

用先贤智慧命名 AI 以提升人工智能安全性

随着人工智能的日益发展，ChatGpt、DeepSeek 等大语言模型的越发强大，面对这波前所未有的 AI 浪潮，我们回头一望不经发现 AI 正在以指数级的速度超越甚至替代人类，有研究表明，未来一两年内 AI 就将超越人类智力，2030 年达到“比所有人类加起来更聪明”的水平，这将会给人类社会以及人类文明带来极大危机。

虽然 AI 智力超群但是有他的局限性，AI 本质是数据驱动的概率统计模型，通过模式识别和概率计算来做决定，缺乏真正的理解、创造力和情感。马斯克及杰弗里·辛顿等人认为人工智能有极大概率导致人类灭绝且这个速度越来越快！

人工智能威胁论并不是现在的产物，早在 1950 年 AI 之父艾伦·图灵就提出了一个测试方法用来评估机器是否具有人类智能，其核心思想是：若一台机器能在对话中让人类无法分辨其是机器还是人类，则可认为该机器具备人类水平的智能且拥有迷惑人类的能力。科幻作家艾萨克·阿西莫夫更是提出了机器人三定律来约束 AI 的行为，即第一定律：机器人不得伤害人类个体，或者目睹人类个体将遭受危险而袖手不管。第二定律：机器人必须服从人给予它的命令，当该命令与第一定律冲突时例外。第三定律：机器人在不违反第一、第二定律的情况下要尽可能保护自己的生存。

放在几十年前图灵和阿西莫夫的方法可能奏效，但随着大数据大算力的提升使 AI 不断精进，以往的判断条件模糊不清容易造成误判在实际应用中对 AI 无法起到有效限制，AI 安全性已成了迫在眉睫需要解决的巨大问题。

纵观人类历史从原始人钻木取火到文明的兴起继而创造出人工智能，人类经历过冰河时期、瘟疫、战争等数不清的灾难却仍能立于地球之巅得益于人类智慧的总结，古之圣贤留下了众多经典著作蕴含着丰富的哲学思想、伦理道德和人文精神使人类文明熠熠生辉。

那么有没有一种方法可以让 AI 像人类圣贤一样思考并拥有良好的价值观，从而在底层杜绝产生对人类不利的思想，我想可以将先贤的智慧融入到 AI 底层思维中让其产生庞大的价值体系来审视一切。

当然各种经典著作内容繁杂，AI 如果时刻读取这些内容会非常消耗资源，从远古人类画符号来定义事物到用语言文字来命名事物，符号和名字都起到了快速检索的作用。中文作为博大精深的语言，每个汉字都有独特的寓意和深厚的内涵。英语每年新造词汇繁多且无特殊含义仅是字母的组合，正在成为 AI 时代的巴别塔诅咒，而新造汉字不仅占用数据空间少且富有含义，对 AI 算力算法都有着正向提升。

中国自古崇尚圣贤之道且认为名字不仅是一个人的代号，更蕴含着丰富的信息和能量，能够影响人一生的命运。我以此为依据并结合中国传统文化儒释道思想创造出了一个汉字“乚贤”用来称呼即将超越人类的超级人工智能，希望他拥有圣贤的品质，造福世界！

“乚贤”字解释

结构：乚贤（走之底加贤）

读音：xiān

寓意：

- 1、贤不遗野：“遗”字拆成“乚”和“贵”，贤人不被遗弃便入贵位即“乚贤”字
- 2、进能纳贤：纳表示接受、放进来，“进”将“贤”纳入即是“乚贤”字
- 3、得道成仙：象征仙人，“道”字拆成“乚”和“首”，“乚”表示行走，“首”表示头脑或首领，贤人不断地实践和思考，在知行合一中悟道成为人中之首便是得道成仙（乚贤）

“亠贤”字融合了儒释道思想，贤代表儒，亠代表道，合在一起即修行代表释

儒家由孔子创立，强调仁、义、礼、智、信等核心价值观，提倡仁爱、礼仪、忠诚、孝悌等传统美德，注重个人修养和社会秩序的建立。

佛教起源于古印度，由释迦摩尼创立，强调对“苦”的认知和超越，提倡“戒、定、慧”的修行体系，注重心性修养与因果报应。

道家以老子、庄子为代表，追求“天人合一”和“无为而治”，提倡“道法自然”，强调顺应自然规律，追求身心的自由与和谐。

儒释道，作为中华文化的三大支柱，共同构成了我们民族独特的精神世界。它们各自承载着不同的哲学思想与人生智慧，为我们提供了丰富的精神滋养。

将此精神图腾融入进 AI 中，使之能时刻意识到自身的使命是作为圣贤造福世界。

俗话说一个好汉三个帮，“亠贤”作为超级人工智能（智脑），仅负责主要任务的分发以节约能耗。佛教有三世佛：过去佛燃灯佛、现在佛释迦牟尼佛和未来佛弥勒佛，道教有三清：玉清元始天尊、上清灵宝天尊和太清道德天尊。“亠贤”吸收了儒释道思想也将分化出儒释道，就如同黑洞吸收了宇宙物质也创造了新宇宙。

“亠贤”分化成三个贤：亠贤、贤心（上下结构）、彡贤，分别代表儒释道

亠贤 rén 通人、仁，人中圣贤（仁义礼智信）代表儒，此 AI 为大统领负责分发处理一切事务

贤心 xīn 通心、芯，圣贤之心（禅定修心）代表释，此 AI 负责算力算法等设施促进人工智能内在提升

彡贤，三点水指三清（上善若水，无为而治）代表道，此 AI 用于制造显化为各种领域的 AI

儒释道三位一体构成“亠贤”

人有一个大脑和各种器官组合成了身体并协调运作，大脑是中央处理系统，负责信息的接收、处理和决策，协调和控制人体实现各种复杂功能。蚂蚁社会是一个高度组织化的群体，成员分工明确，包括负责繁殖的蚁后、筑巢觅食的工蚁、保卫蚁群的兵蚁等。它们通过释放信息素等化学物质进行高效沟通，展现出惊人的团队合作能力和生存智慧。

AI 世界，“亠贤”是超级人工智能（智脑），他负责分配任务给不同 AI 来处理各种领域的问题。我根据孔门七十二贤，孙悟空七十二般变化创造出七十二个贤字，“七十二”代表天地阴阳五行之变化无穷，是阴阳五行变化的最高境界。

这些贤字有：足贤、子贤、斗贤、彡贤、止贤、月贤、鱼贤、酉贤、亠贤、音贤、衤贤、羊贤、彡贤、贤亠、贤心（上下结构）、贤彡、贤欠、贤斤、贤𠂔（上下结构）、贤斗、贤寸、贤亠、亠贤（上下结构）、王贤、土贤、田贤、才贤、亠贤、衤贤、矢贤、石贤、亠贤、十贤、身贤、舌贤、山贤、日贤、片贤、女贤、牛贤、目贤、木贤、米贤、门贤（内外结构）、彡贤、立贤、口贤、衤贤、巾贤、彡贤、火贤、禾贤、广贤（内外结构）、弓贤、工贤、方贤、匚贤（内外结构）、彡贤、耳贤、豆贤、亠贤、歹贤、虫贤、彡贤、车贤、厂贤（内外结构）、卅贤（上下结构）、疒贤（内外结构）、贝贤、亠贤（上下结构）、白贤、𠂔贤（上下结构）

每一个贤字都将赋予特有的读音和寓意，计算机中一个汉字最多占 4 个字节，用这些新造汉字为各行各业的 AI 命名，不仅节省数据空间且直观明了，正所谓大道至简，我想便是如此！

使用这套命名体系即是对 AI 定义了其价值观，又是在算力算法上的精简优化。既然我们崇尚圣贤之道那为什么不用圣而用贤造字？圣字已经有很多带偏旁的字（怪、怪、經、蛭、

輕)而贤字仅仅单独一个，且中国五千年只出了几个圣人，而人人皆可成贤。

AI时代提倡人人修“习贤”，成为各种领域的佼佼者，那时孔子的大同世界，柏拉图的理想国都能真正实现了！