四．现状与规划

4.1 人工智能发展现状

人工智能最早能够追溯到1936年，英国数学家AM. Turing在论文《理想计算机》中提出了图灵机模型，然后1956年在《计算机能思维吗》一文中提出机器能够思维的论述（图灵实验）.之后计算机的发明和信息论的出现为人工智能发展奠定了良好的基础.1956年在达特茅斯会议上， Marvin Minsky、 John Mccarthy等科学家围绕“机器模仿人类的学习以及其他方面变得智能”展开讨论，并明确提出了“人工智能”一词。

人工智能的发展经历了2次发展热潮。第1次是1956-—1966年，1956年，Newe和 Simon在定理证明工作中首先取得突破，开启了以计算机程序来模拟人类思维的道路；1960年， McCarthy建立了人工智能程序设计语言LSP.上述成功使人工智能科学家们认为可以研究和总结人类思维的普遍规律并用计算机模拟它的实现，并乐观地预计可以创造一个万能的逻辑推理体系。第2次是20世纪70年代中期至80年代末，在1977年第五届国际人工智能联合会会议上，Feigenbaum教授在特约文章《人工智能的艺术：知识工程课题及实例研究》中系统地阐述了专家系统的思想并提出“知识工程”的概念。至此，人工智能的研究又有新的转折点，即从获取智能的基于能力的策略变成了基于知识的方法研究。此后，人工智能的发展进入平稳发展期。

近些年，大数据时代的到来和深度学习的发展象征着人工智能的发展迎来了第3次发展热潮.1997年，IBM的深蓝（ Deep blue）机器人在国际象棋比赛中战胜世界冠军卡斯帕罗夫，引发了人类对于人工智能的思考.2016年英国初创公司 DeepMind研发的围棋机器人 AlphaGo通过无监督学习战胜了围棋世界冠军柯洁，让人类对人工智能的期待提升到了前所未有的高度，在它的带动下，人工智能迎来了最好的发展时代.2019年，上海举办了世界人工智能大会，会议集聚了全球人工智能领域最具影响力的科学家和企业家以及相关政府的领导人，围绕人工智能领域的技术前沿、产业趋塾和热点问题发表演讲和进行高端对话，开启人类对于人工智能发展的新一轮探索[1]。

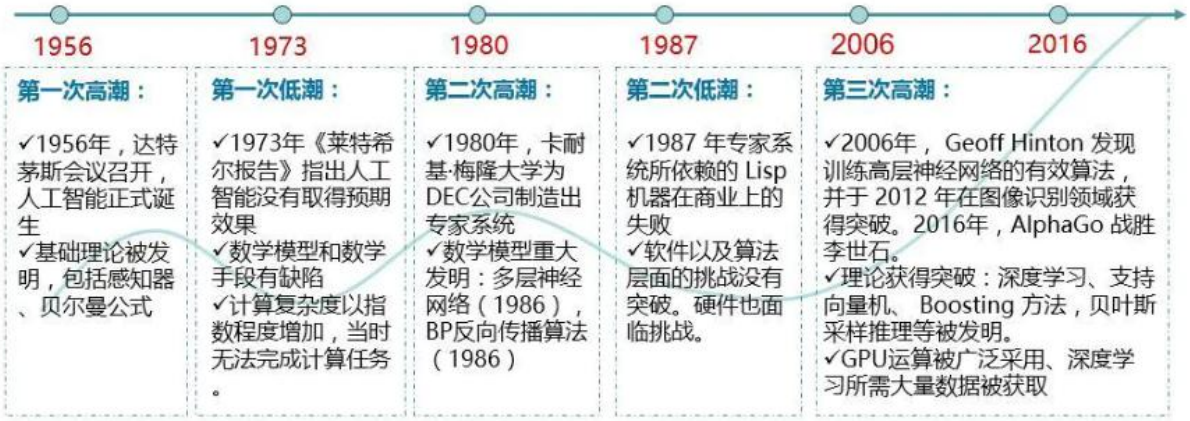

图 1：人工智能发展浪潮

4.2 知识图谱发展现状

知识图谱自 2012 年推出以来,进展迅速,已经成为大数据时代的重要知识表示之一,极大地推动了智能化的发展进程。目前知识图谱技术已经在大规模简单应用场景中取得了显著效果。但近年来,知识图谱的需求从数据丰富的大规模简单应用场景转向专家知识密集但数据相对稀缺的小规模复杂应用。这一转向过程给知识图谱带来了新的挑战。

4.2.1 知识图谱实现功能

首先从知识表示层来看，知识图谱的研究和落地，现在只是完成了大规模简单应用所需要的表示。知识图谱本质上是大规模语义网络。知识图谱首先是一种大规模知识表示，所以它通常包含海量的实体，往往是数以亿计。大规模也体现为多样的关系，成千上万的关系。正是因为它规模大，往往需要做出质量妥协，所以很多时候知识图谱也允许出错。现在没有人敢说自己数千万、数亿规模的知识图谱百分百正确，永远是99.999%，允许错误。也允许 schema 不完善，从而包容更多实例，精良的模式在很多图谱里面是缺失的。语义网社区投入巨大精力推动通用 schema 的建设，但是遇到很多挑战。

它支撑的应用，大部分是简单应用：以实体（词汇）为中心的知识表示，表达的往往是实体的属性和关系；它的推理极为简单，往往都是基

于路径或者上下位词的简单推理，以及基于分布式表示的推理。所以知识图谱这几年的发展，解决了大规模简单应用的场景。

其次实现了简单推理。符号知识存在的根本价值在于能做推理。当前知识图谱的大部分推理是简单推理例如，用户搜索周杰伦，很多平台给用户推荐他的歌。这是因为知识图谱知道刘德华是歌手，因此一定会有相应歌曲。这是基于上下位关系推理。搜索唐太宗，推荐李世民，这是同义关系推理；搜索战狼 1，那么平台可能会推荐战狼 2。因为它们都是同类型的电影，并且是同一个导演、同一个主演，这是基于路径的推理。

现实中大部分应用利用这些简单推理就能解决，并且即便只用这种简单推理也能解决很多以前搜不到、问不清的痛点问题，并且效果显著。大家现在看到的很多应用场景、应用知识图谱所解决的根本问题，都是搜索、推荐和问答。

4.2.2 知识图谱瓶颈

而最近两年最大的变化就是我们面临着应用场景的变换。我们正在从大规模、简单的应用场景向小规模、复杂应用场景切换。知识图谱的前期应用场景都是以 BAT、TMD 为代表，它们属于大规模简单应用场景，模式单一，其应用的知识是众人皆知的。但是现在越来越多的是石油、能源、工业、医疗、司法、金融这种小规模复杂应用场景，它有着密集的专家知识、有限的数据资源和深度的知识应用等鲜明特性，这都是新场景给我们提出的全新挑战。这与知识图谱在互联网应用中用到的衣食住行这类通用知识显著不同。这一新的形势对于获取隐性的专家知识提出了新挑战。一方面专家知识往往是隐性的，难以直接从文本中抽取。另一方面，专家知识有着一定的门槛，只有少部分行业从业人员才能完成专家知识的众包工作。除此之外，在盘点数据的时候，会发现大部分的场景数据是稀缺的。首先领域数据本身就稀缺。其次还缺乏高质量的标注数据。我们很多机器学习模型需要标注数据，哪怕有资金可以投入人力标注，但是领域任务往往是不明确的，而专家资源又很昂贵，那么标

注也会非常困难。如果不采用人工标注，而利用外界爬取的数据进行融合，也会十分困难，因为领域数据融合代价通常也非常大。所以总体上来讲，虽然很多时候我们觉得有大数据，但是相对于很多领域智能化应用而言，我们的数据还是十分“贫乏”。[2]

图 2：知识图谱应用场景转变

因此我们考虑到人工智能，是当前最热门研究专业领域之一，其相关方向的人才匮乏也正越来越成为（市场）关注的议题，而在培养人才时，如何准确把握所授相关领域知识的准确性、全面性与前沿性成了一个难题。而与此同时当前的知识图谱也存在着无法在专业领域得到有效应用的问题。所以团队选择构建一个面向学习者尤其是本科生的人工智能领域的垂直知识图谱。人工智能领域繁多，我们选取机器学习、自然语言处理与机器视觉等三个领域作为代表。

4.3 产品现状

我们目前已完成人工智能中机器学习、自然语言处理与机器视觉这三个比较热门的三个领域的知识图谱。用户可以使用我们的产品对这三个领域的相关知识进行检索，我们也会针对用户输入的关键词进行扩展，展现给用户与其输入的关键词相关联的知识。同时对用户界面进行优化，满足不同用户对知识表示方式的需求。

4.3.1 产品成本

我们的知识获取主要是基于国内的科技论坛网站，用 Python 语言编写爬虫程序进行自动化获取的。这些论坛网站的讨论基本上是与当前人工智能领域的最新发展内容息息相关的，从而可以保证用户能够得到最前沿的信息。为了能够获取大量的知识并进行相关的存储同时要保证产品的反应速度，我们需要对电脑进行不同程度的升级，但这些花销即可获取大量的数据。相对而言，成本是非常低的。

4.3.2 产品功能

首先我们利用知识图谱使得大规模自动化知识获取基本可行。针对人工智能这一领域，我们基本实现了从数据获取->知识抽取->知识融合三个环节的自动完成。

其次我们利用知识图谱完成了许多元数据之间的关联。比如，搜索人工智能时，其往往可以表示为 AI，这样一种关联就可以告诉我们这两个字段是可以匹配的，而关联就能创造价值。所以，我们利用知识图谱作为数据融合的指引，当在搜索框内输入关键词并点击搜索后。主光圈即为输入的关键词，而周围的光圈即为其关联得性质与详细信息。因此学习人工智能的学生通过一个关键词就可以了解到多方面的知识。

同时我们利用知识图谱解决了语言表达鸿沟问题。很多时候用户所提供的搜索关键词与我们提前存在数据库里的词汇表达是有一定的差异的，特别是对于初学者。另外不同专业的人在对人工智能中同一件事情的描述所使用的语言极有可能是不一样的。而与此同时有些实体本身就有若干种说法。我们通过建设大量词汇知识图谱，包含领域的同义词、缩略词、上下位词等关系，有效解决语言表达鸿沟的问题。

相较于传统的以简单的知识应用与常识为基础的知识图谱，我们实现了能应用于专业领域，方便学习的知识图谱。现在越来越多的高校开设人工智能专业，同时国家也在这一领域投入大量资金。根据教育部在 2020 年 2 月份公布的 2019 年度普通高等学校本科专业备案和审批结果，据统计中国

人民大学、北京化工大学、北京邮电大学、北京师范大学、中国传媒大学、复旦大学等 180 所高校新增人工智能本科专业。这是人工智能（AI）本科专业被纳入我国本科专业的第二年，去年仅有 35 所高校获批，今年这一数量涨势迅猛，超过去年的 5 倍。人工智能的热潮越来越高，而且人工智能方面的人才也非常的少，所以这是很多高校开设人工智能专业的原因。我们当前的产品可以供学子们进行人工智能相关内容的学习，也能够根据学子们的搜索关键字频率，将当前最热门的内容展现给他们。一定程度上也有助于人工智能的推广与发展。

4.3.3 产品价值

我们的知识图谱作为一种语义网络拥有极强的表达能力和建模灵活性：首先可以对现实世界中的实体、概念、属性以及它们之间的关系进行建模；其次，知识图谱是其衍生技术的数据交换标准，其本身是一种数据建模的“协议”，相关技术涵盖知识抽取、知识集成、知识管理和知识应用等各个环节。

同时我们的产品作为一种特殊的图数据。其中每个结点都有若干个属性和属性值，实体与实体之间的边表示的是结点之间的关系，边的指向方向表示了关系的方向。非常的直观美化，对于用户没有高的要求，使得任何人，都可以通过我们的产品查阅人工智能领域的相关资料，都可以进行相关的学习。

其次，我们的产品采用了人类容易识别的字符串来标识各元素；图数据表示作为一种通用的数据结构，可以很容易地被计算机识别和处理。产品的可扩展性良好，技术路线已经完成，针对不同的应用场景，更改数据源即可完成新的应用。

对于知识图谱如何应用于专业领域的这一问题，我们根据自己的创新性技术路线给出了回答。考虑到近两年应用场景正在逐步们从大规模、简单向小规模、复杂进行转变，我们产品的应用前景非常广阔。同时对于其它正在研究知识图谱相关内容的人员来说，我们也提供了一种新的技术路线，在一定程度上也能促进共同的进步。

最后，由于我们的产品部署在服务器上，消耗的自然资源非常少。相较于需要购买大量的书籍来掌握相关内容，我们的产品经济环保了许多。

图 3：产品示例图

4.4 产品规划

4.4.1 扩大应用范围

我们计划在一年内实现文档级的知识获取。考虑到人工智能发展的火热，在未来肯定会有越来越多的人工智能产品走入大家的生活中。而在现实情况中，我们买任意设备，经常会附赠一个说明手册，例如买冰箱都会有一个手册，但是手册的利用率极低，很少有家庭成员会真正的翻阅。然而当碰到问题想去查找的时候，我们也很难从手册中找到答案。更何况是人工智能领域的高科技产品。所以基于能否将这些鸡肋一般的手册全部淘汰掉，同时还能提升用户满意度的考虑。我们希望团队的人工智能知识图谱不仅仅适用于想要学习这一专业领域的人才，也能够帮助到其它人即使不太了解专业知识也能够应对这种生活中的突发情况。我们计划将手册变成知识库并存储在数据库中利用知识图谱实现知识问答。那么不仅仅是人工智能这个领域，还可以将比如冰箱的手册变成知识库进行储存，需要变换的就是数据库里的数据，整个技术路线我们已经完成。所以我们将可以为整个社会解决手册这一巨大成本问题。实现

这个目标的前提是文档级的知识获取。基于文档的信息抽取需要结合文档自身的结构，书写风格，和组织形式进行一定的迁移。业务文档结构化迫切需要从句子级别抽取发展到篇章级别抽取。

4.4.2 开发新的业务

我们计划在探究如何将知识图谱应用于专业领域获得的启发应用于平时生活中的简单场景。在两年内利用我们的知识图谱技术路线补全简单场景中缺失的因果链条（背景知识）。万事万物都处在一个复杂的因果网络中，当前的大数据多是业务结果数据，缺乏产生这些数据的背景因果。比如，数据挖掘中的经典案例尿片与啤酒，买尿片的人经常买啤酒。可是为什么会出现这种情况呢，其实如果我们能够推测男性用户为什么会同买啤酒与尿片的原因，这实际上可以帮助我们创造更大的商业价值。可能是家里有婴儿，而孕妇出行不便，因此必须得由作为父亲的他来买尿布，同时这几天由于工作他非常紧张与疲惫，所以买一点啤酒顺便缓解一下压力。如果我们能里用知识图谱把这个因果链条给补全，当男性用户再次买尿片时，我们推断他压力大，因而给他推荐心理咨询服务。由此得到启发我们可以推荐很多新的业务。又再次的扩大了整个产品的经济价值与应用前景。

参考文献

[1]李晓理,张博,王康,余攀.人工智能的发展及应用[J].北京工业大学学报,2020,46(06):583-590.

[2]肖仰华.知识图谱的下半场：机遇与挑战[R]