Deloitte.

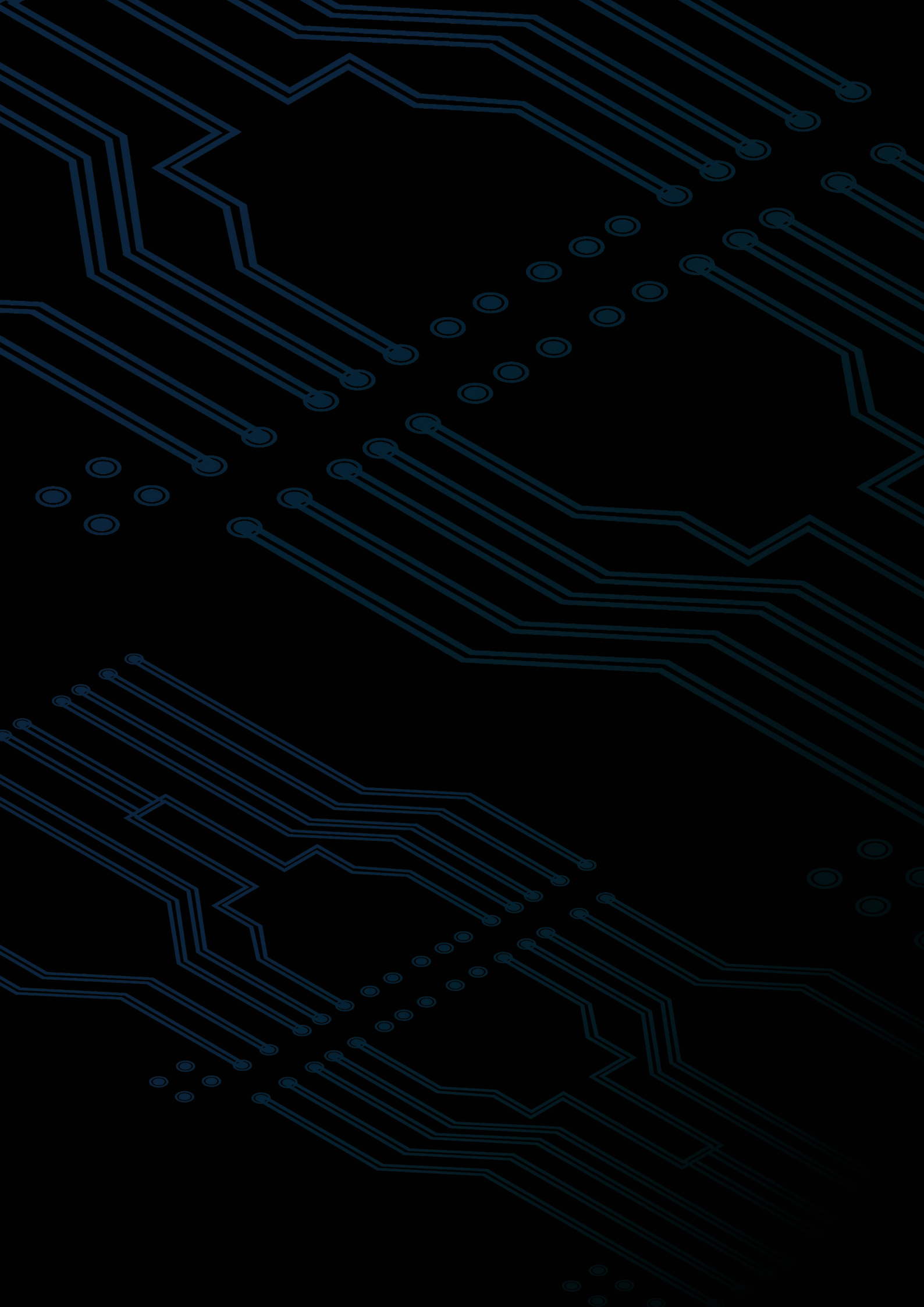
德勤



认知智能大门

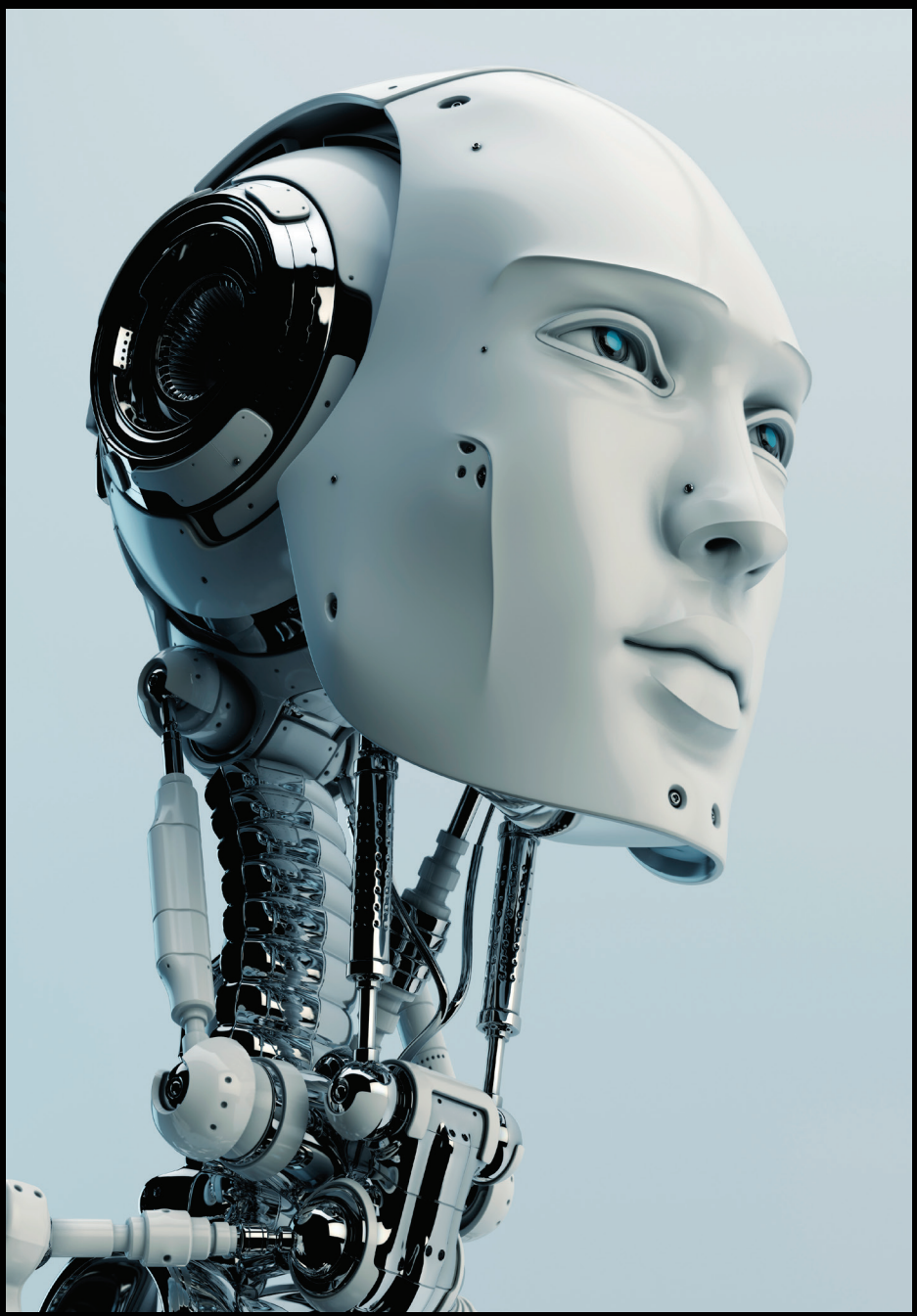
企业如何构建NLP能力

因我不同
成就不凡
始于1845



1 引言

当你看到下面这张照片，你会想到什么？机器人能听懂我说话吗？机器人会像人一样喜怒哀乐吗？这是大家对人工智能(“AI”)最初的幻想，也是AI研究的终极目标。





1.1 得语言者得天下

NLP（Natural Language Processing, 自然语言处理）是AI的子领域，同时涉及语言学和计算机科学，被比尔·盖茨称作“AI领域皇冠上的明珠”。微软全球执行副总裁沈向洋认为在AI领域“得语言者得天下”。OpenAI(2015年成立的一家人工智能公司)开放了GPT-3的访问接口，它可以根据描述生成HTML布局和代码、创建UI设计、生成机器学习代码，将口语化描述转化为法律语言，甚至可以根据关键词进行文学作品的创作，让人们体验到AI的强大之处。

在中国新基建背景下，NLP将迎来更快发展。以AI为代表的新技术基础设施将进一步支撑传统基础设施转型升级。NLP作为AI的核心技术之一，已经成为企业在认知智能时代实现转型升级的关键。

清华大学刘知远教授认为，当今深度学习和知识图谱是NLP技术发展的两大引擎。深度学习技术能够让计算机具备思考的能力。知识图谱能够提供计算机进行思考的知识基础。在知识图谱加持下，计算机已经能进行简单的知识推理和认知，在智能问答、智能搜索和智能风控等方面崭露头角。融合知识图谱的NLP技术已经打开了计算机实现认知智能的大门。



1.2 NLP为企业带来的价值

对于各大企业来说， NLP到底能带来什么？NLP技术是企业在认知智能时代实现转型升级的关键技术，主要用于解决人与人的交互问题、企业间的交互问题、硬件与人的交互问题。企业服务在发展的过程中，唯一不变的追求是“降本”，“增效”与“提质”，NLP能够为企业提供：

数据驱动的业务洞察：通过企业内外服务积累的海量结构化数据和非结构化数据，洞察客户需求和业务痛点，帮助企业找到需要补足的短板，提升业务质量。如通过对互联网上的新闻文章、论坛评论以及客户反馈等海量文本数据进行分析，可以抽取与目标企业的核心话题，分析出其竞争对手的产品策略和发展趋势，了解到客户对企业及其产品的舆论评价等。

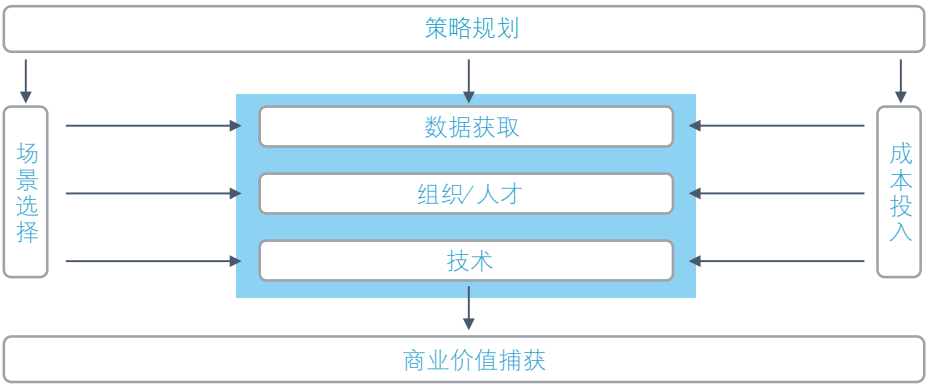
交互客户服务：智能化的交互方式打造全流程无间断的客户服务系统，提高用户体验，并通过分析交互反馈，聆听客户声音，解决客户关心的问题。如当下典型的应用“智能客服机器人”，代替人工座席，帮助回答90%的问题，大大降低客服成本。目前德勤DS COE团队自主研发的Bot-Alpha机器人工厂能够支持业务人员根据自己的业务需求，通过图形界面的方式迅速创建自己的机器人，并应用到实际业务场景，能帮助企业迅速提高运营效率。

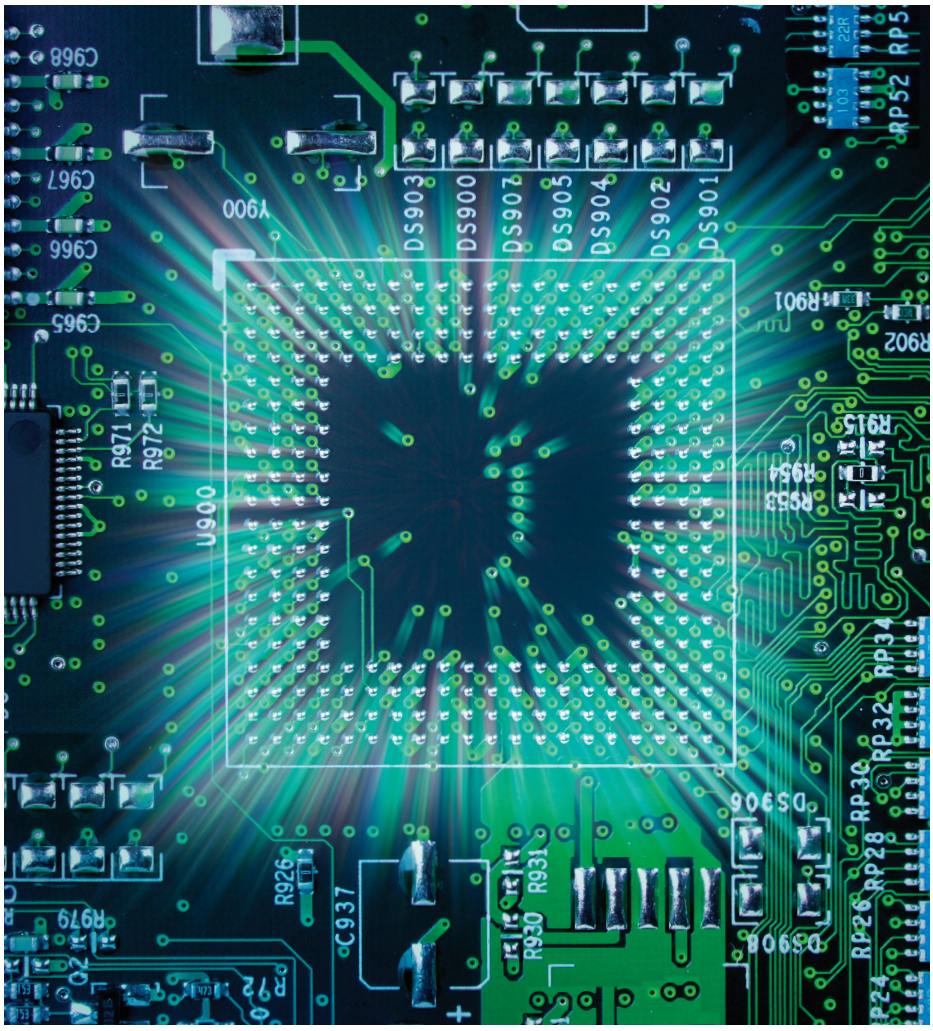
高效运营工具：通过打造智能化的运营工具，帮助企业缩短业务流程，实现业务自动化。在企业运营过程中，很多时候交互需要涉及到语言和文字信息，利用NLP技术可以辅助企业做到文档处理的智能化。如企业运营中会涉及到很多文本合同拟定、校验、审核和归档等，利用NLP技术可以实现自动对文档进行结构化信息抽取，进行合同风险点的识别，法律合规风险的评估以及与标准合同模板的比对等。

重复性劳动的替代：通过AI机器人代替重复性较高的工作，降低企业运营成本。NLP与RPA进行结合是当前AI在企业进行运用的一大亮点，利用NLP技术识别出任务，然后再利用RPA执行任务，可以极大代替重复性劳动。如发票报销是一个耗时耗力的工作，可以利用OCR+NLP技术识别出发票内容和报销目的，然后利用RPA进行报销内容的填写和报销流程的处理，能够完全实现发票报销的自动化。

1.3 NLP能力建设成功的六大要素

NLP在企业的落地过程中有诸多挑战，本文将从策略规划、场景选择、成本投入，数据获取，组织及人才以及技术六个方面，分别阐述并提出应对方案。





2 策略规划：结合自身业务实际，制定NLP能力建设战略

在底层数据与技术基础未扎实的情况下，NLP能力建设规划不可盲目求快、求全。制定策略规划时，需要综合考虑企业NLP能力、业务发展重点和数字化现状，科学有序地建设企业NLP能力。

2.1 企业NLP能力分成“能用”、“好用”和“智用”三个层次

制定企业NLP能力建设战略时，需要根据获取的难易程度认识NLP的能力层次。企业NLP能力可以分成“能用”、“好用”和“智用”三个层次。需要在进行NLP能力建设的时候，按照逐层递进的路径开展实践。

能用：具有基础的开发环境，能针对特定场景定向开发，使企业具有NLP应用定向开发能力。企业可在此阶段对具体的业务场景进行“一对一”的智能提升，初步享受NLP能力带来的红利。

好用：具有统一的NLP开发平台，面向所有业务场景进行统一开发部署，实现模型复用，使企业具有NLP平台化开发能力。能够统一NLP技术能力，并不断丰富模型以快速响应业务需求。

智用：具有统一的知识服务平台，将企业所有结构化和非结构化数据进行融合，实现知识复用，使企业具有NLP知识服务能力。企业在此阶段能够源源不断地将知识渗透到企业各方，实现知识驱动业务创新。

2.2 NLP的实施路线图应该从“远处着眼，近处着手”

明确了NLP能力层次后，企业可根据自身业务情况，制定分步执行路线图。首先要确定企业应用NLP技术的愿景，然后根据自身业务特点与NLP技术结合。具体路线图可以参考：

第一步，规划远期目标，对比愿景和当前的技术能力，分析全业务线，规划NLP的潜在场景。

第二步，NLP开发的基础设施建设，选择典型业务场景，进行应用定向开发。

第三步，搭建AI平台实现NLP应用开发的平台化、自动化和全生命周期管理，实现NLP模型的共享。

第四步，搭建知识中台，基于NLP和知识图谱，整合内外部数据，统一企业知识服务。

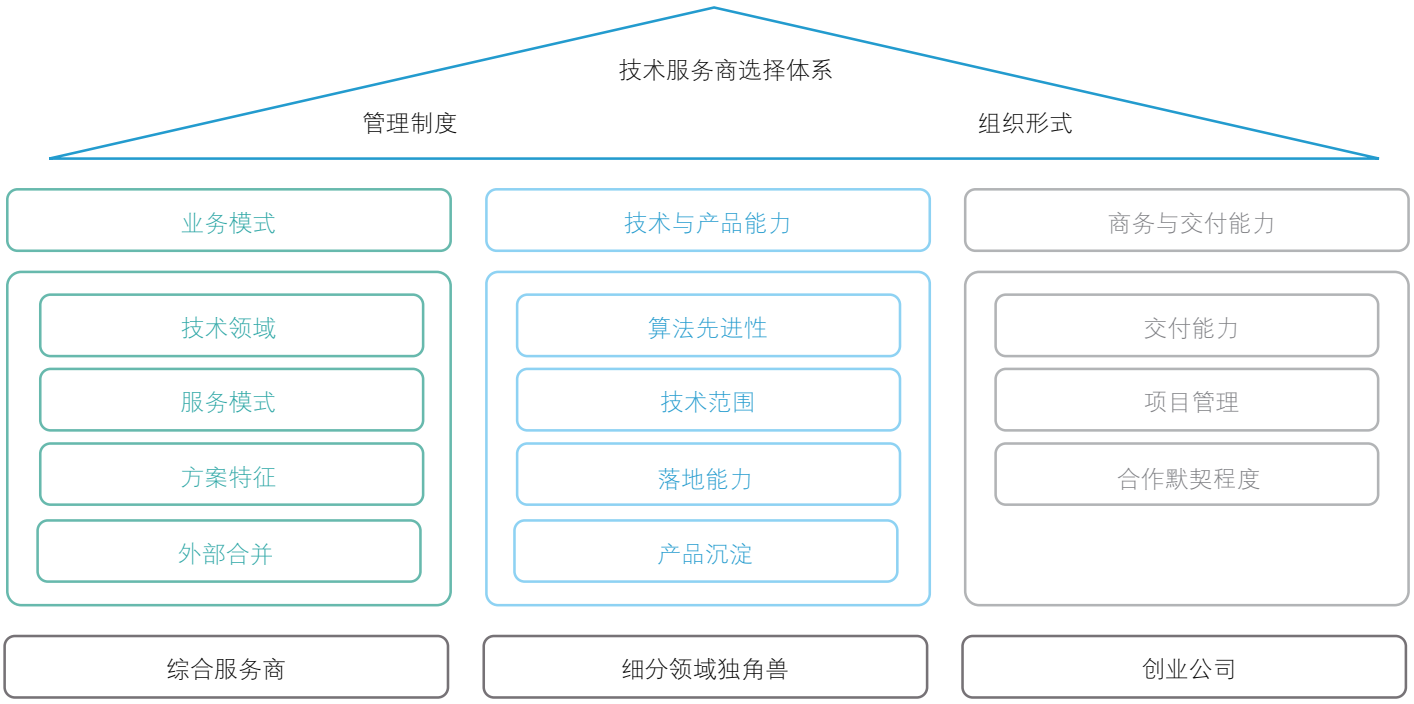
第五步，全面实现知识驱动业务创新的发展模式，将知识服务渗透到企业生产运营决策的细节中。



2.3引进合适的技术服务商合作共建才能又快又好的建设NLP能力

由于NLP技术的复杂性和专业性，传统企业往往缺少技术积累。另外，每个企业的业务领域和商业模式千差万别，企业也难以依靠软件购买的方式一次性完成NLP能力的建设。因此，需要找到合适的技术服务商，与企业共建NLP能力。

然而，NLP技术服务商鱼龙混杂，选择和管理合作伙伴是一个挑战。选择时，需要运用有效的方法论对服务商进行评估，帮助企业建立一个稳定的生态圈，以避免合作伙伴频繁更换而造成损失。本文从技术服务商的业务模式、技术与产品能力以及商务与交付能力三个维度，构建了完整的技术服务商评估体系，见下图。



3 场景选择：典型业务场景应用是NLP能力构建实践的最佳起点

3.1全面梳理业务场景，构建潜在的NLP应用场景库

应用场景挖掘是AI项目成功的关键因素。战略定点完成后，需要NLP技术部门牵头，与业务部门一起整理潜在的NLP应用场景，最终形成NLP应用场景开发部署蓝图。在场景梳理的过程中，若没有NLP技术人员参与，可能会存在场景过度乐观、脱离技术或者漏掉支持场景的情况。

下图是某消费品企业涵盖企业全价值链的NLP潜在应用场景举例：

研发管理
- 行业发展趋势
- 消费者意见洞察
- 竞争对手动态
- 产品概念挖掘

生产制造
- 维修经验检索
- 障碍关联分析

供应链管理
- 单据识别
- 地址对齐
- 合同信息抽取

营销管理
- 热门事件聚集
- 消费者反馈分析
- 客户画像

售后服务
- 客服机器人
- 客诉分析
- 公司舆情监督

3.2 选择典型业务场景进行试点开发

典型业务场景需要企业结合业务价值和技术可行两方面进行选择。业务价值需要关注预期收益、投入成本、业务边界以及部署阻力；技术可行需要评估技术难度、数据基础、计算资源和开发周期。综合考虑以上要素后，企业可以选择1-2个典型业务场景进行试点开发。具体步骤可以参考如下：

概念验证：根据当前的业务需求和资源基础，规划较小的实施范围，进行POC开发以完成概念验证，原则上不超过3个月。

分期上线：将长期的业务目标进行阶段划分，每个阶段以层层递进的方式进行功能升级，每个阶段需控制在3-6个月。

经验总结：每个阶段都需要进行经验总结，并在项目完成后形成企业独有的NLP项目实施方法论和流程。

4 数据获取：持续的数据更新是项目成功的基础

数据数量和质量是AI项目成功实施的基础，NLP项目也不例外。保证数据数量和质量需要做好数据需求评估、数据持续供给和数据质量监控。

数据需求评估需要特别注意数据的获取成本、技术难度和法律风险。尤其是NLP数据缺少通用的数据集，在针对某项业务进行数据获取时，需要综合考虑，准备数据标注和数据增强的工作。避免数据获取方案脱离成本和技术限制，或者有合规风险。

数据持续供给需要形成完整、流畅的数据供给机制，包括数据的收集、清理、标注、加工和集成等环节。具体执行中，首先要调研数据现状，结合可视化的方式，形成数据的标注方案，获取模型启动的第一批数据。其次，根据第一批数据，得到基线模型，依据模型表现，对数据标注进行调整，持续优化模型训练和数据标注工作，获得达到上线标注的NLP模型。NLP模型上线以后，需要收集用户反馈，形成自动化的数据反馈和标注机制，持续学习。

数据质量监控需要融合在数据持续供给的过程中，可以通过数据分析和数据采样检查的方式，发现数据质量的具体问题，并提出改进的方法。



5 结合NLP能力的层次打造稳步推进的技术路线

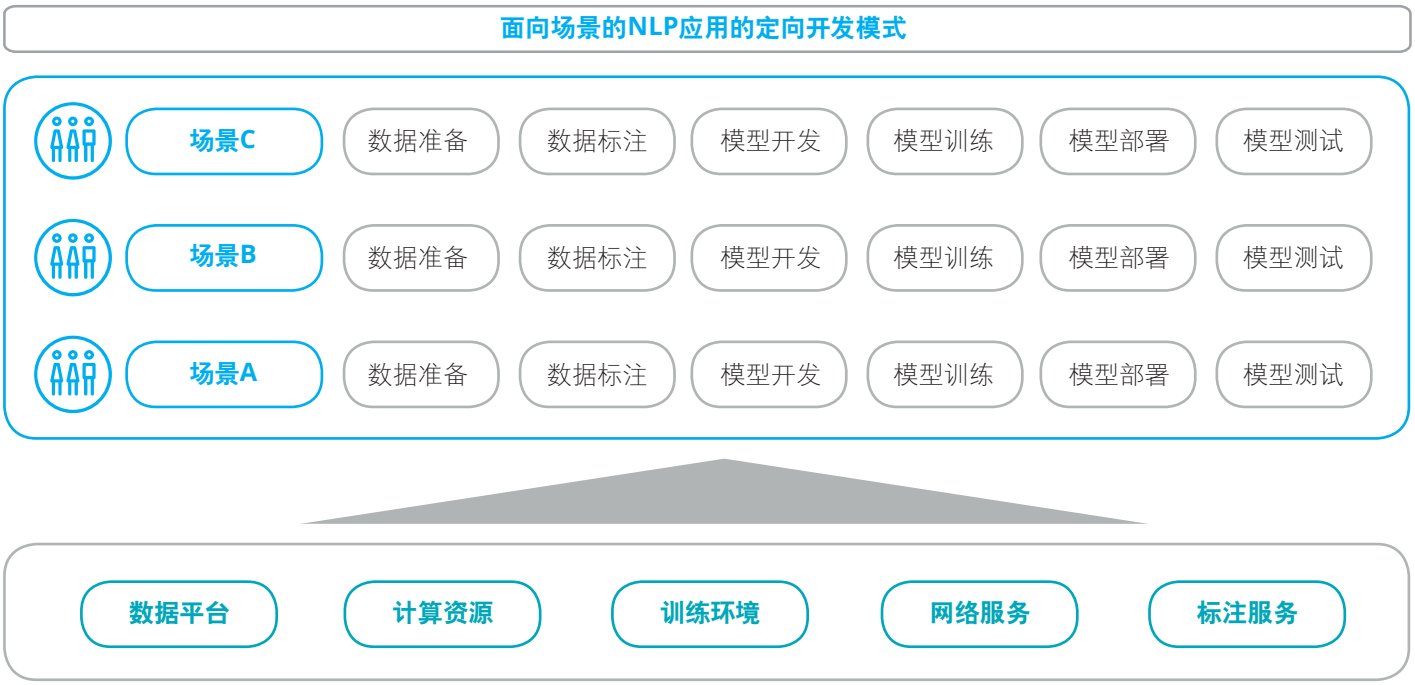
前文提到NLP能力分成“能用”、“好用”和“智用”三个层次，本部分将对实现三个层次分别给出具体的技术行动方案。

5.1 面向特定场景需求，进行NLP模型的应用开发是NLP能力建设的初级阶段

在NLP初级阶段，企业应该专注NLP基础建模能力的建设，包括软、硬两方面的实力。

硬实力方面，首先建设数据平台，实现多方数据的接入、存储、管理和统一的对外服务。其次，准备计算资源，通过购买物理GPU服务器或租用云服务器。接着，搭建NLP模型训练环境，在GPU服务器上搭建深度学习框架、NLP算法库和模型库。再次，搭建网络环境，实现线上模型发布。最后，通过自建或者外包的方式，搭建标注服务。

软实力方面,应组建核心NLP技术团队，研究技术发展的最新进展，掌握最新技术，并深入企业业务场景，透析NLP技术的应用点。该阶段会形成如下图的开发模式，NLP技术人员可以在基础环境下针对特定的业务需求进行定向开发。



5.2平台化的NLP开发，是一个企业NLP能力成熟的标志

初级的NLP定向应用开发，支持企业对特定业务场景的NLP应用，但这种模式存在如下问题："烟囱式"开发导致项目之间缺少共享、不易集成，重复开发、成本高，能力缺乏沉淀；底层资源缺少统一的资源调度和管理平台，容易造成资源浪费；缺少统一的模型发布和监控平台，模型管理难。

基于AI平台开发NLP的方案应运而生，其核心目的是打通多条业务线之间的开发需求，为业务层打造一个共享的NLP模型服务，并支持平台化、自动化和全周期管理。下图给出了一个AI平台（NLP部分）的建设功能架构示例：



NLP应用的开发上线，涉及到业务支持，数据处理，模型学习和模型监控四个方面。平台支持按业务需求设计实施方案，通过模板来规范方案设计。平台支持统一的数据获取、数据标注、特征工程以及数据集划分等，以数据处理服务的方式统一进行发布。模型学习包括线下模型训练、模型评估以及可复用模型库、算法库管理。模型监控包括模型自动部署运行、性能监控和对外服务接口管理等。

99 5.3 知识服务代表着企业NLP能力驱动业务创新的未来

企业如何从海量数据中获取信息和知识，最终达到企业智慧，决定了企业最终的智能化程度。知识中台通过整合企业内外多方面的数据，并进行统一加工处理形成企业的知识，以知识服务的方式支撑企业的智能化运营和业务创新，支撑企业决策的每个细节，快速响应内外部的变化。

NLP和知识图谱是构建企业知识中台的关键技术，其核心功能包括知识表示、知识生产、知识加工和知识应用四个部分。下图给出了知识中台的功能架构示例：



6 有效的组织和丰富的人才资源是企业NLP能力建设的保障

6.1 避免NLP团队沦为业务部门的技术外包

利用NLP技术进行智能化的企业转型，不仅是一种技术的引进应用，更多的是企业思维和运营模式的变化。NLP项目的实施过程中会遇到各种困难或阻力，包括企业对NLP能力水平的误解，项目实施牵扯的各方利益再分配以及项目投入和预期之间的不匹配等。需要建立一种自上而下的项目推进制度，成立由CIO，甚至CEO直接领导的部门，由该部门统一协调各个业务部门完成项目，避免NLP团队成为业务部门的外包开发人员。

6.2 NLP团队建设需要以内部培养为主

在NLP技术团队建设方面，团队至少应包含行业专家、NLP数据科学家、NLP算法工程师和NLP产品经理四个角色。其中行业专家负责提供特殊行业的业务知识；数据科学家基于业务需求开发相应的NLP算法；算法工程师负责总体的系统架构，功能开发和NLP模型上线工作；产品经理负责理解业务需求，制定并推进开发计划，保证项目交付。

针对四类人才需求，团队建设需要结合人才招募和内部培养。然而，目前AI人才匮乏和自身业务的独特性，企业更需要加强人才内部培养，通过项目实施“以战代练”的方式进行团队的磨合和培养，打造适合企业自身业务的技术团队。

7 成本投入：有计划的持续投入和良好的成本管控是NLP能力建设的基础

NLP能力建设的战略实施需要有配套的资金与时间投入。在NLP能力建设的初期，需要搭建软硬件资源，组建具有专业能力的技术团队，同时引入技术服务商。在项目实施过程中，需要获取相关数据，对数据进行标注和处理，优化算法的性能，以达到业务部门的实际需求。在项目上线运行后，需要持续监控算法性能，并引入自学习机制不断优化。最终，搭建AI平台，更好地支持NLP应用的规模化开发。

为了保证NLP能力建设的可持续性，所有的花费都需要进行提前规划和审查。首先需要进行详尽的成本核算，细分成本类型，明确成本结构，避免资源的重复投入；其次，要建立项目预算的审核审批制度，明确审批人责任；同时，充分调研市场，获取合理的价格信息；最后，制定合理的KPI考核，激发团队的战斗力，提高项目实施效率。



8 企业成功案例分享

在利用NLP技术推进企业智能化方面已有很多成功落地案例，并帮助企业极大提高了生产效率，例如：

AI智能培训：某国际手机生产厂商在全国有2000多家门店，3万多名门店导购人员，由于门店导购人员相对不太稳定和厂商培训内容经常发生变化，传统通过人工培训的方式，很难满足课程内容实时更新和保证培训效果的需求，该企业通过部署智能培训机器人，利用机器人扮演多种顾客角色，并将培训内容编排到培训机器人知识库的方式，成功实现了7*24小时实时培训和培训内容迅速更新的目标。

兴趣发现与推荐：某国有大型商旅集团通过对其旗下历史积累的2万多条线路的宣传文稿，利用NLP技术对其进行分析，构建了具有2000多个标签的旅游特色标签体系，很好地描述了其旗下产品每条线路的旅游特色，并利用该标签体系结合历史数据，能够准确的描述每个旅游用户的兴趣，从而实现了对旅游用户做到千人千面的产品推荐和旅游服务。

业务尽调及审查审批：对于某大型国有商业银行的小额信贷业务，通过搭建基于知识图谱的信贷业务风控平台，将原有来自多业务、多渠道、多管理系统的复杂多元异构数据，抽离成符合银行自身业务特色的各类实体、关系及属性。结合指标、规则、模型体系，对客户准入、尽调强化、违约识别、不良预警、组合风险管理等业务场景提供了多维度分析。授信业务办理时间缩短至30分钟左右；显著提升贷后水平，将不良贷款率控制在4%以内，相较平台上线前至少降低了70%。实现了银行普惠金融场景下各类信贷业务贷前、贷中、贷后的全面覆盖，有效增强授信业务全流程的风险管理能力。



9 结语

AI发展今天，已经走过了计算智能和感知智能阶段，开始迈入认知智能。NLP技术作为认知智能的核心，

能够帮助企业转型升级，驱动企业业务创新，享受知识服务带来的红利。企业从“能用”到“好用”再到“智用”的NLP能力建设过程中，需要“远处着眼，近处着手”，先进行战略定点，后逐步开展场景定向开发能力建设，AI平台建设和知识服务建设，通过自上而下推进的方式和持续人才培养的战略，保证项目的顺利推进。同时，谨慎合理的选择恰当的技术服务商也关系到项目的成败。



联系我们



尤忠彬
德勤管理咨询中国数据科学卓越中心
领导合伙人
电子邮件：zhyou@deloitte.com.cn



王鹏
德勤管理咨询中国
总监
电子邮件：wangpeng@deloitte.com.cn



Deloitte（“德勤”）泛指一家或多家德勤有限公司，以及其全球成员所网络和它们的关联机构（统称为“德勤组织”）。德勤有限公司（又称“德勤全球”）及其每一家成员所和它们的关联机构均为具有独立法律地位的法律实体，相互之间不因第三方而承担任何责任或约束对方。德勤有限公司及其每一家成员所和它们的关联机构仅对自身行为及遗漏承担责任，而对相互的行为及遗漏不承担任何法律责任。德勤有限公司并不向客户提供服务。请参阅 www.deloitte.com/cn/about 了解更多信息。

德勤是全球领先的专业服务机构，为客户提供审计及鉴证、管理咨询、财务咨询、风险咨询、税务及相关服务。德勤透过遍及全球逾150个国家与地区的成员所网络及关联机构（统称为“德勤组织”）为财富全球500强企业中约80%的企业提供专业服务。敬请访问 www.deloitte.com/cn/about，了解德勤全球约330,000名专业人员致力成就不凡的更多信息。

德勤亚太有限公司（即一家担保有限公司）是德勤有限公司的成员所。德勤亚太有限公司的每一家成员及其关联机构均为具有独立法律地位的法律实体，在亚太地区超过100座城市提供专业服务，包括奥克兰、曼谷、北京、河内、香港、雅加达、吉隆坡、马尼拉、墨尔本、大阪、首尔、上海、新加坡、悉尼、台北和东京。

德勤于1917年在上海设立办事处，德勤品牌由此进入中国。如今，德勤中国为中国本地和在华的跨国及高增长企业客户提供全面的审计及鉴证、管理咨询、财务咨询、风险咨询和税务服务。德勤中国持续致力为中国会计准则、税务制度及专业人才培养作出重要贡献。德勤中国是一家中国本土成立的专业服务机构，由德勤中国的合伙人所拥有。敬请访问 www2.deloitte.com/cn/zh/social-media，通过我们的社交媒体平台，了解德勤在中国市场成就不凡的更多信息。

本通讯中所含内容乃一般性信息，任何德勤有限公司、其全球成员所网络或它们的关联机构（统称为“德勤组织”）并不因此构成提供任何专业建议或服务。在作出任何可能影响您的财务或业务的决策或采取任何相关行动前，您应咨询合资格的专业顾问。

我们并未对本通讯所含信息的准确性或完整性作出任何（明示或暗示）陈述、保证或承诺。任何德勤有限公司、其成员所、关联机构、员工或代理方均不对任何方因使用本通讯而直接或间接导致的任何损失或损害承担责任。德勤有限公司及其每一家成员所和它们的关联机构均为具有独立法律地位的法律实体。

© 2021。欲了解更多信息，请联系德勤中国。
Designed by CoRe Creative Services. RITM0625301

这是环保纸印刷品