

# 打造教育人工智能大脑:教育数据中台技术实现路径

李爱霞<sup>1</sup> 舒杭<sup>2</sup> 顾小清<sup>1</sup>

(1. 华东师范大学 教育信息技术学系, 上海 200062; 2. 江南大学 教育信息化研究中心, 江苏无锡 215000)

**[摘要]** 在智能时代,基于教育数据打造驱动教育不断发展的“教育人工智能大脑”,是一个前沿而紧迫的问题。数据中台的出现为实现教育数据的快速流转与价值发掘,构建最强的教育大脑带来可能。本研究首先分析了教育数据中台之于“教育是什么”“为什么”“将来会怎样”以及“如何应对”等教育应用的前景及 OneData 的核心运行机制;然后从教育数据采集层、数据存储层、数据开发层、数据服务层和数据治理层等方面剖析教育数据中台的技术架构;最后结合当前教育数据中台建设的基本情况,从顶层数据意识引导、基础设施建设和智能服务生态三方面提出教育数据中台的实现路径,以期为教育数据中台的落地做铺设。

**[关键词]** 教育人工智能大脑;数据中台;智能化;大数据

**[中图分类号]** G434

**[文献标识码]** A

**[文章编号]** 1007-2179(2021)03-0096-08

## 一、问题提出

在智能时代,融合 AI、大数据技术构建“教育大脑”可实现对教育系统的全局性即时分析,促进教育数据资源、教学资源、人力资源等有效调配,进而达到精准的、个性化的可持续性教育价值挖掘与开发,是实现智慧教育体系的关键。近几年,国家强调教育智能化建设(国务院,2017),大力推进智能教育,推动人工智能在教学、管理等方面的全流程应用,利用智能技术加快推动人才培养模式、教学改革,通过人工智能等技术为教育赋能的诉求愈发紧迫。如何将智能技术与教育融合,顾小清等(2021)提出人工智能大脑的隐喻,指出人工智能大脑为教育数据治理与教学创新提供了方向、过程以

及结果,进而形塑智能化、精准化、適切化的教育生态系统。

但怎样构建教育大脑?数据中台的出现及其在数据融合、处理、分析和管理工作方面的优势,特别是在教育全息画像、教育诊断、教育预测、教育干预等方面的技术突破,使构建“教育人工智能大脑”成为可能。

总体而言,数据中台是一种组织战略,连接前后台,使前台能够快速响应业务变化,即能够有效赋能前台的公民数据用户,再利用后台的数据进行决策。高德纳(Gartner,2016)的分层应用策略报告,将网络业务系统分为前台、中台、后台,中台的核心作用是提升业务系统响应能力与速度。付登坡等(2020)认为数据中台的本质是“数据仓库+数据服务中间

**[收稿日期]**2021-01-26

**[修回日期]**2021-05-05

**[DOI 编码]**10.13966/j.cnki.kfjyyj.2021.03.010

**[基金项目]**2019 年国家社科重大项目“人工智能促进未来教育发展研究”(19ZDA364);2019 年华东师范大学优秀博士生学术创新能力提升计划“学习分析视域下学习者在线学习粘性测度及干预机制研究”(YBNLTS2019-040)。

**[作者简介]**李爱霞,博士研究生,华东师范大学教育信息技术学系,研究方向:学习分析、学习设计研究(943084713@qq.com);舒杭,博士,讲师,江南大学教育信息化研究中心,研究方向:学习分析;顾小清,教授,华东师范大学教育信息技术学系,研究方向:教育大数据与学习分析、技术支持的教学创新。

**[引用信息]**李爱霞,舒杭,顾小清(2021).打造教育人工智能大脑:教育数据中台技术实现路径[J].开放教育研究,27(3):96-103.

件”。数据中台通过数据技术,对海量数据进行采集、计算、存储、加工,同时统一标准和口径(项阳,2020),生成数据资产及数据服务。数据中台的核心是可持续地“让数据用起来”,使得数据来源于业务,反哺于业务,不断循环迭代,以促进数据的可见、可用和可运营。数据流转不仅能降低重复建设、减少“烟囱式”协作的成本,还能实现系统发展的可持续优化迭代(卜意磊等,2020)。

教育数据中台是数据中台在教育中的应用,是为了满足教育教学的特定化需求,对教育数据进行融合、处理,使之成为可理解、可使用、可管理的数据资产,并将教育数据服务于教育体系(见图1)。

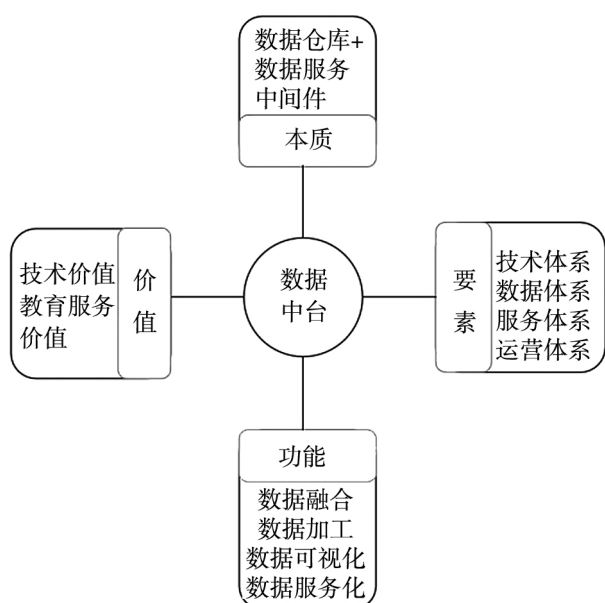


图1 教育数据中台概念解构

教育数据中台的本质是“数据仓库+数据服务中间件”。这也是数据中台与数据仓库的区别。数据中台不仅存储数据,还将数据服务于教育(付登波等,2020)。比如,它通过分析学生行为数据支持学生画像刻画,还可以根据教育需求提供针对性教育报告。教育数据中台不是单纯的技术叠加,不是技术化的大数据平台。大数据平台关心技术层面,比如研发效率、大数据处理等,针对的是技术人员,而数据中台的核心是数据服务能力,结合实时的教育需求,通过数据建模赋能教育教学。数据中台也是个不断更新的体系,在教育服务过程中,持续迭代技术、数据建模。

数据中台由四部分组成:技术体系、数据体系、服务体系和运营体系。技术体系主要实现大数据的存储、处理、管理与应用,以及支持中台的构建;数据体系是实现数据资产化的核心,使离散的数据成为可用的服务型数据;服务体系是实现数据到产品落脚的关键,通过数据的可视化,实现用户的画像刻画、管理、评价等,使数据与业务匹配;运营体系根据数据及用户需求,实现产品的更新迭代及创建。

数据中台的功能包括数据融合、数据加工、数据可视化、数据服务化。其中,数据融合回应教育数据孤岛问题,数据中台提供统一、适配的一站式数据收集标准与方法,实现教育数据的收集与转换;数据加工功能应对教育数据资产化,通过数据处理,打通教学、学习与学校管理等的全域数据流,以统一的数据标准和质量体系服务未来的教育;数据中台智能化的数据管理方法应对全域数据可视化展示,为教育数据使用者提供可视化数据图谱;数据业务化应对数据流转问题,能运用机器学习、自然语言等人工智能方法将教育数据应用于具体教育问题的解决,比如学习监督与预测、学习画像分析等。

总之,教育数据中台是将教育数据转变为教育生产力的机制。

## 二、数据中台开辟教育数据价值挖掘新体系

### (一) 可能应用

教育数据中台的价值主要体现在两个方面:技术价值和数据服务价值。技术方面,教育数据中台融合 Hadoop、MySQL、Oracle、Spark 等技术架构,满足教育场景对多数据并发处理的高量级数据处理需求,为教育数据的跨主题域访问、量级数据标签化、数据多线程并发处理等提供支持;教育服务方面,教育数据中台以个性化学习为理念,通过对学生、教师等行为数据的全视域采集融合、处理分析,为学校教与学提供精准的决策支持,最终实现从“教育是什么”“为什么”“未来怎样”以及“怎么应对”的教育全域问题的精准解析,即教育全息画像、教育诊断、教育预测、教育干预等技术突破,构建“教育人工智能大脑”。

#### 1. 教育数据中台明晰“教育发生”的真谛

教育数据中台以教师、学生行为数据为来源,通过成熟的画像技术实现教师、学生的多模态画像智

能生成,可视化呈现教师教育成果、教育过程及学生学习效果、学习过程等,告诉人们教育中发生了什么。其中,教师画像包括教师特征模型、教研心理模型、社会交互模型、教研行为模型和教研成果模型(胡小勇等 2019)等,学生画像主要有学生的认知、非认知、学习习惯等全息画像。以上画像是实现精准化教学素养提升以及个性化助学的基础,也是智能教育决策的前提,推动了教育由被动向主动转变。

## 2. 教育数据中台支持教育诊断

教育数据中台的另一应用是实现智能的教育诊断:通过访问智能硬件设备,管理具有集成服务的教育平台及设备并从中获取教育数据;使用数据规则对教师教学、学生学习等活动进行及时的数据转换,形成实时的智能洞察报告,进而诊断教育过程问题。比如,汇聚了学生学习知识图谱数据后,数据中台可跟踪及可视化报告学生的知识完整度与不足,为教师全方位、立体地了解学生学习效果及干预提供参考。

## 3. 教育数据中台助推教育数据挖掘

教育中台的更大价值在于对教师和学习者模型的挖掘和预测。教育数据挖掘研究的重点是建立从学生数据中提取隐藏知识的模型,从而提高学生学习成绩。利用教育数据挖掘技术,可将教育系统的原始数据转化为有价值的信息,供教师、学生、家长、教育研究者、教育软件开发人员等使用。其教育应用主要有聚类、预测、分类等。例如,数据中台可通过机器学习模型“学习”每个学生,找出他们的缺点,并确定改进的方法,如学习更多的课程或练习,也可以用来预测某一课程的入学人数(Yadav et al. 2012),预测传统课堂教学模式的异化(Akinola et al. 2012),检测在线考试中使用的不公平手段,检测学生成绩记录的异常值,预测学生成绩等。简言之,数据中台的数据挖掘和预测功能为教育提供了预测未来的“法眼”,是教育者有效教学的导航灯。

## 4. 教育数据中台助力教育决策

解决教育问题,完全依赖经验的模式已被淘汰。如何对教育进行智能的、个性化的决策支持是当前教育面临的重要课题。教育数据中台在基于全面的教育服务获取丰富的教师及学生数据,以及对他们全面画像的基础上,总结教育规律,实现教育资源的

循环丰富完善、教育推荐路线的精准匹配等,为教育管理者、教师以及学习者提供精准的教育服务。具体而言,数据中台可以根据学生画像提供个性化学习资源推荐,如根据学生的认知特征、学习风格等进行准确定位,分析其潜在的学习需求,进而提供有针对性的学习资源、学习服务支持。例如,教育数据中台能够通过 Apriori、DEA-BP 等算法智能地从教学评价数据中提取潜在的规律和知识,为教学评价决策提供支持(Ma et al., 2021);还可集成神经网络、专家系统、遗传算法等模型,实现对教育质量和教育投入的监测,实现教育物质资源、人力资源等的优化配置及管理,提高教育服务质量。

## (二) 运行机制

数据中台打破传统的“烟囱”式的产品应用管理模式,疏通数据流转通道。数据中台拥有提供适配数据采集、转换的完整的大数据软硬件工具,能够实时、大批量地实现数据的采集和交换,且能够根据业务层次需求,部署数据采集的来源与类型,从而协助定位、理解数据,根据统一的标准工具与方法实现数据的标准化转换。这与当前教育急需解决的数据零散、无法统一的需求紧密相关。

### 1. OneData 数据资产化与服务化

建设数据中台的最终目的是让“数据用起来”,最终实现产品的创新,更新迭代。系统的技术、理念及方法都是可复制的。数据中台的根本创新是将数据资产化,然后将资产化的数据作为生产资料应用于业务价值的创造,持续产生价值。数据中台不仅仅是技术,更是一种从“技术优先”到“数据优先”的思维转变,目标是让数据持续用起来,通过数据中台提供的工具、方法和运行机制,把数据变为服务能力,让数据更方便地被业务所使用(付登波等 2020)。数据中台遵循 OneData 的核心方法(见图 2),通过统一的数据标准收集、处理数据,及对数据进行清洗、加工,使零散的数据变成可以用于教育管理、教学及学习场景的可应用数据,即数据资产。

### 2. OneModel 和 OneID 实现数据的双打通

1) OneModel 打通跨平台数据融合通道。中台是统一数据格式、数据接口打通平台间数据融合的通道。该方法实现数据的资产化构建与管理,通过对数据的标准定义、数据质量与安全控制将数据收

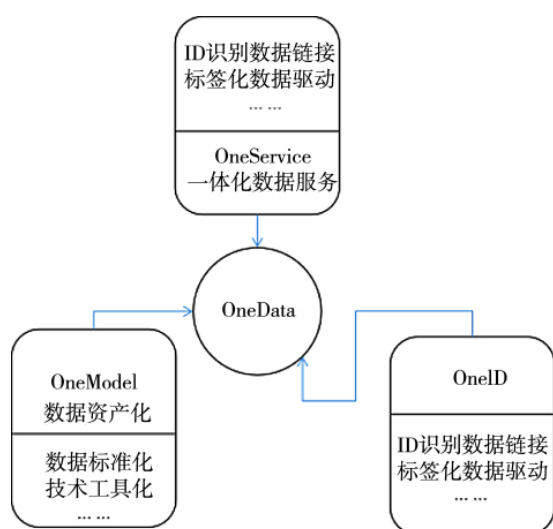


图2 数据中台 OneData 核心方法

集处理等技术形成完整的体系,进而对大批量数据进行智能化建模、梳理。

2) OneID 打通跨平台的个人数据融合衔接通道。OneID 的个人数据收集方法,使得不同平台间的个人数据无缝衔接,可加速个人数据标签化处理,实现全方位的个人画像刻画。两种方式的结合可消除数据孤岛,驱动数据价值化,更可实现智能化的数据管理。这与当前数据流通的需求不谋而合。

3) OneService 一体化数据服务。中台针对产品

的一体化服务体系,为数据到业务的落地提供技术保障。在数据资产化处理,伴随而来的是如何将数据应用到实际教育场景,中台的后续系统能够通过 ID 实现多维数据链接,并根据标签化的数据驱动业务提升,其中包括服务质量、产品升级等,智能化地根据教学需求优化教育环境、教学策略、教育管理等。

### 三、教育数据中台的技术架构

教育数据中台作为促使教育数据用起来机制,能实现“教育服务数据化”到“数据服务教育化”的循环。为了实现教育数据的大规模、高效率处理需求,数据中台拥有 PB 级大规模数据管理能力,支持穿透数据库、Hadoop、大规模 MPP 集群,实现 PB 级结构化数据、半结构化及非结构化数据的多样化海量数据的统一存储、管理和分析。如何保证数据中台持续、稳定地运行,强大的技术架构是基础,其中不仅包括基础硬件设备技术,还包括数据采集、存储、转换及开发的软技术,以及数据管理的理论技术。

教育大数据的数据采集、数据处理、数据分析与应用服务流程与环节,融合以教育数据中台不仅能够支撑以上功能,并且设有数据治理层,支持教育数据的循环、序列及融合分析需求。本文将分别从教育数据采集层、数据存储层、数据开发层、数据服务层和数据治理层,分析教育数据中台的技术架构(见图3)。

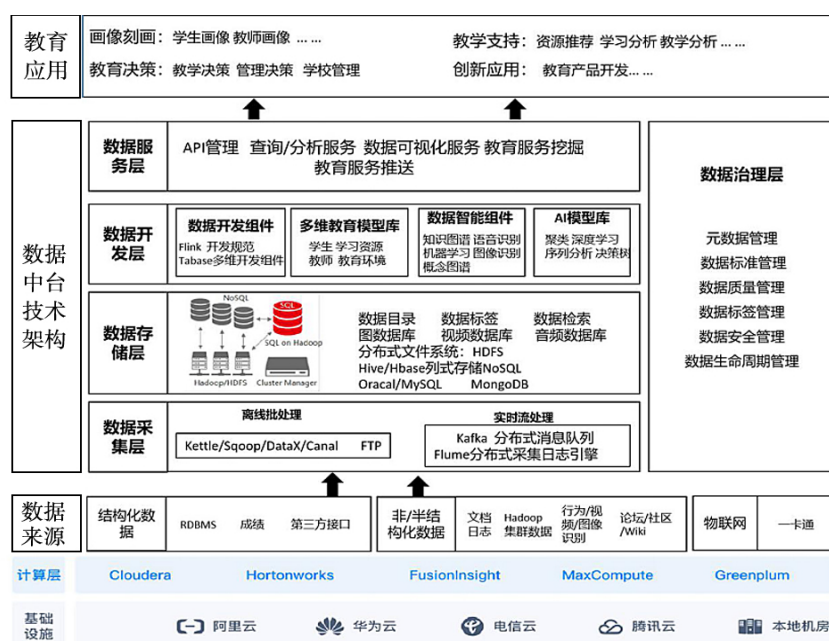


图3 教育数据中台技术架构

### 1. 数据采集层:全方位实时数据获取

数据融合指数据中台按照一定的规范,收集、关联、整合不同的教育数据类型,为后续的数据资产化做准备。教育数据来源的多样性、数据类型的复杂性是教育数据融合的难点。当前教育数据来源主要有线上数据、线下数据和物联感应数据,每种来源对应的数据类型也不同:按教育场景分有课堂师生行为数据、户外学习数据、网络社交活动数据、成长经历数据;根据数据类型分有结构化数据(如学生成绩等数据)、半/非结构化数据(如视频、音频、行为序列、文档日志、文本等数据)、物联感应数据(如一卡通或生理感应数据等)。针对多样化的数据来源,数据中台具有适配性的数据采集及融合手段;从时效看,主要包括离线批处理和实时流处理数据;从数据类型看,有结构化数据和非结构化数据。

离线批处理一般用于对低时效的海量教育数据的周期性迁移,实现数据的全量或增量式数据存储。离线批处理数据收集技术主要有开源工具(主要有 Kettle、Sqoop、DataX)和服务器协议(FTP)两种。Sqoop 是解决结构化数据和分布式文件系统(Hadoop Distributed File System, HDFS)之间数据转移的软件,可将关系型数据库(如 MySQL、Oracle、Postgres 等)与 Hadoop 的 HDFS 的数据互相传出。Kettle 是开源的数据抽取-转换-存储工具(Extract-Transform-Load, ETL),适用于处理轻量级数据,对大规模数据的清理则可能会因数据量大和清洗逻辑复杂导致数据传输效率打折扣。另外,该模式还可能清洗掉未被开发的数据。DataX 是对结构化数据进行插件式离线交换的工具,特别是异构数据源的转换、流量转换、进度监察等。

实时流处理主要用于采集 APP、服务器日志、小程序、各种 API 接口及数据文件等实时数据。其中,记录师生教与学的行为日志数据居多,且数据结构多样,来源环境复杂。针对这类数据,研发人员常用实时日志采集引擎 Flume,这是一款由 Cloudera 开发的,主打高并发、高速度、分布式海量日志采集的技术。Flume 支持在日志系统中定制各类数据并发送,用于采集数据,且支持对数据的简单处理,并写到各种数据接收方。Wiki、论坛等产生的海量消息类数据,可由分布式消息队列技术 Kafka 收集和转

换。Kafka 为实时数据提供统一、高吞吐、低延迟的平台。

综上,教育数据中台既满足异构存储、异构数据类型的交换需要,还满足不同时效数据的互通,但没有哪种工具能兼顾所有需求。对此,数据中台采用内置离线和实时数据同步的策略,满足复杂数据的融合。

### 2. 数据存储层:海量数据的“仓库”

教育数据汇集后的海量级、多类型数据对传统的数据存储方式带来巨大冲击。以往教育中以关系型数据库为主要数据存储方式,显然不能满足大数据环境下的多类型、海量数据存储以及复杂的数据挖掘和分析操作需求。对此,数据中台具备的融合性数据存储架构,对大数据的存储方式拥有比较成熟的技术支撑。它能针对不同的数据来源、数据类型及未来的数据应用场景,以分而治之的策略适配不同的数据库和数据存储技术。为了后续数据开发的便捷,教育数据中台按不同的数据类型将数据库分为数据目录、数据标签、数据检索、图数据库、视频数据库和音频数据库。对应的存储方式有分布式文件系统(HDFS)、非关系型数据库 NoSQL、关系型数据库(见图4)。

分布式文件系统 HDFS 是以 Hadoop 为基础的对超大集高吞吐量数据的存储,容错性高,可为教育中产生的日志、会话、知识结构以及物联环境感知等高并发数据提供实时存储,还可为大数据融合提供底层的数据存储能力支撑。非关系型数据库 NoSQL 可满足图、文档等超大规模非关系型数据的存储需求,具有易扩展、大量级、高性能等特点,主要技术包括 MongoDB、HBase、Hive 等。关系型数据库是成绩、知识点等结构化数据的主要存储方式,主要以行和列的形式存储数据,具有规范化的数据格式,能充分节约数据存储空间,还便于用户对数据的理解和检索,主要技术包括 Oracle、MySQL 等。

总之,三类数据存储方式所支持的数据类型与来源不同,但是它们之间以 OneID 为主要的融合桥梁支撑数据开发和挖掘。

### 3. 数据开发层:教育数据价值提炼工厂

数据开发是将教育融合的原始数据资产化的转化工厂。数据开发是一套包含数据加工算法和过程管理的工具(付登波等,2020),根据数据的时效性

与类型,分配不同的数据处理方法。比如,通过离线开发组件计算分析一定规模的数据,实现非实时的、批量教育数据的挖掘;通过实时开发对实时流数据进行“跟踪式”处理,挖掘教育数据价值。另外,数据中台的数据开发系统还可用于开发内部算法与教育模型,以满足不断变化的教育数据服务需求。综上,教育数据中台开发层主要由数据开发组件、多维教育模型库、数据智能组件和 AI 模型库四部分组成。

数据开发组件的设置一般根据数据来源不同而有别,实时产生的流数据和历史数据,分别由 Flink 和 Tabase 多维开发组件与之对应。其中,Flink 作为第四代数据计算引擎的代表,是可扩展的批处理和流式数据处理平台,可实现数据的高效率开发。Flink 的数据流 API 还支持有界或无界数据流转换,支持 20 多种不同类型的转换和流计算。Tabase 多维开发组件是对历史数据的处理技术的组合,主要有 MapReduce、Spark、Redis、Impala 等。安通过多种技术的融合完成教育数据的批量计算、查询以及交叉分析。多维教育模型库设置为数据开发指明了方向。按照分析对象的不同,主要教育模型有学生、教师、学习资源、教育环境四类。教育数据中台根据模型库的指标对数据进行处理、挖掘,进而生成教育服务模型。算法是数据挖掘的推动力,其在教育中已有比较成熟的应用,比如根据算法功能分为语音识别、图像识别、聚类、序列分析、知识追踪等;根据支持数据量级有深度学习、机器学习等。另外,教育数据中台的算法框架不仅支持数据的挖掘,还支持算法本身的开发,以满足更复杂数据的分析。

#### 4. 数据服务层:链接教育服务与数据的桥梁

数据服务体系是实现教育数据与教育服务对接的关键,包括 API 管理、查询/分析服务、数据可视化服务、教育服务挖掘、教育服务推送。

API 管理功能主要是对众多数据来源、存储、处理等软件数据接口的管理,通过对各系统数据接口的统一处理,为数据安全、数据调配做导航。数据的查询/分析服务以及数据的可视化模块是利用检索与可视化的方式向开发者、管理者展示数据结构,便于他们监管与运用数据。

教育服务推送是为学生、教师及管理者推荐教

育服务功能的技术,让他们更好地了解、运用信息化教育系统的功能,赋能教育教学,形成教育服务生态圈。教育服务挖掘是建设教育服务生态的关键之一,也是连接教育与数据应用的核心。在大量数据与技术的融合下,数据中台深入挖掘教育规律,描绘教育现象,以响应多样化的数据服务需求。

#### 5. 数据治理层:为教育数据稳定运转护航

随着数据的积累与开发,数据逐渐成为教育资产的一部分。教育数据中台对教育元数据、数据标准、数据标签、数据质量、数据安全以及数据生命周期进行管理。其中,元数据管理主要包括对数据元的配置、数据模型管理以及元数据属性管理;数据标签模块是对数据资产的贴源标签、数据标签进行监管;数据质量管理主要是对数据质量的校验与管理;数据安全模块是监测全局数据的隐私、共享异常,实现对全局数据的安全保障;数据生命周期是监督系统,对有意义数据持续保存,将无意义数据缓释的功能模块。

总而言之,数据治理是通过可视化监控元数据、数据质量、数据周期、数据标准和数据周期等,实时把控学校教育数据资产质量,为教育数据中台的稳定运转护航。

## 四、实现路径

通过数据中台构建教育数据大脑,对助力未来智能化教育发展具有现实意义,故促进教育数据中台落地,是建设教育大脑的核心工程。教育数据中台的建设主要包括教育数据意识的启发、基础设施及标准的建设、教育服务生态的驱动三个环节。其中,数据意识的顶层引领为教育数据中台建设指明方向,基础设施建设与规范为教育数据中台铺设基石,教育服务生态为教育数据中台的驱动提供动力和原料。

#### (一) 强化数据意识的顶层引领

数据作为教育数据中台的“血液”,为智能化教育发展提供“燃料”。掌握、运用有效的教育数据,是学校优化教育发展的必备技能。教育数据化的实现要以顶层数据意识引领,实现从“经验”为主导到靠“数据驱动”的教育决策理念的转变。引领顶层数据意识主要回答“是什么”“为什么”“怎么用”三个问题,即提升学生、教师及管理者懂数据、用数据



的意识与能力。

首先,注重教育大脑、大数据等政策、价值的宣传,提升用数据服务教育的意识,养成教育数据价值意识。另外,以经典的应用案例或以教育大脑为核心的数据赋能教育服务体系的构建,让师生、管理体会大数据带来的技术红利。

其次,坚持以用数据为中心,切实推动教育数据融合发展。培养教师、管理者用数据助力教学的意识和能力,让他们在理解数据价值的基础上善于根据自身诉求获取、分析、运用、管理教育数据,不断提升利用数据推进工作的意识;鼓励教育数据共享共用,推动公共数据资源汇聚融合。

## (二) 注重基础设施建设与规范

大量级数据的采集、存储与处理是数据中台的基本能力,支撑教育数据中台功能发挥的根本是扎实的基础设施架构。基础设施建设主要包括制定数据标准、建设硬件设备以及搭建云平台等。

数据规范与标准是融合教育数据的前提。这要求:统一基础平台标准和数据元标准,实现对基础平台数据的深度融合挖掘,以保证个性化教育数据服务需求;执行统一的数据共享标准,保障数据采集、整合、共享协议一致,实现数据互联互通和开放共享;执行统一的管理标准,保障数据管理规范、安全可控。

硬件设备主要指加强学习、教学、管理、校园建设等基础数据库和网络等基础硬件设施的建设。高性能、大容量数据存储设备,以及高速的网络带宽是保证数据采集、上传时效的基础,也是保证数据高效运转的“底座”。

考虑到建设量级数据处理的时效性以及硬件设备建设成本,依托云平台建设数据中台服务体系是最佳选择。云平台综合硬件资源和软件资源的服务,为数据中台的计算提供有力保障,可促进各系统数据资源交换共享。

## (三) 构建智能化数据服务生态

数据中台的最终功能是实现教育数据的价值挖掘,但是挖掘教育数据价值的目的、动力及“原料”来源是教育教学的服务需求,即多样化的教育服务功能为数据中台输送全面、立体化的教育数据,反过来,教育数据中台助力教育服务功能的完善。因此,开发全方位的教学、学习服务功能,打造数据驱动的

智能教育服务生态圈,形成数据运转闭环,是推动教育数据中台深入发展的助燃剂。

当前研究较多的教育服务包括画像刻画、学习分析、学习环境设计等,距离全息的智能化教育体系蓝图还很远,故全面的教育服务生态开发是必要的。对用户需求开展调研是产品设计比较常见的方式,然而该方式中用户顶层设计能力的缺失是弊端,特别是在智能化时代,人们对教学、环境功能更迭效率需求无法用人工调研方式来满足。因此,构建数据智能和场景驱动的教育数据服务体系,通过大数据智能分析帮助开发者深挖学生、教师、教育环境等的深层特征,即以数据化的“望、闻、问、切”,实现教育数据中台精准化的教育数据采集、教育服务支持。智能化数据服务生态构建主要从三个方面着手:1) 加强教师平台、学生平台、校园管理、校园学习空间等智能化、数字化服务教育的软硬空间建设,促进教育空间服务数据化;2) 提升教师、管理者数据服务教育的意识,养成懂数据、用数据的教育服务理念,实现教育教学、管理服务数据化;3) 关注智能技术数据驱动的教育服务挖掘模型构建,运用技术与数据相结合的手段,实现教育服务的自动化产出。

总之,以教育数据为基础构建教育大脑,赋予智慧教育可能是当下的重点工作。但目前我国的教育数据中台建设还处于起步阶段,不能一蹴而就,数据积累、设施配备、环境建设等都是今后教育数据中台建设的重要方向。具体而言,教育数据中台的建设应注意:1) 教育数据中台不是千篇一律的,教育机构需根据自身需要定制中台规模;2) 提高数据意识,尤其是要重视小数据,贴合教育中“小数据带动大数据”的现实问题,使教育数据中台深入教育规律和本质;3) 重视数据中台的价值挖掘,完善教育理论模型和数据模型的融合,开发更完善、精确的教育模型,并应用于实践。基于以上分析,从技术层面加强教育建模与表征的技术开发体系以及智能的多模态教育数据收集、融合与分析算法;从理论层面完善教育模型框架,挖掘行为数据与真实教育意义的关系,这些是未来教育数据中台落地的关键,也是完成教育人工智能大脑的必经之路。

简言之,教育数据中台的发展和构建对未来智能教育的发展具有重要意义,除了基础设施、顶层意识、技术规范等基本支撑,实现教育理论层面和数据

驱动的教育分析的融合、迭代和落地,也是未来教育数据中台建设的重点。

#### 【参考文献】

- [1] Akinola, O. S., Akinkunmi, B. O., & Alo, T. S. (2012). A data mining model for predicting computer programming proficiency of computer science undergraduate students [J].
- [2] Gartner (2020). Pace-layered application strategy [EB/OL]. <https://www.gartner.com/en/information-technology/glossary/pace-layered-application-strategy>.
- [3] Ma, J. (2021). Intelligent decision system of higher educational resource data under artificial intelligence technology [J]. International Journal of Emerging Technologies in Learning, 15(5). e.
- [4] Yadav, S. K., Bharadwaj, B., & Pal, S. (2012). Mining education data to predict student's retention: A comparative study [J]. arXiv preprint arXiv:1203.2987.
- [5] 卜意磊, 庞文迪 (2020). 基于数据中台的市场监管数据资源中心架构设计 [J]. 电子技术与软件工程, (9): 155-159.
- [6] 付登波, 江敏, 任寅姿, 孙少忆, 武凯, 沈金, 蒋珍波 (2020). 数据中台: 让数据用起来 [M]. 北京: 机械工业出版社: 20-21.
- [7] 顾小清, 李世瑾 (2021). 人工智能教育大脑: 以数据驱动教育治理与教学创新的技术框架 [J]. 中国电化教育(1), 80-88.
- [8] 国务院 (2017). 新一代人工智能发展规划 [EB/OL]. [http://www.gov.cn/zhengce/content/2017-07/20/content\\_5211996.htm](http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm).
- [9] 胡小勇, 林梓柔 (2019). 精准教研视域下的教师画像研究 [J]. 电化教育研究, 40(7): 84-91.
- [10] 项阳 (2020). 数据中台元年 [J]. 中国教育网络, No. 251(1): 26-27.

(编辑: 徐辉富)

## Building an Educational Artificial Intelligence Brain: the Technical Architecture and Implementation Path of the Educational Data Middle Platform

LI Aixia<sup>1</sup>, SHU Hang<sup>2</sup> & GU Xiaoqing<sup>1</sup>

- ( 1. Department of Education Information Technology, East China Normal University, Shanghai 200042, China;  
2. Research Center of Educational Informatization, Jiangnan University, Wuxi 215000, China)

**Abstract:** There is an advanced and urgent problem in the age of intelligence to build an “educational artificial intelligence brain” that provides data-driven development. However, the emergence of the Data Middle Platform makes it possible to realize the rapid flow and value mining of educational data and build the strongest educational brain. This research first analyzes the educational application prospects of the educational Data Middle Platform in terms of “what is education,” “why,” “what will happen in the future,” and “how to deal with” and its OneData core operating mechanism; Then, the technical framework of the middle platform of education data is analyzed from the aspects of education data acquisition layer, data storage layer, data development layer, data service layer, and data governance layer; Finally, the three aspects of education data asset awareness, education data platform infrastructure construction and standardization, and the cultivation of comprehensive quality talents provide strategies for the application of data middle platform in education, and lay the foundation for the education data middle platform.

**Key words:** educational artificial intelligence brain; data middle platform; intelligent education; big data