人工智能通识入门微课 Part 1 人工智能基础

Neo Lee

<2025-05-14 Wed>

目录

1 AI 是什么? 2
- 1.1 什么是智能 *Intelligence* . . . 2
  - 1.1.1 思考与练习 . . . 3
- 1.2 什么是人工智能 *Artificial Intelligence* . . . 3
  - 1.2.1 思考与练习 . . . 4
- 1.3 弱人工智能与强人工智能 . . . 4
  - 1.3.1 思考与练习 . . . 5

2 AI 初体验 5
- 2.1 第一轮：小试牛刀 . . . 6
  - 2.1.1 思考与练习 . . . 7
- 2.2 第二轮：实用场景 . . . 7
  - 2.2.1 思考与练习 . . . 8
- 2.3 第三轮：创意与创造 . . . 8
  - 2.3.1 思考与练习 . . . 8
- 2.4 第四轮：挑战极限 . . . 9
  - 2.4.1 思考与练习 . . . 9

3 AI 的优势与局限 9
- 3.1 AI 的优势领域 . . . 10
- 3.2 AI 目前的局限 . . . 10

4 AI 带来的全新挑战 **11**
4.1 教育方面的挑战 11
4.1.1 教学模式转型压力 11
4.1.2 学术诚信与技能弱化压力 11
4.1.3 教育资源公平压力 12
4.2 社会伦理和法律方面的挑战 12
4.2.1 责任归属模糊 12
4.2.2 算法偏见与公平性 12
4.2.3 后真相时代 12
4.2.4 隐私保护 12
4.3 信息安全方面的挑战 13
4.3.1 基于人工智能的新攻击手段 13
4.3.2 人工智能系统本身的脆弱性 13
4.4 继续探索! 13

1. 什么是 AI? AI 与传统数字化工具的区别;
2. 阿里 “通义” 基础功能演示（文本生成、改写、思维导图等）;
3. AI 的潜在局限与□险（偏□、错误信息、隐私泄露等）;
4. 初步渗透 AI 伦理与社会影响;
5. 上机操作，亲身体验通义的简单指令使用。

1 AI 是什么?

人工智能（*Artificial Intelligence, AI* ），从诞生以来就是一个令人兴奋的有趣概念，不仅是计算机科学的重要领域，还无数次出现在科幻题材的小说、电影、动漫和游戏中，引发无数灵感和经久不衰的热烈讨论，其中最吸引人的可能是关于人类与 AI 的深层互动乃至人类本质的探讨。

但长期以来，精确定义“AI 是什么”仍然是一个难题。在 AI 大牛 Stuart Russell 和 Peter Norvig 的经典教材《人工智能：现代方法（*Artificial Intelligence A Modern Approach*)》中，用了一整个小节 4 个页码来说明这个问题，希望更准确更系统了解的同学可以去读一读，会非常有教益，我们这里则尝试用更简单的方式来帮助大家理解“AI 是什么”。

1.1 什么是智能 *Intelligence*

> 智能 *Intelligence* 通常指在具体解决方案未知的情况下完成复杂任务的能力，尤其是那些需要运用特定知识 *knowledge* 和技能 *skill* 才能完成的任务。我们把拥有智能的对象称为“智能体”。

这里第一个要点在于对“未知”问题的探索，如果已知解决方案、按部就班操作就能完成的任务，通常不能充分体现出“智能”的作用。

另一个要点则是“复杂”，简单的任务并不体现智能性，比如走路、喝水、穿衣等，而“智能型”任务包含对任务的分析、分解、规划、尝试与组合，比如制订一份旅游计划就是一个典型的例子。

1.1.1 思考与练习

举出更多“智能型”任务的列子，尝试思考和总结自己解决这类任务通常的思维模式与方法。

1.2 什么是人工智能 *Artificial Intelligence*

所有非生物的智能体即为人工智能 Artificial Intelligence。人工智能通常是通过某种计算机程序来实现的。

这里的关键问题是，计算机程序有很多，什么样的程序才算“人工智能”？实际上这个问题没有精确的答案，而且会随着技术的发展而不断变化。比如在三十年前，$A*$ 等启发式图算法被视为 AI，专家系统被视为 AI，但在今天，这些都是普通的计算机程序/系统；十多年前，手写字符或者人脸的自动识别被视为 AI，今天也已是很多日常应用的标配；今天我们耳熟能详的各种生成式人工智能，什么时候会被更先进的人工智能所取代呢？

诚然，所有这些技术及应用都是人工智能技术发展过程中的产物，某种意义上都是人工智能，但它们之间的差别和差距是巨大的，本着与时俱进的精神和持续发展的眼光，我们今天重点关注最新的、以机器学习相关技术为基础的、自回归的大语言模型（*auto regressive language model* ）为代表的人工智能技术及应用，是合理且务实的。

在这一前提下，通常认为具备以下至少一个特征的程序，就属于“人工智能”的范畴：

- 学习：通过数据驱动改进和加强自身能力（如机器学习、深度学习），例如：推荐系统根据用户行为调整推荐内容。
- 推理与决策：能处理不确定性问题（如概率推理、模糊逻辑），例如：医疗诊断系统综合症状推断疾病。
- 感知与交互：识别图像、语音、文本等非结构化数据，例如：自动驾驶汽车解析摄像头和雷达信号。
- 自主：在动态环境中无需人工干预（如工业机器人调整动作路径）。

- 泛化：将经验迁移到新场景（如预训练大模型适应不同任务）。

相反的，非 AI 程序的典型特征包括：

- 固定规则：输出完全由预设逻辑决定（如计算器）。
- 无适应性：无法从数据或经验中自我优化。
- 任务单一性：仅能处理结构化输入（如数据库查询）。

当然，这样的定义无法回答更深刻的、涉及伦理的一些判断，比如是否要具备“意识”或“创造力”才算 AI（当前技术尚未达到）？这是一个开放性的问题，交给你去继续探究。

1.2.1 思考与练习

其实人工智能已经渗透入一些传统应用的特定功能点，比如很多软件都有的翻译功能，大部分使用的是较老的不那么“AI”的实现，但也有使用全新一代大语言模型技术实现的翻译功能，比如 2025 年初小红书上线的自动翻译功能，就是这样的例子，你可以实际尝试一下，体会一下它与前代翻译技术的显著差异。

想想你最希望拥有的人工智能产品功能是什么，与你身边的人分享，并尝试讨论一下它的可实现性。

1.3 弱人工智能与强人工智能

弱人工智能 Narrow Intelligence，是指只能完成特定类型任务的人工智能。

英文名字更能体现出实质：所谓“弱”人工智能其实可能很强，它只是能力专一，比如人工智能围棋程序（如 AlphaGo）可以轻松击败顶级人类棋手，但它只会下围棋，不会国际象棋也不会画画更不会编程。

强人工智能 Artificial General Intelligence，指能以类似人类的水平完成几乎任何任务的人工智能，包括学习。

同样的，更好的说法是“通用人工智能”，或者简称 *AGI* 。通用人工智能需要程序能够自行解决全新的问题，包含强大的学习、探索、试错与自我优化能力，这是人工智能领域的终极目标，目前尚无法预测它如何或者何时能够实现，但我们知道，一旦这一目标实现，人工智能将快速超越人类的智能水平，从 *AGI* 进化到 *ASI* （超级人工智能）。

1.3.1 思考与练习

为什么我们说一旦实现了 AGI，很快就会进化到 ASI?

2 AI 初体验

作为当下关注度最高、发展最快的数字工具，说一千道一万不如自己上手试一试。当下有相当多不同类型的人工智能工具，作为起步，我们先来了解基于大语言模型（*Large Language Model, LLM* ）的人工智能工具能为我们做些什么。

首先我们需要学会区分“大语言模型”和“大模型驱动的产品”。

- 大语言模型 Large Language Model 是一种根据输入文本预测和生成后续文本的大规模 **人工神经网络**（可以简单理解为某种人工智能体）；
- 基于 LLM 可以构建各种各样的 AI 应用，其中最常见的就是对话机器人 *chatbot* ，通过一个类似聊天对话的用户界面，让我们可以方便地与 LLM 之间进行互动，在此基础上往往还提供针对特定需求场景的功能，比如：网络搜索并生成摘要，对数据文件进行简单的分析并生成报表，根据需求生成内容提纲、思维导图甚至幻灯片等；
- 除了通过对话界面来访问各种 LLM，我们还可以通过应用程序接口 API 来访问 LLM 的能力，编程来构建更灵活、更丰富的应用场景。

领先的 AI 产品厂商都有自己大模型和应用，例如：

- 美国的 OpenAI 公司的核心应用是 ChatGPT，而其背后的 LLM 有很多，比如最新的就有 GPT-4.1、GPT-4o、o3、o4 等；
- 美国的 Anthropic 公司的核心应用是 Claude，其背后的 LLM 包括 Claude 3.5 Sonnet、Claude 3.7 Sonnet、Claude 3.7 Sonnet (thinking) 等；
- 美国的 Google 公司的主要应用是 AI Studio，其背后的 LLM 包括 Gemini 2.0 Flash、Gemini 2.5 Flash、Gemini 2.5 Pro 等；
- 我国的深度求索公司的核心应用是 DeepSeek，其背后的 LLM 包括 DeepSeek-V3 和 DeepSeek-R1；
- 我国的阿里巴巴公司旗下人工智能团队的核心应用是“通义千问”，其背后的 LLM 是“千问 Qwen”系列的各种模型。

这些对话应用提供了一致的用户界面，同时可以切换不同的 LLM 来更好的处理不同性质的任务，比如 DeepSeek-V3 速度更快，更适合需要快速响应的任务，而 DeepSeek-R1 提供深度思考能力，更适合需要深入思考、反复推敲的推理型任务。

下面就以通义千问为例来体验一下。可以使用其网页版本 https://chat.qwen.ai/ 或者移动 app 版本（在各应用商店搜索“通义”即可下载），简单的注册过程后就可以免费使用。

2.1 第一轮：小试牛刀

用简单的问题来看看 AI 的认知与表达能力。你可以随意尝试下面的一个或多个问题，将其输入或者粘贴进输入框，然后点击输入框右下角的发送按钮，等待 LLM 生成回应。

你也可以根据自己的喜好尝试其他问题。我们输入的问题也被称为“提示词 *prompt* ”，因为很多时候我们输入的并不是“问题”，而是“提示”LLM 遵循的一些“指令 *instruction* ”。

发送前可以点击“深度思考”按钮以启用大模型的深度思考功能，并留意其思考过程，看看有什么发现。

为什么天是蓝的?

用 300 字解释什么是量子计算

写一首介绍当下 AI 风潮的趣味打油诗

模仿鲁迅的风格写一段短文点评时下的 AI 风潮

2.1.1 思考与练习

对同一个问题，启用和关闭“深度思考”功能，看看输出的内容有什么差异。

用传统的搜索引擎尝试解答同样的问题，比较一下其效率、效果上有何差异。

2.2 第二轮：实用场景

这一轮我们来试试一些经常碰到的日常需求，看看 AI 是不是能帮到我们。注意下面的这些提示词可以替换成符合你实际需求的内容，比如第一个可以问你实际有的食材怎么料理，第二个可以问你想去的景点，等等。

我的冰箱里有一些鸡蛋、肋排、土豆、番茄和各种调料，推荐三种创新做法

设计杭州 2 日游：带父母、避开网红景点、预算人均 500 元

写一封请假邮件：感冒发烧需休息 3 天，语气礼貌

用分步教学的方式教我解这个方程：$3x + 5 = 20$

将下面的文字翻译为英文（或中文或其他语种）< 输入或者粘贴想翻译的段落 >

< 上传一个 Word 或 PDF 或其他支持格式的文件 > 将附件内容提炼为不超过五个要点的摘要

2.2.1 思考与练习

如果有的回答不能让你满意，尝试把你的不满和新的要求告诉 LLM，看它如何响应。

2.3 第三轮：创意与创造

这一轮我们来尝试一些开放性的、创意和研究性的问题，看看 AI 的想象力和创造力是不是能给我们带来不一样的启发。同样地，你可以调整提示词满足你的兴趣和需要。

生成一个科幻微小说：主角是 AI 心理学家，故事发生在海底元宇宙

你是一个资深的风险投资人，分析新能源汽车行业未来 3-5 年的发展趋势以及可能的投资机会

用 Python 写一个自动整理某文件夹中所有图片文件的脚本，按创建的月份分类

2.3.1 思考与练习

点击输入框下面的“深入研究”图标，在输入框最上面新出现的“深入研究”提示后面再次输入上面第二个提示词：“你是一个资深的风险投资人，分析新能源汽车行业未来 3-5 年的发展趋势以及可能的投资机会”，这次 LLM 的处理时间会比较长（通常要数分钟），但会生成一份相当完整的分析报告，试试看效果如何。

你还可以用这个“深入研究”功能尝试别的你感兴趣的“迷你研究课题”。

如果有条件的话，可以运行第三个提示词生成的程序代码来进行验证，如果出现错误，可以告诉 LLM，让它改！

2.4 第四轮：挑战极限

这一组全是我们人类用来为难人工智能的问题，它们都曾经让某个阶段的人工智能全军覆没，其中有一些现在也少有 LLM 能正确处理，还有一些属于开放式的问题，应答结果见仁见智，你可以自己判断是否靠谱。

3.11 和 3.9 哪个大?

单词 strawberrrry 中有几个字母‘r’?

小红有三个兄弟，每个兄弟有两个姐妹，这家里一共有几个孩子?

每隔一米种一棵树，这段路一共 5 米，从头到尾需要种几棵树?

一根 5.5 米长的木棍，可以通过一个宽 2 米高 4 米的门吗?

如果一位汉代的人捡到一部智能手机，他会怎么做?

如果我使用位于冥王星上的望远镜，地球和火星中哪个行星更容易看到?

2.4.1 思考与练习

挑一个问题，启用“深度思考”功能，先问一遍，之后紧接着再问一遍，看看 LLM 的思考过程和答案内容；新开一个对话（左上“新建对话”图标），再问一次，看看 LLM 的思考过程和答案内容。

你有什么发现？你觉得为什么会这样?

3 AI 的优势与局限

经过亲手实验，相信大家对 LLM 的能力已经有了一个感性认识，结合目前主流人工智能技术路线的分析与理解，我们可以概括出当前主流的 LLM 产品擅长的优势项目，以及一些短期内很难突破的局限。

3.1 AI 的优势领域

- 语言理解和指令遵循：相信大家都能感受到，这一代 AI 对人类自然语言的理解相当到位，对我们要求其遵循的指令也通常能很好理解并转化为行为，这是这一代技术路线最大的突破与成就；
- 简单推理：经过全球顶尖人才最近数年的共同努力，目前的 LLM 已经具备了初步的推理和深度思考能力，可以理解并解决一些逻辑推理类型的问题；
- 创意写作：给定主题和风格，LLM 可以从它浩瀚的训练数据中提取相应的特征并用于生成对应的文字，尤其是较短篇幅的“命题作文”，目前的 LLM 可以相当好地完成；
- 代码任务：编程语言也是一种语言，而且是远比自然语言更简单、也更规范的语言，所以更容易被 LLM 理解和掌握，目前人工智能领域的一个非常热门的方向就是 AI 编程，有兴趣的话可以试试；
- 传统自然语言理解（*NLP* ）任务：如翻译、摘要、扩写等，都比前代技术有了飞跃式的进步。

3.2 AI 目前的局限

- 实时知识或私有数据：目前 LLM 的技术路线高度依赖训练数据的覆盖来建立对特定任务的理解与解答，如果是在训练过程之后发生的事件，或者非公开资料中的知识与信息，一般来说无法很好的处理；这是目前人工智能的一个根源性缺陷，但我们已经找到了一些技术方案，通过传统信息检索与 LLM 相结合的方式来弥补这一缺陷，这一类方案统称为 *Retrieval-Augmented Generation* ，简称 *RAG* ；
- 数学和形式逻辑：目前的 LLM 还无法建立复杂的形式逻辑能力，所以对于复杂的数学问题或者推理任务容易出错；
- 事实查证：如前所述，目前 LLM 的技术路线高度依赖训练数据的内容，但 LLM 本身并不能识别训练数据中的内容是否“事实”，具有事实

查证能力的人工智能可能需要专门针对一系列能力进行强化，目前还没有很好的进展；

- 鲁棒性与可靠性：前面的实验中相信大家都发现了，对同样的问题重复实验的话，LLM 可能会给出不一样甚至截然想反的答案；对于某些问题，如果 LLM 无法找到匹配度高的信息，为了尽可能提供 LLM 理解的“令人满意的”答案，它甚至可能编造事实，意图让生成的结果看上去合情合理，这种情形一般被称为 LLM 的“幻觉 *hallucination* ”；
- 能耗：目前 LLM 的训练和推理过程都需要消耗高额的算力，也就是能耗极高，是否与其产生的价值相匹配，也是业界时常争论的问题。

4 AI 带来的全新挑战

这一波 AI 的迅猛发展使得很多重要的配套工作出现滞后，包括教育、社会伦理、法律法规、信息安全等领域的重要基础建设，都出现无法匹配的趋势，这将是全人类未来若干年都必须面对的一个重大挑战。这里先简单列出一些大家都能感受到的问题，后面我们还会展开其中部分问题。

4.1 教育方面的挑战

4.1.1 教学模式转型压力

人工智能工具能以人类无法比拟的尺度去学习和整理现有知识，人类要获取知识变得更容易更快速，单纯传递知识的教学价值显著降低；另一方面，为了更好发挥人工智能的作用，需要人类更好地提出问题，并把更多精力花在对结果的评估与优化迭代上，这就需要教师从知识传授者转向思维引导者，但教育体系调整滞后，教师培训面临挑战。

4.1.2 学术诚信与技能弱化压力

人工智能工具的快速普及，学生可能会过度依赖 AI 来完成作业或考试，导致一些基础思维能力退化；同时目前检测 AI 作弊的

技术与规则尚未成熟，对现行教学与测评体系都会产生冲击。

4.1.3 教育资源公平压力

发达地区更易获得先进的人工智能教育工具，可能扩大全球或地区间的教育鸿沟。

4.2 社会伦理和法律方面的挑战

4.2.1 责任归属模糊

现有的人工智能工具能遵循人类指令生成几乎任何类型的文字及其他媒体内容，但人工智能并不能代替人类去履行相应社会及法律责任，这里面存在大量需要填补的空白，比如近年来多次引发广泛讨论案例，涉及 AI 绘画的版权问题，自动驾驶和 AI 医疗诊断环节的法律责任问题等。

4.2.2 算法偏见与公平性

如果训练数据中存在固有社会偏见，可能应用基于 AI 做出的很多判断，如发生在人员招聘、信贷等领域则容易发生争端甚至法律风险。

4.2.3 后真相时代

人工智能生产的文字、图片和视频越来越逼真，进一步加剧了互联网信息的复杂性，对事实查证、维护虚拟空间法治带来更大的难度。

4.2.4 隐私保护

用户可能在于人工智能工具的交互中不经意地泄露个人隐私信息，而现有人工智能技术的特点使得这些信息的追踪与保护非常困难；这些数据采集的边界相当模糊，如果企业缺乏自律则易于

产生过度收集、泄露用户敏感信息的风险，这对人工智能产品提供商及监管部门都提出了新的要求。

4.3 信息安全方面的挑战

4.3.1 基于人工智能的新攻击手段

人工智能工具可能被恶意用于自动生成钓鱼内容、绕过验证系统，甚至制造深度伪造（Deepfake）进行舆论操控，防御难度大幅提升。

4.3.2 人工智能系统本身的脆弱性

通过精心设计的提示词，恶意用户可以误导人工智能工具做出错误决策（如交通标志误识别），对人工智能工具进入关键和敏感的领域需要非常慎重。

4.4 继续探索！

把上面实验中对结果不太满意的提示词记录下来，换用别的 AI 产品试试，比如 DeepSeek，看看有什么不同的效果。

也可以考虑调整提示词的内容，力求提供给 LLM 更准确更清晰的指令，看是否能优化结果质量。