

# 人与 AI 组队 西游取经修行

▣ 杨斌

科技或产经新闻现在有了个新习（máo）惯（bìng），就是爱给新一年命名成“某某元年”，因为是预见，展望就常有分歧。英文里没有恰好对应元年那种带劲儿的表达，有的多是在过完一年或好多年后追记说，那真是“year of XYZ”，回望容易看得更清楚些。一本书，起名叫做《A Year of No Significance》，也是想以此引起你的好奇——无关紧要，抑或至关重要，历史会有大判断。

但，人们还是希望跨年的时候听吉祥话——2026 is a promising year！说到promising这个词，现在用来形容“充满希望”，而英文里最初只是个“承诺”的动名词形式。也对，不做出承诺并践行承诺，哪会有充满希望的未来。如此看来，跨年时宣告“某某元年”就是一种承诺，承诺这种科技或是产经领域来年会有充满希望的爆发。

去年这时候，人形机器人，co-pilot、AI智能体乃至自动驾驶都被媒体以“元年”来期冀。其时，AI虽经历了两年多一点的超高速演进，更多还是基础大模型在参数和表现上竞相比拼，彼时DeepSeek还没有横空出世，有关智能体落地以及AI组织演化、AI原生组织的讨论都还很少。

现在回望2025这一年，被看好“充满希望”的领域放异彩的程度有大有小，但大模型、智能体确实没有辜负“元年”承诺，比学赶帮超，直冲着要给客户拿结果而不是刷榜单的目标大踏步，要“真出活”而不是“会说话”，大企业多是个采用、应用的态度，而总是新创的AI原生企业更有冲劲儿，有不少鲜活实践。在这些实际经验基础上，再针对

杨斌

清华大学经济管理学院教授、清华经管领导力研究中心主任。开发并主讲清华大学《领导与团队》等精品课程。著有《企业猝死》《战略节奏》（合著）、《在明明德》（合著），译有《要领》《教导》等。



性地去讨论下一步组织中人与AI的关系，热度就在2026年前这一段时间陡增，深度也层层递进。

是啊，人与AI，会发展成怎么个关系呢？这不是个容易回答的问题。如果你能抽身片刻，靠在椅背上闭眼想上一想——在跟AI合作中，你把ta当作什么？程序？工具？员工？助手？专家？还是队友？

这个问题可以有很多指向，答案各不同，也难分对错——因为不同人不同公司其实真的就是有完全不同的“当作”。摆这个问题出来，是期待一个更认真些的内省，因为你对AI在你们之间是个什么角色的认知和定位，会决定着你现在和今后对ta的使用方式，影响着AI的靠谱底线和贡献上限，甚至还塑造着未来碳一硅基社会的样貌、走向。

不敢说ta们有自知之明，但是大模型（们）确实都扪“脑”自问——你们自己是怎么理解自己跟人之间的角色定位的呢？几乎所有主流大模型回答中，都有一个共同部分：AI更愿意把自己说成是一个协作者，而不是一个伺机替代人类的存在。

这个定位其实非常关键——至少对大模型的存在很关键，说错了可能命就没了——当然这是个笑话了。大模型并不是已经有了七十多年工作履历的传统意义上的“程序”——当然它有从程序开始的基于计算的作为工具的优势，信息传递速度、运算速度都是人类无法匹敌的。但它只是亦步亦趋、被动听命的驯服工具吗？



不妨以西游取经团队为类比，围绕 AI 时代中人与 AI 的关系、角色定位做一个思想实验。之所以选西游是因为对中国人来说，故事人人耳熟能详，算是文化上的通识基础，歧义误解少，不用什么解释就可以发问——你觉得 AI 与人，在一个 AI 和智能体进入到组织中迅猛发展的 2026 年，分别会扮演师徒取经团队中的哪一个角色？当然也可以延伸到 2026 年以后——那必将会是更加不好预判的未来。

由易到难。白龙马是一个长期被公认为“道具”的形象。在很长时间内，当我们听到工具这个词的时候，当用“机”来作为人工智能缩写的时候，如果联想到白龙马的话，就会发现它甚至连游戏里的 NPC 都算不上，就是完完全全的道具马——即便讲了很多出身家世也没用，在关键故事里都不起作用，缺少主动参与感，很难算得上角色。这也是为什么我总是说“人机协同”这个词组不适合用来指代人与 AI 的关系，“人机”里的“机”，满满的一种道具感，哪里罩得住 AI。如果把 AI 理解为机器智能的话，那么，简称为“智”，也要比“机”，更靠近 AI 而不是旧式机器一些——不如改叫“人智

协同”。

沙僧则是非常典型的工具人，NPC。西游记的作者给他取“沙”这个姓，而他也就真的如一粒沙般平常普通；沙僧，真算得上西游取经团队中存在感最弱的成员，任劳任怨挑担牵马，几句台词循环播放，“师父被妖怪抓走了”，“师父（师兄）说得对”。作为取经路上的 NPC，沙僧无所欲，维护成本低，符合人类头脑中忠心耿耿服务主人的机器人的模样。巧了，沙还就是硅基物质。但不得不承认，AI 的发展，会让我们觉得沙僧的本领不太配得上高能 AI——以现在的说法，AI 动辄就是在某种单项学科智能上超过 99.9% 以上人类，沙僧在聪明能干这一点上似乎差距还有点儿大。

要是考虑聪明与能干，AI 智能体，会不会更像是大师兄？超级活跃，能动性极强；超级能干，老孙的本事绝对超过了绝大多数最能干的人类，以及大多数的神。它在整个团队中的目标性很明确：发现和解决问题，保护师父平安，取到真经；当然你可以说大师兄体现的也是服务于取经任务的工具理性，但是我们分明在孙行者身上看到时常自然流露



的价值理性，看到了超越工具的自主、决断和创造，解决问题时的主动、积极和涌现。好了，坐实了，不用怀疑了，AI就是大师兄，大师兄就是AI！奇妙的还有，厉害的可不只是大师兄一个而已，需要的时候，大师兄还会拔下许多根汗毛变出来成百上千个AI“分身”——合在一起，就是多智能体的组织智能了呢——协同一致，保护师父取回真经。

不过，就算是大师兄本领如此高强，去西天取经，护佑东土苍生这个目标却不是大师兄自己能承载的，那是唐三藏的使命。唐三藏，“大我”担当，是人类中富有使命感的领导者，是人在持续修行中的崇高化身。

至于八戒呢？大家伙估计都会觉得它可不能代表AI——水平先放在一旁，先说态度——AI哪会像八戒那么的懒惰，意志不坚呢？正相反，AI就是最理性最恒定最耐烦最不怕苦不怕累不偷懒不眠不

休经得住考验的杰出代言人。八戒的“小我”形象深入人心，进化到这一步的人还摆脱不了的俗世本性的东西，都在八戒这里淋漓尽致地反映出来。有趣的是，每位读者都从八戒身上看到部分的自己，虽然未必愿意承认；但恐怕人们是不喜欢自己的合作伙伴、助手，是一个像八戒这样的人的。自己可以偷是，别人这样却不可以，这就是人们对八戒这种“真”所持的态度。

翻译成成人话：组织（唐三藏）的使命，如何能够在人（唐三藏、八戒）与AI（能力超众的大师兄）协同努力下达成。人里头既有超凡脱俗的三藏，也有乐在俗世的八戒。如果看AI和智能体的发展阶段的话，真的是从白龙马（如果还算有过那一段的话）进步为沙僧，然后就在我们眼前以迅雷不及掩耳之势出落成本领卓绝的大师兄。

八戒是人，是人就会犯错，因为各种人的不完美、六根不清净而惹了祸，这账就会记在八戒的头上。可本领高强的大师兄（再加上他的主动性又强），处理复杂问题有时也会惹祸，且因为本领大可能惹的祸也更大。大师兄惹的祸，账要记在谁头上呢？这是个好问题。西游里头，大师兄是一心为了师父取经，一心要铲除妖孽，他的失误，可能就得看看：是否师父也有看管不严的问题？八戒也有协同不力的问题？更主要的，要看把一个大闹天官的猴子打造成大师兄的造物主如来等，是不是要持续负责对大师兄做些迭代矫正。说到底，把账算在孙行者的头上，没用，他毕竟不是人，得去找跟他关联的人，以及开发他出来的大神算账。

三藏（以及作为原创者的如来等）必须能够驾驭好孙悟空，但恐怕有了跟孙悟空并肩战斗的体验，谁也不会再想着说退货给花果山而去流沙河换一个沙僧回来。回不去了，海拔基准水位上去了，就再也下不来。沙僧们早就被淹到水平面以下了。

阿西莫夫三定律，虽然三个徒弟都得遵守，但更像是约束高能悟空的行为规范脚本吧。

再追问一句，八戒这么多毛病，也不是那么有本领，这样的八戒，在有了能干得多的 AI 之后会不会首先就被裁员，以纯洁队伍、降低冗员、减少抱怨、提高效率？

这就要对强 AI 社会的未来——会不会演变为每一个角色都能趋于完美——做些设定了。

2024 年底 2025 年初我正重读诺伯特·维纳的 *The human use of human beings*，跟很多人絮絮叨叨地说，也许中文书名应该叫《人有人本之用》，而不是说着挺顺口也还有点自嘲俏皮的《人有人的用处》。维纳在书中，关于人的自主性有多紧要，说过这么一段：

“我是相信人类社会远比蚂蚁社会有用得多的；要是把人判定并限制在永远重复执行同一职能的话，我担心，他甚至不是一只好蚂蚁，更不用说是个好人了。那些想把我们按照恒定不变的个体职能和恒定不变的个体局限性这一方式组织起来的人，就是宣判了人类只该拿出远低于一半的动力前进，他们把人的可能性差不多全部抛弃掉了，由于限制了我们可以适应未来偶然事件的种种方式，他们也就毁掉了我们在这个地球上可以相当长期地生存下去的机会。”

顺着维纳的这段话往下想，强 AI 时代的问题其实并不是人们都在绞尽脑汁想的“AI 会不会替代人”，而是我们作为人类，是否愿意把自己预设在某一种被替代的姿态里。

当组织习惯于把人定义为流程节点、岗位标签、可预测的执行单元时，它其实已经在为更高效、更稳定、更不抱怨的“非人类”腾位置了——不是因为机器更聪明，而是因为预设人主动放弃了变化的权利。

如果沿着这个角度往下想，那么先被裁掉的未必是八戒，更可能是那个已经把自己活成沙僧的人。

这也许是“AI 次方”质变的某种价值所在：它不是一次能力的跃迁，而是一次对“何为（wéi）人”“人为（wèi）何”这个根本问题的一种压力测试。它更显豁地呈现出，若是人的工作拆解成任务与本领的话，哪些并非人之所长，也不利于人之为人、人之发展；哪些应为立人之本，却已经疏远太久。

当 AI 智能体具备了越来越强的发现问题、拆解目标、跨任务协同的能力时，组织中最稀缺的，不再是“会不会做”，而是“为什么这样做”，“是否值得这样做”，“这牺牲了什么其他可能”。这些问题，恰恰不是超级劳模的大师兄擅长回答的——大师兄可以超级快超级强，但大师兄并不主导意义构建、价值辨析，大师兄也无法真正担责兜底背锅。

所以，极大概率上，唐三藏（“大我”）这个角色，非但不会因为 AI 变强而式微，反倒会愈发彰显自身的核心价值。之所以重要，不是因为他聪明，更能干，更有威严，而是因为他承担着一个不可外包的职责：为目标赋值，为行动定向，为冲突作出价值判断，也为整个摊子出的问题负责、背锅。

再有就是，假设，我们说假设，要是没有了唐三藏，取经团队也许可以以更快速度更低难度达到某个地点，尽管确实存在这么个推理可能性。但这个假设，是个伪命题，压根就没有意义。没有了人类（三藏），再智能，效率再高，意义又何在——就像八戒爱说的那句，“师父都让妖怪给吃了，这还取什么经啊”？！当然，也许硅基生命觉得这是碳基人类的自以为是。

好了，让我们换个角度再问一句：如果不限装束的话，是不是每个人都想要一个大师兄来作为

自己的团队成员、工作伙伴呢？有没有人愿意要八戒呢？

其实，八戒在整个团队生态中的价值，或许值得换一个视角重新审视。

他身上的贪、懒、怕、欲，并不只是缺点，恰恰构成了对“效率至上”的持续扰动。八戒会抱怨、会拖延、会算计眼前得失，但正是这些“不完美”，提醒着团队：取经不是选择一条纯粹最优路径的问题，它是一段必须被人类承受、体验、理解的慢旅程。八戒（也是小“我”）的这些个不完美，各种幺蛾子，也就像很多网站检验是否真人的测试，是比图灵测试厉害的活人感测试过关的保证。

所以问题也许就不该只是围绕大师兄：“AI 会进化到哪一步？都能替代人去做什么？”而应当反躬自省：“当团队中有了无所不能的大师兄，我们是否还允许并能让自己不被碾压成沙僧？我们还能不能包容并欣赏身边的八戒？”当然，也该问：“什么大使命，何种慈悲心，让我能为师父？”并能在关键时刻说出那句——“悟空，休得无礼”。

这需要一种新的协作伦理——既不是主仆（奴）关系，也不是囹圄平权，而是一种“差异与共”。人在其中不以算力高强、不眠不休取胜，而以直觉、判断、忐忑、修正、意义和担责为价值所在。

去年跨年时，我推敲过 co-pilot（副驾驶）这个词，是不是就永远要比 captain（机长）低，

只能是个从属角色。现在想，也许 captain 和 co-pilot 地位高低并不真的重要，重要的是：是否允许在飞行途中根据需要，同担责任、备份支撑，甚至可以保持某种模糊。成熟的“机长”，不以揽总包全为自豪，不以掌控一切得权威，而是知道何时让渡、何时介入、何时为系统的选择负责。

这些个关于人与 AI 在西天取经组队中角色的

碎碎念，答案在路上，可能是个持续协同进化的序列与组合。AI 会在不同阶段，像道具，像工具人，像工具神，像大师兄，像大师兄们。

答案也在心里。看人，是否还能既秉持三藏（师父）的一本正经、使命必达，又保留一点八戒（二师兄）的真实、不装与贵在“有疑”——可能是接下来几年真正需要修行的本领。

真经未必只在西天，人与 AI 要一起走的取经路还很长，是不是“充满希望”，人类对自己做什么“承诺”

其实很重要。可以确定的是：如果人类不愿意继续成为“可变的的存在”，那么再能再强的 AI（大师兄），也不过是在为一个走向闭锁、有限游戏、零和博弈的世界奔波打拼。正如维纳在《人有本之用》中所说：“我们是如此彻底地改造了我们的环境，以致我们现在必须改造自我，才能在这个新环境中生存下去。”

师徒西游取经路，本就是改造自我的修行。人与 AI 展开的这个彻底改造环境的新取经路，唯有人智与共，才能各得其所，修成正果。🙏

