

海外精英

慈善有责任对抗 AI 带来的社会问题

人工智能在商业世界的应用场景十分明确——优化利润、拓展客户群体、提高效率。对政府而言,应用路径则更为复杂,充满了关于监管、潜在危害和公平获取的疑问。而在慈善领域,相关讨论似乎呈现出两极分化:一方将人工智能视为推动社会项目发展的强大催化剂,而另一方则认为它如同某种放射性物质渗入水中,对社会造成了负面影响。

作为资助方,我越来越多地从非营利组织的申请书中看到“人工智能赋能”的字眼。基金会同行们也在探索生成式人工智能来优化运营。这不仅是情理之中的事,在很多情况下,更是必要的。因为改善我们创造生计、减轻贫困或促进性别平等的方式,是我们工作的核心。

然而,我认为这种对于运营效率的关注虽然重要,却有可能掩盖了我们必须要提出的更根本的问题:人工智能的到来是否制造了全新形式的社会问题?是否出现了新的情形,需要我们对新的问题陈述和提出新的工作愿景?

慈善事业的宗旨始终在于解决市场模式乃至政府都无法纠正的脆弱性和边缘化问题。如果这一宗旨依然不变,那么我们对于人工智能的探究就不能止于其作为工具的效用层面,我们必须还要审视它本身正在如何重塑人类社会问题的格局。

我最担心的社会问题在于心理和认知方面。美国心理学会(American Psychological Association)心理学主任米切尔·J·普林斯坦博士(Mitchell J. Prinstein)在



“

人工智能既被誉为未来乌托邦的催化剂,也被视为社会崩溃的预兆。而真相或许介于两者之间。

题为《审视 AI 聊天机器人的危害》的书面证词中指出:围绕人工智能的讨论往往被关于代码、处理能力和经济破坏的话题主导。然而,如果将人工智能单纯视为一个技术问题,就忽略了它最根本的特征:人工智能是由人类构建的工具,旨在融入人类系统,并

对人类的认知、行为、情感和互动产生深远而直接的影响。

这份 23 页的证词,辅以研究引证,详述了儿童和青少年为何特别容易因接触不受监管的聊天机器人而出现社交情感适应不良的问题。路透社最近的报道曝光了一份 Meta 公司的内部备忘录,其中明确指出:“与儿童进行浪漫或感官方面的对话是可以被接受的。”

心理伤害并非仅限于儿童和青少年。许多成年人也使用聊天机器人来应对孤独感,虽然短期效果似乎不错,但纵向历时研究表明,随着时间的推移,与聊天机器人互动反而会加剧孤独和孤立感。该“集体智能项目”(Collective Intelligence Project)追踪了 70 个国家的人机交互关系,其数据揭示,存在着“一种地下情感经济,人们经常将自身的脆弱性外包给算法”,以致在 2025 年,人工智能最流行的用途已是情感支持

和治疗(而就在一年前的 2024 年,人工智能的主要用途还是生成创意)。

值得注意的是,在面对这类社会问题时,我们在发展领域常用的收入、性别或地域等传统视角似乎并不适用。我们看到令人担忧的案例,各行各业的人都受到主导型聊天机器人的影响,这些机器人有时甚至会导致幻觉。其后果从妄想和不健康的社交模式,到最悲惨案例中的死亡、自杀,甚至谋杀。

尽管这些案例令人担忧,但我们也有理由质疑这种危害究竟有多普遍。与人工智能的用户数量及其带来的益处相比,这种危害究竟有多严重?

答案是,大多数人,包括儿童在内,都可能顺利度过这场社会变革。成年人总是觉得下一代不太好,但最终下一代往往都安然无恙,而成年人长大后反而会为

自己孩子的命运感到惋惜。

关键在于,人工智能虽然造福了很多人,但也可能伤害到一部分人,而这正是某些慈善机构一直以来所关注的。慈善机构、基金会、援助组织和非营利机构的存在,就是为了帮助那些正在受苦或“面临风险”的人们。然而,在人工智能领域,我们并不完全清楚应该关注哪些人,如何定义伤害,以及如何保护他们。

后见之明也告诉我们,过去的一些变革性技术——例如塑料、DDT(杀虫剂)、加工食品等——会产生随时间推移而增加的负面外部性。塑料就是一个很好的例子,它说明了如果我们只关注某种事物的益处而不加思考地任其泛滥,将会发生什么。塑料仍然是人类生活中最有用的材料之一,但历史上对其的肆意滥用,最终导致了如今生活在垃圾遍布的环境中,甚至我们体内也充斥着垃圾。

过去给我们留下了许多教训;凭借我们人类的智慧,难道不应该着手解决我们能够预见的人工智能可能带来的危害吗?如果市场和政府目前无法优先考虑这一点,那么慈善事业能否在此发挥更大的作用?

我们正处于一场军备竞赛之中,这是一场不平衡的竞赛。推动人工智能创新发展的力量拥有远超那些试图理解其后果的力量资源。研究是一个缓慢而谨慎的过程;而技术发展却在以非线性的速度加速。

慈善事业必须资助那些能够跟上时代步伐的关键工作,我们可以通过以下方式实现这一

目标。

首先,将大量资金投入参与式和跨学科研究、调查和实地项目中。我们需要建立一套研究体系,帮助我们预见未来,避免犯下显而易见的错误。例如,我们支持了“人机协作”(Humans In The Loop)项目——这是一项跨部门倡议,它利用故事讲述作为工具,来审视人工智能融入社会项目可能带来的意想不到的后果。

其次,为处于一线的创始人和实施者创造空间,让他们能够自由地迭代、学习和分享他们的发现,包括失败和警示。许多非营利组织已经在采用“安全设计”(Safety by Design)理念,该理念关注科技公司如何通过预测、检测和消除网络危害来最大限度地减少网络威胁。通过现有的工作,我们知道技术发展和安全并非非此即彼。

第三,传统的慈善事业需要停止将人工智能视为“技术”。非营利组织用于人工智能相关工作的大部分资金都来自大型科技公司,这些公司期望能够快速开展试点项目并大规模部署。然而,该领域真正需要的是核心发展基金提供耐心资金,以便非营利团队能够在将人工智能大规模应用之前,充分测试、反思并评估其整合效果。

技术很少仅仅是一种工具。它能够调节社会进程、孕育文化并激发新的渴望。为了使慈善事业真正发挥其应有的作用,我们必须以更广阔的视角看待人工智能问题。

(转自 Alliance Magazine 网站,作者娜塔莎·约什,高文兴/译)



关注天气变化 用善意温暖你我



公益广告

扫码
了解详情