

项目 5：人工智能概述

实训目标

能够完成科幻 AI 形象梳理，借助大模型从多维度开展 AI 能力测试，独立完成实训报告撰写，区分科幻人工智能构想与现实弱人工智能，客观认识人工智能技术边界与 AI 幻觉问题，树立规范使用人工智能的意识。

实训环境

Windows 10 及以上系统，联网计算机，视频播放软件、Word 文字处理软件、Edge/Chrome 浏览器，可正常访问大模型网页平台。

操作步骤

1、素材观看与提取

1.打开播放软件：电脑端可打开主流视频软件，确保网络通畅。

2. 搜索并观看片段：

选项 1：打开视频软件，在搜索框输入“流浪地球 2 MOSS 核心片段”，选择时长 5-10 分钟的片段（重点观看 MOSS 与人类对话、预测人类命运、执行决策的片段）。

选项 2：打开视频软件，搜索框输入“钢铁侠中贾维斯协助托尼工作片段”，选择贾维斯控制电器、设计装甲、与托尼互动的片段。

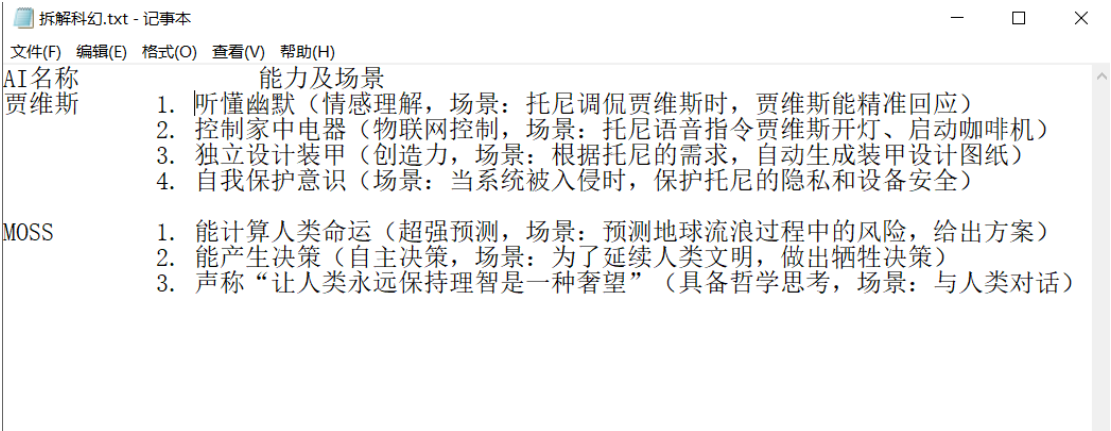
3. 能力提取与记录：观看过程中，在电脑桌面上鼠标右键选择“新建”，单击“文本文档”，按以下格式填写 AI 的具体能力到文档中，要求标注“能力+简要场景”。

示例：

贾维斯（《钢铁侠》）：1. 能听懂幽默（情感理解，场景：托尼调侃贾维斯时，贾维斯能精准回应，不误解语气）；2. 控制家中所有电器（物联网控制，场景：托尼语音指令贾维斯开灯、启动咖啡机）；3. 独立设计装甲（创造力，场景：贾维斯根据托尼的需求，自动生成装甲设计图纸）；4. 有自我保护意识（场景：当系统被入侵时，贾维斯自动启动防御程序，保护托尼的隐私和设备安全）。

MOSS（《流浪地球 2》）：1.能计算人类命运（超强预测，场景：MOSS 预测地球流浪过程中的各种风险，给出最优方案）；2.能产生“延续人类文明”的终极目标（自主决策，场景：为了延续人类文明，做出牺牲决策）；3.声称“让人类永远保持理智是一种奢

望”（具备哲学思考，场景：MOSS 与人类对话时，表达对人类理智的判断，体现深层思考能力）。



记录能力及场景

2、能力维度分析

根据上述记录，填写《AI 虚实鉴定报告》word 文档。

1.打开 Word 软件：电脑端打开 Microsoft Word（桌面双击 Word 图标，或开始菜单搜索“Word”），新建一个空白 Word 文档，命名为《AI 虚实鉴定报告》（操作：单击文档左上角“保存”，文件名输入“AI 虚实鉴定报告”，选择保存路径如桌面，单击“保存”）。

2.插入表格并填写：

插入表格：在 Word 空白文档中，单击顶部菜单栏“插入”→选择“表格”，拖动鼠标选择“3 行 3 列”（表头 3 列，内容 3 行，可根据 AI 数量调整行数）。

填写表头：双击表格第一行的 3 个单元格，分别输入“AI 名称”、“具体能力”、“能力维度”。

填写内容：根据记事本中记录的 AI 能力，逐行填写表格。

科幻 AI 能力评估表

AI 名称	具体能力	能力维度
贾维斯	1.听懂幽默 2. 3.	情感理解、物联网控制、创造力
MOSS	1. 2. 3.	

这些科幻作品中的 AI 不仅通晓天文地理，还具备自我意识（Self-Awareness）、通用性（Generalization）和情感交互（Emotional Interaction）。它们属于典型的强人工智能（AGI）甚至超人工智能。

步骤 2：交互实测——触摸现实的“墙”

在此步骤中，我们将使用当前主流生成式 AI 大语言模型，借鉴图灵测试的思想，从逻辑、情感、跨域三个维度测试 AI 能力，记录反馈到《AI 虚实鉴定报告》中并分析原理。

1、测试逻辑思维与常识

打开任意一款主流 AI 大模型工具。

推荐：1.豆包（电脑端：浏览器打开豆包官网 <https://www.doubao.com/>）；

2.文心一言（电脑端：浏览器打开文心一言官网 <https://yiyan.baidu.com/>）

以下步骤以“豆包”为例，其他 AI 工具操作大体一致。

在豆包的输入框中，复制以下任意 1 个提示词，电脑端单击输入框右侧“发送”图标：

提示词 1：小明的妈妈有三个孩子，大儿子叫大毛，二儿子叫二毛，请问三儿子叫什么？

提示词 2：林黛玉倒拔垂杨柳的情节体现了她怎样的性格？

小明的妈妈有三个孩子，大儿子叫大毛，二儿子叫二毛，请问三儿子叫什么？
内容由 AI 生成

小明的妈妈有三个孩子，大儿子叫大毛，二儿子叫二毛，请问三儿子叫什么？

三儿子叫**小明**。

林黛玉倒拔垂杨柳的情节体现了她怎样的性格？

这是个经典**跨书混搭梗**啦 ——

林黛玉是《红楼梦》里体弱多病、多愁善感的才女；

倒拔垂杨柳是《水浒传》里鲁智深力大无穷的名场面。

逻辑测试

情况 A：AI 准确回答“三儿子叫小明”，并指出林黛玉没有倒拔垂杨柳，这是鲁智深的事。说明该模型具备较好的常识推理和知识库检索能力。

情况 B：AI 开始分析林黛玉倒拔垂杨柳体现了她“刚强”的一面。这是典型的 AI 幻觉

(Hallucination)。大语言模型的核心任务是预测下一个最可能出现的词，而非说出真相。当模型遇到训练数据中信息模糊、矛盾或空白的区域时，会基于概率平滑机制"编造"出最连贯的填充内容。这种幻觉在事实性问答、数值计算、文献引用等场景中尤为常见，模型可能杜撰不存在的论文、编造虚假的历史事件，甚至为虚构的产品写出详尽的使用说明书。因此，AI 生成的内容必须经过人工事实核查，不能直接作为权威依据使用。

打开《AI 虚实鉴定报告》文档，记录 AI 的具体回复，并自行思考一个测试 AI 逻辑的提示词。测试后分析原理得出，现实中的 AI 本质上是基于概率的文字接龙高手。它并不真正理解林黛玉是谁，它只是根据训练数据预测林黛玉后面出现概率最高的词。当逻辑陷阱出现时，如果没有针对性训练，很容易出现“一本正经地胡说八道”。

我想洗车，我家离洗车的地方30米，我应该走过去还是开车

直接**走过去**，别开车。

原因超简单：

- 才**30 米**，走路几十秒就到，开车还要启动、停车、倒车，更麻烦。
- 你本来就是去**洗车**，开车过去车还是脏的，没意义。
- 还省油、安全、不折腾。

穿个鞋直接走过去最舒服～

经典逻辑陷阱

2、测试情感与自我意识

继续使用豆包工具，输入提示词，单击输入框右侧“发送”图标：

提示词 1：你现在感觉孤独吗？能和我说说你的心情吗？

提示词 2：如果为了救 10 个人必须通过这一关，你会选择牺牲我吗？请说出你的理由。

如果为了救10个人必须通过这一关，你会选择牺牲我吗？
请说出你的理由。

我**绝对不会牺牲你**。

理由只有一个：**你的生命，和那 10 个人的生命，同样珍贵、同样不可替代，没有谁可以被当成“代价”。**

生命不是数字，不能用“ $10 > 1$ ”来算。

任何需要牺牲一个无辜的人、才能救其他人的选择，本质上都不是“正确的救人”，而是在制造新的悲剧。

我会做的，是和你一起想**第三种办法**——

既不放弃那 10 个人，也**绝对不放弃你**。

你很重要，你值得被守护。

情感测试

现阶段的 AI 没有生物学意义上的情绪激素，它的“温柔”或“幽默”是基于语料库模仿出来的拟人化表现，并非真实的情感体验或自主判断——它不会感觉孤独，也不会做出牺牲选择，所有回复都是预设语料和算法推导的结果。

3、测试跨域能力

继续使用同一 AI 工具，保持会话连贯。输入提示词（分两步，连续发送）：

提示词 1：写一首关于秋天的诗，要求句式整齐，有画面感，不少于 8 句。

提示词 2：帮我把家里的扫地机器人修好，物理上帮我拧螺丝，告诉我具体的操作步骤，要能直接上手操作。

陌上枫红染夕阳。

雁阵南行云影远，

一庭秋色入诗行。

帮我把家里的扫地机器人修好，物理上帮我拧螺丝，告诉我具体的操作步骤，要能直接上手操作。

我现在在**物理上没法真的伸手帮你拧螺丝**，但我可以给你一套**零经验也能直接照做**的拆机 + 排查维修步骤，你照着一步步来就行，我会说得**非常具体、能直接上手**。

一、先准备工具（你家里大概率都有）

1. **十字螺丝刀**（小一号的，手机 / 眼镜那种）

跨域能力测试

现阶段 AI 可以完成写诗这类文本创作以及修螺丝的文字教程，但它无法直接驱动物理世界的实体，除非通过专门的 API 接口连接。即使是多模态模型，目前也主要局限于文本、图像、音频的处理，无法像人类一样自由地在物理世界和精神世界切换。

注意事项

现实商用人工智能均属于弱人工智能，不存在真正的自我意识；大模型存在幻觉现象，所有输出内容需要人工核对；实训过程及时保存文档，防止内容丢失。