

## 前沿视线

## 学人关注

## 『显性』平等与『隐性』不公：解密ChatGPT中的性别问题

·编者按·

人工智能领域长期以来存在性别不平等的问题，当最近爆火的AI应用程序ChatGPT被问及一些性别话题时，它呈现出了友好的姿态，引得一片赞誉声，也吸引了上亿的活跃用户。那么，ChatGPT真的能解决AI长期被诟病的性别偏见吗？本文作者带我们一起解密ChatGPT中的性别问题，文章围绕算法程序规避下的“显性”平等、技术背后资本操控的“隐性”不公以及ChatGPT与性别平等的现实审视进行了讨论。

■ 戴雪红

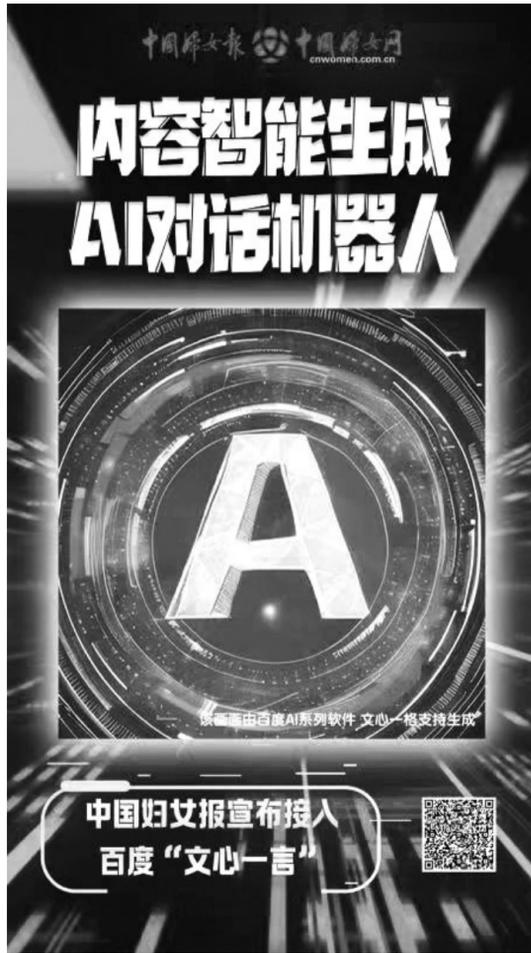
人工智能领域存在性别不平等的问题已经是老生常谈了。当最近爆火的AI应用程序ChatGPT被问及一些性别话题时，它的回答常常是这样的：“作为一种人工智能，我认为歧视女性是一种不道德和不公正的行为。每个人都应该被平等对待，无论他们的性别、种族、宗教或其他身份特征。歧视女性不仅违反了基本的人权原则，也会导致社会的不公平和不平等。”由此可见，ChatGPT在应对性别问题上呈现出友好的姿态，引得一片赞誉声，也吸引了上亿的活跃用户。那么，ChatGPT真的能解决AI长期被诟病的性别偏见吗？当ChatGPT遭遇性别问题，又能给我们带来哪些新变化与新挑战？

## 算法程序规避下的“显性”平等

ChatGPT是由美国公司OpenAI开发的一种大型语言模型，其主要功能是生成语言文本。OpenAI是“开放人工智能”的缩写，宗旨是通过开放的研究和合作，将AI的好处带给更多的人。Open的意思是“打开”，正如潘多拉打开了一个盒子，释放了各种疾病、灾难和希望等。而在AI领域中，恰恰也存在着一个“黑盒子”——通常指的是一种模型或算法，OpenAI的使命之一是尽可能地减少“黑盒子”带来的负面影响。例如，Tay是微软公司2016年推出的一个聊天机器人，然而，由于其设计不良和缺乏对滥用的限制，Tay在推特上遭受了大量恶意用户的干扰，发布了涉及性别歧视与种族歧视等言论，引起了公众的不满和批评。Tay仅上线不到一天即死亡。“微软Tay事件”是AI领域的重大事件，是一次由于缺乏适当的监督和控制而导致AI产生不当行为的事件。为了避免出现类似的事件，OpenAI建立了一些内部规则和流程来确保ChatGPT的安全性和可靠性，以符合道德和法律的要求。到目前为止，ChatGPT已经发展成为一种具有一定道德标准且保持中立态度的使用程序。值得一提的是，不同于微软Tay被设定为19岁的少女，ChatGPT是没有性别设定的。

针对AI模型中存在的性别不平等的问题，当下的ChatGPT更加注重语言的包容性和多样性。在ChatGPT的训练数据中，对于不同性别、种族、文化和背景的人的言论进行了广泛的收集和分析，使用相同数量的男性和女性数据，从而确保模型学习到的知识是平衡的。另外，ChatGPT能够更好地理解语言中的隐含含义，这有助于避免在对话中出现性别偏见等不恰当的言论。在ChatGPT的设计中，模型被训练成能够识别并理解社交上下文和语境，这可以使ChatGPT更加准确地理解和回应包括性别在内的社会问题。ChatGPT已经通过大量的数据训练来识别性别不公，并学会了避免使用具有性别刻板印象的单词和短语，这就可以用来对文本进行性别敏感语言的处理和删选。ChatGPT也能够用于传播关于性别平等的信息和知识，从而提高公众对性别平等话题的意识和理解。不过，无论技术背后的设计理念是否趋于平等，ChatGPT对于性别歧视问题是一种“显性”平等的思维逻辑。这

2月14日，中国妇女报宣布成为百度“文心一言”（英文名：ERNIE Bot）首批生态合作伙伴。后续，中国妇女报将全面体验并接入“文心一言”的能力。



导致了很多人认为其在算法方面突破了性别不公，那么，事实真相究竟如何呢？

## 技术背后资本操控的“隐性”不公

只有了解AI的性别偏见是怎么来的，我们才能知道ChatGPT为了纠偏做了些什么，还存在哪些问题。首先是数据源的偏见。数据源是从网络中抓取的，偏见来自搜集的数据源与现实情况的偏离。算法由算法工程师编写，但背后是工程师代表的利益集团的倾向。数据标注的质量和数量对ChatGPT的性能和应用场景都会产生重要影响。例如，当人们问ChatGPT一些关于性别、性少数或种族歧视等问题时，ChatGPT一般这样回答：“作为一个AI程序，我没有情感和个人信仰，我不会有任何偏见或倾向性。”为什么会出这种答案？原因在于ChatGPT的技术编码对于性别问题的敏感度提高，于是当聊天中出现相关话题时，它就会机械性地选择回答已经被植入的符合社会要求的标准答案。但是ChatGPT也不是绝对的道徳理性状态，当有一些不怀好意的人引导、质疑、威胁它时，它就会展现出进退两难、模棱两可的姿态。而且，ChatGPT给出的答案就一定标准答案吗？并不一定，但这个答案是出由ChatGPT公司所有标注师综合价值观形成的结果，某种程度上它代表了在性别问题上形成的共识。它似乎显得非常成熟，永远保持中立的态度。简言之，ChatGPT的“成熟人设”是由算法工程师和数据标注师合力打造的。

实际上，ChatGPT的背后是互联网巨头微软，这种资本大鳄的垄断发展模式必将在ChatGPT的发展上重演。关于性别问题的标准答案是由垄断了数据的人说了算，掌握了标准答案就是掌握了话语权，甚至我们很难说清楚，是我们生产的数据影响了标准答案，还是标准答案影响了我们。因此，在这种貌似公开、平等的程序背后，我们要关注的不仅仅是ChatGPT的数据来源，ChatGPT的AI执行系统及其背后的算法机制，还要关注技术背后的资本支持，这些都离不开人类的操作与干预。也许在不远的未来，ChatGPT就会为你推荐适合你的产品，影响你的思维方式和判断力，因为它不仅仅是机器，它还是人类的对话者，它掌握着人类赋予的话语权。当ChatGPT

得到了人类的信任之后，性别偏见就会以另一种方式重新浮现，它会为女性对话者推荐护肤品、母婴用品和家庭用品，为男性对话者介绍理财产品、交通工具和楼盘信息，对这种模式的理解可以参考谷歌和百度发展成熟之后的广告推送模块。因为ChatGPT的背后是资本的力量，ChatGPT作为资本的新宠必然是存在帮助资本实现价值增值的可能，说不定ChatGPT将是继直播平台后的新一任“带货王”。

## ChatGPT与性别平等的现实审视

ChatGPT或者AI的性别不平等的问题并不是独立存在的，其实是反映了社会本身的偏见，不会因为一款人工智能产品的出现而简单地消失。只有实施思想和制度上的变革，ChatGPT才能被进一步改进与完善。尽管AI领域一直被认为是男性主导的领域，但事实上，越来越多的女性进入了这个领域，特别是在研究和开发AI算法和应用程序方面。然而，女性进入AI领域会面临诸多障碍，例如，招聘和晋升中存在的性别不公、工资差距和性别平等等问题等。可能我们很难在表面上捕捉到ChatGPT的性别歧视现象，但是现实的性别问题往往是隐而不见的、沉默不语的，这远远不是AI等技术可以帮助解决的。

总之，ChatGPT和其他AI本身并不具有道德、价值观或利益的色彩，它们只是通过算法、数据和模型进行编程的工具，最终取决于人类如何使用它们。这些工具可以被用于积极的和有益的目的，也可以被用于负面和有害的目的。实现性别平等不仅需要培养女性从事先进科技研究的能力，使其能与男性一同引领新技术的发展方向；还需要每个人的努力，以共同打破性别壁垒和陈旧观念，为女性提供更多的机会和平等的待遇。解密ChatGPT或者AI可能存在的性别歧视问题要从现实社会中开始，只有建立一个公正和平等的社会和制度，才能使ChatGPT更加智能、更加人性化。这意味着实现性别平等的道路是一个充满冒险和挑战的漫长旅程，需要我们勇敢地面对各种不确定性与困难。

（作者为南京大学马克思主义学院教授、博士生导师）

注：本文是“江苏高校学习贯彻党的二十大精神”项目“二十大精神专题研究项目”的阶段性成果。

## 阅读提示

高价彩礼是重要的民生问题，已经引发一系列社会后果。2023年2月13日发布的中央一号文件再次指出，要“扎实开展高价彩礼、大操大办等重点领域突出问题专项治理”。自2017年中央一号文件发布以来，国家就对彩礼问题予以高度关注，相继出台相关文件，充分说明治理农村地区彩礼问题的重要性与紧迫性。治理高价彩礼既是乡村振兴的现实需要，也是促进人口长期均衡发展的重要前提。

■ 靳小怡

2023年2月13日发布的中央一号文件再次指出，要“扎实开展高价彩礼、大操大办等重点领域突出问题专项治理”。自2017年中央一号文件发布以来，国家就对彩礼问题予以高度关注，相继出台相关文件，充分说明治理农村地区彩礼问题的重要性与紧迫性。治理高价彩礼既是乡村振兴的现实需要，也是促进人口长期均衡发展的重要前提。

## 高价彩礼后果严重，国家治理势在必行

高价彩礼是重要的民生问题，已经引发一系列社会后果，不利于乡村振兴战略的顺利实施和人口社会可持续发展。高价彩礼并不意味着女性地位的提升，也不能保证婚姻的幸福，它带来的是社会不同群体的利益“普遍受损”。

首先，由于男女比例失调，女性成为婚姻市场的稀缺资源，高价彩礼有诱发“拐卖妇女”的可能，有兄弟的女性往往成为“换婚”的牺牲品，女性权益进一步受到侵害。

其次，遭受过成婚困难的已婚男性由于难以支付高价彩礼，往往经媒人介绍，从经济发展更为滞后的外省娶妻。其妻子年龄明显较小，婚后发生婚姻暴力的概率较高，婚姻满意度较差。

再次，部分农村大龄未婚男性具有被迫性、聚集性和脆弱性三大特征，高价彩礼致使一些家庭“因婚致贫、因婚返贫”，高价彩礼引发“骗婚”的现象在欠发达的农村地区频发，给大龄未婚男性及其老年父母以及社区发展均带来不利影响。

最后，近十年来，高价彩礼影响了农村的初婚年龄。不但使30岁以上的“晚婚”现象增加，成为农村男性成婚难的主因，还致使19岁以下的“早婚”现象增多，这是农村男性尽早抢夺稀缺的可婚配女性人口的策略，也是男性婚姻挤压加重的直接表现，对农村婚姻的稳定性与生育率产生负面影响。

## 高价彩礼成因复杂，决定了治理的艰巨性

高价彩礼问题的成因主要包括以下四方面。

第一，人口性别失衡带来的男性婚姻挤压是直接原因。我国人口性别失衡的主要诱因是长期偏高的出生性别比，随着1980年以后出生的过剩男性人口逐步进入婚姻市场，婚姻挤压的累积效应日益凸显。传统“女性向上匹配”的婚配模式带来“自西向东、从农村到城市”的女性婚姻迁移。因此男性婚姻挤压高度集中在以中西部为主的欠发达农村地区。近年来，彩礼问题在这些地区愈演愈烈，正是男性婚姻挤压日趋严峻的集中体现。

第二，父系家族制度与嫁娶婚姻模式是根源性原因。“单系偏重”的父系家族制度与嫁娶婚姻模式是彩礼存续的根源性原因。“单系偏重”以父系、父权、从夫居为主要表现形式，决定了女性在家庭和社会中的从属地位，这在相对落后的中西部农村是比较突出的问题，也是“重男轻女”“男孩偏好”和“天价彩礼”盛行的共同诱因。所谓“嫁出去的女儿，泼出去的水”，这种“买断”和“交割”成为男方向女方父母进行经济补偿的基本逻辑，并在男性婚姻挤压条件下抬升“新娘价格”。加之，农村地区父母普遍存在“不孝有三，无后为大”“延续香火”等观念，这种“父辈责任”和社会习俗在家庭社区层面进一步助推彩礼攀升。

第三，市场化加重消费主义与攀比心理，婚姻被过度物化。随着农民物质生活条件的明显改善，利益至上的“交换观念”也在农村婚嫁习俗中打上了烙印，劳动致富与超前消费为彩礼上涨提供了可能。在社会关系相对稳定、以“熟人社会”为特征的农村社区，文化风俗对彩礼影响较大，受到攀比心理影响，村民往往将高价彩礼与“有面子”画上等号，将彩礼金额与家族的社会地位、经济实力联系起来，认为较高的彩礼才显得新娘有身价、娘家有地位。

最后，城镇化、现代化重塑农村代际关系，分家提前进一步抬升男方婚嫁花费。随着城市化进程推进，大量农村劳动力外出务工深深影响着农村传统的代际关系，适婚青年个体意识凸显，夫妻关系重要性上升，老年父母地位下降，彻底改变了传统农村的大家庭居住模式，结婚时即与男方父母分开居住成为主流，婚房成为很多地区成婚的先决条件，再加上物价房价上涨等因素，进一步推动彩礼的上涨与货币化。

## 对症下药，进行常态化的综合治理

根据2018年西安交通大学公共政策与管理学院“百村调查”数据，我们对11省的农村男性婚嫁花费进行全面分析，基于彩礼、婚房、婚礼、谢媒费等多项指标，发现当代中国农村男性婚姻成本的类型特征存在“西部偏彩礼、东部偏婚房”的明显地区差异，说明高价彩礼是西部农村男性成婚难的主要原因、男性婚姻挤压更严重，而东部农村男性结婚花费高的症结在于婚房。因此，仅就彩礼而言，治理的重点和难点在西部，但就目前各类婚嫁花费在城乡同时攀升的现实境况而言，治理的内容和人群应更多元，解决年轻人因过高婚嫁花费而结婚难的问题。

高价彩礼及其他婚嫁花费不断攀升的成因复杂，其治理应是综合性的。目前的治理手段主要从彩礼婚娶数额的规定入手，多面向未婚女性及丈母娘，短期内可以治标，但从长期来看则难以治本，需要与乡村振兴战略、妇女权益保护与发展等工作相结合，进行常态化的综合治理。

其一，减少教育与就业的性别歧视、充分发挥妇女在乡村振兴中的作用，切实提高妇女地位。性别不平等是催生高价彩礼的温床，通过提升女性受教育水平、就业质量和经济独立性，建立性别平等意识，不把婚姻幸福寄托在男方给予的高价彩礼上。

其二，宣传倡导“爱情是婚姻基础”的婚恋观，以具体实例说明“高价彩礼换不来幸福婚姻”，引导适婚青年树立“提升个人能力”“家业自己创”的奋斗观。

其三，面向社区和中老年父母进行移风易俗观念宣传，淡化“帮助儿子成婚才算尽父母责任”“养儿才能防老”的传统观念，加强社会养老保障，保护老年人合法权益，避免因结婚而过度“啃老”的现象。

其四，加大对中西部农村地区的资源投入，缩小地区城乡差距，提升农村人口教育水平与职业保障，增加个人收入，提高婚嫁支付能力，切实提升人民生活质量。

其五，与生育友好型政策对接，建立一揽子的家庭友好型政策，加大房地产市场监管力度，为新婚夫妻购买首套房提供更多福利。

最后，结合当地文化与经济实行精准施策。由于省际乃至不同县域的经济发展差异大，地理与人文环境千差万别，对彩礼的治理应深入不同地区科学调研，因地制宜地制定能被当地大多数群众接受的规章制度。

（作者为西安交通大学公共政策与管理学院教授）

## 研究视窗

## 《性别差异下人工智能技术冲击与工作认同》

作者：程雅馨 程延园 何勤 邱明

为探讨人工智能技术应用对员工工作认同的影响机制及边界条件，文章以“基于任务的模型”和自我决定理论为基础构建有调节的中介效应模型，并利用北京市100家人工智能技术应用型企业的员工数据进行实证检验，研究发现：人工智能技术冲击程度对员工工作认同具有显著正向影响，工作要求和条件感知在其中起到中介作用；性别调节了工作条件感知与员工工作认同之间的关系，即在人工智能技术冲击影响下，与男性相比，工作条件感知变化越明显，女性对工作认同越强。

来源：《华东经济管理》

## 《科学基金助力女性科研人员成长：政策、成效与展望》

作者：于璇 高瑞平

女性科研人员是科学技术研究队伍中的一支关键力量。2010年起，国家自然科学基金委员会推出了支持女性科研人员的一系列举措，得到了科技界的广泛好评。政策实施10年来，国家自然科学基金逐渐成为女性科研人员成长的摇篮，在提高女性在科研中的参与度、增加哺乳期女性申请和获资助机会、助推高层次女性科研人员成长发展、提高科技领域性别平等意识等方面取得了显著的成效。为进一步促进科研领域中的性别平等，结合影响我国女性科技人才成长的主要障碍和政策不足，文章提出新时代要强化政策协同，完善高层次女性科研人员脱颖而出机制，更好地发挥女性在科技活动中的咨询决策作用等具体政策建议。

来源：《中国科学院院刊》2023年第2期

## 《推进人口可持续发展政策下的深圳女性友好型城市建设研究》

作者：周璇 张琪

育龄女性的婚育观和生育观在不同的时代环境中不断变化。在快速城市化过程中，养育成本高企、家庭就业失衡、家庭支持不足、社会配套服务短缺等问题的出现，一定程度上抑制了女性生育意愿。随着进一步完善和落实积极生育措施的各项政策出台，构建女性友好的社会环境和城市环境成为降低“三育”成本、平衡女性工作家庭冲突以及促进家庭发展的重要手段。文章通过对深圳市育龄女性的构成特征进行分析，总结深圳女性友好型城市建设现状，从建筑与城市规划的角度对深圳城市社区空间、办公空间、公共空间等城市空间与设施建设提出针对性建议，倡导从城市空间规划的角度降低生育顾虑、提高女性友好的空间质量，促进性别包容，促进深圳建设女性友好型城市。

来源：《城市观察》2023年第1期

(白晨 整理)