

# 马克思主义视角下的 NLP 发展



## 1.引言

相信每个人都有过这样一些经历：问 Siri 天气，它一本正经地回答：“成都今天有中到小雨，记得带伞哟。”；问豆包怎么写作文，它能够在短时间内为你创作出一篇流畅的文章。这些场景都是人工智能中的“语言助手”——自然语言处理（Natural Language Processing，简称 NLP）在发力。它让机器能听懂、能说话、能聊天、甚至能吵架。



NLP 广泛的应用场景

那么，NLP 到底是什么？NLP 隶属于计算机科学、人工智能和语言学的交叉领域，旨在让计算机能够理解、生成和处理复杂且抽象的人类自然语言。简单来说，NLP 就是赋予计算机“听懂”“看懂”人类的语言，以及“说出”“写出”人类能理解的语言的能力。

语言既是人类的最基础活动，同时又是最复杂的信息游戏。**所谓人工智能，就是把人的部分智能智能活动机器化**，而语言无疑是最重要的一部分，无法“学会”人类语言的机器不能被认定是智能的。因此，NLP 领域从发展之初就备受关注。本文将以马克思主义的视角带你走进 NLP 这一关键技术的发展。

## 2.NLP 的演进是“从量变到质变”的过程

**任何事物的变化都有一个量变的积累过程，没有量变的积累，质变就不会发生。**NLP 的发展正是“量变到质变”的生动写照——从机器“听不懂人话”到能写作文、聊日常，背后是几十年技术迭代的厚积薄发。

**量变是保持事物的质的相对稳定性的不显著变化。**在 NLP 领域，这一阶段持续了半个多世纪，经历了“规则化→统计化→早期神经网络”三代技术迭代，每一步都在解决前一代的局限，却始终未跳出“人工依赖”或“功能单一”的框架：

- 基于语法规则（1950s-1980s）

在该阶段，以乔姆斯基为代表的大师们设计了许多形式语言规则，试图将所有语法输入给计算机从而让机器学会说人话。但是，自然语言的复杂性却似乎超出了他们的想像。每个语种都有自己的语法规则，判断标准也不确定，语言还会动态变化，即便相同的句子，在不同时间、场合、人群代表的含义不同，形式语言无法解决这些问题。



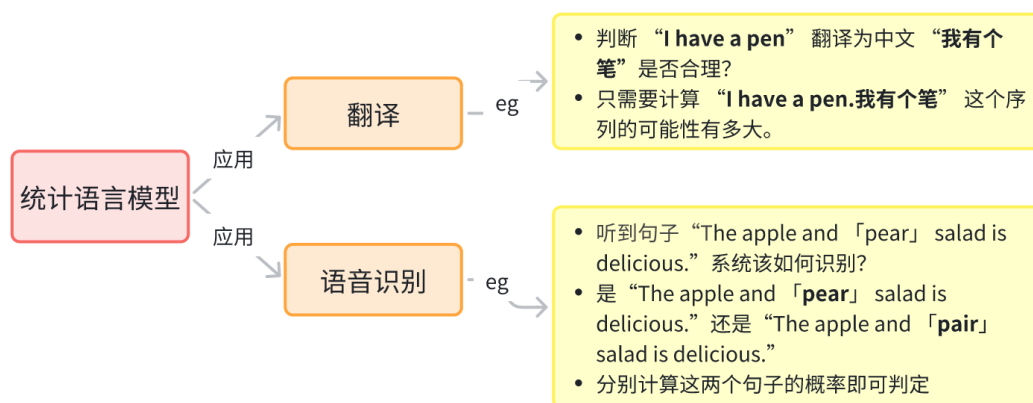
乔姆斯基

- 基于统计的语言模型（1980s-2010s）

**事物的发展不是直线性前进，而是螺旋式上升的。**经历了 20 多年的挣扎后，一部分人开始回到最初的起点，从香农和马尔可夫的理论中寻找灵感，这群人被称为统计派。

马尔可夫假设：任意一个词  $W_t$  出现的概率只同词  $W_{t-1}$  有关

统计派以基于语料库的统计分析为基础方法，更注重使用数学知识，从能代表自然语言规律的大规模真实文本中发现知识，抽取语言现象或统计规律，代表性技术包括：隐马尔可夫模型（HMM）、条件随机场（CRF）、N 元语言模型（N-gram）和支持向量机（SVM）等。



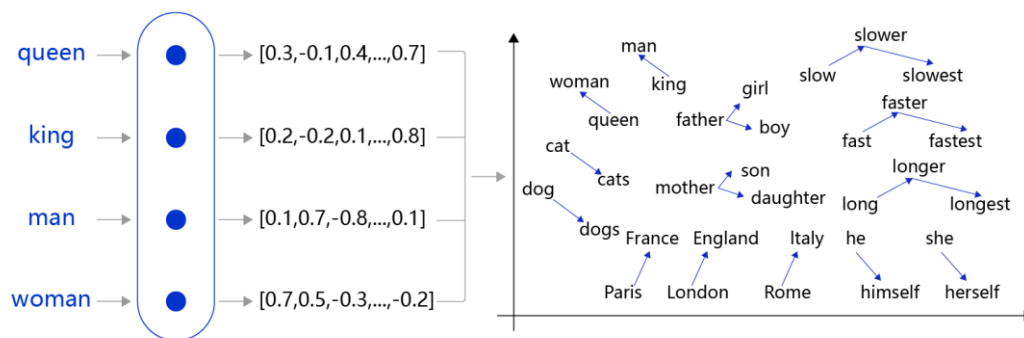
统计语言模型应用场景

这些技术让 NLP 处理的文本量从“几百句”提升到“几十万句”，但本质还是“机器被动接收人类筛选的信息”，属于量变的“效率提升期”。但仅仅通过简单的统计学方法，依旧难以应付大千世界语义的复杂。它们的研究也只能限于感知层面，而无法在认知层面上取得进展。

- 基于早期神经网络的语言模型（2010s-2017）

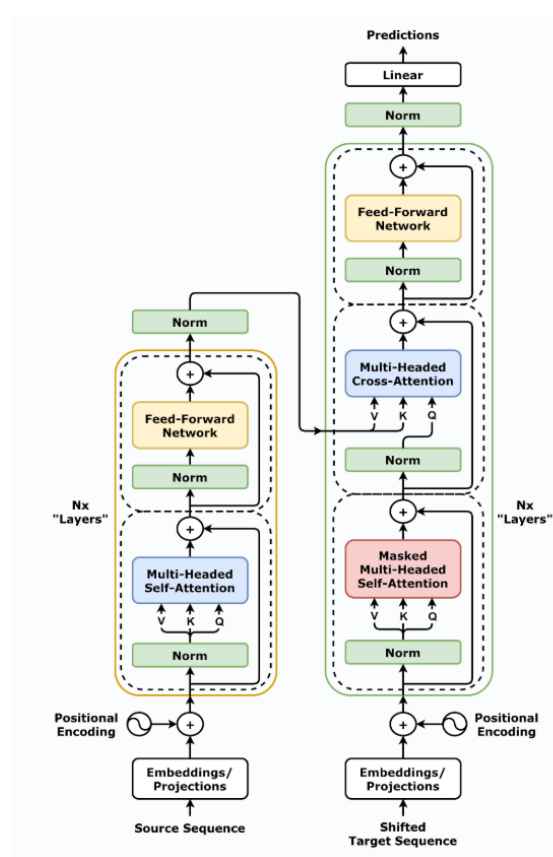
当数据量进一步扩大，“人工筛选特征”越来越吃力。随着深度学习技术的突破性进展，尤其是神经网络架构的优化与大规模算力的突破，NLP 领域进一步发展，开始向语义理解靠拢。

2010 年，Tomas Mikolov 提出基于 RNN 的语言模型（RNNLM）；2013 年，Tomas Mikolov 再接再厉，引入层次 softmax 和负采样，解决了 RNNLM 训练问题，设计出来第一个真正意义上的神经网络语言模型 Word2Vec，将每个词向量化表示，并在向量空间中捕捉到了语义关联。这类语言模型已经具有了一定语义理解能力，但能力还是相对有限，处于量变的“加速积累期”。



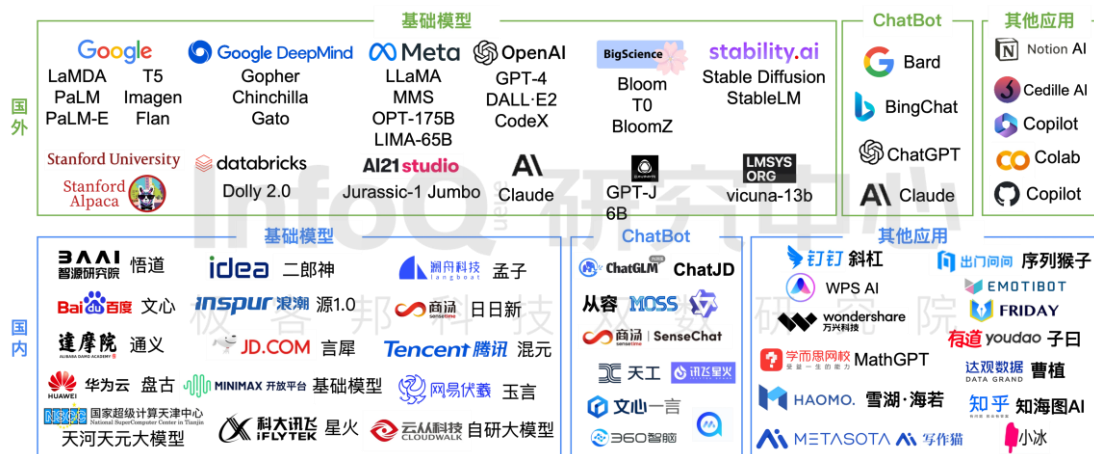
通过词向量捕捉语义关联

单纯的量变不会永远持续下去，量变达到一定程度必然引起质变：2017 年，Transformer 架构以划时代的姿势横空出世，它彻底改变了 NLP 的模型架构，让 NLP 从“被动处理文字符号”跃升到“主动理解语义语境”，完成了领域生产力的革命性突破。



Transformer 架构图

2018 年底，Google 推出以 Transformer 架构为基础的 BERT 模型，横扫 11 个榜单，成为当时的王者，NLP 任务模式全面进入 BERT 为主的预训练语言模型时代；2020 年，GPT-3 出现，又是一个分界点，从此后进入大语言模型时代。此后的故事，就是我们正在经历的现在……



大语言模型时代

物质世界处于永恒的运动之中，而物质世界的运动中内在地包含着事物的变化和发展。从规则驱动到统计学习，从 Word2Vec 的语义向量到大模型的通用生成能力，NLP 发展的历史揭示：技术进化并非范式替代，而是技术栈的持续融合。人工智能正在不断突破认知边界，机器对人类语言的本质正在不断重新诠释。

### 3.NLP 的发展重塑劳动形态与社会关系

生产力与生产关系、经济基础与上层建筑之间的矛盾，是人类社会基本矛盾。生产力与生产关系矛盾运动的规律、经济基础与上层建筑矛盾运动的规律，是人类社会发展的基本规律。NLP 作为新时代的“智能生产力”，正不断推动着生产力与生产关系的“适应与矛盾”。



科学技术上的发明创造，会引起劳动资料、劳动对象和劳动者素质的深刻变革和巨大进步。过去，海量的非结构化文本信息（如病历、报告、新闻、合同、邮件）的处理，高度依赖人类的阅读、理解和归纳。NLP 技术的迅猛发展，让信息处理领域的劳动发生了质的变化。它正在重塑各行各业的生产方式，将人类从繁琐、重复的信息



劳动中解放出来，迈向更富创造性的工作阶段。

很具象的一个例子是，在传统诊疗模式中，医生需要花费大量时间翻阅厚重的纸质病历或在不同电子系统中查找患者历史，诊断很大程度上依赖于个人经验和记忆。NLP 系统可以瞬间解析患者的这些内容，甚至辅助提示可能的诊断结果和推荐治疗方案。医生得以从“信息挖掘工”转变为“诊断决策者”，将时间和精力集中于与患者的深度沟通、复杂病情的综合研判和手术等上。从生产力层面上看，NLP 是“劳动工具的智能升级”。



NLP 带来诊疗模式转变

为了进行生产，人们相互之间便发生一定的联系和关系；只有在这些社会联系和社会关系的范围内，才会有他们对自然界的影响，才会有生产。仔细留意，我们也能从 NLP 中感受到生产力与生产关系的有机统一：

- 生产力决定生产关系：

为适应并服务于新的 NLP 生产力，生产关系正在发生一系列积极的、结构性的演变。NLP 技术并未简单地消灭工作，而是重塑了劳动分工，催生了一系列新兴职业，如 AI 训练师、算法工程师、算法审核员等，以支撑、维护和优化智能系统的运行。这

些新岗位的出现，标志着生产关系为匹配智能生产力而发生了适应性变革。



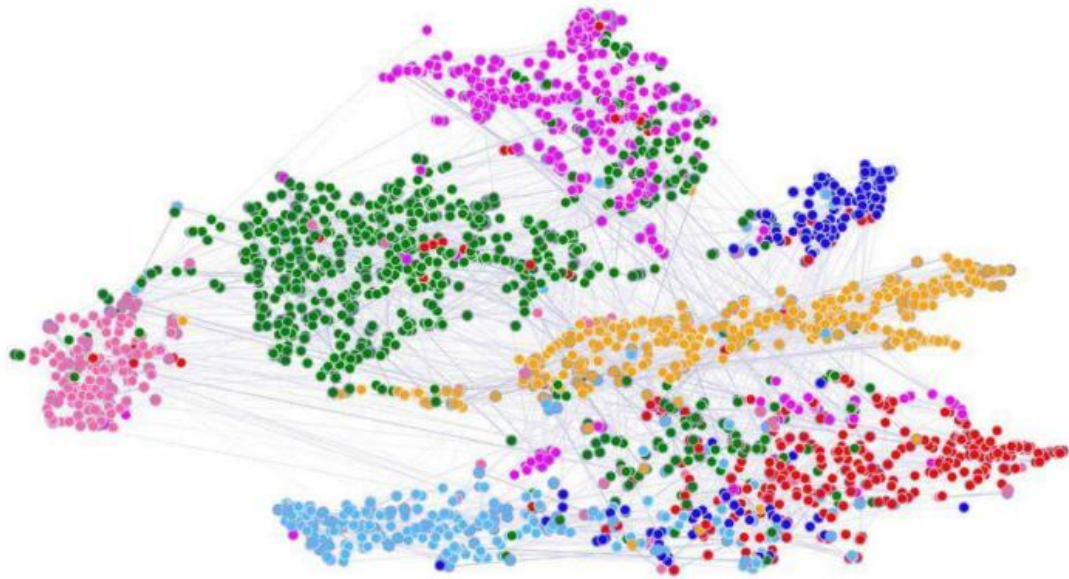
AI 工程师等岗位正在兴起

- 生产关系又反作用于生产力：

生产关系的调整往往滞后于生产力的迅猛发展，从而引发剧烈的社会矛盾。当生产关系不适合生产力发展的客观要求时，就会阻碍生产力的发展。NLP 能够自动完成翻译、数据录入、基础法律文书审阅等重复性、程式化的信息处理工作，相关传统白领岗位很大一部分被替代。

然而，社会层面的劳动力教育体系却未能跟上。被替代的劳动者所掌握的技能，如简单的信息整理、报告撰写等，在新时代迅速贬值，而他们向新岗位的转型却面临极高的学习成本和门槛。这种矛盾最终需要通过社会制度的主动变革，如建立终身职业技能培训体系等加以解决，以促成生产力与生产关系在新的水平上达到统一。生产力与生产关系的这种矛盾运动循环往复，不断推动社会生产发展，进而推动整个社会逐步走向高级阶段。

从阶级视角分析 NLP 带来的影响，我们能感受到一种新型生产关系的形成。在农业时代，核心生产资料是土地；在工业时代，是机器与厂房；而在信息时代，数据，特别是经过清洗、标注的高质量数据，已成为驱动智能经济运行的“新石油”。



大型科技公司对海量数据进行利用和占有

科技公司免费或有偿地收集、占有和控制着全球数十亿用户产生的行为、社交和内容数据，并将其转化为训练专有模型的“燃料”。通过垄断这些数据，训练出强大的模型，进而提供智能服务、获取巨额利润、巩固市场地位。广大普通用户是这些海量数据的实际创造者。我们的每一次搜索、点击、评论、分享、上传照片，都是在无偿地生产和再生产着训练 AI 所必需的生产资料。在创造这些价值连城的数据时，我们通常并未获得与之匹配的经济回报。

NLP 的发展将古典政治经济学中的核心矛盾——即生产资料的私人占有与社会化生产之间的矛盾，以一种崭新的形式清晰地展现出来。可以预见，当数据成为关键生产资料，而占有权高度集中时，科技进步所带来的巨大红利仍将面临分配极度不公的风险。

## 4.结束语

NLP 的发展是技术进步与社会矛盾的辩证统一，它既是生产力飞跃的体现，推动劳动工具与劳动分工改变，也同样引发了生产关系的不适配，暴露生产资料占有与价值分配的不平等矛盾。当我们睡醒询问 Siri 天气如何、当我们“江郎才尽”时让豆包帮助我们写出一篇流畅的作文时，是否想到了在其背后的生产力的进步，又是否会加以思考：如何让 NLP 的发展更好的适配社会，让更多人更好的享受到这样的技术红利。**矛盾是推动事物发展的动力，社会领域也不例外**，生产力与生产关系矛盾运动推动社会发展，NLP 的未来不仅在于技术的不断迭代革新，更在于人类通过制度创新利用矛盾，让智能生产力真正成为推动社会进步与人的发展的动力。