



网上办公 (<http://oa.ia.ac.cn>) | 联系我们 (<http://www.ia.cas.cn/gkjj/lxwm/>) | English (<http://english.ia.cas.cn/>) |

中国科学院 (<http://www.cas.cn/>) |

新闻中心

[科研动态 \(../\)](#)

[近日要闻 \(../jryw/\)](#)

[媒体扫描 \(../mtsm/\)](#)

[头条新闻 \(../ttxw/\)](#)

[学术活动 \(../xshd/\)](#)

[成果转化 \(../cgzh/\)](#)

[信息公开 \(../xxgk/\)](#)

当前位置: [首页 \(../..\)](#) > [新闻中心 \(../\)](#) > [科研动态 \(../\)](#)

自动化所参与撰写联合国教科文组织《人工智能伦理建议书》

发表日期: 2021-11-30 【大 中 小】 【打印】 【关闭】

2021年11月25日,联合国教科文组织在法国巴黎发布了《人工智能伦理建议书》,这是全球首个针对人工智能伦理制定的规范框架,是迄今为止全世界在政府层面达成的最广泛的共识,也是全球人工智能发展的共同纲领,将进一步形成人工智能有关的国际标准、国际法等提供强有力的参考。

《人工智能伦理建议书》于2018年春季立项,由全球遴选的24人专家团撰写,并经历了193个成员国之间超过100小时的多边谈判修订完成,于今年11月24日在联合国教科文组织第41届大会上获得通过。联合国教科文组织总干事阿祖莱在当天举行的发布会上表示,这是多边主义取得的胜利。

由中国联合国教科文组织全国委员会协同国内多部委推荐,并经过联合国教科文组织遴选,中国科学院自动化研究所中英人工智能伦理与治理研究中心主任、国家新一代人工智能治理专委会委员曾毅研究员入选特设专家组,深度参与此项工作。曾毅研究员在专家组中担任3个工作组中伦理工作组联合组长,并参与了全部撰写过程。随后,联合国教科文组织开展了区域性专家咨询会(曾毅研究员代表联合国教科文组织特设专家组专家向亚太地区各国政府专家代表宣讲人工智能伦理建议书草案)、政府间专家磋商(曾毅研究员作为中方专家代表参加)、各成员国意见征集等环节,并在此基础上对文本进行不断完善。


UNESCO

Artificial intelligence | In action | Stories & ideas | Resources

We need international and national policies and regulatory frameworks to ensure that these emerging technologies benefit humanity as a whole.

We need a human-centred AI. AI must be for the greater interest of the people, not the other way around.

On 24 November 2021, UNESCO has adopted a comprehensive global standard-setting instrument to provide AI with a strong ethical basis. It will not only protect but also promote human rights and human dignity, and will be an ethical guiding compass and a global normative bedrock allowing to build strong respect for the rule of law in the digital world.



[Read the Recommendation](#) [Evolution of the Recommendation](#) [Ad Hoc Expert Group](#)

联合国教科文组织《人工智能伦理建议书》这一历史性文本确定了共同的价值观、伦理原则及相关的政策建议、监测与评估，用以指导建设必需的法律框架来确保人工智能的健康发展。其中价值观主要涉及4项内容：（1）尊重、保护和促进人权和基本自由以及人的尊严；（2）环境和生态系统蓬勃发展；（3）确保多样性和包容性；（4）在和平、公正与互联的社会中共生。伦理原则主要涉及：（1）相称性和不损害；（2）安全和安保；（3）公平和非歧视；（4）可持续性；（5）隐私权和数据保护；（6）人类的监督和决定；（7）透明度和可解释性；（8）责任和问责；（9）认识和素养；（10）多利益攸关方与适应性治理和协作。在政策建议部分，主要内容包括：（1）伦理影响评估；（2）伦理治理和管理；（3）数据政策；（4）发展与国际合作；（5）环境和生态系统；（6）性别；（7）文化；（8）教育和研究；（9）传播和信息；（10）经济和劳动；（11）健康和社会福祉；最后对监测与评估进行了建议和说明。

在联合国教科文组织特设专家组撰写草案期间，曾毅研究员提出“和谐共生”的价值观，倡导人类社会与人工智能技术和谐发展，人工智能赋能人类、社会、生态之间的和谐发展，号召各个国家通过人工智能促进人类命运共同体的构建。这一主张体现着由国家新一代人工智能治理委员会制定的《新一代

人工智能治理原则》以及《新一代人工智能伦理规范》中以“和谐友好”等形式和内容为代表的核心思想，是中方发展人工智能技术，处理技术与社会关系的总纲。韩国、南非等国的专家还提出，“和谐共生”的思想根植于不同的文化中，有意义并应得到更广泛的认同。在征求各国政府意见以及政府间专家磋商阶段，这一“和谐共生”的核心要义得以部分保留和采纳。例如建议书中强调：“人工智能系统在整个生命周期内都有可能为所有生物之间及其与自然环境之间的相互关联作出贡献。”以及“即每个人都属于一个更大的整体，当这个整体中的组成部分都能够繁荣兴旺时，整体才会蒸蒸日上”等。相关思想被各国专家认为是人工智能伦理建议书中融入非西方价值观的重要体现。

我国《新一代人工智能治理原则》以及《新一代人工智能伦理规范》将“可持续发展”作为新一代人工智能的总体愿景和赋能全球人工智能发展的重要途径。在《人工智能伦理建议书》撰写阶段，曾毅研究员提出，虽然可持续的理念隐含在文本中的若干方面，但是作为全球共同愿景，发展人工智能必须赋能可持续发展目标的实现，并且要密切关注人工智能应用于可持续发展相关目标领域时是否带来阻碍该目标实现的风险，寻求发展与治理的协同。该意见被建议书不同阶段广泛接受。在政策部分，专家们还特别针对环境和生态系统领域如何通过人工智能赋能实现可持续性给出了具体建议。

在政府间专家磋商阶段，部分国家特别关注中国是否会同意文本中充分强调人权的做法，甚至直接提出中国是否会反对这种提议。曾毅研究员特别指出尊重人权是写在《中华人民共和国宪法》中的，特别是“国家尊重和保障人权”，此外《新一代人工智能治理原则》和《新一代人工智能伦理规范》都对尊重人权做出了明确规定。这化解了部分国家对发展人工智能及其伦理的误解。

人工智能作为推进社会发展的颠覆性技术和赋能技术正深刻影响着社会和全球变革，与此同时也为社会发展带来了巨大的挑战与风险。面向未来，更多严肃的公共讨论与全球性的参与将共同推动人工智能造福社会。曾毅研究员说：“目前联合国教科文组织已经开始着手筹备帮助人工智能不同的相关方以及政府针对人工智能伦理建议书中提出的价值观、原则、政策进行落地，而这项努力则更需要全球各个国家通力协作、互鉴共赢”。



ARP (<https://ia.arp.cn/>) | 科技网邮箱 (<https://mail.cstnet.cn/>) | 联系我们

(<http://www.ia.cas.cn/gkjj/lxwm/>) | 违纪违法举报

(http://www.ia.cas.cn/qtgn/kslj_1/201609/t20160920_4665869.html)

此网站支持IE9及以上浏览器访问

1996 - 2021 中国科学院 版权所有

备案序号：京ICP备14019135号-3 (<https://beian.miit.gov.cn>) 京公网安备110108003079号

地址：北京市海淀区中关村东路95号 邮编：100190 Email: casia@ia.ac.cn (<mailto:casia@ia.ac.cn>)



(<https://bszs.conac.cn/sitename?>

method=show&id=08D8E9015DA3450AE053022819AC2F0E)