

浙江省普通本科高校“十四五”重点教材

# 大语言模型通识

---



# 什么是ChatGPT

OpenAI是一家人工智能研究实验室，进行AI研究的目的是促进和开发友好的AI。OpenAI系统运行在世界上第五强大的超级计算机上。马斯克、微软都是该组织的捐助者。

2020年OpenAI发布GPT-3，这是一种在大型互联网数据集上训练的语言模型，旨在用自然语言回答问题，但它也可以在语言之间进行翻译并连贯地生成即兴文本。

ChatGPT+Midjourney



# 什么是ChatGPT

2021年OpenAI推出DALL-E，这是一种深度学习模型，可以从自然语言描述中生成数字图像。

2022年12月，OpenAI推出基于GPT-3.5的新型AI聊天机器人ChatGPT，受到媒体的广泛报道。在发布仅两个月后就拥有1亿用户（成为史上用户增长最快的应用）。

2023年3月14日，OpenAI发布了GPT-4聊天机器人语言模型。相比GPT3.5，GPT 4.0访问更加稳定，数据更加准确，并且更加符合我们人类的思维方式。简而言之，更智能！

ChatGPT+Midjourney



# 什么是Midjourney

Midjourney（简称MJ）是AI绘图里实用性最强的软件，其开发团队是美国旧金山的小型自筹资金团队，专注于设计、人类基础设施和人工智能。

MJ是一个致力于探索新的思维方式并扩展人类的想象力的人工智能绘图平台。它于2022年7月12日首次公测，并于2022年3月14日正式以架设在 Discord 上的服务器形式推出，用户注册Discord并加入MJ服务器即可开始AI创作。

ChatGPT+Midjourney



# ChatGPT+Midjourney

一双丹凤三角眼，  
两弯柳叶吊梢眉，  
身量苗条，体格风骚，  
粉面含春威不露，  
丹唇未启笑先闻。

谷歌翻译：王熙凤

A pair of Danfeng triangle eyes,  
Two curved willow leaves hanging eyebrows,  
Slim and coquettish,  
Powder noodles contain spring power,  
Dan didn't smile and smell first.

## ChatGPT+Midjourney



# ChatGPT+Midjourney

一双丹凤三角眼，  
两弯柳叶吊梢眉，  
身量苗条，体格风骚，  
粉面含春威不露，  
丹唇未启笑先闻。

谷歌翻译：王熙凤

A pair of Danfeng triangle eyes,  
Two curved willow leaves hanging eyebrows,  
Slim and coquettish,  
Powder noodles contain spring power,  
Dan didn't smile and smell first.

## ChatGPT+Midjourney



大语言模型通识

# 第1章 概述

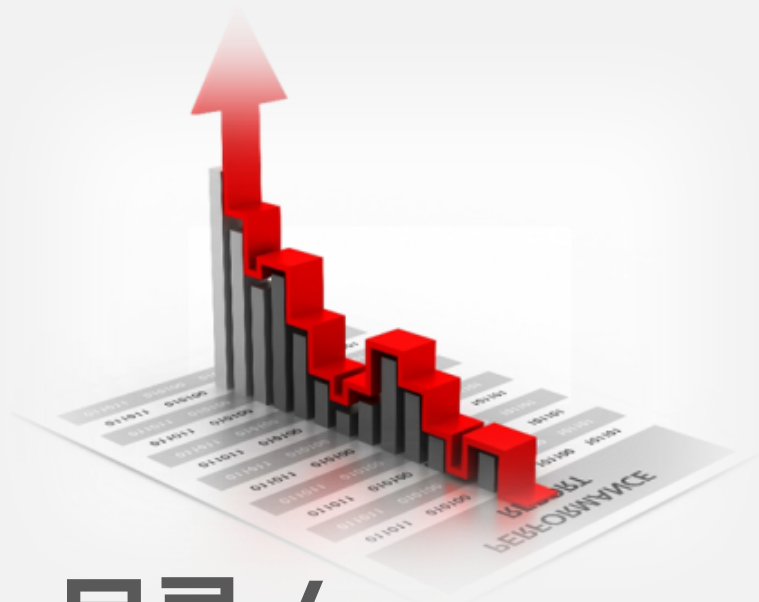
# 第1章 概论

人类，又称智人，即有智慧的人，这是因为智能对于人类来说尤其重要。几千年来，人们一直在试图理解人类是如何思考和行动的，也就是不断地了解人类的大脑是如何凭借它那小部分的物质去感知、理解、预测并操纵一个远比其自身更大更复杂的世界。



# 第1章 概论

“数据成为新生产要素，算力成为新基础能源，人工智能成为新生产工具。”2023年被称为生成式人工智能（Generative AI）元年，以ChatGPT为代表的生成式人工智能技术的涌现获得了前所未有的关注。大型科技公司、各类创业公司迅速入场，投入海量资源，推动了大语言模型（Large Language Model, LLM，简称“大模型”）能力和应用的快速演进。



## 目录 / CONTENTS



01  
人工智能  
基础



02  
大语言模型  
的定义



03  
大语言模型  
形成基础



04  
通用  
人工智能



PART 01

---

# 人工智能基础

## 1.1 人工智能基础

人工智能是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术科学，是一门自然科学、社会科学和技术科学交叉的边缘学科，它涉及的学科内容包括哲学和认知科学、数学、神经生理学、心理学、计算机科学、信息论、控制论、不定性论、仿生学、社会结构学与科学发展观等。

## 1.1.1 人工智能的定义

作为计算机科学的一个分支，人工智能专注于创建“智能系统”，这些系统具有推理、学习、适应和自主行动的能力。人工智能是一个多元化的领域，围绕着设计、理论、开发和应用能够展现出类似人类认知功能的机器而展开。具有人工智能的机器努力模仿人类的思维和行为，包括但不限于理解自然语言、识别模式、解决问题和做出决策。

## 1.1.1 人工智能的定义

人工智能研究领域的一个较早流行的定义，是由约翰·麦卡锡在1956年的达特茅斯会议上提出的，即：**人工智能就是为了让机器的行为看起来像是人类所表现出的智能行为一样。**另一个定义指出：**人工智能是人造机器所表现出来的智能性。**总体来讲，对人工智能的定义大多可划分为四类，即机器“像人一样思考”“像人一样行动”“理性地思考”和“理性地行动”。这里“行动”应广义地理解为采取行动，或制定行动的决策，而不是肢体动作。

## 1.1.1 人工智能的定义

尼尔逊教授对人工智能下了这样一个定义：“**人工智能是关于知识的学科——怎样表示知识以及怎样获得知识并使用知识的科学。**”而温斯顿教授认为：“**人工智能就是研究如何使计算机去做过去只有人才能做的智能工作。**”这些说法反映了人工智能学科的基本思想和基本内容。即人工智能是研究人类智能活动的规律，构造具有一定智能的人工系统，研究如何让计算机去完成以往需要人的智力才能胜任的工作，也就是研究如何应用计算机的软/硬件来模拟人类某些智能行为的基本理论、方法和技术。

## 1.1.1 人工智能的定义

可以把人工智能定义为一种工具，用来帮助或者替代人类思维。它是一项计算机程序，可以独立存在于数据中心、个人计算机，也可以通过诸如机器人之类的设备体现出来。它具备智能的外在特征，有能力在特定环境中有针对性地获取和应用知识与技能。人工智能是对人的意识、思维的信息过程的模拟。人工智能不是人的智能，但能像人那样思考，甚至也可能超过人的智能。



## 1.1.1 人工智能的定义

20世纪七十年代以来，人工智能被称为世界三大尖端技术之一（空间技术、能源技术、人工智能），也被认为是21世纪三大尖端技术（基因工程、纳米科学、人工智能）之一，这是因为近几十年来人工智能获得了迅速的发展，在很多学科领域都获得了广泛应用，取得了丰硕成果。

## 1.1.2 人工智能的实现途径

对于人的思维模拟的研究可以从两个方向进行，

一是结构模拟，仿照人脑的结构机制，制造出“类人脑”的机器；

二是功能模拟，从人脑的功能过程进行模拟。

现代电子计算机的产生便是对人脑思维功能的模拟，是对人脑思维的信息过程的模拟。

## 1.1.2 人工智能的实现途径

实现人工智能有三种途径，即强人工智能、弱人工智能和实用型人工智能。强人工智能，又称多元智能。研究人员希望人工智能最终能成为多元智能并且超越大部分人类的能力。有些人认为要达成以上目标，可能需要拟人化的特性，如人工意识或人工大脑，这被认为是人工智能的完整性：为了解决其中一个问题，你必须解决全部的问题。即使一个简单和特定的任务，如机器翻译，要求机器按照作者的论点（推理），知道什么是被人谈论（知识），忠实地再现作者的意图（情感计算）。因此，机器翻译被认为是具有人工智能完整性。

## 1.1.2 人工智能的实现途径

强人工智能的观点认为有可能制造出真正能推理和解决问题的智能机器，并且这样的机器将被认为是有知觉的，有自我意识的。强人工智能可以有两类：

(1) 类人的人工智能，即机器的思考和推理就像人的思维一样；

(2) 非类人的人工智能，即机器产生了和人完全不同的知觉和意识，使用和人完全不同的推理方式。

强人工智能即便可以实现也很难被证实。为了创建具备强人工智能的计算机程序，我们首先必须清楚了解人类思维的工作原理，而想要实现这样的目标，还有很长的路要走。

## 1.1.2 人工智能的实现途径

弱人工智能，认为不可能制造出能真正地推理和解决问题的智能机器，这些机器只不过看起来像是智能的，但是并不真正拥有智能，也不会有自主意识。它只要求机器能够拥有智能行为，具体的实施细节并不重要。深蓝就是在这样的理念下产生的，它没有试图模仿国际象棋大师的思维，仅仅遵循既定的操作步骤。计算机每秒验算的可能走位就高达2亿个，就算思维惊人的象棋大师也不太可能达到这样的速度。

## 1.1.2 人工智能的实现途径

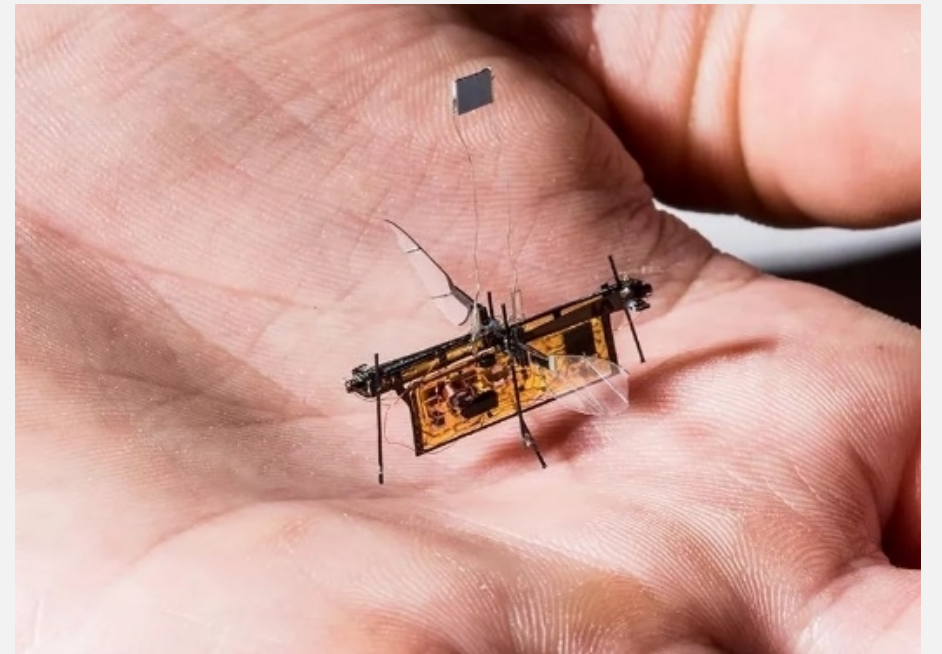
人类拥有高度发达的战略意识，这种意识将需要考虑的走位限制在几步或是几十步以内，而计算机的考虑数以百万计。就弱人工智能而言，这种差异无关紧要，能证明计算机比人类更会下象棋就足够了。

如今，主流的研究活动都集中在弱人工智能上，并且一般认为这一研究领域已经取得可观的成就，而强人工智能的研究则处于停滞不前的状态。

## 1.1.2 人工智能的实现途径

第三种途径称为实用型人工智能。研究者们将目标放低，不再试图创造出像人类一般智慧的机器。眼下我们已经知道如何创造出能模拟昆虫行为的机器人。机械家蝇看起来似乎并没有什么用，但即使是这样的机器人，在完成某些特定任务时也是大有裨益的。比如，一群如狗大小，具备蚂蚁智商的机器人在清理碎石和在灾区找寻幸存者时就能够发挥很大的作用。

图1-1 华盛顿大学研制的靠激光束驱动的RoboFly昆虫机器人



## 1.1.2 人工智能的实现途径

随着模型变得越来越精细，机器能够模仿的生物越来越高等，最终，我们可能必须接受这样的事实：机器似乎变得像人类一样智慧了。也许实用型人工智能与强人工智能殊途同归，但考虑到一切的复杂性，我们不会相信机器人会有自我意识。



## 1.1.3 机器学习和深度学习

机器学习是人工智能的一个关键子集，是一种能够根据输入数据训练模型的系统。它的主要目标是让计算机系统能够通过训练模型，使其能够从新的或以前未见过的数据中得出有用的预测。换句话说，机器学习的核心是“使用算法解析数据，从中学习，然后对世界上的某件事情做出决定或预测”。这意味着，与其显式地编写程序来执行某些任务，不如教计算机学会如何开发一个算法来完成的任务。

### 1.1.3 机器学习和深度学习

在机器学习中，我们不是直接编程告诉计算机如何完成任务，而是提供大量的数据，让机器通过数据找出隐藏的模式或规律，然后用这些规律来预测新的、未知的数据。机器学习可以根据所处理的数据自主地学习和适应，大大减少了对显式编程的需求。通常将人工智能看作是自主机器智能的广泛目标，而机器学习则是实现这一目标的具体方法。

### 1.1.3 机器学习和深度学习

比如，如果我们通过代码告诉电脑，图片里红色是玫瑰，有说明的是向日葵，那么程序对花种类的判断就是通过人类直接编写逻辑达成的，不属于机器学习，机器什么也没学。但是如果我们给计算机大量玫瑰和向日葵的图片，让电脑自行识别模式、总结规律，从而能对后来新输入的图片进行预测和判断，这就是机器学习。

## 1.1.3 机器学习和深度学习

深度学习是机器学习的一个子集，其核心在于使用人工神经网络模仿人脑处理信息的方式，通过层次化的方法提取和表示数据的特征。虽然单层神经网络就可以做出近似预测，但是添加更多的隐藏层可以优化预测的精度和准确性。神经网络是由许多基本的计算和储存单元组成，这些单元被称为神经元。神经元通过层层连接来处理数据，并且深度学习模型通常有很多层，因此被称为“深度”学习。深度学习模型能够学习和表示大量复杂的模式，这使它们在诸如图像识别、语音识别和自然语言处理等任务中非常有效。

## 1.1.4 监督学习与无监督学习

机器学习有三种主要类型，即监督学习、无监督学习和强化学习。其中，监督学习就像一个有答案的教科书，模型可以从标记的数据中学习，也就是说，它有答案可以参考学习；而无监督学习则更像一个无答案的谜题，模型需要自己在数据中找出结构和关系。此外，还有一种介于两者之间方法称为强化学习，其模型通过经验学习执行动作。

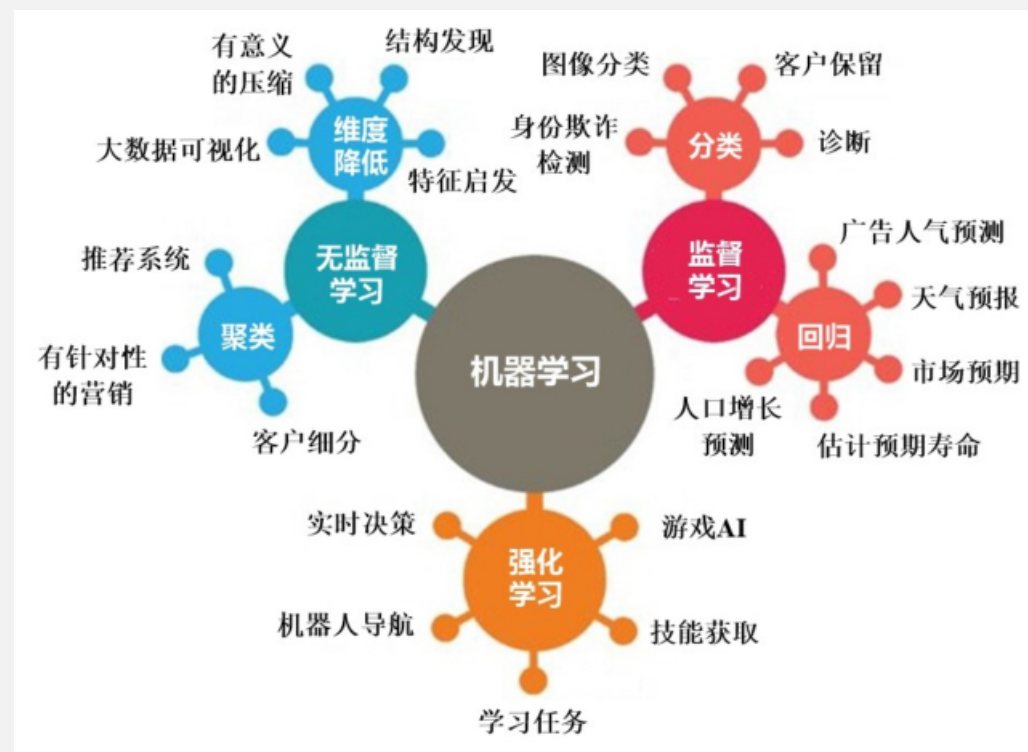


图1-2 机器学习的三种主要类型

## 1.1.4 监督学习与无监督学习

(1) **监督学习**，也称有导师学习，是指输入数据中有导师信号，以概率函数、代数函数或人工神经网络为基函数模型，采用迭代计算方法，学习结果为函数。

在监督学习里，机器学习算法接受有标签的训练数据（标记数据），标签就是期望的输出值。所以每一个训练数据点都既包括输入特征，也包括期望的输出值。计算机使用特定的模式来识别每种标记类型的新样本，即在机器学习过程中提供对错指示，一般是在数据组中包含最终结果 (0, 1)。通过算法让机器自我减少误差。

## 1.1.4 监督学习与无监督学习

监督学习从给定的训练数据集中学习出一个函数，当接收到一个新的数据时，可以根据这个函数预测结果。算法的目标是学习输入和输出之间的映射关系，从而在给定新的输入特征后，能够准确预测出相应的输出值。

监督学习的主要类型是分类和回归。在分类中，机器被训练成将一个组划分为特定的类，一个简单例子就是电子邮件中的垃圾邮件过滤器。过滤器分析你以前标记为垃圾邮件的电子邮件，并将它们与新邮件进行比较，如果它们有一定的百分比匹配，这些新邮件将被标记为垃圾邮件并发送到适当的文件夹中。

## 1.1.4 监督学习与无监督学习

在回归中，机器使用先前的（标记的）数据来预测未来，天气应用是回归的好例子。使用气象事件的历史数据（即平均气温、湿度和降水量），手机天气预报APP可以查看当前天气，并对未来时间的天气进行预测。

例如，拿一堆猫猫狗狗的照片和照片对应的“猫”“狗”标签进行训练，然后让模型根据没见过的照片预测是猫还是狗，这就属于分类。拿一些房子特征的数据，比如面积、卧室数量、是否带阳台等和相应的房价作为标签进行训练，然后让模型根据没见过的房子的特征预测房价——这就属于回归。



## 1.1.4 监督学习与无监督学习

(2) **无监督学习**，又称无导师学习、归纳性学习。在无监督学习中，学习的数据是没有标签的，是指输入数据中无导师信号，采用聚类方法，学习结果为类别，所以算法的任务是自主发现数据里的模式或规律。典型的无导师学习有发现学习、聚类、竞争学习等。

无监督学习通过循环和递减运算来减小误差，达到分类的目的。在无监督学习中，数据是无标签的。由于大多数真实世界的的数据都没有标签，这样的算法就特别有用。比如，拿一堆新闻文章，让模型根据主题或内容的特征自动把相似文章进行组织。

## 1.1.4 监督学习与无监督学习

无监督学习分为聚类和降维。聚类用于根据属性和行为对象进行分组。这与分类不同，因为这些组不是你提供的。聚类的一个例子是将一个组划分成不同的子组（例如，基于年龄和婚姻状况），然后应用到有针对性的营销方案中。降维通过找到共同点来减少数据集的变量。大多数大数据可视化使用降维来识别趋势和规则。

## 1.1.4 监督学习与无监督学习

(3) **强化学习**。是让模型在环境里采取行动，获得结果反馈。从反馈里学习，从而能在给一定情况下采取最佳行动来最大化奖励或是最小化损失。例如刚开始的时候，小狗会随心所欲做出很多动作，但随着和驯犬师的互动，小狗会发现某些动作能够获得零食，某些动作没有零食，某些动作甚至会遭受惩罚。通过观察动作和奖惩之间的联系，小狗的行为会逐渐接近驯犬师的期望。在很多任务上，比如说让模型下围棋，获得不同行动导致的奖励或损失反馈，从而在一局局游戏里优化策略，学习如何采取行动达到高分。



PART 02

---

# 大语言模型的定义

## 1.2 大语言模型的定义

大语言模型是深度学习的应用之一，尤其在自然语言处理领域，这些模型的主要目标是理解和生成人类语言。为了实现这个目标，模型需要在大量文本数据上进行训练，以学习语言的各种模式和结构。例如，ChatGPT就是一个大模型的例子，它被训练来理解和生成人类语言，以便进行有效的对话和解答各种问题。

## 1.2 大语言模型的定义

### 1. 大模型和生成式人工智能存在交集

生成式人工智能是一种能够创造新的内容或预测未来数据的人工智能技术。这个技术包括用于生成文本、图像、音频和视频等各种类型的内容的模型。生成式人工智能的一个关键特性是，它不仅可以理解和分析数据，还可以创造新的、独特的输出，这些输出是从学习的数据模式中衍生出来的。

## 1.2 大语言模型的定义

### 2. 大型通用语言模型预训练和微调

大模型可以进行预训练，然后针对特定目标进行微调。以训练狗为例，可以训练它坐、跑、蹲和保持不动。但如果训练的是警犬、导盲犬和猎犬，则需要特殊的训练方法。大模型的训练也采用与之类似的思路。大模型被训练来解决通用（常见）的语言问题，如文本分类、问答、文档总结和文本生成等。

## 1.2 大语言模型的定义

(1) 文本分类：大模型可以通过对输入文本进行分析和学习，将其归类到一个或多个预定义的类别中。例如，可以使用大模型来分类电子邮件是否为垃圾邮件，或将博客文章归类为积极、消极或中立。

(2) 问答：大模型可以回答用户提出的自然语言问题。例如，可以使用大模型来回答搜索引擎中的用户查询，或者回答智能助手中的用户问题。

(3) 文档总结：大模型可以自动提取文本中的主要信息，以生成文档摘要或摘录。例如，可以使用大模型来生成新闻文章的概要，或从长篇小说中提取关键情节和事件。



## 1.2 大语言模型的定义

(4) 文本生成：大模型可以使用先前学习的模式和结构来生成新的文本。

例如，可以使用大模型来生成诗歌、短故事、或者以特定主题的文章。

另一方面，大模型可以基于特定领域的小规模数据集进行训练，来定制化解决不同领域如零售、金融、娱乐等的特定问题。

## 1.2 大语言模型的定义

### 3. 大模型的特征

在“大模型”的上下文中，“大”主要有两层含义。一方面，它指的是模型的参数数量。在这些模型中，参数的数量通常会非常大，达到数十亿甚至数百亿。这使得模型能够学习和表示非常复杂的模式。另一方面，“大”也指的是训练数据的规模。大模型通常可以在来自互联网、书籍、新闻等各种来源的大规模文本数据上进行训练。

## 1.2 大语言模型的定义

在大模型中，“通用”这个词描述的是模型的应用范围。通用语言模型在训练时使用了来自各种领域的数据，因此它们能够处理各种类型的任务，不仅限于某一个特定的任务或领域。这使得这些模型在处理新的、未见过的任务时具有很强的泛化能力。

## 1.2 大语言模型的定义

预训练和微调。在预训练阶段，模型在大规模的通用文本数据上进行训练，学习语言的基本结构和各种常识。然后，在微调阶段，模型在更小、更特定的数据集上进行进一步的训练。这个数据集通常是针对某个特定任务或领域的，例如医学文本、法律文本，或者是特定的对话数据。微调可以让模型更好地理解 and 生成这个特定领域的语言，从而更好地完成特定的任务。

## 1.2 大语言模型的定义

### 4. 大模型的优势

单一模型可用于不同任务：由于大模型是通用的，并且具有强大的泛化能力，所以它们可以处理各种类型的任务，比如文本分类、命名实体识别、情感分析、问答系统、文本生成等。这意味着可以使用同一个预训练模型来处理不同的任务，只需要进行相应的微调就可以。这大大减少了开发和维护不同模型的复杂性和成本。

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：  
<https://d.book118.com/547001066114006144>