

人工智能伦理教育： AI是“问题”也是“解方”？

黄龙翔
吕赐杰

由ChatGPT陡然在普罗大众之间掀起的“AI（人工智能）旋风”和“AI焦虑”，至今已一年余。有关AI伦理的讨论也从未停止过。在主政者和业界还在纠结于AI开发应用的伦理规范和法规如何制订，能否有效执行之际，教育界也忙于跟学生斗智，防止后者滥用AI，并逐步推出伦理课程，为学生做好迎向AI世界的准备。

只是，人说教育乃百年树人，若把AI伦理孤立起来进行“教学”或培养，是否有“头痛医头”之嫌？或许，该把AI伦理纳入全人教育的语境，探讨AI伦理与各种既有的课程，及21世纪核心素养等的教学目标之间的关联性。

先看AI伦理本身，学生该培养的价值观和技能是什么？对学生来说，AI伦理是双向的，一方面得学习及防范AI对自己的福祉、隐私等造成威胁，另一方面又得建立正确的应用AI的伦理观念和行为——AI伦理教育该双管齐下。

厘清基本范畴，便可看出AI伦理是位于AI素养和品格与公民教育的交叠之处。品格与公民教育着重于培养学生正确的普世价值和行为——诚信、正真、尊重、公平、包容、体恤、社会责任，以及对于相关议题、事件的反思、批判能力；而AI伦理，正是要把这些价值观迁移到使用、应对，以至于开发AI的情况下。AI素养则还涵盖对AI的功能、局限、对人类可能造成的正负面影响等的理解，及AI技术及工具的操作和实际运用，这些都是“防范”及“善用而不滥用”的基础知识。原来，学AI（素养及伦理），（也该）就是学做人——不单是学做事。

其实，AI或能成为“载舟”之水，为防止“覆舟”尽一份力。可善用如ChatGPT来辅助品格与公民教育及AI素养教育，与学生针对伦理、公民议题本身，或相关的特定时事、生活语境进行讨论、辩论，或角色扮演，引导学生进行探究和反思，协助他们厘清思路、澄清疑惑。例如，把一篇新闻稿输入ChatGPT，让它消化内容后，跟学生开讲这个话题，如同读报教学。ChatGPT甚至能有限度地执行苏格拉底式对话，如不单只是被动地接受及回复学生的提问，还能反问学生、挑战学生的观点。学生不仅学做事、学做人，还学思考。

例如，ChatGPT可采用以下的生活语境作为比喻，引导学生思考AI伦理：“想象你组成小组参加校际智能机器人设计比赛。在比赛中，一个小组的机器人表现超乎预期，但后来发现他们在程序中作弊。你会怎么做？你认为这是否公平？”“想象一下，在未来的高科技课堂里，学生配戴智能眼镜上课。有一天，你发现你的智能眼镜被其他同学远程入侵，对方看到了你的笔记和个人隐私。你怎么想？”ChatGPT先把学生置于受害者的角度来思考，再联结“己所不欲，勿施于人”，讨论学生本人的为与不为。

作为优良的华族传统价值观，“己所不欲，勿施于人”却在这个资讯时代日渐淡泊？我们联想起前年本地事关“华文课程30年未变，除尊老爱幼、保护环境、种族和谐等亚洲价值观，基本上没其他内容……枯燥乏味”的论争。后来有人分析了小学六年课程的所有主课文，涉及亚洲或普世价值观的仅占五分之一。我们不认为语文课还要刻意及过度承担起品格与公民教育的责任。但这19篇主课文及围

绕着它的各种学习活动的设计，始终紧扣“让学生能在真实生活中以华文沟通”的目标；用以说教的语境，其实也是驱动学习特定语言知识和技能媒介。ChatGPT或能给这些跨学科的学习目标提供新契机。

AI代语文课传授伦理价值

如小三课文《留张便条》，几个小孩在楼下踢球，打破邻家花盆，似要“肇事逃逸”。原本没有科技介入的后续活动，可以是学生在课堂中模拟情境，学用连接词“反正（没人知道）……”“要不（留张便条）……”讨论解决方案。但也可以尝试提示ChatGPT与学生角色扮演（ChatGPT可扮演屋主、学生家长，或另一名参与踢球的孩子）、讨论或辩论。我们把这个语境输入ChatGPT，请ChatGPT模拟出一段它自己与学生对话的情景：

ChatGPT：小明，听说你和几个朋友在踢球时，不小心打破了邻居的花盆。你说怎么办？在回答时，记得用上“反正”和“要不”。

小明：反正，邻居还没发现，要不我们溜走。

ChatGPT：好，这可能是一个办法，但我们还可以找更负责任的方法。如果你们溜走，那你们可能会失去邻居的信任，也没法从错误中得到教训。还有其他的解决办法吗？

小明：要不我们先留张便条给邻居，跟他说发生什么事和说对不起。

ChatGPT：很好，这样子更负责任。有没有其他的想法呢？

小明：要不我回家跟爸爸妈妈讲。可是我怕爸爸妈妈骂我。

ChatGPT：小明，听起来你有点担心。但是向爸爸妈妈求助是对的，爸爸妈妈最关心的是你的安全和健康。他们可能会生气，但他们也会明白你是愿意认错的。你可以坦诚地向他们说对不起，并向他们展示你的诚意，让他们知道你想做对的事。

小明：好吧！我试试看。

ChatGPT：很好！你提出了几个不同的办法，每个都有它的好处和坏处。我们可以一起讨论一下哪个办法更好。

这个模拟“剧本”或许不符合本地一般小三生的华文水平，但它示范了ChatGPT的可能应用方法，以课文中的贴近生活经验的语境为引子，把语言学习、品格教育和思维训练共治一炉。

目前公开让公众使用的主要聊天机器人，本来就有伦理和内容上的把关机制，会拒答不合适的话题。在一般应用下，它们不太可能对用户造成伤害，尤其它们会严守普世价值的底线；就算是较有争议的价值观，也会主动提醒说各地文化有异，并对各种观念进行思辨。教师或专家可通过设计新提示，“训练”聊天机器人，为它们规划跟学生互动的范畴和规则，而不是天南地北地聊到出状况。

科技可能制造问题，但也可以提供解方。如何在这个风云万变，尤其是被AI“包围”的人世间安身立命，培养并坚守正面的价值观，不单是AI伦理教育的责任，或许更应该对整体教育内容，进行全方位的反思和调适。

作者黄龙翔是南洋理工大学
国立教育学院高级研究科学家
吕赐杰是南洋理工大学终生荣誉教授
香港教育大学讲座教授