

从“通用”到“垂直”:人工智能思政大模型的定制化应用

张国启,段港回

(华南理工大学马克思主义学院)

摘要:加快人工智能思政大模型的定制化应用,是开辟思想政治教育发展新赛道、塑造思想政治教育发展新优势的应然之举。就其内涵而言,人工智能思政大模型是以通用大模型为基座,以强相关的数据或专业知识为训练参数,能够体现思想政治教育立场、观点和价值属性的专业智能体。正是由于人工智能思政大模型受训于高质量数据或强相关的专业知识,而通用大模型只受训于泛在的数据或通用知识,才造成了二者之间的显著差异。相较于通用大模型,人工智能思政大模型的表现更具专业性、准确性、可靠性和安全性。在定制路径的选择上,依托全架构通用大模型进行垂直领域增强的方案,因其具备定制门槛低、算力消耗少等优势,更契合思想政治教育精智化发展的现实要求,该方案主要通过抽取通用大模型中已有的思想政治教育知识、向通用大模型输入新增思想政治教育知识、基于思想政治教育混合专家评估的方式增强其在思政领域的现实表现。面向未来,人工智能思政大模型建设应着眼于“开发-部署-应用-反馈”的全流程体系,通过聚焦能力建设,增强人工智能思政大模型开发“效能”;通过明确应用场景,彰显其部署“效用”;通过锚定工具导向,规范其应用“效度”;通过注重风险防治,优化其反馈“效果”。

关键词:数智技术;人工智能;思想政治教育;大模型;数智化转型;人机交互

以 DeepSeek 为代表的通用大模型,虽然已具备广泛的通用知识和跨领域迁移能力,但因缺乏知识深度而无法满足特定领域的精细化要求。与通用大模型不同,垂直大模型又称“垂类大模型”或“行业大模型”,是根据特定领域的需要和应用而定制的人工智能大模型。2025年,《教育部等九部门关于加快推进教育数字化的意见》明确提出,要“加快建设人工智能教育大模型”“推动思政、科学教育、美育、心理健康等领域及数学、物理等基础学科专题大模型垂直应用”^[1]。《中华人民共和国国民经济和社会发展第十五个五年规划纲要》(以下简称

“十五五”规划)再次强调,要“抢占人工智能产业应用制高点,全方位赋能千行百业。”^[2]这些重要论述一方面体现了人工智能与思想政治教育深度融合的时代要求,另一方面指明了推动思想政治教育专题大模型垂直应用的实践方向。当前,学术界主要探究了 DeepSeek、ChatGPT 等通用大模型赋能思想政治教育的逻辑机理与实践路径^[3],尽管也有少数学者对人工智能思政大模型(以下简称“思政大模型”)的应用图景、逻辑向度与实践进路等进行了探讨^[4],但在论及大模型时,相关研究并没有严格区分通用大模型与垂直大模型的不同。不可

引用本文:张国启,段港回.从“通用”到“垂直”:人工智能思政大模型的定制化应用[J].河海大学学报(哲学社会科学版),2026,28(2):21-29.

基金项目:国家社会科学基金一般项目(21BKS150)

作者简介:张国启(1975—),男,教授,博士,主要从事马克思主义与当代意识形态研究。E-mail:zhangq1216@scut.edu.cn

忽视,随着模型微调、基于人类反馈的学习微调、检索增强生成等技术的发展,通用大模型与垂直大模型在特定领域的表现差异已日益凸显,加快思想政治教育专题大模型的垂直应用已显得尤为必要和迫切。为此,本研究从思政大模型的理论内涵及技术准备出发,选择定制门槛低、算力消耗少的全架构通用大模型进行垂直领域增强^①,并基于思政大模型“开发-部署-应用-反馈”的全流程体系提出实践策略,以期为加快思政大模型的落地与应用、推动思想政治教育数智化发展提供理论参考。

一、人工智能思政大模型的内涵及特征

生成式人工智能、大语言模型等技术的蓬勃发展为人工智能与思想政治教育的深度融合带来了新契机,思政大模型并非大模型技术系统与思想政治教育内容的简单叠加,而是兼具技术属性、价值属性与教育属性的智能复合体。当前,学界对思政大模型的探索仍处于起步阶段,厘清思政大模型的内涵特征是把握其本质规律、更好发挥其技术功用的基本前提。

1. 人工智能思政大模型的内涵释义

对思政大模型内涵的把握,可从理论层面和技术层面两个维度入手:理论层面,构建思政大模型是出于一种理论自觉,是为了推动思想政治教育精智化发展并满足思想政治教育者的智能需要;技术层面,正是由于模型微调、基于人类反馈的学习微调及检索增强生成等技术的完善和更新,思政大模型才能进一步提升智能体与思想政治教育的适配性。

理论层面,思政大模型是以强相关的数据和专业知 识为训练参数,通过一定技术架构生成的,能够体现思想政治教育立场、观点和价值属性的专业智能体。就其形态而言,思政大模型是大模型技术与思想政治教育深度融合的结果,属于一种人造“机器”,虽然表现出一定的“类人化”智慧和“主观能动性”,但其始终需要人的驱动作为“开关”。就其功能而言,思政大模型集通用智能和专业智能为一体,多模态理解和输出能力使其能够精确理解思想政治教育主客体的意思表示和智能需要,便于开展日常活动。而思政大模型的“智能涌现”,即“随着

模型的规模扩展达到临界点后,模型展现出新性质的现象,其中一种表现就是模型性能大幅提升”^[5],这能够使其经过训练后具备“专家级”智慧,从而应用于思想政治教育的智慧教学、学术研究等专业场景。就其属性而言,思政大模型具有鲜明的意识形态属性,聚焦思想政治教育,受社会主义意识形态的价值统摄,服务于社会主义意识形态的教育和引导。

技术层面,思政大模型是以通用大模型为基座,经过各类技术改进并服务于思想政治教育的智能体。对大模型相关技术的理解是推动其与思想政治教育深度融合的必要前提。一般而言,为节约成本、提升效率,垂直大模型的构建通常以通用大模型为基座,而后通过一定的技术微调开展模型定制。具体而言,模型微调技术是指在通用大模型的基座上,通过微调或添加部分参数提升大模型专业表现性能,以使其符合特定领域的应用需要。这一技术可大幅降低思政大模型的定制成本,满足其多场景部署的应用需要。基于人类反馈的学习微调技术是指“将人类偏好转换为数字奖励信号,即奖励模型接收一段文本序列并输出标量奖励值”^[6],进而提升模型的表现性能。这一技术可以使大模型根据思想政治教育主客体的反馈,深度理解该领域的相关内容和个体需要。如,检索增强生成技术是将检索外挂数据库中的信息以自然语言的形式放入提示中,以增强模型生成回答的能力,这是为了降低数据和知识训练成本,通过接入外挂知识库的方式提升模型回答的质量和相关性。这一技术的应用,可使思政大模型增强模型输出内容的专业性和时效性。以上技术的完善和更新为构建思政大模型提供了技术前提。

2. 人工智能思政大模型的显著特征

思政大模型与通用大模型的表现存在显著差异,对这些差异的分析可进一步理解思政大模型的专业表现。马克思恩格斯认为:“物质的原始形式是物质内部所固有的、活生生的、本

① 所谓垂直领域增强,是指以强相关领域知识注入为手段,对通用大模型进行优化,提升其在垂直领域应用能力的过程。这里的垂直领域,特指思想政治教育领域。

质的力量,这些力量使物质获得个性,并造成各种特殊的差异。”^[7]在马克思恩格斯看来,正是物质内部的固有力量造成了物质之间的差异,数据或知识是影响大模型表现的“固有能力”,正是由于思政大模型受训于高质量数据或强相关的专业知识,而通用大模型只受训于泛在的数据或通用知识,才造成了二者之间的显著差异。

其一,相较于通用大模型,思政大模型的表现更具专业性和准确性。一方面,思政大模型的表现更具专业性。相较于通用大模型,思政大模型数据的训练或知识的注入均与思想政治教育领域强相关。通过模型微调技术,垂直大模型对这一领域的内容理解更具深度、输出的结果也更为聚焦,这能有效减少通用大模型的“幻觉”问题,即为满足使用者需要,但由于知识深度不足而编造的看似可信的虚假信息。特别是检索增强生成技术的应用,通过进一步整合外部特定信息和数据进行内容生成,能显著提升思政大模型的专业表现。另一方面,思政大模型的表现更具准确性。为起到“降本增效”的作用,思政大模型需要接入外挂知识库,这能有效减少不必要的数据训练和知识注入,降低智能体的开发成本和应用成本。如,中国共产党思想理论资源数据库、读秀、中国知网等外挂知识库会对习近平总书记关于思想政治教育领域的最新论述、学界关于思想政治教育领域的最新成果等进行定期更新,接入这些外挂知识库,能够使思政大模型及时访问最新的数据和知识,进而对矛盾知识和不一致数据进行校准,确保输出内容的准确性和时效性。

其二,相较于通用大模型,思政大模型的表现更具可靠性和安全性。一方面,思政大模型的表现更具可靠性。应用检索增强生成技术,可使思政大模型的输出内容均可追溯到特定信息源。如,使用者在利用思政大模型进行相关信息检索和提问时可要求其提供准确的网络地址或者信息来源,这为使用者验证思政大模型输出内容的可信度提供了方式。通过明确模型如何使用特定数据生成特定答案,增加了管理决策者对人工智能生成的分析和建议的理解和信任,降低了算法厌恶^[8]。另一方面,思政大

模型的表现更具安全性。通用大模型受训于公开的数据集或知识库,对数据隐私和知识保护的安全性要求较低,而思政大模型受训于特定的数据集或知识库且具有鲜明的意识形态属性,加之该领域的数据或知识较为敏感,因此对数据隐私和知识保护的安全性要求较高。针对思政大模型内部的数据集或知识库,需采用控制访问与加密技术,根据特定人群和访问需要严格控制外部访问权限,并对意识形态问题、网络舆情等敏感数据的传输和存储过程进行加密,这将有效防止相关敏感数据和信息的泄露。

二、人工智能思政大模型的定制方法

一般而言,垂直领域大模型的定制方法包括基于全架构通用大模型的垂直领域增强、基于预训练模块的垂直领域大模型改造及无预训练模块的垂直领域大模型全架构构建3种^[5]。其中,基于全架构通用大模型垂直领域增强的定制难度与算力要求较低,该定制方法旨在解决垂直领域的核心业务,而非追求“大而全”,这符合思想政治教育精智化的发展定位。

1. 抽取通用大模型已有的思想政治教育知识进行垂直领域增强

思政大模型是以通用大模型为基座的智能体,以国产通用大模型 DeepSeek 为例,该模型功能较为全面,适用于科研、管理、教学等多种任务场景,千行百业正在进行广泛部署。若以该模型为思政大模型的基座,则不再需要修改模型的内部架构,因为其基本具备思想政治教育所需要的数据模态,而只需通过注入思想政治教育知识的方式对其进行垂直领域增强即可。DeepSeek 在本地训练或网络部署的过程中,有可能已经接触过部分思想政治教育知识,通过微调提示词的方式可以更好地抽取模型本身所具有的思想政治教育知识,从而增强其在思政领域的表现,这种提示词的微调分为硬提示和软提示。

其一,通过设计“硬提示”,调用已有的思想政治教育知识。硬提示是指由思想政治教育相关主体根据场景需要或任务要求制定的人工提示词,这些提示词可能是特定的思想政治教育话语或其他能够指导模型理解思想政治教育的

特定任务的嵌入量,模型会根据这些人工提示词生成相关结果。优化提示词的工作往往需要具备思想政治教育专业知识的相关主体进行操作和探索。如,思想政治教育学术研究场景的提示词可能更具专业性,而思想政治教育课堂教学场景的提示词可能更偏向生活化。无论是何种场景的提示词,都需要相关主体与思政大模型进行更为深入的交流体验,以测试不同提示词对模型回答效果的影响程度,从而寻找最优解。由于硬提示是根据特定场景或任务设计的,这种提示词的灵活性较低。

其二,通过设计“软提示”,调用已有的思想政治教育知识。软提示是指通过模型自主学习的模式来代替固定的或人工设计的提示词,具有更大的灵活性。软提示的核心是将提示词的学习纳入模型的自主训练中,模型不仅需要在思想政治教育实践中学习如何生成最佳提示词,还需要学习如何即时响应特定任务。这种不涉及模型参数改变的提示词微调,不仅保留了预训练模型学习到的丰富知识,还减少了计算资源和储存资源的需求^[5]。如,大模型可以基于思想政治教育相关主体的交互数据,自主抓取教师、学生、科研工作者在不同场景中使用的提示词,而后通过模拟思想政治教育管理者与被管理者、思想政治教育施教者与受教育者、思想政治教育专家与普通用户之间的交流进行深度学习。通过这种端到端的模拟训练,模型可以根据语义理解和逻辑推理自主调整提示词的嵌入量并完成特定任务。

2. 向通用大模型输入新增思想政治教育知识进行垂直领域增强

通用大模型可能由于本地化部署或进行联网搜索而具有部分思想政治教育知识,但其储存的思想政治教育知识类型为公开且泛在的知识,如百度百科、新闻资讯、公众号推文等开放平台的思想政治教育知识。增强通用大模型在思政领域的表现效果,不仅需要这些公开泛在的思想政治教育知识作支撑,还需要抽取特定知识库中的高质量知识作标注。如,中国知网、中国共产党思想理论资源库等知识库中的思想政治教育知识就需要付费使用。此时,就需要输入新增的思想政治教育知识,以引入问题的

相关背景信息,进一步提升大模型完成特定任务的表现,这种输入方法包括人工输入与接入外挂知识库两种。

其一,通过思政大模型部署者、思想政治教育工作者手动上传思想政治教育新知识。以人工方式输入思想政治教育新知识是一种最为直接的方式,该方式使思政大模型的本地端更符合个人的风格和偏好。一般而言,现有的大模型支持用户以文件、图片等多模态形式进行知识传输,用户不仅可以在大模型的本地端传输思想政治教育教材、文献等常规文件,还可上传具有鲜明思想政治教育特征的主题图片。在传输完成后,大模型根据任务要求进行思想政治教育信息读取、文献分析及图像识别,以供用户进行选用。在这一过程中,思想政治教育新知识不仅会以数据形式保存于本地端的个人自建库,还会以“投喂”形式不断训练大模型。在无数用户的持续“投喂”下,大模型逐渐形成符合思想政治教育特征的叙事风格,逐渐具备理解思想政治教育的语义能力和逻辑推理能力。

其二,通过接入思想政治教育外挂知识库的方式输入新知识。当用户无法通过人工方式上传足够的思想政治教育新知识时,就需要模型部署者通过接入外挂知识库的方式,即通过检索增强生成技术实现思想政治教育新知识的注入。接入外挂知识库应注重思想政治教育知识的整体性,即需要接入“包括基础理论型知识、方法论型知识、实践应用型知识与政策规范型知识”^[9]在内的外挂知识库。目前而言,中国共产党思想理论资源库、中国知网、读秀、各高校思想政治教育论文自建库、思想政治教育典型案例库、各科研单位与思想政治教育相关的智库,基本涵盖了上述知识范畴。针对这些外挂知识库,应根据知识的应用频率和使用对象进行分类管理:对于中国知网、读秀等常用的外挂知识库,应确保接入的稳定性与连续性;对于各高校的论文自建库及相关智库等,则应确保接入的隐私性与安全性。

3. 基于思想政治教育混合专家的评估进行垂直领域增强

调用通用大模型已有的思想政治教育知识及向其注入新增的思想政治教育知识后,通用

大模型在思想政治教育领域的表现将显著增强。此时,大模型就完成了从“通用”到“垂直”的蜕变,思政大模型由此诞生。但受多种因素的影响,思政大模型可能在初期具有表现的不稳定性,这就需要对这一智能体进行专业评估,以通过人类反馈学习的方式对其性能进行优化。值得注意,对这一智能体的表现进行专业评估是一项较为复杂的任务,与自然科学强调可量化、可验证不同,思政大模型的评估往往取决于价值性、逻辑性、深刻性等主观标准。因此,具有深厚理论积淀、丰富实践经验的思想教育混合专家对思政大模型的评估就变得尤为重要。

其一,科学构建思政大模型多维评估指标。科学的评估指标是思想政治教育专家开展内容评估的前提依据,应围绕价值正确性、内容适配度、表达规范性等维度细化指标。其中,价值正确性属于首要维度,应重点检验思政大模型输出内容是否符合党和国家的大政方针、是否存在历史虚无主义等问题、是否符合主流意识形态要求、是否存在权威表述错误等情况;内容适配度属于核心维度,主要检验输出内容是否与思想政治教育教学、科研、评价等场景相适配,是否符合特定教育阶段学生的认知规律和成长要求,是否能满足思想政治理论课的教学要求等;表达规范性属于重要维度,主要检验生成的思想政治教育话语是否规范、流畅,学术观点是否严谨,引用的文献是否真实可靠,来源是否属于官方权威等。

其二,严格规范思政大模型的评估流程。在评估的准备阶段,应组建一支包括思政课教师代表、思想政治教育学科专家、思想政治教育科研代表、思想政治教育管理部门代表、思政大模型技术代表等在内的混合专家团队。按照应用场景或任务要求整理大模型生成样本,明确告知专家评估的依据、目标及范围。在评估的实施阶段,可采用图灵测试评估^[10],即由相关专家对思想政治教育内容样本进行独立审阅和匿名打分,针对价值正确性、内容适配性及表达规范性等具体评估指标,明确给出“优”“良”“差”的判断,并阐述存在的具体问题。在评估研讨阶段,应对思政大模型生成的争议内容进行

充分论证,以民主集中制形成统一评估意见。在评估反馈阶段,应形成标准化、专业化的评估报告,涵盖专家组成情况、评估依据、评估争议点、评估结果、改进建议等,而后将报告反馈至思政大模型的技术研发团队、本地部署方,以实现模型的持续优化。

三、人工智能思政大模型的应用策略

思政大模型的应用是一项复杂的系统工程,需围绕开发、部署、应用、反馈的全流程运行体系展开。其中,聚焦能力建设是前提环节,旨在增强智能体的开发“效能”;明确应用场景是基础环节,旨在彰显智能体的部署“效用”;锚定工具导向是关键环节,旨在规范智能体的应用“效度”;注重风险防治是保障环节,旨在优化智能体的反馈“效果”。

1. 聚焦能力建设,增强人工智能思政大模型开发“效能”

思政大模型的开发质量既取决于底层算法和训练数据,还取决于其接收提示(问题表达)的有效性^[11]。因此,要高度重视底层算法的价值性与逻辑性,充分关注训练数据的质量而非数量,注重不同应用场景提示词的适配性,全面增强思政大模型开发“效能”。

其一,聚焦思政大模型的算法设计能力建设。思政大模型既需要呈现鲜明的社会主义意识形态属性,又需要具备对思想政治教育相关概念与话语的精确理解能力,这就需要强化相关算法的设计。具体而言,要将社会主义意识形态转换为可量化、可校准的算法规则,开发专用的校准插件,对模型输入和输出内容进行价值审核,确保主流价值驾驭算法。同时,要优化语义理解和逻辑推理算法,将思想政治教育权威语料作为训练材料,开展算法逻辑对抗训练,以此强化算法对思想政治教育信息的深层理解和逻辑推演。

其二,聚焦思政大模型的数据标注能力建设。强相关、高质量的思想教育数据标注直接影响思政大模型的表现,为此应组建“思政专家+技术人员”复合型标注团队。思想政治教育专家往往对领域内的思政数据较为熟悉,可根据思想政治教育发展历程、积淀资源及

获取经验等方面制定数据标注标准,而后引导技术团队对思想政治教育语料、图文等进行数据标注并展开合规性校验,以确保数据质量。

其三,聚焦思政大模型的提示交互能力建设。在大模型时代,问题定义能力直接影响模型生成内容的质量与方向。提升思政大模型的提示交互能力需聚焦思想政治教育基本答疑场景,通过循环对话优化提示词,引导大模型沿着提示词方向进行逻辑推理。为此,一是要聚焦思想政治教育导向型内容生成场景,通过细化调整提示词以提升生成目标的针对性;二是要聚焦思想政治教育材料优化与拓展场景,通过提示词的切换把握思想政治教育内容转化程度。

2. 明确应用场景,彰显人工智能思政大模型部署“效用”

《国务院关于深入实施“人工智能+”行动的意见》明确提出,“探索模型应用新形态,提升复杂任务处理能力,优化交互体验。”^[12]思政大模型的多模态融合、语义理解及动态交互功能深刻影响着思政课教学、思想政治教育科研及评价,为此应积极探索思政大模型在智慧教学、知识共创及数智评价等方面的应用新形态,彰显其在多维场景处理复杂任务的“效用”。

其一,应用于以思政课教学为主的基础性场景。课堂教学是开展思想政治教育的主渠道。相较于通用大模型,思政大模型在思政课中的应用已不再局限于基础教学辅助,而是拓展至专业情境的模拟及批判性思维的引导。其多模态融合功能,能够创设高度符合思想政治教育叙事风格的专业情景,使师生能够“通过观之有形、叙之有物、触之有感的普遍性形式”^[13]体悟理论知识。此外,强相关且高质量的数据训练使其具备了思想政治教育专家级的“类人智慧”,这使其具有引导师生在课堂追问中养成批判性思维的能力。

其二,应用于以思想政治教育科研为主的发展性场景。科学研究是推动思想政治教育高质量发展的重要方式。相较于通用大模型,思政大模型在学术研究中已不再局限于进行文献检索与辅助分析,而是能够“在研究全过程以认知协作者进行深度参与”^[14];生成环节,思政

大模型通过对强相关数据的处理,引导思想政治教育科研工作者发掘符合本学科研究方向的新路径;批判环节,研究者通过对思政大模型输出内容的质疑和发问,引导模型作出深层修正,最终在与模型的不断交互中推动思想政治教育学术研究的共创。

其三,应用于以思想政治教育评价为主的保障性场景。质量评价是确保思想政治教育得到持续优化的重要保障,相较于通用大模型,思政大模型在质量评价中已不再局限于某种结果性指标,而是转向全过程动态评价。这一多场景部署,能将思想政治教育实践活动产生的“文本、图像甚至人本身在内的一切内容被纳入评价范围”^[15],而后对其发展趋向进行实时可视化呈现。这种转变强调对思想政治教育主客体的持续跟踪,使质量评价更为精确和客观。

3. 锚定工具导向,规范人工智能思政大模型应用“效度”

习近平总书记指出,要“坚持以人为本、智能向善”^[16],这一论述既锚定了人工智能的工具导向,也为其发展明确了人本价值导向。思政大模型的应用必须将人的主体性放在核心位置,以促进人的全面发展为价值目标,将其作为拓展人的能动性、增强人的自主性及激发人的创造性的工具,规范其应用“效度”。

其一,将思政大模型作为拓展人的能动性的工具。思政大模型的应用催生出新的教学模式、科研范式及评价方式,这对思想政治教育主客体的数字素养提出了更高要求。这种技术浪潮倒逼思想政治教育主客体积极发挥主观能动性,通过主动提升数字素养以适应思想政治教育的数智化转型。具备较高数字素养后,思政教育主客体能有效防范“AI幻觉”、避免技术异化,在人工智能技术与思想政治教育深度融合中发挥人的能动性。

其二,将思政大模型作为增强人的自主性的工具。“时间实际上是人的积极存在”^[17],时间自主是人具有自主性的重要标志。思政大模型通过释放思想政治教育主体的可控时间增强人的自主性。具体而言,这一智能体不仅可以应用于思想政治教育教案生成、课件制作、文献搜集与分析、事务管理等日常性工作,还可辅

助相关主体开展科学研究、情感疏导等高阶工作,这将有效释放人力时间,提升思想政治教育主体从事高阶工作的效率和质量。

其三,将思政大模型作为激发人的创造性的工具。发散和收敛两种思维对于创造性潜能的激发都是必要的^[18],思政大模型的应用可以有效触发用户的发散思维和收敛思维。以思想政治教育学术场景为例,当用户输入特定提示词后,这一智能体调动强相关的数据进行学术热力图谱呈现并提供多元研究视角和思路,促使用户进行发散思考。在思考过程中,用户需要对生成的多元观点进行收敛,即进行批判性吸收,以形成“以我为主、为我所用”的自主模式。

4. 注重风险防治,优化人工智能思政大模型反馈“效果”

作为现象级的技术创新,思政大模型的部署可能会对思想政治教育系统产生适应性挑战。为优化思政大模型反馈“效果”,应及时扭转“事后响应”的治理思维,转向对技术、制度、伦理等的“提前介入”,以推动这一智能体“向上向善”。

其一,注重思政大模型的技术加固。思政大模型本身属于一种技术范畴,对其进行技术加固往往会在风险防控中起到基础性作用,为此应从访问控制、数据加密、隐私保护等方面构建闭环防护体系^[19]。针对思想政治教育者、受教育者及管理者等访问人员,应通过多重身份认证赋予其不同的访问权限,实现相关数据可用而不可见。针对涉意识形态的敏感数据及涉访问者隐私的相关数据,应建立覆盖端到端后的动态加密技术,确保数据隐私与安全。

其二,注重思政大模型的责任追溯。思政大模型作为一种技术人工物,责任溯源往往会在风险防控中起到关键性作用,为此应完善从模型开发方、本地部署方到具体使用方的连带责任体系,采用举证责任倒置原则,要求相关责任方进行自证,并根据责任认定进行相应处置。如,具体使用方利用智能体生成思想政治教育相关内容时应在发布端悬挂合成内容标识;本地部署方应提前进行伦理审查,而模型开发方则应进行技术调试和价值嵌入。

其三,注重思政大模型的伦理对齐。思政

大模型的最终目的是实现对人的价值引领,注重伦理对齐往往会在风险防控中起到保障性作用,为此,应根据思政大模型应用的不同场景提升伦理审查的针对性。如在思想政治教育场景中,应针对该场景交互性强、政治要求高的特点,组成包括技术开发者、思政课教师及学生代表等在内的伦理审查团进行相应审查,还可通过开发虚拟现实模拟系统,让审查成员体验伦理失序的相应后果,而后通过“人类反馈强化学习”对这一智能体进行伦理对齐。

四、结 语

通用大模型与垂直大模型的根本区别在于数据训练不同,这种差异在大模型内部表现为某种参数或规则,在大模型外部则表现为生成的知识。换言之,以通用大模型为基座定制思政大模型,关键在于是否向通用大模型输入了强相关、高质量的政治教育数据或知识。基于全架构通用大模型的垂直领域增强方法,在不改变通用大模型内部架构的情况下,从数据或知识层面对定制思政大模型作出了有益探索,并着眼于“开发-部署-应用-反馈”全流程体系,为加快这一智能体的应用提供了增强“效能”、彰显“效用”、规范“效度”及优化“效果”的策略。面向未来,随着相关技术的成熟和人才队伍的建设,也可尝试运用“基于预训练模块的垂直领域大模型改造”与“无预训练模块的垂直领域大模型全架构构建”两种方法定制思政大模型。虽然这两种方法十分复杂且对算力要求较高,但其追求“大而全”的专业表现也将更加出色。不可否认,当前对智能体的认识仍处于探索阶段,在实践过程中可能会面临技术依赖、数据泄露、伦理失范等风险,为此应主动研判并提前介入相关风险,推动思政大模型为形成强大思政引领力、培育更多时代新人作出更大贡献。

参考文献:

- [1] 教育部等九部门关于加快推进教育数字化的意见 [EB/OL]. (2025-04-11) [2026-02-10]. https://www.gov.cn/zhengceku/202504/content_7019045.htm.

[2] 中华人民共和国国民经济和社会发展第十五个
五年规划纲要[M]. 北京:人民出版社,2026:41.

[3] 王思佳,张瑜. 生成式人工智能介入思想政治教
育的基本逻辑及矛盾调适——从 ChatGPT 到
GPT-4o[J]. 思想教育研究,2024(12):52-58.

[4] 刘润. 高校思政大模型:应用图景、逻辑向度与实
践进路[J]. 思想教育研究,2025(11):109-115.

[5] 陈浩洸,陈罕之,韩凯峰,等. 垂直领域大模型的
定制化:理论基础与关键技术[J]. 数据采集与处
理,2024,39(3):524-546.

[6] 籍欣萌,咎红英,崔婷婷,等. 大模型在垂直领域
应用的现状与挑战[J]. 计算机工程与应用,
2025,61(12):1-11.

[7] 马克思,恩格斯. 马克思恩格斯文集:第一卷
[M]. 中共中央马克思恩格斯列宁斯大林著作编
译局,编译. 北京:人民出版社,2009:331.

[8] 李游,梁哲浩,常亚平. 用户对人工智能产品的算
法厌恶研究述评及展望[J]. 管理学报,2022,19
(11):1725-1732.

[9] 侯勇,肖洋. 生成式人工智能赋能思想政治教育
学知识生产的向度、限度及进路[J]. 探索,2025
(2):141-152.

[10] 王金林. “生成式理性”:大语言模型引发的知识
生产范式转型——以生成式哲学模型为基础的
初探[J]. 社会科学,2025(8):184-192.

[11] 喻国明,曾嘉怡,黄沁雅. 提示工程师:生成式 AI
浪潮下传播生态变局的关键加速器[J]. 出版广
角,2023(11):26-31.

[12] 国务院关于深入实施“人工智能+”行动的意见
[EB/OL]. (2025-08-26) [2026-02-10]. [https://
www.gov.cn/zhengce/content/202508/content_7037861.htm](https://www.gov.cn/zhengce/content/202508/content_7037861.htm).

[13] 张国启,汪丹丹. 新时代我国主流意识形态图像
叙事的优化理路[J]. 华南理工大学学报(社会
科学版),2024,26(4):51-57.

[14] 邓水光. 大模型赋能哲学社会科学研究:范式与
路径初探[J]. 浙江社会科学,2025(9):4-13.

[15] 张国启,段港回. 流量文化的意识形态风险及其
治理[J]. 理论探索,2025(3):73-79.

[16] 习近平向联合国贸易和发展会议成立 60 周年庆
祝活动开幕式发表视频致辞[N]. 人民日报,
2024-06-13(01).

[17] 马克思,恩格斯. 马克思恩格斯全集:第三十七卷
[M]. 中共中央马克思恩格斯列宁斯大林著作编
译局,编译. 北京:人民出版社,2019:161.

[18] 王春丽,陈艳艳,顾小清,等. 协作学习中生成式
人工智能促进创造性潜能的机理研究[J]. 电化
教育研究,2024,45(11):76-83.

[19] 王正超. 数智时代政务大模型数据安全风险及治
理机制研究[J]. 中国科技论坛,2025(11):151-160.

From “Universal” to “Vertical” : Customized Application of AI Ideological and Political Model /

ZHANG Guoqi, DUAN Ganghui (School of Marxism, South China University of Technology)

Abstract: Speeding up the customized application of AI ideological and political model is an inevitable move to open up a new track for the development of ideological and political education and shape new advantages for the development of ideological and political education. As far as its connotation is concerned, the AI ideological and political model is a professional agent based on the general AI model, with strong relevant data or professional knowledge as training parameters, which can reflect the positions, views, and value attributes of ideological and political education. The AI ideological and political model is trained based on high-quality data or strong related professional knowledge, while the general AI model is only trained in ubiquitous data or general knowledge, which leads to a significant difference between the two. Compared with the general AI model, the performance of the AI ideological and political model is more professional, accurate, reliable, and safe. In terms of the choice of customization path, the vertical domain enhancement scheme relying on the full architecture general AI model has the advantages of a low customization threshold and less computational power consumption, which is more in line with the practical requirements of the intelligent development of ideological and political education. The scheme mainly enhances its practical performance in the ideological and political field by extracting the existing ideological and political education knowledge from the general AI model, inputting new ideological and political education knowledge into the general AI model, and integrating expert evaluation with ideological and political education. In the future, the construction of an AI ideological and political model should focus on the whole process system of “development, deployment, application, and feedback”, and enhance the “efficiency” of AI ideological and political model development by focusing on capacity building. Additionally, it should highlight the “utility” of its deployment by clarifying the application scenarios, standardize the “validity” of its application through anchoring tool orientation, and optimize the feedback “effect” by focusing on risk prevention and control.

Key words: digital intelligence technology; Artificial Intelligence; ideological and political education; large language model; digital intelligence transformation; human-computer interaction