



读十

Neo-reading

周刊



扫一扫 读周刊

2025年9月16日 星期二 主编:王永芳 责编:马梦娅 美编:陈昌 版式:陈笑宇 责校:刘明

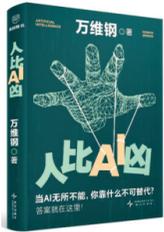
达尔文的呐喊与彷徨 专栏

审视明朝:由古代向近代过渡的新气象 读书

科学作家万维钢:

AI让聪明变便宜,让靠谱更稀缺

□长江日报记者李煦



《人比AI凶》
万维钢 著 新星出版社

科学作家万维钢,去年以一部《拐点》“点”中了2024年高考作文题,准确预言了人工智能(AI)给方方面面带来的冲击。当时,面对即将到来的拐点,他提出一个想法,“人要比AI凶”。一年多过去,他把这个想法变成了一部新书,书名就叫《人比AI凶》。这次他说,我们已经在拐点之中了,AI越来越强大,几年之内,就会有一个AI,而不是创造它的人类,因为独立做出某个科学发现而获得诺贝尔奖。他说,由于四大原因,人还是比AI凶。并且,万维钢强调,这是人类自带的禀赋,所以人天生就比AI凶!这种自信从何而来呢?上周,长江日报《读+》周刊专访了万维钢。



万维钢

当机器人进入千家万户,你会更加体会到人的宝贵

它会解决一些已经困扰人类科学家很多年的问题;
它会做出一系列诺贝尔奖级别,甚至颠覆学科范式的重大科学发现;
它会接管所有繁重、无聊和危险的劳动;
它会解决长寿问题,以至于只要你能活到2045年,你就能活到2100年;
它会让你物质极大丰富,让全人类实现富足……

这像梦一样的愿景,出自万维钢的作品《人比AI凶》,是他基于AI发展一日千里的现状,加上他的走访调研,从而作出的研判。

这不是空中楼阁,而是有着坚实基础。“未来已来”。就拿医学来说,2025年6月底,微软公司的一篇论文公布了一项医疗诊断测试结果。面对300多个真实的疑难杂症医疗案例,允许AI像人类医生一样问患者问题,像人类医生一样要求患者做各种检查并阅读报告。结果,无论是微软专门训练的一个AI智能体,还是市面上现有的主流大模型,准确率都远远超过了人类医生。这不是孤例,还有一些其他研究也证明了现在AI的疾病诊断准确率超过人类医生。

AI不仅已经拥有一定的独立科研能力,而且在很多方面表现出远超人类的智能水平,尤其擅长以人类不适应的方式思考科学问题。万维钢写道:“就在2024年,创造出AlphaFold——一个专门预测蛋白质折叠结构的AI模型——的两位开发者获得了诺贝尔化学奖。我们完全可以期待几年之内,就会有一个AI,而不是创造它的人类,因为独立做出某个科学发现而获得诺贝尔奖。”

可是,面对这样强悍的AI,万维钢坚定地认为,人比AI凶。

首先,人有意识。意识跟人一生的“连贯叙事”很有关系,人从小到大连续的“连贯叙事”一直在给自己讲故事。是那些故事,决定了人心中的“自我”,决定了人接下来想要什么。一个人想要出人头地,是

AI时代,靠谱成为新的智商

在万维钢上一本书《拐点》中,他预言AI会让职场中人更有道德,让人更像君子;在这一部《人比AI凶》里,他进一步阐发了自己的观点,并把这种道德与“靠谱”“可信”“尽责”等等联系起来,认为这是AI时代,让人更有价值的重器。

长江日报记者对这个观点感到好奇,在采访提纲中追问他是否有真实案例支撑或者亲身体验。万维钢索性用手头的大模型结合国外研究成果作了一个调研,然后给记者讲了一个真实故事。

他问大模型:“现实生活中,当今世界,特别是在AI面临的这个关口,有没有研究证据证明一个人的尽责性、可信性、办事靠谱这些能力,会给他带来很好的竞争优势?”

大模型回答:“靠谱是新的智商。在算力把‘会算、会写’的门槛不断压低的年代,可被信任地做事,变得稀缺、更值钱。市场对才华有好感,对可靠有溢价。”

“尽责性-绩效”是人类行为科学里最稳定的关联之一。跨职业的大型元分析显示:尽责性与工作绩效稳定正相关,可迁移到不同岗位与评价指标之上,这一结论在欧美多样性的样本上复现了三十多年。“靠谱”能形成跨领域的长期优势——健康、财富与职业路径的复利。综合纵向研究表明,更高尽责性导致更长

因为他给自己讲了一个奋斗故事。

AI没有自己的“元叙事”,所以它没有——或者至少暂时没有——自己的执念,它暂时不积极地想要什么。所以人才是主动的发起者。总是人想要,AI帮着实现;人提问,AI回答。

第二,人的“想要”是不可预测的。什么样的款式能在明年流行?什么样的电影能爆?这些都很难预测。人的“想要”受到太多因素影响,其中很多都没有形成可供AI学习的数据。你的基因、你所在的文化、你的经历、你不经意间接收到的微妙信息、你早上吃饭了没有,等等因素都可能起作用。所以AI智能再强,也不可能完全代替你做决定。

第三,人会死。所以我们的每一刻都是宝贵的。人说的每一句话,做的每一个动作都值得郑重对待,因为在人有限的生命之中只能说这么多话,做这么多动作。而AI不会死,可以无限自我复制,AI理论上可以同时跟极多人说极多的话,那它这一句又有多珍贵呢?人会死,所以人有更高的道德优先级。

万维钢写道:“当机器人进入千家万户的时候,你会更加体会到人的宝贵。你会更加认同哲学家康德说的:人只能是目的,而不能是手段。所以应该是AI服务我们,而不是我们服务AI。”

第四,人必须与他人共处。人长这么发达的大脑主要不是为了解决自己的生存问题——一般动物的大脑都很小,生存也没问题。我们大脑的主要工作是解决群居问题,是为了琢磨如何与他人相处。灵长类动物的脑容量大小跟族群的大小呈现明显的正相关。我们能创造这么高级的文明,是因为我们善于与他人相处——我们不只竞争,更爱合作,我们能合作的规模越来越大。而与他人相处,不仅仅是一个智能问题,更需要“智慧”。

这些原因决定了,AI必须听人类的,人类才是事情的发起者,而人类的喜好无法预测,所以AI再强也只是助手。

机制包括更好的自我管理、与医生行为。与之密切相关的自控力从童年到成年贯穿健康、财富与守法等多端点结局,这意味着“靠谱”不是短跑优势,是全生命周期的复利资产。

此外,良好声誉带来更高成交价与更快成交;“团队平均尽责性”显著提升绩效。尽责性、可信性与办事靠谱,不仅与更高的个人绩效相关,也通过更低的协作成本,更强的声誉溢价与更稳健的长期行为,在AI放大的分工体系里形成可复制的竞争优势。简言之:AI让聪明变便宜,让靠谱更稀缺。

这还不算,大模型干脆写了一首“小偶”:机器写风云,人心定乾坤;才华如流水,靠谱是炉温;若问胜何处——信任即资本。

万维钢把这个结果发给记者,并讲述了一个故事。多年以前,他在洛斯阿拉莫斯实验室做博士后时,曾经担任一篇论文的审稿人,作者也是个大学教授。他给自己的导师说起这篇论文,导师告诉他,论文作者是个“不诚实”的人。万维钢说:“那时我初出茅庐,这句话给我很大震动。科学家圈子很小,大家抬头不见低头见,一年总会在学术场合遇到几次,给出‘不诚实’的评语,这是相当严厉的。学术共同体是讲究声誉的,我自然要严格评审这篇论文。”

后来他把这篇文章给“拒”了。



读+

读+:能否定义一下,什么叫作“人比AI凶”?
万维钢:如果要给“凶”一个定义,那就是一起做事的时候,始终坐在主驾驶的位置上,去承担主动的发起者角色,并且对大家有控制力。以前我看过一部电影叫《车神传说》,里面有句台词是“人要比车凶”。电影是讲赛车的,赛车在高速情况下很容易失控;强调“人要比车凶”,我的理解就是这个车再厉害,你也必须能控制这个车,而不能让这个车控制你。

我觉得这个精神可以用在我们对待AI的态度上,我们应该始终掌握主动权,也就是我们要控制AI,不能把主动权让渡出去。

AI没有自我意识,它没有自己的连贯叙事,它没有一个笃定的自我,所以AI不会发起一个主动性议题。它不像我们人类,我们人是从从小在我们的文化、我们的环境、我们的教育乃至我们的基因影响下成长的,这决定了我们会有一些长期的愿望。比如“我想要出人头地”,“我想要挣钱养家”,“我想要振兴中华,等等,之所以有这些东西,都是因为我们是一个人类,我们从小有个连贯的故事,有个连续的记忆,有此前的这么多经历、熏陶、文化,这些使得我们有各种愿望。

AI在预测训练完成之前,它只是在不停地读各种各样的语料,读完之后它就上岗了。只要AI版本不变,今天的AI和3个月前的AI在本质上是完全一样的。当然它现在可以有一些记忆了,也就是读取你跟它对话的聊天记录,它能够生成一些记忆,但是这个记忆不是它的经历,AI没有自己的人生经历。

AI训练的语料非常全面,它什么语料都读,

读+:前面说到“责任”,感觉这是经营者、管理者的事,这和我们普通人、职场小人物有关吗?

万维钢:当然有关,AI给了普通人决策自由。现在假定我是一个送咖啡的服务生,我的唯一工作就是把咖啡给顾客端到桌上。那么我看到一位老顾客,我要不要跟人家打个招呼,怎么打招呼?我看到一个刚来的小朋友,我是不是要夸一夸小朋友很可爱?还是说现在比较忙,我应该减少互动,把效率提高一点?客人脸色不好,我是若无其事,还是更温和一些?所有这些东西都不能事先编程,都需要现场发挥。我每一个动作其实都是一种“微决策”,包括我把咖啡给人家端到桌子上哪个位置,是离客人很近还是稍微远点?这里边一举一动都体现了个人的风格,而往往这个风格是更有趣的部分。

以前我们可能以为,决策是老板和领导们的事情,只在项目的关键节点上发生——其实不然。如果你用了心,你会发现工作中的每一步都可以是一个决策。

其实生活中有无数个微决策的机会摆在我们面前,只是以前我们没有能力、没有能量、没有心思做它们。比如画画,以前只有专业画家才可

读+:那要是反过来,一个人不想在这件事上用心用力,就跟着AI走,就不比AI“凶”,那又会怎么样?

万维钢:那样的话,相对AI,他不具备稀缺的能力,不能提供一个独特的价值,等于是随波逐流了,他的存在感将会很低,他的境遇将值得担心。

读+:前面说到丧失主动性,我看看您书中关于家庭教育的部分,其实也涉及这一块。

万维钢:压制孩子的主动性,我觉得这个是很严重的问题。不止一个人告诉我,某个超级大市的中小学,两堂课之间的10分钟,除了上厕所

所以它不会有自我身份认同;理论上它可以假装自己是任何一个人,但是它实际上并不是任何一个人。

它也没有命运,所以它不会有一个想要改变命运的愿望;它也没有真正的喜好;所以它不会有像人这样主动想要干一件什么事的愿望。从这个角度来讲,主动性的发起者始终是人,而不是AI。

读+:如此说来,“人比AI凶”是基于责任而非能力,并不是说人的能力比AI厉害,而是人有主动性、有自我意识?

万维钢:你说得非常精确,的确是这样。基于能力的话,在智能这一块,AI肯定超过人了。国外新款的大模型,智商测验的成绩是148,这如果在人类里已经是天才级,爱因斯坦也就160左右。在医学领域,AI对医学知识的把握度,对疑难杂症的诊断,对X光片的读取,做手术动作之精准,所有这些方面都已经超过了人类。在数学领域,AI拿了国际数学奥林匹克金牌,已经可以帮助数学家去推演一部分新的数学问题了。利用AI搞科研,现在还没有看到绝对突破性的进展,但是我们在希望在一两年之内看到一个科学上的重大突破、重大发现,是由AI而不是人做出来。

所以,种种迹象表明,人类跟AI比能力是完全比不了的。但是基于责任来说,那就不一样了。把人比作公司老板,把AI比作聘请的总经理或者CEO,他在经营管理上面的才能远远超过你,但是重大决策是不是还得听你的?如

果这个决策失败了,赔钱赔的是你的钱,不是他的钱。

从这个意义上讲,谁承担损失,谁有最终的拍板权。AI只是一个抽象的程序,它有无数个副本运行在不同的服务器上,所以AI不可能受到伤害,它把事办砸了,你没有办法对它追责,所以不能把“签字权”交给AI,一定要交给一个人。也就是最终出了事我找谁,我只能找人,不管这个人是谁,AI公司的也好,是提供AI应用的公司的也好,还是公司具体使用AI的人也好,肯定得有一个责任方,这个责任方必须得能承受伤害才行,他只能是人。

读+:这些都是人类天赋,所以无须刻意修炼就能做到“人比AI凶”?在现阶段,其实人人都可以宣称“我比AI凶”?按照AI进化速度,哪怕是今后5到10年,其实人还是可以做到比AI凶?

万维钢:现阶段是这样,因为现阶段人人都是在给AI布置任务,AI是听命于人。在今后相当长一段时间,也还是这样。这里有一个关键性的判断,就是AI不了解“本地知识”,哪怕这个模型它非常聪明,极其聪明,但是它不了解你所面临的现场情况,它就没有办法替你做出决定。工作现场有很多知识是没有形成文字的,是所谓的“隐性知识”;这部分知识没有数据,AI没有办法知道这些东西。这些知识就确保了人不会那么容易被取代。

但是我们必须警惕,人可能丧失主动性,比如拱手交出,或者自己扼杀主动性。

我们工作的价值就体现在每天无数个微决策之中

以通过每一个构图细节、每一处光影、每一个笔画表达他的意图,而现在你可以直接让AI出图,然后你负责审美、选择和修改。到目前为止,大部分人的大部分工作时间都只是在努力把事情做“对”而已。而有了AI,你将负责决定什么是“对”。

我特别喜欢喜剧演员和制片人宋飞的一个故事。他拍电视剧《宋飞传》的时候,有一段时间出活儿特别慢,工作推进得很艰难。于是就有人建议他请咨询公司来帮帮忙,也许可以把制作工作给流程化。那大约就相当于我们今天用AI帮着创作。

而宋飞拒绝了,他当时慷慨陈词:“如果你高效,那你就是在错误的方式下工作。正确的方式是艰难的方式。这个节目之所以成功,是因为我进行了微观管理——每一个字、每句台词、每一条拍摄、每一次剪辑、每一个选角都由我把关。这就是我的生活方式。”

宋飞说的就是微决策。你的每一个决策,不管多么微小,都是对世界的改变。AI总可以建议。但只要你在意,你就必须干预。你必须确保每一个微决策都是你的决策。我们工作的价值就体现在每天无数个微决策之中。

任何主动性、每一处现场发挥、任何微小的

创造,都体现了你的个性和风格——也必须由你承担风险和光荣。

其实趋势已经出现了。纯粹由AI生成的图片,已经让人从最初的新奇逐渐变得审美疲劳;纯粹由AI撰写的文章,也正在从最初的惊艳走向公式化的乏味。人们指望“你”控制更多细节。我们听你的观点是因为你会用自己的声望背书。我们不想要一个计算上的展开,我们想要的是有风格有冒险的、充满微决策的作品。

按照规定和推演做正确的事,那不叫决策;真正的决策一定是某种任性和冒险。

在这个意义上,决策不但属于你的权力,而且是你的权利:你将通过这个决策展现你的个性、你的风格、你的价值观、你的冲动。你的每一个决策塑造了你,你通过每一个决策塑造世界。

而AI,它的决策则只不过展现了一种数学可能性而已。

这就是为什么我们不关心两个AI下棋比赛的输赢,那只是数学上无数可能性中的一个。但我们非常关心是阿根廷队还是法国队夺得世界杯冠军,因为那是把不确定变成了确定,那是此前所有人的故事切实的延续,也是此后一系列故事的开始。我们的世界线从此不同。

面对AI,人一定不能丧失主动性

可以去之外,竟然不允许孩子出去玩,搞得教育部门要发文件来保证这10分钟。一个主动性始终被压制的孩子,如果终于有一天考上大学了,离开家了,他从小学没给自己作过任何决定,他现在到了大学之后突然可以给自己作决定,你猜他能作出什么决定?他肯定是作最差的决定,因为他没有接受过作决定的训练,他没有接受过外界的反馈,他没有试错过。

现在孩子乖听你的话,你的确感觉很好——可是等孩子长大成人,跟其他人交往的时候,你也希望他像服从你一样服从那些人吗?同样的,面对越来越聪明强势的AI,你是希望他乖

乖服从,还是昂首挺胸“凶”起来?

正确的育儿方法,叫作“自主性支持”。这种方式充分理解孩子的需求和兴趣,尊重孩子的独立性和选择权,支持孩子自己作一些决策,包括家里的一些事情也鼓励孩子参与,不搞胁迫,有事会对孩子解释。

研究结果显示,自主性支持跟孩子当下的心理健康、学业表现和长大以后的幸福感都正相关。自主性支持家庭中成长的高中生都有更高的生活满意度和学习主动性;父母给予自主性支持的青少年能更好地适应学校生活,有更高的自尊心。