

揭秘出版企业 AI 工具开发成功之道

编者按 摩根士丹利分析师今年年初曾预测,2024年将是AI应用者之年。至今,一些大型出版企业和技术公司已走在前面,致力于面向核心用户以及为出版各流程提供更精准、快捷、低成本、高效的产品和服务。从本报采访的几家企业的实践经验可以看到,AI在出版业的应用遵循几个准则:制定负责任的AI原则,确保内容准确性,保护用户版权及隐私,选出合适的AI技术开发商,确保用户数据不用于训练大语言模型,每次处理完用户请求后会删除相关数据。



叶莲娜·塞沃

您在励讯集团的主要职责是什么? 励讯集团在一些细分市场的数字化收入占比已经很高,如风险板块电子业务贡献了99%的收入,科技与医学和法律业务板块电子业务贡献了90%的收入,未来数字收入是否会持续增长?

■作为励讯集团首席战略官,我的职责是帮助业务部门制定战略并将其转化为行动。作为一家公司,我们的宗旨是通过开发能帮助众多关键行业的解决方案来造福社会。我们帮助研究人员推进科学知识传播;帮助医生和护士改善患者的生活;帮助律师赢得案件;帮助企业和政府预防欺诈;帮助消费者获得金融服务和保险;帮助客户了解市场并完成交易。这一宗旨推动着我们时刻专注并保持创新。我们工作的核心是利用信息和数据,将其与技术和分析相结合,创建解决方案,帮助客户作出更好的决策,提高工作效率。我们主要通过内生发展(即创新)来实现这一目标。

我们的战略始终如一:过去20年里,我们已将业务从以印刷为基础的出版商转变为以信息为基础的分析集团。2003年,印刷业务创造了近60%的营收。到了去年,印刷业务仅占营收的5%左右。我们将继续以客户为中心,开发更先进的增值工具——增加更广泛的数据集、更复杂的分析,并利用更强大的技术——来推动业务增长。

励讯集团近年来很重视技术的发展,在技术方面的投资越来越大,2023年在技术方面的投资高达17亿美元。在励讯集团的3.6万名员工中,有将近1/3是技术人员,而其中超过一半的技术人员是软件工程师,这意味着励讯集团有5500多名软件工程师。软件工程师必须具备哪些专业背景或技能?

■我们将丰富的数据集、技术基础设施、如何使用新一代创新知识以及深入了解客户

确保数据和隐私安全 持续创新满足客户需求

■叶莲娜·塞沃(Jelena Sevo,励讯集团首席战略官)

□渠竞帆(中国出版传媒商报记者)

需求相结合,为客户创造有效的解决方案。

以客户为中心是我们成功的关键。由于终端市场的多样性,我们工程师的背景和技能组合也将是多种多样的。例如,在律师联讯,技术和产品团队不仅包括用户体验设计师、产品经理和软件工程师,还包括主题专家、了解法律和法律语言的知识产权律师以及近200名数据科学家和机器学习工程师。团队拥有专业知识非常关键,它对于我们在大语言模型(例如我们的法律业务)方面的成功至关重要。同样,在我们的风险业务部门有很多保险和金融专家,在我们的科技与医学业务部门有医生、护士以及博士研究生。我们投资的一个重要方面是公司在提高员工技能方面所做的努力。现在掌握这些新技能的人并不多,所以我們也在投资培训自己的员工。

励讯集团近期推出了三款生成式AI工具,分别面向律师、研究人员和临床医生。您能分享一下励讯集团是如何满足不同客户需求吗?

■开发AI应用工具,不仅要针对不同客户,还要针对不同的用例进行试验和关注,这是关键所在。例如,就Lexis+ AI而言,这是一个多模型平台,如果你是终端用户,将根据提出的问题类型,在单个传输中同时发送到Azure Open AI的ChatGPT-4和亚马逊AWS的Anthropic。我们将选择最佳的大语言模型,针对客户提出的特定问题提供最佳答案,并不断尝试不同的模型和第三方供应商。

准确性、质量和速度至关重要。我们的客户(如法律专业人士和临床医生)需要作出高价值的决策,因此对不准确或幻觉内容的容忍度为零。我们的模型是由主题专家根据可信来源的数据和内容建立和训练的。Lexis+ AI利用我们最新的庞大专有法律内容库,最大限度地降低了编造内容或幻觉的风险,并根据我们强大的法律引文工具“谢泼德”(Shepard's)检查所有引文,以确保

引文的有效性。同样,ClinicalKey AI建立在爱思唯尔医学事业部旗舰产品ClinicalKey的基础上,该产品已被全球数千家医院系统和医学院使用,以满足医疗点和深度参考的需求。

对于生成式AI,数据隐私也是我们客户最关心的问题。会话托管在一个专用、安全的云环境中,系统会在每个用户会话结束时清除所有上传的文档、查询和提示,而且不会将任何数据或内容用于在开放网络上训练核心人工智能模型。

我们对数据安全和保护隐私有着坚定的承诺,并在负责任地使用先进技术方面有着长期良好记录。我们所有的产品和服务都是在人为监督下开发的,并符合励讯集团制定的《负责任的AI原则》。

在与Open AI、谷歌、亚马逊、微软和其他大语言模型开发者的合作过程中,能分享一些保护励讯集团利益和开发AI工具的有效做法吗?

■我们在利用先进技术和分析来帮助客户解决复杂问题方面有着悠久历史,我们所有产品的内容和技术都以励讯集团的《负责任的AI原则》和《隐私原则》为基础。我们对AI技术的负责任使用将继续推动我们的增长和创新,提升我们为客户创造价值的

能力。

未来还会开发其他AI工具吗? 您能否给那些对AI工具开发感兴趣的出版商提供一些建议?

■在励讯集团,一切都始于了解客户的需求和期望的结果。他们追求更好的结果和更快、更明智的决策,这推动了我们的专注和创新。我们的方法是持续创新。我们不断开发新产品、数据和技术平台,为客户创造有效的解决方案。我们每天都会进行实验,将一些产品投放市场,看看它是否有效,并不断学习、迭代,然后进一步开发。



托马斯·明库斯



托马斯·考克斯

的原则。我们的目标不是替代人类在出版流程中的工作岗位,因此我们建议所有客户在将洞见平台生成内容用于工作流程前,一定要阅读、编辑并核对其准确性。

您如何看待及处理AI生成内容的版权所有权,以及版权内容未授权被使用等争议性问题?

■对于AI生成内容的法律地位及其版权可保护性,各个国家有不同的法律规定。这个新出现的问题目前在全球引起重视,并在法庭上引起争论。我认为,如果人类大量修改或创造性地修改了AI生成的内容,那么它应被视为衍生作品。但是,这些衍生作品的版权是否受到保护,要取决于当地的版权法以及对AI生成内容的编辑程度。目前全球各地的法律系统都在讨论这一问题。

为保护出版商的IP,“洞见”平台采用强大的加密标准,所有上传到我们系统的出版机构数据在传输和存储时都会加密。我们的核心管理技术获得了ISO 27001认证。此外,“洞见”平台通过应用程序编程接口(API)访问大语言模型。我们与合作的AI提供商已签署相关协议,承诺不会使用其API接收或生成的数据来训练未来的模型。他们每次处理完请求所需的必要时间之后也会删除数据。

您能否为那些对AI工具感兴趣或正在尝试使用的出版商提些建议?

■我会建议出版商要了解AI工具将如何使用他们的数据,并与积极保护版权的知名公司合作。

利用AI以精益敏捷方式提供专家解决方案

■斯塔西·凯伍德(Stacey Caywood,威科集团医学与全球增长市场首席执行官)

您在威科集团的主要职责是什么? 威科数字出版业务占比有多大? 与纸质出版相比,数字出版有哪些异同?

■我是威科集团医学与全球增长市场首席执行官。作为威科医学事业部负责人,我的职责是领导一家值得信赖的临床技术和循证解决方案的全球供应商,这些解决方案服务于临床医生、患者、研究人员、学生和下一代医务工作者。数十年来,威科为专业人士提供服务,帮助他们优化工作流程并作出明智决策。我们的专家解决方案结合了深度的专业知识和先进的技术与服务,为客户提供更好结果、分析和提高生产力,并解决复杂的问题。现在,威科94%的收入来自数字和专家解决方案,我们希望继续将数字信息产品转化为专家解决方案。

威科过去20年进行了数字化转型,已成为向医学界、学术界、专业共同体和科学界提供数字平台和分析工具的市场领导者。通过数字化,我们改变了产品开发及交付方式,从印刷向数字内容的转型,主要通过增加多媒体和交互式内容增强用户体验。通过专注于内容的数字化交付,我们改变了医学研究的生产、分发、获取和消费方式,适应了学术出版持续变化的格局,并满足了受众持续变化的需求。威科还提供数字工具,促进研究人员和从业者之间的合作,从而简化研究过程和研究结果的传播。

近年来威科集团对技术的投资日益增加,在聘用技术人员方面有哪些要求? 如何进行工具研发?

■在威科内部,我们的技术人员在一个专门的职能部门工作。该部门的使命是推动更快、更高效的创新,并利用我们的技术投资为客户创造价值。该职能部门目前由4500名技术人员和60多名AI专家组成,为威科集团的所有5个事业部提供可扩展的服务和技术支持。

我们设立了6个卓越中心,包括用户体验、人工智能、知识产权和专利、架构和资产重用、质量工程以及应用程序安全。这些卓越中心是我们从信息产品向专家解决方案转型战略的核心。

通过对客户的了解,结合我们值得信赖的内容和深度专业知识,我们正在利用生成式AI,以精益敏捷的方式提供新的专家解决方案。

威科是否为这些核心用户开发一些AI工具? 请分享一些应用场景和案例? 如何找到开发AI应用的技术合作伙伴? 如何在确保用户利益的同时与大语言模型开发商更好合作? 技术人员发挥了哪些作用?

■威科将AI应用到产品中已有10多年历史。新的生成式AI技术是我们利用高质量内容使客户受益的又一强大工具。今天,威科约50%的数字收入来自应用了某种形式AI的产品。

AI在我们的几个产品中得到应用,它为在复杂专业领域工作的专家带来了益处。我们在许多解决方案中使用自然语言处理、机器学习、深度学习 and 虚拟助手来增强和简化客户的工作流程,并提供新的或改进的见解。

我们的医学解决方案将AI应用于各种医疗挑战,从识别药物转移到帮助护理专业学生个性化学习。我们致力于为负责任的生成式AI设定标准,以应对诊疗一线

今年早些时候,我们宣布了UpToDate AI Labs的几个关键里程碑,这是一个协作资源,将生成式AI的力量引入威科市场领先的临床决策支持解决方案中。AI Labs现在可以访问UpToDate的全套循证临床内容和超过25个医学专科的分级建议。它是唯一的、完全由UpToDate可靠内容提供支持的

大语言模型。通过提出自然语言问题,AI Labs的功能可以提供有针对性的答案,这些答案完全来自于UpToDate持续更新的内容,避免了过时或已被取代的证据的陷阱。此外,AI Labs还可以记住查询内容并提供相关的证据引用。

我们对UpToDate的目标是,不断改进产品呈现和交付内容的方式,以最好支持诊疗过程中的知情决策。将UpToDate与生成式AI结合起来,可以帮助临床医生和患者创造价值。有了这种先进的能力,我们可以实施生成式AI,帮助临床医生做出更好、更明智的决策,在任何地方都可以提供最佳诊疗。

在与Open AI、谷歌、亚马逊和微软等公司的大语言模型合作时,如何保护威科的利益并采取有效措施开发出最有价值的AI工具?

■威科集团正在与技术领导者、医疗专家、学术合作伙伴和客户合作,共同开发AI和生成式AI产品。在这些合作关系中,为防止知识产权数据泄露,我们只与经批准的供应商合作,执行版权、合理使用和隐私条款,识别和披露任何未经授权的第三方使用生成式AI的情况,并进行第三方风险评估。

此外,威科致力于通过以下准则实现负责任的AI:一是隐私和安全。在产品和服务中(包括设计、开发和部署AI流程),确保创建的AI系统更加安全、稳妥和可靠。二是透明度和可解释性。设计和开发具有高度透明度和可解释性的AI系统,使用户能正确理解并使用该系统。三是治理和责任。遵循各项开发标准和流程,确保对AI系统及其结果负责。我们在设计、开发过程中及部署后均会进行风险管理和问题修复。四是公平。威科认识到在设计

和开发AI产品及服务时公平、无歧视地对待他人的重要性。五是以人为本。威科致力于创建以人为本的AI系统,专注于解决业务问题并使客户受益,同时考虑AI系统可能对社会和环境产生的潜在影响。

威科未来是否会开发其他AI工具? 您能否给有意于开发AI工具的出版商提出一些建议? 对于AI工具在出版业大规模应用怎么看?

■在威科集团内部,我们专注于在未来的产品中使用AI和生成AI的一些关键领域。我们认为信息数字化是未来几年整个出版行业的趋势之一。在编辑、营销和销售职能方面,我们的目标是将单调和重复的任务自动化,使我们专业且才华横溢的编辑、营销人员和销售人员更加高效,并将他们的专业知识提升到价值链的上游,更专注于客户体验,提高客户投资回报率。对于传统的纸质出版业务,我们希望从可持续发展的角度,更聚焦纸张的生产和印刷过程,以及纸质产品的生产和运输。

我们对医疗的未来非常乐观,因为前瞻性的组织将继续拥抱技术——包括数据分析、基于云的软件和AI在内的技术,以满足患者和消费者现在和未来的需求。

以最强大的加密标准保护版权和隐私安全

■托马斯·明库斯(Thomas Minkus,“真实舞台”联合创始人)

□渠竞帆(中国出版传媒商报记者)

“真实舞台”(Veristage)出版公司在ChatGPT发布不到一年的2023年8月创立,是如何赶上第一波AI的浪潮的?

■“真实舞台”是致力于为全球出版商提供AI服务的独立公司,我与公司另一位创始人托马斯·考克斯(Thomas Cox)在IPR License公司工作时结识,多年一直保持联系。考克斯当时担任IPR公司首席技术官,早就意识到AI在图书行业

的巨大潜力。我们讨论了AI对于改进出版工作流程方面的潜力之后,去年8月创建了这家公司。

我们相信,在不远的未来,那些采用AI的机构将在多个方面实现加速发展,而不使用AI的机构可能会落后于时代。

“洞见”平台是如何运作的? 目前的业务重点是什么?

■我们的核心产品是基于AI的“洞见”平台(Insight)。这个平台使用多种大语言模型,比如ChatGPT和Claude,分析文件、生成英文、法文、德文、意大利文和西班牙文的重要讯息和报告,未来我们将开发更多语言版本。平台支持从书稿评估到营销与销售的整个出版周期,可以基于不同的AI工具提取和描述文档中的图像,生成营销、宣传和销售资料,以及生成有声读物和下载报告,从而面向海外市场进行多语种版权销售。在安全系统的保障下,通过与AI对话实现与文档的交互,并将个性化写作风格用于AI生成的内容中。

我们为出版机构提供契合其业务和流程的基于AI的底层技术,还向出版商提出AI最新发展的建议,以及将AI技术智能化、负责任地融入业务中的做法。目前,我们聚焦于扩大“洞见”平台的功能,并使我们的功能与出版商的特别需求和应用场景相匹配。

“洞见”平台是如何研发AI工具的? 如何找到最合适的合作伙伴,利用好大语言模型,同时又能保护好客户的利益和诉求?

“洞见”平台的研发是将在出版业几十年来的真实体验与出版商的持续对话相结合完成的。我们了解出版商的需求和面临的挑战。“洞见”平台的AI赋能的工具可以精简业务流程,提升工作效率。但是真正的潜力在于出版商拥抱AI时出现的新的业务机遇。我们正在将“洞见”平台开发成出版商的战略盟友,了解流程中真正遇到的阻碍,同时也能看到AI为出版业呈现的令人激动的新世界。

我们很早就决定,要把市场领先的多个大语言模型整合到“洞见”平台的文件处理和对话环节。主要基于两个原因:一是每个大语言模型在处理不同的任务方面各有特色,我们希望平台能使用每个模型的最佳功能。二是我们希望提供给出版商访问各种大语言模型的安全通道,这样他们也可以熟悉最新的AI可用技术。

我们已经筛选出一些不会合作的AI提供商,因为他们的服务条款和数据处理的做法不符合我们内部对于保护出版商数据的标准。

“洞见”平台对用户采取什么样的商业模式? 出版商有何反馈?

■我们实行基于用户数量和出版机构规模的订阅制。此外,我们还提供定制化的企业服务。目前已有多家出版机构在使用“洞见”平台,包括法律类、少儿类和综合类出版机构。我们从用户那里得到了非常积极的初步反馈。一旦人们开始将AI融入日常工作,就会发现AI在提升工作效率方面能发挥多大的作用。

AI可能带来虚假、错误、偏见甚至幻觉等方面问题,你们如何应对此类问题? 是否因此增加技术人员?

■我们的平台将使用出版商内容作为减少潜在错误及不真实信息等问题的基础。我们与领先的AI提供商合作,这些提供商正致力于制定开发负责任的、安全的大语言模型