公众算法认知调研报告（2022）

支振锋

中国社会科学院法学研究所研究员 博士生导师

中组部国家万人计划哲学社会科学领军人才

中宣部文化名家暨四个一批人才

全国杰出青年法学家提名奖

目录

CONTENTS

01 前言

02 问卷调查结果基本分析

03 不同群体对算法感知的差异分析

04 公众算法认知的影响因素

05 调研结论及政策建议

前言

数字时代

随着人工智能技术的广泛使用，依托各类商业和公共场景中多元用户的海量数据，算法已经渗透进经济社会生活的方方面面。

算法问题

算法滥用成风，算法不规范应用、不合理应用引发的“信息茧房”“算法歧视”“大数据杀熟”等问题层出不穷

数字化转型

近年来，我国在商业和公共领域广泛应用算法，将其作为应对信息过载危机、促进企业数字化转型和推动国家治理体系和治理能力现代化的重要力量。

算法治理

加快建立和完善算法治理体系，完善算法治理规则，提高算法治理能力，已成为推动建设网络强国、数字强国的重要内容。

前言

●研究背景

➢2022年3月1日，《互联网信息服务算法推荐管理规定》正式开始施行，算法推荐迎来了一场前所未有的变革，算法推荐野蛮生长的时代宣告结束。

●问卷调查目的

➢清楚准确地掌握用户对算法的感知情况

➢详细了解算法治理过程中亟须完善的侧重点

➢为算法治理路径的完善探索提供建议

●问卷模块设计

➢核心问题与用户个人基本信息

问卷调查结果基本分析

- 调查方式：问卷调查，线上线下相结合，兼顾自由填写与定向推送
- 调查范围：涵盖各个年龄段、地区、收入水平和学历层次
- 调查时间：2022年11月12日至11月30日，历时18天
- 调查结果：共回收总样本量3523份，有效样本3522份，样本有效率达到99%以上
  - 其中，通过部分基层社区、扶贫干部、建筑工队、家政公司等的协助，特意在东北、西南、西北、北部及中部等不发达地区定向推送问卷1070余人/次

问卷调查结果基本分析

- 样本基本情况

➢调查对象的性别、年龄和区域分布

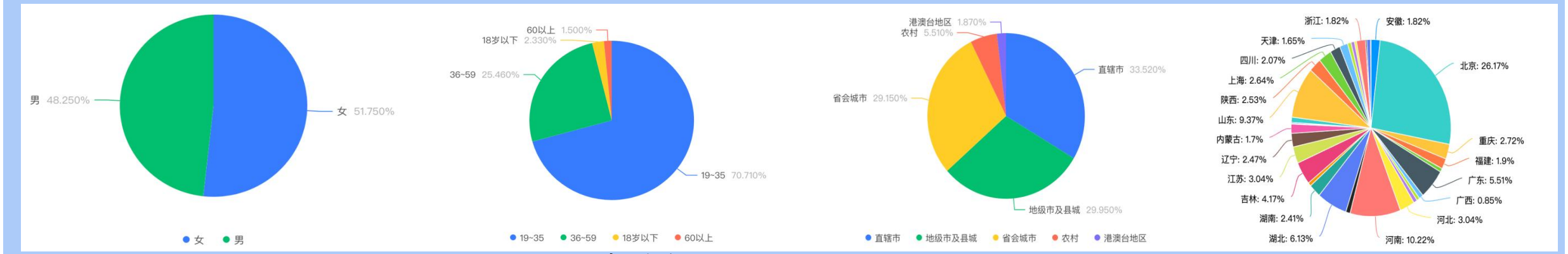

图1-1调查对象的性别分布

回收的样本数量中性别比例较为均衡。

图1-2 调查对象的年龄分布

本次调查对象的年龄集中在19~35岁之间，而18岁以下和60岁以上的调查对象占比之和仅为3.83%，存在明显的“数字鸿沟”。

图1-3 调查对象的地区分布

九成以上的受调查者居住于县城以上地区，其中居住地在直辖市的占比最高。这显示出我国不同群体居民数字素养或算法素养上存在的巨大差异。

图1-4 调查对象的省份分布

受调查者常住省份在北京市的比例最高，其他区域的总体分布还较为平衡。

问卷调查结果基本分析

➢调查对象的学历、职业和收入分布

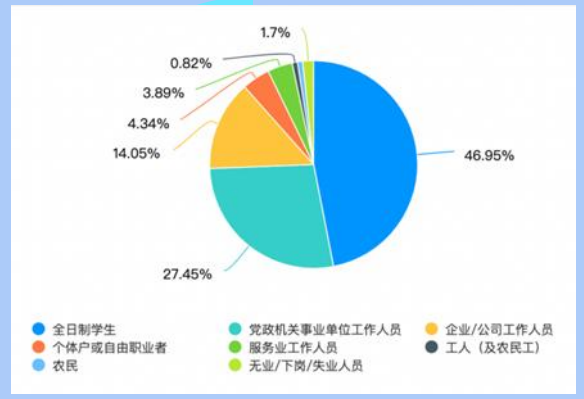

图1-5 调查对象的学历分布

超九成的受调查者拥有大专以上学历，其中硕士及以上的高学历受调查者占比最高；中专或职业学校的学历占比最低。

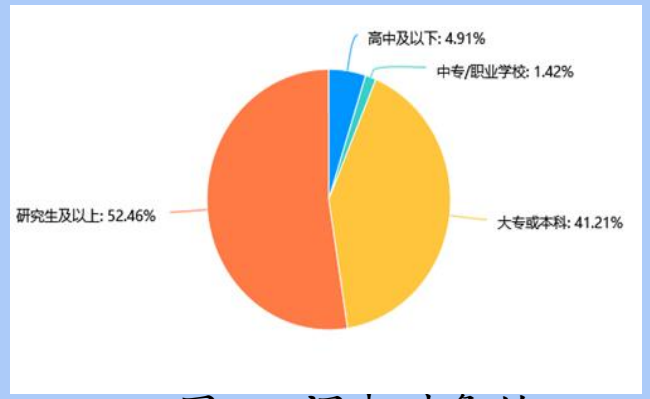

图1-6 调查对象的职业分布

全日制学生占比最高，工人、农民的比重最低，这一职业分布可能与受调查者的算法应用的使用情况有关，也可能反映出不同群体的受调查意愿存在较大差别，说明“数字鸿沟”不仅客观存在，而且比较严重。

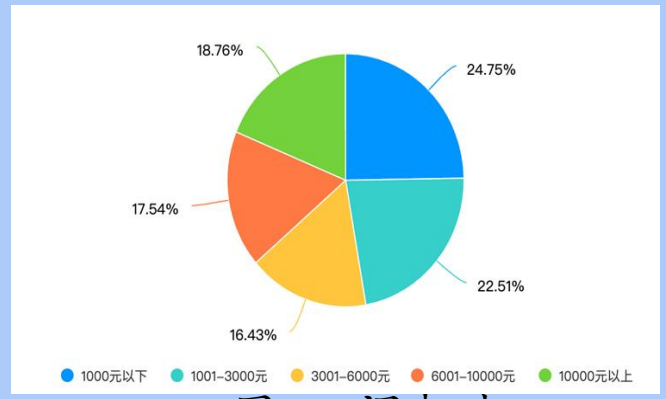

图1-7 调查对象的收入分布

调查对象的收入结构分布较为平均，其中月收入1000元以下的比重最高，这主要和在受调查者中有部分是全日制学生有关。

问卷调查结果基本分析

➢关于算法本身的认知、态度和行为

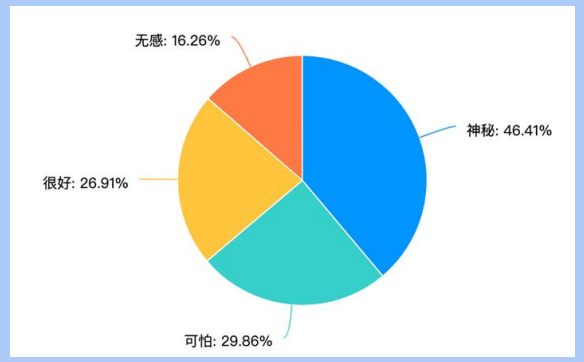

图2-1 调查对象对算法的感知度和基本了解

感到“神秘”的比例最高，对算法“无感”的比例最低。大多数受调查者能感受到“算法”的存在但不了解，这说明算法在某种意义上已经成了大众“日用而不知，习焉而不察”的事物了。

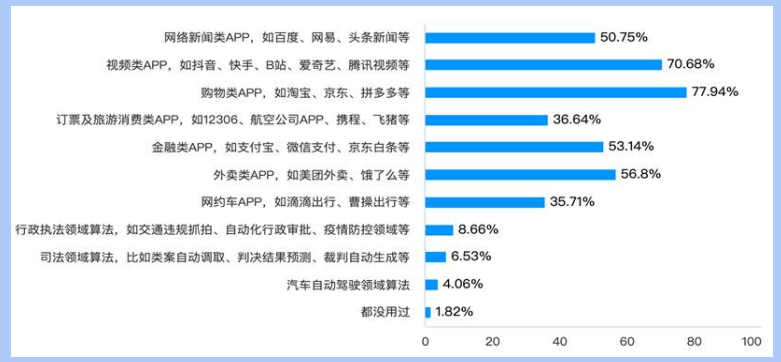

图2-2 关于常用算法应用

超七成的受调查者使用购物类APP和视频类APP，受调查者的常用算法应用集中于生活场景，主要是为了满足接受资讯、娱乐或消费的需要，专业化应用场景并不突出。

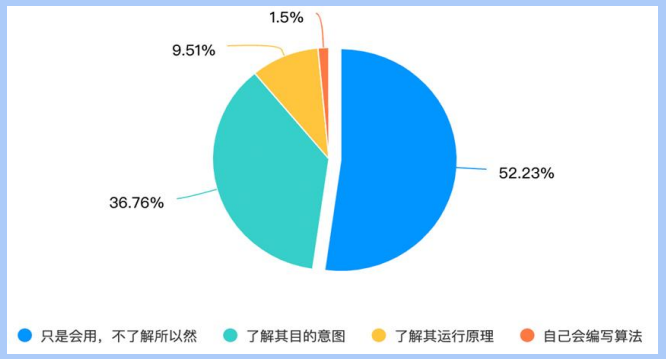

图2-3 常用的算法的总体了解程度

受调查者对常用算法应用的总体了解程度偏低。逾一半的受调查者表示“只是会用，不了解其所以然”。

问卷调查结果基本分析

- 调查结果的整体分析

➢对算法影响的知识、态度和行为

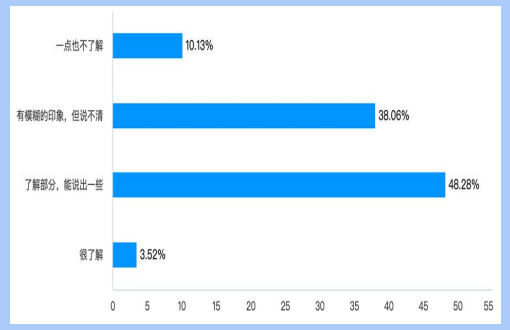

图2-4 算法带来的好处了解程度

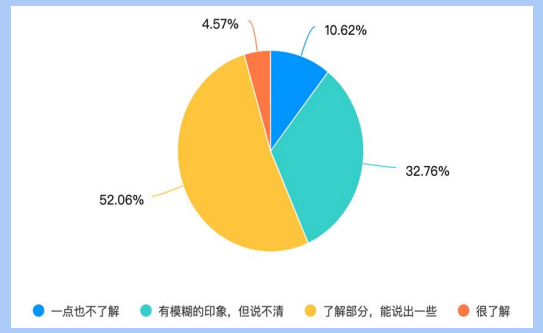

图2-5 算法带来的弊端了解程度

受调查者对算法带来的弊端与对算法带来的好处的了解程度大致相同，一点也不了解与很了解占比较小，模糊了解与部分了解占比较大，这说明公众对算法有感知但理解不深。

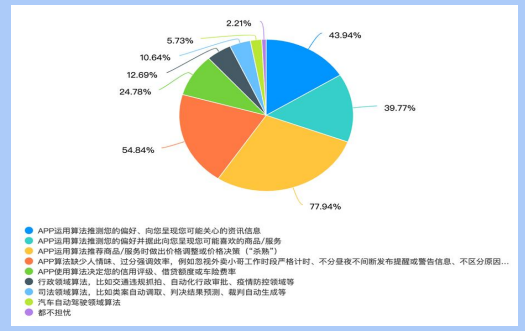

图2-6 最担忧的算法风险分布

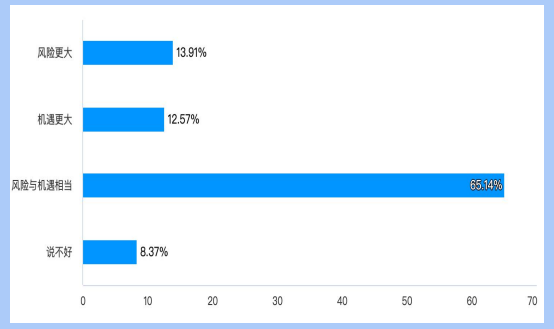

图2-7 算法对人类社会的影响

绝大多数的受调查者关心可能存在的算法风险，其中担心APP运用算法“杀熟”的比例最高。整体来看，使用越频繁、对个人信息利用越多的算法场景，受调查者就越担忧其算法风险。

但仍有超六成受调查者认为算法技术风险与机遇相当，这反映了更多受调查者具备一定的科学素养，对技术创新和应用创新有开放的心态。

问卷调查结果基本分析

➢对算法的监管的认知

- 知情权

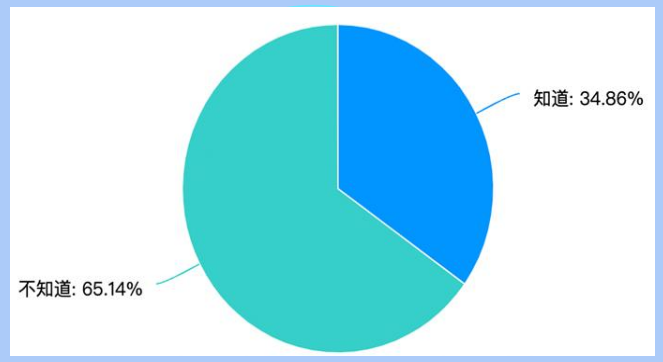

图2-8 您是否知道有法规要求算法服务提供者向您明示算法的目的、原理和机制？

过半数的受调查者不知道自己享有对算法的目的、原理和机制的知情权，这可能与算法相关的法律法规的普法宣传力度不够有关，也可能受算法认知程度的影响。

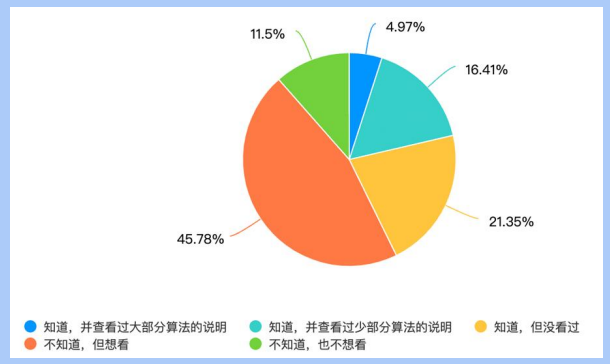

图2-9 您知道去哪里查看算法服务提供者对算法的解释或说明吗？

只有少部分受调查者查看过算法服务提供者对算法的解释或说明，大部分受调查者不知道如何查看关于算法的说明但是表示想要了解，这说明多数受调查者虽然没有专业知识，但对于算法的内容、运行机制抱有一定的好奇心，存在一定的相关诉求。

问卷调查结果基本分析

- 知情权

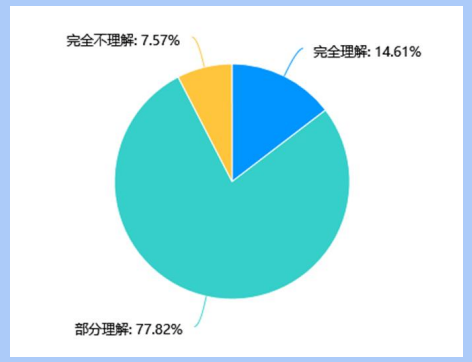

图2-10 对“算法的解释”的了解程度

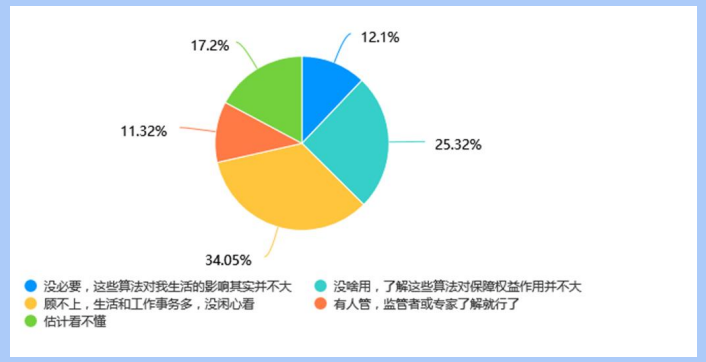

图2-11 您为何不去看对算法的解释？

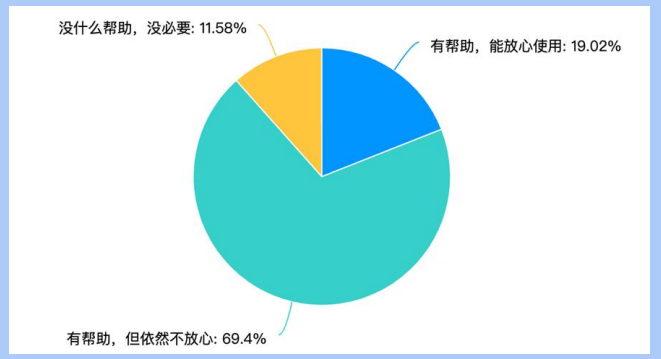

图2-12 算法服务提供者对算法的解释是否有助于您放心使用该算法？

问卷调查结果基本分析

- 救济权

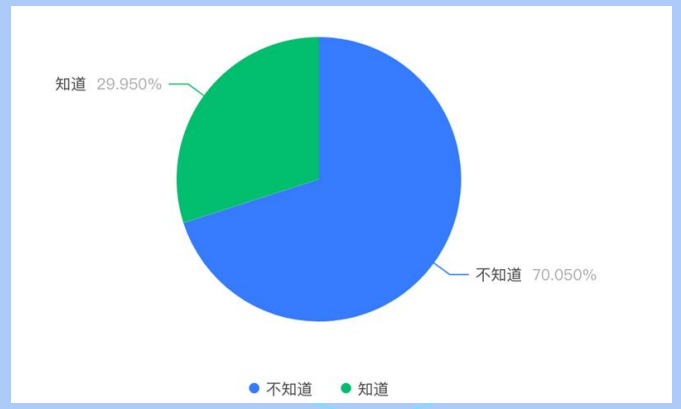

图2-13 如果您觉得算法侵害了您的权益，您是否知道依法该经哪些部门或渠道投诉或举报？

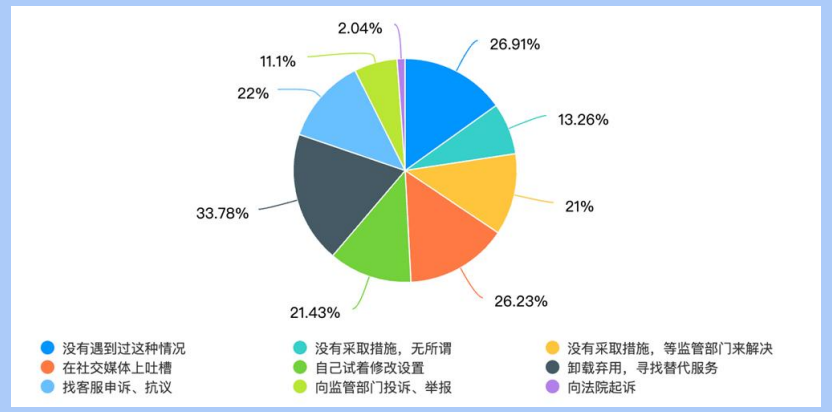

图2-14 当算法侵害了您的权益时，您采取过哪些措施？

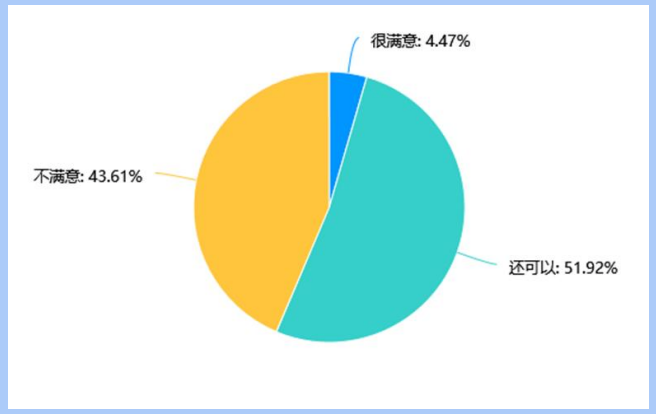

图2-15 您对如此处理后的情况满意吗？

根据算法侵权后对投诉路径了解情况的频数分析，结果显示：选择不知道的所占百分比远大于知道的。这说明算法时代，用户维权意识较高，算法开发者应该更多地考虑到用户权益。

选择“卸载弃用”，“寻找替代服务”的比例最高，向法院起诉的比例最低，大部分都选择会采取措施。这启示算法开发者应当更加注重用户体验，注重建立多种反馈渠道，倾听用户心声，这有利于提高算法的质量，保留用户，并增强竞争力。

很满意的比例极低，这说明公众对算法治理措施满意度一般。算法治理的决策往往是黑箱运作的，对外界难以透明。这可能导致公众对算法治理的决策缺乏了解，并产生疑虑情绪

问卷调查结果基本分析

- 救济权

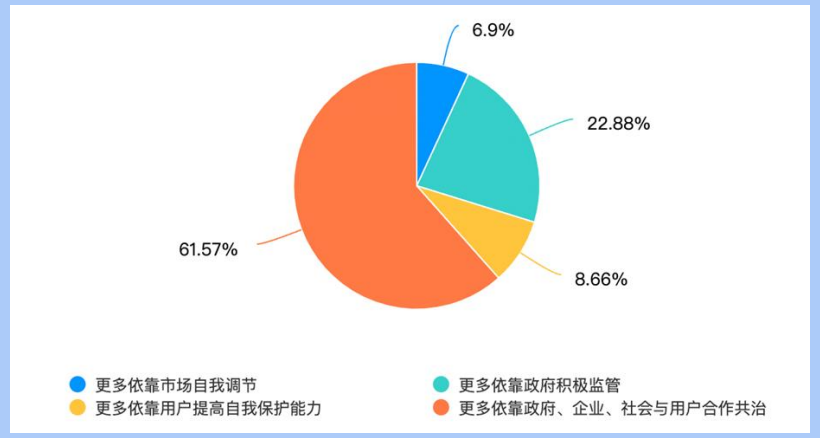

图2-16 对于社会上的算法问题，您认为相对最可靠的治理方式是什么？

大多数人认为，更多依靠政府、企业、社会与用户合作共治是解决问题的最佳方式。这说明算法问题可能涉及政策、法律、道德、伦理等多个方面，因此可能需要政府、企业、社会和用户的共同努力才能解决这些问题。

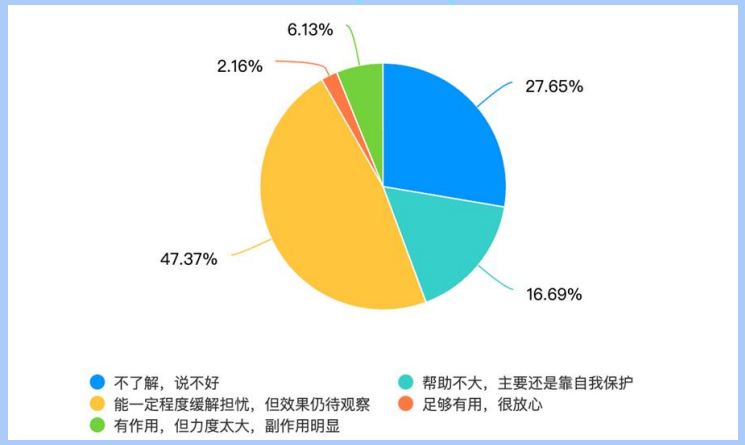

图2-17 对监管部门治理措施效果的评价

能一定程度缓解担忧，但效果仍待观察的答案频率最高。
足够有用，很放心的答案频率最低。

问卷调查结果基本分析

➢对具体算法场景问题的认知与态度

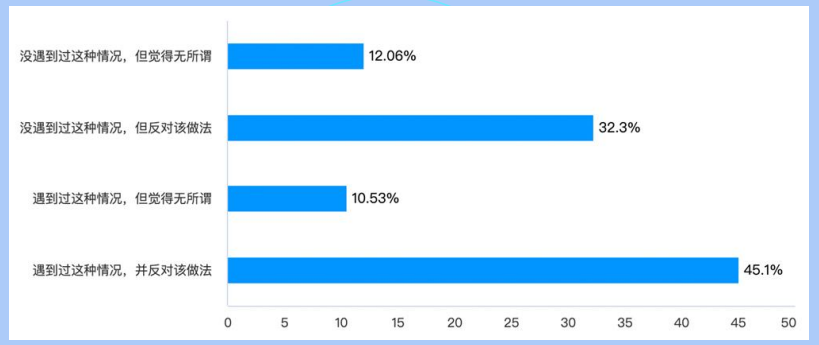

图2-18 对“大数据杀熟”的评价

遇到过这种情况，并反对该做法的频率最高。大数据杀熟是一种利用大量数据来了解用户偏好并进行针对性报价的做法。通过使用大数据来了解用户的消费习惯、喜好和行为，商家可以给不同的用户提供不同的报价。

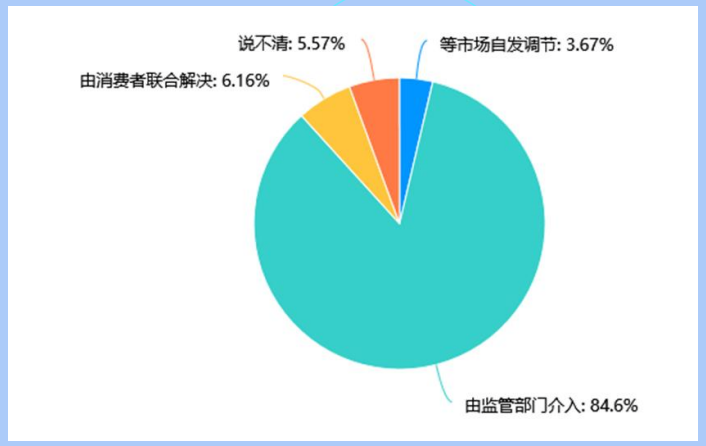

图2-19“大数据杀熟”的期待解决方法

选择由监管部门介入的比例最高。公众这样选择一个可能的原因是，监管部门可以通过制定相关法律法规来规范大数据杀熟的行为，并对违法违规行为进行处罚，以维护用户的合法权益。另一个可能的原因是，监管部门可以对商家使用的大数据进行审查，以确保数据的真实性和合法性，防止商家过度收集用户数据并侵犯用户隐私。

问卷调查结果基本分析

➢对具体算法场景问题的认知与态度

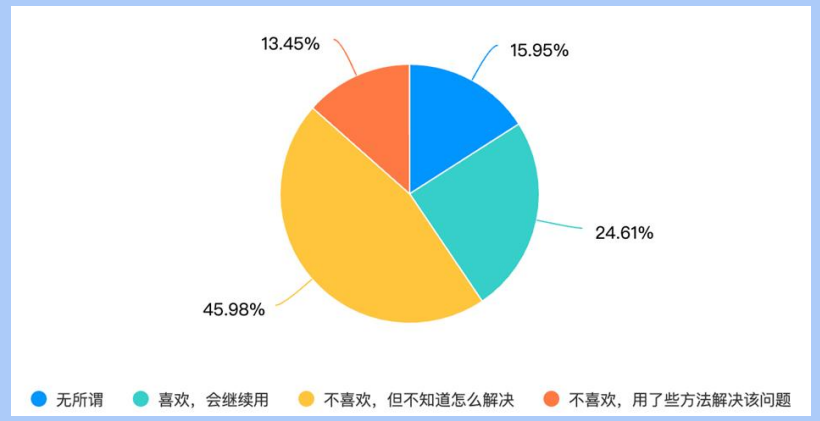

图2-20 对“个性化推荐”的评价

选择“不知道怎么解决”的最高，选择“不喜欢，用了一些方法解决该问题”的最低。

“推荐偏见”是指用户只能看到与他们的兴趣和偏好相匹配的内容，而不会看到其他内容的现象。这种偏见可能会导致用户偏向于狭隘的观点，无法从不同的角度来看待问题。这既是所谓“信息茧房”“同温层效应”，往往强化社会的极化心理，助长社会认知撕裂。

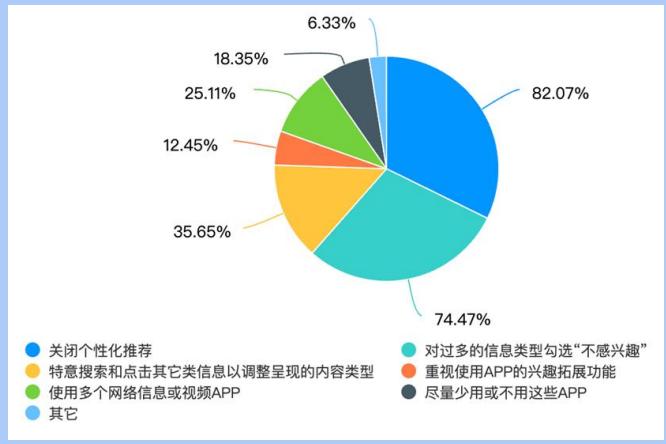

图2-21 应对个性推荐的方法

应对个性化推荐的方法，关闭个性化推荐的比例最高。推荐偏见可能会导致用户陷入狭隘的思维模式中，从而影响他们的思考能力和判断力。因此，公众认为关闭个性化推荐可能是一个好的选择，以避免推荐偏见的产生

问卷调查结果基本分析

➢对具体算法场景问题的认知与态度

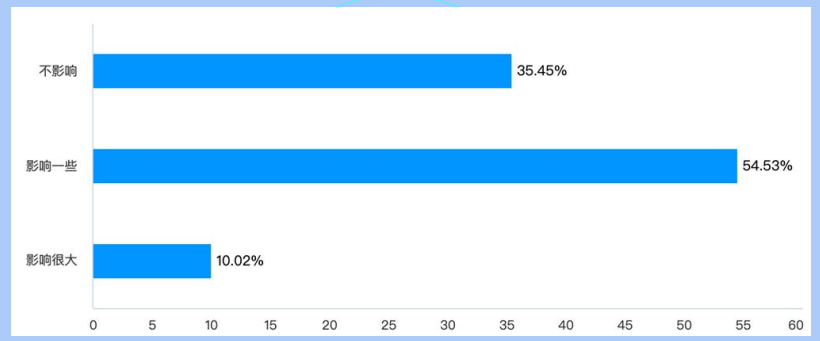

图2-22“恶意刷评”对商品评分信任程度影响

刷好评是指商家通过不正当手段来提高自己的好评率。这种做法通常包括购买假好评、对购买产品的顾客施压要求好评等。淘宝刷好评是违法的，会对商家的声誉造成极大损害，并有可能被网站官方严惩。数据显示，刷好评行为对消费者的行为产生了较大影响。

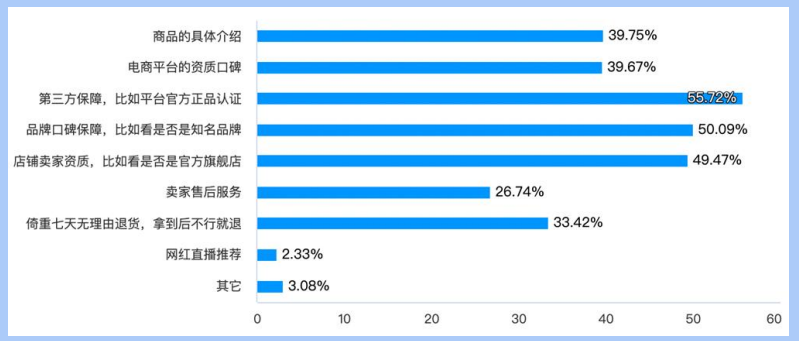

图2-23为了“弥补评分缺陷”，可替代选品方法情况

选择“第三方保障”的比例最高，选择“网红直播推荐”的比例最低。详情见图2-23。“第三方保障”指的是，在购物平台上发布的商品评价之外，还有其他机构或第三方提供的保障。

问卷调查结果基本分析

➢对具体算法场景问题的认知与态度

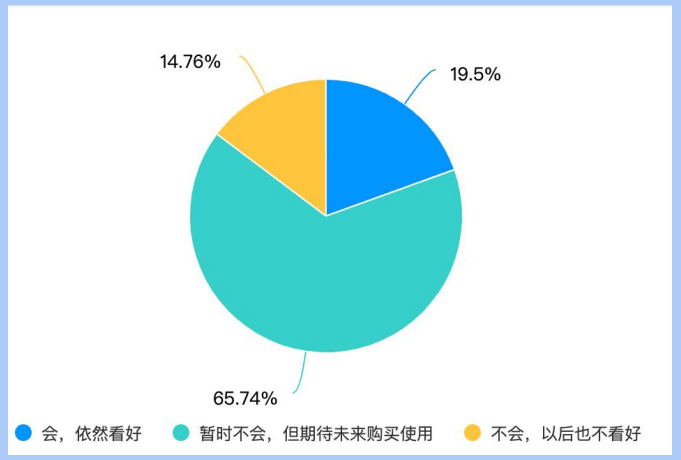

图2-22“汽车自动驾驶”的看法

选择“暂时不会，但期待未来购买使用”占百分比65.73%；会，依然看好占百分比19.506%；不会，以后也不看好占百分比14.764%。
这说明用户对新技术和新应用创新大多还保持着开放和期待的态度，可以增强相关市场主体进行技术创新的信心。

问卷调查结果基本分析

➢对具体算法场景问题的认知与态度

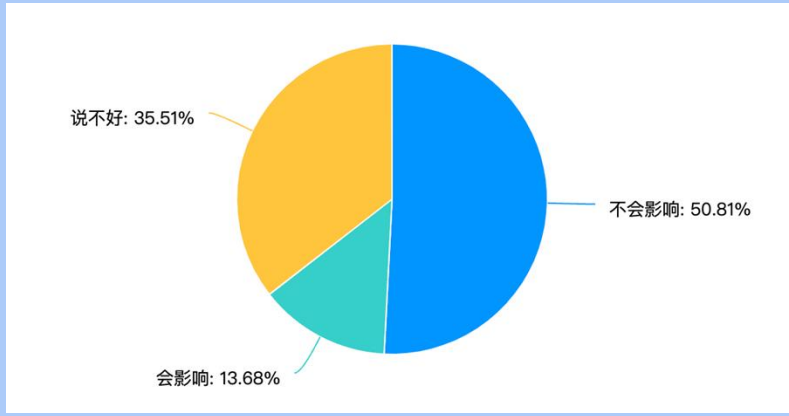

图2-25“一种算法极大提高了您的生活效率，但是您无法知道它的原理和运行机制”，该情况对用户使用行为的影响

选择“不会影响” 占百分比50.795%；选择“说不好” 占百分比35.52%；选择“会影响” 占百分比13.685%。这再次印证了受调查者或者说中国消费者对新技术与新应用创新的开放和支持态度。

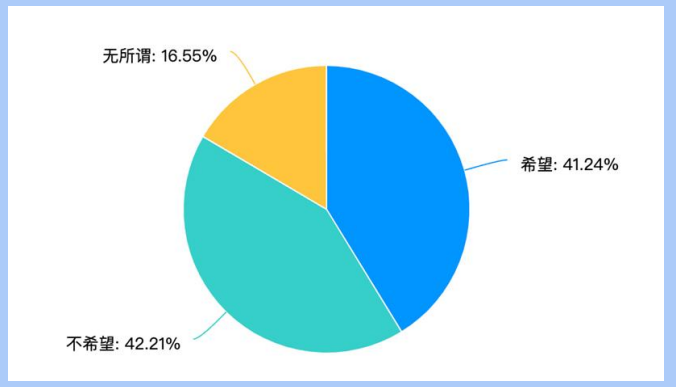

图2-26一种算法极大提高了政府的工作效率，但是公民无法知道它的原理和运行机制

选择“不希望” 占百分比42.22%；选择“希望”占百分比41.255%；选择“无所谓” 占百分比16.525%。选择“希望”与“不希望”的比例悬殊不大，反映了受调查者中间存在的疑虑心理。政府等公共部门提供公共产品，事关社会公平正义，在选择使用算法时应该更加注重公开、透明和可解释性。

问卷调查结果基本分析

➢对具体算法场景问题的认知与态度

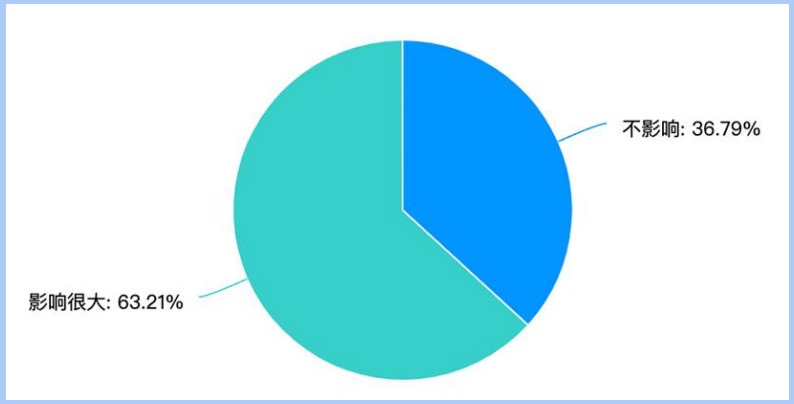

图2-27“使用算法时，发现算法存在错误决策”对算法信任程度的影响

选择“影响很大” 占百分比63.203%；选择“不影响” 占百分比36.797%。这说明算法决策出错可能会影响消费者对算法的信任度，可能影响算法在市场上的发展。

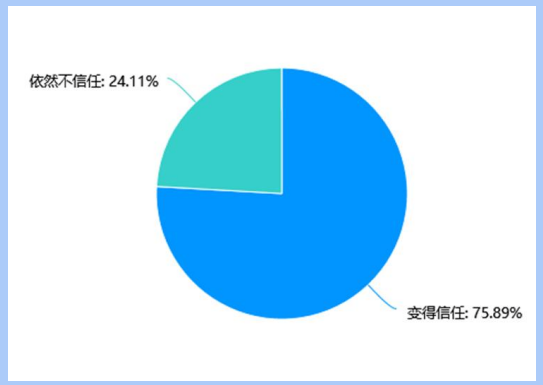

图2-28“用户可以了解该算法的原理，并能够亲自微调和控制该算法的运行时”对应用信任情况的影响

选择“变得信任”的比例最高，达到75.89%。了解算法的工作原理和决策方式，可以让用户更好地判断算法的可靠性和准确性，从而提高对算法的信任度。同时，参与算法的设计和实现，可以让用户更好地提供建议和反馈，从而提高算法的质量和性能，也可以增加对算法的信任度。某种程度上，这也算是一种“算法民主”。

问卷调查结果基本分析

➢调查后读者对算法的态度变化情况

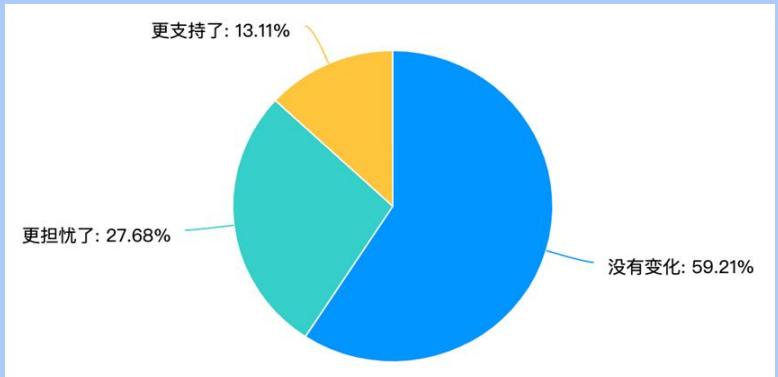

图2-29 调查后读者对算法的态度变化

选择“没有变化” 占百分比59.228%；选择“更担忧了” 占百分比27.655%；选择“更支持了” 占百分比13.118%。

由于我们在设计算法问卷时，特别注重问题的通俗易懂，使得填写问卷在某种意义上成为一次关于算法常识的普及活动，从而对超过四成的受调查者产生了明显的影响。这本身就说明了对算法进行科普，以及增强算法透明性、可解释性的重要意义。了解和参与会显著增强公众对算法的信任程度。

不同群体对算法感知的差异分析

- 本部分采用卡方检验的方法。

- 卡方检验主要是比较定类变量与定类变量之间的差异性分析。通过统计样本的实际观测值与理论推断值之间的偏离程度，实际观测值与理论推断值之间的偏离程度就决定卡方值的大小。

- 如果卡方值越大，二者偏差程度越大；反之，二者偏差越小；若两个值完全相等时，卡方值就为0，表明理论值完全符合。

- 问卷在基本信息部分询问了被调查者的性别、年龄、地区、学历、职业和收入情况，基于此信息，课题组又对调查结果进行了不同维度的分组交叉分析，得到了一些差异性结论。

不同群体对算法感知的差异分析

- 不同性别的算法感知差异

交叉(卡方)分析结果 数字(占比)

题目	名称	30、性别(%)		总计	χ^2	p
		女	男			
3、【单选】算法总体了解程度	了解其目的意图	659(36.17)	635(37.35)	1294(36.74)		
	了解其运行原理	115(6.31)	220(12.94)	335(9.51)		
	只是会用，不了解所以然	1032(56.64)	808(47.53)	1840(52.24)	64.798	0.000**
	自己会编写算法	16(0.88)	37(2.18)	53(1.50)		
总计		1822	1700	3522		

* $p<0.05$ ** $p<0.01$

表3-1 性别对于算法总体了解程度的差异关系

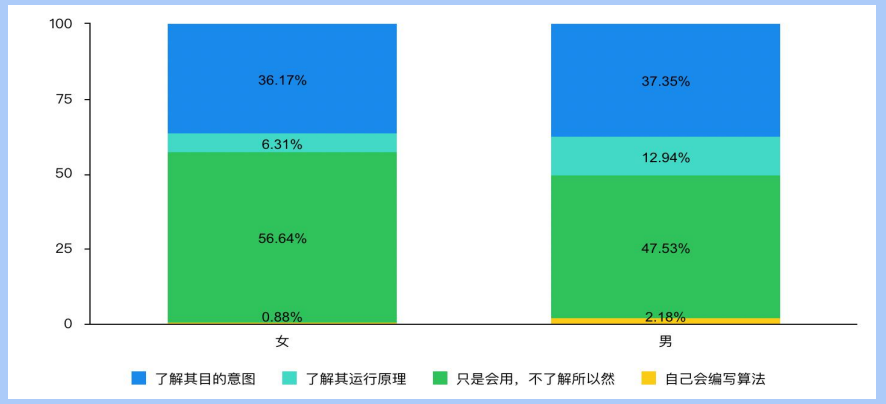

图3-1 性别和算法总体了解程度的交叉图

从表3-1可以看出：不同性别样本对于算法总体了解程度呈现出显著性($p<0.05$)，意味着不同性别样本对于算法总体了解程度共1项均呈现出差异性，可结合括号内百分比进行差异对比。
性别对于算法总体了解程度呈现出0.01水平显著性($chi=64.798, p=0.000<0.01$)，通过百分比对比差异可知，女选择只是会用，不了解所以然的比例56.64%，明显高于男的选择比例47.53%。
总结可知：不同性别样本对于算法总体了解程度全部均呈现出显著性差异。详见图3-1。

不同群体对算法感知的差异分析

● 不同年龄的算法感知差异

交叉(卡方)分析结果 数字(占比)

题目	名称	31、年龄(%)				总计	χ^2	p
		18岁以下	19-35	36-59	60以上			
3、【单选】算法总体了解程度	了解其目的意图	28(34.15)	989(39.72)	266(29.65)	11(20.75)	1294(36.74)		
	了解其运行原理	2(2.44)	248(9.96)	82(9.14)	3(5.66)	335(9.51)		
	只是会用，不了解所以然	49(59.76)	1211(48.63)	542(60.42)	38(71.70)	1840(52.24)	56.579	0.000**
	自己会编写算法	3(3.66)	42(1.69)	7(0.78)	1(1.89)	53(1.50)		
总计		82	2490	897	53	3522		

* $p<0.05$ ** $p<0.01$

表3-2 年龄对于算法总体了解程度差异关系

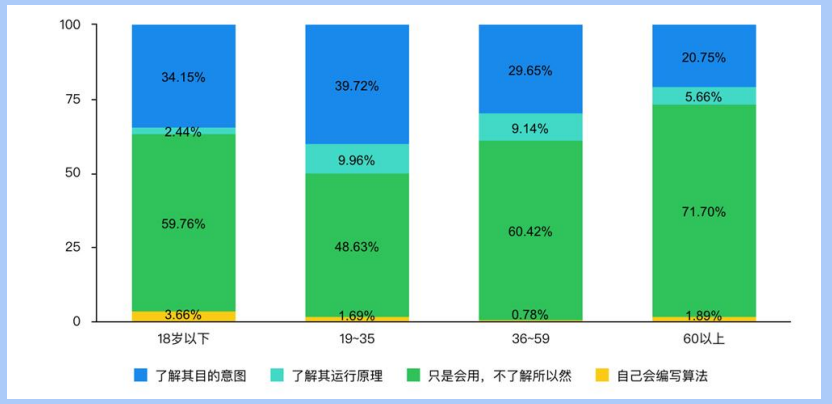

图3-2 年龄和算法总体了解程度的交叉图

从表3-2可以看出：不同年龄样本对于算法总体了解程度呈现出显著性（$p<0.05$），意味着不同年龄样本对于算法总体了解程度共1项均呈现出差异性，可结合括号内百分比进行差异对比。

年龄对于算法总体了解程度呈现出0.01水平显著性（chi=56.579, $p=0.000<0.01$），通过百分比对比差异可知，60以上选择“只是会用，不了解所以然”的比例71.70%，会明显高于平均水平52.24%。

36~59选择只是“会用，不了解所以然”的比例60.42%，会明显高于平均水平52.24%。

总结可知：不同年龄样本对于算法总体了解程度全部均呈现出显著性差异。详见图3-2。

不同群体对算法感知的差异分析

●不同学历的算法感知差异

交叉(卡方)分析结果 数字(占比)

题目	名称	32、学历(%)				总计	χ^2	p
		中专/职业学校	大专或本科	研究生及以上	高中及以下			
3、【单选】算法总体了解程度	了解其目的意图	13(26.00)	509(35.08)	743(40.21)	29(16.76)	1294(36.74)		
	了解其运行原理	1(2.00)	136(9.37)	184(9.96)	14(8.09)	335(9.51)		
	只是会用，不了解所以然	36(72.00)	794(54.72)	884(47.84)	126(72.83)	1840(52.24)	66.558	0.000**
	自己会编写算法	0(0.00)	12(0.83)	37(2.00)	4(2.31)	53(1.50)		
总计		50	1451	1848	173	3522		

* $p<0.05$ ** $p<0.01$

表3-3 学历对于算法总体了解程度的差异关系

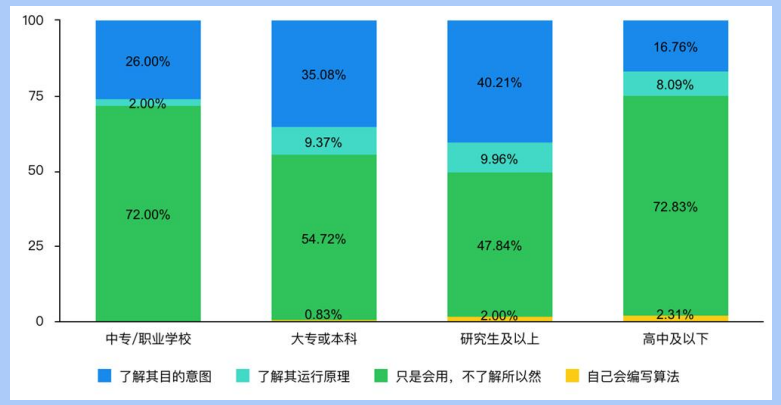

图3-3 学历和算法总体了解程度的交叉图

从表3-3可以看出：不同学历样本对于算法总体了解程度呈现出显著性（p<0.05），意味着不同学历样本对于算法总体了解程度共1项均呈现出差异性，可结合括号内百分比进行差异对比。
学历对于算法总体了解程度呈现出0.01水平显著性（chi=66.558, p=0.000<0.01），通过百分比对比差异可知，高中及以下选择“只是会用，不了解所以然”的比例72.83%，会明显高于平均水平52.24%。
中专/职业学校选择“只是会用，不了解所以然”的比例72.00%，会明显高于平均水平52.24%。
总结可知：不同学历样本对于算法总体了解程度全部均呈现出显著性差异。详见图3-3。

不同群体对算法感知的差异分析

●不同职业的算法感知差异

交叉(卡方)分析结果 数字(占比)

题目	名称	33、职业(%)								总计	χ^2	p
		个体户或自由职业者	企业/公司工作人员	党政机关事业单位工作人员	全日制学生	农民	工人（及农民工）	无业/下岗/失业人员	服务业工作人员			
3、【单选】算法总体了解程度	了解其目的意图	43(28.10)	192(38.79)	335(34.64)	667(40.35)	6(21.43)	8(27.59)	13(21.67)	30(21.90)	1294(36.74)	105.353	0.000**
	了解其运行原理	10(6.54)	75(15.15)	76(7.86)	153(9.26)	5(17.86)	0(0.00)	6(10.00)	10(7.30)	335(9.51)		
	只是会用，不了解所以然	98(64.05)	213(43.03)	548(56.67)	810(49.00)	17(60.71)	21(72.41)	37(61.67)	96(70.07)	1840(52.24)		
	自己会编写算法	2(1.31)	15(3.03)	8(0.83)	23(1.39)	0(0.00)	0(0.00)	4(6.67)	1(0.73)	53(1.50)		
总计		153	495	967	1653	28	29	60	137	3522		

* p<0.05 ** p<0.01

表3-4 职业对于算法总体了解程度的差异关系

图3-4 职业和算法总体了解程度的交叉图

从表3-4可以看出：不同职业样本对于算法总体了解程度呈现出显著性（p<0.05），意味着不同职业样本对于算法总体了解程度共1项均呈现出差异性，可结合括号内百分比进行差异对比。

职业对于算法总体了解程度呈现出0.01水平显著性（chi=105.353, p=0.000<0.01），通过百分比对比差异可知，工人（及农民工）选择“只是会用，不了解所以然”的比例72.41%，会明显高于平均水平52.24%。服务业工作人员选择“只是会用，不了解所以然”的比例70.07%，会明显高于平均水平52.24%。

总结可知：不同职业样本对于算法总体了解程度全部均呈现出显著性差异。详见图3-4。

不同群体对算法感知的差异分析

●不同月平均收入的算法感知差异

交叉(卡方)分析结果 数字(占比)

题目	名称	34、月平均收入(%)					总计	χ^2	p
		10000元以上	1000元以下	1001-3000元	3001-6000元	6001-10000元			
3、【单选】算法总体了解程度	了解其目的意图	256(38.73)	326(37.43)	296(37.33)	191(32.99)	225(36.41)	1294(36.74)		
	了解其运行原理	112(16.94)	65(7.46)	78(9.84)	36(6.22)	44(7.12)	335(9.51)		
	只是会用，不了解所以然	273(41.30)	476(54.65)	405(51.07)	344(59.41)	342(55.34)	1840(52.24)	96.406	0.000**
	自己会编写算法	20(3.03)	4(0.46)	14(1.77)	8(1.38)	7(1.13)	53(1.50)		
总计		661	871	793	579	618	3522		

* p<0.05 ** p<0.01

表3-5 月平均收入对于算法总体了解程度的差异关系

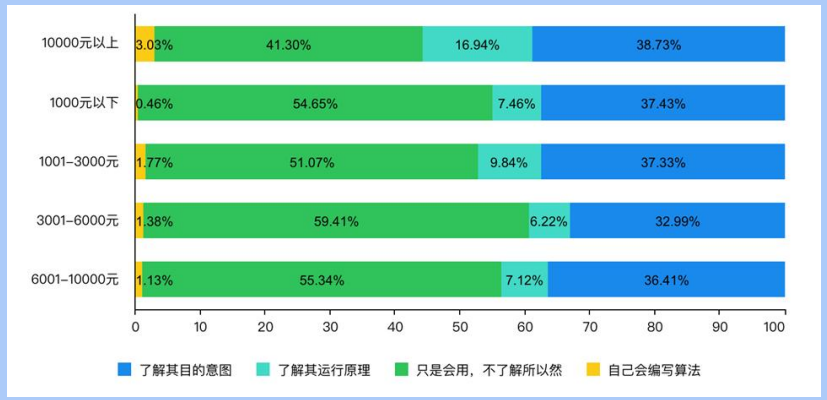

图3-5 月平均收入和算法总体了解程度的交叉图

从表3-5可以看出：不同月平均收入样本对于算法总体了解程度呈现出显著性（p<0.05），意味着不同月平均收入样本对于算法总体了解程度共1项均呈现出差异性，可结合括号内百分比进行差异对比。

月平均收入对于算法总体了解程度呈现出0.01水平显著性（chi=96.406, p=0.000<0.01），通过百分比对比差异可知，3001-6000元选择“只是会用，不了解所以然”的比例59.41%，会明显高于平均水平52.24%。

总结可知：不同月平均收入样本对于算法总体了解程度全部均呈现出显著性差异。详见图3-5。

不同群体对算法感知的差异分析

●不同常住区域的算法感知差异

交叉(卡方)分析结果 数字(占比)

题目	名称	35、常住区域(%)					总计	χ^2	p
		农村	地级市及县城	港澳台地区	直辖市	省会城市			
3、【单选】算法总体了解程度	了解其目的意图	39(20.10)	357(33.84)	26(39.39)	464(39.29)	408(39.77)	1294(36.74)		
	了解其运行原理	14(7.22)	69(6.54)	7(10.61)	139(11.77)	106(10.33)	335(9.51)		
	只是会用，不了解所以然	138(71.13)	622(58.96)	29(43.94)	562(47.59)	489(47.66)	1840(52.24)	90.008	0.000**
	自己会编写算法	3(1.55)	7(0.66)	4(6.06)	16(1.35)	23(2.24)	53(1.50)		
总计		194	1055	66	1181	1026	3522		

* $p<0.05$ ** $p<0.01$

表3-6 常住区域对于算法总体了解程度的差异关系

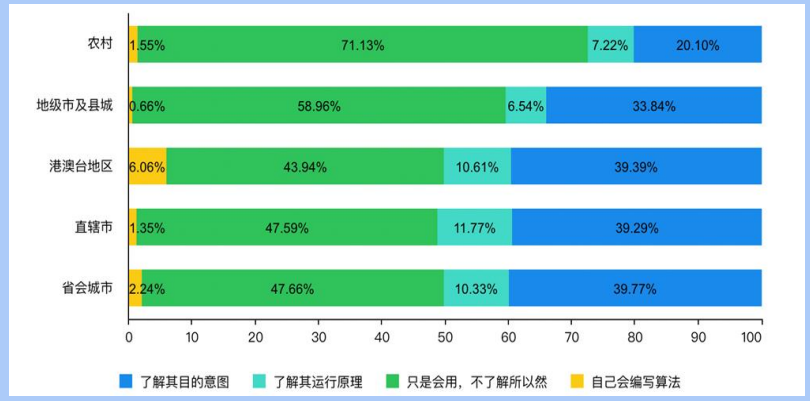

图3-6 常住区域和算法总体了解程度的交叉图

从表3-6可以看出：不同常住区域样本对于算法总体了解程度呈现出显著性（$p<0.05$），意味着不同常住区域样本对于算法总体了解程度共1项均呈现出差异性，可结合括号内百分比进行差异对比。

常住区域对于算法总体了解程度呈现出0.01水平显著性（chi=90.008, $p=0.000<0.01$），通过百分比对比差异可知，农村受调查者选择“只是会用，不了解所以然”的比例71.13%，会明显高于平均水平52.24%。

地级市及县城选择“只是会用，不了解所以然”的比例58.96%，会明显高于平均水平52.24%。

总结可知：不同常住区域样本对于算法总体了解程度全部均呈现出显著性差异。详见图3-6。

不同群体对算法感知的差异分析

●不同常住省份的算法感知差异

题目	名称	36、常住省份																																		总计	X^2	校正X^2	P
		北京	浙江	山东	陕西	河南	内蒙古	广东	河北	辽宁	天津	湖南	江苏	安徽	四川	吉林	江西	重庆	甘肃	福建	湖北	新疆	黑龙江	上海	山西	香港	广西	海南	西藏	云南	澳门	青海	宁夏	贵州	台湾				
3、【单选】算法总体了解程度	只是会用，不了解所以然	434	32	205	54	216	33	86	65	42	27	44	47	30	36	88	10	51	15	37	106	14	13	44	22	2	13	15	12	10	16	5	2	12	2	1840	150.874	150.874	0.001***
	了解其目的意图	371	27	94	26	123	20	78	32	33	26	33	44	25	30	48	10	34	6	24	85	9	14	32	11	3	12	8	11	4	10	2	1	8	0	1294			
	自己会编写算法	10	0	4	3	5	1	3	0	2	1	3	2	2	1	2	0	0	2	1	2	0	1	4	0	1	0	1	1	0	0	0	0	0	1	53			
	了解其运行原理	107	5	27	6	16	6	27	10	10	4	5	14	7	6	9	4	11	1	5	22	1	3	13	1	3	5	2	1	1	1	0	0	2	0	335			
合计		922	64	330	89	360	60	194	107	87	58	85	107	64	73	147	24	96	24	67	215	24	31	93	34	9	30	26	25	15	27	7	3	22	3	3522			

注：***、**、*分别代表1%、5%、10%的显著性水平

表3-7 常住省份对于算法总体了解程度的差异关系

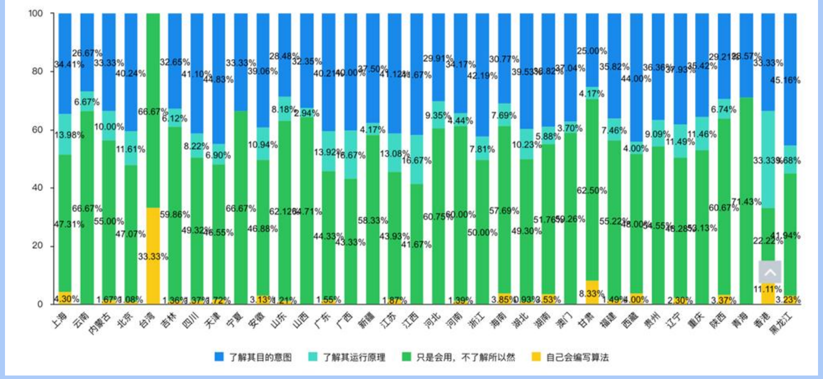

图3-7 常住省份和算法总体了解程度的交叉图

从表3-7可以看出：不同常住省份样本对于算法总体了解程度呈现出显著性（p<0.05），意味着不同常住省份样本对于算法总体了解程度共1项均呈现出差异性，可结合括号内百分比进行差异对比。

常住省份对于算法总体了解程度呈现出0.01水平显著性（chi=150.874, p=0.001<0.01），通过百分比对比差异可知，黑龙江选择“了解其目的意图”的比例45.16%，会明显高于平均水平36.74%。天津选择“了解其目的意图”的比例44.83%，会明显高于平均水平36.74%。青海选择“只是会用，不了解所以然”的比例71.43%，会明显高于平均水平52.24%。云南选择“只是会用，不了解所以然”的比例66.67%，会明显高于平均水平52.24%。

总结可知：不同常住省份样本对于算法总体了解程度全部均呈现出显著性差异，但背后更具体的原因有待挖掘。

公众算法认知的影响因素

- 本部分采用有序逻辑回归的分析方法

➢有序Logit回归分析用于研究X对于Y的影响关系。有序逻辑回归是对逻辑回归的拓展，用来处理有序分类问题。有序逻辑回归的基本原理和逻辑回归相同，都是使用Sigmoid函数来预测分类结果

➢有序Logit回归分析时，首先进行模型平行性检验，如果p值大于0.05，说明满足平行性检验，如果p值小于0.05，说明不满足平行性检验

➢满足平行性检验后，接着再具体研究影响关系情况即可，比如是正向影响还是影响关系等

➢除此之外，还可以研究有序Logit回归分析的模型构建公式，以及模型的预测准确率情况等

公众算法认知的影响因素

➢将性别，年龄，学历，职业，月平均收入，常住区域，常住省份作为自变量，并将算法总体了解程度作为因变量进行有序logistic回归分析，并且使用Logit连接函数进行研究。详情见表4-1。

名称	选项	频数	百分比
算法总体了解程度	了解其目的意图	1294	36.74%
	了解其运行原理	335	9.51%
	只是会用，不了解所以然	1840	52.24%
	自己会编写算法	53	1.50%
	总计	3522	100.0

表4-1 有序logistic回归分析因变量频数分布

公众算法认知的影响因素

- 首先对模型整体有效性进行分析（模型似然比检验），从表4-2可知：此处模型检验的原定假设为：是否放入自变量（性别，年龄，学历，职业，月平均收入，常住区域，常住省份）不影响模型质量结果；分析显示拒绝原假设（chi=79.908,p=0.000<0.05），即说明本次构建模型时，放入的自变量具有有效性，本次模型构建有意义。

模型	-2倍对数似然值	卡方值	*df*	*p*	AIC 值	BIC 值
仅截距	7001.736					
最终模型	6921.828	79.908	7	0.000	6941.828	7003.496

表4-2 有序logistic回归模型似然比检验

公众算法认知的影响因素

➢从表4-3可知，将性别，年龄，学历，职业，月平均收入，常住区域，常住省份共7项为自变量，而将算法总体了解程度作为因变量进行有序logistic回归分析，从上表可以看出，模型伪R平方值（McFadden R方）为0.011，意味着性别，年龄，学历，职业，月平均收入，常住区域，常住省份可以解释算法总体了解程度的1.1%变化原因。以及模型公式如下：

项	项	回归系数	标准误	z 值	Wald χ^2	p 值	OR值	OR值95% CI
因变量阈值	了解其目的意图	0.123	0.282	0.438	0.192	0.661	0.884	0.509 ~ 1.535
	了解其运行原理	0.526	0.282	1.866	3.481	0.062	0.591	0.340 ~ 1.027
	只是会用，不了解所以然	4.915	0.315	15.593	243.138	0.000	0.007	0.004 ~ 0.014
自变量	性别	-0.172	0.066	-2.603	6.773	0.009	0.842	0.739 ~ 0.958
	年龄	0.327	0.066	4.989	24.893	0.000	1.387	1.220 ~ 1.577
	学历	0.054	0.056	0.952	0.906	0.341	1.055	0.945 ~ 1.178
	职业	0.070	0.025	2.786	7.764	0.005	1.073	1.021 ~ 1.127
	月平均收入	0.056	0.024	2.301	5.296	0.021	1.058	1.008 ~ 1.109
	常住区域	-0.121	0.026	-4.731	22.383	0.000	0.886	0.843 ~ 0.932
	常住省份	0.005	0.004	1.428	2.039	0.153	1.005	0.998 ~ 1.013

McFadden R 方：0.011
Cox 和 Snell R^2：0.022
Nagelkerke R^2：0.022

表4-3 有序logistic回归模型分析结果汇总

公众算法认知的影响因素

- logit[P（算法总体了解程度〈=了解其目的意图）/（1-P（算法总体了解程度〈=了解其目的意图））]=0.123-0.172*性别+0.327*年龄+0.054*学历+0.070*职业+0.056*月平均收入-0.121*常住区域+0.005*常住省份
- logit[P(算法总体了解程度<=了解其运行原理)/(1-P(算法总体了解程度<=了解其运行原理))]=0.526-0.172*性别+0.327*年龄+0.054*学历+0.070*职业+0.056*月平均收入-0.121*常住区域+0.005*常住省份
- logit[P(算法总体了解程度<=只是会用，不了解所以然)/(1-P(算法总体了解程度<=只是会用，不了解所以然))]=4.915-0.172*性别+0.327*年龄+0.054*学历+0.070*职业+0.056*月平均收入-0.121*常住区域+0.005*常住省份

公众算法认知的影响因素

➢总结分析可知：年龄、职业、 月平均收入、性别和常住区域会对算法总体了解程度产生显著的影响关系；学历，常住省份对算法总体了解程度产生影响较小。详情见图4-1。

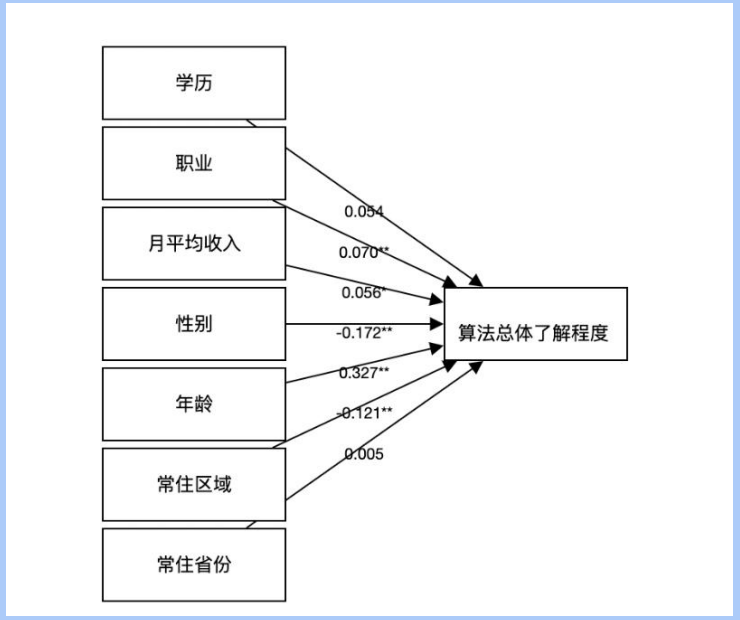

图4-1 模型结果图

调研结论及政策建议

- 从调查结果的整体分析中可以看出，在算法领域存在信息不对称，大部分使用者对算法缺乏深入了解，这种信息不对称会导致机会主义行为和市场失灵，具体存在两个问题：

➢ 第一，隐藏特征会导致逆向选择。

➢ 第二，隐藏行为会导致道德风险。详情见表5-1。

完全信息且无合同法下的博弈

200 200	250 100
0 0	0 0

完全信息且有合同法下的博弈

200 200	150 200
0 0	0 0

信息不对称且有合同法下的博弈

200 200	250 100
0 0	0 0

表5-1 博弈分析表

调研结论及政策建议

- 在完全信息条件下，合同法的赔偿可以激励企业保护消费者个人信息，但是在信息不对称条件下，合同法失灵，企业存在滥用消费者个人信息的激励，这同时也会使社会总福利下降。所以为了消除这种机会主义行为、恢复竞争市场效率，就必须有一部对算法进行监管的法律。结合本次调查情况，报告建议：

➢ 一是积极完善算法问责制，建立健全算法技术使用和管理的制度体系。

- 首先要建立全周期的算法监管制度。
- 其次，更好应对算法给法律带来的挑战。

➢ 二是提高算法的透明度。

- 首先，完善算法可解释要求。

调研结论及政策建议

- 其次，完善算法的事前备案和审查机制。
- 最后，鼓励和督促企业和行业建立自律机制，探索算法审计监督、商业第三方评估以及算法保险补偿等机制，强化对算法的监督和保障。

➢ 三要立足我国行业发展实践，不断提高算法监管水平。

- 首先，政府部门要厘清监管职责分工，适时建立算法或者人工智能的专门技术指导部门。
- 其次，要优化算法监管手段，采取多层次的监管措施。

➢ 四是以实现公平正义为目标推进算法发展，引导算法应用向上向善。

- 首先，政府在算法监管过程中，要坚持正确政治方向、舆论导向、价值取向，规范信息分发行为和秩序，推动企业借助算法加强正能量传播。

调研结论及政策建议

- 其次，企业要明确算法应用的目标，制定明确的算法应用目标和方向，恪守公平、正义、无偏见的伦理原则。
- 最后，企业应当把握正确方向，增强责任意识。

➢ 五是倡导多元社会主体的参与。

- 多元社会主体的参与是实现“善治”的必要条件。
- 同时，培育公民的算法素养，使公众更加关注和理解算法应用情况和治理实践，从而实现算法的协同共治。

感谢倾听，欢迎指正！

支振锋 研究员 教授

中国社会科学院法学研究所

22820199@qq.com