法学研究 Chinese Journal of Law
Vol.42，No.1（2020）

人工智能与事实认定

栗　峥*

内容提要：人工智能与司法的深度融合，既体现在案件的法律适用环节也体现在事实认定环节，而事实认定是法律适用的前提。人工智能对案件事实认定的介入，需要将证据数据化、对数据进行运算整合、输出人可以理解的结论。在证据数据化环节，需要对证据进行结构化的数据改造，并克服语言障碍。在数据整合环节，人工智能主要以概率推理而不是因果推理作为逻辑推理方式，其算法也需要面对可计算性与复杂性两大难题。在结论输出环节，需要解决机器学习如何深化、信念如何建立与机器如何表达等难题。人工智能融入案件事实认定所面临的这些主要难题，可以尝试通过“小数据”训练，逐步构建人工智能“心智微结构”去慢慢攻克。

关键词：人工智能　事实认定　心智微结构　小数据

引　言

从世界范围看，人工智能正从四个层面由浅入深地介入司法。一是逐步代替人力从事法条查询、案例检索、案件分类、庭审录入等相对简单机械、重复性高的工作。二是作为助理，完成司法咨询、合同审核、证据获取、裁判文书辅助生成等辅助性任务。三是对海量裁判文书进行统计分析并作出预测，为法官裁判提供参考信息。四是直接参与决策或进行局部裁判，比如进行再犯风险评估、嫌疑人逃脱可能性判断、合理量刑测算等。2016 年，美国法院通过对“威斯康星州诉卢米斯”（Wisconsin v. Loomis）案的判决，正式承认了人工智能参与量刑裁判的合理性与正当性。〔1〕由于量刑本身就是裁判的一部分，这意味着目前

* 中国政法大学教授。
本文系国家社会科学基金重大项目“加强预防和化解社会矛盾机制建设研究：逻辑建构、现实考量和运行机制”（18VZL008）和国家社会科学基金重大项目“增强法治思维、运用法律手段领导和治理国家研究”（17VZL009）的阶段性成果。

〔1〕 State of Wisconsin v. Eric L. Loomis，2016 Wis. 68，881 N. W. 2d. 749（2016）.

的人工智能已经不仅仅充当一种辅助性工具，它对司法的影响正逐渐进入实质层面，试图介入司法的核心——裁判。

早在 20 世纪 70 年代，美国学者阿马托就提出了“人类法官能否以及应否被机器法官所取代”的问题。〔2〕该问题“在美国学术界一直争论不休，有时还上升至哲学层面讨论”。〔3〕在该问题上存在两种截然不同的立场：一种观点认为，人工智能会改变并代替法官裁判，机器裁判超越法官裁判的“奇点”必然会到来。〔4〕“当这一时刻到来之际，机器可能就会成为证人、陪审员和法官。”〔5〕另一种观点认为，人工智能不会影响裁判，人工智能法官纯属无稽之谈，辅助性始终是人工智能在司法中的根本定位。〔6〕除此之外的其他观点大多游移在这两者之间。针对上述争论，有学者评论认为，“理论界有关法律人工智能的研究虽然热闹，但仍处于开拓阶段，尚缺乏对法律人工智能运用现状与未来的深刻把握与思考”。〔7〕

上述问题可以拆解为三个层面。其一，人工智能对裁判的影响究竟是辅助性的还是主导性的。随着智能技术的迭代进化与机器决策能力的指数级提升，人工智能的司法应用范围会越来越大，专属于人类的裁判空间会被渐次压缩。如果人工智能对裁判的影响应限于辅助性影响，那么辅助的限度在哪里？人工智能究竟要在哪些层面或者哪些方面止步？其二，判断人工智能在司法活动中的地位和角色，有必要弄清目前人工智能在司法裁判中究竟能做到什么、做不到什么，哪些是人工智能的专长、哪些是它的瓶颈。只有找准人工智能司法应用中的具体动力和阻力，清晰辨别人工智能的“能”与“不能”，才能客观中肯地评价人工智能对司法的真实影响力。其三，人工智能在司法领域内的拓展不能仅靠立场表白与主观想象，它更需要论证与方案，更需要填充足够的原理与设计。当前切实需要的是，在人工智能当下所“不能”之处提供某种破解难题的建设性方案，尝试回答如何使“不能”变成“能”，以推进人工智能的司法应用能力，为智慧司法的发展提供动力。

要深入探讨人工智能对司法裁判的影响，裁判过程中的事实认定环节是一个较为理想的观察场域。这不仅是因为，在裁判过程中，事实认定是法律适用的基础，对于裁判结果的得出尤为关键，更重要的是，事实认定的过程系日常化的司法场景，相较于法律适用这种专业化场景，其较少受到知识壁垒的限制，具有开放性，易于同人工智能技术进行对接，也便于人工智能充分发挥作用。正因如此，国内外有关法律人工智能的研究，许多也是从司法裁判的事实认定维度切入，结合其中的某一具体步骤进行探讨，或预测人工智能的应

〔2〕 See Anthony D'Amato, *Can/Should Computers Replace Judges*, 11 Ga. L. Rev. 1277 (1977).

〔3〕 左卫民 《关于法律人工智能在中国运用前景的若干思考》，《清华法学》2018 年第 2 期，第 110 页。

〔4〕 参见马皑、宋业臻 《人工智能“法官”的一种实现路径及其理论思考》，《江苏行政学院学报》2019 年第 2 期，第 90 页。

〔5〕 See Andrea Roth, *Trial by Machine*, 104 Geo. L. J. 1304 (2016).

〔6〕 参见何帆 《我们离“阿尔法法官”还有多远》，《浙江人大》2017 年第 5 期，第 47 页。

〔7〕 前引〔3〕，左卫民文，第 112 页。

用前景，或提出建构性方案，或进行解构性批判。〔8〕本文亦选择聚焦人工智能与事实认定的融合关系，尝试为法律人工智能领域的诸多难题与困惑新增一个答案版本。沿着人工智能介入事实认定的三个主要步骤，即输入可供分析的数据、对数据进行整合生成、表达运算结果并完成输出，本文将依次讨论人工智能应用在各步骤各节点上的进展情况与亟待解决的难题，在此基础上，提出法律人工智能在现阶段及未来的努力方向。

一、证据数据化：结构化改造与语言难题

人工智能介入事实认定的第一步，是供给人工智能可分析的数字化信息，即将全案信息转化为数据。然而，案件事实基础是证据，证据是实体而非数据。证据数据化是人工智能介入事实认定的先决条件。

（一）证据如何变数据

就人类认知而言，我们可以直接感知并理解实体证据，无论它以怎样的形态呈现。人类的感官系统、思维系统、表意系统等为接受各类证据提供了工具，使我们抓取证明信息不受证据样态的限制。但对人工智能而言，证据形态将成为信息输入的第一障碍。计算机如何理解一张照片的含义或一段视频的内容？“一把带血的刀子”如何进入电脑，使电脑意识到这把刀子的证明信息及价值？解决这一问题须深入证据结构之中。

现实世界里的证据都是“块结构”的，即不可拆分的完整一块。一个证据就是“一个整块”。比如，“一把带血的刀子”作为一个独立的证据，不能拆分成“血液”“刀子”。当然，该证据包含丰富的证明信息，如被害人的血液、加害人的指纹、加害人用刀杀害被害人的因果关联等；该证据与其他证据关联还会产生新的信息，如被害人伤口与刀刃的吻合情况。一个证据究竟包含多少证明信息、哪些对裁判有效、哪些无效或应予排除，依赖裁判者的自我感知与主观判断，不同的人会作出不同的考量和取舍。由于内含信息的不确定与变动性，该证据从证明内容上并不能被随意拆分，出于任何理由的拆分都只是基于个体化的认知，不具有普适性。最好的方法是将这一整块证据展示到裁判者面前，由其自由心证。由于裁判者无需解释是否采信、采信多少，只要该证据曾呈现于裁判者面前，就视为

〔8〕这一领域的代表性成果，参见魏斌、郑志峰《刑事案件事实认定的人工智能方法》，《刑事技术》2018年第6期；罗维鹏《人工智能裁判的问题归纳与前瞻》，《国家检察官学院学报》2018年第5期；李飞《人工智能与司法的裁判及解释》，《法律科学》2018年第5期；纵博《人工智能在刑事证据判断中的运用问题探析》，《法律科学》2019年第1期；赵艳红《人工智能在刑事证明标准判断中的运用问题探讨》，《上海交通大学学报（哲学社会科学版）》2019年第1期；刘品新、陈丽《数据化的统一证据标准》，《国家检察官学院学报》2019年第2期；周慕涵《证明力评判方式新论——基于算法的视角》，《法律科学》2020年第1期；吴习彧《司法裁判人工智能化的可能性及问题》，《浙江社会科学》2017年第4期；A. Valente, J. Breuker & P. Brouwer, *Legal Modelling and Automated Reasoning with ON-LINE*, 51 International Journal of Human-Computer Studies 1079 – 1126 (1999); Michael Riesen & Gursel Serpen, *Validation of a Bayesian Belief Network Representation for Posterior Probability Calculations on National Crime Victimization Survey*, 16 Artificial Intelligence and Law 245 – 276 (2008); H. Prakken, *An Abstract Framework for Argumentation with Structured Arguments*, 1 Argument & Computer 93 – 124 (2010); M. Abraham, D. Gabbay & U. Schild, *Contrary to Time Conditionals in Talmudic Logic*, 20 Artificial Intelligence and Law 145 – 179 (2012); Mark A. Lemley & Bryan Casey, *Remedies for Robots*, 86 University of Chicago Law Review 1311 – 1396 (2019); Ashley Deeks, *The Judicial Demand for Explainable Artificial Intelligence*, 119 Columbia Law Review 1829 – 1850 (2019).

裁判者已经穷尽了对该证据的全面思考，充分吸收了它的证明价值于裁判之中。可见，证据的“块结构”既符合物理形态上的视觉空间要求，又满足内在证明信息上的完整性标准，更有利于遮蔽司法裁判上的种种被动与不确定。

但是，这把刀子摆在计算机面前并不会产生上述效果。现阶段，计算机尚无法直接理解“块结构”的证据，只能运用计算机可理解的语言将“块结构”击破打散，形成数字化、可识别的结构性数据，以获得其中的有效信息。计算机可理解的语言一般有三层：以 0 和 1 为基础的二进制机器语言，可识别并运行机器指令或操作指令的汇编语言，以及作为高级编程语言的 C 语言。无论哪一层语言形态，以怎样的语法指令和程序运行，其实质都是计算。为了完成计算，证据必须数字化。

对此，人工智能的解决路径是运用智能识别技术。这项技术在人脸识别、指纹识别等领域的应用已日趋成熟，其基本技术要素包括：基点检测、关键点定位、特征提取、向量集成、相似度排序、图形反馈与矫正等。以人脸识别为例，智能识别技术的主要原理在于特征比对，即在获取海量对比图片数据的共同规律的基础上，设立若干面部基准焦点，将这些相对不变的“锚点”（如鼻尖）作为勾勒的框架固定值，然后寻找关键点周边的像素深浅变化，用从高到低的向量箭头取代以形成梯度分布，再对向量进行集成，就可以将人脸图像转换为结构表达式，最后与预先存储的人脸特征数据库进行相似度排序，得出与输入图像相似度最高的反馈结果。

虽然智能识别技术在人脸识别中相当奏效，〔9〕但它目前在司法领域中的应用仍相当有限。其一，这种识别需要基于海量数据，从千万图像中寻找下一张，其参照并不唯一，而司法案件中的每一个证据都专属于该案，独特且唯一。其二，基于人脸要素与构成的相似性，可以提炼出足够多的关键点作为识别基准，其规律较易获取与归纳。司法案件却没有如此高度的相似性，即使同类案件，其证据也千差万别，难以形成固定的关键点和有价值的向量梯度分布。其三，更为重要的是，即使识别出人脸，如张三，仅表示检测图像与对比图像具有一致性，并不代表电脑理解张三是谁。这仅是对形式外观的识别，并不涉及对内容的理解，它与一个人认出了张三，意义完全不同。司法证明需要的，是能够探知并理解证据内容与涵义的智能方式。

证据证明的核心是证明力，证明力作用于裁判者的内心，转化为影响心证的力量。如果能将每一个证据转化为指向事实证成、能为心证作出贡献的某种证明概率，那么证据就可以实现数据化。比如，对于一把带血的刀子，根据其沾染的被害人血迹、刀把上犯罪嫌疑人的指纹、两者的因果关联性等信息，裁判者可以在综合考量后得出该证据用以证明有罪事实的某一评估值，即它能证成最终事实的贡献度。这个值可以以百分比计，其既反映该证据的证明力，也代表裁判者的认可度。由此，证据被转化为影响裁判心证的数字，以此方式，所有证据都可以转化为一系列概率值。〔10〕将这些数值输入计算机，采用一套科学算法（如贝叶斯决策）累积求和，便能获得对案件事实的整体评价值。将评价值与证明标准相比较，即可得出相应裁判。

〔9〕 See Yann LeCun, Yoshua Bengio & Geoffrey Hinton, *Deep Learning*, 521 Nature 436 – 444 (2015).

〔10〕 参见 [匈] 拉卡托斯 《数学、科学和认识论》，林夏水等译，商务印书馆 2010 年版，第 237 页。

人类通常可以对单个证据作出较为中肯的评价，面对众多复杂证据时却很难精准驾驭，人脑难以并行思考多元证据间的叠加交融，此时，将证据转化为数据交由人工智能来完成，可兼具科学性与高效性。

（二）数据的结构化改造

证据变数据的前提是对传统证据进行符合计算机认知模式的结构改造。传统证据难以被计算机分析，因为目前人工智能并没有具备人类如此多元而强大的感知力，证据必须被转化为电脑可识别、可感知的格式。这种将证据转化为格式化数据的方法是建立法律人工智能的关键操作，可以称为“结构化改造”。

数据按格式可分为结构化数据与非结构化数据。结构化数据是以二维逻辑表达的数据。它符合数据格式与长度规范，提供关系型数据认知的基础结构。现有技术已开发出访问结构化数据的通用语言——SQL 语言，可用于数据查询、检索、合并、分析、处理等。绝大部分数据，包括文本、文档等半结构化数据和声音、图片、视频等多媒体非结构化数据，都达不到模范格式的要求。就事实认定而言，证据主要是以文本、实物、声音、图片和视频等形式存在。虽然这些格式及其转换格式可以输入计算机，但目前的技术仅实现了数据的存储，难以对数据进行破解与分析，故证据无法被直接利用。人工智能需挖掘原始数据，解构非结构化数据，从中提取特征，将其重构为结构化数据。

结构化改造的有效路径是模拟人脑处理信息的方式。人脑并非直接对所感知的事物和信息予以处理，而是对其进行特征提取，在接收到刺激信息后，经由一个复杂层状网络结构来处理信息，识别并认知事物。人的感知系统与大脑处理系统具有可以有效过滤并加工信息的层次结构，使人类得以精准把握事物本质特征。人工智能同样需建立一种“多层神经网络”，〔11〕仿制人脑神经网络，自动抓取事物特征，通过组合低层特征，生成逐层抽象的高阶特征，以形成表征数据的分布式表达，进而逐步逼近人类认知证据的智能水平。

证据是特征的集合。特征对结果的影响很大，抽取怎样的特征决定形成怎样的认知。〔12〕现阶段人工智能技术对特征的提炼一般通过人工来完成，其效率相对低下，受主观影响较大，并且难以应对复杂情形。而“多层神经网络”可通过组合低层特征、提炼高层特征从而自动选择特征，无需人工辅助。〔13〕其原理为：假如有一个系统 E，它有多个层次（E_1……E_n），它的输入是 F，输出是 G，即 F→E_1→E_2→……→E_n→G，如果输入 F 等于输出 G，即输入 F 经过这个系统变化之后没有任何的信息减损，则意味着每一层 Ei 都保持了全部信息，都可以作为原有信息的特征表达。现假设有一个证据输入 F（如一个图片证据或文本证据），并且设计了一个证明系统 E，通过调整系统中的参数，使得它的输出仍然是 F，那么就可以自动获取输入 F 的一系列特征，即 E_1……E_n。这一思路的用意在于叠列多层级结构，使上一级的输出成为下一级的输入，进而逐层拆解证据特征，直至一个证据的全部有效证明特征被逐级分解并分级表达，由此确保证据证明信息提取的完整性和准确性。通

〔11〕 See Davide Castelvecchi, *Can We Open the Black Box of AI?*, 538 Nature 20 – 23 (2016).

〔12〕 See Paul de Laat, *Big Data and Algorithmic Decision-making: Can Transparency Restore Accountability?*, 47 ACM SIGCAS Computers and Society 44 (2017).

〔13〕 See United States (2016) Executive Office of the President, *Preparing for the Future of Artificial Intelligence*, Technical Report, National Science and Technology Council, Washington D. C. 20502, Oct. 2016.

过这种方式，证据以一种可被计算机识别的形式获得了人工智能的“认知”。

（三）语言难题

现阶段，上述结构化改造方式会遭遇语言难题，因为最难提取特征的证据形态正是语言。一把带血的刀子，可以拆解出血迹、指纹、刀痕等清晰明了的特征数据，但一段文字或话语，由于语义的多元与阐释的差异，可生成诸多表意版本。“人类特有的建模工具就是语言”，〔14〕语言对应着丰富的世界，蕴含着人类最精湛的智慧，如何将语言的奥妙传达给人工智能，这是一个很大的难题。无论是证人证言、当事人陈述等言词证据，还是口供笔录、文书合同、勘验笔录、鉴定意见等文本证据，司法证明中最普遍的证据形态恰恰是语言。人工智能想要在事实认定上取得进步，就必须对此有所突破。

目前，有关语音识别和文字翻译的技术已日趋成熟，但是，机器对语言的识别并不意味着对语义的理解。例如，人工智能对语音的识别依赖于参照系的丰富与比对的准确。因字词的发音是固定的，发音的声调、平仄、节奏、韵律等均有规律可循。将字词对应的发音输入电脑，电脑只需将被检测发音与参照发音进行比对。这种比对具有较强的机械性，比对频率越高，各种发音训练越多，电脑的识别力就越强。但显然，这种智能目前仍处于统计学意义上大概率、机械性比对的熟练化层面。就事实认定而言，在证人说出“我看见一把带血的刀子”时，我们需要计算机能够认知并理解“血”和“刀子”背后的含义，以及这两个概念对于“杀人”的意义。

为此，人工智能需要掌握语义和语法。语义是一种高度浓缩的文化概念，传达的是某种约定俗成的信息。当人们提及“血”或“刀子”时，不用再去对其进行定义、解释和描述，它承载着人与人认知交往的共识，是在长期共同生活中形成的自然提炼。〔15〕机器并不能获得这种实践性积累，机器的学习脱离了人类生活实践与交往的具体情境，其学习的基础主要是大数据。语义的上一层次是主题。主题是一段自然语言潜在的中心，是一个语义组合的灵魂，但它不一定直白地体现在字词上。比如，我看见“甲拿着一把带血的刀子”，其主题在于描述证明“犯罪工具”。“犯罪工具”并没有出现在字句中，机器能否通过字面意思关联到这句证言所暗指的犯罪要素，这关系到语义理解的有效性，决定着人工智能能否在事实认定中发挥实质作用。主题的上一层次是倾向。语义隐含着倾向，褒或者贬，利于控方还是辩方。倾向包含“正面”“中性”“负面”三类，尤其在司法程序中，证据被置于控辩审三方张力之中，人工智能能否经由语义与主题的探知达至倾向的意会，这关系到人工智能介入司法的能力。除了语义层次，人工智能还需解构句法结构。句法结构的基本单位是词语，上一层次是构词法，再上一层次是词性，然后是语法，最后由多个句子组成篇章。相应地，人工智能的任务对应由词根和词形还原分析构词法，再由词性标注获得词性，最后由句法分析获得句法结构，以完成句法结构解析。

面对上述难题，将人工智能应用于事实认定的设计方案需完成如下任务：第一，搭建语料库。丰富的语言材料是提升深度学习能力的基础。首先需要收集汇总足够多的案例，

〔14〕［意］苏珊·彼得里利、奥古斯托·蓬齐奥《打开边界的符号学：穿越符号开放网络的解释路径》，王永祥等译，译林出版社 2015 年版，第 401 页。

〔15〕参见［意］多梅尼科·帕里西《机器人的未来：机器人科学的人类隐喻》，王志欣等译，机械工业出版社 2016 年版，第 132 页以下。

将海量案例中的语言证据（无论书面还是口语）集合入库，形成可供比对参照的“学习池”。第二，对语料库进行结构化处理，将完整的篇章、段落信息解构为结构化信息，即将原始文本语料转化为人工智能可识别的矩阵数据构造。其转换工具可依赖“向量表示”。〔16〕“向量表示”是一种表义字词的数字化方法，它把每一个词表达为一个很长的向量，向量的维度达到词典大小，其中绝大多数维度的值为0，只有一个维度的值为1，这个维度代表了该词，比如“血”表示为［1，0，0，0……0］，“刀子”表示为［0，1，0，0……0］。采用稀疏方式存储记忆，相当于给每个词配备了一个专属ID。通过向量表示，可以形成词条排列关系矩阵。这样一句或一段话语就可以形成一种结构化的矩阵模型。第三，向量距离计算。基于语义间的远近，可以运用相似度公式与空间向量线性代数来计算语词间的关联情况，从而为计算机提供语义判断上的参考，即数据结构分析。

需要指出的是，即使完满完成上述工序，也很难确保人工智能对语词的理解就是人类创造该词汇时赋予其的本义。尤其是，当若干语词组合成一句或一段完整意义时，更难保证人工智能的认知与人类的理解一致。因为，即使在人类之间，针对同一段话语也经常有不同的解读。

二、数据整合：推理与算法

在实现数据输入后，下一步要解决的问题是，以何种方式组织这些数据使其得出结论。目前主要有两种进程：一种是基于逻辑学，以推理作为动力推进智能化；另一种是基于数学，以算法作为动力推进智能化。两者各具特色，各有优劣。

（一）推理：因果与概率

事实认定的精髓并不是识别与认知证据，而是组合证据以生成事实。人工智能迈向事实认定的高阶，离不开对证据间关系链条的有效处理——推理。司法证明过程包含多种推理形式，但面对众多证据，最关键的莫过于寻找彼此的因果联系，建立符合常识认知的合理事实叙事，即因果推理。因果推理一直是事实求证中组织编排证据、构建事实图景的核心逻辑。〔17〕

因果解释是人类把握世界的根本方式。人脑在一堆错综复杂的证据谜团中厘清思路与顺序，依赖的正是因果推理的线性方式。〔18〕人脑无法直接感知到全部证据并一揽子地平行理解多重线索。一般而言，它主要是在接收证据的原始信息之后，通过对比、修正、协调、排列事件探寻时间轴上的某种规律，努力形成自身可理解的故事版本。事实的复杂性和证据的多样性均要求人们寻求某种可解释的思考轨迹，将证据一个接一个地附加认知关联，进而捋清埋于事实深处的线路。在这个过程中，每一个证据都要被追问背后的原因，每一种关联都被用以形成推断式的因果链条。因果流不断地从此证据涌向彼证据，将分散多元

〔16〕 参见前引〔9〕，LeCun等文，第440页以下。

〔17〕 参见［德］马克斯·韦伯：《社会科学方法论》，李秋零、田薇译，中国人民大学出版社1999年版，第83页以下。

〔18〕 See Michael Rescorla, *The Computational Theory of Mind*, https://plato.stanford.edu/archives/spr2017/entries/computational-mind/, 2019年12月31日最后访问。

的局部片断逐步整合成一个连贯的整体。当证据 A 与证据 B 之间能够建立起某种因果联系，证据 A 与证据 B 的证明信息即获得充分提取，并提升为单独的证据 A 或 B 所不具备的复合推断。复合推断驱动着事实论断的前进，并持续对感觉、判断、行为与意识进行合理化解释。因与果作为人们融入案件事实场景的捷径，帮助裁判者个体感知被告的行为逻辑与行事动因，以此构筑合理化的心理过程。因果关系串连出事实认定的连贯论证秩序，并建构出对案件的个体叙事。通过因果推理，我们得以以一种连续性的路径轨迹去完整地看待事实，可以取舍并重建过去图景，再现关于被告动机、行为以及场景的全景叙述。因果关系的内在连续性确保了因果推理可以作为缝合证据、焊接事实的有效工具，确保人们能够把诸多证据通过意义编织的形式加以认识，并使认定的事实保持协调一致，促使人脑生成统一结论。

目前，人脑完成上述因果推理的心证历程，在人工智能中还难以实现。人工智能并不依赖因果，确切地说，它无法理解因果，〔19〕它的推理依靠概率。人工智能的本质是计算，计算关系建立的是 A 与 B 之间的数量关联，即被量化的两个数据值之间的数理关系。这种关联实质上反映了相关性的概率强弱：当数据值 A 增减，如果数据值 B 也随之明显变化，则意味着相关强概率；当数据值 A 增减，如果数据值 B 发生很小变化，则视为相关弱概率。当然，可以用复杂多样的算法或公式来表示数值间的多变关联，但无论以怎样的计算模式呈现，计算机表达元素间逻辑的本质依旧是某种概率推理。

概率推理是完全不同于因果推理的逻辑形式，即使获得相同结论，概率推理也体现出截然不同的分析路径。它知道“是什么”，但不知道“为什么”。比如，在“A 与 B 争吵”与“A 杀害了 B”之间，人们可以轻易地建立起因果关联：A 因争吵产生愤怒而杀害了 B。但对于机器而言，它无法体验人类争吵中的愤怒情绪，无从理解“愤怒”，也难以用“愤怒”来连接因与果。人工智能主要是从大量案例数据中发现，当“争吵”增加时，“加害”概率也会增加，“争吵”与“加害”之间存在某种概率关系，由此建立两者的联系。

在很多领域中，知道“是什么”就足够了，并不需要知道“为什么”。〔20〕但是，司法审判是精致操作，不仅需要“知其然”还必须“知其所以然”。裁判的精华是裁判理由而不是裁判结论。同时，裁判也不能仅作为预测性判断，我们不能在某种概率性关联得出之后就终止思考，相反，我们仍要继续思考，直至得出适合于案件具体情境的甚至是唯一的结论。

（二）算法

人工智能的核心智慧依赖的是算法。算法是在限定条件下以运算方式将输入转换成输出的问题解决机制，体现为机械化的运算过程。算法将问题情境转换为限制条件，将问题要点抽象为计算变量，将整个问题切换为数学模型，通过公式化运算求解答案。算法的优势在于：一方面把问题模型化，提炼出普遍规律与一般特征；另一方面将求解科学化，以可验证的数学原理保障证明的严肃性。何种算法适用于事实认定？目前，虽已有正在设计或应用的多种算法试图解决此问题，但仍然需要人工智能专家与法学工作者共同合作，去

〔19〕 See Jason Millar & Ian Kerr, *Delegation, Relinquishment and Responsibility: The Prospect of Expert Robots*, 70 SSRN Electronic Journal 532 – 534 (2013).

〔20〕 See Ryan Calo, *Artificial Intelligence Policy: A Primer and Roadmap*, 51 U. C. Davis L. Rev. 414 (2017).

开发专属于该领域的算法。在这之前，只能借助通用人工智能来尝试解决事实认定问题。

一种可行的进路是使用“贝叶斯决策”搭建“拟人思考”的决策框架。贝叶斯决策建立在概率基础上，它描述先验概率演化成后验概率的轨迹。以被告口供证据 E 为例，在被告供述前，法官会预先形成对被告口供真假性的事先估计，结合对被告历史、态度、行为、状态等方面的综合印象，判断被告说真话的可能性，此即先验概率，记为 P（H）。法官在听取被告供述后产生的对口供真实性的判断，即后验概率，记为 P（H | E）。法官根据自身经验与认知，权衡口供 E 真假两种可能的概率：当被告供述为真，得到 E 的概率，记为 P（E | H）；当被告供述为假，得到 E 的概率，记为 P（E | －H）；然后将两者作概率比，即获得真假比系数，后验概率就是先验概率乘以这个系数，〔21〕即：

$$P(H/E) = P(H)\frac{P(E/H)}{P(E/-H)}$$

这个贝叶斯决策公式表达了概率间的相互关系，描述了证据 E 出现后，先验概率如何被修改调整生成新的后验概率。由于证据 E 的注入，先前的判断得到加强或弱化，融合新证据 E 后的后验概率取代了先验概率，决策获得了进化。贝叶斯决策能够吸收新信息，并把对新信息的判断转化融入后验概率，实现微观上的决策推进。从理论上看，人工智能完全可以借助贝叶斯决策完成对每一个证据的吸收与积累，进而得出对事实的整体评价。

在贝叶斯决策的基础上发展出的序贯决策方法，可以进一步优化人工智能的证据判断。计算机在进行决策后会面临新信息的注入，新信息的注入又会产生新决策，接着又出现新信息，形成新决策，如此反复，形成一个序列。在这个过程中，信息常以随机或不确定的动态形式呈现，每次决策的下一步都不确定。在未出现新信息前，决策总是最优的，但新信息能够将决策更新，直至程序终止。序贯决策方法通过这种动态调适机制，保持着决策的最优化与灵活化。法官需接收来自控辩双方的证据与信息，其中的不确定性与不可预期性完全符合序贯决策的理论情境。因此，序贯决策方法能够极大助力人工智能对法官裁判的学习与模仿。以算法为起点，以贝叶斯决策及发展出的序贯决策方法为路径，两者相结合就形成了可适应多种环境的通用人工智能。

从原理上说，运用通用人工智能在事实认定中建立人工智能框架具有一定的可行性。但在现阶段，实际设计与实践操作可能会存在两大难题。

其一，可计算性难题。通用人工智能作为一种思路表达式，往往不可实际计算，只具描述性，不可用以求证。即使通用人工智能转化为某种可计算的变种形式，它的论证能力会在多大范围内适应事实与法律的需求仍属未知。可计算是人工智能通往司法的基石，但事实与法律并不具备可计算基础。司法裁判的有序性本质并非计算出来的。算法的广泛应用需要将大量司法案件所处的情境进行充分地数字模型化改造。然而，数字模型化的代价是忽略复杂干扰项，将问题理想化，以诸多假定来框定范围，将问题提纯到固定模态上加以演算。此时，高度提炼的模型很可能已与实际情境相去甚远。而且，我们很难判定哪些因素足够重要或不重要，任何省略的因素都可能是影响问题判断的关键。

其二，复杂性难题。复杂性表征解决问题的困难程度。人具备高度的应变力与灵活度，

〔21〕 参见栗峥 《司法证明的逻辑》，中国人民公安大学出版社 2012 年版，第 82 页以下。

在遇到困难时，并不需要区分比较难或者很难的差异等级。对于机器，难易是分等级的。应对复杂问题需要更繁琐的公式与更深涩的理论，有时这样的公式或理论甚至还没有出现。〔22〕问题的难度每增加一级，算法需求可能增加几倍。人类法官面对一堆不知如何着手的证据时，可以始终“保持清醒，从常识与经验出发，走一步看一步”。但对于算法，从实物证据到口供再到其他证据，无论是动机、行为还是结果，要对每一步设计出妥当的程序相当困难。因为预料不及的情况实在太多：案件没有实物证据；动机无法探知，只是猜测；行为缺乏证据支撑，等等。算法是依照菜谱对齐备的食料与配料进行加工，无法根据现有食材即兴烹饪。现实世界不仅不可轻易计算，而且它远比算法模型复杂。肖恩·莱格指出：“对于任何包含初等数论的形式系统 T，存在某个复杂性水平 c，对于任何高于 c 的复杂性水平 n，形式系统 T 都无法帮助我们找到可以逼近任何复杂性不超过 n 的环境的通用模型，尽管这种模型是确实存在的。不严格地说，强大的智能体必然复杂，复杂且强大的智能体是存在的，但只要它足够复杂，形式系统将无法帮助我们找到它。”〔23〕也就是说，目前通用人工智能及其各类算法虽然为人工智能的发展提供了工具，但也设置了上限。面对复杂性水平为 n 的问题，一种可计算的通用算法模型是存在的，但其算法本身的复杂性不会小于 n。我们可以设计足够强大的智能去应对足够复杂的现象，但这种智能系统也会足够复杂，以至于应对复杂智能系统一点不比应对复杂现象轻松，甚至我们能够发现足够复杂的现象，却不一定能够找到或建立足够复杂的系统。〔24〕

三、智能的深化与输出：学习、信念与表达

完成智能分析的基础框架后，人工智能仍需不断进化才能实现真正意义上的应用，这一过程主要通过深度学习与信念建构来推进。最终，人工智能需要依赖机器表达实现智能的输出，以完成全部智能过程。

（一）机器如何深度学习

机器的智能晋级主要依赖“深度学习”，〔25〕学习的深度主导着智能的程度。目前，在深度学习技术尚未有效落地的司法领域，“深度学习”的修饰性远远大于其应用性。我们总说“人工智能通过深度学习可以实现”，但其实并不清楚如何通过深度学习实现人工智能，也不确定通过深度学习能否实现。〔26〕确切地说，我们并不了解深度学习的“深度”，以及这种“深度”所能解决问题的“深度”。

深度学习需要建构复杂的数字模型来帮助机器仿效人的思考轨迹。数学建模是进行机器学习的基础。建模有两种路径：人工设计与自主训练。人工设计将通过人类思维机制的“临摹”提炼抽象化的认知模式赋予机器，使其具备人类思考的诸多特征，比如决策树模型以人脑决策时形成的树结构为机理，形成枝叶清晰、归属分明的决策图式。例如，将证据

〔22〕 See Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 Fordham L. Rev. 1138 (2018).

〔23〕 Shane Legg, *Machine Super Intelligence*, PhD thesis, Department of Informatics, University of Lugano, p. 105 (2008).

〔24〕 前引〔11〕，Castelvecchi 文，第 20 页以下。

〔25〕 See Jürgen Schmidhuber, *Deep Learning in Neural Networks: An Overview*, 61 Neural Networks 85 – 86 (2015).

〔26〕 参见前引〔11〕，Castelvecchi 文，第 20 页以下。

基于“类别”的根结点分为言词证据与实物证据，将言词证据基于“来源”这一分支结点分支出证人证言、被告人供述与被害人陈述，将证人证言基于“方向”这一叶结点分支出趋向证明有罪的证人证言与趋向证明无罪的证人证言等。如此，通过类别、来源、方向三个维度判断某一证据的归属，进而形成识别认知。决策树采用“分而治之”策略，完成思维上的递归，它能够在短时间内抓取大型数据源的有效数据并进行分类，如果给定一个观察模型，它还可以很容易地推导出相应的逻辑表达式。目前，人工智能已发展出数以百计的类似数学模型，它们极大地提升甚至实现了局部智能，使学习逐步迈向深度。

数学模型作为一种对现实的高度抽象，其适用上的有效性依赖大量严格的前提条件的保障，也需要对诸多影响性因素有意忽略。在满足条件与排除干扰的情况下，它的机械化运算的优势可以发挥得淋漓尽致。这一点，在棋类比赛中被充分放大。〔27〕棋类是一种脱离现实的典型思维游戏，其步法规则通过数条机器指令即可实现，并且游戏策略具有高度的模仿性与复制性，上一步“下法”与下一步“下法”保持着可供计算机运算的某种必然性联系。说到底，棋类本身就是一种形象化的数学模型，每种棋盘布局都是一张典型的拓扑数学结构图，棋法本质上就是算法，将棋法策略切换成公式运算，不但没有任何障碍与顾虑，且能够达到完满程度上的统一。计算机超越人脑的机械化运算的能力可以产生强大的“试错”概率以权衡效果，人脑却难以精准地复现“下一步棋”的足够多的可能性。在“下棋”方面，人工智能超越人类具备原理上的天然优势。〔28〕但是，一旦被投放到案件事实认定的具体情境中，棋类的上述优势便很难发挥。案件事实的情境与条件复杂多样且不断变动，难以满足数学模型严格的标准，且建模所排除或忽略掉的诸多干扰因素很可能恰是问题的核心。

相比于人工设计，强调自主训练法的无监督学习直接借助大数据，由机器自主认知、自我巩固，最后自行归纳来进行学习。设想有一批照片证据，我们将图片数据输入无监督学习的模型中自主训练算法，使计算机试图理解图像证据，由机器自行识别出各个证据的信息及特征。实现这一点至少需要两项保障：其一，自主训练要类似于人类的学习方式。人类认知事物首先是基于生存与生活的实际需要。例如，当父母指着一只小猫告诉孩子这是“猫”时，孩子是在对真实猫的观察以及与猫的相处中，通过发现、观察、互动与对比，调整对“猫”的理解。这种认知包含了复杂的生存交往、行为方式、情感喜好等，可以说，孩子理解“猫”就是理解世界的一部分。但是，机器并不存在生存与生活上的认知前提，也无理解世界的任务与需求。机器只能单纯理解“猫”，或者说为了理解“猫”而理解“猫”。为此，计算机需要通过千万张图片记住“猫”的各种形态与样貌，我们需要足够数量的数据“喂”给电脑，告诉它什么是“猫”，这个过程需要耗费极大的人力、物力、财力以及时间成本。其二，输入的数据还须经由过滤加工等工序，以形成机器学习可以消化的材料。自主训练的好坏取决于用于训练的数据的好坏。如果是大量未经甄别、良莠不齐的数据，机器自主学习的效果也会受到影响。高质量的智能依赖高质量的“教育”与“培养”，这一点对于人工智能也不例外。

〔27〕 参见姚海鹏、王露瑶、刘韵洁 《大数据与人工智能导论》，人民邮电出版社 2017 年版，第 127 页。

〔28〕 前引〔20〕，Calo 文，第 432 页。

无论是人工设计还是自主训练，即便从某种程度上实现了深度学习，也仍然只是深度学习中的基础学习，因为这类学习主要集中在“判别式学习”上。学习模型分为“判别式学习”与“生成式学习”。〔29〕判别式学习是指通过某种模拟的高维感官输入映射出一个类别标签结果。但我们期待的并不是被人工智能告知某一客观事物的名称，而是帮助我们发现我们发现不了或还没发现的世界的内在结构、案件事实的内在规律，并在此基础上进行推理与创新。这种能够反映现象内在特征与规律并生成全新的自主模型的“生成式学习”，才是深度学习的深层阶段。

生成式学习在推进最大化自主智能的同时，也为当下实践带来挑战：一是事实建模需要大量先验知识积累，若输入的质量不高，其生成的结论就有可能偏离常识；二是真实案件的证据复杂多样，拟合模型所需的计算量呈几何倍数增长。生成式学习主要依靠生成对抗网络（generative adversarial network，GAN）技术支持，〔30〕它采用“左右互搏”的原理，通过对抗训练机制进行练习，形成生成器与判别器，依赖判别器的提升激发生成器，依赖生成器的结果刺激判别器，使两者“搭梯子式”地相互促进，从而大大提高应用效果。目前，这种方法已在医学影像识别、图像处理、语言处理中得到一定程度的应用，但在司法领域尚处于理论构想期，有待寻求应用上的对接路径。〔31〕

（二）机器如何建立信念

司法需要信念，裁判就是裁判者信念的表达与体现。事实认定需要形成对“真实”的信念，人工智能在完成识别、认知、理解、推理、学习等环节后，最终也要输出一种对事实真实的判断，这种判断需要达到与法官同样程度的“排除合理怀疑”或“内心确信”。

人类的司法裁判是大脑与心灵的结合。大脑产生裁决，心灵相信裁决，事实认定基于脑与心的统一。每一项证据与每一个证明环节均出自人脑的智能并伴随着内心的信念，证据与事实的判断杂糅了推理演算的脑力劳动与拿捏权衡的心理活动，两者彼此牵涉、并行而进。但是，机器并没有这两个器官，人工智能对事实认定的“信念”究竟是要复制大脑还是要造就心灵？这是两种截然不同的路径，复制大脑依赖算法式，造就心灵依赖启发式。

没有推演出结论，就没有相信结论的基础。人工智能建立信念首先需要算法式，也就是通过计算获得信念。但问题在于，“内心确信”或“排除合理怀疑”，是一种难以言说与解释的状态。“一个人相信一件事”，通常是不可测量的，但人工智能不得不对信念有所测度。〔32〕

要运用算法式计算信念，前提是信念可赋值。在完全相信与完全不相信之间建立测量区间，依赖于主观概率。完全相信视为 100% 概率，完全不相信视为 0% 概率，两者之间存在无限程度变化的相信度，即信念值。这样一来，虽然仍存在因人而异的差别，但所有事实证明所凭借的证明标准均可以对应出大致客观的某一等级区间的信念值，比如“内心确信”可以设定为 90% 左右的相信度。这为计算机获得结论后的信念判断提供了可行的比对参照标准：满足 90% 即为“相信”，达不到 90% 归于“不相信”。由此，所有算法的努力都汇集在得出这一概率值的大小上。但是，这种操作仅仅是为了满足算法需要的纯技术路线，

〔29〕 李德毅主编 《人工智能导论》，中国科学技术出版社 2018 年版，第 126 页。

〔30〕 同上。

〔31〕 See Harry Surden，*Machine Learning and Law*，89 Wash. L. Rev. 87（2014）.

〔32〕 参见前引〔18〕，Rescorla 文。

达到90%视为相信，并非如人类来自意志的自主自觉的坚定决心与真诚信念一般，而仅属程序化的输出方式，它并没有经历心灵权衡的过滤。计算机不会理解90%的真正意义，也没有形成这一程度上的信仰。

进一步而言，对事实的信念不是一蹴而就的，最终信念是由无数个小信念聚合而成的。在证明过程中，每一个证据与证明环节都需要裁判者注入相信与不相信的某种心理判断，这意味着人工智能必须实现信念的微分化拆解，对各个证据赋予信念值，并且能够累积加和所有值，生成信念总值。从技术层面看，计算各个证据与各证明环节信念值依赖复杂的算法，案件难度与复杂性也会使信念算法难度以指数倍数增加。就目前而言，简单案件的算法通过针对主观概率累加的贝叶斯决策等组合可以胜任，但对于复杂疑难案件，算法设计将面临难度几何倍数以上的复杂性难题。同时，算法虽然可以解决运算过程，却仍然需要面对初始值的确定问题。一个证据的初始信念值是裁判者对这个证据的原初印象分。这个分值基于裁判者个体阅历、经验、常识、认知等背景自然而然地形成。它是人通过该证据对世界的看法，或者说是人与物的交往方式。人工智能难以做到这一点。为了保持中立与客观，机器设定的初始信息值必然是50%，而这显然从一开始就偏离了裁判的实质。

可见，算法式路径通过概率测度实现了信念的数值化，使其能够成为人工智能的计算基础。然而，纯粹对应于概率之后，作为包含人类丰富感知、理解、情绪等因素的“信念”被抽干简化成生冷的数字，机器只负责去完成它，并没有真正相信它。

鉴于算法式的弊端，启发式力图抛开概率，寻求建造一种直接衡量信息的纯粹心理学方法，以贴近人类的心灵轨迹。启发式不同于追求算法的求解方式，它模仿人类理解的开启方法，注重直觉、顿悟、学习等非结构形式的技能增长，借助反复训练成就与人类相仿的“真实”信念。〔33〕当然，启发式仍然绕不开对信念的衡量，它的第一种方法是假定信念的程度可以被感觉到，即相信度就是感觉的强度。“我们可以假定一个信念的程度是可以被有这个信念的人所感觉到的，例如，某种感觉会伴随着信念一起出现，这种感觉可以被称为信念感觉或确信感觉，不同的信念由强弱程度不同的感觉所伴随，而相信度就是指这种感觉的强度。”〔34〕这种方法其实使相信度依赖于更难操作的感觉度上，且“也是不符合事实的，因为实际上我们最强烈的信念经常都没有伴随任何感觉；没有人对他视为理所当然的事情有强烈的感觉”。〔35〕恰恰是人们习以为常的惯常式行为，最为理所当然却最为平淡无奇、波澜不惊、毫无感觉。

因此，启发式转向第二方案，以外在具体行为作为信念强度差异的识别标准。由于在相信与不相信之间存在太多模棱两可的纠缠状态，其中细微的内在感觉很难有效体察与衡量，追求外在具体行为的对应性是标识内心差异的显性办法。当主体愿意据此行动时，我们视为它相信；反之，视为不相信。如此一来，不同程度的信念度的差异体现在“我们应该在多大程度上根据这些信念而行动……在许多情况下，……我们关于我们的信念程度的判断实际上是关于在假设的场合我们应该怎样行动的判断”。〔36〕采用这种方案，启发式将信

〔33〕 See Jeffrey M. Lipshaw, *Halting, Intuition, Heuristics, and Action: Alan Turing and the Theoretical Constraints On AI-Lawyering*, 5 Savannah L. Rev. 156 (2018).

〔34〕 Frank P. Ramsey, *Truth and Probability*, 1 Readings in Formal Epistemology 28 (2016).

〔35〕 同上。

〔36〕 同上。

念的培养与辨识放置于日常的行为情节之中，通过普遍的生活经验形塑相信与不信。这一方案固然贴近人类内心与行动逻辑，但是让人工智能实现起来颇为繁复，因为机器没有日常生活，也没有针对行为的内心回应，所有常识经验都需要提前输入电脑。而且这一点无法依赖大数据。上述种种信念带有强烈的个体色彩，是基于个人经验与常识的生成提炼，只能通过长期反复的模拟训练，慢慢让人工智能建立起自身的“生活逻辑”。

启发式采用打赌的方法来训练人工智能的日常信念系统。它具有一定的合理性，人类在现实生活中采取某种行动也恰是打赌相信它。人工智能被安排到各种生活场景中去练习“可能输也可能赢”的判断，进而构建独立自主的信念体系。这一方法在原理上并无不妥，但鉴于目前的技术，在操作层面会碰及两大难题：其一，通过打赌建立信念最重要的是给出结果，赢或输是纠正打赌信念的唯一有效工具。只有给出赢与输，信念才会因此调整：或基于赢的结果而增强判断、总结经验，或基于输的结果而改变策略、调整方案。在现实世界中，生活一定会给出输赢结果，且这个结果实际而客观。但在模拟世界中，所有标准答案仍然还是一种假设，人工智能并没有遭遇真正的输的痛苦或赢的刺激。其二，输入什么样的生活场景，将直接决定建立什么样的信念体系。即使输赢结果可逼真还原，满足上述各条件的计算机经过训练后所形成的信念也是个体信念，与个人所形成的信念无异，它并不代表人工智能的普遍化信念。采用这种方法，每一台计算机均会形成一套不同于其他计算机的、存在一定偏差的独立信念体系，〔37〕就如同一个人的信念体系不同于另一个人的一样。果真如此的话，建立由人工智能完成的信念体系的意义就变得相对有限了，因为我们期待的是人工智能能够提供超越人类个体信念的更客观、更准确、更科学的，具有普适规律的信念系统。

（三）机器如何表达

司法最终输出裁判。裁判是针对证据与事实进行回应与展示证明的一系列表达。人工智能要像人类智能一样在事实认定中发挥作用，就必须在完成推理学习、形成信念之后，输出智能结果以实现有效表达。

人类借助于语言实现表达。语言是全域性的，不仅可以与客观世界一一对应，且几乎可以无障碍地传递出内心世界的全部，即全息表达。〔38〕人类的表达异常丰富，能够呈现千万种概念，并借助概念激活关联概念，经由思想延展相邻思想，既可传送出有穷内涵，又可浮现出无限意义。借助于成熟的语言系统，人类的表达实现了众多概念与思想的“并行处理”。但是，作为计算类机器，人工智能输出的通常是一个或一组数值。一般而言，一项数值作为一个结论只表达一种意义，数字展现不出语言的丰富世界，基于公式推导出的数字通常都是单线条的，表明一种单向上的状态。若使人工智能集中呈现多元数据，算法必须升级到足够高阶的程度。因为叠加数值并不像排列话语即可生成意义那般简单，它需要复杂公式的支持，每一步都离不开被证明的公理。展现多重数据化结论所依赖的复杂公式算法的成倍追加，实现起来并不轻松。

为了保持与人类对话的简约与流畅，人工智能在输出结果时也需要依赖语言。〔39〕语言

〔37〕 See Batya Friedman & Helen Nissenbaum, *Bias in Computer Systems*, 14 ACM Transactions on Information Systems 332 – 345 (1996).

〔38〕 参见梅勇 《语言与世界：语言哲学的研究范围》，《外语学刊》2009 年第 5 期，第 27 页以下。

〔39〕 See L. Thorne McCarty, *How To Ground a Language for Legal Discourse in a Prototypical Perceptual Semantics*, 511 Mich. St. L. Rev. 526 (2016).

包含语法、语义与语用三个层面。〔40〕在一定程度上，人工智能可以掌握并处理语言，这主要指语法层面。语法本身就是一种操作话语编排的规则体系，它实质上是程序化、形式化的，与计算机的本质相符。将字词处理为可识别的符号形式，将语法抽象为计算应用形式规则，这时计算机即可依规则指令设计程序处理加工符号信息。它可以基于形式逻辑的运算与操作，准确识别并表达出字、音、词等符号形式。但是，机器无法理解并表达语义。语义强调符号与所指之间的意义关系。〔41〕人工智能表达符号是出于规则操作的结果，机器虽然表达出某个符号，但难以通过符号操作理解符号的意义。计算机无法具有理解能力，因为实现表达依靠的仅是形式逻辑的运算，而纯形式操作本身并不产生理解所需的内容。程序可以代替输出，却无法代替理解，而事实认定的核心恰恰在于对证据与事实的理解。这种困境具体表现在：其一，语法与语义之间存在断裂，语法产生不了语义。塞尔曾说："人心不仅是语法的，它还有一个语义的方面。计算机程序就不可能替代人心，其理由很简单：计算机程序只是语法的，而心不仅仅是语法的。心是语义的，就是说，人心不仅仅是一个形式结构，它是有内容的。"〔42〕其二，心之所以产生语义，是源自个体心灵与其环境之间长久信息关系的互构。语义发自情境中行为依赖的"刺激—反应"关系，这是人们在"身临其境"时本能的反应，所以，语义由心智派生，是心智的一面影射镜子，而不是一种模型建构。其三，语义所反映的心智具有生物自然主义立场。心智是一种生物现象，其载体为人体中的神经蛋白，这是心—物关系建立的基石，而电脑中的金属与硅不具有这个特性，人工智能的物理构成缺乏大脑物质生化特性，这是计算机难以模拟大脑生成理解力的关键所在。

在语用层面，即符号与解释者之间的关系上，人工智能与人的对话沟通也存在意向性障碍。〔43〕无论采用哪种语言，交流传导的核心是意向。智能的传导说到底就是意向的传递，正是意向性使表达不仅是一个程序操作结构，更是一种表述思想与信息内在能动的活动。但是，人工智能难以具备意向性功能，〔44〕因为意向不可计算，无论怎样的算法或程序都不能解释心灵自然产生的想法。这种想法根源于人的生物学现象，具有天然智能的进化特性，是人工智能无法模拟的。同理，任何通过语言符号表达出的理念或思想，也不能仅仅凭借它展示出计算机算法或程序就当然推断其具有意向性，表示并不等于领悟。这种不一致在正常操作状态下可能不易察觉，但在出现语言失误时就变得异常明显。〔45〕假定人工智能在输出事实认定结论时，跳出一个令人难以理解的怪异词汇。我们并不清楚这个词汇究竟属于机器运行错误，还是机器就想表达这个词汇。如果属于前者，那么我们担心的是，那些看上去正常的词汇是否也存在错误，到底哪些词汇正确表达了人工智能的本意；如果属于后者，那么我们疑惑的是，为什么人工智能选择使用这个词汇，是机器希望通过这个

〔40〕 参见 [美] 马克·布尔金 《信息论：本质·多样性·统一》，王恒君等译，知识产权出版社 2015 年版，第 179 页。

〔41〕 参见 [意] 翁贝尔托·埃科 《符合学与语言哲学》，王天清译，百花文艺出版社 2006 年版，第 66 页以下。

〔42〕 [英] 约翰·塞尔 《心、脑与科学》，杨音莱译，上海译文出版社 2006 年版，第 23 页。

〔43〕 参见前引〔33〕，Lipshaw 等文，第 166 页。

〔44〕 参见 [荷] 彼得·阿德里安斯、约翰·范·本瑟姆主编 《爱思唯尔科学哲学手册：信息哲学》下册，殷杰等译，北京师范大学出版社 2015 年版，第 724 页以下。

〔45〕 参见前引〔5〕，Roth 文，第 1276 页。

词汇改变我们的认知，还是机器并没有搞清楚这个词汇的正确内涵或适用情形。一个词汇就可以在人与机器间制造诸多疑问，这早已远离表达助力沟通的本意了。

由此可见，现阶段的人工智能能够解决语法问题，但尚未逾越语义和语用两道屏障，处理语言是一回事，深谙其义并以心灵的名义准确表达又是另一回事儿。人工智能对事实认定结论的表达目前尚属一种理想，其实现路径仍有待开拓。

结　语

经由上述分析可以发现，伴随法律人工智能的不断更新与发展，实体形态的证据正越来越多、越来越容易地被人工智能“解码”和“理解”。基于数据的概率推理已产生强大的规模累积效应，算法的技术迭代也持续拓展着人工智能在司法领域中的应用疆界，使其能够发掘出人脑所不及的统计意义上的问题与现象，揭示出计算意义上的逻辑与规律。机器的深度学习能力更是使得上述进步呈指数化加速。无论人们是否愿意接受，司法裁判中的相当一部分工作正在转移给人工智能，人工智能还将“拿走”更多。但同时也应看到，人工智能要深度介入司法，仍有许多瓶颈亟待突破。例如，在将证据转化为数据的过程中，证据信息可能丢失，影响事实认定的准确性；目前人工智能尚不能有效解析并认知人类语言，极大限制了其理解证据的能力；人工智能在因果推理方面还存在障碍，而法律案件中的因果关联大多较为复杂，限制了人工智能的适用空间；算法面临的可计算性难题和复杂性难题，又构成人工智能发展的自我设限；人工智能难以像人类一样建立信念并准确地完成表达，要解决这一难题，有待智能技术发展到更高阶段。尽管我们有足够多的理由相信，随着智能技术的发展，上述难题终究会被一一攻克，但客观来讲，在这些技术瓶颈取得突破之前，我们将一直处在法律人工智能的初级阶段。

就当下而言，笔者认为，可以尝试基于“小数据”训练探索法律人工智能的心智微结构。所谓探索“心智微结构”，是针对特定难题模拟人的心智轨迹，在极小的微观层面建立模型关系，通过微型建模规划结构、塑造行为、训练认知、建立信念、完成表达。由于人工智能对事实认定过程中的复杂问题化解能力有限，搭建心智微结构能够缩小问题范围，以各个击破的方式，将大块难题分解为小块问题，在小区域内提升智能密度，进而最大化增强人工智能，突破潜在阻力。实现心智微结构的具体路径是运用“小数据”训练。〔46〕“小数据”是指“能为人类所理解的数量足够小的数据，它是一种在容量和格式上都便于访问和操纵，含有用信息的数据”。〔47〕之所以强调小数据，而不是大数据，原因在于，“‘大数据’这个术语是关于机器的，而‘小数据’是关于人的。这是说，可以一眼看清或比如只有五个相关数字的就是小数据。小数据是我们以前认为的数据”。〔48〕小数据聚焦于个体智能发挥的有限范围，以单独问题的求解为目标，旨在推动人工智能在细微处的进化与提升。小数据以具有针对性的数据切割、编排、修正等精细化处理方式，力图克服人工智能

〔46〕 参见 ［美］ 克莉丝汀·L. 伯格曼 《大数据、小数据、无数据：网络世界的数据学术》，孟小峰等译，机械工业出版社 2017 年版，第 3 页以下。

〔47〕 范煜 《数据革命：大数据价值实现方法、技术与案例》，清华大学出版社 2017 年版，第 58 页。

〔48〕 同上。

在事实认定上的种种难题。

运用小数据训练心智微结构的意义在于：第一，运用小数据训练心智微结构是人工智能与事实认定的关联点与着力点，培育微观智能是防止人工智能激进化的缓冲阀，也是人工智能在事实认定细节处真正供给有效智慧的助推器。人工智能介入事实认定的主要困难并不是宏观架构上的不可通约性，而是在细小环节上与事实认定具体问题的融贯性。心智微结构旨在克服细节上的计算障碍，使人工智能实现迈向事实认定的“一小步”，并通过这“一小步”促成宏观问题的逐步化解。第二，运用小数据训练心智微结构是一种有关人工智能与法治匹配度的研究。它着力于搭建两者深层结构的对接渠道，寻找同构性的基础与条件，克服彼此的排异反应，建立相互融贯的一体化模型，进而推动实然层面的细节改进。第三，运用小数据训练心智微结构是一种发现问题并解决问题的路径尝试。它着力分析人工智能影响司法的内在机理，以人类可控的方式发展法律人工智能，既为人所用又为人所限，遵循良性且有序的路径。而达到这一点，就需要重构一套使问题的发现与解决走向深入的结构性路径与精细化方案。对此，无论是现有的法学分析框架还是目前的人工智能技术，还未从根本上对所呈现出的现实与种种可能给出有说服力的解释与回应。这激励我们去寻找属于人工智能法学界或法律人工智能界的独有且自主的理论，并完成系统的理论建构。

Abstract: The deep integration of artificial intelligence and justice is embodied in both the application of law and fact-finding. The latter is a precondition of the former. To intervene in the fact-finding in a case, artificial intelligence needs to digitize the evidence, integrate the data, and output conclusions that can be understood by human beings. On the step of digitalization of evidence, it is necessary to carry out structured data transformation of evidence and overcome language problems. On the step of data integration, artificial intelligence mainly uses probabilistic reasoning, rather than causal reasoning, as its logical reasoning model and needs to face two major problems of computability and complexity in its algorithm. On the step of conclusion output, it needs to solve the problems of how to deepen machine learning, how to build belief, and how the machine expresses itself. The main problems faced by the integration of artificial intelligence into the fact-finding in cases can be solved through “small data” training and the gradual building of “mind-microstructure” of artificial intelligence.

Key Words: artificial intelligence, fact-finding, mind-microstructure, small data