

人工智能立法背景下人工智能的法律定义

陈亮 张翔

摘要：科学、合理地界定人工智能的法律定义，是人工智能领域体系化立法的逻辑前提。在本体论意义上，人工智能首先是一种信息处理系统，而后才有人工智能科学、人工智能产业等社会性认知。从内涵看，人工智能是依托算力基础设施，通过控制系统的算法处理输入的数据，以软件或硬件等多元集成方式嵌入系统后输出，或者直接在具体场景中输出对人类某种理性功能的模拟状态，在环境中进行交互，并在目标约束下经过反馈修正，最终完成预设任务的信息系统，该种智能系统的法律本质为“理性智能体”。从外延看，人工智能既包括单纯的智能软件或硬件，也包括软件和硬件集成嵌入的智能体，而“强弱”之分和“通用专用”之分不适宜在外延中直接列举，而是需要随着技术发展通过评价性的法学判断来决定是否将其纳入人工智能的法律意涵。

关键词：人工智能立法；法律定义；提取公因式；理性智能体；系统一控制论

中图分类号：D922.17 **文献标识码：**A **文章编号：**1000—8691（2023）05—0162—09

法律概念是法律规范和法律制度的建筑材料^①，具有“表达”“认知”和“提高法律合理化程度”三大功能^②。清晰的法律概念是科学表达法律规范、正确认知法律内容、增强法律适用确定性的基本前提。“人工智能法”作为重点新兴领域立法，其调整对象界定、制度框架搭建，以及具体规则设计，皆需要科学、合理的人工智能法律定义充当识别社会关系、定性法律关系的第一道“滤网”。早在2017年，国务院在《关于印发新一代人工智能发展规划的通知》中就提出人工智能立法的“三步走”战略目标，其核心要义是建构完善的“人工智能法律体系”；2023年6月6日，国务院办公厅发布《国务院2023年度立法工作计划》，正式将“人工智能法草案”列入预备提请全国人大常委会审议草案的范围。在人工智能立法趋向“体系化”以及立法进程日益加快的背景下，首先需要解决的便是人工智能立法体系的逻辑前提——人工智能的法律定义——应当如何确定的问题。虽然人工智能的定义方式非常多元，国内外人工智能领域立法实践中都有单独设置定义性条款来界定人工智能法律概念的先例，但这些定义有其特定的立法背景和体系语境，未必适合当前中国“人工智能法”的立法需求。

从丰富的立法素材中“提取公因式”是生成法律概念的重要方法。本文遵循人工智能的“规范认知”和“学术认知”线索，全面梳理人工智能的各种定义模式、总结共识和分歧，并基于体系性思维检视定义方式之优劣；引入“系统一控制论”原理解决界定人工智能法律定义时存在的“概念不确定性”问题，并尝试从“属概念”“内涵”和“外延”三个维度重构人工智能的法律定义。

作者简介：陈亮，男，西南政法大学人工智能法学院院长、教授、博士生导师，主要从事人工智能法学研究。

张翔，男，西南政法大学人工智能法学院博士研究生，主要从事人工智能法学和法律方法论研究。

① [德]魏德士：《法理学》，丁晓春、吴越译，北京：法律出版社，2013年，第91页。

② 张文显主编：《法理学》（第5版），北京：法律出版社，2018年，第114页。

一、人工智能定义的多维面向

（一）域外规范性文本对人工智能的界定

域外政策法规文本中，只有少部分国家或组织^①对人工智能的界定沿袭了词典的“功能型”定义模式，即强调人工智能作为计算机科学的分支，其具有复制、模拟、协助甚至替代人类智能能力^②，大部分国家和组织采用“发生定义”，从被定义概念反映、代表、指称事物的发生或来源方面来揭示种差。^③该类定义模式可以概括为：“目标控制下的输入—输出模型”，其中“目标/任务/问题”通常由人为设定，“输入”的是可供算法处理的数据，而“输出”的则是能够影响环境的内容、预测、建议或决策。除前述共性外，针对“是否需要在人工智能定义中列举外延，以及如何对人工智能进行分类”等问题，域外规范性文本的处理方式不尽相同。主要分类方法有两种，一是根据人工智能应用领域的宽窄，将其分为“通用人工智能”和“专用人工智能”。譬如《西班牙人工智能研发战略》认为：“当前的人工智能不是实现通用智能，而是专注于所谓的特定人工智能，……但通用人工智能仍然是要实现的主要目标，即创建一个具有智能多任务系统的生态系统”。又如《人工智能：澳大利亚的伦理框架》具体界定了二者的特征：“狭义人工智能执行特定功能，通用人工智能在一系列领域都可以与人类智能相媲美。”二是根据人工智能与人类的相似程度，将其分为“弱人工智能”和“强人工智能”。譬如德国《人工智能战略报告》指出：“强人工智能意味着人工智能系统具有与人类相同的智力，甚至超过人类；弱人工智能专注于使用数学和计算机科学的方法解决特定问题，从而开发的系统能够自我优化。”尽管某些国家或组织在规范性文本中没有明确提及人工智能有强弱之分，但使用了“自主/自我”等表征人类意识的心理学措辞。譬如欧洲议会和委员会通过的《人工智能法》草案妥协修正案（以下简称“欧盟AIA”）以及经合组织发布的《人工智能委员会的建议》均承认人工智能具有不同程度的“自主性”。又如欧洲人工智能高级别专家组发布的《人工智能的定义：主要能力和科学学科》指出：“人工智能系统……决定为实现给定目标而采取的最佳行动……分析环境如何受到先前行为的影响来调整自己的行为。”而阿联酋国家政府发布的《人工智能伦理原则与准则》则认为，人工智能具有“自我完善”能力。

（二）中国规范性文本对人工智能的界定

表1 三部文件中人工智能定义性条款比较

	深圳人工智能条例	上海人工智能条例	人工智能术语国标
属概念	侧重行为论，其属概念是“模拟、延伸或者扩展”。	侧重本体论，其属概念是“理论、方法、技术及应用系统”。	侧重知识论，其属概念是“研究和开发”。
两部条例的种差与属概念之间呈现出“基本互换”的特点		国标较之于两部条例的三个特征	
种差	利用计算机或者其控制的设备，通过感知环境、获取知识、推导演绎等方法。	利用计算机或者计算机控制的机器模拟、延伸和扩展人的智能，感知环境、获取知识并使用知识获得最佳结果。	特征1：强调人工智能系统的“目标/任务—输出”结构，而“上海人工智能条例”则采用了评价性用语“获得最佳结果”，没有客观陈述“输出”端，而是侧重“目标/任务”端并对其赋予了模糊的价值限定。 特征2：更加具体化地在技术维度揭示了研发端使用人工智能方法的行为特征——开发用于执行任务的（数据、知识、过程等）模型；并且明确展示了人工智能系统的输出类型，包括内容、预测、推荐或决策等。 特征3：特别强调人工智能系统具备不同的自动化级别。
两部条例在种差方面的共识与分歧			
共识1：都将“计算机或者其控制的设备/机器”作为人工智能的载体。 共识2：都将“感知环境、获取知识”作为人工智能的方法。 分歧1：“深圳人工智能条例”在两种方法之外提出了“推导演绎”方法，而“上海人工智能条例”在两种方法之外强调“使用知识获得最佳结果”。 分歧2：“上海人工智能条例”在附则中不完全列举了七种关键的人工智能技术外延。			
两部条例和国标在种差方面的共识			
共识：“人工智能术语国标”指出“使用人工智能方法开发知识模型”，两部条例也关注人工智能对知识的获取和使用。			

① 域外政策法规的梳理范围，主要参考了欧洲议会在2020年7月14日发布的由安德里亚·贝托里尼博士撰写的政策研究报告 *Artificial Intelligence and Civil Liability*（《人工智能与民事责任》）第一部分，与该文选取的国家和地区基本一致。

② 譬如丹麦《人工智能战略报告》、芬兰《人工智能战略报告》、印度《国家人工智能战略》以及日本《人工智能战略》等对人工智能的界定。

③ 陈波：《逻辑学是什么》，第247页。

在中国涉及人工智能的政策法规文本中，鲜见针对人工智能本质特征和特有属性的明确定义，大多数文件只是在涉及人工智能的语句中零散地给出特征描述词或者类型归属词，难以准确阐明何谓人工智能。^①对人大、政府和法院出台的三类文件“合并同类项”，可以归纳出人工智能的如下特点：其一，人工智能的属概念主要有“技术”“领域”“产业”“学科”“产品”和“权益”六类。技术范畴强调“变革性信息技术和智能硬件”特征，领域范畴强调“前沿性、创新性、战略性”特征，产业范畴强调“数字经济”特征，学科范畴强调“领域性、基础性、新范式”特征，产品范畴强调“科技创新成果”特征，权益范畴强调“数字知识产权和数字财产权”。其二，对人工智能特征的直接描述有四种，分别是“基于大数据、物联网、5G等技术”“深度学习、跨界融合、人机协同、群智开放、自主操控等新特征”“具有自主学习能力”和“具备高可解释性、强泛化能力”，主要从人工智能的基础层（硬件设施）、技术层（数据和算法）和场景层（泛在性）三个方面概括出人工智能的关键特质。

此外，对比《深圳经济特区人工智能产业促进条例》（以下简称“深圳人工智能条例”）《上海市促进人工智能产业发展条例》（以下简称“上海人工智能条例”）和GB/T 41867-2022《信息技术 人工智能术语》（以下简称“人工智能术语国标”）三部文件中人工智能的定义性条款^②，其特点如表1所示。

二、人工智能定义的法学审思

法学界对人工智能法律定义的专门研究较少，在涉及人工智能的法学著述中开辟独立板块来确定问题讨论前提已实属不易，大部分著述只是简单引用人工智能的某种概念以服务于作者预设的价值立场，但该种定义模式为何正当或者具有合理性，却缺乏细致论证。总体来看，根据属概念不同，中国法学界对人工智能的定义大致可以分为两类：主张人工智能是一种非实体性客观存在的“科学技术论”，以及主张人工智能是一种实体性客观存在的“系统（程序/机器）论”和“类人论”。“科学技术论”虽然倾向于以人类为参考系，但主要从“功能或目标”角度来描述人工智能的内涵，而界定外延的分类标准和具体类型则分化为“研究领域”“技术基础”和“应用场景”三种不同出发点。^③“系统论（程序/机器）”观点除了从功能上描述人工智能，还侧重在发生学意义上阐释人工智能的运行原理，并且依据系统的迭

① 以“人工智能”为关键词在“北大法宝”进行法律法规检索，定位关键词所在句后，如果该句式属于“人工智能是X的Y”，则分别摘取特征描述的关键词“X”和表征类型归属的关键词“Y”。根据政策法规制定主体不同，将这些关键词按照“法律及法律性文件（人大）”“行政法规及规范性文件（政府）”和“司法解释性质的文件及两高工作文件（法院）”进行归类，统计结果如下。一是法律及法律性文件（人大）中描述人工智能的关键词，主要有：新技术、新应用；现代技术手段（现代信息技术）；科技；科技创新项目；重大创新领域；前沿领域；新兴数字产业；基于大数据、物联网、5G等技术的人工智能；人工智能是引领新一轮科技革命和产业变革的战略性技术；关键领域；新技术、新领域；新型基础设施（“人工智能”和“新一代信息基础设施”也并列出现）；科技创新和高技术领域；前沿战略性领域。二是行政法规及规范性文件（政府）中描述人工智能的关键词，主要有：人工智能加速发展，呈现出深度学习、跨界融合、人机协同、群智开放、自主操控等新特征；经济发展的新引擎；新一轮产业变革的核心驱动力；人工智能是影响面广的颠覆性技术；具有自主学习能力的人工智能；人工智能技术属性和社会属性高度融合的特征；实现具备高可解释性、强泛化能力的人工智能；学科专业；智能化技术；领域科学技术；现代科技手段；信息技术和智能硬件；新技术应用；变革性技术；新一代信息技术；新职业新技能；重大科学基础工程。三是司法解释性质的文件及两高工作文件（法院）中描述人工智能的关键词，主要有：新领域新业态；信息化手段；新技术；新科技革命和产业变革前沿领域；现代科技；新兴科技；信息技术；新业态新领域；新技术新思维（人工智能+）；新类型数字化知识产权财产权益；前沿技术；新知识新技能；科技创新成果。以上统计截止时间为2023年6月16日。2022年8月，深圳市人大常委会通过首部以“人工智能”命名的地方性法规《深圳经济特区人工智能产业促进条例》；2022年9月，上海市人大常委会紧随其后，也通过了《上海市促进人工智能产业发展条例》；2022年10月，国家标准化管理委员会发布GB/T 41867-2022《信息技术 人工智能术语》。目前，主要是这三部文件设计了专门界定人工智能的定义性条款。

② 主要是《深圳经济特区人工智能产业促进条例》第二条第一款、《上海市促进人工智能产业发展条例》第七十一条，以及GB/T 41867-2022《信息技术 人工智能术语》第3.1.2、3.1.8、注1、注2。

③ 包括但不限于吴汉东：《人工智能时代的制度安排与法律规制》，《法律科学（西北政法大學學報）》2017年第5期；叶明、朱静洁：《理性本位视野下智能机器人民事法律地位的认定》，《河北法学》2019年第6期；龙卫球：《人工智能立法规范对象与规范策略》，《政法论丛》2020年第3期。

代阶段来划分人工智能外延的类型。^①“类人论”观点的共识在于都选用了“自主性”特征，但对自主性的表现形式、人类特质的主要类型以及人工智能对人类智能的替代程度等问题的看法尚存分歧。^②概言之，相较于政策法规的规范性认知，中国法学界对人工智能认知方式的差异和观点分化，与之基本一致。

（一）检视依据：逻辑一致、脉络关联和价值融贯

体系方法是法律人规整法律素材、建构制度体系、维护法秩序统一、系统性解决社会问题时必不可少的元方法，一直被法学研究与法律实践予以践行。^③综观法律体系化思想嬗变和中国学者对体系化方法的转介历程^④，体系化立法的理想状态可以提炼为“立法要素的内在逻辑保持一致”（以下简称“逻辑标准”）“立法脉络能够维系整体性意义关联”（以下简称“脉络标准”）以及“立法价值在规范中融贯地表达”（以下简称“价值标准”），这三个标准也是判断法律定义是否科学、合理的重要依据。

一是逻辑标准。法律概念应当全面映射涉法事实，概念的内在特征之间能够保持互斥，也即法律概念以“特征增减”方式并通过“涵摄”来实现内容固定，确保概念外延至少在逻辑层面能够穷尽涉法事实。^⑤在此标准要求下，人工智能的法律定义应当统合技术实践中已经出现的人工智能具体类型，也即在逻辑学层面抽象出属概念的“最大公约数”，并且在各类属概念之下明确人工智能的种差，在立法技术上可以采用“有限列举+等字兜底”方式，确保法律概念在相对稳定的基础上对未来技术变迁保持开放，也即法律定义具有适当程度的包容性。^⑥

二是脉络标准。法律概念应当形成谱系，不仅能够通过形式逻辑确保各层次的法律概念之间互相推演，而且能够维系各要素在整体上的意义脉络关联，表现出积极的“内聚性”。具言之，通过形式逻辑对必要的抽象法条（规则）采取一种概念层系的建构，在体系关联上认知彼此互为条件、又相互衍生的法条，以便能够追溯单个法条的“谱系”。^⑦法律概念作为法律体系的基本要素，其并非制定法偶然形成或者任意维系的集合^⑧，也不是对规范素材简单地摘录、汇编，而是应当经由规范解释，对相关内容进行整合、抽象，形成无矛盾的、和谐的有机体系^⑨。因此，人工智能的法律定义应当能够与既有人工智能领域立法的基础概念保持有机衔接，譬如“深度合成”“自动化决策”“生成式人工智能”和“决策式人工智能”^⑩等，确保概念体系融洽。

① 包括但不限于季卫东：《人工智能开发的理念、法律以及政策》，《东方法学》2019年第5期；陈亮：《繁荣人工智能法学 助力人工智能法律法规体系建设》，《中国高等教育》2020年第2期；曹伟：《技术迭代视角下人工智能的法律分析》，《人大法律评论》2021年第1辑。

② 包括但不限于马长山：《人工智能的社会风险及其法律规制》，《法律科学（西北政法学报）》2018年第6期；刘艳红：《人工智能法学研究的反智化批判》，《东方法学》2019年第5期；刘宪权：《对人工智能法学研究“伪批判”的回应》，《法学》2020年第1期。

③ 杨铜铜：《体系解释方法运用研究》，北京：中国社会科学出版社，2021年，第120页。

④ 法律体系化思想肇端于历史法学派萨维尼以“法律制度”为中枢的“有机体系”论，经普赫塔所代表的概念法学改造，发展为凸显逻辑演绎的“概念金字塔”方法，但因其脱离实践而被利益法学、社会法学批判。其后，拉伦茨的法学方法论历经两次转向，建构出由“类型序列”联结的“内外部体系”，试图以评价法学思路弥合规范与现实之间的“裂痕”。由此，潘德克顿法学的“法律体系化”方法以“实质理性—形式理性—实质理性”的循环，实现了“否定之否定”式发展，为现代法教义学的成熟留下了宝贵的思想遗产。中国法学界主要经台湾法学家转介该种体系化方法，不同领域法学对其继受程度不尽相同，但或多或少都受拉伦茨对“内部体系”和“外部体系”相区分观点的影响；在自主引介、吸收施塔姆勒、德沃金、拉兹、阿列克西、佩岑尼克和麦考密克等法学家的体系化思想并经本土化改造后，法学界基本认同达致理想法律体系的三个标准：“逻辑一致性”“价值融贯性”和“维系整体意义脉络关联”。

⑤ [德]卡尔·拉伦茨：《法学方法论》（全本·第6版），黄家镇译，北京：商务印书馆，2020年，第33页。

⑥ Jonas Schuett (2019). *A Legal Definition of AI*. Retrieved on 17th Aug. 2023 from: <https://arxiv.org/abs/1909.01095v1>.

⑦ 舒国滢：《格奥尔格·弗里德里希·普赫塔的法学建构：理论与方法》，《比较法研究》2016年第2期。

⑧ 朱虎：《萨维尼的法学方法论述评》，《环球法律评论》2010年第1期。

⑨ 杨代雄：《萨维尼法学方法论中的体系化方法》，《法制与社会发展》2006年第6期。

⑩ 丁磊：《生成式人工智能 AIGC 的逻辑与应用》，北京：中信出版社，2023年，第6页。

三是价值标准。“法律概念也承担着法的调控任务，受制于目的论。”^①在定义中内嵌价值理念是通过法律概念划定调整对象、调控社会关系的第一道“滤网”，因而人工智能法律定义的价值倾向也会直接影响立法目的能否准确、恰当且有效地实现。具言之，法律概念的形成和内容解释必须回归法律原则，概念并非仅仅以抽象方式形成，也是基于对原则的考量而形成，并且在对原则的回溯中表达其内容。^②因此，人工智能的法律定义应当保持适度的弹性解释空间，为“人工智能法”的价值嵌入预留解释学通道；同时，对人工智能法律定义的理解不能仅停留在文义层面，而是需要依据法的价值和法律原则来充实其实质意涵。

（二）求同存异：对现有立法素材“提取公因式”

人工智能的定义具有一张“普罗透斯似的脸”，其多元化和不确定性折射出使用术语“人工智能”的困难之处，似乎在于该术语难以通过下定义方式来确切描述；即便其可以被定义，一旦机器能够完成“之前只有人类才能完成”的任务，无论是下棋还是人脸识别，那么它就不再被认为是智能的标志，这便是所谓的“AI效应”。^③诚然，“人工”“智能”等概念无法进行词典式的孤立定义，试图分别界定二者意涵并拼接出“人工智能”的定义方式也是徒劳的，原因在于“智能”难以被归结为传统的、孤立实在论道路上的计算问题，但又可以根据“图灵测试”通过行为主义的比较来判定，即它作为数学意义上的判定性问题处于可计算与不可计算的悖论之间；即便如此，规范主义的定义并非不可能，定义一个词的含义或者所有词汇含义的，并非该语言系统中所有真命题的集合——因为这个集合无穷大而无限缠绕——而是该语言系统中判定何种命题成立的标准，因而定义人工智能需要采用系统论观点和复杂性系统理论背景下的公设系统。^④该公设系统应当是在人工智能立法语境下，遵循判断法律定义科学性的一般标准，从前述人工智能定义的立法素材中“提取公因式”，在此基础上结合人工智能立法之目的所在，对“公因式”进行转译和组合。

首先，人工智能的种差主要从发生机理和应用特征两个维度进行描述，在发生机理层面的“公因式”可以提取为：硬件支撑（计算机设备等）+程序操控（目标控制下的输入—输出结构）；应用特征层面的“公因式”可以提取为：自主性、泛在性和交互性。其次，人工智能定义的分歧主要集中在属概念和外延；属概念主要有“实体”和“非实体”（科学、理论、技术、知识、方法、领域、产业、权益等）两大类，其中实体类定义又存在“物质性”（系统、程序、模型、机器、产品等）和“非物质性”（依据对人类行为功能的替代程度差异，分为模拟、延伸、扩展等）两种区分；而整体意义上的人工智能类型，常见分类是“通用人工智能和专用人工智能”以及“强人工智能和弱人工智能”。最后，虽然人工智能定义在发生机理和应用特征方面能够抽象出一些基本共识，但在前述标准检视下，若想生成科学、合理的人工智能法律定义，还需降低人工智能种差的模糊性并选定符合立法目的需要的人工智能外延。

“自主性、泛在性和交互性”等看似内容确定、表达中立的技术性术语，不能直接转化为立法文本，原因在于这些术语背后隐藏着“智能判断的不确定性”问题，关乎法律对社会生活的介入程度（或者说立法价值追求对人工智能的容忍度），如果不加以限定，将会造成法律适用的不确定性。虽然何谓“智能”的问题在一定程度上达成共识——具有实现预设目标的部分计算能力，并且在可实现目标的种类及其实现程度上，人类、动物和智能体存在差异^⑤——但环境变化或者目标改变，会导致同一智能体在不同应用场景中表现出不同程度的智能，因而评估智能体实现预设目标的能力强弱，还需要依据各种计算科学意

① [德]魏德士：《法理学》，第91页。

② [德]卡尔·拉伦茨：《法学方法论（全本·第6版）》，第604页。

③ Shruthi Anand (2017). *Artificial Intelligence-Literature Review*. Retrieved on 17th Aug. 2023 from: <https://cis-india.org/internet-governance/files/artificial-intelligence-literature-review>.

④ 张绍欣：《普罗米修斯精神与人工智能前史——人工智能概念的历史规范主义回顾》，《中国图书评论》2018年第7期。

⑤ Andrea BERTOLINI (2020). *Artificial Intelligence and Civil Liability*. Retrieved on 17th Aug. 2023 from: [https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU\(2020\)621926](https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU(2020)621926).

义上的阈值进行“基准测试”。然而，基准测试的阈值标准并非统一，评估维度也不一定与人类智能相关。根据人工智能科学的四种研究范式^①，至少存在两种智能性判定基准，一是依据理性基础上的“理想性能”进行衡量，二是依据智能体对人类行为的“保真度”进行衡量。从实证研究结果看，人工智能研发人员对前述四种范式的偏好程度也不尽相同，约65%的受访者选择“理性”主导下的人工智能定义，对政策文本的统计分析则结果迥异，约57%的人工智能定义强调与人类思想或行为进行比较。^②因此，同一智能体在两种不同阈值下进行基准测试，既有可能被判定为“智能”，也有可能被判定为“非智能”。

此外，并非所有的人工智能属概念都符合人工智能立法的需求，不同的外延也在一定程度上决定了人工智能的立法模式。以欧盟AIA为例，该法案侧重对人工智能部署和应用风险的分類分级监管，但事实上，仅符合该法案定义的人工智能并不会直接引发使用风险，也即人工智能定义本身对法案适用没有发挥应有的过滤功能；尽管第3条之（1）和附件一的范围几乎包括了每一种人工智能系统，但法案内容大多是专门针对高风险系统的，符合现有的产品安全立法；换言之，高风险的分类不仅取决于人工智能系统所执行的功能，而且还取决于该利用系统的具体目的和模式；人工智能系统预期用途的潜在类别划分，与人工智能系统的提供者或用户利益相关，具有主观性，将造成法律上的不确定性。^③同理，中国《生成式人工智能服务管理暂行办法》第二十二条将生成式人工智能界定为“具有内容生成能力的模型”，已然预设了“防范通用人工智能风险”的基本立场。^④又如，“上海人工智能条例”偏重工具论，将人工智能归属于“理论、方法、技术及应用系统”之列，其产业发展原则强调“创新驱动和市场主导”；而“深圳人工智能条例”偏重功能论，其属概念比照人类智能，故其原则更加凸显“安全可控”。可见，不同样态和抽象程度的人工智能外延，实质上取决于立法者的价值倾向和立法目的。

三、重构人工智能的法律定义

（一）事物本质是“理性智能体”而非“类人”

若想通过下定义方式来明晰人工智能的法律意涵，需要揭示人工智能的本质特征或关系属性^⑤，而“事物本质”的理念要求立法者根据事理的不同对事物做不同的处理^⑥。这就意味着，在法律上界定人工智能不能完全受制于技术层面人工智能的物理本质，而是应当在“人与人工智能的应然关系”范畴内分析二者“事理之不同”。在判断“智能性”和列举“类型”时，应当以法律意义上的“理性人”而不是自然科学意义上的“生物人”为参照，确定法律所能容忍的智能性程度和法律所欲控制的人工智能类型。事实上，类人理论的“图灵测试”排除了那些不使用自然语言进行交流的智能系统，尽管这些系统也需要监管（例如，自动驾驶汽车）；那些未能实现其目标或者未能实现最大化性能度量的系统也被排除在外，尽管它们可能具有重大风险。^⑦况且，目前主流AI技术的进展无助于真正制造出具有AGI基本特点的智能机^⑧，现阶段的人工智能只具有模拟人类外显性智慧的能力，迭代发展至今的人工智能没有体现出“自

① [美]斯图尔特·罗素、彼得·诺维格：《人工智能：现代方法》（第4版），张博雅、陈坤、田超、顾卓尔、吴凡、赵申剑译，张志华审校，北京：人民邮电出版社，2023年，第2—5页。

② P.M.Krafft, Meg Young, Michael Katell, Karen Huang & Ghislain Bugingo. (2020). Defining AI in Policy versus Practice. In Machinery (Eds.), *AIES '20: Proceedings of the AAI/ACM Conference on AI, Ethics, and Society* (72-78), New York: Machinery.

③ Hannah Ruschemeier (2023). AI as a challenge for legal regulation—the scope of application of the artificial intelligence act proposal. *ERA Forum*, 23, 361-376.

④ 《国家互联网信息办公室有关负责人就〈生成式人工智能服务管理暂行办法〉答记者问》，中国网信网，http://www.cac.gov.cn/2023-07/13/c_1690898326863363.htm，2023年8月17日。

⑤ 温惊雷：《定义性规范立法研究》，成都：西南财经大学出版社，2020年，第81页。

⑥ [德]卡尔·拉伦茨：《法学方法论》（全本·第6版），第525页。

⑦ Jonas Schuett (2019). *A Legal Definition of AI*.

⑧ 徐英瑾：《人工智能哲学十五讲》，北京：北京大学出版社，2021年，第82页。

主学习并独立运行”的隐性智慧能力^①。所以，“智能性”基准测试的阈值标准应当确定为理性基础上的“理想性能”而不是“人类行为保真度”，在人工智能法律定义中将“自主性程度”作为识别人工智能和类型划分的标准也为时尚早。较之于人工智能科学的其他研究范式，“理性智能体”进路还拥有四大优势：一是比“理性思维”的逻辑主义更具普适性，因为正确的推断只是实现理性的几种可能机制之一；二是更适合科学的发展，理性标准在数学上是明确定义且完全普适；三是理性标准能够被证明可否实现智能体设计，而模仿人类行为或思维的设计在很大程度上是难以证伪的；四是理性智能体最初建立在逻辑基础之上，为了实现特定目标而需要制定明确规划，但随着概率论和机器学习方法的发展，智能体能够在不确定性下作出决策以获得最佳期望结果。^②此外，在AI与法律的交叉研究领域，agent应当译为“行为体”，这就为AI中agent作为“行为体”在法学研究中涉及“法律行为”（act in the law 或 juristic act）时埋下了循名责实的线索，预留了名正言顺的空间。^③因此，将人工智能的法律本质确定为“理性智能体”，不仅符合科学发展规律，而且能够在法律规制和法学理论层面，为人工智能监管和人工智能法学研究留下能够接入既有法律制度谱系和法学理论体系的空间。

（二）理论构造以“系统—控制论”为支撑

在法律意义上界定人工智能，还需要通过法学上的构造“将制定法中现存的某种规则体或者行为模式整合到体系之中以产生无矛盾的脉络关联”，但检验某种构造理论的可靠性时，不能单纯按照自然科学方式将其“追溯到普遍的自然法则”，而是需要依据法学特有的方式——通常是“逻辑性衡量”和“目的性衡量”，并且“最终具有决定性的是从理论中获得的与规范效力有关的结论语句在事理上的妥当性”^④。那么，人工智能的法律定义究竟应当以何种理论为基础，才能实现行为整合和目的嵌入？“长期以来，有无目的一直是划分生物科学与技术科学的水分岭，而智能研究的新方向便是在两种科学中找到其共性，并将生物系统和技术系统连接起来；它们的目的性在负反馈的基础上得到统一，而控制论所讲的目的就是系统与环境间的反馈，也是一种智能表现的过程。”^⑤维纳开创的控制论在系统科学中引入休谟的“应然”范畴，促成系统论从“单纯探索客观规律”向“研究目的性并强调发挥能动性”发生重大转向。^⑥自维纳提出控制论以来，在学术界创造出贯穿“目的”与“行为”等特殊内涵的五组范畴与概念体系，主要包括：（1）状态、变换、过程——强调系统“内部关系”；（2）信息、反馈、控制——强调“主客体关系”；（3）输入、输出——强调“外部关系”；（4）施控、受控、控制系统——凸显“目的”与“关系”；（5）信息与信号——揭示“系统运行的载体”；同时，维纳也创造出控制论范式下的新科学方法：功能模拟、黑箱方法以及信息反馈。^⑦

之所以选择“系统—控制论”作为界定人工智能法律定义的理论支撑，主要原因有二：一是，控制论的基本范畴能够以较为合适的抽象程度覆盖人工智能全生命周期，并且对现存争议具有较强包容性，而“理性智能体”作为一种通用范式，也适用于“系统—控制论”原理。^⑧该范畴既能实现对主体间关系的整合、评估行为的风险程度（譬如作为主体概念的“施控者”“受控者”以及作为行为描述的“输入”“输出”等），还能顾及用户对人工智能的一般认知（譬如“输出”涉及对外部环境的影响、“功能模拟”也可以直观感受）。二是，责任分配的底层原理在某种程度上也能和“系统—控制论”相关范畴实现耦合，法律责任能够通过“目的性衡量”来合理确定。以“汉德公式”为例，“只有在潜在的致害者预防未来

① 曹伟：《技术迭代视角下人工智能的法律分析》。

② [美]斯图尔特·罗素、彼得·诺维格：《人工智能：现代方法》（第4版），第4页。

③ 寿步：《人工智能中agent的中译正名及其法律意义》，《科技与法律（中英文）》2022年第3期。

④ [德]卡尔·拉伦茨：《法学方法论》（全本·第6版），第554、566、567页。

⑤ 王黔玲：《控制论的智能观》，《社会科学研究》2000年第5期。

⑥ 高剑平、齐志远：《系统哲学思想史研究》，北京：科学出版社，2022年，第120页。

⑦ 高剑平、齐志远：《系统哲学思想史研究》，第122、132页。

⑧ [美]斯图尔特·罗素、彼得·诺维格：《人工智能：现代方法》（第4版），第5页。

事故的成本小于预期事故的可能性乘预期事故损失时，他才负过失侵权责任”^①。与之对应，控制论研究的“输出对输入的响应特性”表征因果关系，其具有模糊性和不确定性特征，而控制论研究正是为了精确描述这种不确定性以便确定地“趋达目标”，其中另一个关键范畴——基于环境响应的“反馈”——更是为了及时修正信息来增强算法处理的确定性，控制论所采用的复杂数学函数表达在概率论意义上能够实现相应的“预防成本”核算。因此，以“系统—控制论”原理为支撑界定人工智能的法律概念，契合“发展负责任的人工智能”^②治理目标，通过构建新型行为模式实现对人工智能全生命周期下相关主体的权责分配。

四、结 论

首先，就人工智能法律定义的属概念而言，可以从人工智能全生命周期厘清各个阶段的关系。人工智能并不直接指涉“科学或者学科”，其首先是一种信息处理“系统”，“程序、模型和机器”则是该系统的软硬件组成部分；对智能系统的科学研究或技术研发，称为人工智能科学，其范畴包括人工智能知识、原理、方法和技术等；科学成果转化、投产后，形成人工智能产业；而人工智能作为一种“权益”，受制于特定的政策背景和法学理论话语，具有较强的评价性色彩和逻辑上的后置性，不能成为界定人工智能法律定义的术语。此外，如果对人工智能科学进行立法，则着眼于研发端的科技伦理导向；如果对人工智能产业和人工智能产品进行立法，则着眼于市场端的生产激励、公平秩序维护和消费者权益保障；而“人工智能法”着眼于系统运行、维护端的社会性风险防控，只有“系统—控制论”的构造才符合人工智能立法需要。因此，“人工智能法”应当定位为人工智能领域立法的基本法、一般法，《生成式人工智能服务管理暂行办法》则是人工智能领域的专门立法，而“人工智能科技促进法”“人工智能产品质量法”等定性为科技法、产品质量法等领域立法的特别法较为适切。

其次，就人工智能法律定义的内涵而言，应当结合功能型定义模式和发生型定义模式，以期同时对“理性智能体”的“目标控制”和“行为控制”。具言之，人工智能是依托算力基础设施，通过控制系统的算法处理输入的数据，以软件或硬件等多元集成方式嵌入系统后输出，或者直接在具体场景中输出对人类某种理性功能的模拟状态，在环境中进行交互，并在目标约束下经过反馈修正，最终完成预设任务的信息系统。该种信息系统的法律本质为“理性智能体”，是界定其法律地位的主要依据。该定义的法学构造主要包括：网络、数据、算法、软件/硬件、目标/任务、反馈、输出等相互影响、互为制约的基本要素，以及这些要素在“系统—控制论”原理下的各种内生性或外显性的行为（譬如“黑箱”“集成”“功能模拟”“变换”等）。此时，制定法中既有的涉人工智能规范，能够被人工智能的一般性法律定义依据“系统—控制论”脉络实现彼此之间的有机规整。值得注意的是，《个人信息保护法》《数据安全法》《互联网信息服务算法推荐管理规定》等针对人工智能某一要素的立法，如果相关条款只是单纯分配数据权益、保护个人隐私或者施加数据安全义务等，则不属于“人工智能法”范畴；只有内嵌“系统—控制论”构造的条款——譬如《个人信息保护法》第二十四条对“利用个人信息进行自动化决策”的规制——才能作为“附属人工智能法”归入“人工智能法”的法律渊源。

最后，就人工智能法律定义的外延而言，既包括外观上模仿生物人但本质上是集成软硬件的“智能机器人”，也包括单纯的智能系统（如 ChatGPT）或硬件设备（如智能传感器），以及嵌入这些硬件的其他“智能体”（如自动驾驶汽车）；至于“强人工智能”“通用人工智能”能否纳入人工智能的类型，不适合通过“外延”列举方式来确定，而是应当通过价值判断在法律概念中为其保留裁量空间，也即“强弱”之分、“专用通用”之分并非立足于客观现实的规范性法律判断，而是基于未来技术发展的评价性

① 冯珏：《汉德公式的解读与反思》，《中外法学》2008年第4期。

② 《发展负责任的人工智能：新一代人工智能治理原则发布》，中华人民共和国科学技术部，https://www.safea.gov.cn/kjbgz/201906/t20190617_147107.html，2023年9月1日。

法学判断。即便人工智能的法律定义中没有直接列举这两种类型，也不会导致定义外延不周延，因为法律调整对象的确定还需要在“定义初筛”的基础上再结合立法目的、法律原则进行决断，这种“留白技术”是在立法的确定性和前瞻性之间、立法的安定性和执法司法的适应性之间寻求平衡的必要选择。

诚然，本文试图给定一个相对而言内涵确定、外延清晰的人工智能法律定义，但这并非一劳永逸的工作，也不意味着该法律定义就可以包罗万象、顺利“涵摄”一切涉人工智能法律现象；法律定义也具有时代局限性，其对社会现象的容纳度不仅取决于立法者的权力决断，也受制于技术变迁，但这并不能否定为人工智能下定义的可能性，因为科学、合理的法律定义可以通过法律解释方法增强自身的社会适应性，除非人工智能实现颠覆性突破并且引发全生命周期的根本性变革，此时人工智能的法律定义便需要动用立法手段予以修正以适配科技发展规律。

Legal Definition of Artificial Intelligence in the Context of Artificial Intelligence Legislation

CHEN Liang & ZHANG Xiang

(School of Artificial Intelligence and Law, Southwest University of Political Science & Law, Chongqing, 401120)

Abstract: Scientifically and rationally defining the legal definition of artificial intelligence is the logical prerequisite for the systematic legislation in the field of artificial intelligence. In an ontological sense, artificial intelligence is first an information processing system, and then there are social cognitions such as artificial intelligence science and artificial intelligence industry. From the perspective of connotation, artificial intelligence relies on the computing power infrastructure, processes the input data through the algorithm of the control system, embeds it into the system in multiple integrated ways such as software or hardware, and then outputs it, or directly outputs some rational functions of human beings in specific scenarios. It is an information system that simulates the state, interacts in the environment, and undergoes feedback correction under the constraints of the goal, and finally completes the preset task. The legal essence of this kind of intelligent system is “rational agent”. From the denotation perspective, artificial intelligence does not only include pure intelligent software or hardware, but also intelligent agents integrated and embedded in software and hardware. The distinction between “strong and weak” and “general and specific” is not suitable to be listed directly in the denotation. With the development of technology, it is necessary to decide whether to include it in the legal meaning of artificial intelligence through evaluative legal judgment.

Keywords: Legislation on Artificial Intelligence, Legal Definitions, Factoring out Common Elements, Rational Intelligent Agent, System-Control Theory

[责任编辑：陈慧妮]