引用格式：李刚，葛子豪．从指标到体系：中国智库评价的逻辑演进、实践困境与对策建议．中国科学院院刊，2026, 41(5): 951-964, doi: 10.3724/j.issn.1000-3045.20260403007.
Li G, Ge Z H. From indicators to systems: Logical evolution, practical challenges, and policy recommendations of think tank evaluation in China. Bulletin of Chinese Academy of Sciences, 2026, 41(5): 951-964, doi: 10.3724/j.issn.1000-3045.20260403007. (in Chinese)

从指标到体系：中国智库评价的逻辑演进、实践困境与对策建议

李 刚* 葛子豪

1 南京大学 信息管理学院 南京 210023
2 南京大学 中国智库研究与评价中心 南京 210023

摘要 《关于加强中国特色新型智库建设的意见》印发至今已逾10年，中国智库数量持续扩张，评价活动也随之从自发性探索走向制度化建设。文章从智库建设的供给与需求两端审视评价问题，系统梳理10年来中国智库评价的实践历程与理论探索。在辨析智库评价概念内涵的基础上，提出评价的关键不在指标设计，而在于厘清评价的初衷与目的。研究发现，当前评价体系理论前提尚未充分厘清，统一化指标难以适应智库类型多样性与发展阶段差异，年度考核的短周期导向忽视了数据积累、方法论储备、人才培养等隐性基础能力的建设，评价过程中的单向运作削弱了评价的诊断与建设功能等问题。针对上述问题，文章从拓展评价功能定位、融合关键绩效指标（KPI）和目标与关键成果（OKR）管理理念、引入回应式评价的协商机制、构建差异化动态指标权重体系等维度提出优化路径，并提出中国智库评价应从追求单一的完美指标转向构建多方协同的评价生态体系，以评促建、以评促改，服务于中国特色新型智库的高质量发展。

关键词 中国特色新型智库，智库评价，回应式评价，治理模式，绩效管理

DOI 10.3724/j.issn.1000-3045.20260403007
CSTR 32128.14.CASbulletin.20260403007

*通信作者
修改稿收到日期：2026年5月13日

中共中央办公厅、国务院办公厅印发的《关于加强中国特色新型智库建设的意见》（以下简称《意见》）明确将推动形成定位明晰、特色鲜明、规模适度、布局合理的中国特色新型智库体系列为战略目标，并要求建立与之相配套的评价机制①。10多年来，国内智库数量持续扩张，党政机关、社会科学院（以下简称“社科院”）系统、高校、企业等不同类型的智库机构相继兴起，智库评价活动也随之从自发性探索走向制度化建设。站在承上启下的重要节点，系统梳理这一时期智库评价的实践历程、审视其中积累的问题，既是对过去10年制度建设的客观总结，也为下一阶段评价体系的优化提供依据。

1 智库评价的内涵与价值

讨论中国智库评价的实践历程之前，有必要先厘清一个基础性问题，即智库评价本身究竟在评什么、为何而评。本节尝试对智库评价的内涵与功能加以系统梳理，为后续的实践回顾与问题分析提供基本的分析框架。

1.1 智库评价的概念与类型

在厘清智库评价的内涵之前，必须先明确智库本身的定义边界。《意见》将中国特色新型智库界定为以战略问题和公共政策为主要研究对象、以服务党和政府科学民主依法决策为宗旨的非营利性研究咨询机构，从研究对象、组织属性和功能3个维度确定了中国特色新型智库的基本特征，构成智库评价的对象前提①。

*关于智库评价的概念，学界已有若干界定。*从评价的内容来看，智库评价是确定智库在政策开发、分析、决定及评估等各环节对政策议程发挥了何种影响力的过程[1]；从评价的作用来看，有学者将其理解为对智库本身及其成果、影响力等因素的评估，是促进智库共同体内部会话与沟通，推动智库工作不断创新的重要途径[2]；也有学者结合评价主体和标准进一步提出，智库评价是评价主体根据一定的评价标准和评价指标，对智库组织、智库成果、智库影响力、智库人才等因素进行全方位全过程的系统评价过程[3]。综合以上论述，本文将智库评价界定为，评价主体依据特定的评价标准和方法体系，对智库的研究能力、成果产出、政策影响力及组织管理状况进行系统性价值判断，以服务于管理决策、激励引导和公共问责等多重目标的评估活动。

*智库评价活动的核心在于价值，价值是评价的前提，任何评价行为都建立在某种价值观的基础上。*价值是客观存在的，但评价主体对价值的认识却带有主观性。正因如此，不同的主体面对不同的客体可能依据不同的价值观念，得出不同的评价结论。从评价主体的角度来看，智库评价大体可划分为3种类型，三者在功能定位、服务对象和运作逻辑上各有侧重。

*第一方评价，即智库内部自我评估与绩效管理，其功能定位是质量管理和经验教训总结。*智库作为知识生产组织，需要在日常运营中持续检视自身的研究方向是否契合战略规划，团队协作和资源配置是否合理，已完成项目的经验和不足是否得到系统梳理。内部评价的价值在于帮助机构对过往工作进行审视，使智库工作的质量改进成为一个持续推进的过程。

*第二方评价，即上级行政主管单位开展的监管性考核，其功能定位是投入产出评估和行政问责。*各级宣传部、哲学社会科学界联合会、哲学社会科学规划办公室等智库业务的主管部门，向下属智库拨付建设经费和运行资金，因此有对资金使用效果进行评估的必然需求。第二方评价回答的核心问题是“投入了公

① 中共中央办公厅、国务院办公厅印发《关于加强中国特色新型智库建设的意见》. (2015-01-20)[2026-02-14]. https://www.gov.cn/zhengce/2015-01/20/content_2807126.htm.

共资源，智库到底产出了什么、产生了多大效果”，这类评价直接关联经费分配与机构存续，是当前中国智库评价体系中约束力最强的一类。在党管智库的要求下，政治方向的正确性是第二方评价中具有一票否决效力的底线标准。

第三方评价，即独立学术机构开展的外部评价，其功能定位是回应社会对智库公共属性的知情期盼。智库作为使用公共资源的非营利性机构，社会公众有权了解其运营状况、研究产出和公益性职责的履行情况，第三方评价以南京大学中国智库研究与评价中心的中国智库索引（CTTI）测评体系、中国社会科学评价研究院的“吸引力-管理力-影响力”（AMI）指标体系、上海社会科学院智库研究中心的影响力评价等为代表，不依托行政隶属关系，评价方法与指标体系相对独立，发布结果面向社会公开，在方法论探索和评价框架构建方面最为活跃。

一个运转良好的评价生态应当是三方各司其职、协同互补的格局。第一方评价为智库提供持续的内部质量改进机制，第二方评价确保公共资源的使用效果接受监督，第三方评价为社会提供观察智库公益性的窗口，三者结合为智库发展创造一个既有约束又有激励、既有监督又有空间的制度环境。各方评价也应存在一定边界，第三方评价具有社会监督功能和学术价值，但如果数量过多、榜单泛滥，反而可能消耗智库的精力，助长围绕排名的恶性竞争，偏离了评价服务智库发展的初衷；第二方评价作为行政监管的重要手段，其约束力最强，但如果过于依赖关键绩效指标（KPI）导向的量化考核，就容易陷入“指标至上”的陷阱。

1.2 评价的价值应当从智库建设的供给与需求两端加以审视

理解智库评价的功能和价值，不能仅从评价系统自身的技术逻辑出发，而应当将评价置于智库建设的整体格局中加以考察。智库评价的根本目的，是促进智库更好地履行知识供给职责，提升对国家治理和社会发展需求的回应能力。因此，评价的功能定位应当从供给与需求两端来审视。

从需求侧来看，党和政府的科学民主依法决策对高质量的政策研究和知识产品存在持续性需求，社会公众对公共政策的知情权和参与权也要求智库承担起面向公众的知识传播和舆论引导职能。评价活动的设计，应当以这些真实的知识需求作为参照基准，考察智库是否在咨政建言、理论创新、舆论引导、社会服务、公共外交等方面有效回应了决策体系和社会发展的知识期盼。

从供给侧来看，智库作为知识生产组织，自身在研究能力建设、人才培养、数据积累和方法论储备等方面存在持续改进的需求。评价对于供给侧的价值，在于帮助智库准确识别能力短板和发展方向上的偏差，提供有针对性的诊断信息。如果评价仅仅停留在排名和打分层面，而无法向被评价机构反馈具有建设性的改进建议，评价的供给侧功能就难以实现。

因此，评价体系建设的关键问题，不只是设计出一套精密的指标，也是厘清评价的初衷和目的。每一类评价都应当清楚自身服务的对象和所要回答的核心问题，在此基础上选择与之匹配的方法和指标，而不是用一套标准化的框架去回应所有评价需求。从指标到体系的转变，首先是从“以指标为中心”到“以评价目的为中心”的认识转变。

2 中国智库评价的实践回顾与比较分析

10多年来，中国智库评价实践在第一方、第二方和第三方3个层面均有不同程度的展开。第一方和第二方评价均主要以建设水平和产出水平为指标进行评价，第一方的内部评价主要服务于机构的日常管理和人员激励，评价结果通常仅在机构内部流转；第二方的行政主管部门考核以绩效指标为导向，直接关联经费拨付与机构存续，是当前约束力最强的一类评价形

式，但指标设计和结果认定均由主管方单方决定，智库在这一过程中缺乏实质性的参与空间。

相较而言，第三方评价以学术机构为主体，不依托行政隶属关系，评价方法与指标体系相对独立，发布结果也面向社会公开，因而在方法论层面的探索最为活跃，在评价框架的自觉构建与反思方面积累了较多可供分析的实践素材，是理解中国智库评价演进逻辑的重要切口。

自2014年以来，上海社会科学院智库研究中心[4]、中国社会科学评价研究院[5]、四川省社会科学院与中国科学院成都文献情报中心[6]、南京大学中国智库研究与评价中心、浙江大学信息资源分析与应用研究中心②、清华大学公共管理学院智库中心[7]③等机构相继发布智库第三方评价报告，国际上美国宾夕法尼亚大学麦甘（James G. McGann）团队的《全球智库报告》亦对国内智库评价实践产生了重要的参照影响④。

2.1 评价思路上呈现主观性与客观性的两极分化

*在评价思路上，各家机构大致沿一条从专家主观评议向客观数据驱动演进的谱系分布，但具体路径选择存在明显差异。*宾夕法尼亚大学麦甘报告以邀请全球新闻记者、政策制定者、捐助人及各领域专家对智库进行打分排序为核心方法，评价结果高度依赖参与者的主观印象，同一机构在不同年份的排名波动较大，评价的可重复性受到学界的持续质疑。上海社会科学院智库研究中心的“4+1”评价模型同样采用问卷调查与专家评议相结合的方法，邀请政府官员、研究人员、媒体记者等多类群体参与打分，结果难免受到评价专家群体的构成与认知偏向的影响，但其通过实地考察的形式弥补了这一方面。

*随着评价实践的推进，主客观相结合的方法路径逐渐成为主流。*中国社会科学评价研究院构建AMI体系，综合运用实地走访、问卷调研与数据采集等多种渠道；四川省社会科学院则依托“中华智库研究大数据平台”采用互联网信息抓取技术，进一步减少了对专家主观判断的依赖，两者均在方法论客观化方面较早期的专家打分模式有所推进。南京大学CTTI的“治理-资源-成果-活动-影响力”（MRPAI）测评体系以自主填报加后台审核的众筹机制采集数据，兼顾了信息覆盖的广度与来源的多元性，但评价结果在一定程度上受制于各智库提交信息的完整性与准确性。浙江大学信息资源分析与应用研究中心的“资源-影响力-公共形象产出”（RIPO）指标体系代表了方法论客观化的另一个极端，以“数据公开、面向世界、评价透明、计算可重复”为设计原则，所有3级指标均来源于机构财报、年报、Web of Science数据库、中国知网（CNKI）等可查询的公开途径，完全排除专家评议，在方法论透明度和可重复性方面具有优势，但当采集的数据无法完全解释指标时，会产生不符合业界认知的评价结果。清华大学公共管理学院的《中国智库大数据报告》以社交媒体平台的传播数据为评价依据，与RIPO体系同属全客观数据驱动的路径，但仅采集社交媒体这一单一维度数据，也导致评价结果缺乏说服力。

从整体来看，主观评议路径能够调动业内人士对声誉与学术地位的经验判断，但评价结论的稳定性高度依赖参与者构成；客观数据路径在可重复性方面更具说服力，但受制于公开数据的可获取范围，难以触及机构运营的隐性层面。

② 浙江大学信息资源分析与应用研究中心. 全球智库影响力评价报告(2017). 杭州: 浙江大学信息资源分析与应用研究中心, 2017.

③“中国智库大数据评价研究”课题组. 中国智库大数据报告(2016). 北京: 清华大学公共管理学院, 2017.

④ McGann J G. 2020 Global go to think tank index report. Philadelphia: University of Pennsylvania, Think Tanks and Civil Societies Program. [2026-03-08]. https://repository.upenn.edu/think_tanks.

2.2 指标维度存在“影响力”和“资源产出比”2种测度重点

在指标维度的构建逻辑上，各机构形成了影响力测量、资源与产出并重考察的测度重点，不同层次之间的评价逻辑存在实质性差异。指标维度的选择反映了评价设计者对“什么是好智库”的理论预设，不同的理论视角驱动着不同的指标架构。

上海社会科学院和四川省社会科学院的评价体系：其在维度构成上最接近纯影响力导向。上海社会科学院的“4+1”体系借鉴约翰·加尔通（J. Galtung）的社会结构理论，结合中国智库影响力围绕决策核心由近及远呈圈层结构分布的经验特征，将影响力拆解为决策、学术、社会、国际4个维度，其中决策影响力居于核心圈层，学术和社会影响力构成中间层，国际影响力属于外围层。“智库成长能力”虽有向内延伸的意图，但在实际操作中仍以外部可感知的产出为主要依据，整体上属于对显性影响力的多维度分解。四川省社会科学院的评价指标体系以系统理论为基础，将智库视为具有特定功能和构成要素的组织系统，通过解析智库的系统特性和重要功能来识别影响力的生成机制，在此基础上综合运用理论分析法和频度分析法，构建了决策、专业、舆论、社会、国际影响力5个维度，依托大数据平台强化了对新媒体传播渠道的追踪，使舆论影响力的测量更为细化，但5个维度均属外显影响层面，研究过程、资源积累等基础性内容未进入评价视野。两者的评价逻辑均以智库对外产生的可见影响作为唯一的考察维度，评价结论呈现了智库影响力的宽度与活跃程度，但并未体现机构的综合研究能力。

浙江大学信息资源分析与应用研究中心的RIPO体系：以信息计量学方法为理论基础，强调评价数据的公开性和评价过程的可重复性。其将计量学中对学术期刊和科研机构的评价思路迁移到智库领域，在影响力测量的基础上引入了资源（R）维度，形成了以影响力为主、辅以资源端指标的混合结构。资源维度的设置有向机构投入端延伸的意图，但由于全客观数据驱动的方法论约束，资源维度所能采集的内容主要是机构预算、人员规模等可公开统计的外部信息，机构在研究方法论积累、数据库建设、人才培养机制等方面的实际状态，仍无法通过现有公开数据渠道加以捕捉。从指标体系的整体重心来看，知名度（P）和影响力（I）占据主导，资源维度的贡献相对有限，RIPO体系仍属以影响力为主的评价范式，资源指标的加入更多扮演补充性角色。

中国社会科学评价研究院的AMI体系和南京大学CTTI的MRPAI框架：其是各家评价机构中指标覆盖最为均衡、向综合能力考察延伸最为明显的两套体系。AMI体系以“吸引力—管理力—影响力”三力理论模型为架构，试图从智库运营的完整链条出发构建评价逻辑。其中“吸引力”（attractive power）维度考察智库获取声誉、人才和资金等关键资源的能力，反映智库对外部环境的资源汇聚水平；“管理力”（management power）维度考察管理者配置和运用资源推动机构发展的能力，聚焦组织结构、制度规范和战略执行等内部运营环节；“影响力”（impact power）维度则是吸引力和管理力的最终外在体现，考察智库的政策影响、学术贡献和社会传播效果。3个维度之间形成了从资源获取到内部转化再到外部产出的递进关系，在维度层次上的推进较为明显。南京大学CTTI的MRPAI框架，以“投入—过程—产出”的分析逻辑为理论基础，智库的运行可以沿“制度保障—资源投入—成果产出—活动扩散—社会影响”的链条加以理解，从治理结构（M）、智库资源（R）、智库成果（P）、智库活动（A）、媒体影响力（I）5个维度展开，治理结构维度尝试回应智库运作规范性这一长期被忽视的评价盲区，智库资源维度考察机构的知识储备与人力投入，智库活动维度将研究产出之外的知识传播与社会互动纳入考察，整体呈现出过程与结果并重的

评价取向，在现有国内评价实践中对机构综合能力的覆盖最为系统。宾夕法尼亚大学麦甘报告虽以专家评议为核心方法，但在指标设计上采用了“资源—效用—产出—影响”四维分析框架，明确区分了产出与影响2个层次，从维度构成来看AMI和MRPAI体系的综合能力取向与其较为接近，只是由于方法论上高度依赖专家主观打分，其指标设计的实际落地效果受到较大限制。

总体来看，各个体系在显性产出和外部影响维度上的覆盖已相对完善，但有一类能力在所有评价体系中普遍缺位，数据库积累的规模与质量、研究方法论储备的厚度、长期田野调查网络的建设，以及支撑这些工作的组织学习机制与知识管理体系成为“冰山以下”的隐性基础能力，既无法从公开数据渠道精准采集，也难以通过专家打分加以准确判断，导致各类评价体系无论在方法论上如何取向，最终所呈现的都主要是可观察、可统计的显性能力，而那些支撑智库长期竞争力的基础性积累，在现有评价框架中依然处于覆盖缺失的状态。

2.3 智库评价的持续性与制度化生态化水平存在明显差异

从各机构评价活动的持续时间与发布规律来看，10年间智库评价格局经历了明显的分化。宾夕法尼亚大学麦甘团队的《全球智库报告》是国际上持续时间最长、影响最广的智库评价项目，但在2021年麦甘本人辞世后，该报告未再延续，10多年的连续报告发布画上了句号。四川省社会科学院与中国科学院成都文献情报中心联合发布的《中华智库影响力报告》自2015年起持续发布，至2021年共发布6版，此后未再次更新。清华大学公共管理学院智库中心的《中国智库大数据报告》2017—2020年共发布4版；浙江大学信息资源分析与应用研究中心自2017年起每年发布《全球智库影响力评价报告》，至2022年共连续发布6年，此后亦暂未见新一轮成果。清华大学与浙江大学均设有专门的研究机构和稳定的团队，评价工作具备持续运作的组织条件，但在项目安排上并未将持续发布智库评价报告作为工作重点。

相较而言，上海社会科学院智库研究中心、中国社会科学评价研究院和南京大学中国智库研究与评价中心，在评价工作的连续性与制度化程度上走得最远。上海社会科学院智库研究中心自2014年起每年发布《中国智库报告》，至2024年已发布第10本，并在2017年推出英文版，持续性与规律性在国内第三方评价机构中较为突出。

中国社会科学评价研究院的AMI综合评价以约4年为1个周期推进，持续推进中国智库综合评价研究项目，2025年公布项目评审结果，共产生“AMI智库入库单位”556家、“中国特色新型智库建设十年回顾特色案例”64个及“中国特色新型智库建设十年回顾参考案例”60个；该院还设有智库建设与评价高峰论坛、“智库评价日”等配套机制，将周期性综合评价与学术交流活动相结合，形成了较为完整的评价工作体系。南京大学中国智库研究与评价中心围绕CTTI平台建立了多层次的评价机制，2016—2019年每年发布《CTTI智库报告》，以及2020年与2022年各发布一版。在评价的基础上，还推出来源智库认证、高校智库百强榜、智库最佳案例与优秀成果推介等配套工作，将评价、认证与荣誉供给3种机制相互配合，形成了国内第三方评价实践中生态化程度较高的工作格局。

从上述对比来看，各机构在评价工作的持续性上存在明显差异，但成因各不相同。部分评价项目因核心推动者的变动而中断，部分则因阶段性任务完成而暂告一段落。能够长期持续运作的评价，大多依托专门的研究机构，并在指标迭代、数据积累和评价形式多样化上形成了持续投入的能力。评价的制度化程度越高，越能在智库共同体中积累稳定的信誉和参照价值；评价形式越多元，就越能通过认证、荣誉供给等

机制与被评价机构建立长期的信息交换关系，逐步形成更加良好的评价生态。

3 中国智库评价面临的现实困境与矛盾

回顾10年实践可以发现，中国智库评价体系在制度化程度和覆盖范围上均取得了显著进展，但评价活动存在的问题也在实践推进中逐渐显露。

3.1 智库评价的理论前提尚未充分厘清

任何评价活动都以评价对象的清晰界定为逻辑起点，而智库评价在这一基础环节上就面临着理论层面的困难。西方学界普遍强调智库的独立性、非营利性和政策导向性，但对这些特征的具体理解因国别制度环境不同而存在显著差异。《意见》从研究对象、组织属性和服务宗旨3个维度给出了工作性定义，但在操作层面，党政智库、高校智库、社会智库、企业智库之间在组织形态、运行机制和资源获取方式上的差异极为悬殊，将不同类型的机构群体纳入同一评价框架加以衡量，本身的合理性就存在问题。概念边界的模糊也会影响评价实践，导致评价标准的制定缺乏稳固的逻辑基础。

在评价内容方面，智库评价究竟应当评价什么，也是一个需要探讨的问题。从已有的评价实践来看，评价内容大致包括研究成果产出和政策影响力评价，以产出为中心的评价倾向于将论文数量、研究报告、获批课题等作为核心指标，操作简便但容易将智库价值等同于可计量的文本产品；以影响力为中心的评价最贴近智库的社会功能，但影响力的测度至今仍是评价方法论中的难题。智库的影响力具有滞后性、间接性和弥散性等特征。以政策影响力为例，政策建议从提出到被决策者采纳往往存在较长的时间差，且影响路径通常经由学术讨论、媒体传播、官员认知形成等多重中间环节发生作用，难以在因果关系上直接归因于某一机构的特定研究成果。当前国内外的影响力评价普遍依赖问卷调查与文献计量相结合的间接方法，以代理指标替代直接观测，这种方法在学理上存在效度不足的风险，评价结论反映的可能是智库的可见度和传播活跃程度，与其对政策过程的实际作用之间仍有相当距离。

3.2 统一化的评价标准难以适应智库的多样化发展

当前大多数智库评价体系采用统一的指标框架衡量所有类型的智库，未能充分考量不同智库在组织属性、功能定位和发展阶段上的内在差异。党政智库处于行政系统内部，距离决策层相对更近，研究成果以内部报告为主要形式，评价的核心标准是对决策的直接贡献程度；高校智库以科研人员为主体，学术生产逻辑与政策咨询逻辑并行交织；社会智库以市场化方式运营，经费来源与成果传播渠道均有别于前2类；企业智库则将服务公司作为重要基本职责，研究选题与报送机制各有差异。用一套指标框架统一评判上述类型各异的机构，必然使部分智库处于劣势，而另一些机构则凭借体制机制便利在指标得分上占据优势，评价结果并不能呈现研究能力与政策贡献的真实差距，而是反映了机构属性与指标设计之间的契合程度。

从智库的生命周期来看，指标统一化往往也忽视了不同阶段智库的特点。一般而言，初创期的智库优先需要的是基础设施建设、数据积累和研究方向的明确；成长期的智库开始形成稳定的研究产出，咨政成果逐步增加；成熟期的智库则依靠品牌效应和知识存量发挥影响力，知识积累、文化积淀与制度传承成为核心竞争力。然而，现行评价体系对上述阶段性差异几乎没有体现，无论机构成立几年、规模大小、历史积累深浅，都按照同一套标准进行打分排名。初创阶段的机构往往将大量资源投入到基础性工作中，短期内产出有限但长远潜力可观，统一标准下的低分排名既无法准确反映其真实价值，也可能对其资源获取和发展空间产生负面影响。

标准统一化也导致了评价实践中对智库功能的认

识过于狭窄，在相当程度上将智库等同于政策研究组织（PRO）来考核。从广义上讲，政策研究组织属于智库的范畴，承担着智库最基本的咨政建言职能；但从狭义上看，智库的功能远比政策研究组织丰富。《意见》明确指出，中国特色新型智库在咨政建言、理论创新、舆论引导、社会服务、公共外交5个方面发挥重要功能。应当承认，当前阶段国内大多数智库机构实际承担的核心职能仍然集中在政策研究和咨政建言方面，舆论引导、社会服务和公共外交等功能在多数机构中尚未充分发挥。针对以政策研究为主要职能的机构，评价重心落在研究产出和咨政实绩上是合理的，但对这类机构最具评价效力的第二方评价虽然能够获取内部咨政数据，但评价结果通常不对外公开；而第三方评价机构面向这类以内部报告为主要成果形式的机构开展评价时，往往无法获取核心的咨政产出数据，只能依赖公开发表的论文、媒体引用等间接指标进行替代性测量，评价结论的可信度因此打了折扣。

与此同时，对于那些已经在理论创新、社会服务或公共外交方面形成特色的智库，现行评价体系同样缺乏与之匹配的评价维度。5项功能中仅有咨政建言得到较为系统的评价覆盖，理论创新和舆论引导的衡量方式仍较为粗放，社会服务和公共外交则几乎未进入评价视野。

从智库发展定位的宏观层面来看，《意见》确立的5项功能是国家对新型智库的整体期待，也是智库区别于一般性政策研究组织的核心标志。评价体系如果长期仅围绕咨政建言这一单一维度展开，实际上是在用政策研究组织的标准衡量智库，无法引导智库在理论创新能力、公共传播能力和国际对话能力等方面形成差异化发展格局，也就无法回应国家治理对多元化知识供给的深层需求。评价框架与智库完整功能谱系之间的覆盖缺口，使被评价机构自然将资源向被考核的方向集中，不同类型智库在评价压力下趋于同质化，压缩了智库生态的多样性和整个体系的创新潜力。

3.3 年度考核的短周期评价忽视了智库“冰山以下”的能力

行政主管部门和高校对智库的考核周期大多以年为单位，经费拨付往往附带明确的绩效指标约束，要求被考核机构在规定时间内产出可量化的成果，资源配置的重心自然向短期内能够完成并形成文字记录的工作倾斜。政策咨询报告、公开发表的论文、媒体采访记录、获批立项数量等，均属于当期可统计的产出，因而获得较为充分的制度激励；数据库开发、调查网络建设、方法论积累、研究人才的系统培养等工作，则因周期长、见效慢，难以在考核年度内转化为分数，长期处于相对低优先级的状态。

智库研究与政策咨询本身具有较长的转化周期。重大议题的研究往往需要持续数年的数据跟踪和理论积累，政策建议从形成到被采纳、从采纳到产生可观测效果，往往需要相当时间。年度考核机制将评价窗口压缩到12个月，客观上使智库在选题策略上倾向于时效性强的热点议题，期望快速产出可见度高的成果，导致专业化、基础性、开拓性的中长期研究则缺乏持续投入的制度保障，缺少长线的隐性能力提升。将冰山模型应用于智库评价分析，智库的能力结构区分为“冰山以上”的显性部分和“冰山以下”的隐性部分[8]。显性部分是可观察、可测量的知识和技能，包括研究报告、政策建议、政府批示、学术论文、媒体引用和获奖情况等，可见性强、可量化程度高，构成现行评价体系的主要考察对象；隐性部分则是价值观念、内在动机和深层特质，包括数据库建设与维护、研究方法论储备、人才培养与团队凝聚力、机构制度建设等，支撑着智库的长期竞争力，但由于难以在短期内转化为可观测成果，往往在年度绩效考核中受到忽视。当前评价体系过度集中于“冰山以上”部分，对“冰山以下”的隐性积累关注不足，是导致智

库评价偏短期、偏显性的重要制度原因。

3.4 评价过程中的单向运作削弱了评价的建设功能

评价理论的发展大致经历了4个代际，第1代评价以测量为标志，将评价等同于标准化测试工具的运用，评价者的工作主要是选择工具、组织测量、提供数据；第2代评价以描述为标志，泰勒的工作奠定了将教育目标与成果加以系统描述的基础；第3代评价以判断为标志，评价者开始将价值判断纳入评价任务[9]。Guba和Lincoln[10]于1989年提出第4代评价理论，主张评价应从回应各方利益相关者的诉求出发，通过评价主体与客体之间的协商对话，形成共同建构的价值共识。就中国智库评价实践而言，无论是行政考核，还是第三方排名，当前的运作模式在相当程度上仍停留于第2代或第3代评价的范式框架内，以预设目标或专家判断为主轴，尚未将智库自身的发展诉求和利益相关方的多元价值纳入评价过程的核心设计。

在这种单向评价的模式下，评价主体制定指标、采集数据、发布结果，被评价的智库机构处于被动接受的地位，在评价指标的制定过程中缺乏实质性参与，对评价方法与结果呈现方式也基本没有协商空间。第二方评价的行政考核体系中这一特征尤为明显，指标框架由主管部门决定，智库的任务是提供数据并接受结论；第三方评价机构虽然在体制上相对独立，但在评价方案设计和结果解释上同样较少引入被评价方的反馈与回应。这种缺乏双向沟通的评价格局，使得评价过程难以真正理解被评价机构的处境、优势与面临的约束，评价结论的针对性和改进参考价值也因此受到限制。

此外，单向评价模式的评价结果主要服务于排名公布和资源分配决策，改进导向的功能相对薄弱。智库从评价中获取的主要信息，不应是自身在榜单上的相对位次，而是针对其具体发展短板的系统性分析和改进建议。从评价发挥建设性功能的角度来看，一个理想的评价过程应当帮助被评价机构更清晰地认识自身的优势与不足，并就改进路径形成有参考价值的指引。

4 中国智库评价体系优化的对策建议

优化评价体系，既需要在理念层面拓展对评价功能的认识，也需要在方法论层面引入更具弹性的评价工具，还需要在制度设计层面为多元主体的协商参与提供正式渠道。

4.1 拓展智库评价的认知边界与功能定位

当前智库评价实践中存在一种功能认知上的收窄，即将评价等同于排名和行政考核，将评价目标窄化为监管与控制。因此，有必要从概念层面重新审视智库评价的外延与功能定位。就外延而言，完整的智库评价体系应当涵盖3个层次：智库机构内部的自我评估与绩效管理、行政主管部门实施的监管性考核，以及第三方学术机构开展的外部评价。3个层次各有侧重，内部评价服务于机构战略调整与资源分配，行政考核服务于公共问责与政治方向把控，第三方评价服务于社会认知与横向比较，三者应协调运作，共同构成立体的评价生态体系。就功能定位而言，评价应当从单一的监管考核功能走向监管、激励、诊断与问责的有机统一。监管功能确保智库坚守政策研究的专业方向，对资金使用负责；激励功能通过正向激励引导智库提升研究质量，鼓励创新探索；诊断功能帮助被评价机构识别发展短板，提供有针对性的改进参考；问责功能则使社会公众能够了解使用公共资源的机构如何履行其知识供给职责。

建议借鉴美国非营利组织的透明化管理经验，出台智库注册和管理相关法规，着重推动智库信息公开制度建设。美国《国内税收法》相关条款要求免税非营利组织每年向美国国税局提交“990表”，完整披露机构收入来源、主要支出项目、高管薪酬、董事会成

员等核心财务信息，并向公众公开查阅⑤。中国智库大多使用公共资源，向社会公众承担相应的问责义务，但目前信息公开程度普遍偏低。推动智库建立年度报告制度、定期公开财务与运营状况，不仅是回应公共问责、构建社会信任的必要条件，也是第三方评价机构在客观评价层面获取可靠数据的重要来源。透明度建设与评价机制的完善相互支撑，才能共同构成智库治理和评价现代化的制度基础。

4.2 推动评价方法从单一考核走向多元融合

关键绩效指标（KPI）以组织战略目标的层层分解为逻辑起点，通过自上而下的指标分配，将组织目标拆解为部门和个人层面的可量化考核标准，强调结果导向和组织控制，考核结果通常与薪酬激励直接挂钩[11]。KPI的优势在于目标明确、操作简便，适合规律性、程序性的日常工作评价；其局限则在于容易使被评价者将精力集中于完成指标，而忽视指标之外的贡献，且对于难以量化的创造性、探索性工作难以公正评价。

在KPI体系之外引入目标与关键成果（OKR）的管理理念，是优化智库评价方法的重要路径。OKR则以目标导向为核心，强调管理者与员工共同确定具有挑战性的目标，关键成果服务于目标实现，绩效薪酬与OKR得分不直接挂钩，更注重团队内部的过程沟通与协商[12]。OKR适合以创新和探索为主的工作类型，但对管理者和员工的主观能动性有较高要求。智库的研究工作兼具2种性质：年度政策报告、常规调研等具有较强的规律性，适合KPI考核；大型数据库建设、跨学科研究体系构建、新研究方向的探索性投入等，则更适合OKR框架下的目标驱动管理。二者有机结合，能够为不同类型的智库工作提供与之相匹配的评价逻辑。

引入OKR理念还需要对评价周期作相应延长，当目标设定为3—5年的阶段性工作，评价窗口就不应局限于12个月，而应与工作的实际节奏相匹配。这有利于鼓励智库将资源投向基础性能力建设，对“冰山以下”的数据积累、方法论储备、人才梯队等进行持续投入，而不是年复一年地寻找当期可统计的显性产出。在评价形式上，有必要进一步丰富评价体系的层次结构，将荣誉供给型评价⑥和认证性评价⑦纳入整体设计。荣誉供给型评价和认证性评价不以排名为结果形式，有助于缓解排名导向所带来的竞争异化问题，为智库提供多元化的发展激励信号。

4.3 引入回应式评价理念增强评价过程中的对话协商

回应式评价（responsive evaluation）是第4代评价理论的核心主张，其基本立场是将评价界定为利益相关各方通过协商共同建构意义的过程[13]。将这一理念引入智库评价实践，需要对评价流程的若干关键环节进行调整。评价启动阶段，应通过与被评价智库的预先沟通，了解机构的发展定位、阶段性目标和面临的主要约束，将评价视角与被评价者的真实处境相结合，避免以统一框架误读特殊情境下的机构表现；在评价实施阶段，应建立初步结论的征求反馈机制，就评价方法和判断标准与被评价方进行交流，识别可能存在的理解偏差；评价结束后，应根据评价结论与机构共同商定改进路径，将评价成果转化为可操作的发展建议。

美国联邦资助研究与发展中心（FFRDC）的绩效

⑤ United States. Exemption from tax on corporations, certain trusts, etc. [2026-02-20]. https://www.law.cornell.edu/uscode/text/26/501; Internal Revenue Service. About form 990, return of organization exempt from income tax. [2026-02-20]. https://www.irs.gov/forms-pubs/about-form-990.

⑥ 设立并评选智库研究奖项，对具有代表性的优秀成果给予正向激励。

⑦ 对智库的资质条件、治理规范和运营能力进行认证，作为质量基准的标志。

管理实践可以为智库评价提供可参照的制度案例。美国兰德公司2022年发布的专题报告指出，有效的FFRDC绩效管理系统须具备若干关键前提，包括对机构优先事项和目标的清晰界定、信息在战略规划与绩效评估之间的双向流动，以及与各类利益相关方的有效沟通协商机制。报告进一步列出了建设性评估过程的操作标准，其中包括评估目的与范围须经由利益相关方共同界定、评分标准保持一致适用、证据来源须提供充分可靠的信息，以及在全面推行重大流程变更前先行开展试点验证等。在绩效问责机制上，FFRDC采用以沟通协商为核心的回应式评估方式，由资助机构每5年组织1次综合评审，通过专家组实地考察和多利益相关方参与，形成关于合同续签的综合判断，并设有年度绩效跟踪机制。这些做法体现了将评价对象纳入评价过程设计的协商精神，与回应式评价理论在实践层面高度契合。将类似的协商机制引入智库评价流程，需要克服评价主体与被评价对象之间的权力不对等问题，要在制度设计上主动改革，为智库机构参与评价方案的讨论提供正式渠道。

4.4 构建差异化、动态化的评价指标与权重体系

克服评价指标同一性带来的局限，需要在评价体系设计层面引入差异化原则，根据智库的类型属性和发展阶段，分别设置与之相适应的指标维度和权重结构。*在类型差异方面*，党政智库的核心贡献在于政策咨询质量和决策支撑能力，评价指标应向咨政成果的转化实效和政策研究的深度倾斜；高校智库兼具学术研究和政策咨询2种功能，评价指标应在成果质量与政策影响之间保持平衡；社会智库的运营逻辑与前两类差异较大，评价框架应更多考量其财务透明度、独立性维护和公共传播能力。*在阶段差异方面*，可以将生命周期分析框架系统引入指标设计，初创期机构侧重考察基础设施建设、治理结构完善程度和研究方向的清晰性；成长期机构侧重考察成果产出质量和资源运用效率；成熟期机构则将影响力广度、国际参与度和机构文化的传承能力纳入评价重点。

在指标维度的拓展方面，应根据智库的功能类型进行分类评价设计。当前大量智库机构的核心职能和智库评价的重点仍集中于政策研究和咨政建言，需要在理论创新、舆论引导、社会服务、公共外交方面完善相对应的考察维度，将理论创新成果的原创性和学术引领力、舆论引导的议题设置能力和公众触达范围、面向社会的专业培训和知识普及活动、国际学术网络建设和公共外交参与等纳入评价框架，使评价体系能够识别和激励智库功能谱系的全面发展，避免所有机构在单一的咨政产出维度上同质竞争。分类评价的前提是对智库功能定位的准确判断，评价设计方在启动评价之前应当充分了解被评价机构的功能重心，据此选择与之匹配的指标组合和权重结构，使评价真正发挥引导智库差异化发展的作用。

在权重设置方面，权重结构应具有弹性，不宜机械化地固定，而应根据评价的目的、情境和被评价机构的特殊性进行调整。权重的设定本身是一个价值判断过程，不同的权重分配反映了对智库功能的不同理解，应当通过利益相关方的协商来确定，避免由评价方单方面裁量。同时，可以探索引入开放式（不设上限）的评分机制，与封顶式评分相比，开放式评分有助于识别智库在某一特定领域的突出优势，使功能聚焦型机构的专业特色能够在评价结果中得到体现，不被综合加权平均所淡化。在指标的选取上，应适当向“冰山以下”的基础性能力倾斜，将数据库建设水平、调研网络覆盖、方法论储备的厚度和人才梯队的结构纳入评价视野，以此引导智库在中长期基础能力建设上加大投入[14]。

推进中国智库评价体系建设，归根到底要坚持以习近平总书记关于中国特色新型智库建设的重要论述为根本遵循，紧紧围绕服务党和国家决策、提升国家治理能力的目标展开，以评促建、以评促改，使评价机制成为推动中国特色新型智库高质量发展的制度保

障。中国智库建设的高质量发展，离不开智库评价体系与智库建设工作的深度匹配。任何单一的指标，无论技术上如何精细，都无法涵盖智库发展的全部面向。从指标到体系的转变，意味着评价优化的方向不是追求一套完美指标，而是构建由多方主体、多种形式、多个周期共同组成的评价生态体系。评价的最终目的，并不是限制和束缚智库的发展，而是实现正反馈的作用，帮助智库成为真正意义上的知识生产机构，从供给侧检视自身的能力短板，引导不同类型智库在咨政建言、理论创新、舆论引导、社会服务、公共外交等功能领域形成差异化发展格局，发挥公共理性对于治理质量的长期支撑作用。

参考文献

1 帕瑞克·克勒纳, 韩万渠. 智库概念界定和评价排名：亟待探求的命题. 中国行政管理, 2014, (5): 25-28.
Kellner P, Han W. The definition and evaluation ranking of think tanks: A proposition urgently to be explored. Chinese Public Administration, 2014, (5): 25-28. (in Chinese)

2 李刚. 建立智库全层次全要素评价体系. 决策探索(下半月), 2017, (2): 45-46.
Li G. Establishing a full-level and full-element evaluation system for think tanks. Policy Research & Exploration, 2017, (2): 45-46. (in Chinese)

3 汤建军, 郑代良, 黄渊基. 中国特色新型智库评价体系初探. 湘潭大学学报(哲学社会科学版), 2018, 42(4): 12-16.
Tang J J, Zheng D L, Huang Y J. A preliminary study on the evaluation system of new-type think tanks with Chinese characteristics. Journal of Xiangtan University (Philosophy and Social Sciences Edition), 2018, 42(4): 12-16. (in Chinese)

4 上海社会科学院智库研究中心. 2013年中国智库报告：影响力排名与政策建议. 上海：上海社会科学院出版社, 2014.
Think Tank Research Center of Shanghai Academy of Social Sciences. 2013 China think tank report: Influence ranking and policy suggestions. Shanghai: Shanghai Academy of Social Sciences Press, 2014. (in Chinese)

5 荆林波. 全球智库评价报告(2015). 北京：中国社会科学出版社, 2016.
Jing L B. Global think tank evaluation report (2015). Beijing: China Social Sciences Press, 2016. (in Chinese)

6 方茜, 廖冲绪. 中华智库影响力报告(2015—2017年). 北京：人民出版社, 2018.
Fang Q, Liao C X. China think tank influence report, 2015-2017. Beijing: People's Publishing House, 2018. (in Chinese)

7 "中国智库透明度研究"课题组, 朱旭峰, 韩万渠. 中国智库透明度报告. 智库理论与实践, 2017, 2(1): 63-82.
Research Group of China Think Tank Transparency Research, Zhu X F, Han W Q. China think tank transparency report. Think Tank: Theory & Practice, 2017, 2(1): 63-82. (in Chinese)

8 淡晶晶, 陈光春, 马尊武, 等. 基于冰山理论的科研机构管理人员能力结构模型构建. 管理现代化, 2019, (6): 52-55.
Dan J J, Chen G C, Ma Z W, et al. Construction of competency structure model for scientific research institution managers based on iceberg theory. Modernization of Management, 2019, (6): 52-55. (in Chinese)

9 张民选. 回应、协商与共同建构——"第四代评价理论"评述. 外国教育资料, 1995, (3): 53-57.
Zhang M X. Response, negotiation and co-construction: A review of "fourth generation evaluation theory". International and Comparative Education, 1995, (3): 53-57. (in Chinese)

10 Guba E G, Lincoln Y S. Fourth generation evaluation. California: Sage, 1989.

11 田五星, 王海凤. 大数据时代的公共部门绩效管理模式创新——基于KPI与OKR比较的启示与借鉴. 经济体制改革, 2017, 204(3): 17-23.
Tian W X, Wang H F. Performance management model innovation in the public sector in the era of big data: Enlightenment and reference based on the comparison of KPI and OKR. Reform of Economic System, 2017, (3): 17-23. (in Chinese)

12 西楠, 彭剑锋, 曹毅, 等. OKR是什么及为什么能提升团队

绩效？——柔性导向绩效管理实践案例研究. 科学学与科学技术管理, 2020, 41(7): 116-138.

Xi N, Peng J F, Cao Y, et al. What is OKR and why can it improve team performance? A case study of flexibility-oriented performance management practices. Science of Science and Management of S.& T., 2020, 41(7): 116-138. (in Chinese)

13 胡中锋, 李奕霏. 现代教育评价往何处去: 兼评第四代评价理论. 教育研究与实验, 2025, (6): 126-134.

Hu Z F, Li Y F. Where is modern educational evaluation heading: A review of the fourth generation evaluation theory. Educational Research and Experiment, 2025, (6): 126-134. (in Chinese)

14 张兰, 孙旌凯, 薛东波. 混改背景下国有企业绩效考核体系创新研究——基于KPI与OKR比较的借鉴与启示. 税务与经济, 2022, 245(6): 101-106.

Zhang L, Sun J K, Xue D B. Research on innovation of performance appraisal system in state-owned enterprises under the background of mixed ownership reform: enlightenment and reference based on the comparison of KPI and OKR. Taxation and Economy, 2022, 245(6): 101-106. (in Chinese)

From indicators to systems: Logical evolution, practical challenges, and policy recommendations of think tank evaluation in China

LI Gang[1,2*] GE Zihao[1,2]

（1 School of Information Management, Nanjing University, Nanjing 210023, China;

2 China Think Tank Research and Evaluation Center, Nanjing University, Nanjing 210023, China）

Abstract Over a decade has passed since the issuance of the “Opinions on Strengthening the Construction of New Types of Think Tanks with Chinese Characteristics”. During this period, the number of think tanks in China has continued to grow, and evaluation activities have evolved from spontaneous exploration to institutionalized practice. Examining think tank evaluation from both the supply and demand sides of think tank development, this study systematically reviews the practical progress and theoretical exploration of think tank evaluation in China over the past decade. Building upon an analysis of the conceptual connotation of think tank evaluation, the study argues that the key to evaluation lies not in indicator design, but in clarifying the fundamental purposes and intentions of evaluation. The study identifies several challenges confronting the current evaluation system: the theoretical premises of think tank evaluation have not been fully clarified; standardized evaluation criteria are inadequate for accommodating the diversity of think tank types and developmental stages; the short-cycle orientation of annual performance assessments neglects the cultivation of latent foundational capacities such as data accumulation, methodological reserves, and talent development; and the unidirectional operation of evaluation processes undermines the diagnostic and developmental functions of evaluation. In response to these issues, this study proposes optimization pathways from multiple dimensions, including broadening the functional positioning of evaluation,

*Corresponding author

integrating KPI and OKR management concepts, introducing deliberative mechanisms informed by responsive evaluation theory, and establishing a differentiated and dynamic system of evaluation indicators and weightings. The study further argues that Chinese think tank evaluation should shift from pursuing a single set of perfect indicators toward building a collaborative evaluation ecosystem involving multiple stakeholders, so as to promote improvement through evaluation and serve the high-quality development of new types of think tanks with Chinese characteristics.

Keywords new types of think tanks with Chinese characteristics, think tank evaluation, responsive evaluation, governance model, performance management

李　刚　南京大学信息管理学院教授、博士生导师，南京大学中国智库研究与评价中心主任、首席专家。研究方向：智库评价理论与方法、智库信息系统与知识管理、图书馆与档案学理论基础。E-mail: ligang@nju.edu.cn

LI Gang Professor and doctoral supervisor at the School of Information Management, Nanjing University; director and chief expert of the China Think Tank Research and Evaluation Center, Nanjing University. His research interests include theory and methods of think tank evaluation; think tank information systems and knowledge management; and theoretical foundations of library and archival science. E-mail: ligang@nju.edu.cn

■责任编辑：文彦杰