arXiv:2304.04309v1 [cs.SE] 9 Apr 2023

面向业务流程管理的大语言模型：机遇与挑战*

Maxim Vidgof**[1][0000−0003−2394−2247]、Stefan Bachhofner**[1][0000−0001−7785−2090]和 Jan Mendling[1, 2, 3][0000−0002−7260−524X]

[1] 维也纳经济与商业大学，Welthandelsplatz 1，1020 维也纳，奥地利 {first_name.last_name}@wu.ac.at[2] 柏林洪堡大学，Unter den Linden 6，10099 柏林，德国 {jan.mendling}@hu-berlin.de[3] 魏岑鲍姆研究所，Hardenbergstraße 32，10623 柏林，德国

摘要。 大语言模型是具有大量参数的深度学习模型。这些模型在大量任务上取得了显著进展，因此能够作为宝贵且用途广泛的工具，服务于各种应用。它们的能力也为业务流程管理带来了机遇，然而，这些机遇尚未得到系统研究。在本文中，我们通过聚焦BPM生命周期中的各项管理任务，探讨这一研究问题。我们提出六个研究方向，重点关注使用大语言模型时需要解决的问题，其中包括面向实践者的使用指南。

关键词： 自然语言处理 · 大语言模型 · 生成式预训练 Transformer · 深度学习 · 研究挑战

1 引言

近期，基于大语言模型（LLM）构建的应用一经发布，便迅速被广大用户采用。ChatGPT 仅用 2 个月就吸引了 1 亿用户 [31]。其迅速普及的关键原因在于适用范围广，使其成为一种通用技术。研究中的许多任务也可以借助大语言模型应用来完成，包括寻找同行评审人、评估手稿和资助申请、润色手稿，以及总结文本 [32]。因此，有人认为，大语言模型——尤其是对话式大语言模型——是“科学领域的游戏规则改变者” [6]。

目前，关于 ChatGPT等应用的大部分讨论都集中于一个问题：它现在以及未来的表现会有多好。我们认为，这

* 本研究获得了 Teaming.AI 项目的资助。该项目属于欧盟地平线2020研究与创新计划，资助协议编号为 957402。Jan Mendling 的研究获得了柏林爱因斯坦基金会资助（项目编号 EPP-2019-524），以及德国联邦教育与研究部资助（项目编号 16DII133）。

** 贡献相同提交至 *arXiv* 的预印本

这个问题需要以明确界定的任务为出发点来探讨。从任务本身入手，将使讨论从各种滑稽或令人不安的错误和偏见 [31] 转向如何组织人类专家与大语言模型应用之间的协作。此外，这也提供了了解特定失败类别的机会，最终有助于以系统化的方式改进这项技术。

本文探讨了一个研究挑战：如何将大语言模型应用集成到业务流程管理的不同阶段。为此，我们参考 BPM生命周期 [8] 及其各项管理任务 [13]。我们的研究采用探索性方法，制定了将大语言模型应用集成到特定业务流程管理任务中的策略。我们发现了多种颇具前景的应用场景，并指出了未来研究面临的挑战。

本文结构如下。第 2 节讨论深度学习（DL）和大语言模型（LLM）的核心概念及其与业务流程管理（BPM）实践的关系。第 3 节识别并讨论业务流程管理中以及不同 BPM 生命周期阶段的大语言模型应用。基于这些应用，第 4 节阐述六个核心研究方向，内容涵盖大语言模型如何改变业务流程管理项目的动态和执行方式，以及业务流程管理特定的数据集和基准。第 5 节指出了使用大语言模型时面临的挑战。此外，我们还展望了大语言模型未来可能的发展。

2 背景

大语言模型应用的兴起，为众多新的业务流程管理相关应用铺平了道路。迄今为止，业务流程管理已采用自然语言处理 [1],人工智能 [7], 和知识图谱 [14] ，以支持各种应用场景。本节将介绍深度学习（第 2.1 节）和大语言模型（第 2.2 节）的基础知识，以阐明它们各自的具体能力。

2.1 深度学习

近期的大语言模型应用建立在机器学习和深度学习模型之上，例如循环神经网络（RNN）和变换器网络。机器学习（ML）研究的算法“能够基于自身以往的经验进行学习，从而提升任务表现” [15]。从本质上说，机器学习技术采用监督学习、无监督学习和强化学习等范式。其中有几种与大语言模型相关。

在监督学习中，机器学习算法接收一组组合作为输入，其中每组由表示某个概念的特征以及一个标签组成。重要的是，该标签与任务相关，并编码了算法应当学习有关这些概念的内容。例如，对于垃圾邮件分类器，标签可以是垃圾邮件 和非垃圾邮件 ，也可以是带有注释的图像边界框。对于大语言模型而言，有两种相关的监督学习情形：少样本学习和零样本 学习。少样本 学习是指机器学习算法仅凭少量标注数据适应新情境，而零样本 学习则是指算法完全无需标注数据即可做到这一点。例如，可以向语言模型提供少量输入-输出对，模型无需更改任何参数即可反转映射函数。在无监督-

监督学习中，算法仅接收一个概念的特征张量作为输入，而所需的输出是未知的。然后，算法会寻找特征张量中所呈现概念的结构属性。一个典型应用是降维，例如使用自编码器。在强化学习中，算法接收一个概念的特征张量作为输入，并据此产生输出，然后通过奖励对该输出进行评估。接着，算法利用这些反馈来改进其参数。ChatGPT 使用一种称为深度强化学习的强化学习形式来提升其语言生成能力，尤其是“从人类反馈中学习总结” [29]。ChatGPT使用一种评估其生成响应质量的奖励信号进行微调，目标是随着时间推移最大化奖励信号。模型从奖励信号中学习的能力使其能够生成越来越相关且连贯的响应。

深度学习（DL）是一种基于神经网络（NN）的机器学习（ML）方法。一般而言，深度学习模型由多个层层叠加的神经网络构成，能够学习多层表示 [10]。重要的是，这些表示可以在没有监督的情况下学习。只有一个隐藏层的网络称为浅层网络。与浅层网络相比，深度网络能够处理更复杂的问题。大量数据的可用性、优化加速方法的改进以及强大的计算资源，使这些网络能够得到有效训练。在自然语言处理领域，深度学习在机器翻译、情感分析和命名实体识别等任务中尤为有效。事实证明，深度学习从输入数据中学习多层表示的能力对这些任务特别有用。这是因为自然语言处理涉及处理词语和字符序列，而这些序列之间的关系往往复杂且具有多层次结构。大量标注训练数据和强大的计算资源使深度学习模型在许多自然语言处理（NLP）任务中取得了最先进水平的成果。例如，Vaswani等人在论文“*Attention is All You Need*” 中提出的Transformer架构 [33] 已成为许多NLP任务（包括语言翻译和语言建模）的标准架构。

近年来，业务流程管理研究在很大程度上整合了深度学习的能力，用于流程预测。概述请参见 [16]。近期还有利用自动流程发现 [28], 从手绘草图生成流程模型 [27], 以及进行异常检测 [17]的应用。

2.2 大语言模型

大语言模型是使用海量文本数据训练的深度学习模型，用于执行各种自然语言处理任务。这些模型的参数规模通常从数亿到数十亿不等，旨在捕捉人类语言的复杂性和细微差别。最大的模型，如 GPT-1 和 GPT-3，能够生成类似人类的文本、回答问题、翻译语言和计算机代码。这些模型的训练过程

涉及处理海量文本数据，并利用这些数据学习词语和短语之间的模式与关系。随后，这些模型会利用所学信息，预测特定语境下给定词元或词元序列的可能性。这使它们能够生成连贯且符合语境的文本，或执行其他与语言相关的任务。大语言模型的兴起推动了自然语言处理领域的重大进展，并广泛应用于各种场景，包括聊天机器人、虚拟助手和文本生成系统。它们的优势之一是能够通过基于提示的学习执行少样本和 零样本学习 [11]。

2018年，Radford 等人在其关于”通过生成式预训练提升语言理解能力” [23] 的论文中介绍了GPT-1（有时也简称为 GPT）。生成式预训练变换器（GPT）-1 指的是作者所训练的最大模型（1.1 亿个参数）。在该论文中，作者研究了分两个阶段训练的变换器网络在语言理解方面的能力。在第一阶段，他们训练一个变换器网络，根据一组此前出现的词元来预测下一个词元（也称为无监督预训练、生成式预训练，或在统计学中称为自回归）。在第二阶段，变换器网络通过监督学习在各项任务上进行微调（也称为判别式微调）。总而言之，他们的主要发现是：在第一阶段结合任务无关的无监督学习，然后在第二阶段使用该模型，通过监督学习对其进行任务微调，可以带来性能提升——从文本蕴含任务中的 1.5% 提升到常识推理任务中的 8.9%。

2019年，Radford 等人在其论文 ”语言模型是无监督的多任务学习者”[24] 中介绍了 GPT-2。同样，GPT-2 指的是他们所训练的最大模型。因此，就模型规模（15亿个参数）以及训练数据规模而言，GPT-2 是 GPT-1 的扩展版本。具体而言，GPT-2 的参数数量大约是 GPT-1 的十多倍，其使用的训练数据量也大约是 GPT-1 的十多倍。他们报告了两项主要发现。首先，在 GPT-2 的训练数据中不包含特定任务数据集的情况下，无监督的 GPT-2 仍能胜过使用特定任务数据集训练的语言模型。其次，GPT-2 似乎能够从未标注的文本数据中学习任务（例如问答）。不过，在这两种情况下，其性能都未达到最先进水平。总之，他们的主要发现是，只要拥有足够的未标注训练数据，大语言模型就无需针对这些任务进行训练，也能学习相关任务。

2020年，Brown 等人发表了论文”语言模型是少样本学习者”[5]，并在其中介绍了 GPT-3。与上述两种情况不同，GPT-3 指的是作者训练的所有模型，即一个模型系列。作者训练的最大模型是 GPT-3 175B，这是一个拥有 1750 亿个参数的模型。在论文中，作者展示了像 GPT-3 这样的语言模型仅凭少量示例就能学习任务，因此标题中包含”少样本学习器”。作者通过在各种任务上对 GPT-3 进行微调来展示这一能力，这些任务包括问答和语言翻译，且仅使用少量示例。

2023 年，OpenAI 推出了GPT-4 [21]。与 GPT的早期版本不同，该版本是一个多模态模型，因为它可以处理作为输入的文本和图像，并生成文本。该模型是一次重大进步，因为它在众多基准测试中表现更佳；然而，它存在可靠性问题、上下文窗口有限，以及无法像之前的 GPT 模型那样从经验中学习等缺陷。不过，这一发布版本有别于之前的 GPT模型，因为 OpenAI 对”架构（包括模型规模）、硬件、训练计算量、数据集构建、训练方法或类似信息”讳莫如深。我们对该模型仅知晓的是：它采用 Transformer 风格的模型架构，在公开可用且未披露的许可数据上进行预训练，通过预测下一个词元来训练，随后使用来自人类反馈的强化学习（RLHF）进行微调。尽管有所不同，作者仍在报告中纳入了关于模型可扩展性预测的研究结果。他们尤其报告了如何根据计算量预测损失，以及在给定训练方法的情况下，如何根据计算量预测平均对数通过率（衡量代码样例通过单元测试数量的指标）。在这两种情况下，他们发现，基于计算量显著更少（少1.000 到 10.000倍）所生成的数据，都可以高准确度地预测相应指标。他们还发现了某项任务的逆向缩放规律，即任务性能会先随着模型规模的增大而下降，随后在达到某一特定模型规模后上升。

2022 年，OpenAI 推出了名为 ChatGPT [19] 的对话式大语言模型。作为一种模型，ChatGPT 的首个版本基于 GPT-3.5，是 InstructGPT 的同系列模型。GPT-3.5 是一种 GPT-3.0 模型，其训练数据集包含截至 2021 [18] 年第四季度的文本和软件代码。InstructGPT 在 ”训练语言模型遵循指令并接受人类反馈 ” [22], 中首次推出；它是一个 GPT-3 模型，首先通过监督学习进行微调，随后通过基于人类反馈的强化学习进行微调 [29]。因此，ChatGPT 是一种针对与用户进行对话式交互而微调的 GPT-3.5 模型。换句话说，用户通过一系列文本（即对话）与模型交互，以完成某项任务。例如，我们可以将一段文本复制并粘贴到 ChatGPT 的输入框中，让它为文本生成摘要。我们甚至可以提出更具体的要求，例如指定摘要应为 10 句话，并采用偏好的文风撰写。重要的是，如果我们对结果不满意，可以让 ChatGPT 优化它自己生成的摘要，而不必复制粘贴需要总结的文本。在撰写本文时，ChatGPT 可以使用 GPT-4 作为后端大语言模型。

此外，还有其他大语言模型。2022 年，Zhang 等人通过论文 *“OPT*：开放式预训练 *Transformer* 语言模型” [34]介绍了开放式预训练 Transformer（OPT）。该论文的主要贡献是向感兴趣的研究人员开放包括九个模型在内的所有相关成果。这些模型在参数规模和性能方面均属于 GPT-3级别的模型。另一个开放式大语言模型是 BigScience 大型开放科学开放获取多语言语言模型（BLOOM）（1760 亿参数），由 BigScience 研讨会开发 [26]。

2.3 大语言模型的采用

上文简要讨论了大语言模型，其中我们尤其关注 GPT模型家族，因为我们推测它们是最受欢迎的大语言模型。从 GPT-3 到 GPT-4 的转变非常重要，因为这在各种基准测试中带来了巨大的提升，尤其是在学术和职业考试中 [21]。正如我们所论证的，这些自然语言处理任务中的性能提升源于对自然语言的理解，并对可实现自动化的范围——自动化前沿——产生了巨大的影响。当自然语言理解与插件软件组件相结合时，这一前沿无疑会进一步拓展。事实上，在撰写本文时，GPT 背后的公司正在试验 ChatGPT 插件。目前提供的插件包括 Klarna、Wolfram、用于信息检索的向量数据库集成，以及用于Python 的嵌入式代码解释器 [20]。这对机器人流程自动化（RPA）产生了影响，更广泛地说，也影响了包括业务流程管理系统（BPMS）在内的业务流程自动化，并且更普遍地影响了工作的开展方式。

3 大语言模型与业务流程管理生命周期

在本节中，我们确定大语言模型在业务流程管理中的应用。我们沿着业务流程管理生命周期的各个阶段，即识别、发现、分析、重新设计、实施和监控阶段，系统地探讨这些应用 [8]。通过这种方式，我们补充了近期构建大语言模型应用总体清单的相关工作，例如其他领域中的数据挖掘领域 [4]。

3.1 识别

业务流程管理生命周期始于识别。通常，在此阶段，公司中没有太多结构化的流程知识可用，相关信息必须从异构的内部文档中提取。这正是大语言模型大放异彩的地方，因为它们可以快速扫描并总结大量文本，突出显示重要文档，或直接输出所需信息。

从文档中识别流程其思路是将组织中现有的所有相关文档作为输入提供给大语言模型。这些文档可以包括法律文件、职位描述、广告、内部知识库和手册。随后要求大语言模型识别组织中正在开展哪些流程。还可以进一步指示其根据输入文档所描述的流程对文档进行分类。多模态大语言模型可以进一步改善结果，因为图表、演示文稿和照片也可以直接用作信息源。

流程选择 还可以要求大语言模型根据提及相关流程的文档数量和类型等信息评估流程的战略重要性，并从流程描述中提取这些信息。如果能够访问

[4] 例如，可参阅 OpenAI Cookbook GitHub 存储库，其中提供了 OpenAI API 的代码示例。

支持该流程的信息系统或其他关键绩效指标，大语言模型也可以评估流程的健康状况。最后，只要将必要的信息（例如近期技术报告）作为输入，理论上也能够评估可行性。根据这些标准，大语言模型可以确定流程的优先级，以便进一步改进。

3.2 发现

BPM生命周期的第二个阶段是流程发现。在此阶段，会选择一种或多种流程发现方法来生成流程模型。谈到自动化流程发现时，通常指的是流程挖掘——一种从支持流程执行的信息系统留下的事件日志中提取流程模型及其他相关数据的技术。不过，借助大语言模型，其他发现技术也能实现（至少部分）自动化。

基于文档的流程发现 除了流程挖掘之外，文档分析也是一种成熟的流程发现方法。在这种方法中，流程分析师会利用内部文档、招聘广告、手册等异构来源中的信息。搜索这些文档可能需要大量时间和精力。大语言模型非常适合这项任务，因为它们能够以简洁且结构化的方式总结大量文本。更具体地说，它们可以按照所需格式（纯文本、编号列表等）输出流程描述。还可以指定详细程度，例如输出是否只应包含活动和事件，还是也应包含资源及其他信息。最后，由于一些大语言模型还能够处理XML 等结构化文档格式，实际上甚至可以自动生成业务流程模型与标记法模型。

从通信日志中进行流程发现 另一种可用于基于证据的流程发现的信息来源是通信日志，即流程参与者之间的电子邮件和聊天记录：内部员工，以及外部合作伙伴和客户。大语言模型可以从这些通信日志中提取模式，这些模式可视为流程中的各种步骤。随后，它们也可以类似地生成流程描述或模型。

访谈聊天机器人大语言模型在流程发现中的潜在应用还可以超越基于证据的发现。另一种常见的发现方法是与领域专家进行访谈。在这些访谈中，流程分析师会询问有关流程的问题，并根据多次访谈生成流程模型。通常，需要与不同领域专家分别进行多次访谈，才能生成流程模型的第一个版本。之后，还会进行其他轮次的访谈，以获取并纳入反馈，并执行验证。在最糟糕的情况下，领域专家可能对流程有相互冲突的看法，此时，解决这些冲突对于流程分析师和领域专家而言都会成为一项非常困难且耗时的任务。

大语言模型可以通过为领域专家提供聊天机器人界面来解决这一问题的部分难题。这样，领域专家便可在聊天中回答问题。

这能带来许多优势。首先，领域专家无需为访谈预留大段时间，而是可以按照自己期望的节奏与聊天机器人交流。其次，反馈周期会缩短，因为大语言模型可以在与领域专家交谈后，甚至交谈过程中直接生成流程模型并更新模型，从而使验证能够与模型创建同步进行。最后，如果多位领域专家同时（且独立地）与聊天机器人互动，而聊天机器人又能在对话中利用所有这些输入，那么这些优势还会进一步扩大。不过，后一种方案实施起来更加困难。

组合式流程发现 所有流程发现方法都有各自的优点和缺点。通常，人们会结合使用这些方法，以取得最佳效果。然而，这种组合受到分配给流程发现任务的资源限制。上文介绍的发现方法能够提供有价值的结果，同时所需资源少得多。因此，可以同时采用更多方法，以获得更好的结果。除了传统的流程挖掘或“手动”流程发现之外，还可以使用这些方法的组合，从而获得最全面的洞见。尽管不同方法的结果可能存在一些不一致之处，需要加以修正，但修正工作也可以通过（半）自动化的方式完成。

流程模型查询 由于大语言模型似乎能够“理解”以 XML 序列化的流程模型，因此可以用来回答有关模型的问题。这对于质量保证非常有用。首先，它可以用于检查语法质量。虽然已有工具能够完成这项工作，而且开销要小得多，但在大语言模型中具备这一功能仍然很方便，因为与其他方法相比，大语言模型或许还能检查其他质量方面。例如，它也可以检查语义质量。事实上，流程分析师可以同时向大语言模型输入访谈记录和流程模型，大语言模型便能根据访谈检查模型的有效性和完整性。当然必须指出，这只有在访谈记录本身具备有效性和完整性的前提下才会奏效。检查模型语义质量的另一种方式是进行流程仿真，例如，明确询问大语言模型给定的流程模型是否能够生成某个执行序列，或者请它给出模型可能生成的执行序列。众所周知，大语言模型能够模拟 Linux shell，因此，只要提供足够的输入，它们或许也能模拟 BPMS 执行引擎。最后，只要将某些准则的定义（例如 7PMG）作为输入，大语言模型也能在一定程度上检查模型的实用质量。还必须指出，大语言模型不仅能发现这些质量问题，还能提出修正建议。

3.3 分析

下一阶段是流程分析。在这一阶段，会分析已发现的流程，以找出问题和瓶颈。尽管这项任务需要大量认知投入，但在某些方面可以利用大语言模型协助人工分析师。

问题发现 如果流程中存在问题，很可能已经有人对此提出过投诉。根据公司、产品和流程的不同，投诉者可能是客户、合作伙伴或员工；投诉也可能出现在不同的平台上，包括社交媒体、客服服务或内部沟通工具。大语言模型擅长总结大量非结构化文本并发现其中的模式，因此可以利用这一能力完成这项任务。只需从这些平台抓取文本，并将其作为输入提供给大语言模型，再配上类似“找出客户投诉过的所有事项”这样简单的提示即可。

问题定位 发现流程中的问题后，下一步是找出导致该问题的流程环节。在某些情况下，这可能是一项艰巨的任务，尤其是在流程复杂时。这里的思路是将所有流程模型（如果已知只有一个流程导致问题，且明确知道是哪个流程，则提供相关流程的模型）以及已发现的问题交给大语言模型。大语言模型的任务是通过分析任务名称和描述，提出可能导致问题的任务。在更高级的情况下，大语言模型甚至可能有能力提出一些修复方案。这些方案可能很简单，例如建议将耗时过长的人工任务自动化；也可能是更复杂的流程重新设计建议，前提是将重新设计方法作为额外输入提供给大语言模型，或使用这些方法对其进行训练。

3.4 重新设计

BPM生命周期的第四个阶段是流程重新设计。在这一阶段，会根据已发现的问题和通用的流程改进方法制定流程改进建议，并对这些建议进行评估，最终形成未来流程模型。

业务流程改进一个显而易见但很有前景的用例是让大语言模型重新设计流程。如前所述，大语言模型可以解决由单项活动引发的简单问题。但它的能力不止于此，理论上仅取决于提供给大语言模型的输入质量。事实上，如果向其提供关于流程的全面信息（包括详细的流程模型以及流程或任务的描述），并提供某种重新设计方法的详细说明（或者用某些重新设计方法对其进行训练），那么重新设计业务流程就如同告诉大语言模型将该方法应用于流程一样简单。不过，这一做法还可以进一步完善。首先，可以将上一阶段发现的问题描述作为额外输入，以引导流程重新设计，并优先解决这些问题。其次，可以指示大语言模型应用不同的重新设计方法，并分别列出每种方法给出的建议，以便分析人员从中选出最佳方案。此外，还可以让大语言模型自行挑选最佳建议并说明选择理由。当然，必须指出，只有在提供了充分输入的情况下，这种做法才能奏效。例如，对于内向型重新设计方法，既需要方法本身，也需要详细的流程信息；对于外向型方法，除了这些信息外，还必须有足够的外部信息，和/或有办法让大语言模型与外部世界进行有效沟通。

3.5 实施

BPM生命周期的下一阶段是流程实施。该阶段涵盖为改变流程参与者的工作方式所需的组织和技术变革，以及为未来流程提供的IT支持。

带有纯文本的*BPMN*模型解释如前所述，大语言模型可以处理以XML 序列化的 BPMN 模型。我们已经讨论过，大语言模型如何操纵流程模型，以提高质量，并提出或纳入重新设计方案。为了完成这一闭环，大语言模型可以生成对BPMN模型的文本解释。更有趣的是，还可以控制详细程度。因此，根据目标受众的不同，大语言模型既可以生成文本概览，也可以生成模型的详细描述。如果BPMN模型本身包含足够的细节，它还可以将其转换为软件开发人员所需的需求。

*BPMN*模型聊天机器人 在前一个用例的基础上，模型描述也可以针对每位特定用户进行定制。这样一来，给定一个模型，或者更好的是，给定一个包含附加文档的模型仓库，大语言模型就可以为例如流程负责人以及个体参与者准备特定描述，并对他们负责的所有具体任务进行详细描述和解释。此外，在此用例中，还可以加入用户与大语言模型之间的交互。这样，用户可以针对自己未理解的部分请求澄清，或者在需要指导时，通常也可以请求更多细节。

流程编排器 大语言模型可以通过 API访问，同时也可以自行访问 API，从而带来各种各样的巨大机遇。前者意味着它可以用于自动化任务，并由编排器调用；后者则意味着从理论上讲，它本身可以成为编排器：在将可执行流程模型和附加约束作为上下文，并将所需的实例数据作为输入的情况下，它在调用其他 API 并分配任务时，理论上可以比传统编排器以更加灵活的方式执行流程。

3.6 监控

BPM生命周期的最后一个阶段是流程监控。在此阶段，已实施的流程将投入执行，其绩效也会受到监控。在这一阶段收集的观察结果既用于运营管理，也作为生命周期后续迭代的输入。

流程仪表板聊天机器人仪表板是一种强大的工具，可在单个屏幕上概览流程中最重要的关键绩效指标。然而，其最终目标是告诉查看者流程状态是否良好，而数字和颜色大多只是作为中介媒介。大语言模型可以省去这一中间步骤，让用户直接了解流程的状态。

4 研究方向

本节提出研究方向。我们将研究方向分为三组。第一组研究大语言模型在实践中的应用及其应用场景。这包括在公司的业务流程管理项目中使用大语言模型，或将其作为信息系统（IS）的一部分（第 4.1 节）；为从业者和研究人员制定使用指南（第 4.2 节）；以及梳理业务流程管理任务（第 4.3 节）及其对应的数据集（第 4.4 节）。第二组研究如何将大语言模型与现有的业务流程管理工具，以及更广泛而言的业务流程管理技术相结合，以提升用户体验（第 4.5 节）。至关重要的是，这一组将借鉴第一组的研究发现。第三组也是最后一组，专门为业务流程管理开发大语言模型，使这些模型能够理解业务流程的语境和语言，并支持流程发现、监控、分析和优化等各种任务（第 4.6 节）。同样，这一组也建立在第一组的研究发现之上。

4.1 大语言模型在业务流程管理实践中的应用

第一个研究方向探讨大语言模型在实践中的应用。一个亟待回答的重大问题是，大语言模型可以用于哪些任务。在第 3 节中，我们列出了一系列可使用大语言模型完成的任务。不过，这份清单可能并不完整，而且其中一些任务最终可能会被证明用处不大。与此相关的问题是，哪些任务能够为组织带来价值，最终又有哪些任务能带来最大价值。下一个重大问题是，任务与实现预设价值所需的模型属性之间有何关联。这里的一个问题是，哪些任务可以借助现有模型完成。另一个值得研究的问题是，是否每项任务都需要使用规模最大、因而准确度最高的模型。我们推测未必如此。最后，也是最重要的一个待解问题是，大语言模型将如何改变业务流程管理项目以及主动管理的流程中的工作开展方式。例如，我们推测，对话式大语言模型可能会取代著名的鸭子方法 [5]中的“鸭子”。这是一个社会技术系统问题，因此我们坚信，业务流程管理领域的学界，以及更广泛的信息系统领域的学界，尤其有能力为解答这一问题作出贡献。

4.2 研究人员和从业者的使用指南

第二个研究方向旨在为业务流程管理研究人员和从业者制定使用指南。此类指南需要回答的一个问题是：给定组织背景、生命周期阶段和任务的流程背景，应推荐使用哪种大语言模型以实现预期价值。此外，此类指南还会系统地收集创建提示的最佳实践。例如，在业务流程管理生命周期的流程实施、监控与控制阶段，公司可以考虑在受管理的流程中使用大语言模型。假设这家公司是一家银行，希望自动处理客户咨询回复任务，使用

[5] 橡皮鸭调试

大语言模型。那么，这份指南会针对流程实施推荐一种具体的大语言模型，说明其参数数量，给出创建提示模板的示例，展示如何将客户背景信息填入模板，最后说明如何将客户咨询整合进提示模板。对于流程监控与控制，由于生命周期阶段背景有所不同，指南可能会推荐使用另一种模型来分析不同的咨询类别。举例来说，大语言模型首先将每条咨询归类为正面或负面情绪，然后分别列出两类咨询中排名前五的原因。这个研究方向建立在第一个研究方向的基础上，因为第一个研究方向会确定大语言模型原则上可用于哪些任务，除此之外还会开展其他工作。

4.3 业务流程管理特定任务变体的创建、发布和维护

该研究方向构建并维护两份不同的任务清单。第一份清单将通用的自然语言处理任务映射到业务流程管理中的任务。例如，考虑文本摘要这一通用的自然语言处理任务。在业务流程管理中，文本摘要可以涉及对一组流程描述或任务描述进行摘要。我们可以将这份清单视为一对多映射：一方面是自然语言处理任务，另一方面是业务流程管理任务。第二份清单列举业务流程管理所特有的任务。该研究方向利用第 4.1 节和第 4.2 节所介绍的研究方向中的发现。

4.4 数据集和基准的创建、发布与维护

公共数据集和基准对于大语言模型研究的进展至关重要，因为它们使研究人员能够衡量进展。此外，它们对实践者也很重要，因为它们定义了数据集属性（例如元信息），而这些属性很可能正是他们在微调模型时自身需要的。因此，数据集和基准需要与业务流程管理的自动化需求适当对齐。Blagec 等人的观点与我们类似，但其研究对象是临床专业 [3]。在他们的研究中，他们分析了 450 个自然语言处理数据集，并发现“与临床直接相关的人工智能基准十分稀缺，且未能覆盖临床医生希望得到解决的大多数工作活动”。因此，业务流程管理社区的一个研究方向是开展同样的研究。一个值得研究的问题是，现有的自然语言处理数据集和基准是否与业务流程管理相关，例如，它们是否涵盖业务流程管理研究人员和实践者的活动。这一研究方向建立在第 4.3 节所述研究方向的基础之上。

4.5 大语言模型与业务流程管理工件

这一研究方向探讨大语言模型、业务流程管理制品与业务流程管理任务之间的相互作用。其目标是了解哪些制品是大语言模型及其多模态后继技术生成有用输出所必需的。因此，这可以被视为提示工程的一个特例，我们或许可以称之为面向业务流程管理的多模态提示工程。由于大语言模型的输出质量在很大程度上取决于所提供上下文的质量和数量，因此这是一个重要的研究方向

。换言之，上下文越丰富、每项上下文的质量越高，大语言模型的输出质量就越高。因此，我们认为这一方向应被视为一个独立的研究方向。以之前提到的客户咨询流程为例。在这种情况下，我们可以设想，大语言模型的上下文取决于咨询内容。某种情况下，客户可能会在咨询中附上一张图片。也可以想想咨询流程的重新设计阶段。在这一阶段，人们会创建各种制品，例如在白板上绘制的流程图、在文字处理软件中对这些流程所作的评论，以及记录在音频文件中的数据可用性和访问权限说明。当我们询问另一个可能不同的大语言模型，为什么尚无法回答客户关于当前特惠活动的咨询时，这些信息或许会有所帮助。原因可能是尚不存在用于存储特惠活动信息的中央系统。这一研究方向建立在第 4.1 节和第 4.2 节所述研究方向的基础之上。

4.6 面向业务流程管理的大语言模型的开发与发布

该研究方向探讨如何构建用于业务流程管理任务的大语言模型，前面讨论过的所有研究方向都是这一方向的基础。与通用语言模型相比，该方向旨在构建能够适应业务流程管理特定挑战和需求的大语言模型。这包括专门针对业务流程管理构建的模型，以及在业务流程管理领域进行微调的通用语言模型。该方向的一个重要方面是将所创建的大语言模型开源，例如 OPT [34]。这很重要，因为研究人员可以在研究中使用该模型，企业也可以免费将这些模型用于实际应用场景。

5 讨论

本节将讨论大语言模型面临的挑战、结合的力量与过高的期望，并以展望与未来工作作结。

挑战 大语言模型的使用既带来机遇，也带来挑战。例如，它们可以帮助理解艰深的研究，但也会将训练数据集中的缺陷（包括事实错误）带入其生成的文本中 [32]。Borji 对这些错误进行了系统研究，分析了 ChatGPT 的错误并对其进行分类——作者还进一步概述并讨论了此类模型的风险、局限性及其社会影响[6] [4]。作者识别出的错误类别包括推理、事实、数学和编码。 [2], 也开展了类似的缺陷研究，但这些作者关注的是大语言模型整体。《自然》的一篇新闻特稿讨论了这些问题以及使用大语言模型的风险 [9]。对教育而言，一个可能的影响是，作为作业的论文或许应重新审视 [30]。

组合的力量与管理预期 ChatGPT的重大创新并不是引入了一项新技术，而是将

[6] 请参阅 ChatGPT 故障档案库（GitHub）以获取最新列表

已存在的方案以及易于使用的用户界面 [12]。这种组合的影响超越了大语言模型、自然语言处理或机器学习创新的范畴。例如，OpenAI目前正在尝试将ChatGPT与软件插件集成，这甚至可能在短期内为其平台带来一个软件市场[7]。因此，我们建议并倡导在上述研究方向中研究并构建这些与现有 业务流程管理技术的组合，而不是只专注于开发新技术。在本文中，到目前为止，我们已经论证了大语言模型所实现的机遇，简要讨论了它们的不足，并指出了在一个领域内部以及跨越领域边界结合技术的重要性。然而，我们在此也强调，管理由这些非常近期的发展所驱动的预期，甚至可能是过高的预期，是多么重要。例如，关于 GPT-3 后继者可能具备的能力的猜测，在炒作的推动下不断升温，以至于“人员正苦苦哀求着失望” [12]。

展望与未来工作 大语言模型已被应用于商业产品中，并且未来还将继续应用于拥有大量用户的商业产品中。我们推测，这将对研究产生影响，因为资助机构可能会增加对该研究领域的资助申请数量。在我们看来，与大语言模型（直接或间接）交互的用户群体不断扩大因此是不可避免的。未来，我们计划致力于拓展超出本文范围的研究方向。我们预计，大语言模型将影响工作的开展方式（见第2.3节和第4.1节）。但这可能产生远超本文所涵盖内容的影响，例如对业务流程管理能力的影响，即战略、治理、信息技术、人员和文化 [25]。

6 结论

本文提出了用于研究和构建面向业务流程管理的大语言模型的六个研究方向。我们借助BPM生命周期提出大语言模型的应用，以展示这些模型的影响。

参考文献

1. Van der Aa, H., Carmona Vargas, J., Leopold, H., Mendling, J., Padró, L.：在业务流程管理中应用自然语言处理的挑战与机遇。载于：COLING 2018：第27届计算语言学国际会议：会议论文集：2018年8月20日至26日，美国新墨西哥州圣菲。第2791–2801页。计算语言学协会（2018年） 2. Bender, E.M., Gebru, T., McMillan-Major, A., Shmitchell, S.：论随机鹦鹉的危险：语言模型会不会太大？载于：2021年ACM公平、问责与透明度会议论文集。第610–623页。FAccT '21，计算机协会，美国纽约州纽约（2021年）。https://doi.org/10.1145/3442188.3445922, https://doi.org/10.1145/3442188.3445922

[7] https://openai.com/blog/chatgpt-plugins

3. Blagec, K., Kraiger, J., Frühwirt, W., Samwald, M.：推动人工智能发展的基准数据集未能满足医疗专业人员的需求。《生物医学信息学杂志》 p. 104274（2022） 4. Borji, A.：ChatGPT 失败案例分类档案（2023）。https://doi.org/1003494 548550230203494, 2302./ARXIV..https://arxiv.org/abs/.. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., Amodei, D.：语言模型是少样本学习者。载于：Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H.（编）《神经信息处理系统进展》。第 33 卷，第 1877–1901 页。Curran Associates, Inc.（2020年），
https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

6. van Dis, E.A.M., Bollen, J., Zuidema, W., van Rooij, R., Bockting, C.L.：ChatGPT：研究的五项优先事项。《自然》**614**(7947), 224–226（2023年2月） 7. Dumas, M., Fournier, F., Limonad, L., Marrella, A., Montali, M., Rehse, J.R., Accorsi, R., Calvanese, D., De Giacomo, G., Fahland, D., 等：人工智能增强型业务流程管理系统：一份研究宣言。《ACM管理信息系统汇刊》**14**(1), 1–19（2023年） 8. Dumas, M., La Rosa, M., Mendling, J., Reijers, H.A.：《业务流程管理基础》，第2卷。施普林格（2018年） 9. Hutson, M.：机器人作家：语言生成式人工智能的兴起及其风险。《自然》**591**(7848), 22–25（2021年3月） 10. LeCun, Y., Bengio, Y., Hinton, G.：深度学习。《自然》**521**(7553), 436–444（2015年） 11. Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., Neubig, G.：预训练、提示与预测：自然语言处理提示方法的系统综述。《ACM计算调查》**55**(9), 1–35（2023年） 12. Loizos, C.：StrictlyVC与Sam Altman的对话，第二部分（OpenAI）。https://www.youtube.com/watch?v=ebjkD1Om4uw（2023年1月），Connie Loizos的YouTube频道 13. Malinova, M., Mendling, J.：运用集成业务流程管理框架识别应做与不应做之事。《业务流程管理杂志》（2018年） 14. Miller, J.A., Mahmud, R.：使用知识图谱和机器学习进行流程建模与挖掘的研究方向。载于：Qingyang, W., Zhang, L.J.（编）《服务计算——SCC 2022》。第86–100页。Springer Nature Switzerland，Cham（2022年） 15. Mjolsness, E., DeCoste, D.：科学领域的机器学习：现状与未来展望。《科学》**293**(5537), 2051–2055（2001年） 16. Neu, D.A., Lahann, J., Fettke, P.：关于流程预测先进深度学习方法现状的系统文献综述。《人工智能评论》第1–27页（2022年） 17. Nolle, T., Seeliger, A., Thoma, N., Mühlhäuser, M.：Deepalign：使用循环神经网络进行基于对齐的流程异常纠正。载于：《高级信息系统工程：第32届国际会议，CAiSE 2020》，法国格勒诺布尔，2020年6月8日至12日，会议论文集。第319–333页。施普林格（2020年） 18. OpenAI：研究人员模型索引。
h
t
t
p
s://
platfor
m.openai.com/docs/model-index-for-researchers/model-index-for-researchers

19. OpenAI：ChatGPT：优化用于对话的语言模型。https://openai.com/blog/chatgpt/（2022年11月） 20. OpenAI：ChatGPT插件。https://openai.com/blog/chatgpt-plugins（2023年3月） 21. OpenAI：GPT-4技术报告。https://cdn.openai.com/papers/gpt-4.pdf（2023年3月） 22. Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C.L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., 等：通过人类反馈训练语言模型遵循指令。arXiv预印本 arXiv:2203.02155（2022年） 23. Radford, A., Narasimhan, K., Salimans, T., Sutskever, I., 等：通过生成式预训练改进语言理解。https://openai.com/blog/language-unsupervised/（2018年） 24. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., 等：语言模型是无监督的多任务学习者。https://openai.com/blog/better-language-models/（2019年） 25. Rosemann, M., vom Brocke, J.：业务流程管理的六大核心要素。载于：《业务流程管理手册》第1卷，第105–122页。施普林格（2015年） 26. Scao, T.L., Fan, A., Akiki, C., Pavlick, E., Ilić, S., Hesslow, D., Castagné, R., Luccioni, A.S., Yvon, F., Gallé, M., 等：Bloom：拥有1760亿参数的开放访问多语言语言模型。arXiv预印本 arXiv:2211.05100（2022年） 27. Schäfer, B., Van der Aa, H., Leopold, H., Stuckenschmidt, H.：Sketch2process：基于神经网络的端到端业务流程模型与标记法草图识别。《IEEE软件工程汇刊》（2022年） 28. Sommers, D., Menkovski, V., Fahland, D.：使用图神经网络对流程发现技术进行监督学习。《信息系统》第102209页（2023年） 29. Stiennon, N., Ouyang, L., Wu, J., Ziegler, D., Lowe, R., Voss, C., Radford, A., Amodei, D., Christiano, P.F.：利用人类反馈学习摘要生成。载于：Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H.（编）《神经信息处理系统进展》。第33卷，第3008–3021页。Curran Associates, Inc.（2020年），
h
ttps://proceedi
ngs.neurips.cc/paper/2020/file/1f89885d556929e98d3ef9b86448f951-Paper.pdf 30. Stokel-Walker, C.：人工智能机器人ChatGPT能写出高明的论文——学者是否应该担忧？《自然》（2022年12月） 31. Teubner, T., Flath, C.M., Weinhardt, C., van der Aalst, W., Hinz, O.：欢迎来到ChatGPT等开启的时代：大语言模型的前景。《商业与信息系统工程》第1–7页（2023年） 32. Van Noorden, R.：语言生成式人工智能如何改变科学。《自然》**605**(7908), 21（2022年） 33. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.：Attention is All You Need。载于：《神经信息处理系统进展》。第5998–6008页（2017年） 34. Zhang, S., Roller, S., Goyal, N., Artetxe, M., Chen, M., Chen, S., Dewan, C., Diab, M., Li, X., Lin, X.V., 等：OPT：开放式预训练Transformer语言模型。arXiv预印本：https://arxiv.org/abs/2205.01068（2022年）