传统的将审判权作为一个整体考量的思维定式，继而建构审判权可分性理论，这是建立审判权分享机制从而实现 AI 法官角色替代并获得正当性的理论基础，也是确立 AI 法官职能与权限的法律边界的逻辑前提。

整体上，法院作为司法机构所承担的司法职能，换言之，由法院行使的审判权实际上分为两大板块，即由法官专享的核心职能（或专属审判权），以及由法官以外的各类司法辅助人员分享的审判辅助职能。但这种划分不仅是抽象的、粗糙的，而且从比较法视角和动态视角来看只是相对而言的。因为各国对于“法官”的定义及其与司法辅助人员的划分标准差异很大，在一个法域中被列入正式司法体系的（较低等级的）法官，在另一法域被列入正式或非正式司法体系的司法辅助人员，有可能行使着相同或类似的审判权。比如，普通法系国家的治安法官（peace justice）、审裁官（magistrate）等司法人员，虽然是被排除在正式司法编制和司法统计数字之外的“法官”，却行使着大陆法系国家普遍由基层法官行使的审判权。这些无论名称是否相同但实质工作相同的审判人员，均可归类为“限权法官”。

从 AI 法官的替代角色研究价值来看，值得关注的不是其名称是否为法官，而在于对限权法官的权限范围及其权限行使方式作具体区分。限权法官审判权限范围的界定，既可能是按“案件”类型进行划分，也可能按同一案件中由限权法官所处理的“事项”进行划分。笔者根据限权法官的权力行使方式及其权限的独立性存在的差异，将可以在法律规定的有限审判权范围内独立行使审判权的那一类“限权法官”称为案件受限型的限权法官或事项受限型的限权法官（如美国的小额法官和中国审判人员分类改革前的助理法官）；而那些明确规定或默认权限法官必须在法官监督或指导下履行职务——无论该职责范围是否由法律以明确或模糊的文字予以规定——而未赋予其在法定权限范围内“独立行使”其有限审判权、相应地也在这一权限范围独立承担该职责的那一类“限权法官”，称之为非独立的限权法官或辅助型限权法官（如作为美国法官私人助理的 law clerk 和中国审判人员分类改革后的法官助理）。

此外，在某些司法制度中，同一（类）限权法官可能同时履行这两类职责（但运行良好的制度会对这两类职能的范围及其相应权限的运行方式进行明确界分），甚或同一正式法官也可能兼具权限法官的角色从而兼具这两类职责。而法律并非依特定“人”而是依据特定“角色”履行特定职责的需要配置相应的权限，比如美国一些州法官同时也担任小额法官（周末法官或夜间法官），但同一法官在普通诉讼程序中（作为正式法官）所享有的对藐视法庭行为的裁判权，在小额诉讼程序中（作为治安法官）就不能行使。

在动态视角下，即使在同一法域内，法官与司法辅助人员的分野及其承担的司法职能在近几十年来正在发生制度性或惯例性的移转，总体趋势是传统的由法官独享的司法权日益转移给司法辅助人员，或

者换个角度说，司法辅助人员在司法职能中担当的角色日益趋近于法官，从而在某些职能上完全替代了法官的角色。比如，美国初审法院中审裁官（magistrate）日益扩张的裁判权、某些联邦上诉法院中 staff attorney 对于占全部上诉案件 70% 的“不发表判决书”的案件享有的近乎独立的裁判权。而且这些权限的扩张是制度性或惯例性的，甚至是成建制的、体系性的，如美国的某些州、加拿大的某些省近几十年来将治安法院收编入正式司法体系。中国的发展动态表面上与前述各国的普遍趋势形成反差，如审判人员分类改革之前本属于法官系列的助理审判员，却在改革后转变为法官助理从而被列入司法辅助人员。但是就改革前由法官承担的司法职能更多地转移给司法辅助人员这一动态趋势而言，中国与前述其他国家的情况并无二致，只是改革后的法官助理在与改革前的助理审判员同样承担传统上由法官独享的司法职能时，其制度性的角色定位更加模糊，制度层面的独立性也更少。[①]

尽管法官与司法辅助人员的划分及其权限配置模式存在比较法上的多样性和发展动态上的角色交错，但关于审判权的可分性与可替代性问题仍存在一些基本共识。首先，终局性裁判权被视为核心审判权，只能由被定义为“法官”的审判人员专享；案件受限型的限权法官在法律规定的案件范围内适用同一准则，而无论其名称如何。其次，事项受限型限权法官应当在法定职能范围内享有独立的权限，并承担独立的职责，即使其自身被归类为司法辅助人员，或其履行的职能被定义为司法辅助事务，如某些法域明确赋予司法辅助人员（书记官）就诉讼费用及其分担，以及审前案件管理中的程序性分流（如在“多门诉讼”机制下将案件分流到不同程序轨道）等类似程序事项的独立决定权。此外，财产保全也被各国普遍归入权限法官独立决定的权限范围，但被我国归入行为“保全”体系的禁令却属于法官的实体裁判权，因而即使做出非终局性的中间禁令和临时禁令，也属于法官的核心审判权。为此，在赋予权限法官对小额诉讼和轻罪案件完整审判权的司法制度中，案件受限型限权法官与（员额）法官的职能区分，尽管在审判人员分类改革时用于界定各自的职权和职责时有意义——且这种界定主要取决于“法官”与司法辅助人员的定义标准，但对于 AI 法官与人类法官的角色区分意义不大，因为就其独立裁判所依赖的人性要素而言，小额案件与大额案件、简易案件与复杂案件、非讼案件与诉讼案件之间并没有普遍性、规律性的差异。

因此，以下关于 AI 法官审判主体权限的讨论将聚焦于 AI 审判者与人类审判者之间的职能划分：一方面，目前由法官专享的裁判权——事实判断与认定、法律选择与解释，以及基于上述所得出的裁判结论——可否由 AI 法官替代人类法官行使（而不论法官行使该裁判权的案件类型或程序）；另一方面，那些并非核心审判权的司法辅助事项，虽然可由司法辅助人员分享，但其中仍有些职能只能由人类司法辅助人员承担或采用人机合作模式承担，而不能由 AI 独立承担或替代人类完成。不过，司法辅助职能的全面 AI 化无论是基于技术上的可行性还是价值伦理上的可欲性，主要取决于 AI 能力发展的阈值。限于篇幅，此处不赘述。下文仅在最狭窄的范围内讨论目前法官专享的裁判权可由 AI 法官替代的技术限度与法律边界。

AI 法官审判主体权限的技术可能与制度边界

AI 法官的审判主体权限受到技术、伦理、法律及司法实践需求等多重因素的制约。在技术维度上，AI 司法功能的阈值应由客观技术条件决定，即 AI 法官在核心审判事务上的参与程度取决于 AI 发展所提供的技术可能性。[②]技术乐观主义者坚信，随着相关技术的成熟，将大部分甚至全部人类法官的审判职能移交给狭义的 AI 法官终究是切实可行的。[③]目前法律职

① 傅郁林：《全责的法官与迷失的法官助理》，《人民司法》2019 年第 22 期。

② Carl Benedikt Frey, Michael A. Osborne, “The Future of Employment: How Susceptible are Jobs to Computerisation?” *Technological Forecasting and Social Change*, vol.114, 2017, pp.254-280.

③ Eugene Volokh, “Chief Justice Robots,” *Duke Law Journal*, vol.68, 2019.

业群体对司法 AI 的保守态度实质上是法学教育塑造的思维定式与技术发展之间的张力，认知能力缺陷、推理能力不足、价值判断缺位、特殊情境失能等，[①] 实际上都是技术成熟度在司法场景中的客观投射，并不能在技术维度上否定 AI 法官全面替代人类法官在理论上的可能性。

然而，当对审判“职能”的关注点由“事项”转移到“权限”时，算力的提升、模型的优化、AI 的进化，能否将人类法官的权限逐步转移给 AI 法官，直到 AI 法官独立行使审判权，从而使人类法官最终被 AI 法官替代，正如原本作为司法辅助人员的治安法官逐步被收编为正式法官，不仅独立地承担原来由法官承担的某些职能，而且也享有相应的独立权限？笔者的答案是，基于本文第一部分所述的司法伦理，AI 法官由于缺乏以人性为基础的正当化机制，因此赋予其独立于人类审判人员的裁判权，即使在技术上是可能的，在法律上也不能突破受审判权由人类法官独享、由 AI 法官承担的审判职能受人类法官监督、审判行为与结果责任由人类法官承担这一制度边界。换言之，AI 法官可以承担审判职能，但不能享有审判权限；AI 法官承担核心审判职能的可能性或限度受 AI 技术的阈值制约，AI 法官被赋予的审判权限的可能性或边界受人类社会的法意识和正当性基础制约。

制度上，基于审判职能的划分，在人类法官与人类审判辅助人员之间进行审判权配置时，法官与司法辅助人员应当在各自的法定职责范围内“独立”行使权限、履行责任。审判辅助人员即使履行的是审判辅助职能，原则上也应被法律赋予独立的审判人格、享有独立的审判权限；如果法律规定某些审判辅助职能由审判辅助人员在法官的指导、监督下行使，但未对各自的权限边界、行为规范和主体责任进行明确规定，则该职能（事项）的权限与责任的最终归属应为法官。但 AI 法官的底层逻辑与之不同。基于审判职能的划分，在人类法官与 AI 法官、人类司法辅助人员与 AI 司法辅助人员之间进行审判权配置时，人类法官与 AI 法官共同承担核心审判职能，即裁判权，但审判权力与职责只能由人类法官承担；人类司法辅助人员与 AI 司法辅助人员共同承担非核心审判职能，其中纯技术性、程序性事项可由 AI 独立承担；但那些虽被划定为司法辅助职能的事项，若需基于人类的心智进行判断、须借助人性资源而获得正当化的部分职能或事项，即使 AI 的职能行为也应受到人类司法辅助人员甚或直接受人类法官的指导和监督，其权限和责任归属于人类审判人员。

具体而言，作为司法辅助人员的 AI，其任务配给机制一部分是系统性的 AI 化，如由法院统一建构的随机分案系统、庭审排期系统、档案管理与查阅系统、法律检索与类案参考系统等；另一部分则是个案化的，如诉讼材料的送达、程序之间的转换、诉讼费用的分担等。但作为法官的 AI，其任务配给机制只能是个案化的，如证据的调查、整理、比对，案情事实的分析与归纳，本案的法律选择与解释的多选项提示与分析，指导案例及其他相关案例的检索与比对，以及根据人类法官的意见或裁判要点制作裁判文书，等等。

总之，AI 法官是否具备审判主体性特征，首先是技术可能性的问题，目前 AI 的发展距离担当起核心审判职能所需的法律专业化水平还相去甚远，但理论上仍具有技术可行性；但在制度上赋予 AI 法官以审判主体权限，除技术可能性的考量外，还有关于技术理性担当司法职能的伦理基础及政治哲学基础等多方面的考量。而基于司法本身的社会属性与评价体系的人性依赖性，除非硅基生命完全替代碳基生命，否则给 AI 法官赋予人类法官的审判权，将因为其缺乏人类法官的人格而导致正当性缺失，故此种做法在法律意义上是不可欲的。

① 代表性文献参见钱大军：《司法人工智能的中国进程：功能替代与结构强化》，《法学评论》2018 年第 5 期；左卫民：《中国计算法学的未来：审思与前瞻》，《清华法学》2022 年第 3 期。

数字矫正正义的实现维度

王福华，上海交通大学凯原法学院教授

数字技术，尤其是人工智能技术，在为人类提供便利和智能化服务的同时，也带来虚假信息传播、数据安全损害和版权侵权等风险，进而造成物质性和非物质性损害等“数字不正义”的结果。为实现可持续的数字化转型，必须将数字

权利保护置于核心位置，不仅要鼓励技术创新，还要警惕数字技术的潜在风险，实现数字矫正正义。从哲学视角来看，亚里士多德的正义理论仍可为破解数字社会正义难题提供分析框架。他将正义区分为分配正义与矫正正义，其中，矫正正义的核心在于通过纠偏机制恢复被破坏的平等关系。[①]矫正正义通过纠正数字不公、修复数字侵害、制衡技术权力、实现程序正义和促进数字包容多种途径，被视为继“数字分配正义”之后的“第二代数字正义”。[②]

数字矫正正义的制度维度

（一）数字矫正正义的制度功能

数字技术催生了新型侵害形态，对其识别与分类是建构数字矫正正义理论的核心问题。从技术手段出发，数字侵害可以分为以下几类：一是数据滥用型侵害；二是技术操纵型侵害；三是技术漏洞型侵害。根据严重程度，上述数字侵害可以区分为数字妨害和数字损害。数字妨害是指基于风险或过程的抽象性侵害，尽管在个案中可能较为轻微，但由于算法等数字技术的广泛应用，导致其具有集合性、累积性和系统性的特征，可能带来广泛而深远的负面影响。数字损害则是基于结果的具体侵害，对传统部门法所保护的权益造成了明确的损害后果。[③]解决上述“数字不正义”问题，需要以修复价值为起点，以社会关系重建为进阶，最终实现数字生态系统的协同优化。

首先，修复目的。在数字时代，风险治理需要构建预防与矫正并重的双轨机制。面对技术复杂性带来的现实挑战，传统的预防机制无法完全实现数字正义的目标。当预防机制失效时，数字矫正正义机制应发挥关键作用。其以恢复因错误行为而被破坏的公正为基本理念，追求实现“最大程度接近损害未发生状态”的理想目标，包含四个方面的修复价值：一是为遭受数字侵害的受害者提供实质性救济，重塑受害者与加害者的关系；二是修复特定行为引起的数字权力失衡状态，扭转数字弱势群体的不利地位；三是修复因数字技术的不当使用而导致受损的社会信任关系及人际或人机协同关系；四是修复通过数字技术构建的社会运行规则，以及技术伦理秩序，包括法律提供的规范性秩序与社会自发形成的秩序。

其次，溯源目的。通过追踪数据流、算法决策链或技术滥用路径，尽可能精准地定位产生“数字不正义”的决策环节，揭示产生“数字不正义”的根源，使得程序正义理念嵌入技术架构。这不仅是实现问责与救济，以及透明治理的实操工具，本身也应作为数字矫正正义的基础价值。同时，对溯源价值的认识既要防止绝对化，承认技术局限下溯源的不完美性以及特殊领域的特异性；又要避免完全否定溯源的价值而引起责任崩塌，故宜采取以“最小必要性”为原则的弹性溯源框架。

最后，参与目的。作为数字社会的直接参与者和

① 亚里士多德：《尼各马科伦理学》第5卷，苗力田译，北京：中国社会科学出版社，1999年，第88—102页。

② 关于“数字分配正义”的阐释，可参见马长山主编：《数字法理学》，北京：法律出版社，2025年，第298页。

③ 王莹：《算法侵害责任框架刍议》，《中国法学》2022年第3期。

风险的潜在受害者，公众在实现数字矫正正义方面的作用不可或缺，具体表现为有权就数字侵害提出异议、获得相应的救济等。然而，在现实中，数字技术的专业知识和决策权通常集中在科技巨头和权力机关手中。由于数字技术的复杂性，加之公众数字素养有限、经济和语言障碍以及“参与剧场”等问题，数字鸿沟可能进一步加剧，导致弱势和边缘群体的声音难以被听见，其关切也难以得到有效解决。强化公众的数字参与权，不仅能够改变数字权力失衡的结构，而且本身也是矫正正义的体现。

（二）数字矫正正义的制度框架

数据驱动侵害的救济，一方面需以技术机理为起点，以损害性质为分层依据，通过预防性制度（算法备案与审计、行业合规整改等）与补偿性机制（民事赔偿与刑事追责）联动，完成权益恢复与系统治理的双重目标。另一方面，数字矫正正义必须以人权保护为核心价值，通过技术性正当程序保障与司法行政协同治理，实现对数字驱动型损害的全面矫正。

首先，以人权保护为核心的数字矫正正义制度框架。数字空间是物理空间的延伸，数字正义也是传统正义理念的延伸，故必须体现对人类主体地位和选择权的尊重，人工智能监督、隐私和数据治理、透明度及问责机制均应基于尊重人权的原则建立。在表现形式上，数字驱动型损害具有隐蔽性、复杂性和影响滞后性的特点，其危害程度常被低估或忽视，导致个人权益受到慢性侵蚀。这不仅使人们在享受数字红利的同时不知不觉让渡核心权利，还可能引发数字化生存的危机，从根本上侵犯人权。因此，数字权利的保护或救济不应仅限于传统民事权益范畴，而应拓展至人权领域。

其次，以正当程序为保障的数字矫正正义制度框架。技术性正当程序与传统纠纷解决机制中的正当程序有所区别，其要求在数字化过程中保障基本程序性权利不受忽视或侵犯。政府和企业必须在各种数字技术应用中建立可及且可执行的程序性保障机制，将算法效率与程序公正的核心要素有机结合。具体来说，在部署人工智能时，应构建申诉技术架构，建立可及且可执行的程序性保障机制，以及透明的决策流程，并持续评估算法偏见的风险。无论是在刑事还是民事程序中，基本保障措施都应通过数字技术手段得以实现，应注重数字纠纷解决的实际效果，而不仅仅关注纠纷处理的数量。

最后，司法与行政协同的数字矫正正义制度框架。针对数据驱动造成的各种“数字不正义”，行政机关可以通过扩大有害数字内容的认定范围、强化平台责任、明确内容审核要求等监管措施进行事前预防与监管。与此同时，鉴于数据驱动的损害救济需求，司法机关也应发挥诉讼机制的作用：通过民事诉讼，司法机关可以为受害者提供有效救济，判决侵权者赔偿其精神损失和经济损失（如因数据泄露所致）。对于严重的侵权行为（如网络诽谤、算法操纵竞争），司法机关还应追究相关主体刑事责任。相较于行政机关通过规则强制力实现数字侵权的前端防控，司法机关通过司法权威性完成的后端矫正更为重要，二者共同构成数字时代监管与救济的治理闭环。

数字矫正正义的归责维度

在传统法律框架下，矫正正义的实现依赖向责任人施加相应责任。然而，数字时代的“多头侵害”现象对传统归责制度提出了严峻挑战。首先，行为主体的匿名性增加了身份识别的难度；其次，侵害行为的多方参与性导致责任主体分散；最后，责任链条的稀释和断裂使得因果关系的认定变得异常复杂。这些因素共同导致数字侵害难以被准确归责，面对上述难题，应从以下方面着手。

（一）扩展责任主体

传统的责任主体通常指向服务或技术的提供者与使用者，但大语言模型等技术架构催生了新型的责任主体，数据收集方、

数据标注方、数据交易平台、算力供应商、基础模型开发者、模型微调者、API 集成商、终端部署方等数据层、算法层与应用层的各个参与者均可能造成侵害。由此使得责任主体在三个维度上扩展：一是纵向深化，从终端应用层向基础研究层追溯；二是横向拓展，覆盖算力等新型基础设施；三是时间延伸，即技术全生命周期责任覆盖。对于数字侵害民事责任的分配，应根据各方对技术系统的实际掌控程度、风险收益平衡与侵害贡献度来确定：对于技术开发运营使用高度一体化的情形适用连带责任，对于模块化可分的情形适用按份责任，对于明显存在责任层级的情形则适用补充责任。

（二）确立差异化归责原则

首先，传统侵权法的过错责任依然适用于数字侵害，如果能够证明技术的设计者、应用者或部署者因故意或过失导致侵害，则应适用过错责任。比如，“白箱”人工智能的决策逻辑和决策过程具有可解释性和高度透明性，造成侵权时可采取产品责任致害的归责路径。对于过错的认定，应采取主客观相结合的标准，以比例原则为指引，从合规性、谨慎性、合理性等角度标准综合推断上述主体的主观心理状态，判断其是否尽到合理注意义务。

其次，过错推定原则可以矫正技术权力的不对称，使得技术控制者处于最佳风险防控位置。比如，《个人信息保护法》第 69 条第 1 款规定：“处理个人信息侵害个人信息权益造成损害，个人信息处理者不能证明自己没有过错的，应当承担损害赔偿等侵权责任。”与此同时，也应设置可反驳推定的抗辩事由，如证明已采用行业领先的防护措施，损害完全由第三方恶意行为导致，或是受害人故意诱发侵害等。

最后，需从技术风险性、损害可逆性、主体控制性三方面衡量是否设立无过错责任。比如，高风险的自主人工智能系统的潜在危害性高，且受害者通常难以举证。若造成损害，从域外立法实践来看，欧盟认为责任应归属于技术控制方。2024 年，欧盟《人工智能法案》对高风险人工智能系统采取了严格责任原则，旨在确保技术开发者、部署者和使用者对人工智能系统造成的损害承担无过错责任，这代表了全球最严的人工智能监管方向。

（三）缓解因果关系认定困境

数字侵害在因果关系确立方面的困难主要体现为：数字生态的复杂性导致多因一果的现象普遍存在，多方行为叠加形成错综复杂的因果关系网络；受害人难以举证损害发生的完整因果链条，比如深度学习模型等人工智能的决策过程具有不可解释性，以及受害人因无法获取关键算法日志而面临举证困难；时空分离特征导致因果关系关联程度模糊，数字侵害通常存在显著的时间延迟效应，从技术缺陷形成到损害实际显现可能间隔数月甚至数年，增加了因果关系的判断难度。有鉴于此，应构建阶梯式因果关系认定标准。对于可溯源的侵害，坚持必然因果关系标准，借助区块链存证等技术手段实现精准归责；在算法自主决策、侵害原因无法被科学验证或解释时，采用条件因果关系理论，若满足行为是损害发生的必要条件，并显著提升损害发生概率，即可认定因果关系成立。

此外，应积极探索举证责任减轻机制，如针对不同的技术复杂度、是否有理解与控制技术可能性设置阶梯性的证明标准体系；通过事案解明义务促使不负举证责任的对方当事人提供协力；以及采取因果关系推定等。在完善归责制度的基础上，也要配套建设技术基础设施，如开发算法可解释性工具、区块链存证体系、沙盘推演系统等支持机制，提升归责制度的精准性、效率性与公平性。

数字矫正正义的救济维度

各国在数字矫正正义中追求矫正性、恢复性和惩罚性目标，但同样面临着损害识别能力薄弱、问责机制缺失及救济渠道匮乏等挑战。

（一）数字侵害救济的层次性

对数字侵害的救济应从程序和实质层面予以考虑。数字侵害受害者的救济措施需根据侵害类型、程度、受害者需求及社会治理要求而定，依功能可分为个案矫正实现即时正义与提炼规则预防未来风险。

数字侵害的层次救济体系能够应对数字侵害复杂性、多样性与扩散性特点，兼顾个体即时救济与系统性治理需求，实现预防、发生时及发生后的“全周期”覆盖，确保救济效果的全面性。为此，应有多种多样的救济措施可供选择。（1）恢复性措施。直接消除侵害后果，如删除错误个人信息、撤销歧视性自动化决策。（2）赔偿性措施。对无法恢复原状的损害进行经济赔偿，涵盖物质损失、精神损害、机会成本及道德损害。（3）康复性措施。提供医疗、心理干预、社会支持与专门援助，帮助受害者康复或扭转数字弱势群体的不利地位。（4）精神性与象征性措施。非经济手段补救精神道德层面，如正式道歉恢复名誉、制裁责任人、设立受害者基金等。（5）停止侵害与预防措施。包括发布禁令阻止侵害继续，以及要求侵害方改变政策、流程或战略的结构性救济。（6）系统性变革措施。针对广泛性、结构性侵害，通过完善立法、制定政策、加强行业监管推动整体变革。

（二）数字侵害救济的系统性

为保障救济的可及性与实效性，“救济生态系统”理论倡导整体性、系统化设计，强调关注法律、政策、机构、行为主体等关联要素构成的系统如何协作提供救济，厘清各主体作用。

首先，是司法程序的基础作用。作为公众权利的最后一道防线，司法在数字侵害救济生态中仍具基础性地位，并构成其他机制（尤其非国家救济）的有力后盾。除传统私益诉讼外，应重点发展适应技术特性和治理需求的公益诉讼。公益诉讼能破解个体维权困境，应对规模化侵害，是重塑“技术—权力—权利”关系的基础制度。未来应通过扩展主体资格、改造诉讼规则、创新责任形式、改革执行监督、加强能力建设等，推动其向“事前预防、制度变革、治理引擎”转型。

其次，是行政监管的核心响应。公安机关、网信部门等数据安全监管机构的公共执法对数字矫正正义目标的实现至关重要。行政机关兼具主动性、高效性、系统性、灵活性与专业性，不仅可运用从约谈、合规审计等柔性手段到责令整改、处罚等刚性措施，还能细化规则标准、发布指南。作为国家层面的第一响应人，其能有效弥补司法救济的滞后性。

最后，是非国家机关的韧性补充。非国家机关可填补公权力救济的不足，增强系统韧性。企业和社会责任的强化至关重要：平台、技术服务提供者或使用者可能造成侵害，应践行尊重数字人权责任，建立便捷、高效、低成本的内部申诉响应机制。这不仅是最高效的救济方式（减轻公权力负担），更是预警机制，为改进技术、模式、政策提供方向，预防未来侵害。

（三）数字侵害救济的有效性

对数字侵害的救济必须满足有效性，其衡量包含三个核心维度：一是及时性，数字信息传播高效且扩散迅速，救济价值随时间骤减，救济不应被拖延；二是比例性，即救济措施需与侵害的严重程度及受害者实际损害成比例；三是充分修复性，即救济应以能充分有效修复侵害的方式提供。

非官方救济机制尤须注重有效性，以发挥其在解决数字侵害纠纷中的优势。联合国全球公共政策项目为此提供了全面的评估框架。（1）合法性：程序公平，赢得信任。（2）可得性：利益相关方知晓机制，并为有障碍者提供充分帮助。（3）可预测性：程序清晰、阶段时间明确、结果与监督方式确定。（4）公平性：受害方能合理获取所需信息、建议和专业知识，在公平、知情、受尊重条件下参与。（5）透明性：各方及时了解进展。（6）兼容性：结果符合国际公认标准。（7）持续改进性：借鉴经验改进机制，预防未来侵害。（8）参与和对话：咨询利益相关者，使受影响者有机会参与程序设计，通过有意义的协商形成共识。

人工智能司法应用的限度及其制度规制

唐力，西南政法大学法学院教授

人工智能的应用极大改变了司法过程，甚至对司法的一些价值、原则形成了挑战。本文从人工智能司法应用的基本原理出发，探讨人工智能司法应用的技术限度和伦理限度，并从制度构建的意义上讨论人工智能司法应用的规制。

人工智能司法应用的技术限度与伦理限度

对人工智能司法应用限度的讨论，首先需要厘清其功能定位。从司法实践的探索来看，人工智能的司法应用最初是将诉讼的相关环节技术化数据处理以后，辅助法院电子化程序的运行，这一阶段是现代数字技术在司法领域应用的初级阶段，并未实质改变司法程序物理化运行场景中的基本原则、基本制度，遵循传统的司法运行规律，法官是案件事实认定和法律适用的主导者。随着“数字司法”“智慧法院”建设进程的推进，数字技术已从“辅助司法”的角色逐渐向“参与司法”的方向发展。人工智能在文书内容的生成、类案检索等方面表现出了超凡的能力，突破了案件事实认定、法律适用的传统思维模式，从人类大脑的“生物思维”逐渐向机器大脑的“数字思维”转变，人工智能正在逐步向司法的核心领域“渗透”，甚至大有替代人类法官的趋势。[①] 数字技术司法应用从“辅助”走向“主导”的发展趋势，引发了人们对人工智能司法应用前景的担忧。因此，数字技术的司法应用要确定必要的限度，以维护实质司法正义的实现。

（一）数字技术应用与司法规律

传统司法过程是以法庭的物理场域为载体，遵循一定的司法原则、制度和规则构建起来的一套精致的程序规则，以法官居中、两造对审的“诉（控）”“辩”“审”构建的程序结构，诉讼以“主张”→“辩论”与“证明”→“判断”的程序运行方式展开，法官在案件审理中根据证据认定事实并适用法律做出判决，每一个案件的审判都是法官寻找个案法律正义的司法过程。在这一过程中，案件事实由当事人主张并加以证明，法官则综合案件的全部信息做出判决。

传统司法过程既是一个逐渐探明案件事实的过程，也是法官针对个案解释法律，实现司法正义的过程。然而，传统案件的审理受裁判者个体的职业素养、生活阅历、价值观等影响，案件裁判的公正性受个体因素影响较大并难以克服，容易导致民众对司法公正信任度不高的消极后果。人工智能被广泛应用于司法实践，得益于法院系统推行的法院数字化及智慧法院工程。人工智能应用于司法的过程，是技术与经验结合的过程，人工智能通过海量案件数据，以要素化、模块化的技术方法形成“类案识别”应用于待决案件，并且其判断决策是基于预先设计的算法来实施的。理论上，人工智能的司法应用能有效克服人类法官的上述不足。

人工智能广泛应用于司法领域，重塑了司法过程。传统的司法模式注重事实发现和法律适用的过程性，而数字司法则以符号化、模块化并通过“生成式人工智能强大的数据分析能力和反应速度，实现审理过程、办案程序、决策输出方面的指引和监督”。[②]

① 有学者以数字技术发展的前瞻性思维，认为“AI法官”将会替代人类法官。参见尤金·沃洛克：《AI首席大法官》，载周少华主编：《数字法学》2024年第3辑，北京：社会科学文献出版社，第279页。

② 雷磊：《数字司法的理论反思：意义、问题与监管》，《交大法学》2024年第6期。

作为司法过程性载体的程序，是最为公开、透明、公正的，其要求任何决定都必须经历对抗性辩论和证明的洗礼。[①]特别是为保障当事人诉权以及裁判公正性的审级设计，能有效防范事实认定和法律适用方面可能产生的错误。然而，在通过算法自动生成案件判决结果的数字司法中，程序的价值和意义将被弱化甚至忽略。

（二）数字技术的应用与司法权的人类专属性

随着人工智能司法应用的实践发展，司法主体与技术的关系从技术辅助逐渐走向技术主导，并大有滑向技术依赖的趋势。[②]这种技术主导或技术依赖所带来的风险是司法权的不当让渡，算法“法官”取代人类法官、技术公司取代司法机关，获得司法裁判权。这种风险或者担忧并非空穴来风，从人工智能技术的发展和人民法院信息化建设目标来看，如果不进行必要的技术限制的话，大概率会成为现实。

人工智能司法应用的过度扩展，必然会带来算法主导司法和人类法官依赖技术的消极后果。司法场景中人工智能技术的全面应用，重塑了司法过程，“在智慧司法中本为辅助司法流程开发的技术应用，最终却导向了技术主导下的司法制度变革。因此，更合适的问题不是‘技术应用是否会重塑司法制度’，而是‘何时’以及‘在何种程度上’技术会重塑司法制度”。[③]技术重塑司法过程，改变了司法的本质，国家司法权面临着让渡于掌握算法技术的技术公司的风险，可能会形成司法的“技术垄断”的严重后果。我们可以把司法中的这种“技术垄断”称之为“技术权力”,那么，我们应当思考的问题是：“技术赋予谁更大的权力、更多的自由？谁的力量和自由又会被削弱？”[④]

（三）数字技术的应用与人权保障

现代司法注重人权保障，注重司法程序中主体作为人的基本权利保障。人工智能的司法应用涉及算法伦理问题，其包括技术应用中的道德边界、人类主体性以及社会公平性等核心问题。人工智能的司法应用是基于海量司法数据的训练以及算法设计，此即可能存在数据采选的局限性和对算法设计的主观性而形成算法歧视与偏见。在大规模数据采集和训练过程中，可能存在数据的过度采集而导致数据滥用和侵犯个人信息的情况。基于人类自身天然的偏好和兴趣，算法可能会通过“信息茧房”“行为预测”等手段强化对人的控制，削弱个体自主决策能力。

算法技术应用所带来的伦理风险，在司法领域可能会导致人的基本尊严受到损害，以及产生不公正的裁判结果等风险。算法技术的司法应用，应当坚持以人为本的原则，即算法设计应当以人类尊严和人的主体性为底线，避免因数据采集偏差带来不公平的结果；克服算法设计的主观性可能产生的算法歧视；司法参与主体应当对司法决策具有可参与空间，即通过算法决策的可追溯性和可解释性来避免“算法黑箱”或者“算法霸权”现象。2021 年 9 月 25 日，国家新一代人工智能治理专业委员会发布《新一代人工智能伦理规范》(以下简称《人工智能伦理规范》), 第 1 条明确强调将伦理道德融入人工智能全生命周期，促进公平、公正、和谐、安全，避免偏见、歧视、隐私和信息泄露。《最高人民法院关于规范和加强人工智能司法应用的意见》(以下简称《人工智能司法应用意见》) 第三部分也明确规定了人工智能司法应用的“安全合法”“公平公正”“透明可信”“公序良俗”等基本原则。这些针对人工智能的伦理宗旨、规范，是人工智能司法应用的道德底线。

（四）数字技术的应用与司法正义

传统司法模式对实现司法的个案正义具有积极意义，法官能够根据个案的具体情况发挥其主观能动性，实现司法判决的法律效果、政治效果和社会效果的有机统一。算法技术主导下的司法模式，是以标准化的代码和数据类型化、案件要素化为基本框架，根据技术公司预先设计好的“方法”作出决策。算法技术主导下的司法模式克服了传统司法模式案件裁判受法官个

① 季卫东：《人工智能时代司法权之变》,《东方法学》2018 年第 1 期。
②③ 张凌寒：《智慧司法中技术依赖的隐忧及应对》,《法制与社会发展》2022 年第 4 期。
④ 尼尔·波斯曼：《技术垄断——文化向技术投降》, 何道宽译，北京：中信出版集团，2019 年，第 10 页。

人因素的影响，“通过数据标注识别、案件要素抽取、知识图谱构建来进行算法建模。相较人类法官而言，算法的使用被认为与公正呈正相关性，因为它减少了人类决策人员的偏见和主观性”。[①]

但是，我们也应当清楚地认识到，算法能够减少人类决策人员的偏见和主观性，是指算法技术在司法应用过程中只能根据设计好的运行程序、设计参数和指令行事，其可抑制偏见和主观性。然而，算法的偏见或者主观性可能早在设计时就已经被“植入”。在此情况下，算法所带来的只能是所谓的司法“技术正义”而非司法的法律正义。而且，算法在形成个案决策过程中，法官、当事人无法参与其中，只能被动地接受算法“算出”的结果，有违程序正义的要求。同时，算法也会忽略案件的个体差异，无法替代人类作出价值判断，会用所谓的“算法正义”来替代个案的法律正义。

人工智能司法应用的制度规制

人工智能的司法应用应当保持必要的限度。最高人民法院发布的《人工智能司法应用意见》，对人工智能司法应用的指导思想、总体目标、应当遵循的原则、应用范围、系统建设以及综合保障等方面作了较为详细的规定。在制度构建方面，应从人工智能司法应用的“人机关系”、应用领域、司法伦理，以及安全保障等方面进行规范。

（一）人工智能司法应用的伦理准则

《人工智能伦理规范》第 3 条规定了增进人类福祉、促进公平公正、保护隐私安全、确保可控可信、强化责任担当、提升伦理素养等 6 项基本伦理要求。[②]具体而言，在人工智能司法应用中，应当从制度建设层面把握好以下几个问题。

一是以人为本，维护司法程序中人的尊严与主体性。司法人工智能不能替代司法参与者的主体地位，更不能将其“客体化”，要保障司法参与主体拥有充分的自主决策权。二是促进司法公平正义。首先是程序平等，任何参与司法程序的主体都应当得到平等对待，特别是要警惕数字技术的应用可能形成的“数字鸿沟”对当事人程序平等权利的侵蚀，确保“人工智能产品和服务无歧视、无偏见，不因技术介入、数据或模型偏差影响审判过程和结果的公正”。[③]三是增强算法透明性和可解释性以提升司法人工智能的可信度。算法应当具备透明性，算法的决策逻辑、数据来源以及信息处理过程，应当可被追溯和验证；算法应当具有可解释性，能够以人类可以理解的方式说明决策过程和依据。四是保护隐私安全。制度层面应当明确人工智能司法数据处理的合法性边界，充分尊重个人信息知情、同意等权利，数据采集、存储、使用应当遵循“合法”“正当”“必要”和“诚实信用”原则，保障个人隐私与数据安全。五是构建可追溯的问责机制。人工智能司法应用涉及多主体的责任承担的分配机制，其包括技术的开发者、服务提供者与用户（使用者）各自应当承担的法律责任。基于人工智能应用的辅助性定位，应当坚持以“裁判职权始终由审判组织行使，司法责任最终由裁判者承担”的基本原则，构建人工智能技术设计开发、服务提供和使用的归责体系。

（二）人工智能司法应用的“人机关系”

人工智能司法应用的一个关键问题是，人（法官）与机器（算法）的关系。从上述对人工智能司法应用限度的分析可以发现，人工智能自身也存在“数据茧房”“算法黑箱”“算法歧视”等问题，特别是过度依赖技术可能产生司法权的不当让渡、削弱人类法官自主决策权等严重违背司法规律的问题。对此，应从制度层面严格定义“人机关系”，明确人工智能司法应用的定位。《人工智能司法应用意见》明确“无论技术发展到何种水平，人工智能都不得代替法官裁判，人工智能辅助结果仅可作为审判工作或审判监督管理的参考”，这有利于充分保障审判人员的自主决策权。

从制度构建层面出发，人工智能司法应用的“人机关系”应当把握好以下三个问题。一是用户自主决

① 马克·舒伦伯格、里克·彼得斯：《算法社会——技术、权力和知识》，王延川、栗鹏飞译，北京：商务印书馆，2023 年，译者序。

②《新一代人工智能伦理规范》第 3 条规定。

③《最高人民法院关于规范和加强人工智能司法应用的意见》（法发〔2022〕33 号）第 4 条规定。

策权与人工智能的辅助性地位。无论人工智能技术发展到何种水平，其只能起到司法辅助性作用而非替代性作用。二是裁判权的专属性与权责统一。人工智能司法应用必须坚持裁判权专属于审判机关的原则，并对人工智能可介入的司法领域进行制度限制。与此同时，依据“权责相当”原则，应当从制度上设计人工智能司法应用中的归责机制，并坚持“让审理者裁判，由裁判者负责”的司法责任制核心原则。三是各类用户的选择权。人工智能司法应用作为司法审判辅助工具，意味着其不是强制性应用，在“人机关系”中作为用户的程序参与者拥有选择权。裁判者有权决定是否利用司法人工智能提供的辅助，当其发现人工智能的应用存在问题或者虽未发现问题但认为其有碍自主判断时，有权随时退出与人工智能产品和服务的交互。从制度设计上，应当确保用户的这种选择权的正当行使，建立良好的“人机关系”。

（三）人工智能司法应用的范围

与欧美国家人工智能司法应用较为谨慎的态度不同，[①]我国对人工智能司法应用表现出较为积极的政策导向。我国《新一代人工智能发展规划》（国发〔2017〕35号）明确提出“建设集审判、人员、数据应用、司法公开和动态监控于一体的智慧法庭数据平台，促进人工智能在证据收集、案例分析、法律文件阅读与分析中的应用，实现法院审判体系和审判能力智能化”。司法实务在上述政策引导下，积极推动“智慧司法”工程建设，《人工智能司法应用意见》采取“开放”式的司法应用场景，人工智能应用覆盖司法全流程、全领域，包括诉讼服务、案件管理、司法决策等领域。基于我国上述政策导向，人工智能司法的风险性可能被掩盖，须引起学界的高度重视。

基于人工智能司法应用可能带来的各种系统性风险，对其规制最有效的方法即是从制度上明确其司法应用的准入领域以及禁止介入的领域。数据计算最为擅长的是重复类行为，因此人工智能司法应用“原则上，司法领域中的重复性、可替代性工作均可由人工智能承担”。[②]这是从“准入”方面对人工智能司法应用场域所作的规范。此外，规制人工智能司法应用范围还应建立“禁入”的应用范围，对于可能影响司法公平与正义、可能因算法技术介入影响裁判者独立判断、自主决策以及关涉当事人重大程序权利与实体权利等领域，应当限制人工智能应用的介入。

（四）人工智能司法应用的技术安全保障

人工智能司法应用的安全保障应以“安全可控、公开透明”为要旨，加强技术安全建设。数据和算法构成了人工智能的两大支撑：一是严格把控数据质量，建立数据审查机制，实行数据的分类分级管理，采取差异化的数据收集、存储、传输策略。应当在制度构建层面，加强人工智能数据采集、存储、管理和应用的法律法规等制度规制，确保数据质量和安全。二是构建算法评估机制。《人工智能司法应用意见》明确提出技术、服务、运行的透明性原则，应强化对算法设计的制度约束，建立事前、事中、事后全过程的算法评估机制，建立人工智能开发、运行、应急全生命周期的监管制度体系；三是健全伦理审查机制。人工智能安全保障，需要从人工智能伦理要求方面健全伦理审查机制。通过设立司法人工智能伦理审查委员会，对数据采集、存储、管理和使用，以及算法设计进行伦理审查，防范伦理道德风险。

人工智能司法应用的健康发展，技术安全是基础，制度建设是关键，伦理规范是引导，需要通过技术加固、制度规制、伦理引导、权责明晰等多维度的共同作用，实现对人工智能司法的有效规制。尊重司法规律，明确人工智能司法应用的辅助性定位、确保人类法官的自主决策权，建立良好、可信任、和谐的“人机关系”。

① 李训虎：《刑事司法人工智能的包容性规制》，《中国社会科学》2021年第2期。

② 雷磊：《数字司法的理论反思：意义、问题与监管》，《交大法学》2024年第6期。

智能化赋能金融案件“四大检察”：挑战与对策

许多奇，复旦大学法学院教授、智慧法治实验室主任

2024年7月，《中共中央关于进一步全面深化改革 推进中国式现代化的决定》前瞻性地提出，建立风险早期纠正硬约束制度，筑牢有效防控系统性风险的金融稳定保障体系。在立法推进过程中，《金融稳定法（草案）》明确提出需强化

金融风险的源头管控，秉持市场化、法治化原则协同高效地处置风险。[①]随着“法律监督”内涵融入新的理念，“四大检察”格局得以构建。[②]在这样的背景下，智能化赋能金融案件“四大检察”，成为提升金融检察监督效能、保障金融安全的重要手段。2022年6月，全国检察机关数字检察工作会议作出部署，要求激活虽有流动但总体处于休眠状态的各类数据，通过关联分析与深度挖掘，为强化法律监督、深化能动履职提供前所未有的线索与依据。[③]在金融领域，上述顶层设计均指向同一核心目标，即构建反应敏捷、运转高效、智能化的现代化金融案件“四大检察”全新格局。

金融案件检察监督的必然性

（一）金融案件的场景描述

近年来，资本市场中存在的虚假陈述现象、互联网金融领域出现的风险无序扩张问题、资管行业长期存在的刚性兑付弊病、房地产企业呈现的“脱实向虚”倾向，以及城投公司背后潜藏的隐性债务等情况，均对我国金融稳定构成了严峻挑战。以下结合具体场景展开分析：

场景一：部分金融控股集团被实际控制人违规利用。这些实际控制人借助复杂的股权代持及关联交易隐匿真实身份，致使金融机构的内部控制与风险防控体系失效，为系统性金融风险埋下重大隐患。

场景二：金融机构从业人员与外部金融掮客相互串通，骗取巨额资金。近年来，在贷款、票据、担保、债券、信托等领域，相继出现涉案金额高达数亿元甚至上百亿元的重大犯罪案件，给国家和金融机构造成了巨大损失。

场景三：金融犯罪手段持续更新，呈现出高度智能化与跨境化的特征。例如，犯罪团伙借助虚拟货币开展“兑换操作”，把境内的非法所得转换为虚拟货币后，经由境外平台出售以获取外汇，达成资产的跨境转移与非法洗白。此类新型犯罪模式显著加大了追诉与追赃的难度。

场景四：近年以P2P为代表的网贷平台集中“爆雷”，严重冲击金融和社会秩序。这些平台常常以“互联网金融创新”作为幌子，违法违规设立资金池、编造虚假项目，并以高额回报为诱饵实施欺诈行为，其本质属于“借新还旧”的“庞氏骗局”。当风险暴露时，巨大的资金缺口与庞大的受害群体，使后续的处置工作面临极大的困难。

上述金融犯罪案例，不仅深刻揭示了当前金融领域违法活动所呈现出的跨区域、涉及面广、隐蔽性强等特征，也充分展现了案件背后错综复杂的利益链条和盘根错节的关联关系。从案件侦办过程中可以看到，这些金融犯罪往往横跨多个省市，涉案金额巨大、受害群体广泛，且犯罪分子通常采用高科技手段和精心设计的作案手法来掩盖其违法行为，使得案件侦破难

①《中华人民共和国金融稳定法（草案）》的发布与征求意见是一个持续的过程，参见《草案》（二次审议稿）第4条、第6条、第18条。

② 王海军：《“法律监督”概念内涵的中国流变》，《法学家》2022年第1期。

③ 邱春艳：《深入贯彻习近平法治思想以“数字革命”驱动新时代检察工作高质量发展》，《检察日报》2022年6月30日，第1版。

度大大增加。另一方面，此现象亦促使我们对现行社会治理体系及风险防控机制展开反思：为何在风险萌芽阶段未能及时发出预警？为何在问题累积进程中未能实施有效干预？这表明当前治理体系在风险识别、预警与处置环节仍存在亟待改进的制度性短板。

我国《宪法》明确赋予检察机关法律监督的重要职责。这一定位不仅为刑事、民事、行政和公益诉讼"四大检察"职能奠定了坚实的宪法根基，更从国家治理体系现代化的层面，确立了检察机关在维护法律统一正确施行、防范化解重大风险中的制度性作用。人民检察院的"四大检察"职能并非各自孤立，而是以法律为基础，围绕"法律监督"这一核心职能铺展开来。从理论层面，应当构建起相互衔接、互补协同的检察工作体系，共同服务于法治国家建设。然而，为何在金融案件中该体系无法有效发挥作用？在实施过程中又遇到了哪些阻碍？金融监管与金融检察之间为何难以实现协同？需要深入分析。

（二）金融监管与金融检察的"双向耦合"

金融监管职能与金融检察职能之间存在内在统一的"双向耦合"关系，这为完善金融治理体系提供了理论依据与实践指引。金融监管作为"一阶观察"，其核心功能聚焦于"事前预防"。它依托制定规则、市场准入、现场与非现场检查等微观审慎及行为监管举措，深度参与金融市场运行，力求将风险遏制在萌芽状态，以实现宏观调控目的。然而，"一阶观察"不可避免地存在局限性。在此情形下，金融检察作为"国家法律监督机关所承担的'二阶观察'角色"，其独特价值得以凸显。所谓"二阶观察"，即对"一阶观察"的再度审视。[①]当风险未能得到有效防控而演变为违法犯罪行为时，金融检察可以介入并实施"事后归责"以及"微观"层面的精准打击，但其价值不止于此：在具体案件办理进程中，检察机关可凭借独特视角，通过刑行反向衔接机制，审视前端行政监管漏洞以及法律适用偏差。

"二阶观察"可对"一阶观察"形成反向赋能效应。检察机关在案件办理过程中所察觉的制度性漏洞以及普遍性风险点，能够为行政监管部门调整政策、完善规则提供直接且具时效性的依据，进而提升事前预防的精准度，达成从"事后补救"到"事前防范"的转变。反之，当二者的耦合关系处于顺畅状态时，治理效能将得以显著提升。例如，检察机关在办理非法集资等上游犯罪案件时，若发现犯罪所得去向存疑，可主动与中国人民银行反洗钱部门开展协作；中国人民银行依托其专业优势，通过分析可疑交易、穿透资金链条，为司法机关提供关键证据，从而实现对下游洗钱犯罪的连带打击，推动"打财断血"与"行刑双罚"的协同推进。因此，重塑并激活金融检察与金融监管的"双向耦合"关系，构建从微观层面到宏观层面、从事后处理到事前预防的良性循环，是破解当前金融治理困境的重要举措。

智能化赋能"四大检察"面临的核心挑战

新时代"四大检察"的全面协调、充分发展，为承载并达成金融协同治理的使命提供了系统性的制度框架与有力支撑；而要使这一框架切实发挥效能，还需借助"数智赋能"。

（一）数智赋能"四大检察"

其一，运用数智化手段赋能刑事检察工作，以精准高效的态势打击金融犯罪活动。重大金融刑事案件往往与金融风险的扩散存在深度耦合关系，若处理不当，极有可能引发系统性金融风险。在办理金融案件的过程中，应秉持"在办案中监督，在监督中办案"[②]的原则，这就意味着刑事检察工作必须依托数智化手段。具体来说，需要构建专业化的金融案件知识数据库，以便为检察官精准推送相关法律法规、典型案例以及金融专业知识。同时，应搭建数字化的行政执法与刑事司法衔接平台，实现跨部门信息共享、案情通报以及案件移送的自动化与智能化，从根源上解决"有

① "二阶观察"是对观察的观察，指的是对社会系统自身或其他社会系统观察的反思性观察。即当一个系统不仅观察其环境中的事件（一阶观察），而且还观察自己或其他系统是如何观察这些事件的时候，就发生了二阶观察。Niklas Luhmann, "Deconstruction as Second-Order Observing," *New Literary History*, vol. 24, 1993, pp.763-782。

② 余钊飞：《"四大检察"与执法司法制约监督体系之构建》，《法律科学（西北政法大学学报）》2021年第1期。

案不移、以罚代刑”这一长期存在的问题。

其二，借助数智化手段为民事检察赋能，于海量案件中精准捕捉风险信号。金融领域的民事案件不仅数量众多，而且专业性极高。如前所述，在 P2P 平台“爆雷”前，相关民事诉讼量通常会呈现出急剧增长的态势，此乃风险的“数据先兆”。若缺少大数据分析工具，仅依赖法官、检察官开展个案审查，难以发觉其中潜藏的规律。此外，针对新型金融法律问题裁判标准不统一、金融领域民事虚假诉讼频发等现象，亦需借助大数据比对与关联分析来达成有效识别与监督。

其三，借助数智化手段为行政检察赋能，提高金融监管的法治化程度。金融行政监管所涉及的法律规范繁杂多样，行政机关享有较大的自由裁量权。检察机关若要对金融领域行政执法、行政复议、行政诉讼等全流程实施有效监督，就需深度介入并处理相关案件信息，凭借技术手段发现执法与裁判中的不当问题，通过抗诉、提出检察建议等途径，推动依法行政与司法公正。

其四，借助数智化手段赋能公益诉讼检察工作，以维护金融消费者的公共利益。公益诉讼检察亦归属于“大公诉权”的范畴。[①]“拓展公益诉讼案件范围”是强化法律监督的关键方向，金融秩序领域理应成为检察公益诉讼的重点关注领域。以上海为例，地方检察机关已在金融秩序维护领域积极探索开展公益诉讼工作。在处理涉及众多投资者的群体性纠纷时，传统办案模式已难以契合实际需求。检察机关唯有依托金融系统的大数据融合支持，方可有效履行公益诉讼职能，为广大金融消费者提供权益保障。

（二）智能化赋能中的核心挑战

数字检察孕育于突破法律监督“被动性、碎片化、浅层次”困境的进程之中。上述问题不仅是当下法律监督质效欠佳的集中表征，更是长期掣肘检察机关化解监督职能虚化、弱化等难题的结构性缺陷。[②]尤其是在新时代背景下，公众在安全、环境、理财等领域的需求持续增长，传统法律监督模式的上述弊端更为凸显。然而，数字检察在化解固有困境的同时，也滋生了一些新问题。

其一，技术层面存在“短板”式瓶颈。一是数据壁垒问题。有效推进金融检察工作的关键在于掌握数据，唯有通过对大数据进行碰撞、比对与分析，方可发掘潜在的金融案件监督线索。然而，当前金融机构、司法机关以及监管部门之间的数据尚未实现互联互通，“数据孤岛”现象导致职能分析结果出现偏差。此外，内外网相互隔离、模型参数与硬件设施存在差异，使得实验室成果难以在内网进行落地应用。二是算法局限性。金融案件通常涉及多层嵌套的交易架构、跨市场的风险传导以及新型金融工具的复杂运用，其法律关系与事实认定具有高度专业性与模糊性。若算法模型基于有限样本或简化逻辑构建，极易因难以精准呈现此类复杂性而产生误判。同时，算法决策过程的不透明性可能形成“技术黑箱”，既妨碍监督过程的可追溯性，也可能掩盖模型自身的逻辑瑕疵，降低监督结论的可信度。三是数据安全隐患。金融案件数据往往包含账户信息、交易流水、商业秘密等敏感信息，其智能化处理需历经采集、传输、存储、分析等环节。若缺乏全流程加密与访问控制机制，极易因技术漏洞或操作失误导致数据泄露，既侵害金融主体的信息权益，也可能违反《数据安全法》关于核心数据保护、风险防控的强制性要求，引发法律责任与信任危机。

其二，机制层面存在“梗阻”现象。法院的审判数据、市场监督管理部门的投诉数据、金融监管部门的监管数据以及公安机关的刑事犯罪数据相互分离，使市场内的金融风险表征未能得到及时整合、分析与预警。当前存在协同的技术障碍，即检察院、法院、司法行政机关的系统网络相互独立，数据交换在很大程度上依赖线下方式或通过隔离网闸进行，效率低下且信息损耗严重。这些技术障碍的背后，是更为严峻的制度障碍：部门之间存在不愿共享、不敢共享，甚至不会共享的情形。简言之，一是协同欠佳。“四大检察”内部智能工具标准不统一，如证据格式、分析模型存在

① 陈军：《“四大检察”改革背景下的检察权能配置探析》，《政法论丛》2020 年第 5 期。

② 贾宇：《论数字检察》，《中国法学》2023 年第 1 期。

差异，跨检察环节衔接不够顺畅。二是权责不明晰。智能化决策与检察官自由裁量权的界限模糊，引发“算法主导办案”的司法决策权争议。三是人才短缺。既精通金融与法律，又掌握智能技术的复合型检察人才稀缺，技术应用示范落地存在困难。

其三，伦理与规范层面存在“潜在风险”。一是价值冲突问题。金融案件往往涉及多方利益的博弈、复杂的交易安排以及特殊的行业背景，实现个案正义需要兼顾法律适用与金融实践的特殊性。若过度依赖算法开展标准化处理，可能会因机械套用模型逻辑而忽略个案争议的独特性。因此，需要在数智化所带来的效率导向与司法公正原则之间寻求平衡，避免技术理性对实质正义的消解。二是法律滞后问题。当前法律体系针对智能技术在金融检察监督中的应用缺乏针对性的规范，如模型开发者与使用方的权责划分、算法决策失误时的责任归属等核心问题尚未明确，导致实践中存在规范模糊区域，可能对数智化监督的合规推进产生制约。

“四大检察”数智赋能的破解路径

构建服务于金融案件“四大检察”的跨部门协同平台，需以技术突破作为支撑，在破除数据壁垒、精准适配算法、筑牢安全机制等核心难题的基础上，同步消除金融案件办理过程中跨域协作、权责划分、规范适用等制度性障碍，最终实现技术赋能与制度完善的双向协同，为全方位提升金融检察监督效能提供支撑。

第一，技术优化：构建“安全+高效”的智能支撑体系。首先，破除数据壁垒。借助安全多方计算、隐私计算、联邦学习等前沿技术，在保障数据安全与隐私合规的基础上，搭建一个涵盖公检法司与金融监管部门的业务协同与数据中台，达成“精准需求+关键技术”模式下的“四大检察”数据交互。依托该平台，能够实现以往难以达成的业务流程。其次，算法迭代更新，研发适配金融案件的“可解释性算法”。在拟构建的金融风险预警模型中，系统可自动解析法院的裁判文书，提取涉诉主体、案由、异常高息等关键要素，并与检察院的起诉书信息、金融监管部门的行政处罚信息开展碰撞分析。最后，加强安全防护。运用联邦学习、加密技术、隐私计算等关键技术确保数据传输与存储安全，划定智能工具的“数据使用边界”。

第二，机制重构：完善“四大检察”协同赋能流程。一是统一标准。拟定“四大检察”智能化工具运用规范，如证据筛选智能体、风险评估指标等，明晰各环节智能辅助的权责清单。二是构建“技术+法律”交叉学科研究与实践团队。借助检察官技术培训等内部培养途径，以及聘用数据工程师、开展校检合作等外部引进方式，突破人才瓶颈。三是试点先行。挑选近两年来案件数量居前的金融案件，如非法吸收公众存款罪、集资诈骗罪、骗取贷款罪等罪名，开展智能化协同办案试点工作，总结具备可复制性与可推广性的经验。

第三，规范保障：筑牢“法治+伦理”的制度防线。一是完善立法工作。积极推动与智能检察相关的规范性文件出台，明确算法审查机制、数据使用边界以及责任追溯规则。二是开展伦理审查。针对金融案件智能工具的应用，应设立跨领域伦理委员会。该委员会需重点围绕智能工具是否侵犯当事人财产信息、交易隐私等权利，以及算法决策是否符合金融监管伦理与司法公正原则，开展系统性“正当性评估”。三是强化主动担当与监督制约。将智能化办案纳入检察监督体系，这并非在发现金融监管漏洞后仅制发一份检察建议就宣告工作结束，也不是包揽所有事务，而是要在依托大数据开展金融类案监督的同时，基于金融案件监督职能主动“担当”，[①] 积极协同并推动其他职能部门协同“共治”。定期核查算法运行的公平性与透明度，防范技术滥用，共同构建金融风险的隔离屏障。

① 贾宇：《论数字检察》，《中国法学》2023年第1期。

人工智能对民事司法的挑战与应对

张兴美，吉林大学法学院教授

民事司法领域人工智能的表现正在引发学界广泛关注。当下，人们通过向 Lexis + AI 机器人发问便可以了解法律规则、获得判例摘要、评估诉讼风险和洞察诉讼策略；北京高院推出的“睿法官”系统也可以帮助法官智能分析案

情、自动生成庭审提纲和裁判文书，并对裁判结果进行类案对比进而做出偏离度预警。人工智能的这些新近应用表明，技术不再仅仅是诉讼交往的媒介，而是真正成为智能的“第四方”，[①]在实质的纠纷解决层面发挥更大的效用。然而，当我们惊喜于人工智能所带来的无限可能性之余，也要穿透技术的迷雾，对人工智能的运行原理及其作用于民事司法的潜在风险保持清醒的认识，如此方能防范技术至上主义对民事司法的异化，实现人工智能和民事司法的良性互动。

现代人工智能的运行逻辑

界定人工智能并非易事，对于法学者来说尤为如此，人工智能作为一个泛化的概念，它可能指代不同的系统和模型。尽管如此，我们仍然可以从人工智能创设的初衷及其技术现状中大致把握现代人工智能的运作逻辑。

人工智能萌芽于 20 世纪 50 年代，艾伦·图灵（Alan Turing）开创性地提出，如果机器能够回答人类的问题而不被辨别出其机器身份，那么该机器就具有了智能。约翰·麦卡锡（John McCarthy）等人在 1956 年正式命名“人工智能”（Artificial Intelligence）一词时也旨在用这一概念指代机器模仿人类智能的有关问题。[②]由此，人工智能可谓是人类智能的机器实现，意在使机器能够针对人类设定的目标进行感知、记忆、计算、学习、判别、预测或决策等。然而，传统的以符号主义为核心的人工智能在上述目标的实现方面是非常脆弱的，传统人工智能依赖领域专家构建的静态知识库和人类程序员预先编码的逻辑规则运行，有限的知识获取虽然保障了智能过程的可解释性和逻辑严谨性，却极大限制了传统人工智能的现实适应力。换言之，传统人工智能只能处理简单且直接的问题，一旦遇到复杂或模糊的情境，传统人工智能就会遭遇知识获取瓶颈。除此之外，传统人工智能离不开相关领域大量专家的投入，手动编程的过程及其更新耗时耗力，这些也严重影响了传统人工智能在实践中的可扩展性。[③]

数字驱动系统的出现使人工智能迎来新的奇点，机器学习、自然语言处理等技术让人工智能的发展突飞猛进。在现代人工智能语境下，机器不再依赖人类的经验或直觉获取知识，而是以数据（特别是连接互联网的实时数据）代替人类的感知、经验和记忆。机器对人类学习规律的模仿也不再依托显示编程的方式，而是以统计算法训练数据进而构建预测函数模型。目前以大数据和人工神经网络技术为核心的深度学习进一步提高了机器学习的广度、深度和自主性，这极大促进了现代人工智能的适应性和有效性。机器通过新数据与预测函数之间的相关性映射智能地做出预测和判断。在整个输入与输出过程中，自然语言处理技术使人类能够使用日常语言与机器进行交互。

① Ethan Katsh and Janet Rifkin, *Online Dispute Resolution: Resolution Conflicts in Cyberspace*, San Francisco: Jossey-Bass, 2001, p. 93.

② 连师友编著：《人工智能导论》（第 2 版），北京：清华大学出版社，2025 年，第 23 页。

③ 爱德华·A. 费吉鲍姆、帕梅拉·麦考黛克：《第五代：人工智能与日本计算机对世界的挑战》，汪致远等校，上海：格致出版社、上海人民出版社，2020 年，第 98—109 页。

概言之，现代人工智能的运作以数据为基础材料，以相关性统计为核心方法，根据历史数据权衡概率进行模式分类和回归分析，进而对人类的目标作出回应。这种运作方式向民事司法投射后，将很大程度上改善司法的可及性，人们将以更便捷的方式获取法律信息、以更易被理解的方式表达观点；人工智能的结果预测能力可以帮助人们采取适当的法律行动，防范司法过程中的风险和不必要的成本支出；法官也可以借助历史判例的统计分析把握裁量尺度，提高审判质量。

现代人工智能对民事司法的挑战

欧盟《人工智能法案》指出，我们需要对民事司法引入人工智能所产生的“高风险”保持警惕。这些风险存在于人工智能运行的基础、过程和结果三个方面。

（一）人工智能运行基础层面的挑战

现代人工智能的基石是数据，在民事司法领域具体表现为案由、当事人、案情、裁判结果和程序规则等信息。这些信息本身的真实性是现代人工智能有效助力民事司法的前提。然而人工智能自动获取的数据信息并非总是准确的。其一，数据本身就可能是错误的。例如，互联网上很多错误的法条表述都可能成为人工智能的数据源。其二，人工智能会产生数据“幻觉”，它会将一些看似令人信服的文本串联在一起提供一些虚假的数据信息。例如，美国的两名律师将使用 ChatGPT 搜集到的案例提交给了法院，而事实上 ChatGPT 提供的这些案例根本不存在，法院最终驳回了此案，并对律师进行了惩处。[①]其三，人工智能也会存在曝光偏差，即当在特定数据集上训练的函数模型接触其他数据时，很可能表现不佳，无法准确解读新数据，而这些有偏差的数据又会反向传播成为下一个预测训练的基础数据。这些数据失真问题将在源头上影响人工智能适用的准确性，而且这种影响相较于人类感知错误对民事司法秩序和公信力造成的冲击更大，因为人类的感知错误通常只影响个体处理案件时的准确性，而人工智能的数据错误将波及整个系统。

另一个基础层面的挑战是，现代人工智能可能会侵犯个人隐私和信息。人工智能在收集和处理数据时会对用户的交互信息进行抓取、记录和学习，甚至与第三方共享信息，在未征得相关主体意愿，特别是当数据量过大、算法比较复杂，以至于无法提前告知个人获得他们的同意时，人工智能在助力司法解决纠纷的同时也会制造新的纠纷。

（二）人工智能运行过程层面的挑战

人工智能的运行过程依赖统计算法，当把法律上的概念和因果关联转化成算法可处理的变量用以训练数据时，人工智能就可能因历史数据结构失衡、特征权重分配不当、代理变量失范或数据代表性缺失等原因产生偏见或歧视。例如，如果用以预测抚养权纠纷案件的训练数据普遍将抚养权判给母亲，那么人工智能就很可能延续这一偏见优先推荐母亲作为监护人；在处理离婚财产分割案件时，如果人工智能用“收入差异”变量代替衡量“家庭贡献度”，就会片面得出家庭中主导经济的一方获得更多财产的结论；人工智能在使用自然语言处理技术分析合同纠纷案件时，也可能由于缺少地方方言或少数民族语言的训练，而将此类合同识别为意思表示不清或条款模糊，进而影响相关合同主体的利益。综上，算法偏见或算法歧视不仅会对司法公正造成影响，也会在种族、性别、年龄、地域或经济等方面引发更为深远的司法伦理问题。

除此之外，人工智能的运行过程也会对司法公开构成挑战。司法应当公开，法官裁判的事实和法律依据必须明确，法官是如何权衡当事人陈述的观点和所提出的证据，又是如何适用法律的，应当以看得见的方式呈现。然而，现代人工智能与这些司法公开的要求之间存在现实张力。现代人工智能在借助大模型和相关性概率推理增强其扩展性和适应性的同时，也产生

① Christopher L. Griffin Jr., Cas Laskowski, Samuel A. Thumma, “How to Harness AI for Justice,” *Judicature*, 2024.

了可回溯性和可解释性障碍，即所谓的“算法黑箱”。算法看不见、摸不着，人们很难对机器感知数据、意识先验、表征学习和知识库推理的过程形成具象化的认识，即便是开发者也难以追踪和解释机器模型的每一个分析步骤，但如果人工智能无法清晰地提供其输出背后所考虑的因素和决策的依据，那么当其向司法投射后，司法的合理性和正当性就会受到质疑。此外，由于人们尚不清楚人工智能算法的分析过程，自然也就难以识别算法是在何时以及如何出错的，其结果是人们对算法决策提出异议的可能性也会受到影响。

（三）人工智能运行结果层面的挑战

在运行结果层面，人工智能会对司法亲历性和主体性构成挑战。人工智能对民事司法的预测是基于大量类案参数的分析，这种泛化的结果有客观技术支持、有历史判例“背书”，又具有“群体”的效果，人们出于“从众”心理，或是对大模型的盲从，自然也就很容易跟从数据统计的结果，由此便形成了数据依赖。数据依赖现象大大削弱了法官的能动价值，容易导致“机械司法”。例如，Compas 是在美国使用较为广泛的司法人工智能系统，很多法官比较信任 Compas 的风险评估结果，但当法官遵循 Compas 系统建议裁判时，往往会忽视对个案情节的审查，固化系统潜在的偏差，因而 Compas 系统遭致诸多非议。有鉴于此，审理威斯康星州诉卢米斯案（State v. Loomis）的法官在上诉判决中强调，尽管 Compas 系统仍然可以用于案件审理，但其适用有局限性，Compas 系统不应该成为法官裁判的决定性因素。[①]

① State v. Loomis, 2016 WI 68.

民事司法应对现代人工智能挑战的路径

现代人工智能因数据失真、算法偏见、算法黑箱和数据依赖等原因会对民事司法构成挑战，挑战的背后是技术主义与司法价值之间的碰撞。在技术与司法之间，司法的主体性才是根本。因此，在技术迭代发展的今天，我们有必要重申司法的主体观，并从法官参与、司法公开和系统监管等维度规范人工智能适用，进而有序引导技术赋能民事司法现代化建设。

（一）秉持司法主体性理念

人工智能的分析和预测能力的确会对法官的裁量与判断构成影响，但人工智能的定位仅仅是模仿人类，其在民事司法领域无法替代人类，更无法超越人类。首先，如上文所述，人工智能本身即存在“错误进，错误出”或者“偏见进，偏见出”等问题，这离不开法官的审查和纠正。其次，人工智能的本质是“统计概率”而不是“临床诊断”，它是将严谨的司法演绎推理简化为输入与输出之间的相关性判断，这种判断的结果虽然有助于从历史的维度把握案件的趋势，无法兼顾个案的具体情境，更无法结合个案的特殊性权衡法律的价值、诠释模糊的规则。因此，尽管人工智能有助于提升司法裁判的准确性，但对个案正义的精准回应仍有待人类法官来完成。再次，人工智能是对历史经验的归纳和总结，它无法面向未来完成真正的创新，而能动地解决司法中的新问题恰恰是法治发展的不竭动力。最后，人工智能不涉及意识和精神，它只是形式化地对问题作出判断，却无法理解问题处理背后的真意，更无法像人类亲历交流那样建立同理心，获得当事人对裁判结果的认同感。

综上，在人工智能的时代，人工智能只是一种技术辅助，司法相关主体应该思考的是如何利用人工智能提升独立判断的能力。当然，需要说明的是，坚持民事司法的主体性，并不妨碍借助人类和人工智能各自的相对优势来优化民事司法在解决纠纷方面的分工，人工智能更擅长处理一些重复性的、常规类的案件，而人类法官更需要把握新型的、疑难复杂的案件。如此通过人工智能和人类法官的协同交互，便可能化挑战为助力，共同促进民事司法的优化建设。

（二）注重法官参与

人工智能回应民事司法的有效性很大程度上取决

于训练数据的质量、规模和包容度。为此，各地法官就需要参与到司法人工智能系统的建设中，表达需求、训练数据，并尽可能使人工智能系统兼容不同地方的司法差异，以此降低出现数据错误和算法偏见的可能性。一个可能的方向是，整合现有的司法信息公开平台，打造专业的司法人工智能系统。相较于一般的人工智能系统，专门为司法设计的人工智能系统是从已知、可靠且权威的数据库获取数据信息，而且定制化的系统通常依托专用服务器运行，严格控制了数据与第三方共享，这有助于最大程度保障数据隐私和安全。

除此之外，法官的参与还可以体现在对人工智能系统处理结果的监督与复核上。在司法实践中，人工智能系统给出的建议或判断，最终仍需法官依据法律专业知识、社会伦理道德以及具体案情进行综合考量与决策。法官通过审查人工智能系统的分析过程与结论，能够及时发现并纠正可能存在的逻辑错误或偏离法律精神的情况，确保司法裁判的公正性与准确性。

（三）坚持司法公开

尽管人工智能的运行过程不易被理解，但作为司法的辅助，人工智能仍然要遵循正当程序，司法公开作为正当程序的核心要义，应当得到最低限度的坚守。具言之，人工智能系统的功能和目的需要向社会公开，人工智能的分析和决策过程应该以可被理解的方式向受其影响的人们解释。当然，现实情况下，法院受制于专业限制会将技术外包给第三方，而私营主体经常以知识产权或商业秘密保护为由拒绝信息披露。对此，正如审理休斯顿教师联合会诉休斯顿独立学区（Hous. Fed'n of techs. v. Hous. Indep. Sch. Dist.）案件的法官所指出的，虽然保护商业秘密应该受到肯定，但没有向受判决不利影响的人们告知算法决策背后的原因，构成对正当程序的严重违反。[①]进言之，人工智能运行过程中所要求的司法公开，并非是要公开全部技术和数据，而是要向法院和当事人提供信任人工智能系统的适当信息，使其能够识别系统潜在的错误并有机会采取适当的司法救济措施。

（四）加强系统监督

虽然人工智能实现了人机交互的自动化，但人工智能预期功能的实现仍然离不开人工监督的介入，特别是在高风险的民事司法领域。对人工智能系统的监督可从事前、事中和事后三方面展开，一是民事司法引入人工智能系统前应当根据《人民法院在线运行规则》《最高人民法院关于规范和加强人工智能司法应用的意见》等法律规范对该系统的安全性、合法性和伦理性进行评估；二是人工智能系统适用过程中，司法及相关机构定期对系统的运行效果和适用情况进行监测与审查，监控算法模型的偏差率和准确度；三是允许受人工智能影响的主体提出异议，并根据人工智能的表现和相关主体的反馈意见更新和升级系统。上述人工介入监督人工智能系统的相关举措，虽不能完全消除现代人工智能对民事司法的不利影响，但及时的人为干预无疑会提升人工智能的可控性，使人工智能尽可能在规范的限度内服务于民事司法。

① Houston Fed'n of Teachers, Local 2415 v. Houston Indep. Sch. Dist., 251 F. Supp. 3d 1168.

结语

面对人工智能的蓬勃之势，我们既要扎根于技术本身，了解人工智能的运行原理，也要回归司法的本质，审视人工智能的局限，以构建司法人工智能的最佳方案。本质上，现代人工智能是数据驱动下的相关性统计分析，该分析范式虽有助于增进司法公正、提升司法效率，但其也会因数据失真、算法偏见、算法黑箱或数据依赖等因素给民事司法带来挑战。面对人工智能嵌入民事司法的潜在风险，民事司法应当始终秉持人类的主体性，并从法官参与、司法公开和系统监管等方面把控人工智能的安全有效性。以司法为主导、以技术为辅助定位和发展司法人工智能，方有助于推动新时代民事司法现代化行稳致远。

司法领域智能科技的介入标准

黄泽敏，上海财经大学法学院副教授

司法与科技的交叠发展已经成为智能时代的主旋律。2022 年 12 月出台的《最高人民法院关于规范和加强人工智能司法应用的意见》指出，到 2025 年，我国要基本建成较为完备的司法人工智能技术应用体系，为司法为民、公正司法提

供全方位智能辅助支持。时至今日，一系列政策、文件的出台都无一不宣示着智能科技正逐步有计划、有规模地开始介入司法领域。曾经的正当性争议（司法领域是否应当允许智能科技介入）已悄然淡出学术视野。再没有人愿意极力宣称智能科技在司法领域毫无作为。同样，亦少有人坚持智能科技能够完全取代司法人员处理所有司法问题，如取代法官裁断案件等。两个极端对立的主张在隐去的同时，浮现出的是另一个议题——限度。

既然智能科技有所作为，但又无法全面介入，那么它的应用势必就有一个限度。在司法实践中，一方面以人工智能技术为核心的智能科技融合了大数据、云计算、区块链、大模型等相关技术，能够在辅助量刑、文书自动生成、类案推送等领域发挥其优势，[①]对于解决司法效率低下等问题成效显著。[②]另一方面，智能科技在司法中的应用局限和异化也显而易见，如算法偏见导致司法不公等。实践中，甚至在有些场域已经出现了司法辅助科技，为了智能而智能的异化局面。因此，在司法领域，仅确立智能科技的辅助地位是不够的。唯有划出一条可行的界限，即树立司法领域智能科技的介入标准，智能科技方能真正有效助力司法。

智能科技介入标准的提出

什么样的界限是可行的？当下智能科技的种类、数量繁多，且随着技术的发展，更新换代亦是常态，因而为某一种智能科技的介入划定标准并不可行。宏观上看，智能科技的介入其实就是以智能科技取代司法领域中的固有行为事项，而界限无非就是明确哪些行为事项可以被取代。在司法领域中，行为事项以一种多样态的方式呈现，如根据司法人员分类，可以分为法官的行为事项、司法辅助人员和司法行政人员的行为事项；而根据司法职能分类，又可以分为审判的行为事项、强制执行的行为事项、司法宣传以及司法监督等行为事项。不同的分类甚至可以在其之下细分为更为具体的行为事项，如审判之下，可以细分为庭审活动的各种行为事项等。就介入的标准而言，若要在这样的分类逻辑之下确定哪些行为事项可由智能科技介入、取代，同样不可行，因为行为事项分类无法封闭式合理穷尽。

无疑，介入标准的确立，首先需要对司法有一个一般化的理解。根据已有的司法理论，司法是一种以审判为中心的高度职业化、程序化的法律活动。[③]而审判的本质是一种法律判断，它外化为一种根据法律规则裁判纠纷的行为。在整个司法领域内，一切活动都由不同的司法行为串联而成，有的行为关涉法律判断，如法律发现、法律解释、证据合法性的认定等，而有的行为则无关乎法律判断，如证据的提交、文件的传送、文字的记录、账户的冻结等，除此之外别无

① 郭春镇、黄思晗：《刑事司法人工智能信任及其构建》，《吉林大学社会科学学报》2023 年第 2 期。

② 赵杨：《人工智能时代的司法信任及其构建》，《华东政法大学学报》2021 年第 4 期。

③ 周玉华主编：《中国司法学》，北京：法律出版社，2015 年，第 16—19 页。

其他。因此，司法领域中的行为事项可以被合理划分为涉及法律判断的行为事项和不涉法律判断的行为事项。智能科技的介入无非就是介入这样两类行为事项。考虑到法律判断的独有特性，可以提出一个一般性的介入标准。

强式标准：凡涉及法律判断，智能科技不可介入。这是一项否定性标准，它并不排斥所有智能科技介入司法领域，而是强调涉及法律判断之行为事项，不可交由智能科技予以处理。之所以是强式，是因为它阻却了对关涉法律判断的行为事项进行智能科技化的一切可能。然而，强式标准是不是唯一合理的标准，仍有一定的讨论空间。从现实的角度看，智能科技对一些关涉法律判断之行为事项的介入、取代不可谓没有成效，如最高检推广的一些普通犯罪的大数据监督模型、法院系统的案件繁简分流智能化平台等。而从理论层面看，有些法律判断确实存在被介入的可能，如是否达到《刑法》第 17 条刑事责任年龄的判断。这个判断虽是法律判断，却是一种纯粹的数字判断。除此之外，有的法律判断是一种概念判断，如手枪是否属于《刑法》第 297 条“非法携带武器、管制刀具、爆炸物参加集会、游行、示威罪”中的武器。对此，通过概念涵摄的方式就可以得到一个确定的判断。涵摄，即将外延较窄的概念划归为外延较宽的概念之下，是概念法学的自然产物，它的运作包括两个部分：一是定义处于上位的构成要件的概念和处于下位作为陈述的案件事实的概念；二是确定上位概念的全部要素是否在下位概念中全部重现。所以，概念判断的本质是一种形式判断，它与智能科技主要的形式主义内核是相契合的。正因如此，强式标准就存在修改的余地，形式的法律判断应当从中予以剔除，而介入标准就转换成中等标准。

中等标准：凡涉及非形式的法律判断，智能科技不可介入。中等标准是否现实可行，关键在于能否有效区分法律判断的形式与非形式。理论上，法律判断不可能是单一的形式判断，它一定复合了其他判断要素，因为法律不是概念的形式堆砌。以上述示例中手枪是否属于武器来看，虽然能够通过形式的概念要素予以判断，但主张手枪属于武器同时就一定意味着判断同样符合了《刑法》第 297 条背后的价值主张，以及整个刑法体系和法体系背后的价值要素。对此，即使是数字判断也是如此。或许于现实而言，通过形式的法律判断可以解决绝大部分的简易问题，因而中等标准可以成为一个合理的介入标准，但在极端个例中，中等标准就失去了有效性，如法律规则：任何车辆禁止进入公园，判断救护车是否属于车辆。若沿着形式判断的路径，则很可能得出错误的主张。中等标准的适用在这里遭遇的困境并非是理论上能否区分形式法律判断和非形式法律判断，而是无法确切知晓在什么条件下应当进行概念的形式判断，因为它与个例的具体场景有关。在这个意义上，中等标准和强式标准之间就存有竞争，而竞争的本质似乎是司法效率与司法正义之间的矛盾。若坚持中等标准，绝大多数的法律判断问题将得到智能科技的加持，但与之相对的结果是容易牺牲极端个案的公正。然而，个案公正的泯灭，其所带来的对司法公信力的破坏无法预估。因此，中等标准需让位于强式标准。

强式标准的确立折射出一种司法与智能科技和谐共生的理想状态。但现实是，智能科技发展的雄心并不止于此，它们意图不加区分地介入到所有法律判断之中，如法院的类案推送系统，其毫无疑问不仅违反了强式标准，同时也违反了中等标准。所以，在智能科技浪潮之下，一条介入的最低限度标准就应当予以划定。考虑到司法领域的所有法律判断中，有的法律判断直接决定了权利（权力）和义务（责任）的分配，而有的不是，如裁判结论和为得出裁判结论而涉及的一系列的法律判断。对此，如果介入是有底线的，那么这些法律判断就绝对不可让渡，继而就构成如下弱式介入标准。

弱式标准：凡涉及直接决定权利（权力）和义务（责任）分配的法律判断，智能科技不可介入。弱式标准的提出是一种无奈，它既不优于中等标准，也不优于强式标准。当智能科技试图逾越法律判断的形式与非形式之分，甚至逾越法律判断与非法律判断之分，司法行为就将被彻底解构。而为了防止司法的畸化，弱式标准就是司法的最后一道尊严。

智能科技介入标准的证成

介入标准虽分为强式、中等和弱式3种，但中等和弱式其实理应让位于强式标准，它们的成立在上文已经予以论述，所以唯一的证成就是对强式标准的证成。根据强式标准，证成必依赖对法律判断的理解。强式标准之所以成立，智能科技之所以不可介入涉及法律判断的行为事项，是因为除了形式要素之外，法律判断还复合了价值、道德、情感和权力等其他判断要素。当下智能科技不可能也不应当介入。

首先，法律判断是一种价值判断。法律虽是由概念构成的，但能够把这些概念整合起来的恰恰是价值。价值隐匿在法律之后，但却是法律的灵魂。通常情形下，手枪是不是武器，或许可以转换成为一个概念问题。然而，一旦碰上盐酸是不是武器的问题，它就不再是一个简单的概念判断，而是需要深入到概念构成的规则背后，通过规则所意图实现的价值来寻求支持。作为一种价值判断，智能科技似乎对此有些力不从心。以可计算论辩模型为例，可计算论辩模型背后的理论依据是将司法推理化约为经验层面的事实要素比对。[①]这一技术将法律判断的推理过程简化，仅表现为通过对比概念要素之间的相关性实现判断，而将概念背后所蕴含的价值因素排除在外。于是，基于计算推演的智能推理实际上就减弱乃至剔除了司法过程中的价值判断，未能完全有效体现司法推理的法理逻辑。对此，尽管有观点主张可通过“价值计算”的方法将涉及价值评价的要件及其权重进行数值转化，并建立计算模型，以期解决价值判断的难题，[②]但其发展仍存在不可逾越的障碍。

其次，法律判断是一种道德有涉的判断。法律判断不仅是概念判断、价值判断，还是一种关涉道德的判断。一些重要的道德理念，如“公序良俗”“保护弱者”等贯穿于民法等诸多法律领域，成为法律规范的重要指引。这也表明法律判断不可避免地蕴含着对道德因素的考量和判断。在处理每一个具体案件时，法律的适用、解释都会受到不同程度的道德因素的影响，如在涉及“安乐死”“同性恋”等敏感案件时，法律条文往往难以提供明确且无争议的解决方案，法官必须依据社会主流道德观念以及人类对生命尊严、自由选择等方面的道德认知来进行综合判断，以实现法律适用与道德观念的平衡。而智能科技就难以实现道德判断，这是由智能科技的数字本质和机器理性决定的。一方面，道德判断往往涉及对人类文化、历史传统等复杂因素的综合考量，而这些都不是可以通过简单的数据和逻辑关系来精确描述的；另一方面，道德判断需要主体具备自我意识、自由意志以及对他人的同理心等。智能科技本质上是一种工具，它没有自我意识，也无法像人类一样主动地进行道德思考和选择，其只能根据预设的程序和算法对外部信息作出反应，而无法对自身的行为及其后果进行道德反思和责任承担。所以，法律判断的道德属性已经决定了其无法也不应交由智能科技来完成。相反，在道德判断方面的局限性恰恰证成了人类主体在司法领域的不可替代性。

再次，法律判断是一种包含情感的判断。法律判断绝非冷冰冰的概念式判断，它是有温度的。一个法律判断的作出不仅依赖法律知识和法律经验，还牵涉判断者自身的情感。不仅牵涉自身，还关乎社会公众普遍的情感反应。法律判断包含情感，意味着法律判断不再是简单的是非曲直判断，它还包含了在不同情感取向之间权衡的过程。面对司法过程中的人性化需求，如同情、共情等情感因素，智能科技往往难以准

① 陈子君：《智能裁判系统的法律推理逻辑》，《四川师范大学学报》（社会科学版）2024年第2期。

② 李婷：《构建适配人工智能辅助价值计算的核心价值观裁判说理机制》，《法律方法》2022年第2期，第260—279页。

确把握，因为这些情感体验是人类所独有的。现有的智能科技一般只能根据已有的数据和规则进行判断，由于缺乏情感理解和情境感知的能力，因而难以捕捉到案件背后的情感脉络，导致判断过于机械化，缺乏人性关怀。而法官在遵循法律的同时，能够兼顾人情世故，做出既合法又合乎人性的法律决定。这种综合考量的能力是现有的智能科技水平难以达到和复制的。人类的判断力、直觉、移情能力以及对道德准则的遵守，都是建立在长期的社会生活经验和情感体验基础上的。[①]智能科技的情感观念若仅通过训练数据中的统计规律和模式来体现，那么就注定不可能与人的情感能力相比拟，即使技术不断发展，在包含情感的法律判断方面仍可能存在偏差或错误，从而影响裁判结果的公正性和可接受性。

最后，法律判断还是一种具有权力架构属性的判断。在司法领域，法律判断并不是一种非涉他的个人主观表达，它最终一定指向具有法律效力的决定。效力是应然权力的现实投射。有权作出具有法律效力的判断主体只能是那些被特定权力结构正式授权的主体，包括审判人员、检察人员、监督人员等。授权行为本身即是对判断资格的排他性分配。司法权的本质其实就是判断权。司法部门既无军权又无财权，不能支配社会的力量和财富，不能采取任何主动的行为。故可正确断言：司法部门既无强制又无意志，而只有判断。[②]法律判断是司法权的核心，它的生成始终镶嵌在一套由国家垄断并精心设计的权力网络之中；它的内容不仅反映既定权力关系，界定权利、义务与责任，更在个案中不断重塑这些关系。智能科技虽然在技术层面可以提升司法效率，但它既无法理解司法权力架构，更不应注入权力属性。当下智能科技的介入已经不自觉地形成了“算力即权力”的新型权力形态，它可能将司法纳入更隐蔽、更泛在、更多元的技术权力的侵蚀之中。技术权力以科技为手段，以全流程司法数据的获取为基础，借助代码、算法和架构将支配深入到司法运作各个环节的毛细血管中，其结果就是审判的独立性被技术取代，司法人员的主体地位被不断削弱。[③]因此，以智能科技介入甚或取代法律判断就值得警惕。

结语

法律判断的复合并不止于以上四个要素，但这些要素的存在已经足以证成介入标准的合理性。四个要素所展现出来的智能科技的局限性并不能通过技术的更新迭代加以解决。以既有的技术发展路径来看，科技智能与人类的智慧并不是同种类的替代物，人类的经验、直觉和感知能力无法被任何技术所代替，这是由智能科技与人脑系统的底层运作逻辑所决定的。在人类主体层面，司法人员在进行法律判断时，会综合考虑情感、道德、伦理等因素；而智能科技至少目前主要是基于预先设定的算法和数据进行判断，它无法真正理解人类情感的细腻之处。在法律层面，法律不仅仅是规则的集合，它背后蕴含着社会的价值观念、公平正义的理念；而智能科技在进行法律判断时，则呈现出一种僵硬的状态，侧重于对法律条文的机械适用，而忽视法律背后的精神和目的。在法治层面，法治的权威性依赖人们对法律制度和司法过程的信任；智能科技一旦出现错误或者不公正的法律判断，由于其特殊性，很难像追究法律工作者责任那样去纠正和问责，而这种责任的模糊则会降低人们对法治的信任。

科技新时代虽已到来，但法治时代并不因此而新。以上种种不能虽然以目前的视角审视似乎只是技术不能的问题，但即使有一天技术实现了突破，智能科技已经到达与人相同的水准，我们同样需要谨慎对待。只要人依旧与机器相区别，法律判断就不应当被介入。如果将法律判断交由智能科技处理，这将是对人的主体资格的否定，同时也是对法律和法治的否定。

① 邱昭继：《人工智能、法律解析与未来法律实践》，《政法论丛》2022年第4期。
② 李拥军：《司法的普遍原理与中国经验》，北京：北京大学出版社，2019年，第9页。
③ 王禄生：《司法大数据与人工智能技术应用的风险及伦理规制》，《法商研究》2019年第2期。

人工智能辅助裁判文书生成的适用与限制

季平平，上海政法学院民法方法与案例研究中心研究员

裁判文书是司法权力的重要载体，既记录案件审理经过和结果，也展现司法裁判的理由。[①]提高裁判文书质量一直是我国法院改革关注的重点。从历史演变来看，我国传统民事裁判文书的格式沿袭1992年最高人民法院发布的《法院诉讼文书样式（试行）》所确定的模板，经过2003年《民事简易程序诉讼文书样式（试行）》、2006年《关于加强民事裁判文书制作工作的通知》、2016年《人民法院民事裁判文书制作规范》等文件的细化，目前的裁判文书结构以“92式”模板为基础，在此之上派生出简易程序等少数特殊样式。这一传统格式提供了统一框架，但实践中已难以满足司法实际需要。

首先，固定分段模式造成结构混乱和重复冗余。同一事实材料在原告诉称、被告辩称、查明事实和本院认为中反复出现，判决书篇幅冗长却主线不明。[②]读者需要在各部分间辗转查找才能拼凑出完整情节和论证链条，大大增加了理解判决理由的难度。其次，许多裁判文书的证据分析不清、法律论证不透。部分裁判文书以一句“上述事实有……为证”概括全部证据，未展示证据与事实认定的推理过程。对关键证据为何采信、存疑证据为何未采，法院缺乏交代，裁判说理失去坚实基础，当事人难以信服。法律适用部分也常流于草率，有的判决仅援引法律条文就下结论，未解释条文为何适用、相关原则在本案中的具体考量，用空泛的大道理替代推理，难以令人信服。

我国传统裁判文书体例和写作模式存在事实与理由割裂、结构重复冗杂、证据和法律论证不充分等问题，需要通过裁判文书样式和说理方法的改革予以优化。随着人工智能技术的快速发展，法院系统也逐步引入智能工具辅助法官。本文拟在分析我国裁判文书撰写主要问题的基础上，探讨人工智能介入裁判文书撰写后，提升文书说理质量、优化行文风格的路径与边界。

要素式裁判文书对传统裁判文书的改革

要素式裁判，核心是在判决书中以案件“要素”为中心组织内容。对于能够提炼出固定争议要素的案件，判决书不再机械地分为原告诉称、被告辩称、法院查明、法院认为等传统板块，而是围绕特定要素，将原被告的主张、相关证据以及法院对该要素所认定的事实和理由统合陈述。这种写作模式在陈述每一要素时同步穿插法律评析。要素式裁判根植于要件事实理论和请求权基础理论。每个诉讼请求背后对应一定的实体法规范依据，需要满足若干构成要件。要件事实理论主张将这些法律要件转化为具体的可争议事实要素，以此为单位组织审理和裁判。据此，在要素式裁判中，法官围绕诉讼请求涉及的法律要件列出要素清单，逐一审查相关事实是否得到证明，最终使裁判论证严密对应于法律规定—事实要件—裁判结论的推理链条。[③]

（一）要素式裁判文书的革新

要素式裁判的推行针对传统文书弊端提出了矫正方案。

其一，结构合理性。要素式裁判通过按要素分段、

① 胡云腾：《论裁判文书的说理》，《法律适用》2009年第3期。

② 曹志勋：《对民事判决书结构与说理的重塑》，《中国法学》2015年第4期。

③ 冉博：《民事司法中同案同判的智能化实现路径——以要件事实审判思维为基本遵循》，《江汉论坛》2022年第6期。

夹叙夹议将事实与说理有机融合，每一争点要素从当事人主张、证据到法院认定和法律适用，一气呵成地论述，紧贴审理思路且层次清楚。如某地方法院的改革经验所示，要素式文书确保不遗漏任何诉讼请求，又能使法院对争议焦点的认定一目了然，裁判结论与诉讼请求、争议焦点之间建立了明确对应关系，增强了判决结构的严谨性。①

其二，说理完整性。长期以来，不少民事判决书的“本院认为”部分说理不充分，只笼统援引法律条文，未对争议焦点逐一论证，削弱了判决的说服力。要素式裁判强化了逐点说理，对有争议的问题重点论述，详尽阐明依据和理由；对无争议或次要的问题则尽量简化。如此既不遗漏重要论证环节，又防止了内容空泛冗长。尤其要素式文书的说理严格对应实体法要件，法官对各项主张、抗辩是否成立的评判在判决书中都有理有据地展现，大大提高了裁判说理的严谨性和充分性。

（二）要素式裁判文书撰写引入人工智能的可能性

要素式裁判文书高度格式化、规范化，每份判决按统一结构列示法律要件和案件要素，使裁判理由表达更加标准、清晰。同时，以要件事实思维构建的法律知识图谱为机器学习提供了知识基础和训练语料，使人工智能系统能更有效地学习和模拟法官推理过程。因此，要素式裁判文书不仅提升了人工智能撰写说理的质量，也为人工智能辅助裁判创造了坚实的结构基础。②

人工智能与要素式裁判文书在形式逻辑、规范结构和法律认识论层面高度契合。要素式裁判以案件要素为中心，将案件事实、证据和法律要件有机融合成清晰的推理链条，这一点与人工智能决策所依赖的形式化推理方式不谋而合。尽管目前人工智能尚难直接处理疑难复杂案件，但对于可以要素化审理的标准化案件却能发挥显著作用。这种要件事实型裁判方法与司法人工智能的生成规律内在契合，二者都强调从预设的规范要件出发进行演绎推理和信息提取。因此，人工智能可被视为法律形式推理机制的技术延伸，能够满足要素式审判中法律命题建构、规范演绎和论证结构的需要。

人工智能介入要素式裁判的方式

现阶段人工智能介入要素式裁判的方式主要有以下三种。

（一）事实要素抽取

要素式裁判的第一步是从繁杂案情中提炼关键事实要素，而语料的总结与提炼则是人工智能技术在现阶段较为成熟的应用。要素式审判预先根据案件类型设定基本事实要素，当事人围绕这些要素陈述与举证，法官据此认定事实。案件事实结构化使人工智能的介入成为可能，人工智能系统可以聚焦特定要素，从海量材料中自动提取相关的事实和证据，而无需逐字处理冗余信息。通过自然语言处理和机器学习，人工智能能够高效识别卷宗材料中的事实要素。事实要素经人工智能抽取后，裁判所依凭的关键事实更加清晰，每项证据均归入对应要素下论证，不再杂乱无章。人工智能对事实要素的提炼是对法官分析案情思路的技术再现，弥补了以往因事实认定不清导致的说理不足。

（二）要件要素匹配

事实要素明确后，需要将认定事实纳入相关法律规定的构成要件框架下审查。要素式裁判文书的一大优点在于规范结构与案件事实的精准对接，文书按照法律要件逐项阐述对应事实是否符合，实现了法律规范与案件事实的无缝衔接。这种逐条对应的规范结构令人工智能的参与变得顺理成章。智能裁判系统本质上遵循形式推理的逻辑路径，依赖将法律规则和法学知识逻辑化、代码化，把法律要件转变为机器可识别的条件。采用要素式审判后，每一法律要件（如民事责任的构成要件）都被拆解为明确的判断要素，案件的具体事实又与这些要素一一对应。人工智能据此可以充当逻辑推理辅助工具，将输入的案件

① 林遥：《民商事类型化案件要素式审判机制研究——以C市法院民事庭审优质化改革情况为样本分析》，《法律适用》2018年第15期。

② 高翔：《人工智能民事司法应用的法律知识图谱构建》，《法制与社会发展》2018年第6期。

要素与预设的法律条件进行匹配和演绎推理，在很大程度上促进了裁判过程的规范化和标准化。有研究指出，要件事实审判思维是目前最有效对接类案智能裁判的方案，能够促进裁判标准统一、裁判文书格式化和表述规范化。[①]在我国司法强调同案同判、公平统一的背景下，只有将人工智能深度嵌入法律论证过程，才能真正发挥其作用，而要件式审判思维正提供了这种融合路径。

（三）说理逻辑生成

完成事实认定和法律要件匹配后，裁判文书需要给出完整、透明且有说服力的论证过程，解释为何依据认定的事实和法律要件得出了具体裁判结论。要素式裁判文书的说理部分通常按要素逐一论证，依次阐述各法律要件所对应的事实和证据，论证该要件是否满足，再综合各要件的判断形成结论。这样的论证结构严谨清晰，非常适合借助人工智能来生成和完善。人工智能在这一环节的作用主要体现在以下两方面。

一方面，通过模板化技术自动生成基础说理内容，确保每个必要要素的讨论不被遗漏；另一方面，辅助法官对论证过程进行质量检视和优化。由于要素式文书本身具有高度格式化的特征，人工智能可以根据既定的要素框架，提示法官需要进行着重论证的要件，从而避免传统裁判文书中常见的说理不透、论证跳跃等问题——人工智能不会省略推理步骤，每一步结论都能在文书中得到逻辑推演支撑。这种“机器提示＋人工撰写”的模式还能减小不同法官说理能力差异导致的不一致，使裁判说理更为客观中立。[②]更重要的是，有了人工智能生成的初步引导，法官再行充分说理，可以将法律论证提升到更高层次。

人工智能介入要素式裁判的边界

从法律认识论来看，法律推理不是机械的信息查询，而是一种富有人文性的技艺。因此，即便在高度格式化的框架下，裁判说理仍不能交由机器自动完成。法律推理中渗透的人类理性、价值判断和辩论艺术，需要通过法官的参与才能充分展现。要素式方法将隐含的推理步骤层层外显，为司法论证提供了逐层展开的平台，人工智能则能在这一平台上有序执行演绎和检索。但最终赋予裁判说理正当性和说服力的，仍然是法官关于法律与事实融合的洞见。因此，有必要进一步探讨人工智能介入要素式裁判在法理功能和法律价值两方面所面临的边界：哪些司法职能是人工智能难以替代的？在哪些价值维度上人工智能的应用可能引发正义与公平的困境？

（一）法理边界

从法理角度，尽管人工智能在要素式裁判中展现出上述优势，其应用仍受到事实复杂性、规范适配性和价值判断等方面的制约。首先，在事实层面，人工智能对复杂案件事实的理解力有限。要素式裁判依赖将纷繁案情拆解为固定要素，但现实中的案件事实千差万别，很多细节与背景超出既有要素库的覆盖范围。当前主流人工智能司法以海量数据驱动相关性分析，但样本的结构性缺失、潜在因素标记不足及低质数据大量存在，往往难以满足技术所需的数据充分性。[③]对比简单标准化案件，疑难、非常规案件中的事实模式可能超出智能系统的学习经验，导致提取要素时遗漏关键情节或误判次要事实的重要性。此外，人工智能在技术上也存在语义理解的障碍，如对文本情节的自然语义识别准确度不足，对复杂描述的细微差异难以把握，影响要素抽取的精度。

其次，在规范层面，人工智能难以完全适配法律适用中的弹性与多变。法律规范表面上确定清晰，但深层充满潜在冲突与模糊地带，需要法官运用专业解释予以澄清。要素式裁判虽然将法律要件形式化，有助于机械地比对事实与规则，但面对法律适用中的歧义、冲突及漏洞，形式逻辑本身无力解决。人工智能基于既有规则进行演绎推理，一旦遇到大前提（法律规范）的选择

① 周维栋：《生成式人工智能类案裁判的标准及价值边界》，《东方法学》2023年第3期。

② 苏晓宏：《人工智能生成裁判文书的模型构建》，《学习与探索》2025年第7期。

③ 赵泽睿：《法律如何计算？——赋予法律议论递归性的司法程序》，《求索》2025年第3期。

与解释问题，即难以自主判断适用何种规范以及如何解释。在司法实践中，同案不同判的成因之一在于地域司法经验差异和价值考量不同。智能系统若一味追求类案裁判的普遍化标准，通过深度学习构建统一的模型，可能忽视不同地域、时间背景下法律规范运作的细微差异，将丰富的司法经验简化为封闭独立的机械体系。因此，过度依赖概念同一性和归纳—演绎形式计算模型来理解司法审判，容易将复杂多变的案件情形简化为法律概念的组合迭代，剥离法官和律师的主观能动作用，使法律计算陷入形式悖论与循环论证的困境。①

具体而言，人工智能的算法模型固然能通过提取要素形成标准化裁判知识图谱，为法律适用提供一般性指引。但是法律规范的适用离不开目的解释、利益权衡等价值考量，如果机械套用统一要素标准而不顾个案特殊情境，可能违背立法本意，造成形式正义与实质正义的紧张。②尤其要注意的是，司法裁判中经常出现个案中的规范变通和例外，法官基于公平正义可能对一般规则作出限缩或扩张解释，这些创造性运用并未体现在历史数据中。人工智能缺乏理解规范弹性的直觉，难以及时适配新颁布的法律和最新司法解释，更无法像人类法官一样对法律条文进行创新性发展。换言之，在规范适配性方面，人工智能面临先天的知识局限和滞后性，如果缺少人工的监督与调整，可能导致机械司法甚至“算法歧视”等风险。

（二）价值边界

在价值层面，人工智能介入司法应当避免直接决定裁判结果，以免遮蔽具体案件中应有的个别价值考量。司法裁判既追求同案同判的规则平等，又必须维护个案公正。有鉴于此，司法裁判不能交由人工智能直接生成，人工智能也绝不能取代法官的价值判断与决断。即使在类案推送等辅助场景下，人工智能建议也只能作为参考供法官斟酌，而不能成为束缚法官判断的硬性约束。机器缺乏人类的情感和道德直觉，无法对案件所涉的善恶、责任、救济等价值问题作出实质性评判。审判过程中不可或缺的价值衡量和利益权衡需要法官运用自由裁量予以实现，既无法被算法精确模拟，更不应被技术手段取代。简言之，在裁判结论形成的核心环节，人工智能不宜也不能越界介入价值判断，否则将使裁判异化为缺乏温度与正义含量的“自动售货机”。③人工智能的价值应体现在促进裁判尺度的透明一致，而不是通过僵化复制先例结果来追求机械统一。通过类案要素分析，人工智能能够揭示类似案件在事实情节和法律要件上的共性关键点，推动同案同判向更透明的方向发展，为实现形式正义提供技术支撑。

然而，这并不意味着应对所有类案机械套用相同结论。绝对化的类案同判具有很大局限性，现实生活复杂多样，完全统一的规则难以周延，每个案件可能都有特殊因素，不加区分地追求结果一致，势必出现法律适用的僵化，反而损及实质司法公正。④而且，生成式人工智能模型由于训练数据偏差可能引入系统性误差，如无视这些误差而盲目复制先例结论，将导致同案不同值的不公。为防止类案适用过度僵化导致司法不公，法官必须运用个案裁量对人工智能推荐的类案进行审查，如果发现待决案件有特殊情由与既往案例存在实质性差异，就应果断不采取人工智能建议，作出符合同案同判本意的差异化裁判。

此外，过度依赖人工智能还可能产生机器权威偏误，即当法官迷信人工智能建议时，其自身偏见会与人工智能结论相互强化，形成“偏见回声”。法官因过度信赖机器仅进行表面修改、缺乏深入反思，也会弱化裁判应有的理性审查。⑤可见，若让人工智能直接决定或过度干预裁判结果，司法裁判的价值判断和理性对话空间将被压缩，不仅个别考量被遮蔽，还可能因缺少人类审慎而加剧不公。人工智能应当仅作为辅助手段提供多维信息参考和要素分析，提升裁判透明度和一致性，而不能让技术支配司法裁决。唯有人类在决策回路中保持主导地位，才能避免人工智能单向输出对司法公正的侵蚀，守住裁判的价值底线。

① 李学尧：《大语言模型应用中的司法偏误与认知干预》，《政治与法律》2025年第5期。
② 王禄生：《司法大数据与人工智能开发的技术障碍》，《中国法律评论》2018年第2期。
③ 蒋超：《法律算法化的可能与限度》，《现代法学》2022年第2期。
④ 刘韵：《智能要素式审判的本土实践及路径优化》，《法治现代化研究》2023年第5期。
⑤ 聂友伦：《人工智能司法的三重矛盾》，《浙江工商大学学报》2022年第3期。