

高等教育中人工智能教育的若干挑战

让-加布里埃尔·加纳夏

(索邦大学, 法国)

张婧妍 赵琳 译

(清华大学 教育研究院, 北京 100084)

摘 要:在高等教育领域,人工智能的重要性已日益凸显。这不仅对于志在成为工程师及人工智能专家的学生而言,所有希望发展面向 21 世纪信息社会素养的人都要学习人工智能基础知识和如何使用多样化的人工智能工具。只有熟悉掌握人工智能工具,才能适应未来社会发展。人工智能素养教育,既要求学生了解现有的人工智能技术及其应用,还应涵盖人工智能作为一门学科的历史脉络与发展趋势。鉴于此,本文阐述了如何在人工智能教育中,超越技术层面的局限,而旨在为学生应对未来信息社会做好准备。

关键词:人工智能;生成式人工智能;交织的时间结构;伦理与人文问题;人工智能素养;新实践

中图分类号:G434 **文献标识码:**A **文章编号:**1001-4519(2025)01-0010-07

DOI:10.14138/j.1001-4519.2025.01.001007

一、引言

人工智能与数字技术正深刻改变着人类的各项活动,无论是在专业领域还是个人日常生活。人们的工作方式也在发生变化。当前几乎所有人类活动都更加依赖认知技能,而不再是单纯的体力劳动。大量任务正在实现自动化,其余的亦在经历转型,由此催生了新的工作能力需求。新兴职业不断涌现,传统职业也在发生变化,部分职业甚至面临逐渐消亡的命运。人与人之间的关系也在持续演变之中。人们不得不面对新的挑战,如虚假新闻、黑客攻击等威胁。在当今世界,轻信他人所带来的风险日益增加,因此学会对信息进行核实,并将之与其他信息源进行对比以避免被误导,变得日益重要。

受上述变化影响,当前每个人都应具备人工智能和数字技术的基础知识,包括学习人工智能的基本原理并接受相应的培训。这不仅是对那些有意成为工程师或人工智能专家者的必然要求,也是对所有希望积极参与并成为 21 世纪信息社会合格公民者的基本要求。教育工作者肩负着向所有学生传授人工智能知识的责任,不仅要传授相关技术技能,更要传授如何理解人工智能和数字技术在专业、公共及私人领域应用中所引发的人文和伦理挑战。

收稿日期:2024-12-12

作者简介:让-加布里埃尔·加纳夏,索邦大学计算机科学名誉教授,巴黎六大计算机科学实验室成员,研究方向为人工智能、数字伦理学和数字人文。

译者简介:张婧妍,《清华大学教育研究》编辑;赵琳,清华大学教育研究院副院长、《清华大学教育研究》副主编,副编审,研究方向为高等教育学、研究生教育。

作为人工智能领域的学者,我们的目标是要探索出如何教导学生,尤其是要明确在教育教学过程中所要传授的核心概念。其中的挑战在于,人工智能和数字技术始终保持着创新性。由于技术的未来发展趋势难以预测,因此在某种程度上这些领域总是处于前沿。本文聚焦探讨如何让学生对人工智能和数字技术的未来发展做好准备,第一部分分析了教师教授人工智能与数字技术的目的及教学内容;第二部分通过一项回顾性研究简要展示,尽管在长期趋势上存在诸多共识性预判,但是人工智能与数字技术的短期发展却难以准确预测;第三部分探讨由部分科学家和企业家(如埃隆·马斯克)发表的“存在性风险”(existential risk)言论(即人工智能技术或将毁灭人类)所引发的焦虑,指出这些言论的不足之处以及教导学生如何对抗此消极看法的重要性;文章最后将提出一项面向所有学生的有关人工智能教学的初步、入门方案。

二、为何以及如何教授人工智能?

在思考教学问题时,教育工作者首先应当明确的是,传授知识背后的根本目的是什么,为何要从事教学活动。目的可以仅限于让学生掌握技术技能,让学生能够复现专业操作,但也可以旨在引导学生通过自身创造力为社会发展作出贡献。过去,大众教育的一个功能在于传授日常活动所需的基本操作技能,以及传授基本的认知能力,比如写和算的能力。当前,随着我们步入信息社会(也是知识社会),教育的作用比以往任何时期都更为重要。更准确地说,在知识社会中,工作不再仅是重复性劳动。现代工作者必须具备创造力,如若因循守旧,便会被机器轻易取代。因此,教育的目标应当是培养学生的创造力、想象力以及创新能力,使之在各项活动中能够展现出相应素质。

当然,培养出来的学生也要具有充分的技术技能,以确保生产力和生产效能。这一点对于将成为数字技术工程师的学生而言,无疑是正确的。然而,当前的人工智能和数字技术已超出了传统技术的界限,它们不仅在整体上重塑社会,还以嵌入组织的方式改变各个领域。换言之,人工智能和数字技术不仅具有技术性,更具有“社会技术性”(sociotechnical)^①。由此,未来的工程师既需要认识到人的因素在技术发展过程中的重要性,也要接受软技能培训,包括伦理、社会和法律方面的知识和问题处理能力。

然而,仅此尚不足够。实际上,鉴于人工智能和数字技术将在未来扮演关键角色,无论是“硬科学”如物理学、化学、生物学等学科的学生,还是人文与社会科学如法学、心理学、传播学等学科的学生,如果要积极参与未来社会发展而不仅仅作为技术的消费者,或仅作为制造者与执行者,他们就必须接受人工智能和数字技术方面的教育。换言之,为了能够自主使用未来技术作出真正的创新,成为未来世界中的主动者,学生就有必要了解人工智能和数字技术的基本知识,包括围绕这些技术领域的所有软技能。

综上,作为教育工作者,如果我们期望学生成为未来世界的主动者,即在未来世界中扮演主导者而非仅是执行者的角色,那么显而易见,他们必须接受人工智能和数字技术方面的教育。其次,教学内容不能仅限于技术层面的知识,学生还要接受人文、社会、法律和伦理方面的培养,以确保其能在未来生活的数字社会中负责任地行事。接下来将更为详细地探讨人工智能作为一门科学和技术的历史演进过程,以便更好地理解未来可能的发展趋势。

三、人工智能与数字技术的时间结构

在高等教育领域,我们面临的重大挑战之一是确定教学内容,即向学生传授哪些知识。一方面,有学者认为,无论内容为何,学习过程本身就具有形成性(formative),因为它能够训练学生掌握新技能,而且训练过程可重复。另一方面,由于要掌握的知识总量极为庞大,需避免无效努力,以免造成学生的认知负担,这一点至关重要。此外,显而易见的是,知识是不断发展的,我们不能使学生陷于刚开始职业生涯即

^①Kenyon B. De Greene, *Sociotechnical Systems: Factors in Analysis, Design, and Management* (Englewood Cliffs: Prentice-Hall, 1973).

面临所学知识已过时的困境。因此,预测未来可能应用的技术类型并将其纳入教学内容,应是合理的选择。

此种预测在任何领域都始终是一项难题,但在信息技术领域尤为严峻。这是因为信息技术随时间的推移而迅速演变,其发展轨迹难以预测。普遍观点认为,真正的创新由于带来了新事物而往往出人意料。人工智能和数字技术则具有双重创新性:人工智能和数据技术作为具有社会性的技术,在物质层面,会受到意料之外的科技进步的影响;另一方面,在社会层面,又是经济和文化因素交织影响的产物。换言之,信息技术的创新,既不是线性的,也不是周期性的,而是一种交织或编织的过程^①。

为了阐明这一点,可以回顾18世纪中期著名哲学家孔多塞(Nicolas de Condorcet)在其著作《人类精神进步史表纲要》^②中的论述。孔多塞认为,从时间上看,进步是线性的,永远向前推进、不会回头,如同一支飞行的箭矢,沿着其轨迹不断前行,不知终点何在,亦不知何时结束。这种线性源于其累积性。不断更新和累积的现代性,与传统的循环形成鲜明对比,在循环的时间观念中,未来永远是对过去的回归,这意味着高度的可预测性,因为不会出现真正的新生事物。

回到当前情境,在人工智能和数字技术领域,显然不能将时间视为循环的,因为技术与产业的发展带来了持续不断的创新:今日之世界与昨日已大不相同。现代社会要求我们必须现代化,这本身就是现代性所特有的创新性的证据。当下不应被视为对过去的简单回归。

这是否意味着我们应当将当下视作绝对的线性发展?无疑,这是一个进步的时代,不断累积成果,且每一天都展现出新的可能性。就此而言,我们确实会倾向于将其视为线性发展。然而,将时间简化为一条直线可能会造成误解。为了使这一论点更具说服力,让我们回顾机器翻译的发展历程。

(一)机器翻译

在20世纪50年代初期,大量资金被投入到机器翻译的研究中。当时许多人坚信,随着计算机技术的进步,尤其是存储容量的提升以及对字符处理能力的增强,快速且低成本地建造能够自动进行语言翻译的机器成为可能,例如能自动将俄语或法语文本转换为英语。支持这一设想的依据是,设计出应用于机器翻译的词典是可行的。

1966年美国参议院委派语言自动处理咨询委员会(ALPAC)发布了一份报告^③,对当时机器翻译研究的进展提出了批评,认为其前景不容乐观。报告指出,现有的机器翻译设计仅限于类似转码(即将信息从一种编码转换为另一种编码)的过程,而没有考虑到语言知识。该报告随即导致美国所有机器翻译实验室资金告急并相继关闭。尽管有些实验室仍在该领域开展研究,但在接下来的至少35年时间里,融入语言知识(包括词汇、语法、语义等方面)被视为唯一有前景的方法。直至20世纪90年代,特别是21世纪第一个十年,我们才见证了机器翻译领域的新生,即统计翻译的兴起。统计翻译的创新在于在翻译中抽取语言模型并以此来训练算法,同时采用了高效的机器学习技术。

(二)控制论

第二个例子源自控制论,人工智能即起源于此。回顾20世纪40年代初,控制论拥有众多拥趸,他们认为这一理论为研究提供了全新视角。沃伦·麦卡洛克(Warren McCulloch)和沃尔特·皮茨(Walter Pitts)设想的自动机网络^④,以及阿图罗·罗森布鲁斯(Arturo Rosenblueth)、诺伯特·威纳(Norbert

① Jean-Gabriel Ganascia, "The Plaited Structure of Time in Information Technology," *AISB Convention*, (2008): 59-64.

② Marquis de Condorcet, *Esquisse d'un Tableau historique des Progrès de l'esprit humain* (Paris: Chez Masson et Fils, 1822).

③ John R. Pierce et al., *Language and Machines: Computers in Translation and Linguistics* (Washington, D. C.: National Academy of Sciences and National Research Council, 1966).

④ Warren S. McCulloch and Walter Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity," *Bulletin of Mathematical Biophysics* 5, no. 4 (1943): 115-133.

Wiener)和朱利安·比奇洛(Julian Bigelow)提出的反馈和目的论机器概念^①,似乎开启了一片新天地。在许多人看来,信息学在工程科学与脑科学及社会科学间架起了桥梁。许多研究者相信,他们能够发现一系列复杂法则,从而对众多领域作出新的解释,诸如生物现象,特别是生物体的运作机制;物理现象,如晶体中原子的自发组织;政治和社会现象,为政府制定法规奠定坚实依据;还包括产生思维的神经元现象。然而,最初的尝试很快便以失败告终。例如,弗兰克·罗森布拉特(Frank Rosenblatt)在尝试表征人工视网膜(Perceptron^②,即“感知机”)时遭遇了众多障碍,其局限性在60年代末由马文·明斯基(Marvin Minsky)和西摩·佩珀特(Seymour Papert)所指出^③。直到80年代初,许多由控制论开辟的、曾一度看似陷入僵局的研究才得以重启。技术复兴的一个例证便是并行分布式处理(PDP)和反向传播学习算法的出现^④,下文还将阐明这远非个例。

(三)信息技术的时间交织

与机器翻译及控制论类似,人工智能和数字技术的发展同样出人意料,其进步轨迹并非始终有序,有时甚至会出现回归过往的趋势。在相关研究中,有些曾被采纳后又弃置的路径,再度浮现并取得了成功。变革的路径不断分化,以至于时间线似乎呈分叉、扭曲状,而非简单的线性。因此,时间结构并不像货架那般有序,而是错综复杂。换言之,在任何一个特定的时间节点,都并存多种可能性,其中一些可能性正由最杰出的研究者们探索,而另一些则似乎淡出了公众的视野。然而,那些原本看似隐匿的线索不时又会重新浮现,而一度清晰可见的线索却又会消失。因此,发展的时间线就像一个缠绕的线团,又如同一束发辫,有时线头散落至视线之外,但在新的光芒照耀下又会再度显现,随后再次隐匿不见。在这个意义上,我们可以将当前的技术发展时间线称为交织的时间线。

四、技术灾难

多年来,很多人都在预言一场灾难的降临,包括科学家^⑤、哲学家^⑥,科幻作家如弗诺·文奇(Vernor Vinge)^⑦、科里·多克托罗(Cory Doctorow)和查尔斯·斯特罗斯(Charles Stross)^⑧,以及工程师^⑨和技术专家^⑩等,这一灾难是参照勒内·托姆(René Thom)的灾变理论^⑪、以数学术语来定义的。人类社会发

① Arturo Rosenblueth et al., "Behavior, Purpose and Teleology," *Philosophy of Science* 10, no. 1 (1943): 18–24.

② Frank Rosenblatt, "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain," *Psychological Review* 65, no. 6 (1958): 386–408.

③ Marvin Minsky and Seymour Papert, *Perceptrons* (Cambridge: MIT Press, 1969).

④ David E. Rumelhart et al., "Learning Representations by Back-propagating Errors," *Nature* 323, no. 6088 (1986): 533–536.

⑤ Irving John Good, "Speculations Concerning the First Ultrainelligent Machine," *Advances in Computers* 6, (1966): 31–88; Marcus Hutter, *Universal Artificial Intelligence: Sequential Decisions Based on Algorithmic Probability* (Berlin: Springer, 2005).

⑥ Nick Bostrom and Milan M. Ćirković, eds., *Global Catastrophic Risks* (Oxford: Oxford University Press, 2008).

⑦ Vernor Vinge, "The Coming Technological Singularity: How to Survive in the Post-Human Era," <https://cdoras.sdsu.edu/~vinge/misc/singularity.html>.

⑧ Cory Doctorow and Charles Stross, *The Rapture of the Nerds: A Tale of the Singularity, Posthumanity, and Awkward Social Situations* (London: Titan Books, 2013).

⑨ Ray Kurzweil, *The Singularity Is Near: When Humans Transcend Biology* (New York: Penguin, 2005), 652; Vernor Vinge, "The Coming Technological Singularity: How to Survive in the Post-Human Era," <https://cdoras.sdsu.edu/~vinge/misc/singularity.html>.

⑩ Kevin Warwick, *March of the Machines: The Breakthrough in Artificial Intelligence* (Champaign: University of Illinois Press, 2004); Eliezer Yudkowsky, "Artificial Intelligence as a Positive and Negative Factor in Global Risk," in *Global Catastrophic Risks*, eds. Nick Bostrom and Milan M. Ćirković (Oxford: Oxford University Press, 2008).

⑪ René Thom, *Structural Stability and Morphogenesis: An Outline of a General Theory of Models* (Reading: Addison-Wesley, 1989).

展中的快速(几乎瞬时、剧烈且不可逆转的)变化,其根源在于技术进步,更确切地说是新一代机器的崛起。这些机器将具备意识或至少具备感知能力,从而能够自主行动,不再受人类指令的约束。这一现象通常被称作“技术奇点”。在此,“奇点”一词源自数学理论中的奇点概念,它是灾变理论的一个分支。这一概念应与“强人工智能”相联系,即创造出具有意识的机器或“通用人工智能”(AGI)。尽管相关理论之间存在细微差异,但均得出一个共同结论:人类历史将迎来巨变。支持人士认为,新时代的到来要么意味着人类时代的终结,即进入一个由自主机器主导的时代,亦即“后人类”(post-humanity)时代;要么是人类与机器的融合,催生出新的人类形态,即超人类(trans-humanity)。

人类面临的此种重大终结风险,被称作“存在性风险”。近年来,多份公开信^①和声明^②均提及了这一可能对全人类构成威胁的风险。然而,细致审视用以支撑这些担忧的论证,便会发现其存在诸多不足^③。与气候变化研究不同,后者基于严谨的方法论,并通过观察实证来支持全球变暖的结论,而上述风险论证大多仅建立在一般的系统论证之上,如同摩尔定律,缺乏证实或反驳的具体可能性。换言之,按照卡尔·波普尔(Karl Popper)的科学方法标准^④,这些论证无法被证伪,这意味着它们缺乏真正的科学价值。这些对未来的预测,其可信度并不高于精灵、恶魔或其他流行传说。通过渲染关于人类终结的恐惧,这些观点近似于中世纪的世界末日观点(同样设想了世界末日情景),或类似于“千禧年主义”(millenarianism),即在一场大乱之后引发社会的根本变化。因此,向年轻一代清晰地阐释这些观点的错误性和误导性,显得尤为重要。

五、人工智能素养:教什么?

梳理总结在高等教育领域关于人工智能与数字技术教学的如上反思,下文将阐述笔者目前在此方向上的探索。具体而言,受到不同听众尤其是学生们的多次课程启发,笔者为普通读者撰写了一本人工智能教学书籍——《向人类解释人工智能》^⑤,意在普及人工智能知识。该书采用了笔者与四位青少年对话的形式。全书共分为五个主要部分,可以对应于面向所有学生的入门级人工智能文化课程的不同阶段。

(一)关键术语的定义

首先要对一些常被误解的基本概念进行界定,包括计算(computation)、智能(intelligence)、算法(algorithm)、计算机(computer)、机器人(robot)、机器人程序(bot)、智能体(agent)、自主性(autonomy)等。例如,人工智能的设计初衷并非打造具有心智的“智能机器”,而如今已不限于构建以其简单行为为特征的聪明机器,为了理解这一点,澄清“智能”这一概念的不同含义至关重要。“智能”一词源于19世纪末期的心理学领域,在心理学中指属于心智的一系列心理能力,如感知、推理、记忆、交流等。

另一个需要阐释的关键概念是“智能体”,指涉一种能获取信息并采取行动的实体。“智能体”常被误称为“人工智能”,这种用法是令人误解的——应当予以纠正——因为此种表述暗示了有形或虚拟实体具有智能或被赋予了心智,这显然是不准确的。

(二)成功与应用案例

因此,有必要对我们日常正在使用的人工智能应用进行全面了解。第一个应用就是网络(Web),它是互联网与超文本相结合的产物。超文本是一种记忆模型,由特德·纳尔逊(Ted Nelson)于1965年设计,并利用当时的人工智能技术(即列表编程)进行编程。因此,超文本(以及由此衍生的网络)在两个层

①大量此类公开信可见生命未来研究所(Future of Life Institute)网站(<https://futureoflife.org/fli-open-letters/>)。

②也有许多声明刊载于大众媒体,参见:Kevin Roose,“A. I. Poses ‘Risk of Extinction,’ Industry Leaders Warn,”<https://www.nytimes.com/2023/05/30/technology/ai-threat-warning.html>。

③Jean-Gabriel Ganascia, *L’I. A. expliquée aux humains* (Paris: Seuil, 2024)。

④Karl Popper, *The Logic of Scientific Discovery* (New York: Routledge, 1959)。

⑤Jean-Gabriel Ganascia, *L’I. A. expliquée aux humains* (Paris: Seuil, 2024)。

面上与人工智能相关:一是它模拟了心智能力;二是它通过人工智能技术编程得以实现。还有许多其他常用工具也是基于人工智能技术构建的,例如,采用语音识别技术的听写设备、利用指纹或面部识别技术解锁的手机、自动驾驶汽车以及聊天机器人等。

在未来,人工智能将广泛应用于多种实用场景,如参与医疗保健、在宇宙飞船上调节温度、控制并激活机器人、在现代农业中辅助选种及调整灌溉流量、在环境保护领域监测森林火灾等。其潜在应用领域可谓无穷无尽。向年轻学生阐明人工智能应用的范围并非固定不变,且其发展取决于他们,这一点至关重要。

(三)运作方式

要从历史视角对人工智能技术的发展进行梳理。大众对人工智能的历史知之甚少,因此常常存在时代错误(anachronism)和误解。铭记人工智能的多重起源至关重要,不仅应记住相关学科在1956年正式诞生于达特茅斯学院暑期研讨会,也不应忽视莱布尼茨在18世纪首次尝试构建推理机器的努力,以及斯坦利·杰文斯(Stanley Jevons)的“逻辑钢琴”——利用布尔代数模拟逻辑推理。此外特别是控制论,它在1943年便催生了神经网络的概念^①。

回顾最初人工智能方法所依据的原则同样非常重要,它们在20世纪50年代末促成了基于启发式搜索的自动证明定理技术,以及机器学习的早期尝试,特别是阿瑟·塞缪尔(Arthur Samuel)在强化学习领域的开创性工作^②。同样,阐释所谓的“人工智能寒冬”也十分关键,其实际情况并不如人们通常所描述的那般“寒冷”;以及随着“语意学转向”(semantic turn)而来的人工智能复兴,并在20世纪70年代末催生了现代知识表示技术和专家系统。

还需回顾神经网络在1986年伴随着反向传播算法的复兴,以及在20世纪90年代初期遭遇的冷落,直至2012年随着卷积神经网络(CNN)和深度学习技术的发展而再度崛起。此外,不应忽略生成式人工智能在2014年兴起,伴随着生成式对抗网络(GAN)的出现,以及2017年Transformer模型和大语言模型(LLMs)的问世。

同时,应适度简化对不同人工智能技术的介绍,否则将无法涵盖其多样的灵感来源,例如自然进化与遗传算法,或是人类心理学与语义方法及“组块化”(chunking)。

最后,要展开对计算技术未来的探讨,尤其是量子计算,它有可能彻底革新人工智能领域。

(四)机器能否超越人类?

人们普遍担忧机器智能将超越人类自身,甚至发展到无法被控制的境地。正如前文所述,科幻作家们描绘了多种设想,例如“技术奇点”“强人工智能”“通用人工智能”以及“存在性风险”等,在这些设想中,机器的崛起将导致人类的终结。需要明确界定这些观点的含义,以说明它们缺乏科学依据的原因。

同样重要的是,要阐明在许多任务上机器的表现已经超越了人类,并且这一现象由来已久。例如,在乘法运算方面,机器的效率远胜于人类;在国际象棋和围棋上,机器的表现也优于最顶尖的人类棋手,等等。要解释清楚,这些现象并不值得我们恐慌,因为尽管机器在许多任务上比人类更有效率,但人类始终是计算机所执行工作的发起者。

还应提供探讨“图灵测试”和当前人工智能局限性的机会,因为很难对采用人工智能技术编程的机器会发展成何种形态给出明确的解释。需要讨论人工智能的缺陷,例如“幻觉”(hallucination)概念(尽管聊天机器人的错误输出并不是真正的幻觉,因为机器无法感知除人为构造外的任何事物)。

(五)风险、危机与伦理问题

要专门探讨大规模部署人工智能技术所带来的社会风险。这些风险繁杂,涉及大规模监控、深度伪造

^① Warren S. McCulloch and Walter Pitts, “A Logical Calculus of the Ideas Immanent in Nervous Activity,” *Bulletin of Mathematical Biophysics* 5, no. 4 (1943): 115–133.

^② Nick Bostrom and Milan M. Ćirković, eds., *Global Catastrophic Risks* (Oxford: Oxford University Press, 2008).

信息的制作与传播,以及基于人工智能的倾向性决策导致的歧视问题。此外,还存在一种情形,可能出于多种因素考虑,如经济效益、公平性或系统化需求,而选择让机器代替人类做出决策,从而限制了人的自由。

意识到这些风险并采取措施保护自身免受其影响是非常重要的。近年来,众多伦理委员会就相关问题撰写了报告:2019年约有84份报告出炉^①,而据2022年发表的一篇文章,报告数量超过350份^②。这些报告提出了诸多原则与建议,根据布伦特(Brent)的研究,2019年数量超过62项^③;蒂德容(Tidjon)的研究则显示,2022年数量超过了510项^④。然而,这些报告并未聚焦于少数几项建议,且所提出的建议之间甚至存在矛盾之处。

在上述关于人工智能的伦理考量之外,国家和国际组织,如欧盟委员会,也在试图规制人工智能,以限制人工智能系统的发展。自2019年以来已经出台了多项法律,旨在控制人工智能在各项技术中的滥用。

六、结语

众多伦理委员会近期编写的大量报告,以及欧洲法规,特别是针对人工智能的法规即所谓的《人工智能法案》(AI Act),其面临的主要挑战源于大多数建议所依据的概念或风险界定。实际上,英语“risk”(风险)一词的词源来自意大利语单词“risco”,后者又源自拉丁语单词“resicare”,意为切割。刀具有切割能力,虽然有时可能带来危险,但在大多数情况下它非常有用。因此,不能仅因为刀能切割而禁用它。基于人工智能的技术固然存在风险,我们可以设想它们被滥用的危险。然而,真正的危险往往难以预测。技术创新的本质是创新性的,这意味着它们常常出人意料。如前文所述,回顾数字技术和人工智能的创新历程,说明要预测它们的发展非常困难,其发展在某种程度上具有无序性,即不可预测性。因此,关于数字技术和人工智能,真正的伦理问题不在于试图预测风险,而在于准备好应对技术发展的不可预测性并成功地管控技术带来的风险。

Some Challenges in Teaching AI in Higher Education

Jean-Gabriel Ganascia

(Sorbonne University, France)

Abstract: Artificial Intelligence (AI) is now so important today in higher education that every student, not just those who want to become engineers and AI specialists, but anyone who wants to become a gentleman in the information society of the 21st century, must learn the rudiments of this discipline and be trained in the use of many AI tools during their studies. As a consequence, they will have to become familiar with AI in order to be active citizens. This education, which corresponds to an AI literacy, requires not only an awareness of the current AI technologies and how to handle them, but also of the history and emergence of AI as a scientific discipline. To this end, this paper describes how AI can be taught to students, without limiting the education to technical issues. The aim is to prepare them for the information society in which they will have to live in the future.

Key words: artificial intelligence; Generative AI; plaited structure of time; ethical and humanistic issues; AI literacy; new practices

① Brent Mittelstadt, “Principles Alone Cannot Guarantee Ethical AI,” *Nature Machine Intelligence* 1, no. 11 (2019): 501–507.

② Lionel Nganyewou Tidjon and Foutse Khomh, “The Different Faces of AI Ethics Across the World: A Principle – Implementation Gap Analysis,” <https://hal.science/hal-03689128v1>.

③ Brent Mittelstadt, “Principles Alone Cannot Guarantee Ethical AI,” *Nature Machine Intelligence* 1, no. 11 (2019): 501–507.

④ Lionel Nganyewou Tidjon and Foutse Khomh, “The Different Faces of AI Ethics Across the World: A Principle – Implementation Gap Analysis,” <https://hal.science/hal-03689128v1>.