

· 科学基金深化改革重要举措 ·

管理科学部经济科学学科 人工智能指派与分类评审效果分析

陈中飞¹ 汪 锋² 董明放³ 吴 刚^{1*} 刘作仪¹

1. 国家自然科学基金委员会 管理科学部, 北京 100085

2. 重庆大学 经济与工商管理学院, 重庆 400030

3. 西安建筑科技大学 管理学院, 西安 710055

[摘要] 本文对国家自然科学基金委员会管理科学部经济科学学科(G03)2022年面上项目和青年科学基金项目人工智能指派和分类评审的效果进行分析。基于在经济科学学科国际经济与贸易领域的随机对照试验发现,人工智能指派能够高效匹配“小同行”开展通讯评审,有效提升评审项目共识度和上会率。经统计检验,人工智能指派实验组和对照组样本的项目评审得分分布与均值没有显著差异,意味着人工智能指派不会系统性地影响同行通讯评审结果。2022年经济科学学科全面开展基于科学问题属性的分类评审,统计数据表明分类评审能够有效使同行评议专家按科学问题属性形成共识,明显缩小原创、前沿、交叉类项目同需求类项目之间上会率的差异。

[关键词] 人工智能指派;分类评审;管理科学部;效果分析

国家自然科学基金委员会(以下简称“自然科学基金委”)是我国支持基础研究的主渠道之一,成立30多年来,在国家科技事业中发挥着重要作用^[1,2]。为应对基础研究跨越式发展面临的挑战,近年来自然科学基金委持续探索深化改革,在管理机制上不断推出新举措,满足新一轮科技革命对国家科研资助体系的新要求。

同行评议是国家自然科学基金资助项目申请的主要机制^[3-5]。自然科学基金委十分注重技术手段在同行评议组织管理中的应用,积极开展将人工智能技术(Artificial Intelligence, AI)引入到同行评议的试点工作中,并在实践中不断完善智能辅助指派系统^[6-8]。2022年管理科学部经济科学学科开展了人工智能指派试点工作。我们在选取学科试点领域之后,采用随机对照试验的方法,将研究对象随机分组,在同行通讯评审阶段,对实验组采取人工智能指派,对照组采取通常的分组指派,通过比较通讯评审结果,科学分析采取人工智能指派对申请项目上会率的影响。



吴刚 研究员,国家自然科学基金委员会管理科学部三处处长。早期在中国科学院从事能源安全与能源环境复杂系统研究,获教育部一等奖2项、第十四届中国青年科技奖,入选国家“百千万人才工程”。



陈中飞 副研究员,国家自然科学基金委员会管理科学部三处经济科学学科项目主任。前期在暨南大学从事金融经济研究,入选中组部“国家万人计划”青年拔尖人才,获广东省自然科学基金杰出青年科学基金项目。

1 经济科学学科人工智能指派效果分析

为了做好人工智能指派试点工作,经济科学学科在集中评审之前经过多轮调研和讨论,设计了人工智能指派效果分析的实验方案和技术路线。

1.1 人工智能指派试点方案设计

(1) 认真分析学科各领域申请情况和特点,选取了学科代码 G0306“国际经济与贸易”面上项目和青年科学基金项目作为人工智能指派试点领域,采取随机对照试验的方法观测人工智能指派的实施效果。“国际经济与贸易”领域近年来面上项目和青年科学基金项目申请量稳定在 200 项以上,领域内学者整体研究水平较高,研究方法创新性较强,研究话题相对集中,因此,被选中作为人工智能指派试点的学科领域。

(2) 按学部编号最后一位为单数或双数将“国际经济与贸易”面上项目和青年科学基金项目分为实验组和对照组。

(3) 对实验组申请项目采取人工智能指派进行一对一的单项指派,由科学基金信息系统匹配出人工智能推荐的通讯评审专家,按匹配度从高到低指派专家。对于实验组中经由人工智能推荐评审项目数量低于 5 项的专家,由项目主任进行调整,选择匹配度稍低的专家进行通讯评审,保证通讯评审专家本年度评审项目数至少在 5 项以上,便于专家通过横向比较对申请项目进行客观评价。

(4) 对照组按分类评审要求和研究主题相关度进行分组,并综合运用“分组信息推荐”“关键词匹配推荐”“专家熟悉领域推荐”等手段,通过智能辅助指派系统整组指派通讯评审专家。

(5) 本年度通讯评审结束后,组织学科处工作人员开展专题研究,统计实验组和对照组评审结果的差异,检验实验结果的科学性,总结人工智能指派的实施效果和改进措施。

1.2 效果分析

效果分析的过程和结果如下:

(1) 人工智能指派保障了评审组织工作的效率和有效性。项目主任使用人工智能指派系统进行通讯评审专家指派过程中,普遍反映人工智能指派在单项指派中效果较好,能够迅速列出研究方向相近的通讯评审专家,并对通讯评审专家的研究主题和学术网络进行画像,有助于精准找到“小同行”开展项目评审工作。

(2) 基于实验组和对照组的通讯评审结果对比来看,人工智能指派更精准邀请“小同行”开展项目评审工作,提升了项目的共识度,大幅提升了项目上会率。表 1 统计了参加随机对照试验的“国际经济与贸易”实验组和对照组的上会率结果。实验组采用人工智能指派,面上项目比对照组上会率提高了

10.38%,青年科学基金项目比对照组上会率提高了 14.01%。一对一地使用人工智能指派评审专家使得“小同行”对学术前沿和关键科学问题了解更加深入,增加了评审项目的共识度,进而增加了优秀项目的上会率。而对照组采用通常的分组指派的方式开展通讯评审,上会率同学科其他领域的平均上会率基本持平。因此,单个项目的人工智能指派更精准的匹配“小同行”,且不受成组指派回避的困扰,共识度大幅提高。

1.3 统计检验

上述统计分析可以发现,人工智能指派有效提高了同行评议专家对申请项目的共识度,提高项目的上会率。那么这些统计结果是否科学?是否存在统计偏差?我们还需要对人工智能指派的随机对照试验进行若干检验,包括对实验组和对照组的得分均值进行 t 检验、Wilcoxon rank-sum 检验。检验的目的均是帮助判断人工智能指派是否改变了科学基金项目的得分分布和平均得分。如果是,则表明人工智能指派会产生系统性的得分偏差;反之,则表明人工智能指派试点方案设计是合理的,即它不会改变科学基金项目的得分。

通过 t 检验、Wilcoxon rank-sum 检验,可以发现人工智能指派并没有系统性改变评审结果,确保了评审的科学性。表 2 展示了实验组与对照组的均值统计及 t 检验结果。在通讯评审过程中,每个申请项目通讯评审总得分由综合评价和资助建议相加得到。我们发现,实验组青年科学基金项目总得分的平均值为 3.176 分,对照组为 3.083 分,两者相差 0.093 分。根据 t 检验得出,实验组与对照组的均值差异所对应的 t 值为 0.516, p 值为 0.607,这一结果表明两者差异在统计上不显著。综合评价、资助建议,以及面上项目也存在类似的差异在统计上不显著的结果,意味着人工智能指派并不会改变实验组和对照组在通讯评审中的总得分。

从统计学上看, t 检验需要满足正态分布条件,

表 1 人工智能指派的上会率统计(2022 年)

	面上项目	青年科学 基金项目	合计
实验组(采用人工智能指派)	申请总数(项)	35	76
	上会数(项)	12	20
	上会率(%)	34.29%	26.32%
对照组	申请总数(项)	46	65
	上会数(项)	11	8
	上会率(%)	23.91%	12.31%

而科学基金项目的函评得分并不一定满足正态分布条件。对此,本文采用放宽假设的 Wilcoxon rank-sum 检验来判断实验组和对照组的得分是否存在显著性差异,具体结果展示于表 3。结果发现,实验组和对照组在总得分、综合评价、资助建议等方面的差异均没有通过显著性检验,即两者的得分差异并不显著。

2 经济科学学科分类评审效果分析

基础研究具有差异化的问题属性,无法用一把尺子衡量其研究价值和意义。自 2019 年以来,自然科学基金委全面部署基于科学问题属性的分类评审试点工作,不断探索完善相关机制。经济科学学科连续 2 年开展了基于科学问题属性的分类评审试点工作。

自 2018 年科学基金深化改革以来,为了更好地帮助科学家凝练科学问题,自然科学基金委采取与科学问题属性相匹配、相适应的明确资助导向的分类评审机制,将科学问题分为“鼓励探索、突出原创(属性 I);聚焦前沿、独辟蹊径(属性 II);需求牵引、突破瓶颈(属性 III);共性导向、交叉融通(属性 IV)”四类^[9]。2021 年,经济科学学科在部分学科代码采取随机对照试验方法,对分类评审试点的效果进行分析,发现分类评审能够有效提高同行评议专家对原创、前沿和交叉类项目的共识度^[10]。2022 年,管理科学部将分类评审推广至经济科学学科全部面上项目和青年科学基金项目,再次检验明确资助导向的分类评审机制的实施效果。

表 2 实验组与对照组的评审结果均值统计及 t 检验结果

		实验组	对照组	均值差值	t 值	p 值
青年科学基金项目	总得分	3.176	3.083	0.093	0.516	0.607
	综合评价	2.518	2.495	0.023	0.238	0.813
	资助建议	0.658	0.588	0.070	0.821	0.413
	样本数	76	65			
面上项目	总得分	3.376	2.991	0.384	1.332	0.187
	综合评价	2.620	2.396	0.224	1.393	0.168
	资助建议	0.756	0.596	0.160	1.224	0.225
	样本数	35	46			

表 3 Wilcoxon rank-sum 检验

	统计量	总得分	综合评价	资助建议	样本数
青年科学基金项目	Z 值	0.162	0.156	0.365	141
	p 值	0.872	0.876	0.715	
面上项目	Z 值	1.357	1.362	1.315	81
	p 值	0.175	0.173	0.188	

近年来,经济科学学科的面上项目、青年科学基金项目 and 地区科学基金项目三类申请量稳定在 3 000 项左右,如图 1 所示。

2019—2022 年,经济科学学科的面上项目、青年科学基金项目和地区科学基金项目以属性 III 和属性 II 占主导,约占 70% 和 20%。属性 I、属性 IV 的占比则逐年呈下降趋势(如图 2 所示)。

2.1 分类评审方案设计

2021 年经济科学学科采用随机对照试验方法,选取了 G0306“国际经济与贸易”、G0307“金融经济”、G0311“农林经济管理”三个学科代码的面上项目和青年科学基金项目为研究对象,并区分实验组和对照组,分析基于科学问题属性的分类评审模式对评审质量的影响^[10]。

2022 年经济科学学科在所有学科代码面上项目、青年科学基金项目中开展基于科学问题属性的分类评审。具体的实施方案如下:

(1) 在分组指派过程中,根据申请书选择的 I、II、III、IV 四类科学问题属性进行分组。对于按科学问题属性分组后申请项目较多的组,再按研究主题相似性继续进行聚类、分组;对于按科学问题属性分组后申请项目较少的组(小于等于 5 项申请项目),选择相同学科代码中属性 III 类项目之外的其他科学问题属性的申请项目,保持不同分组项目数量基本一致。最终,青年科学基金项目属性 III 类分成 65 组,属性 I、属性 II 和属性 IV 类项目合计 29 组;面上项目属性 III 类分成 41 组,属性 I、属性 II 和属性 IV 类项目合计 16 组。

(2) 对于属性 I、属性 IV 类项目,在指派通讯评审专家时优先遴选活跃在学术一线的资深学者。

(3) 通讯评审开始时,发送分类评审试点工作的函,事先告知通讯评审专家,经济科学学科将在 G03 经济科学学科面上项目、青年科学基金项目同行评议阶段全面开展分类评审试点工作,明确告知评审专家按照划分四类科学问题属性的标准对申请书进行评审。

2.2 效果分析

2022 年度经济科学学科全面开展分类评审,一方面,与 2020 年经济科学学科未全面开展分类评审的情况进行对比,能够客观地观察到不同的评审模式对各类科学问题属性项目上会率的影响;另一方面,2022 年度经济科学学科分类评审的上会率也能进一步验证 2021 年基于随机对照试验经济科学学科分类评审试点效果分析的结果,得到更加稳健的

分类评审效果结论。

表4统计了2022年全面开展分类评审的经济科学学科各科学问题属性项目上会率,以及2021年参加随机对照试验的三个学科代码实验组和对照组的上会率和2020年末全面开展分类评审的上会率结果。

从表4中可以得到以下几点发现:

(1) 分类评审有利于评审专家形成共识,增加了属性I、属性II和属性IV类项目的上会率。由于2022年全面开展按科学问题属性分类评审,经济科学学科面上项目和青年科学基金项目属性I、属性II类项目上会率比整体上会率分别高1.83%和1.86%,属性IV类项目上会率低于整体项目上会率1.80%。而2020年末实施科学问题属性分类评审

时,属性I、属性IV类项目上会率比整体上会率低1.23%和5.34%,而属性II类则高于整体上会率1.71%。

需要强调的是,对于属性I类项目,2021年开展随机对照试验实验组的上会率仅为14.29%,低于实验组项目整体上会率。而经过全面开展分类评审,2022年属性I类项目上会率超过了项目整体上会率。

对于属性IV类项目,在2021年的对照实验中,分类评审的效果非常显著,实验组的上会率高于对照组15%,且高于实验组整体上会率。而在2022年将分类评审全面推广到所有学科领域时,属性IV类项目的上会率略低于整体上会率水平,但其情况远好于2020年末开展分类评审的。这一

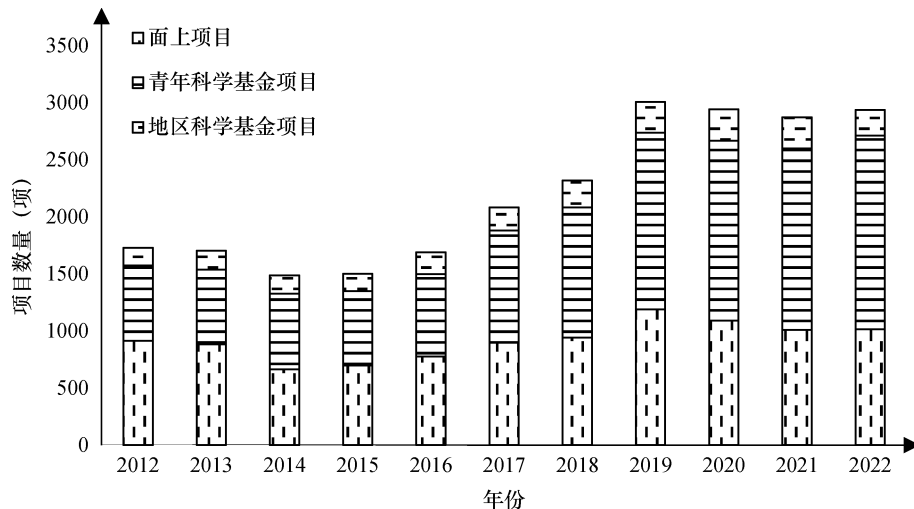


图1 经济科学学科三类项目历年总申请量(2012—2022年)

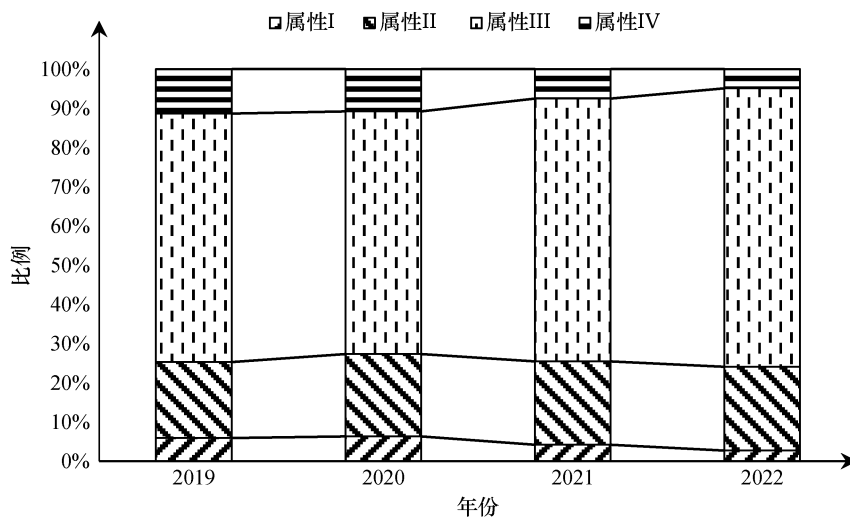


图2 经济科学学科三类项目科学问题属性分类占比(2019—2022年)

表 4 四种类型项目的上会率统计

			属性 I	属性 II	属性 III	属性 IV	合计
2022 年	全面开展分类评审 (面上项目、青年科学基金项目)	申请总数(项)	70	592	1 928	121	2 711
		上会数(项)	17	144	423	25	609
		上会率(%)	24.29%	24.32%	21.94%	20.66%	22.46%
2021 年	实验组	申请总数(项)	35	104	392	48	579
		上会数(项)	5	26	86	12	129
		上会率(%)	14.29%	25.00%	21.94%	25.00%	22.28%
2021 年	对照组	申请总数(项)	6	103	392	30	531
		上会数(项)	1	21	86	3	111
		上会率(%)	16.67%	20.39%	21.94%	10.00%	20.90%
2020 年	未开展分类评审(面上项目、青年科学基金项目)	申请总数(项)	163	581	1 640	280	2 664
		上会数(项)	30	124	329	40	523
		上会率(%)	18.40%	21.34%	20.06%	14.29%	19.63%

现象可能主要受 2021 年遴选三个学科方向进行试点所导致。

(2) 分类评审缩小了属性 I、属性 II、属性 IV 类项目同属性 III 类项目之间的上会率差异。2020 年, 属性 III 类项目上会率高于整体上会率 0.43%, 2021 年对照组中属性 III 类项目上会率高于整体上会率 1.04%; 相反, 实验组属性 III 类项目上会率低于整体上会率 0.34%, 2022 年属性 III 类项目上会率低于整体上会率 0.52%。

3 主要结论和启示

3.1 主要结论

(1) 人工智能指派提高了科学基金评审组织工作效率, 尤其是在遴选匹配的评审专家方面。经济科学学科的试点结果表明, 人工智能指派遴选的“小同行”通讯评审专家提升了项目共识度, 进而大幅提升项目上会率。

(2) t 检验和 Wilcoxon rank-sum 检验表明, 人工智能指派对整体评审得分没有显著的系统性影响, 确保了人工智能指派对评审工作的科学性没有影响。与此同时, 本次对比分析仅局限于国际经济与贸易这一单一学科领域, 人工智能指派具体效果还需要进一步通过扩大试点范围来进行验证。

(3) 2022 年的全面分类评审的统计分析数据再次验证, 分类评审促使评审专家对属性 I、属性 II 和属性 IV 类的项目形成共识, 不再用“一把尺子”度量

所有的项目, 有利于消除属性 I、属性 II 和属性 IV 类项目同属性 III 类项目之间的上会率差异。基于科学问题属性的分类评审效果明显, 学界已初步形成区分科学问题属性撰写项目申请书和评审申请书的共识。

3.2 研究启示

(1) 结合试点实践和评审专家的意见反馈, 今后还需要对相关工作机制开展进一步改进。如人工智能指派是基于历史知识给予的研判, 这对属性 I 和属性 IV 类项目的评审形成了一定的挑战, 如何确保学术创新点的萌发和学术生态多元化, 还需进一步探讨。

(2) 分类评审工作还需要继续深入探索, 包括不同分类属性项目的申请、评审、结题、后评估等不同环节应该进行差异性设计, 从而更好地指引科学家明确资助导向, 更好地设计和执行研究任务, 通过完善的机制设计帮助科学家基于科学问题属性凝练科学问题。

参 考 文 献

- [1] 李静海. 构建新时代科学基金体系夯实世界科技强国根基. 中国科学基金, 2018, 32(4): 345—350.
- [2] 李静海. 大力提升源头创新能力, 构建面向新时代的科学基金体系. 求是, 2018, 22: 32—34.
- [3] 郭碧坚, 韩宇. 同行评议制——方法、理论、功能、指标. 科学学研究, 1994, 12(3): 63—73, 2.
- [4] 胡明铭, 黄菊芳. 同行评议研究综述. 中国科学基金, 2005, 19(4): 251—253.

- [5] 殷嘉珺, 罗惠文, 庄建辉. 原创性科学研究同行评议的国际经验及对原创探索计划项目的启示. 中国科学基金, 2021, 35(4): 567—572.
- [6] 江虎军, 徐岩英, 孙瑞娟, 等. 科研项目的第三方专业评议与智能辅助指派. 中国科学基金, 2015, 29(3): 216—218.
- [7] 江虎军, 郝艳妮, 徐岩英, 等. 国家自然科学基金项目同行评议的智能化探讨. 中国科学基金, 2019, 33(2): 149—153.
- [8] 窦豆, 李萃, 江虎军, 等. 科学基金同行评议智能指派的实践探索. 中国科学基金, 2021, 35(3): 458—461.
- [9] 李静海. 全面深化科学基金改革更好发挥在国家创新体系中的基础引领作用. 中国科学基金, 2019, 33(3): 209—214.
- [10] 吴刚, 陈中飞, 汪锋, 等. 基于随机对照试验的管理科学部三处分类评审试点效果分析. 计量经济学报, 2022, 2(2): 228—236.

The Effect of Using AI Assignment and Category-specific to the Review of National Natural Science Fund: Piloting by the Economic Science

Zhongfei Chen¹ Feng Wang² Mingfang Dong³ Gang Wu^{1*} Zuoyi Liu¹

1. Department of Management Sciences, National Natural Science Foundation of China, Beijing 100085

2. School of Economics and Business Administration, Chongqing University, Chongqing 400030

3. School of Management, Xi'an University of Architecture & Technology, Xi'an 710055

Abstract This study evaluates the effect of the AI assignment and category-specific review in 2022, which covers the General Program and the Youth Scientists Fund in Economic Science (G03) of Division III in the Department of Management Sciences. Based on the randomized controlled trials in the field of International Economics and Trade, we find that the AI assignment can effectively match experts to carry out peer-review. With AI assignment, the experts' consensus on applications has been improved, which increases the probability of being on the conference review. With statistical tests, there is no significant difference in the mean value of the scores between the treatment and control groups, which implies that the AI assignment has not systematically affected the results of the peer-review. All applications in Economic Science (G03) have been under category-specific review based on scientific problem attributes this year. The statistics again show that category-specific review can effectively enable experts to form consensus, and significantly reduce the difference in the rate for the applications being on the conference review among the Original Exploratory program, the Cutting-edge program, the Transdisciplinary program, and the Application program.

Keywords AI Assignment; category-specific review; Department of Management Sciences; effect evaluation

(责任编辑 吴征天 张强)

* Corresponding Author, Email: wugang@nsfc.gov.cn