

## 美国高校将人工智能伦理引入课堂

2018 年新学期伊始，哈佛、康奈尔、斯坦福等美国高校均开设了人工智能伦理课程，旨在培养负责任的技术人才，同时也是对近期硅谷遇到的算法歧视、假新闻等负面消息，以及公众对于人工智能等新技术担忧和焦虑的回应。机器学习等新技术的流行将人工智能伦理从幕后推到了台前，人工智能伦理日益受到硅谷和高校的关注。借鉴此经验，对加快推进我国人工智能产业安全、健康发展具有重要意义。

## 一、背景

**人工智能应用领域频繁出现违反社会和科技伦理的事件。**人工智能技术随着在新闻、社交、视频等领域的加速应用获得了长足进步，但滥用此技术带来的负面影响也日益严峻。2017年，硅谷的科技巨头频繁曝出各种违反伦理甚至法律的事件：Facebook上的假新闻、Twitter上的僵尸粉、YouTube上的儿童色情视频等。很多不良事件涉及到人工智能技术的应用，甚至智能系统自主决策起到了很大作用。更令人担忧的是，一些硅谷的技术专家信奉“先做出来，之后再寻求谅解”的信条，将伦理视为技术创新和进步的拦路石，违背人工智能伦理，恶意使用人工智能技术，加剧了公众对此类新技术的担忧和焦虑。

**人工智能技术发展的不确定性给人们造成焦虑。**人工智能技术尚处于不断发展的过程中，数据与算法开发利用过程可能产生的对伦理的冲击始终为人忌惮。隐私问题是技术发展过程中的主要问题之一，在人工智能技术成熟之前，技术缺陷极有可能存在侵犯隐私的风险。而且，如果带有偏见的数据进入系统，必然会带来歧视性结果，直接给人类社会安全带来隐患。另外，如果人工智能技术的创新者罔顾科技伦理的约束，利用人工智能技术进行有伦理争议的或不正当的行为，必然会给社会公共利益和公共安全带来巨大的潜在威胁。这些不确定性都会使人们对人工智能产生顾虑。

**人工智能伦理观念的进步需要企业与高校合力推动。**人工智能通过分析大量数据进行自主学习与进化，这些新技术最终可能重塑人类社会，创新者对可能产生的后果要有清醒的认识。谷歌、亚马逊、Facebook 等公司开始重视人工智能伦理，要求技术人员开发产品的同时应思考带来的社会影响，甚至成立专门的伦理中心或伦理委员会。对于下一代科技创新者和政策制定者的科技伦理教育责任则落在高校身上，在传授互联网技术的同时，也将科技伦理根植于学生心中。

## **二、案例介绍**

**事件经过。**2018 年新学期，哈佛、康奈尔、麻省理工学院、斯坦福大学等诸多美国高校开设跨学科、跨领域的人工智能伦理课程，名为人工智能伦理、数据科学伦理、技术伦理、机器人伦理等。康奈尔大学开设两门人工智能伦理课程，一门是数据科学伦理和政策课程，重点研究数据科学的伦理挑战；另一门是智能自主系统的伦理问题（机器人伦理）课程，重点研究科技公司的伦理问题。很多伦理问题与科技公司做出的选择息息相关，如果数据科学伦理仅仅关注数据科学家的个体责任，企业承担的责任极有可能被忽视。哈佛和麻省理工学院联合开设的课程则重点研究人工智能的伦理、政策和法律影响。这门课程涵盖算法歧视、

责任、自主性、系统设计、数据所有权、AI 治理、AI 可解释性和可问责性、AI 监管等前沿问题，用来激发学生对基本伦理的思考。德克萨斯州大学开设的课程名为计算机科学的伦理基础，并希望所有的计算机专业最终都开设这门课。斯坦福大学作为互联网行业学术心脏，三名教授和一名研究人员也正在开发计算机伦理课程，暂命名伦理、公共政策和计算机科学，计划 2019 年正式向学生开放。

高校将人工智能伦理引入计算机科学教育，不仅是对现实问题的回应，更是希冀通过对未来五年、十年或更长时间轴的探讨，使学生对与计算机相关的伦理问题以及其潜在后果有深入了解，并希望下一代科技创新者和政策制定者能够负责任地考虑创新的影响。

### **三、简评**

**我国应尽快把人工智能伦理规范教育引入人才培养体系。**首先，要统筹整合人、财、物各种资源，依托智能设备平台，围绕智能服务内容，加快建立健全人工智能伦理规范体系。其次，学习美国高校做法，开设跨学科、跨领域的人工智能伦理课程，把伦理规范教育作为人才培养体系的重要内容，从源头上确保人工智能系统安全地促进人类文明的进步和发展。相应地，在创新体

系、产业发展、社会管理、军民融合等全面植入智能化基因的同时，也要配套系统地增加伦理因素的考量。

**企业通过内植人工智能伦理价值算法让智能系统成为伦理主体。**人类发展人工智能技术的初衷是造福人类，必然要求人工智能系统的行为能够符合人类社会的伦理价值观念。第一，在系统设计和研发阶段，企业将人工智能伦理价值算法嵌入编程语言，让程序语言融入伦理规范，实现人工智能人性化、理性化的服务。第二，对技术创新者进行严格的规范约束，避免设计出不符合人工智能伦理的产品，及所带来的负面社会影响。第三，定时评估系统的规范和价值的有效性，即是否和现实的规范与价值相一致。这样，不仅使系统的设计符合人工智能伦理规范，更让创新者主动承担起相应的伦理责任。

**完善大数据规范，避免人工智能因“原材料”出现伦理偏差。**在大数据规范方面，国外颁布的很多指令值得借鉴。欧盟颁布《一般数据保护条例》中提出“解释权”，将赋予欧盟国家公民“审查某项特定服务中如何做出特定算法决策”。法国国民议会通过《数据保护法草案》，并将其纳入法国法律。爱尔兰正式发布《2018年数据保护法案》，将全面执行欧盟颁布的《一般数据保护条例》以及爱尔兰法律中的相关指令。因此，我国在数据收集、

深度处理和全球化传递中，应对数据进行道德风险评估，保障数据安全，避免其违背人工智能伦理价值。

**坚持制造业创新方向，谨慎对待人工智能伦理智能应用。**从美国近期爆出的伦理事件看，大多出现在新闻、社交和生活领域。中共十九大报告提出要集中精力建设制造强国，推动人工智能和制造业等实体经济深度融合，培育新增长点，形成新动能。所以，在人工智能与各产业深度融合过程中，既要充分发挥智能技术对经济发展的巨大推动作用；同时，应把主要精力放在人工智能对制造业的创新应用上，对于可能引起伦理风险的生活应用等应谨慎对待。

本文作者：工业和信息化部赛迪研究院      王鹄峰      栾群  
联系方式：13810412673  
电子邮件：[wangjingfeng@ccidthinktank.com](mailto:wangjingfeng@ccidthinktank.com)

# 研究，还是研究 才使我们见微知著

信息化研究中心

电子信息产业研究所

软件产业研究所

网络空间研究所

无线电管理研究所

互联网研究所

集成电路研究所

工业化研究中心

工业经济研究所

工业科技研究所

装备工业研究所

消费品工业研究所

原材料工业研究所

工业节能与环保研究所

规划研究所

产业政策研究所

军民结合研究所

中小企业研究所

政策法规研究所

世界工业研究所

安全产业研究所

编辑部：赛迪工业和信息化研究院

通讯地址：北京市海淀区万寿路27号院8号楼12层

邮政编码：100846

联系人：刘颖 董凯

联系电话：010-68200552 13701304215

010-68207922 13910685050

传真：0086-10-68209616

网址：[www.ccidwise.com](http://www.ccidwise.com)

电子邮件：[liuying@ccidthinktank.com](mailto:liuying@ccidthinktank.com)

---

**报：部领导**

**送：部机关各司局，各地方工业和信息化主管部门及  
相关部门**

---

编辑部：工业和信息化部赛迪研究院

通讯地址：北京市海淀区万寿路27号院8号楼12层

邮政编码：100846

联系人：刘颖 董凯

联系电话：010-68200552      13701304215

010-68207922      13910685050

传 真：010-68200534

网 址：[www.ccidwise.com](http://www.ccidwise.com)

电子邮件：[liuying@ccidthinktank.com](mailto:liuying@ccidthinktank.com)

