

数据智能技术普惠化的理论与实践

徐 明

摘 要：数据智能是人工智能技术的一个重要分支，也是我国可以形成相对竞争优势的技术领域。数据作为新型生产要素，是企业能否开展数字赋能和数字化转型的关键。研发面向平民和消费者的数据智能平台工具，是实现 AI 技术民主化，消除大数据带来的数字新鸿沟的有效途径，也是实现 AI 普惠化和自动化的一个选择。随着 ChatGPT 推动的消费者人工智能阶段的到来，企业应主动拥抱低代码 DSML 和 ChatGPT 等新一代人工智能技术，加速实现企业内的 AI 技术民主化，推进 AI 产品的部署和应用，实现企业的高质量发展和数字化转型。

关键词：人工智能；数据智能；数字化转型；AI 技术民主化

中图分类号：TP18 **文献标识码：**A **文章编号：**1673-5706 (2023) 03-0067-06

人工智能的发展和应用展现了一个新模式：感—知—用。感，即基于泛在感知设备或泛在网络收集数据；知，即利用算法对数据进行分析，解析其中的结构、模式和规律；用，即将算法分析结果与背景知识结合，形成解决方案。这种模式的核心是数据。徐宗本院士将这种基于数据形成智能决策解决问题的人工智能技术称为数据智能^[1]。随着大数据和人工智能技术的飞速发展，数据智能技术正在工业应用和科学研究领域发挥着日益重要的作用。然而，数据智能技术的应用仍面临着应用门槛过高的挑战。因此，寻找可行的数据智能普惠化路径，推动实现 AI 技术民主化，消除大数据带来的数字新鸿沟，对于提倡高质量发展和数字化转型的今天，显得尤其重要。

从产业界的角度，中国领先的数字化市场研究咨询机构爱分析将数据智能定义为“以数据为

生产要素，通过融合大规模数据处理、数据分析与挖掘、机器学习、可视化等多种大数据和人工智能技术，从数据中提炼、发掘具有揭示性和可操作性的信息，从而为企业提供数据驱动的分析与决策”^[2]。数据智能是人工智能的重要分支，它以数据为基础，将机器学习、深度学习等技术与领域知识相结合，通过对大量的数据进行处理和分析，从中发现有价值的信息和知识，支持决策和创新。通过数据智能技术，企业可以更好了解自身业务和市场环境，提高决策效率和准确度，实现降本增效，推动业务创新，增强竞争力。目前，越来越多的企业开始认识到数据智能技术的重要性，并将其作为推动企业数字化转型和发展的重要工具。

一、数据智能的落地应用与挑战

（一）数据智能的成功应用

数据智能技术使我们能够更好地利用数据，

实现更高效、更准确的预测和决策。在工业领域，数据智能技术可以帮助优化生产流程、降低成本、提高产品质量；在科学领域，数据智能技术可以帮助科学家更好地理解自然现象、发现新的规律和现象。

1. 数据智能在工业工程中的应用

在工业生产中，机械加工的核心部件是用来切削工件的刀具部件，其对加工工件的质量有最直接的影响。然而，在加工过程中，刀具会随着切削工件而逐渐磨损，导致加工效率和质量的下降。因此，切削刀具的衰退和寿命剩余量的预测对于工作量的影响至关重要。传统的方法是依靠操作人员的经验以及观察切削声等因素进行预测。这种方法容易受人为因素的干扰，且难以实现实时、准确的刀具健康信息评估。利用数据智能技术，以数据为基础，通过机器学习模型即可实现实时、准确的刀具健康信息评估，从而降低工作量与废弃量，刀具和其他部件的健康状况通过可视化来呈现^[3]。

2. 数据智能在科学研究中的应用

随着生物医学和信息技术的快速融合发展，健康医疗领域积累了海量的影像数据、患者报告数据、电子健康记录和数据等，这些数据具有复杂性、异构性和高维等特点，而深度学习有着复杂函数模拟和自动学习特征的能力，能够从复杂的数据中较为精准地提取有效的信息，为医学诊断、药物研发等方面的研究提供高效的技术支撑。确定生物分子的3D形状是现代生物学和医学发现中最困难的问题之一。许多公司和研究机构花费数百万美元来确定分子结构，却常常无果。2021年，美国斯坦福大学博士生Stephan Eismann和Raphael Townshend在该大学计算机科学副教授Ron O. Dror的指导下，利用机器学习技术，开发出一种通过计算预测准确结构来克服这一问题的的人工智能算法。通过几何深度学习，预测出RAN分子的3D结构，助力RNA靶向药研发^[4]。

（二）数据智能的落地挑战

1. 领域专家与科研人员应用数据智能技术的挑战

数据智能涵盖机器学习、深度学习、大数据

等技术，对于领域专家和科研工作者来说，现有的平台工具，需要他们花费很多时间理解并编程实现，导致数据智能落地应用效率低，见效慢。以机器学习为例，一方面，传统机器学习技术门槛高，对人才的要求较高，要成为一名合格的机器学习从业人员，需要掌握算法、模型、数据等多方面的知识。此外，还需要熟练掌握编程语言和数据结构，了解概率论、线性代数、微积分等数学基础知识。另一方面，在传统的机器学习开发中，科研人员需要花费大量的时间写代码，他们需要手动进行模型选择、特征工程、模型训练、模型调优等繁琐的步骤。这些步骤需要依靠编程实现，且需要不断尝试和调整，以获得最好的结果。这个过程需要消耗大量的时间和精力，使得科研人员无法专注于研究本身，影响了科研人员的工作效率和创造力。

2. 数据智能在工业工程领域落地的挑战

斯坦福大学人工智能实验室主任吴恩达曾在IEEE Spectrum《ANDREW NG:UNBIGGEN AI》中提到，“在消费类软件互联网中，我们可以训练少数几种机器学习模型来服务10亿用户。但在制造业，你可能要为1万个制造商构建1万个定制化的人工智能模型”。这是因为制造业的需求和业务场景通常非常复杂和多样化，需要根据不同的制造商、产品和生产流程等因素来设计和构建不同的机器学习模型，以满足不同的需求和业务要求。另外，制造业的数据通常比消费类软件互联网中的数据更加复杂和多样化，包括结构化数据和非结构化数据、传感器数据和图像数据等。这些数据需要更多的预处理和清洗，才能用于机器学习模型的训练和预测。因此，制造业中的机器学习应用需要更加精细和复杂，需要更多的专业知识和经验，也需要更多的投入和支持。

二、数据智能技术的普惠化

推进数据智能技术的普惠化，才能更好地实现数据智能的落地应用，发挥数据作为新型生产要素的作用。吴恩达教授也给出了数据智能技术普惠化的一个思路：“摆脱这种困境的唯一方法是构建工具赋能客户，为他们提供工具来设计数据和表示领域知识，使他们能够构建自己的模

型。”^[5]这种解决方案的核心思想是，将机器学习技术封装成易于使用的工具，赋能给领域专家或普通用户，让他们能够自主地使用这些工具来设计数据和表示领域知识，并构建自己的机器学习模型。

这种方法的优点显而易见。首先，它可以降低机器学习模型的开发门槛，让更多的人能够参与到机器学习模型的开发中来；其次，它可以使机器学习模型更加符合实际需求，因为领域专家更了解自己的业务场景和问题，能够更好地设计数据和表示领域知识；再次，这种方法可以提高机器学习模型的可复用性和可扩展性，不同的客户可以使用相同的工具和方法来构建自己的模型，从而形成一个更加庞大和强大的生态系统。除了构建一个人机交互、更加友好的平台工具外，实现平台工具的零代码或低代码也是推进数据智能技术普惠化的必要条件。

（一）低代码数据科学与机器学习平台（DSML）

国际著名咨询机构 Gartner 给数据科学与机器学习平台下了一个定义：“DSML(Data Science and Machine Learning platform) 数据科学与机器学习平台是由一个核心产品及其所集成的多个数据科学相关产品、组件、框架、库所构成的系统。可以基于集成的多个产品、组件，提供一致的用户体验，让用户在一个平台可以无缝地完成数据读入、数据处理、数据建模和数据应用。”^[6]

中国领先的数字化市场研究咨询机构爱分析给出了这样的定义：“数据科学与机器学习平台是指覆盖数据采集、数据探索、数据处理、特征工程、模型构建、模型训练、模型部署与发布、模型管理与运营等建模全流程的平台，提供一站式建模服务，能显著提升建模效率、降低建模门槛。数据科学与机器学习平台能支持并赋能企业各业务场景实现智能决策，帮助企业打造数据驱动型组织。”^[7]

低代码 DSML 从数据接入、数据处理、数据建模到数据应用的全过程无需编写程序代码或仅仅有极少的环节需要编写程序代码。这样的平台工具，不仅可以让领域专家得心应手，而且可以

实现吴恩达教授期待的让数据智能更好地用于千厂千面的工业场景。

（二）低代码 DSML 的价值

大大降低技术门槛，实现 AI 普惠化。低代码 DSML 提供了可视化的用户界面，降低了技术门槛，让非专业人员也能进行数据分析和建模。用户无需编写代码，或仅需编写少量代码，即可通过简单的拖拽图标的方式实现数据预处理、特征工程、模型训练和评估等任务，无需进行复杂的操作。这样，企业内部不同岗位的人员都能够参与数据分析和建模工作。

知识工作自动化，实现人机协同。低代码 DSML 集成了各种数据科学和机器学习算法、模型和工具，可以将重复的、繁琐的工作自动化。例如，数据清洗、特征选择、模型选择和调优等任务可以自动进行，提高了数据分析和建模的效率。此外，DSML 还提供了可扩展的框架和接口，让用户可以自定义算法和模型，实现更高的灵活性和定制性。

数据模型管理智能化，实现资产价值化。低代码 DSML 还提供了数据资产管理功能，可以帮助用户更好地管理数据和模型。例如，数据质量监测、模型更新和管理、模型监控和诊断等功能，可以让用户发现和解决问题，避免潜在的错误和风险。此外，低代码 DSML 还提供了可视化的数据和模型分析工具，让用户能够更深入地了解数据和模型的特点和表现，从而更好地优化模型和应用。同时，DSML 也支持多角色开发，实现数据和模型的共享和合作，使团队内部和外部的用户可以更好地共同使用和开发数据和模型。

平台支持国产化，实现信创和安全。在当前的国际环境下，作为企业数字化的基础软件，数据科学与机器学习平台必须能够与国产系统适配。低代码数据科学与机器学习平台支持与鲲鹏、飞腾等国产 CPU、麒麟国产操作系统、天数智芯天垓、华为 NPU 等国产加速卡以及华为 GaussDB、优炫数据库等国产数据库的适配。

三、低代码 DSML 在推进 AI 普惠化方面的实践

低代码 DSML 可以适用于百行百业，帮助企业开展数字化赋能或企业数字化转型。在工业过

程企业，低代码 DSML 可以应用于质量控制、故障预测、智能制造等场景。比如，工业企业可以通过数据挖掘技术分析生产线的历史数据，识别生产过程中的异常和缺陷。在金融行业中，低代码 DSML 可以应用于风险控制、投资决策、信用评估等场景。比如，银行可以通过模型预测客户的信用风险，提高信贷审批效率和准确性；保险公司可以利用数据挖掘技术发现欺诈行为，降低保险欺诈率。在零售业中，低代码 DSML 可以应用于推荐系统、营销策略优化、库存管理等场景。比如，电商平台可以通过个性化推荐算法，向用户推荐最合适的商品，提高用户满意度和交易量；零售企业可以通过预测需求和优化库存，降低库存成本和滞销风险。

（一）加氢裂化工艺预测的应用实践

中海油惠州石化是一家集石油化工、煤化工、石化产品销售为一体，配套齐全的大型炼化一体化企业，原油加工能力 2200 万吨/年、芳烃生产能力 100 万吨/年，加工规模和技术水平位居国内炼油企业第一梯队。近年来，企业积极推进信息化工业化两化深度融合，基于工业互联网提出“智能炼厂”，在装置优化、计划优化、供应链优化等环节推进相关应用研究。目前，企业工业互联网已广泛应用于工艺、设备、HSE、能源、经营管理、物流等领域。

与此同时，炼化企业为应对低油价与成品油市场寒冬，纷纷以生产优化为核心深入实施降本增效，进而对生产模型精度、模型开发效率以及模型应用广度等都提出更高要求。而智能建模技术作为“智能炼厂”的核心技术，以炼油生产为“目标函数”，能通过快速创建智能化应用，为安全生产、降本增效提供智能决策，成为石化企业普遍关注的关键技术。加氢装置（包括加氢精制、加氢裂化）是石油炼化企业的关键环节，企业希望通过敏捷建模实现炼化生产智能化。然而，既有建模方式门槛高、难应用，针对生产优化，企业早已应用基于机理的传统优化建模软件如 Aspen、Petro-SIM、ProII，同时也在尝试基于机器学习算法的大数据建模技术，如将炼化机理知识与深度学习算法相结合，已经实现加氢装置原

料油换热器结垢预测、加氢催化剂床层温度预测等场景的智能化应用，并取得了较好效果。但两种建模路径都难以很好地满足业务需求，使企业处于“有数据、无模型、有模型、难应用”的被动局面。具体痛点表现在以下几方面：

一是技术门槛高。基于机理的传统优化建模严重依赖专家经验以及国外厂商提供的优化求解软件，且建模软件模型收敛慢，调优复杂。而机器学习建模过程涉及数据准备、特征工程、模型构建、模型部署，也严重依赖具备专业建模能力的数学家。该企业缺乏懂建模也懂业务的专业复合型人才，同时高技术门槛阻碍了建模技术在企业炼化场景的广泛应用。

二是建模周期长，建模成本高。传统优化建模和机器学习建模都存在着“大数据、小任务”，执行任务单一的特点。其建模过程中数据体量大、开发难度大、研发周期长，而开发的模型仅能适用于单个“小场景”。如果想覆盖炼化过程的更多场景就需要定制化地开发多个模型，导致两种建模方式均投入巨大，难以快速为企业带来效益提升。

三是模型部署难。在企业业务中，当机器学习算法模型训练完毕部署到生产环境中时，数据采集、预处理和边缘推理计算时均受智能设备影响，AI 模型的实际效果与智能设备的选型、接入、点位选择以及人力调试密切相关。然而该化工企业智能设备来源广泛，不同生产领域的设备型号各异，缺少统一的接入标准，且设备安装、接入和调试需要耗费大量人力，导致模型在不同生产领域部署时均需以项目课题形式实施，模型部署难以进一步制约机器学习建模在企业的推广应用。

鉴于机器学习建模在前期实验中的出色效果，企业决定将深度学习算法进行推广应用，将“智能炼厂”推进“深度编码”阶段，通过先进的技术手段，如容器化、分布式计算、低代码等技术，实现对深度学习算法在炼化领域落地的助推。而实现深度学习算法的推广应用，需要大数据建模平台做支撑。企业考虑到自建大数据建模平台成本投入大、技术难度较高，如分布式计算环境对硬件资源要求高、大数据组件的开发与应用难度

大、模型部署对接各个系统复杂性较高等，于是决定部署 DSML 平台（如图 1 所示）。

企业在既有工业互联网平台架构下，针对石化煤柴油加氢装置构建起一站式炼化工业大数据分析建模平台，集成数据预处理、主流机器学习算法、智能优化算法和模型自动训练方法，最终实现生产数据到模型应用的端到端建模平台^{[8][9]}。

（二）企业财务工作自动化的应用实践

财务管理是企业管理的最重要的环境之一，不仅要能够衡量企业的财务状况和经营成果，还要能够帮助企业优化资源配置，实现降本增效助力战略计划落地。如高科技企业，公司业务范围广泛，涉及的数据量庞大，对于财务部门而言，面临着数据量庞大、数据来源复杂、数据分析效率低下等问题。传统的手工数据分析方式已经无

法满足快速决策的需求，需要一种更加高效、准确的方法。具体痛点表现在以下几方面：

财务部门需要花费大量时间和精力手工提取数据，并且数据来源不统一、格式不规范，导致数据质量参差不齐；财务部门需要对大量数据进行分析，但是人工分析效率低下、容易出现错误，且无法发现隐藏在数据中的规律和趋势；财务会计和管理会计在实际工作中会遇到个性化需求，例如，财务会计需要根据公司业务使用不同系统的数据定制个性化报表，而管理会计则需要根据实际情况使用算法来做预警或者稽查，而财务人员平时使用的标准化软件无法实现，如果通过第三方平台定制功能，开发成本非常高。

基于 DSML 平台打造的财务自动化与智能化工具，开发如图 2 所示财务自动化模型，引入“数

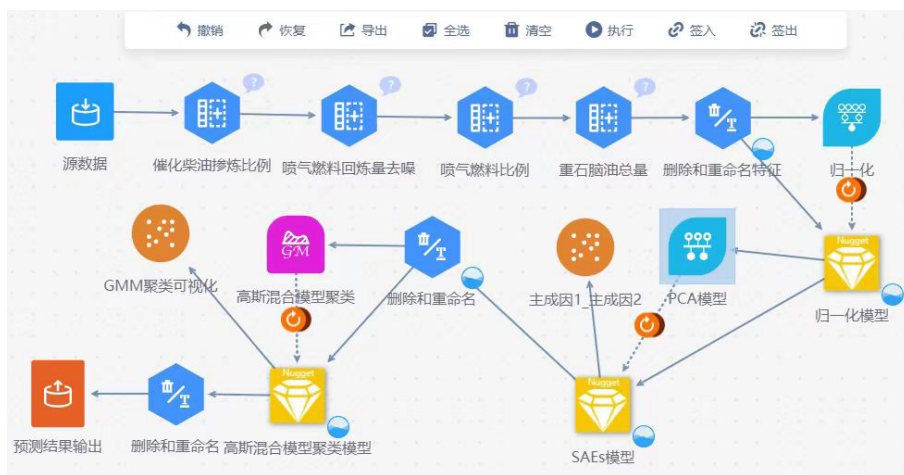


图 1 模型训练的无代码建模算子流组件

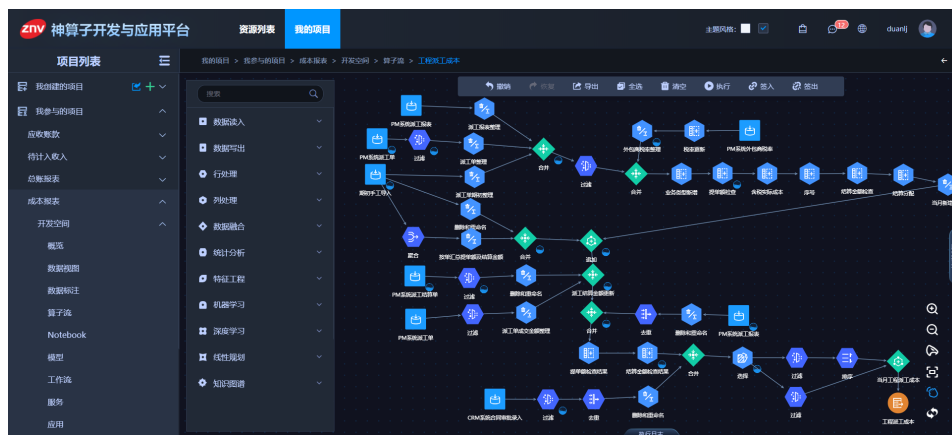


图 2 财务自动化无代码建模算子流

字员工”，从而可以带来如下益处：

在多源数据下自动生成报表，使财务人员从繁琐的报表制作中解放出来，可以专注于业务理解和数据分析；支持分布式存储、计算，帮助财务人员集中且高效处理和管理多源海量的财务数据；支持无代码或者低代码的开发操作，财务人员能够通过简单的培训就能够轻松上手，根据自身业务的个性化需求轻松自开自用，降低财务人员的专业要求。

显而易见，企业通过部署基于 DSML 技术的财务自动化与智能化工具，企业财务人员可以从繁琐的重复工作解放出来，专注于高价值、创造新的工作，提升他们工作的幸福感。

四、推进 AI 普惠化的进一步思考

数据智能技术将越来越多地赋能各行各业。随着数据智能技术的不断成熟和发展，我们可以预见未来这一技术将在更多行业领域发挥关键作用。企业和组织应积极关注数据智能技术的最新动态，了解其在各行业的应用情况，以便及时跟进并将其应用到自身业务中，从而提高企业的决策效率、降低成本、推动创新和增强竞争力。另一方面，AI 技术民主化趋势加速，低代码 DSML 等新一代人工智能技术普惠化的发展，将使得越来越多的非专业人员能够参与到 AI 系统的开发和应用中。这必将有助于缩小技术鸿沟，让更多企业受益于人工智能技术的发展。

伴随着大语言模型技术的进一步大范围使用，ChatGPT 使消费者人工智能时代提前到来。与此同时，国家东数西算的战略部署，将使人工智能的三要素“数据、算法、算力”的发展加速。我国拥有世界最优秀的有为政府，应进一步规划构建基于大语言模型的 DSML 平台，让更多的中小企业共享数据和模型，提升企业的经营效率，助力企业开发数字化智能产品，引领企业走上高质量发展和数字化转型的跃升之路。

参考文献：

- [1] 徐宗本，姚新. 数据智能研究前沿 [M]. 上海：上海交通大学出版社，2019：6.
- [2] 黄勇，孟晨静等. 2022 爱分析·数据智能

厂商全景报告 [R].

[3] Lee Jay, Li Xiang, Xu Yuanming, Yang Shaojie, Sun Keyi. Recent advances and prospects in industrial AI and applications[J]. ACTA Automatica Sinica, 2020, 46(10): 2031–2044. doi: 10.16383/j.aas.200501.

[4] Raphael J. L. Townshend, Stephan Eismann, Andrew M. Watkins, Ramya Rangan, Maria Karelina, Rhiju Das, And Ron O. Dror. Geometric deep learning of RNA structure [J]. Science, 2021, 373 (6558): 1047–1051.

[5] ANDREW NG: UNBIGGEN AI, IEEE Spectrum [Z].

[6] Peter Krensky, Carlie Idoine, Erick Brethenoux, Pieter den Hamer, Farhan Choudhary, Afraz Jaffri, Shubhangi Vashisth, Gartner Magic Quadrant for Data Science and Machine Learning Platforms [R].

[7] 黄勇，孟晨静. 2023 爱分析·数据科学与机器学习平台厂商全景报告 [R].

[8] 王晨，陆鹏飞，阮冰，杨纪，方友，曹晓红. 大数据驱动建模平台技术在加氢裂化工艺预测上的应用 [J]. 石油炼制与化工，2023，54（3）：120–126.

[9] 黄勇，孟晨静. 自主 AI 能力加速企业智能化转型 [R].

作者：徐明，深圳市人工智能产业协会监事长、博士

责任编辑：钟晓媚