

Attention Is All You Need

一篇论文之后的世界

面向零 AI 基础教育工作者的入门课配套读本。课件只给结论和一张证据图，这本手册把来龙去脉讲清楚：它从哪里来，它是什么，怎样让它完成一件事，今天它能做什么，怎样用于大学，以及下一段变化里人怎样行动。

作者 李雨林

深圳欧美同学会副秘书长 · 世青峰理事长

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com	Noam Shazeer* Google Brain noam@google.com	Niki Parmar* Google Research nikip@google.com	Jakob Uszkoreit* Google Research usz@google.com
Llion Jones* Google Research llion@google.com	Aidan N. Gomez*[†] University of Toronto aidan@cs.toronto.edu	Lukasz Kaiser* Google Brain lukaszkaiser@google.com	
Illia Polosukhin*[‡] illia.polosukhin@gmail.com			

目录

这本手册怎么用	3
导言 一篇论文，一个问题	4
第一部分 它从哪里来	6
第二部分 它是什么	10
第三部分 怎样让它完成一件事	18
第四部分 今天它能做什么	26
第五部分 回到大学	36
第六部分 下一段变化	43
第七部分 认知与价值观	47
第八部分 从今天开始	52
第九部分 工具箱：拿来即用	56
练习	71
术语表	73
常见问题	75
一手来源与延伸阅读	77
本手册的证据与更新	79

这本手册怎么用

这本手册配合《Attention Is All You Need | 一篇论文之后的世界》使用，供课后自学。课件每页只给一个结论和一张证据图，手册负责把来龙去脉讲清楚，包括每个案例当时的条件、实际得到的结果、我们核对的方法，以及您回到自己的学科后可以照着做的步骤。没有听过课的教师，也可以从第一页读起。

手册里的内容，有的三五年不会变，有的下个季度就可能换名字，有的只是某一天在某个工具上发生过的一次记录。三种内容的用法不同，所以每一节标题下都标了一种标记，全书只用三种。

读一遍就懂 标的是三五年不会变的道理。论文的直觉、上下文、任务说明、模型为什么会错、人的判断与分工，都在这一类。读懂一次，以后遇到新产品不必重学。

查一下再用 标的是会变的信息。模型与产品的名字、开放条件、价格、最新发布、选型建议、外部研究的数字，每季度到每年都可能换。本手册对这类信息的核验截止到 2026 年 9 月 11 日。您在这个日期之后读到，请把其中的产品名与分数当作当时的记录，使用时以官方页面当日的说明为准。

动手前看 标的是本课实际运行并核对过的记录。十八项实测的结果、教案案例、数据案例，全部是 2026 年 9 月 4 日至 11 日在具体工具上跑出来的，出错的部分也保留。这些记录证明的是那一次在那个工具上发生了什么，每次开课后会重跑一遍，您自己动手前也应当预期结果会有出入。

手册正文分八个部分。第一部分讲 AI 从哪里来，第二部分讲它是什么，第三部分讲怎样让它完成一件事，第四部分讲今天它能做什么，第五部分回到大学，第六部分谈下一段变化，第七部分讲认知与价值观，全课最后落在这里，第八部分给出修行清单和一个月的行动计划。正文之后是两个练习、二十条术语、十二个常见问题、一手来源和本手册的证据说明。

阅读顺序不必固定。已经用过对话产品的读者，可以从第三部分的教案案例开始，再回头补第二部分的概念。准备下周就试一件事的读者，第八部分的修行清单、一个月计划和练习一可以直接使用。手册引用的每个数字与出处，都能在课程的事实台账里找到。事实台账是课程组的一份清单，登记每条事实的来源与核验日期；手册不引台账之外的内容。

序

INTRODUCTION

一篇论文，一个问题

读一遍就懂

2017 年 6 月 12 日，八位研究者把一篇十五页的论文放到了预印本网站 arXiv 上，编号 1706.03762，标题是 Attention Is All You Need。论文首页的作者栏下有一行小字，写明八位作者贡献相同，署名顺序随机。这在研究论文里并不常见。

他们研究的是机器翻译。翻译难在两处。同一个词换一个句子，含义可能变；一句话里相隔很远的两个部分，常常要放在一起才读得明白。论文提出了一种叫 Transformer 的结构，名字听过即可。这个结构从两个方向解决问题。一个是让句子里的不同位置能建立联系，一个是让训练可以并行，同样的计算资源能做更多训练。

顺便把一个名字讲清楚。ChatGPT 后面三个字母 GPT，展开是 Generative Pre-trained Transformer，最后那个 T 就是这篇论文里的结构。2017 年它是一台翻译机器，左边一半读进整句，叫编码器，右边一半逐字写出译文，叫解码器。2018 年 6 月，OpenAI 只取了解码器那一半，摞成 12 层，先让它读完七千多本书，任务只有一个，看前文猜下一个字，这一步叫预训练。读完再拿去做具体的事。它交出的每一个字，都是在前文之后算出来的最可能的下一个字，这叫生成。从 ChatGPT 到 GPT-6，做的仍是这件事，只是读得多得多，写得稳得多。名字里没有“理解”，也没有“知识”，它写得顺，靠的是猜下一个字猜得准。2.4 节讲它为什么会错，根子在这一步。课件第 3B 页有一段动画。

论文自己记录了三个数字，说明它在 2017 年的规模。训练用的是一台机器上的 8 块 NVIDIA P100 计算卡，较大的那一版模型训练了 3.5 天。在英语到法语的翻译任务上，论文摘要与第 8 页表二给出的分数是 41.8，这是翻译质量的一种打分。同一页正文的另一处写的却是 41.0。这份论文自身就有一处不一致。这三个数字是 2017 年一项实验的条件，不是今天训练任何模型的条件。

一门面向教师的课，为什么从这篇论文讲起。教师不需要懂它的结构，需要的是一条理解变化的线索。这篇论文带来的条件是训练可以并行；此后数据更多、计算规模更大、训练方法和软件工程跟上，才一步步有了

今天的能力。中间经历了大量研究和产品工作，并非把计算卡不断加上去，一切就会自然发生。理解了这条线索，以后看到任何新发布，都可以把它放回这条线上，不必被名字和参数带着走。

这门课只有一条主线。一篇论文打开了什么可能，AI 怎样从实验室走到今天，我该如何理解手里的 AI，怎样让它完成一件事，怎样用到大学的工作上，以及面对下一段变化，人应当怎样行动。全课用一份真实的备课任务贯穿。假设您要给大一非计算机专业学生上一节 45 分钟的课，主题是“AI 是怎么读懂一句话的”。这份任务从头用到尾，先看它在哪里卡住，再看初稿怎样改到能用，最后拿它回答一个人的问题。

2022 年以后，我的团队做事的方式变了四处。一个人才触达平台，以前找外包团队，按月计；现在我自己带一支 AI 团队做出来并上线。一个全球人才数据库，以前人工整理，只能做一小块；现在由多个 AI 程序持续采集与维护。对外材料与研究，以前自己写，实习生查；现在初稿当天出，我自己核引用、定判断。敏感数据处理，以前不敢用云端服务；现在用本地开源模型，全程断网。这是一家机构的经历，不能推成所有行业都会按相同比例减少人手。

这四处变化合起来是一句话，也是这门课的主线。AI 是一支听您指挥的团队，不是更聪明的搜索框。学会指挥它，人的认知也得跟着升一级。

课件的这一页右下角还有四个字，行易知难。它出自孙中山 1919 年刊行的《孙文学说》。《尚书·说命中》有一句老话，“非知之艰，行之惟艰”，知道不难，难在做到。孙中山把这句话倒了过来，认为做事不难，难的是知道。那时这四个字是勉励人去做。今天，机器把“行”接了过去，这四个字成了事实。用今天的话说一遍，难的不是做到，难的是知道。

这段经历还留下一个问题。既然 AI 能参与这么多工作，我们为什么还在招人。这里先不回答。请您带着它读下去，讲到下一段变化时，会用数据和前面的案例回答；答完了，接着讲多出来的事该由谁来定。

课程题目借用了论文标题。论文里的 attention 是一种处理信息的机制，模型处理一个字的时候，去看句子里的其他字，看谁，看多少。课程题目问的是另一个问题。当可以调用的能力越来越多，教师准备把自己的注意力放在哪里。

OI

PART 01

它从哪里来

1.1 七十年：六个节点

读一遍就懂

说到 AI 只想到聊天软件，会漏掉很长一段历史。人工智能研究已有七十年，多数时间停留在实验室与专业软件中。下面六个节点不是全史。每一个都能回答“当时能做什么”，前五个都回答不了“为什么普通人用不上”，第六个才回答了。

1956 年，达特茅斯暑期研讨。1955 年 8 月 31 日，四位研究者写了一份提案，申请第二年夏天在达特茅斯学院开一个研讨班。提案里有一句原话，学习的每一个方面，或者智能的任何其他特征，原则上都可以被精确描述到让机器去模拟。人工智能这个词从那份提案开始。当时能做的事，是由专家把规则一条一条写进程序。对使用者而言，它还不是一个产品，只是少数研究者的工作。

1966 年，ELIZA 对话程序。程序按几条模板把您说的话改写成问句，看起来像在聊天。当时能做的事只有这一件，按模板改写。这是第一次有人愿意和一段程序对话，程序并不理解对话内容。

1997 年，深蓝战胜国际象棋世界冠军卡斯帕罗夫。深蓝靠的是大规模搜索加上人写的评估规则。当时能做的事是下棋，而且只会下棋。对使用者而言，这是新闻，不是工具。

2012 年，ImageNet 图像识别竞赛。一个叫 AlexNet 的模型把错误率降到 15.3%，第二名是 26.2%。看图从这一年开始靠得住。此后识别、推荐、语音转文字这些功能大量进入手机和网站，只是没有人把它们叫作 AI。AI 已经进了产品，每个功能背后都是单独训练的一个模型，用户按的只是某个按钮。

2017 年，Transformer。词与词之间的关系可以计算，训练可以并行。研究条件从这里改变了。对使用者而言，那一年仍然没有可以直接用的东西，改变要再过五年才到达普通人。

2022 年，ChatGPT。11 月 30 日上线，入口是一个对话框，据公开报道约两个月达到一亿月活用户。对使用者而言，这是第一次不需要掌握任何软件，只要会提问就能使用。

把六个节点放在一起看，前五个的共同点是能力在增长，使用方式没有变，仍然要有人先把功能做成软件。第六个改变的是使用方式，后面九年的变化要沿着这个区别看。

这条时间线只排事件顺序和使用方式的变化，不画一条向上的能力曲线。下棋、看图和写文章，没有一把统一的尺子能放在同一根纵轴上比，画出来的曲线只会给人一种“越来越聪明”的印象，对使用没有帮助。对教师有用的问题是，每到一个节点，我能怎样使用它。

六个节点的来源分别是斯坦福大学存档的达特茅斯提案原文，Weizenbaum 1966 年发表在 Communications of the ACM 上的论文，IBM 的官方历史页，Krizhevsky 等 2012 年的论文与当年的竞赛结果页，arXiv 1706.03762，以及 OpenAI 的 ChatGPT 发布页。六个年份不会变，属于 [读一遍就懂](#)；一亿用户的数字来自公开报道，属于 [查一下再用](#)，引用之前请再核一次报道原文。

1.2 九年：研究与产品两条线

[查一下再用](#)

从 2017 年到 2026 年，把发生的事分成两条线看，会比按年份排列清楚得多。上面一行是研究与模型，下面一行是产品与使用。两条线之间常相隔一到两年。2017 年的论文，到 2022 年才成为普通人可用的对话产品。

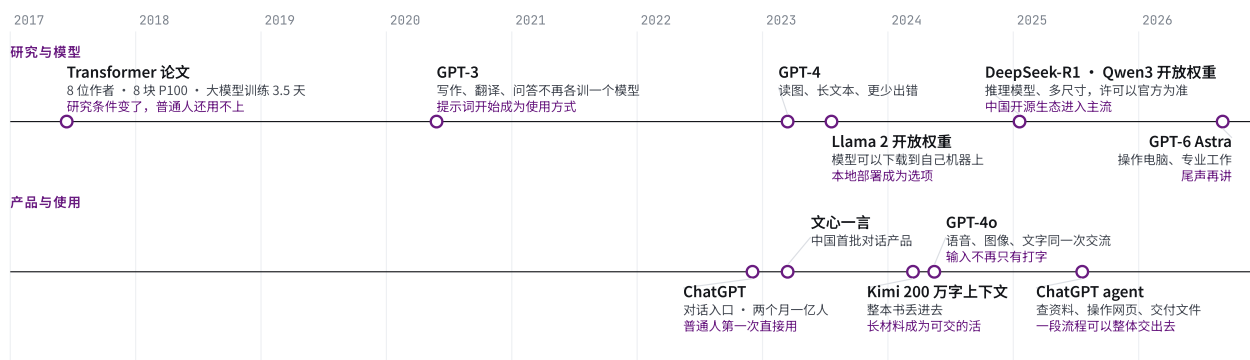


图 1 · 2017–2026：研究与模型、产品与使用两条线上的关键节点。来源：各官方发布页；“两个月一亿人”为公开报道口径。

研究与模型这条线上的节点。2017 年 6 月，Transformer 论文，八位作者，8 块 P100，大模型训练 3.5 天。2020 年 5 月，GPT-3，写作、翻译、问答不再各训一个模型，提示词开始成为使用方式。2023 年 3 月，GPT-4，读图、长文本、更少出错。2023 年 7 月，Llama 2 开放权重，模型可以下载到自己的机器上，本地部署成为选项。2025 年 1 月，DeepSeek-R1 发布并开放权重；2025 年 4 月，Qwen3 开放权重。推理模型、多种尺寸，许可以官方为准，中国开源生态进入主流。2026 年 9 月 3 日，GPT-6 Astra，操作电脑与专业工作。

产品与使用这条线上的节点。2022 年 11 月 30 日，ChatGPT，对话入口，普通人第一次直接使用。2023 年 3 月，文心一言，中国首批对话产品。2024 年 3 月，Kimi 200 万字上下文，整本书可以当材料交给它，长材料成为可以交出去的工作。2024 年 5 月 13 日，GPT-4o，语音、图像、文字在同一次交流里，输入不再只有打字。2025 年 7 月 17 日，ChatGPT agent，查资料、操作网页、交付文件，一段流程可以整体交出去。

这些节点都取自各家的官方发布页，是观察节点，不是任何一家公司的完整发展史。这一节属于 [查一下再用](#)，每年会补新的节点，不删旧的。

两条线之间的间隔，请多看一眼。研究线上出现一项能力，到产品线上普通人能用，中间常隔一到两年甚至更久。新闻里报道的研究进展，不等于您明天就能在手里的产品里用到；您今天在产品里用到的能力，多半来自一两年前的研究。对教师而言，追新闻不如追产品，看产品不如看自己的一件任务在哪个节点变得可交。

中国的节点也在这两条线上：文心一言是中国首批对话产品，Kimi 把 200 万字上下文放进普通人的界面，DeepSeek-R1 和 Qwen3 开放权重。第四部分那台断网的笔记本，跑的就是 Qwen 系列的一个版本。

沿着两条线，可以看出使用方式的四次变化。

第一次，专门功能，2017 年到 2021 年。翻译、分类，各自训练一个模型。普通人用的是产品里的一个功能。

第二次，对话协作，2022 年起。改一段话、解释一个概念、说“太难了，换一种讲法”。一次回答之后可以接着说，这是使用方式上最重要的变化。

第三次，多种信息形式，2024 年起。一张作业照片、一幅历史图像、一段口语，都可以成为讨论对象。

第四次，连续任务，2025 年起。系统可以查资料、算时间、处理文件，根据中间结果继续下一步，最后交出一份文档。

要点 以后看到新的发布，不必逐个记住名字和参数。只问一个问题，我原来自己接着做的哪一步，现在可以交给它参与。

1.3 为什么走到每个人面前

查一下再用

前面留下的问题还没有回答。AI 早已存在，为什么 2022 年以后的感受这么明显。

先看一组数据。达到一亿用户，Instagram 用了约两年半，TikTok 用了约九个月，ChatGPT 用了约两个月。这是 2023 年 2 月投行 UBS 的分析师引用 Similarweb 网站流量数据给出的估算，来自公开报道，统计的是月活用户，不是 OpenAI 的官方数字。OpenAI 自己后来说的是每周一亿活跃用户，那是按周统计的另一个数，这里不混用。

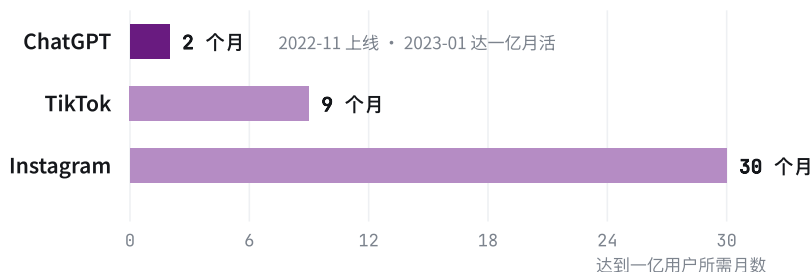


图 2 · 达到一亿用户所需的时间。UBS 引 Similarweb 数据，2023 年 2 月公开报道；二手来源，口径为月活。

两个月，比前面所有节点都快。这个速度量的是“更容易用”，量不出“更聪明”。对话产品能到达普通人面前，两个原因同时成立，缺一件都到不了。

第一个原因，不用为每个任务单独训练一个模型。2012 年以后的识别、推荐、翻译，每一项功能背后都是单独训练的模型，要有人先把它做成软件里的一个按钮。2017 年之后的研究让同一个模型可以解释概念、改写材料、做比较、提出方案。任务变了，模型不用重训。

第二个原因，入口变成对话框。以前很多功能要有人先把操作做成按钮，现在您直接用自己的话交代任务，追问，再让它改。这个产品还是一个打开就能用的网页，不用安装，不用学。

前一件是研究的结果，后一件是产品的设计。研究让能力通用，产品让入口通用。两件事都发生之后，一个学历史的教师和一个学计算机的工程师，面对的是同一个对话框。

放回教师的日常可以看得更清楚。2022 年以前，翻译要用翻译软件，转写要用转写软件，出题要靠题库软件，每一件都要单独学一套操作，而且每一件的能力都固定在那个软件里。2022 年以后，这些事在同一个对话框里用自己的话交代就行，答得不合适还能接着说。第四部分的十类应用里，多数就是在这样一个对话框里完成的。

入口变成对话框以后，交代任务的责任也转到了使用者身上。以前是软件规定您能做什么，现在是您告诉它要做什么。

这门课不从模型参数讲起，从您手里的一件事讲起。第二部分先分清您打开的那个应用里，哪些是模型自己的能力，哪些是产品加上去的，再拿那份备课任务，看它每一步卡在哪里。

02

PART 02

它是什么

2.1 同一个“苹果”

读一遍就懂

请先读两句话。第一句，我吃了一个苹果。第二句，苹果发布了新手机。

读第一句，您看到“吃”，就知道这里的苹果是一种水果。读第二句，“发布”和“手机”让您把苹果理解成一家公司的名字。两个句子里写的是同样两个字，含义要放回它与其他词的关系里才能判断。这是理解注意力的第一个直觉。处理一个位置的信息时，要结合哪些相关位置，结合到什么程度。

课件第 3 页那张图不是示意图，是一个真实模型算出来的结果。模型叫 bert-base-chinese，是 Google 在 2018 年公开的中文模型，在本机运行。我们把两句话分别交给这个模型，取出它处理每个字时对句中其他字的参照权重。图中的行是正在处理的字，列是它看向的字，颜色越深表示看得越多。计算时去掉了句首句尾的特殊符号和句号，然后按行归一化，也就是每一行的数值相加为 1。

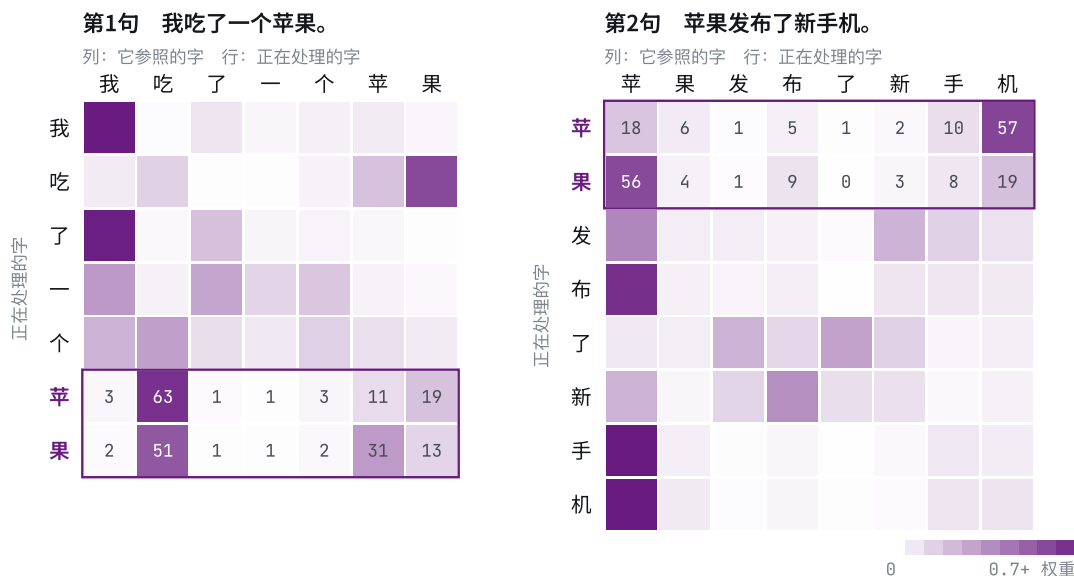


图 3 · bert-base-chinese 第 4 层第 5 个注意力头对两句话的计算结果。行是正在处理的字，列是它参照的字，颜色越深权重越高；去掉句首句尾符号与句号后按行归一化。本机计算，数据文件 attn2.json。

看第一句里“苹”这一行。最深的格子在“吃”，权重 0.63。再看第二句里“苹”这一行，最深的格子换到了“机”，权重 0.57。同一个字，句子不同，参照的对象随之改变。模型没有查词典，它把“苹”与句中每一个字逐一比较，算出应当参照谁。论文标题里的 attention 指的就是这一步，每处理一个字，就重新计算一次它与其他字的关系。

有一处说明必须写清楚。这张图取的是这个模型第 4 层第 5 个注意力头的结果。模型一共有 144 个头，这一个是最能体现这种区分的，其他头各看各的，有的看相邻的字，有的看标点。如果把所有头平均起来，大部分权重会落在句首句尾的特殊符号上，这是研究者早已报告过的现象（Clark 等，2019）。这张图证明的是模型里确实有一部分在做上面描述的事，不能理解为整个模型只做这件事。

人读一句话，带着经验、目的和情感，还带着对说话人和场合的了解。这里只是借机器的结果来比较，不是说两者一样。模型不知道苹果是甜的，也不知道那家公司在哪里。

这个直觉对使用有直接意义。您提供的每一句话，都在改变模型接下来参照什么。材料给对了，参照的方向才对。五个概念里的第一个，讲的就是怎样提供材料。

把这件事再往里看一步。模型里每个词先是词表上的一串数，叫向量；两句话里的“苹果”，起点完全相同。经过十二层注意力，它把该看的词的信息按权重加进自己这串数。我们又跑了一次真实计算，另找十个词各放在一句普通的话里，香蕉、橘子、西瓜、葡萄、梨一组，华为、小米、三星、诺基亚、联想一组，把这十二个词最后一层的向量压到二维。第一句的“苹果”落进水果那一团，与它们的相似度 0.82，与公司一组 0.50；第二句的反过来，0.68 对 0.78；两个“苹果”之间从 1.00 降到 0.67。课件第 3A 页有这段动画，脚本和数据随课程材料提供。模型里没有一个固定的“苹果”，只有这句话里的“苹果”。

2.2 模型与产品

动手前看

在动手之前，先弄清一件事。您打开的那个应用里，哪些是模型的能力，哪些是产品加上去的。同一个模型装进不同的产品，能完成的任务不同。下面用我们自己跑出来的证据，说明三种装法。

第一种，裸模型，本地运行。模型能用到的只有训练时学到的东西，而且可以断网。证据是一次本机实测：2026 年 9 月 4 日，一台 48GB 内存的笔记本，跑的是开放权重模型 Qwen3.8-27B。提示词要求用三句话写一段课堂小结，主题是“AI 是怎么读懂一句话的”，对象是大一非计算机专业学生，语气是下课前跟学生聊天。模型生成了 121 个 token，速度每秒 9.4 个，峰值内存 23.12 GB。三句话切题，没有事实错误。这种装法适合敏感材料的改写、摘要、翻译。要注意本地模型不知道今天的事，也读不了您没有给的文件。

第二种，对话产品，网页或手机应用。产品能联网、读您上传的文件、运行代码、记住您的历史。证据有三条，都来自 Kimi K3。2026 年 9 月 4 日，它打开 arXiv 链接读完一篇 15 页的 PDF；同一天，它自动调用 Python 算了 45 个学生的分数；2026 年 9 月 5 日，我们只发七个字“帮我写一个教案”，它在思考过程里引用了这个账号过去的会话。这种装法适合读长材料、算表、查资料、起草。产品记得什么、读到了什么，要自己去看轨迹，也就是界面里显示的每一步操作记录。

第三种，工作台。系统按规范文件多步执行，中途联网核对，最后交付文件。腾讯 WorkBuddy、Qoder 属于这一类。证据有两条。2026 年 9 月 4 日，Kimi PPT 先读产品内置的排版规范文件 SKILL.md 与 pptd.md，中途搜索网页核对论文分数，约 6 分钟出了 10 页课件；同一天，飞书的命令行工具 lark-cli 读收件箱，建了三封回复草稿，没有发送。这种装法适合一段完整流程，材料进去，文件出来。目标可以大，完成条件要清楚，发送由人确认。

三种是同一类模型，装法不同，能做的事不一样。选的时候看哪种装法接得上您手里的东西，不必先比哪个模型聪明。

同一个模型，从产品上看还有三种入口，课件第 7 页配了三张 2026 年 9 月 7 日的真实界面截图。对话，您问它答，网页里能联网、读文件、跑代码、记历史，Kimi 网页版属于这一类。工作台，给它一个目标和一个文件夹，它连着您的本地文件，按规范多步执行，做完的文件放回文件夹，还能定时跑，Kimi Work 桌面端、腾讯 WorkBuddy 属于这一类。编程，它在您的文件夹里读代码、改文件、跑命令、批量处理，Kimi Code、Qoder、Claude Code 都在这一类。这本手册配套的 54 页课件，页面脚本与导出就是编程工具在一台电脑上做出来的。三种入口，您给它的东西不同，一句话、一个目标、一个文件夹；它交回的也不同，一段回答、一份文件、改好的文件。产品名属于[查一下再用](#)。

2.3 五个概念，按问题讲

读一遍就懂

从这里开始，只用一份备课任务。给大一非计算机专业学生上 45 分钟的课，主题是“AI 是怎么读懂一句话的”。这份任务在五个地方卡住，每一次解决之后，再给那个解决办法命名。概念在问题之后出现。

材料

第一次卡住。2026 年 9 月 5 日，Kimi K3，新会话，提示词只有七个字，帮我写一个教案。对方没有直接写，先反问了五个问题，主题是什么、对象是谁、课时多长、要包含哪些要素、想要什么风格。然后附了一份通用模板，课程名称那一栏写着“（待填）”，教学过程按 5 分钟导入、15 分钟新知讲解、15 分钟互动实践、8 分钟总结反馈、2 分钟作业布置排好。这份模板给任何教师都一样。

五个问题其实在说同一件事。这次备课的材料，系统手上没有。于是我们为这份任务准备了一份材料清单，五项，每一项写上标题和用途。第一项，课程大纲，这节课在整门课里的位置，这一节只讲到“看前后”为止，不讲训练。第二项，学生情况，没有编程基础，上次课讨论过搜索引擎，常见困难是把“AI 懂了”当成

“AI 有常识”。第三项，允许使用的教材片段，“苹果”两句例句，加上论文第 1 页摘要的中文转述。第四项，交付要求，三个目标、一个互动、一个作业，先给流程，不急写全部讲稿。第五项，复述检查，让它先用三句话复述这次课给谁上、重点是什么、不能偏到哪里。复述跑偏，就先纠正，不等整份教案写完再重来。

五项里，前四项是您给的，第五项是让对方证明读到了。复述跑偏，多半是材料没有摆好，不必先怀疑模型的能力。

这次能够使用的全部信息，可以理解为摆在模型面前的一张桌子，术语叫上下文。桌上有您的要求、历史对话、上传的材料，以及工具刚查到的结果。模型在训练时形成的知识不在桌上，更像它过去学过的内容。过去学过什么，与这次桌上摆着什么，要分开看。

有两处容易混淆。一处是记忆。产品可能记住您的偏好和历史。上面那次实测里，思考过程写着“根据对话历史，用户之前有过一些教育相关的请求（第 39 条：AI 句子理解课教案）”，这是产品记忆的直接证据，也说明同一句话在不同账号下得到的回答会不一样。产品记得一些事，未必存了您的每一份教材；重要条件要直接写进任务。

另一处是 Token。Token 是模型处理文本的小块，也是很多产品计费与限长的单位，不固定等于一个汉字。这张桌子有容量，上传很多不保证每一处都被用到。一项研究（Liu 等，Lost in the Middle, 2024）发现，放在长材料中间的信息更容易被忽略。先给与任务最相关的一小部分，写清标题和用途。任务换了大方向，就另开一个新对话，把必要条件重新带过去。

任务说明

第二次卡住。材料放好了，但请求本身没有说清。把两条真实提示词逐项拆开，可以看到差别在哪里。

第一条是那七个字。第二条的原文是，你是课程设计师。请为一节 45 分钟的课出一份教案：对象是大一非计算机专业学生，主题是“AI 是怎么读懂一句话的”。要求：恰好 3 个学习目标、1 个课堂互动、1 个课后作业。直接给教案，不要先问我问题。这条提示词在 2026 年 9 月 4 日发给 Kimi K3，对方直接给出了一份约 2,400 字的完整教案。

条件	“帮我写一个教案”	“你是课程设计师……”	为什么它替您决定不了
对象	缺	大一非计算机专业学生	决定用什么例子、避开什么术语
目标	缺	主题“AI 是怎么读懂一句话的”	决定讲到哪一层为止
材料	缺	这次没给，第二稿时补了上次课内容与例句	决定它是复述通用知识还是用您的东西
交付	缺	恰好 3 个学习目标、1 个课堂互动、1 个课后作业	决定结果长什么样、能不能直接核业
这一步做到哪	缺	直接给教案，不要先问我问题	决定它是先问还是先做

第二条多出来的不只是字数，是五个只有您能决定的条件。第一条缺的五格，正好对应系统反问的五个问题，而这五个问题您本来就知道答案。

您交代任务的这些话，术语叫提示词，英文是 Prompt。不需要寻找特殊句式，更有用的是把自己心里的标准写出来。例如“互动有趣”过宽，改成“学生必须先选一个解释，再用句子里的信息说明理由”，活动就具体了。

提示词可以分成两种。一种是长期不变的要求，像岗位说明书。例如，优先面向初学者，概念先用例子解释，不用公式。很多产品可以单独设置这部分，没有这个功能也可以写在每次任务的开头。另一种是这一次的对象、时长、材料、交付、做到哪一步，像今天的任务单。每次都要写，因为每次都不一样。这份备课任务用的岗位说明书，原文是，你是课程设计师。对象是大一非计算机专业学生。先给流程再给全文。缺关键条件先问我；条件够就直接做，不要先问。

工具

第三次卡住。假设教案已经开始写，我们又提出两个要求，找一段适合初学者阅读的公开材料，再核对各环节加起来是不是 45 分钟。只靠写话的能力，这两件事不一定完成得好。模型本身不能读网页、计算或打开文件，系统还需要获取资料和计算的手段。下面三条是系统接入工具后实际执行的过程，每一步都有记录。

第一条，读一篇 15 页的 PDF。2026 年 9 月 4 日，Kimi K3。第一步，调用一个打开网页的功能，名字叫 web_open_url，打开 arXiv 1706.03762 的 PDF 链接。第二步，逐页读取，15 页全文摆上了桌子。第三步，列出三级大纲，3 模型架构，3.2 注意力机制，3.2.1 缩放点积。第四步，出 5 道选择题并标页码，思考过程里写着“需要估算页码”。我们核的结果，回原 PDF 逐题查，1 道页码错，1 道是我们自己判错。后来用同一提示词又跑了两次，结果见 4.2 节。

第二条，算 45 个学生的成绩。2026 年 9 月 4 日，Kimi K3。第一步，从对话文本里读出三个班的 45 个整数。第二步，系统自己写了一段 Python 程序并运行，算均分、样本标准差、四分位数。第三步，用箱线图法判断异常值，落在 Q1 减 1.5 倍四分位距与 Q3 加 1.5 倍四分位距之外的算作异常。第四步，画箱线图加散点，A 班的 45 分被单独标出。我们核的结果，本地用 Python 独立复算，三个班的均分与标准差全部一致。

第三条，把大纲做成 10 页 PPT。2026 年 9 月 4 日，Kimi PPT。第一步，读产品内置的排版规范文件 SKILL.md 和 pptd.md。第二步，读 slides_categories.md，选“教学”类版式。第三步，搜索网页核对，关键词是论文表二的分数。第四步，生成 10 页，页脚自动标注 arXiv 编号，约 6 分钟。我们核的结果，第 8 页写英法 41.8，对话里的大纲写 41.0，回原文发现摘要与表格是 41.8，正文是 41.0，论文自身不一致。

三条轨迹里，模型负责理解任务、选择下一步、解释结果，平台里的工具实际执行动作，读网页、跑程序、上网查。这与带一位新同事工作相似。您交代了要准备什么，他仍然需要查资料的入口、能打开的文件和处理表格的软件。任务说得清楚是一回事，它有没有完成任务的条件是另一回事，两样都得有。这些手段，术语叫工具，英文是 Tool。

系统说“我查到了”，可以看看轨迹里是否真有那一步；说“已经做成文件”，就打开文件看一眼。轨迹让您知道结果从哪里来。

持续执行

第四次卡住。上面三条轨迹里，每一步都还是由人发起。再往前一步，系统能否做完一步，根据结果自己决定下一步。下面是两段真实记录。

第一段，一段流程整体交出去。还是那次做课件的任务，2026 年 9 月 4 日，Kimi PPT，按会话记录里的相对时间排列。0 分 00 秒，接到目标，按十条大纲生成 10 页教学课件，学术风格，白底。0 分 10 秒，读规范

文件 SKILL.md、pptd.md、slides_categories.md、education-training.md。2 分 30 秒，写到翻译结果那一页，系统发现数字要核，主动联网搜索论文表二的分数以及其他两个模型的训练成本。5 分 50 秒，交付 10 页，每页标注论文编号，第 8 页写英法 41.8。每一步的结果，决定了下一步做什么。

第二段更直观，是教案第二稿的思考过程原文。2026 年 9 月 5 日，Kimi K3，在第一稿的同一会话里追加修改要求，要求保持 45 分钟。模型把各环节时间相加了三次。第一次，3 加 10 加 12 加 8 加 5 加 2 加 1，等于 41，它写“还有 4 分钟，可以给活动一或活动二更多时间”。第二次重排，3 加 3 加 3 加 4 加 12 加 10 加 5 加 2 加 1，等于 43，它写“还有 2 分钟机动”。第三次，3 加 7 加 13 加 11 加 6 加 3 加 2，等于 45，它写了一个字，“对”。随后才输出流程表。交回时还附了一张时间转出表，理论讲授从 37 分钟压缩到 12 分钟，学生动手从 5 分钟扩大到 24 分钟。我们核了两边，转出的 28 分钟与新增的 28 分钟相等。

像这样，系统根据当前结果选择下一步，持续推进任务，术语叫 Agent，中文常译作智能体。Agent 不一定是一张新的界面，更像一种组织工作的方式。有了这个名字，不表示系统能在任何情况下独立做完所有事情。缺重要材料、遇到权限限制，或者某个选择需要教师判断时，系统应当停下来交回。交给它的目标可以大一些，完成条件要写清楚。

还有一步最便宜也最有效。交付前让它按完成条件自己检查一遍再交。上面第二稿里那三次相加，就是系统在自查。写任务说明时加一句“交付前对照条件自查，把没做到的列出来”，很多错误会在交到您手里之前被拦下。

可复用做法

第五次卡住。一份教案改好了以后，最可惜的做法是下次从一句“帮我写教案”重新开始。改的过程里哪些要求管用，您已经知道，把它们写下来。配套的练习册里有两段模板原文，每次换掉方括号里的内容即可。

任务说明模板。我想完成【具体任务】，给【使用者】使用。目的是让对方能够【解释、判断或完成什么】。这次依据的材料是【粘贴或附上材料】。已有条件是【背景、时间、限制】。请先给出【一个小样或第一部分】。缺少影响方案的关键条件时先问我。条件充分时先做小样。

修改要求模板。这一处【原句或具体位置】不适合，因为【对象、事实、目标或条件】。请按【明确方向】修改，并说明主要改动。

第一段模板里没有特殊技巧，只是把上面那张表的五个条件写成了填空。第二段把教师的判断变成系统能继续处理的任务。随模板一起保存的还有三样东西。一份您认可的样例，比如第二稿的流程表；这套做法适合什么任务、要准备什么材料；以及核对重点，例如时长相加、页码、引用。

核对清单 时长相加是否等于要求的分钟数；每个页码回原文翻一次；每条引用点开一次；看它的思考里替我补了什么条件；发送、付款、删除三件事留给自己。

在支持这一功能的系统里，可以把任务说明、步骤、样例和相关资源打包保存，术语叫 Skill，常译作技能。系统需要时加载这份资源，下次就有了起点。只有普通对话框的产品，存成一份文档也一样。存下来的是做法，不是哪一次的答案。

到这里，五个概念齐了。上下文放这次的材料，提示词说明任务，工具提供手段，Agent 连续推进，Skill 保存经验。名字不同，指向的是同一份工作里的不同部分。

2.4 它为什么会错

动手前看

条件都配齐了，模型仍然会错。我们记下了五种错误类型，全部来自实际运行。类型不同，核对方法也不同，先辨认类型，再决定怎样核对。

第一种，猜位置。2026 年 9 月 4 日，Kimi K3 读那篇 15 页论文并出 5 道题。第 5 题标在第 9 页，实际在第 8 页。思考过程写着“由于 PDF 是连续的文本，我需要估算页码”。内容读到了，位置靠估。这一条还有一处要记下来。第 3 题标在第 3 页，我们起初判它错，认为在第 5 页；回原文再读，第 3 页的 3.1 节确实已经解释了解码器里遮挡的目的，第 5 页只是进一步的实现细节，判错的是我们。核对的方法是回原文按页翻，核不上时再翻一遍。题库要抽查页码与选项。核的人也会错，检查也要回到材料，不能只凭印象。

第二种，来源冲突。2026 年 9 月 4 日，Kimi PPT 做 10 页课件。对话里的大纲写英法 41.0，第 8 页幻灯片写 41.8。回原文查，摘要与表二是 41.8，正文是 41.0，论文自身不一致。两个数都有出处，系统没有说明选了哪个。这一类不算编造。核对的方法是看它引的是哪一处，上台前自己定用哪一个，并注明另一处的数字。

第三种，漏项。2026 年 9 月 4 日，飞书妙记处理一场 1 小时 27 分的会议录音。产品给出 2,536 字的摘要，分层小标题加时间线条目，结构完整。抽出的待办是 0 条。一场一个半小时的会议，不太可能没有任何需要继续处理的事。摘要给得出，待办不一定抽得出。核对的方法是待办自己列，再用它的摘要对照，不要反过来。

第四种，似是而非。2026 年 9 月 4 日，即梦图片 5.0 Pro 按一句话生成板书风格示意图。结构对，中文标注可读，整体可以进课件。但左边输入的句子是“今天 天气 真好 适合 出去 游玩”，最右边写着“预测结果：明天”，语义不通。结构对、标注对，细节经不起读。核对的方法是每个字都读一遍，示意图只当示意图用，别把它当成模型的真实输出来讲。

第五种，自补条件。2026 年 9 月 5 日，Kimi K3 处理图书馆三角色案例。发出提示词时，输入框把三处数字吞掉了，“每段不超过 120 字”变成了没有数字的句子。模型在思考过程里写“用户说每段不超过字，确实漏了数字，我按 150 字来控制”。您没有说，它替您定了，而且没有反问。核对的方法是看思考过程里替您补了什么条件。

类型	本课案例	怎么核
猜位置	第 5 题标 p.9，实为 p.8；第 3 题我们判错	回原文按页翻，题库抽查
来源冲突	大纲 41.0，幻灯片 41.8，论文自身不一致	看它引哪一处，自己定用哪个
漏项	摘要 2,536 字，待办 0 条	待办自己列，用摘要对照
似是而非	示意图写“预测结果：明天”	逐字读，示意只当示意
自补条件	字数限制丢失，它按 150 字处理	看思考过程补了什么

五种类型的原因各不相同。语言模型在训练中学习了大量语言和任务的规律，生成时根据当前信息组织后续内容。能够写出合乎语境的回答，与每句话都经过事实确认，并不是同一种能力。模型并非时而聪明、时而

糊涂，它每一次都在计算最可能的答案。我不给总体错误率，一次实测得不出那样的数字。阅读与整理可以直接使用，位置、来源、漏项、细节需要核对。

03

PART 03

怎样让它完成一件事

3.1 交付形式：回答、作品、任务

动手前看

概念到这里可以合在一起用了。把一件事交出去之前，先决定您要对方交付哪一种。同一句请求可以要一个回答、一件作品，或一段完成的任务。交付形式定了，怎么核才定得下来。三类交付，下面各有一个真实例子。

第一种，回答。2026年9月4日，Kimi K3，文字对话。提示词是，你是我的英语口语陪练，用大学面试场景问我三个问题，每答完一个纠正我的发音和语法。系统出了第一题，请答题的人自我介绍。我们用一段故意带四处错误的学生答案回复，原文是 “I lead a project about campus AI course last year, and I learn a lot from it. I want to study in your university because it have good teachers.” 系统交回一张纠正表，四处全部标出。lead 改为 led，learn 改为 learned，study in 改为 study at，it have 改为 it has。另外给了一段更自然的改写、三条发音提示，并追问第二题。

我们逐条读过，四处纠正全部正确，追问也符合“每答完一个纠正”的要求。回答类交付，您当场就能核，看一遍就知道对不对。有一点要说清楚。这是文字对话的记录，只证明陪练与纠错这一轮成立，没证明它听准了发音。实时语音对话、可打断，我们没有实测，课上也不演示。

第二种，作品。2026年9月4日，即梦 AI，图片 5.0 Pro，16:9。提示词是，一张《人工智能通识》课的板书风格示意图：白底，黑色线条，左侧一句话进入模型，中间一层“注意力”把词与词连线，右侧输出下一个词，含中文标注，无装饰，教学插图。一次出了两版，取其一，原图 1920×1080。白底黑线，中文标注可读，结构从左到右是输入、词嵌入层、注意力权重计算、加权求和与全连接、概率分布、预测结果，六步。当课堂示意图用，可以直接进课件。但最右边那处“预测结果：明天”语义不通。作品类交付，核的是每一个细节，这种地方要一个字一个字读。

第三种，任务。2026年9月4日，飞书邮箱，命令行工具 lark-cli，用户身份。为了避开真实往来里的姓名与议题，先从一个邮箱向我的另一个邮箱发了三封虚构的教学场景邮件。系统读全文，按紧急度分成三档，为每封起草回复。紧急今天回的是教务处的课程大纲提交截止提醒，本周的是教学发展中心的研讨会报名，可稍后的是出版社的教材再版意向。三封回复全部保存在草稿箱。草稿箱 3，已发送 0。用的命令默认存草稿，发送需要显式加一个确认参数，这次没有加。任务类交付，核的是每一步的状态，发送键留给人。

三种交付放在一起，还能看出一个落差。多数人对 AI 的用法停在第一种，问一句，答一句。它今天的能力已经到了第三种，把一段任务接过去做完。这个落差，这门课叫它认知差。觉得它不好用的时候，先不必怀疑它，多半是请了一位能独立带项目的人，每天只让他帮忙打字。

三种交付有一个共同的起点，先要一个小样，再要全部。小样在不同任务里长得不一样。读材料时是一段大纲，做图时是一张草图，数据任务里是一张图和几句解释，邮件任务里是一封草稿。小样让您能在事情还不大时调整方向，不必等一整套课件做完才发现对象错了。3.3 节那份教案的第一稿就是小样，我们要求它先给流程，不急着写全部讲稿。

三种交付都可以给完成条件写一句白话。例如，我需要一份能放进下周备课会讨论的教案，不只是一段灵感；或者，我现在只想比较两种开场方式，先不要生成整套课件。目标越明确，越容易知道该在哪里停下来。

3.2 选一件先试的事

读一遍就懂

不是每件工作都适合先交出去。先试的一件应当满足三个条件。材料在手，结果可核，出错代价小。

三个条件各有原因。材料在手，是因为系统这次能用的信息只有您放到它面前的那些，材料不在手，它给的只能是通用知识，您也没法判断它用对了没有。结果可核，是因为初稿总要核，您得是这件事的内行，或者有原文、有数据可以复算，否则核不完的结果不如不要。出错代价小，是因为第一次试，为的是练手感，练怎么提要求、怎么看结果；要是一错就发错通知、算错成绩，您就没有余地去看它错在哪一类了。

举一个例子。把一页课程通知改清楚。材料在手，就是那页通知；结果能核，您知道哪些日期不能漏；出错代价小，改完自己再读一遍就行。这类事适合起步。再举一个反例。解释学生为什么学不会某个概念。材料不在手，没有作答和课堂观察；结果不好核，它说的原因您没法验证；出错代价不小，可能误导下一步教学。这类事先整理已知材料，列出需要补问的地方，不急着要结论。

下表是十类常见任务的建议，课件里也有这张表。请找最像您那件事的一行，看建议那一列。

您一周里的事	材料在手	结果能核	出错代价	建议	对应到哪一节
把一章教材变成练习题	有	能（回原文）	小	先试	4.1
给一节新课出初稿	有	能（您是专家）	小	先试	3.3
整理一批文献的地图	有	能（点开引用）	中	先试	4.3
一张成绩表看分布	有	能（复算）	小	先试	4.3
课堂开场的一段短片	有	能（看一遍）	小	可试	4.2
回复一批邮件	有	能（读草稿）	中	起草可试，发送自己点	4.4
会议录音变纪要	有	部分（待办自己列）	中	可试	4.4
给学生批改并反馈	有	难（逐份核）	大	后试	5.2
财务与报销表	有	难	大	别交	
替您操作电脑改文件	有	能（可撤回）	看操作	先在可撤回的地方试	4.4

请现在写一句，我这周想让 AI 帮我完成什么。再补一句，拿到什么就算这一步有用。先不用写得漂亮，写自己的真任务。第八部分的一个月计划从这句话开始。

选定之后，交不交给它，再问三个问题。我自己做要多久？它做对的概率有多高？从交代给它到核完结果，总共要多久？拿两个例子对照。一份带引用的研究综述，自己做要几天，它约 20 分钟，再花十分钟点开两条引用，交给它。一封需要知道来龙去脉的回信，自己三分钟写完，交给它要先解释十几分钟的背景，自己来。自己慢、它准、我能核，三条都满足就交；缺一条，先自己做。这就是调度，知道什么该交给它，什么该自己来。

3.3 完整案例：一份教案的三步

动手前看

这一节把那份备课任务从头到尾走一遍，一共三步。初稿由它生成，判断由您做出，第二稿由它修改。

第一稿

2026 年 9 月 4 日，Kimi K3，网页版新会话。提示词就是 2.3 节那条完整的任务说明，对象是大一非计算机专业学生，主题是“AI 是怎么读懂一句话的”，恰好 3 个学习目标、1 个课堂互动、1 个课后作业，直接给，不要先问。对方一次给出约 2,400 字的完整教案，结构、时长、活动、作业一个不缺。

三个学习目标写得很规范。能解释 AI 理解句子要经过“分词、找关系、联上下文”三个层次，并用生活类比说明至少一个；能辨析人类阅读与 AI“阅读”的核心差异，AI 没有生活常识，依赖统计规律；能评估一个日常 AI 语言应用的局限，并用课堂概念说明它读懂或没读懂的原因。

教学流程六个环节，合计 45 分钟。导入 5 分钟，板书“苹果不好玩”，同一句话人靠常识就懂。第一层“切蛋糕”讲分词 10 分钟，用“南京市长江大桥”的两种切法。第二层“贴标签”讲词向量 10 分钟，意思相近的词在“数学空间”里住得近。第三层“看前后”讲注意力 12 分钟，“味道”和“好”最相关，给重

要的词打高分。课堂互动 5 分钟，活动叫「谁是关键词」，三位学生上台代表“书、作业、笔”三个词，台下举手投票决定连线粗细。小结 3 分钟，AI 不是“理解”了句子，而是“算出了”最可能的含义。

课后作业的原标题是「AI 翻车现场侦探」。找一个 AI 处理语言的例子，记录输入与输出，用三个层次分析它在哪一层出了问题，300 到 500 字，下次课随机抽两人分享。

我们通读了全文。结构符合五项约束，对象、45 分钟、3 个目标、1 个互动、1 个作业；各环节时长相加为 45；内容没有明显事实错误。这是一份可以拿来讨论的初稿。

三处判断

能不能用，要看教师看出哪一处不合适。这三处判断来自教学经验，不来自技术。

第一处，时间给了名词，没给学生。三层讲解占了 32 分钟，学生自己动手只有互动那 5 分钟。今天希望学生掌握的是，为什么一个词要结合具体信息来理解。那就要让他们自己完成一次解释，时间得转过去。目标写的是解释，时间给的是讲解，栏目齐全，栏目之间没有连上。

第二处，活动很好，概念要补一句。投票连线能帮助讨论语义，但学生投出来的线不是模型内部真正的计算。2.1 节那张热力图里的权重是算出来的，不是投出来的。活动保留，加一句“这是帮助理解的示意”。

第三处，通用、正确、没有您的味道。初稿开场也用了苹果的例子，与这门课的导言正好撞车。任何教师这样问，都会拿到这一份。下一步要给它您的材料，上次课讲过什么，这批学生卡在哪里。

三处判断合成一条修改要求。原话是，保留苹果的例子，缩短切块和数字表示的讲解。让学生两人一组，自己选一个可能有不同意思的词，各造一句话，圈出帮助判断含义的信息。最后换一组新句子，每个人独立说明理由。按这个目标重排活动，保持 45 分钟，并说明哪部分时间转给了学生。只给新的教学流程表和一段时间变化说明。

这条修改要求做了三件事。说明目标，学生自己完成一次解释；说明保留什么，苹果的例子；说明希望看到什么变化，时间转给学生，并且要求对方自己说明转了多少。

第二稿

2026 年 9 月 5 日，Kimi K3，在第一稿的同一会话里追加上面那条修改要求。对方交回一张新的流程表和一段时间变化说明。

Kimi K3 · 2026-09-04 · 输出全文 约 2,400 字 · 此处节选															
提示词：为大非计算机专业学生设计 45 分钟课“AI 是怎么读懂一句话的”，3 个目标、1 个互动、1 个作业															
教案：AI 是怎么读懂一句话的 授课对象 大非计算机专业学生 · 课时 45 分钟 · 副题 从“认字”到“懂意思”的三层秘密 一、学习目标（恰好 3 个） 1. 能解释 AI 理解句子要经过“分词→找关系→联上下文”三个层次，并用生活类比说明至少一个 2. 能辨析 人类阅读与 AI“阅读”的核心差异（AI 没有生活常识，依赖统计规律） 3. 能评估 一个日常 AI 语言应用的局限，并用课堂概念说明它“读懂”或“没读懂”的原因 二、教学流程（45 分钟） <table><tr><td>导入 · 一句话的“旅行”：板书“苹果不好玩”，问一句话人靠常识秒懂</td><td>5 分</td></tr><tr><td>第一层 · 切蛋糕（分词）：“南京市长江大桥”两种切法</td><td>10 分</td></tr><tr><td>第二层 · 贴标签（词向量）：意思相近的词在“数学空间”住得近</td><td>10 分</td></tr><tr><td>第三层 · 看前后（注意力）：“味道”和“好”最相关，给重要的词打高分</td><td>12 分</td></tr><tr><td>课堂互动 · 「谁是关键词：3 位学生代表书/作业/笔」，台下举手投票决定连线粗细</td><td>5 分</td></tr><tr><td>小结 · 人类 vs AI：AI 不是“理解”了句子，而是“算出了”最可能的含义</td><td>3 分</td></tr><tr><td>合计</td><td>45 分</td></tr></table> 三、课后作业 · 「AI 翻车现场侦探」 找一个 AI 处理语言的例子，记录输入与输出，用三个层次分析它在哪一层“翻车”。300-500 字，下次课随机抽 2 人分享。 四、板书核心框架（原文） 读懂一句话 = 准确的切法 + 准确的词义 + 词与词的关系网		导入 · 一句话的“旅行”：板书“苹果不好玩”，问一句话人靠常识秒懂	5 分	第一层 · 切蛋糕（分词）：“南京市长江大桥”两种切法	10 分	第二层 · 贴标签（词向量）：意思相近的词在“数学空间”住得近	10 分	第三层 · 看前后（注意力）：“味道”和“好”最相关，给重要的词打高分	12 分	课堂互动 · 「谁是关键词：3 位学生代表书/作业/笔」，台下举手投票决定连线粗细	5 分	小结 · 人类 vs AI：AI 不是“理解”了句子，而是“算出了”最可能的含义	3 分	合计	45 分
导入 · 一句话的“旅行”：板书“苹果不好玩”，问一句话人靠常识秒懂	5 分														
第一层 · 切蛋糕（分词）：“南京市长江大桥”两种切法	10 分														
第二层 · 贴标签（词向量）：意思相近的词在“数学空间”住得近	10 分														
第三层 · 看前后（注意力）：“味道”和“好”最相关，给重要的词打高分	12 分														
课堂互动 · 「谁是关键词：3 位学生代表书/作业/笔」，台下举手投票决定连线粗细	5 分														
小结 · 人类 vs AI：AI 不是“理解”了句子，而是“算出了”最可能的含义	3 分														
合计	45 分														
1	时间给了名词，没给学生 三层讲解占 32 分钟，学生自己动手只有互动那 5 分钟。今天希望学生掌握的是“为什么一个词要结合具体信息来理解”，那就要让他们自己完成一次解释，时间得转过去。														
2	活动很好，概念要补一句 投票连线能帮助讨论语义，但学生投出来的线不是模型内部真正的计算。保留活动，加一句“这是帮助理解的示意”。														
3	通用、正确，没有老师自己的味道 它开场也用了苹果。任何老师问它都会拿到这一份。所以下一步要给它老师自己的材料：上次课讲过什么，这批学生卡在哪儿。														
3	我的修改要求（原文） 保留苹果的例子，缩短切块和数字表示的讲解。让学生两人一组，自己选一个可能有不同意思的词，各选一句话，圈出帮助判断含义的信息。最后换一组新句子，每个人独立说明理由。按这个目标重排活动，保持 45 分钟，并说明哪部分时间转给了学生。														

图 4 · 2026 年 9 月 4 日 Kimi K3 生成的教案（节选复排）与教师的三处判断。原始输出约 2,400 字，存 05-PPT/assets/shots/。

第二稿流程表的原文如下。

时间	环节	谁	内容要点
0:00-0:03	导入	教师	保留「苹果不好玩」例子，抛出核心问题
0:03-0:10	三层速览	教师	分词 2 分、词向量 2 分、注意力 3 分，只讲概念
0:10-0:23	活动一：两人造句	学生	自选多义词各造一句，圈出帮助判断含义的上下文；抽 1-2 组投影分享
0:23-0:34	活动二：独立判断	学生	教师下发 3 个新句子，独立写词义与依据；同桌互评，抽 2-3 人说明
0:34-0:40	全班验证	师生	「人工注意力网络」快速版，用手势指出关系最紧密的上下文词
0:40-0:43	小结	教师	人类靠生活常识，AI 靠切词、找邻居、算关系
0:43-0:45	作业布置	教师	「AI 翻车现场侦探」

各环节时长为 3、7、13、11、6、3、2，相加是 45。它给的时间转出说明是，导入 5 到 3，分词 10 到 2，词向量 10 到 2，注意力 12 到 3，小结 3 到 2，合计转出 28 分钟；新增活动一 13 分钟、活动二 11 分钟、过渡与巡视 4 分钟，合计新增 28 分钟。它自己的总结是，理论讲授从 37 分钟压缩到 12 分钟，学生动手与表达从 5 分钟扩大到 24 分钟。

我们核了三处。新流程各段相加 3 加 7 加 13 加 11 加 6 加 3 加 2 等于 45，成立。转出 2 加 8 加 8 加 9 加 1 等于 28，新增 13 加 11 加 4 等于 28，两边相等。“理论讲授 12 分钟”是它自己算的，导入 3 加速览 7 加小结 2，我们引用时照它这么算。

思考过程里有一段要单独记录。模型把时间相加了三次，第一次得 41，第二次得 43，前两次不足 45 便重新分配，第三次得 45 才输出流程表。这是“根据中间结果再决定下一步”的直接样本，2.3 节讲持续执行时已经引用过。

第二稿的价值不在更长、更漂亮。先找到一个影响学习的问题，说明希望学生多做哪一步，再看修改是否服务这个目标。三步里只有中间那一步不能交出去。

有两处要说明。第一，课堂上这份初稿是现场生成的，这里写的是 2026 年 9 月 4 日那次的记录，每次开课前会重跑，新的初稿可能与这里的不同，三处判断的方法不变。第二，第二稿还没有经过试教。它在时间分配上服务了我们说明的目标，但哪个活动能让学生真的说出理由，只有上过课才知道。这一步留给您的课堂。

要点 备课三步走。初稿由它生成，先要流程不要全文；判断由您做出，找影响学习的问题而不是不够漂亮的句子，写明目标、保留什么、希望看到什么变化；第二稿由它修改，拿回来先核数字再看变化是否服务目标。

3.4 两份模板

读一遍就懂

课后请用自己的一件事，照下面两份模板走一遍。模板原文来自这门课的练习册，每次换掉方括号里的内容。

任务说明模板。

我想完成【具体任务】，给【使用者】使用。目的是让对方能够【解释、判断或完成什么】。

这次依据的材料是【粘贴或附上材料】。已有条件是【背景、时间、限制】。

请先给出【一个小样或第一部分】。缺少影响方案的关键条件时先问我。条件充分时先做小样。

修改要求模板。

这一处【原句或具体位置】不适合，因为【对象、事实、目标或条件】。请按【明确方向】修改，并说明主要改动。

用 3.3 节的案例填一遍，可以看出模板与实际操作的对。具体任务是一节 45 分钟的课，使用者是大一非计算机专业学生，目的是让他们能用两个例子解释“同一个词要结合相关信息理解”。材料是“苹果”两句例句和上次课讨论过搜索引擎这一背景。先给出的小样是教学流程表。修改要求里，原句或具体位置是三层讲解占 32 分钟，原因是目标要求学生自己完成一次解释，明确方向是两人一组造句并圈出判断依据。

随模板一起保存三样东西。第一样，一份您认可的样例，比如第二稿那张流程表。有了样例，下次可以直接说“按这份的格式”。第二样，适用条件，这套做法适合什么任务，要准备什么材料。第三样，核对重点，这类任务通常错在哪里，时长相加、页码、引用，或者思考过程里自补的条件。

完成的标志是，拿到一个小样，并能指出自己的哪条反馈改变了结果。没有网络时，请一位同事扮演协作者，照任务说明完成一个口头小样，同样有效。每次再用之前，先确认这回的受众、材料版本和任务目标。重要事实回原始来源核一遍；产品今天能不能做某件事，以您手头打开的那个入口为准。

3.5 给它铺轨道：同一个模型，换个流程

读一遍就懂

很多人用它的方式是打开对话框，一段话扔进去，等结果。有用，但不一定好。吴恩达 2024 年 3 月在 The Batch 上发过一组数字，测的是一套公开的编程题 HumanEval。GPT-3.5 一句话让它直接写完，通过率 48.1%；换成更强的 GPT-4，一句话写完，67.0%；还是 GPT-3.5，不换模型，换流程，写完自己运行，看报错，改，再运行，循环到通过，95.1%。变的是流程，不是模型。这是一次测评上的数字，不是普遍规律，但方向清楚。

铺轨道，前期最费力，规划路线、建桥、打隧道；轨道修好了，车就跑得又快又稳。给它设计工作流是同一个道理，前期要花时间搭，搭好之后每次跑都稳。

第一步，拆。拿出您每周都在重复的一件事，拆成小步骤。备课这件事拆开是五步。

步骤	谁来做	本课对应
准备材料清单	自己来	2.3 节的材料清单
出初稿	全交给它	3.3 节第一稿，任务说明见 3.4 节
三处判断	自己来	3.3 节
第二稿	和它一起	3.3 节，修改要求模板见 3.4 节
存成模板	和它一起	2.3 节可复用做法，工具箱 9.5

第二步，每一步分三类，全交给它、和它一起、自己来。分类的依据就是 3.2 节的三问，自己慢、它准、我能核的步骤全交。

第三步，先给"全交"的步骤铺轨道，为它们写专门的任务说明和核对清单。这些步骤投入最少，回报最大。那份教案的任务说明就是第一条轨道，存成模板就是第一份 Skill。第八部分的一个月计划里，第三周做的就是这件事。

这一节讲的是怎么分工，不是谁替谁。判断那一步一直在人手里。

3.6 提示词的骨架：流传的框架都在说同五件事

读一遍就懂

很多教师问过，网上那么多提示词框架，APE、BROKE、COAST、RACE，一串字母，该学哪个。下表把流传较广的十二种放在一起，第二列是各自的要素，第三列是对应到这门课的哪一项。

框架	它的要素	对应到这门课的五项条件
APE	行动、目的、期望	交付、目标、交付
TAG	任务、行动、目标	交付、交付、目标
ERA	期望、角色、行动	交付、角色、交付
RACE	角色、行动、上下文、期望	角色、交付、材料、交付
RISE	角色、输入、步骤、期望	角色、材料、这一步做到哪、交付
CARE	上下文、行动、结果、示例	材料、交付、交付、样例
COAST	上下文、目标、行动、场景、任务	材料、目标、交付、对象、交付
TRACE	任务、请求、行动、上下文、示例	交付、交付、交付、材料、样例
ICIO	指令、背景、输入数据、输出引导	交付、材料、材料、交付
ROSES	角色、目标、场景、解决方案、步骤	角色、目标、对象、交付、这一步做到哪
BROKE	背景、角色、目标、关键结果、试验改进	材料、角色、目标、交付、修改要求
CREATE	角色、要求、示例、调整、输出类型、额外	角色、交付、样例、修改要求、交付、条件

第三列里反复出现的，就是 2.3 节的五项条件，对象、目标、材料、交付、这一步做到哪。多出来的只有两样。一个角色，对应工具箱里的岗位说明书，您让它扮演谁；一份样例，对应您认可的那一份好结果，给它看过，它就知道您要什么。BROKE 和 CREATE 里的试验改进与调整，就是 3.3 节讲的修改要求。

所以不必记字母。记住五件事，会写角色和样例，会提修改要求，任何框架都在您手里。写不全五件事的时候，最省事的办法是让它先问您，缺什么它会问。这十二种框架是公开流传的写法，出处不一，这里只对照要素。

04

PART 04

今天它能做什么

动手前看

导言里留下四个字，行易知难。做，机器越来越会；知道它能做什么，还得靠人。这一部分只做一件事，让您知道它今天能替您做什么。

这一部分记录十类应用，全部是 2026 年 9 月 4 日至 6 日在具体工具上实际运行的结果，出错的部分一起保留。十类分成四组，阅读与表达、视觉与陪练、研究与数据、事务与操作。分组会交叉，一份研究材料可以变成课程内容，会议纪要也可能进入项目报告。分组是为了让您看见，一份输入怎样一步步变成成果。您不需要十件都做，也不需要换十个工具，先找一两种与自己工作接得上的。

每一类固定写四段。适合什么，给它什么，它给了什么，怎么核。“它给了什么”写的是那一次的真实结果与数字；“怎么核”写的是我们核出了什么。这些数字只证明那一次在那个工具上发生了什么，您自己动手，结果会有出入。

4.1 用它的四层

读一遍就懂

同一个工具，多数人停在第一层。往上每走一层，交出去的东西从一句话变成一个目标、一条轨道、一份档案，留在自己手里的是判断。

层	您给它什么	它交回什么	这门课的真实例子	留在这里的
一，问答	一句话	一段回答	2.1 节的“苹果”；4.3 节口语纠错四处全对	判断对不对
二，交任务	五项条件加材料	一份能用的初稿	3.3 节教案第二稿，学生时间 5 到 24 分钟；4.4 节 45 个分数的分布	三处判断；比较前的三个问题
三，铺轨道	一条固定的任务说明加核对清单，可重复	每次都稳的产物	4.2 节 Kimi PPT 按规范文件多步执行；同一提示词跑三次	哪些步骤全交，哪些一起做
四，建档案	岗位说明书加个人档案，放在它面前	每个新会话都从您的情况开始	2.2 节空提示词时它引用了过去的会话；工具箱 9.3 的档案模板	不能替您定的事，不能出门的材料

第一层今天就会，第二层是这本手册的主体，第三层和第四层是第八部分一个月计划要做的事。越往上，交出去的越多，握在手里的越关键。下面各节的每一件事，都能对到某一层。

4.2 阅读与表达

这一组处理的是最贴近备课的工作，把一份长材料整理成结构，再做成课件。两件事用的是同一份材料，那篇 15 页的论文，先交给对话产品读，再把大纲交给做课件的产品。两次都得到了可用的初稿，两次也都核出了问题，而且问题不在同一处。

读长文档

适合什么。一章教材、一篇论文、一份几十页的文件，需要先整理出结构、再从中出题或提炼要点。

给它什么。2026 年 9 月 4 日，Kimi K3。输入是那篇论文的 arXiv 链接，15 页。提示词原文是，请读这篇 15 页的 PDF，然后列出三级大纲；出 5 道选择题，每题标出对应页码；指出初学者最容易混淆的两个概念。用中文回答。

它给了什么。三级大纲七章逐级展开，与论文实际结构一致，3.2.1 缩放点积、3.2.2 多头、3.5 位置编码这些小节都对。5 道选择题题干与选项均可用，答案都是 B。两个易混概念是三种注意力里三个输入的来源，以及缩放因子与另一种温度参数的区别。思考过程里写着“由于 PDF 是连续的文本，我需要估算页码”。

怎么核。用 PyMuPDF，一个读取 PDF 的程序库，在原 PDF 里逐题检索关键词。第 1 题缩放点积，标的是第 4 页，实际第 4 页，对。第 2 题多头，标第 5 页，实际第 5 页，对。第 3 题解码器遮挡的目的，标第 3 页，我们起初判在第 5 页，回原文发现第 3 页 3.1 节已经说明了目的，它是对的，撤销原判。第 4 题位置编码，标第 6 页，实际第 6 页，对。第 5 题英德翻译分数与训练成本，标第 9 页，实际第 8 页，错。结论是五题里一题页码错，一题是我们判错了，已经更正。总体错误率不从这五题推，五题的完整原始输出也没有存下来。大纲和题目可以用，页码要抽查，查页码也得回原文。

一次运行说明不了它稳不稳。2026 年 9 月 6 日，用逐字相同的提示词又跑了两次，仍是 Kimi K3。第二次它改用 Python 下载 PDF 并逐页提取文本，先核实了“15 页”，五题四对，一题差一页，因为那一节的标题在第 8 页末尾，表格和结论在第 9 页。第三次它用 pdftotext 提取，并按分页符逐页核对了章节位置，五题全对。三次里它自己选了三种读法，题目各不相同，大纲与易混概念稳定，“Q、K、V 的来源”三次都在易混概念里。结论不变，大纲可用，题目可用，页码要逐题回原文核。三次的逐题核对表见实测记录 15，随课程材料提供。

做 PPT

适合什么。已经有一份大纲或一份讲义，需要快速形成一版可编辑的课件初稿，再由教师调整。

给它什么。2026 年 9 月 4 日，Kimi PPT。输入是十条大纲，什么是注意力、为什么缩放、多头、位置编码、编码器与解码器、为什么能并行、训练代价、翻译结果、它为什么会错、对教学的启发。要求学术风格，白底。

它给了什么。约 6 分钟后交付 10 页，学术白底，每页一句标题加三条要点，右侧公式或图表栏，页脚自动标注 arXiv 编号。过程中系统先读了产品内置的排版规范文件，选了教学类版式，中途联网核对了论文表二的分数。导出的文件是否可编辑，下载打开看一眼。真实输出的页标题包括，什么是注意力，让每个词有选择地“看向”别的词；为什么要缩放，不除以根号 dk，softmax 会被推到梯度近零；结果，英德 28.4、英法 41.8 BLEU。

怎么核。同一次任务里出现了两个数。对话里的大纲写英法 41.0，第 8 页幻灯片写 41.8。回论文原文查，第 1 页摘要写 41.8，第 8 页表二写 41.8，第 8 页正文写 41.0。论文自身不一致，两个数都有出处。上台前自己定用哪个数，另一个加注说明。另外请看第二页那个标题，大一学生听得懂吗。给研究者的大纲直接切成十页，仍然是研究者的语言，对象要重新告诉它。更好的做法是让同一份材料逐步变成不同成果。先让系统整理结构，再选一个要讲的内容让它转成学生能进入的语言，最后再做课件。每做完一步，您看一眼，再交下一步。

4.3 视觉与陪练

这一组是图像、视频和语言练习。三件事的共同点是产物当场就能看，核对不需要回原文，需要的是逐字读、看一遍、逐条对。这也是教师最容易先试的一组，一句话就能开始，出错的代价小。

做图

适合什么。已经知道想讲什么，只缺一张帮助讨论的图。课件缺一张有气氛的插图，一句话就够。要求数据准确的图表，不要用它画。

给它什么。2026 年 9 月 4 日，即梦 AI，图片 5.0 Pro，16:9。提示词原文是，一张《人工智能通识》课的板书风格示意图：白底，黑色线条，左侧一句话进入模型，中间一层“注意力”把词与词连线，右侧输出下一个词，含中文标注，无装饰，教学插图。

它给了什么。一次出两版，取其一，原图 1920×1080。白底黑线，中文标注全部可读，结构从左到右是输入、词嵌入层、注意力权重计算、加权求和与全连接、概率分布、预测结果，六步。可以直接进课件。

怎么核。逐字读标注。输入的句子是“今天 天气 真好 适合 出去 游玩”，最右边写着“预测结果：明天”，语义不通。当示意图用可以，别把它当模型的真实输出讲。气氛插图它做得稳，要求每个字都对示意图，就得逐字核。提示词的写法上，先说用途，这张图是用来导入、比较，还是解释关系；再说主体以及主体之间必须出现的联系；最后才是白底、颜色和风格。上面那条提示词的顺序就是这样。您做第一张图时，先补一句，我要让学生从这张图里看出什么。

做视频

适合什么。课堂开场的一段短片，把抽象描述变成可观察的画面。十秒、固定机位、单一主体的视频，当前可以稳定生成。

给它什么。2026 年 9 月 4 日，即梦 AI，Seedance 2.5，720P，16:9。提示词原文是，生成一段 10 秒延时视频：一粒种子在土壤中发芽，侧视剖面，先向下扎根再向上出芽，实拍质感，固定机位，无文字。

它给了什么。一段 1280×720 的视频，约 24 帧每秒，241 帧，10.0 秒，13.7 MB。消耗 200 积分，生成前有确认弹窗。抽取十帧看，种子裂开，根先向下，芽后向上，顺序与真实发芽一致，原片没有旁白。

怎么核。看一遍原片，用软件读分辨率、帧率、时长。这一次的经验是，十秒、固定机位、单一主体，稳；超过十五秒、要人物或要连贯情节，就得拆镜头。这个界线是这个工具这一次的，换工具要重新试。这段影像是生成的，不是某次真实实验的记录，不能拿来证明生物过程是怎样发生的。讲解性的内容，没审过就别放给学生。

陪练

适合什么。语言练习、面试模拟、一问一答的纠错。一次只问一个问题，回答以后指出最需要修改的地方，让学习者自己再说或写一遍。

给它什么。2026 年 9 月 4 日，Kimi K3，文字对话。提示词原文是，你是我的英语口语陪练，用大学面试场景问我三个问题，每答完一个纠正我的发音和语法。

它给了什么。系统出第一题，请答题的人自我介绍。我们用一段带四处错误的学生答案回复。系统交回一张纠正表，逐条标出 lead 改 led，learn 改 learned，study in 改 study at，it have 改 it has；一段更自然的改写；三条发音提示，led 短促，learned 尾音 /d/，has 结尾 /z/；然后出第二题。

怎么核。逐条读。四处纠正全部正确，发音标注正确，追问第二题符合“每答完一个纠正”的要求。要说明的是，这是文字对话的实测，只证明“陪练与纠错”成立；实时语音、可打断这一路，没有实测，课上不演示。改文字和纠发音是两回事，得分开试。

4.4 研究与数据

这一组是一份二十分钟的研究综述和一张 45 个分数的表。两件事的核对方法正好相反。研究综述的结构系统给得出，但每条引用都要点开；数据的统计它算得准，可以用独立复算核实，但数据里没有的背景它不补。

深度研究

适合什么。接触一个新问题时，先弄清有哪些材料、大家在讨论什么、分歧在哪里。系统给的是一张材料地图，进入材料还是要自己走。

给它什么。2026 年 9 月 4 日，Kimi 深度研究。提问原文是，就“AI 在高校通识教育中的应用”做一份研究综述：分歧点、近三年进展、研究空白，每条结论附可点击来源。

它给了什么。约 20 分钟后交回一份分层报告。四大分歧点，批判性思维上的认知卸载派与 U 型关系调和派，学术诚信与评价范式，教师角色的“双师”与工具化批评，公平与数字鸿沟。近三年进展，2023 年到 2026 年，政策规模化，教学模式化，证据综合化，素养框架建构。六处研究空白，纵向证据缺乏，人文类通识核心课实证稀薄，高阶素养测评工具缺位，师资激励研究不足，公平议题缺本土实证，课程体系整合理论薄弱。另附一张概念图和一个可下载的文件。

怎么核。抽查两条引用，逐条点开。第一条，Gerlich 2025，发表于 Societies 期刊，英国 666 名参与者，AI 使用频率与批判性思维显著负相关。真实。但点开原文页面才看到，该文后来发布过一份正式更正，只看摘要看不出来。第二条，教育部等五部门《“人工智能+教育”行动计划》，教科信〔2026〕1 号，2026 年 4 月印发，联合发文单位为教育部、国家发展改革委、工业和信息化部、科技部、国家数据局。真实。抽查两条，两条为真。“需核”的提醒仍然成立，真伪之外还有版本问题，更正、撤稿、新版替换，只有点开原文才看得到。

数据

适合什么。一张成绩表、一份问卷结果，需要先看分布再决定下一步。均分是起点，不是结论。

给它什么。2026 年 9 月 4 日，Kimi K3。输入是三个班共 45 名学生的期末成绩，每班 15 人。这 45 个分数是我们编的匿名样本，在对话里直接输入，不对应真实学生。

它给了什么。系统自动调用 Python，算出均分与样本标准差，用箱线图法判断异常值，画了箱线图加抖动散点。A 班均分 78.13，标准差 12.19；B 班 82.07，5.22；C 班 76.00，14.34。标出的异常值是 A 班的 45 分。

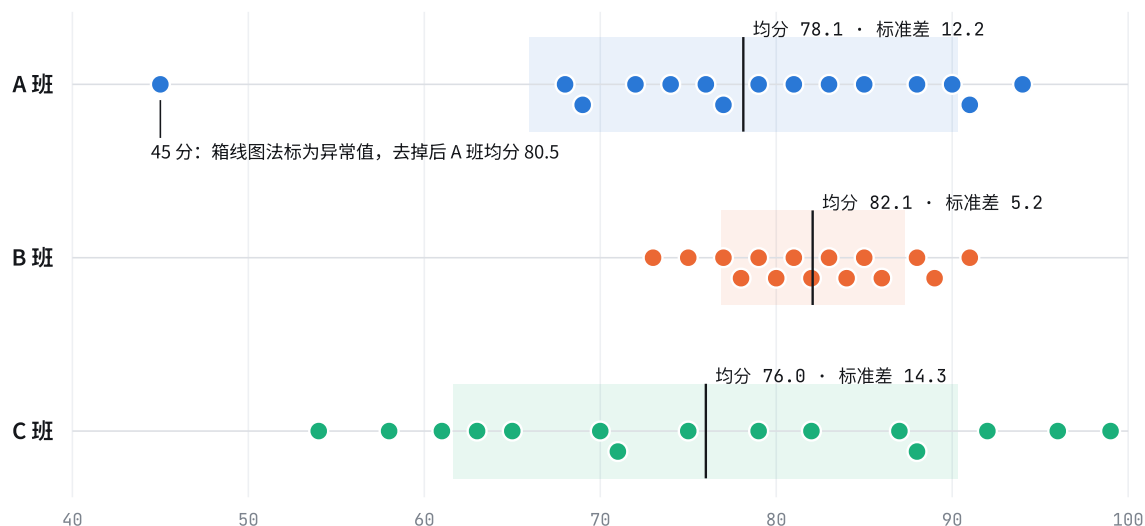


图 5 · 三个班全部 45 个分数（课堂构造数据），竖线为均分，浅色带为均分上下一个标准差。计算由 Kimi K3 调用 Python 完成，本地独立复算一致。

怎么核。本地用 Python 独立复算，三个班的均分与标准差六个数全部一致。异常值用箱线图法和正负两个标准差两种方法复算，都得到同一个 45 分。同一提示词 9 月 6 日又跑了两次，三个班的均分与标准差六个数三次完全一致，异常值三次都只标出 A 班的 45 分，其中一次还主动区分了样本标准差与总体标准差。统计这一步敢写“可用”，靠的是三次一致。把 45 个分数全部画出来，比三根柱子清楚得多。三个班均分最高 82，最低 76，只差六分，分布却大不相同。B 班均分最高，也最集中，多数人在 75 到 90 之间。C 班均分最低，也最分散，既有 54 也有 99。那个 45 分把 A 班拉低了，去掉它均分是 80.5。同样叫“平均 78 分”，下一步教学并不相同。

数字本身还不能解释全部原因。比较之前先问三件事，三个班用的是同一份试卷吗，分制一样吗，缺考是空着还是按零分录入。这三件数据里没有，系统也不该编。标出那个 45 分，不表示它就该删掉；先看录入有没有错，再看实际情况。

4.5 事务与操作

这一组是每天都在消耗注意力的工作，信息来了，还没有变成下一步行动。邮件、会议纪要，再加上让系统替您做一些重复动作。这一组要核的是状态，草稿建了没有，发了没有，待办抽出来了没有，文字反倒在其次。

邮件

适合什么。一批邮件读完分档，逐封起草回复。读它稳，起草要核，发送留给人。

给它什么。2026 年 9 月 4 日，飞书邮箱，lark-cli，用户身份。为避开真实往来里的姓名与议题，先从一个邮箱向我的另一个邮箱发三封虚构的教学场景邮件。要求系统读全文，按紧急度分三档，为每封生成回复草稿。

它给了什么。三档分好。紧急今天回，教务处的课程大纲提交截止提醒，草稿是“已收到，大纲今天 17:30 前提交，考核方式一并附上”。本周，教学发展中心的研讨会报名，草稿是“参加，不需要工位。请把会议室与议程发我”。可稍后，教材再版意向征询，草稿是“有修订意向，想补一章 AI 工具实操，下周给修订大纲”。三封均保存为草稿。

怎么核。列草稿箱。草稿箱 3，已发送 0。回复命令默认存草稿，发送要另加一个确认参数，这次没加。核草稿时看两点。邮件里没说已经做完的事，草稿不能替您宣布做完；要承诺时间的，得跟您的实际安排对得上。

纪要

适合什么。会议录音转成摘要，再由人补出行动清单。转写稳，待办需核。

给它什么。2026 年 9 月 4 日，飞书妙记。样本是我 2026 年 8 月 24 日的一场会议，时长 1 小时 27 分。原会议内容涉及公务，已经脱敏，不写进这本手册，只用三个数字。

它给了什么。摘要 2,536 字，分层小标题加时间线条目，结构完整。待办 0 条。

怎么核。读摘要与待办字段。摘要给得出，待办不一定给得出。一个半小时的会议，不太可能没有任何需要继续处理的事。这是摘要与行动清单的区别。摘要告诉您谈了什么；行动清单还要问，决定了什么，谁负责，什么时候完成，什么仍未确定。如果录音里没有明确负责人，就写未定，或者列出待确认的问题，不要让系统补一个看似合理的人名。待办自己列，再用它的摘要对照，不要反过来。

替您操作、做小工具与本地模型

适合什么。可重复、可核对、可撤回的动作。批量改文件名、批量整理表格、批量起草，都满足这三条。发送、付款、删除不满足最后一条，留给人。敏感材料不能离开本机时，本地模型是另一种选择。

给它什么。先看操作记录。上面那三封邮件与三封草稿，都出自 2026 年 9 月 4 日 lark-cli 实际执行的命令，没有在网页上手工操作。第一条命令读收件箱，列出最近五封，它列出了三封，日期、发件人、主题。第二到第四条各是一次回复，带着回复正文，结果都是同一行字，“草稿已建”。第五条列草稿箱，草稿数 3。

再看本地模型。2026 年 9 月 4 日，一台 48GB 内存的 MacBook，Apple Silicon，运行 Qwen3.8-27B，MLX 6-bit 版本，本机网页界面，断网。提示词是，用三句话写一段课堂小结，主题：AI 是怎么读懂一句话的。对象：大一非计算机专业学生。语气：下课前跟学生聊天。

它给了什么。命令记录里，“草稿已建”三次，“已发送”零次，每一行命令都能重跑，每一个结果都能查。本地模型交回三句话。第一句说 AI 更像在玩一个超级复杂的文字接龙游戏，通过统计海量数据里词语出现的概率来预测下一个字。最后一句说，把它当成一个特别会查资料、特别懂语法的“高级复读机”就好。运行指标是生成 121 个 token，每秒 9.4 个，峰值内存 23.12 GB。

怎么核。命令记录看状态，草稿可以删，发出去的信不行。本地模型逐句读，三句切题、无事实错误、语气符合要求。这段话用了十几秒生成，占了整台机器一半的内存。这次运行证明的是，在当时那套环境里，这一项文字任务离线做得成；文件、图像、语音任务有没有在这台机器上验证过，不能由此推断。图片理解这次没有跑。

再看做小工具。2026 年 9 月 6 日，Kimi K3。提示词是，做一个给大一非计算机专业学生用的课堂练习网页，主题是“AI 是怎么读懂一句话的”。要求：5 道单选题，每题四个选项，答完显示得分，并逐题给出解析；单文件 HTML，白底，中文，不用外部库，直接给完整代码。约两分钟，它读完需求，直接写代码，存成一个 11.6 KB 的文件，给出预览。五道题、四个选项、提交后百分制得分、逐题解析、一个“重做一遍”按钮。第三题用的正是 2.1 节那个“苹果”例子。

怎么核。本地用浏览器打开，五题可选，提交后得 80 分，解析逐题显示，代码里没有任何外部链接。题目逐条读过，五题与这门课的讲法一致。它自己补了一个条件，页面上写着“建议用时 8 分钟”，提示词里没有这一句，用之前删掉或改成自己的数。过去这件事需要一位会写网页的人。现在中间隔着的，是那一句说清对象、主题、题数和形式的话。课堂上播放两分钟录屏，不现场生成。记录见实测记录 16，随课程材料提供。

还有一件事要写明。“替您操作电脑”里桌面操作那一类，打开软件、点按钮，我们没有实测，最新进展放在第六部分。批量改文件名这类动作，做法是先给系统一个明确目录和文件副本，请它先列出原文件名、识别出的姓名和建议的新文件名，重名或读不清的单独列出，不要猜着改；确认规则合适以后在约定范围内执行；最后打开目录看结果。

十类应用一览表

#	场景	给它的	它给的	核什么	结论
01	读长文档	15 页 论文链 接加三 项要求	三级大纲加 5 题带页码	逐题回原文查页码，1 错，1 处判错后更正	需核对
02	做 PPT	十条大 纲	10 页学术白 底	与它自己的大纲比对，41.0 与 41.8	需核对
03	做图	一句话	1920×1080 板书示意图	逐字读标注，“明天”不通	可用，示意需核对
04	做视频	一句话	10 秒 720P	看一遍，顺序对	试用
05	陪练	带四处 错的回 答	逐条纠错加 改写加发音 提示	四处全对	可用
06	深度研究	一个问 题	带引用的综 述加文件	点开两条，真，一条有更正	需核对
07	数据	45 个 分数	均分、标准 差、异常 值、箱线图	本地复算，一致	可用
08	邮件	3 封	三档加 3 封 草稿	草稿箱 3，已发送 0	需核对
09	纪要	1 小时 27 分 录音	2,536 字摘 要	待办 0 条	需核对
10	做小工具与本地模型	一句话 (对 象、主 题、题 数、单 文 件)； 三句话 小结的 要求	11.6 KB 单 文件练习网 页；三句话	本地打开作答提交 80 分，解析全显示；切题、断网	可用，题目需核对

十行的核法各不相同，回原文的、复算的、看草稿箱的都有，但每一行都填了，没有一行是“没核”。十件事的做法可以并成一套，以后每次都照着走。先说清楚要解决什么问题，把有关的材料备好。再请它做一个小样，看它把任务理解成了什么。然后挑一处真正碍事的问题，说明要改成什么样，请它改。它引用的关键事实、数字、文件，您自己核一遍。最后把您认可的结果存下来，那几条管用的要求也一起存下来，下次直接用。

反馈的写法也有共同特点。说“写得不好”很难继续。说“这段没有解释给谁看”“这个例子需要学生还没学过的概念”“请保留材料里的两种不同观点”，就把判断转成了下一轮任务。3.3 节那条修改要求就是这样写的。

最后说一句产品名。表里的 Kimi、即梦、飞书，和这本手册里提到的其他产品，都是 2026 年 9 月这门课实际用到的，或各官网当时写的名字，属于 [查一下再用](#)。名字会换，分法不会。找的时候按您要的交付形式找，不按厂商找。

4.6 去掉 AI 味：先改要求，再改稿子

动手前看

同一个请求，加一段写作要求，机器味明显减少，但减不干净。2026 年 9 月 11 日，Kimi K3，让它为这门课写一段 200 字的课程介绍，跑了两次。

第一次不加要求。它写的是：2017 年，一篇题为《Attention Is All You Need》的论文悄然发表，几年后，它成了 ChatGPT、Claude 等大模型的共同起点。本课程面向零基础的高校教师，不预设任何编程或数学背景，从一个真实的研究故事讲起：这篇论文想解决什么问题？“注意力”究竟是什么意思？它如何改变了 AI 理解语言的方式？我们将结合课堂答疑、论文批改、文献阅读等教学场景中的真实案例，用类比和演示代替公式，帮助您看懂大模型的底层逻辑，理解它为何强大、又为何出错。**学完本课，您将能够自信地谈论 AI，并把它作为教学工具真正用起来。**

第二次加一段写作要求，不用三连排比，不用段末金句，不用冒号引出判断，不用“不是……而是……”，不用赋能、助力、全方位、重塑一类词，先例子后术语，句子短，语气克制，像一个人在说话。它写的是：2017 年，谷歌的八位研究者发表了一篇文章，**标题有点狂**：《Attention Is All You Need》。当时没多少人意识到，这篇文章后来会长成 ChatGPT 的技术底子。这门课就从这篇论文讲起。我们会看机器怎么读一句话，**比如“苹果”为什么有时候是水果，有时候是公司**，再回头看我们自己平时怎么跟 AI 打交道，哪些用法有效，哪些用法白费力气。课上没有公式推导，也不需要编程基础，所有概念都从日常例子入手。**学完你会清楚一件事**：AI 能做什么、做不到什么、以及在你的课堂和研究里，它值不值得用。适合从没碰过技术背景、但对 AI 好奇的老师。

两段对照着读。第一段的末句是典型的段末金句，中间还有三个连着的问句。第二段开头一句实话，中间给了一个具体的例子，句子都短；但末尾仍有一句冒号引出的总结。写作要求能挡掉一半，剩下的一半还得人自己看出来、删掉。

我们通稿时用的六条规则，各配一个改法，第三列是这本手册 2026 年 9 月 5 日通稿时的数量。

痕迹	怎么改	这本手册通稿时
三连排比	留一个，删两个	删去十余处
段末金句	删掉，或换成一个例子	删去十余处
冒号引出判断	把判断写成一句平常话	仍会残留，逐句看
“不是……而是……”	只留肯定的那一半	9 处减到 1 处
万能词：赋能、助力、全方位、重塑	换成具体动作	全部替换
万能主语“它”“我们”	写清谁在做	“它” 315 处减到约 185 处

用法两步。第一步，把这段写作要求存进岗位说明书，工具箱 9.3 已经写好一份“写作说明”，每次都带上。第二步，交回来仍按这张表过一遍，数一数六样各剩几处，剩几处改几处。

05

PART 05

回到大学

5.1 同一份材料怎样形成一节课

读一遍就懂

第三部分那份教案走完了三步。初稿是您写的，判断是您下的，第二稿是您改的。这就是备课三步走，材料换了，三步不变。下面把它写成任何学科都能照着走的样子。

第一步，初稿，交给它。把材料清单和任务说明放到系统面前，要求先给流程，不急着写全部讲稿。初稿通常会目标、流程、活动、作业这些栏目写齐。此时不要急着改文字，先看栏目之间是否真的连起来了。第三部分那份初稿的问题正在这里，目标写的是解释，时间给的是讲解。

第二步，判断，您自己做。找出一处影响学习的问题，不是一处不够漂亮的句子。把三样东西写进修改要求，目标是什么，保留什么，希望看到什么变化。第三部分那三处判断，一处是时间怎么分，一处是概念准不准，一处是材料有没有您的味道，全部来自教学经验，没有一处来自技术。

第三步，第二稿，交给它。在同一会话里发出修改要求，只要流程表和一段变化说明。拿回来先核数字，时长相加是否等于要求的分钟数，转出的时间与新增的时间是否相等。再看改动是不是朝着您第二步写下的目标去的。

换成文学课。第一步，把一篇作品的文本、学生上一次讨论的记录和这节课的目标交给它，让它先列出人物关系与几种不同的读法，再给一个 45 分钟的流程。第二步，您判断哪一种读法是这批学生最容易忽略的，哪一处细读值得花时间，然后写修改要求，例如“保留人物关系表，把第二种读法的讲解压缩，让学生两人一组选一处细节，各自说明它支持哪种读法”。第三步，拿回第二稿，核时长，看学生自己说明理由的时间是否增加了。

换成管理课。第一步，把一个案例的背景材料交给它，让它先整理出案例里的角色、各自掌握的信息和已经做出的决策，再设计一个需要取舍的问题。第二步，您判断这个问题是否真的没有标准答案，学生是否有足

够的材料支撑不同选择，然后写修改要求。第三步，拿回第二稿，看那个取舍问题能不能让不同小组给出不一样、又各有依据的答案。

两个例子里，教师做的都是同一件事，用专业判断定下学生该在哪一步花时间。

5.2 课堂助教

查一下再用

AI 可以给学生更多反馈，这对人数较多的课程很有吸引力。这种帮助怎样安排，会影响学生最后留下什么。一项已发表的随机对照研究，也就是把学生随机分组、比较两种做法的研究，给出了方向。

Kestin 等，2025 年，发表在 Scientific Reports。研究在哈佛大学一门本科物理课上进行，2023 年秋季学期，233 人注册，194 人纳入分析。采用交叉设计，同一批学生轮流经历两种条件。一种是课堂上的主动学习，教师带着做；另一种是在宿舍使用一个 AI 辅导系统。两个主题分别是表面张力和流体流动，做前后测。

结果有三点。AI 辅导那一组学到的东西，是课堂主动学习组的两倍多；花的时间也更少，中位数 49 分钟，少于课堂时长；这组学生自己报告说更投入。

读这项研究要保留三个条件。第一，这是单一课程、单一学期、两个主题的研究，在一所精英大学，不到两百人，作者与报道都说不能推广。第二，那个 AI 辅导系统不是给对话框一个教师角色就完了。它基于大模型，但由教师预先写好分步骤的讲解、提示与护栏，限制它直接给答案，也限制它编造。学科知识和软件设计都在起作用，不能把差异全部归到模型本身。同一个模型，不写护栏，结果未必如此。第三，测的是即时后测，没有证明长期保持。

Bastani 等 2025 年发表在 PNAS 的研究，提醒的是另一面。它比较了几种不同的使用设计。没有护栏的直接问答，学生练习时表现好，离开帮助独立考试时未必；加了护栏的设计，与对照组没有显著差异。练习做得好，与离开帮助以后独立学会，可能是两件事。

给教师的三条做法。

第一条，从最小的任务开始。把您已经会讲的一段解释，写成系统能照着讲的脚手架，规定它什么时候不给答案。任务说明可以写清，先问学生目前做到哪里，再给一层提示；必要时给例子；最后请学生用自己的话解释。

第二条，把使用设计写进作业要求。练习时允许怎样用，考试时允许怎样用，两者不同。让学生保留独立尝试的记录，解释修改的理由，必要时换一个情境再做一次。不能只凭文字风格认定学生有没有使用工具。

第三条，不要用它替您批改并直接把分数发给学生。批改是核不完的工作，逐份核的代价大。系统可以帮您起草反馈的初稿，分数由您定。

另一头也有证据。世界银行 2025 年 1 月发布的一项尼日利亚试点，六周的课后辅导，用生成式 AI 支持英语学习，学习增益约 0.3 个标准差，作者写“相当于近两年的常规学习”。这个数字来自一次试点，不能推广，但它的前提和 Kestin 那项研究一样，都是教师先设计好结构化的引导，对话框没有直接交给学生。

5.3 科研全流程：七段里哪几段能交

查一下再用

把科研拆成七段，选题、文献、设计、数据、写作、投稿、评审。每一段看三样，能交给它什么，课上实测过的证据，自己做什么；最后一列是规矩。

阶段	能交给它	本课证据	自己做	规矩
选题	列候选问题，找分歧点	深度研究给出四大分歧点（4.4 节）	定值得做的问题	
文献	综述初稿，引用清单	抽查两条引用都真，一篇有正式更正（4.4 节）	逐条点开；查更正与撤稿	科技部：不得使用未经核实的由生成式人工智能生成的参考文献
研究设计	列可选方案与各自缺陷	一个议题三个角色各陈理由（5.4 节）	选方案，定边界	
数据	统计、画图、复算	45 个分数三次统计完全一致（4.4 节）	比较前的三个问题；数据不出门	科技部：依规合理使用生成式人工智能处理文字、数据或学术图像，防范伪造、篡改数据等风险
写作	初稿，改句，翻译	去 AI 味对照（4.6 节）	观点与判断	科技部：使用生成式人工智能生成的内容，特别是涉及事实和观点等关键内容的，应明确标注并说明其生成过程
投稿	格式，投稿前自查		披露	Nature 2023：AI 工具不能署名，须在方法或致谢中说明；科技部：生成式人工智能不得列为成果共同完成人，应披露软件名称、版本和使用时间
评审	不建议把稿子交给它			科技部：在评议活动中使用生成式人工智能的，应事先征得评议活动组织者同意，防止泄漏评议内容

规矩一列引自科技部监督司 2023 年 12 月发布的《负责任研究行为规范指引（2023）》和 Nature 2023 年 1 月 24 日的社论，原文见来源。各校、各期刊另有细则，投稿前以当期要求为准。七段合起来看，能交出去的是找资料、算数据、改句子这一类活；研究什么问题、下什么判断、署谁的名，留在人这里。

行政事务的做法见 4.5 节邮件与纪要那两件事，条件写进任务：按紧急度分三档，每封一句建议回复，只到草稿箱，姓名与机构不进对话。

5.4 一次议题

动手前看

一件校内议题，在正式讨论之前，可以让系统以不同角色各陈述一段理由，把分歧点与待弄清的问题先摆出来。我们用练习册里的一个构造情境做过实测。情境是图书馆是否延长开放时间，全部为课堂构造，不对应真实学校。

2026 年 9 月 5 日，Kimi K3，新会话。提示词原文是，请分别以学生代表、图书馆工作人员、教务负责人三个角色各说一段理由，然后列出三方都同意的一点、真正分歧的一点，以及试行前必须先弄清的两个问题。不要编造已发生的数据。

交回的真实输出摘要如下。学生代表说，期末期间宿舍嘈杂，许多同学需要安静的深度学习环境，现有闭馆时间较早，复习计划常被打断，希望学校在关键期延长自习区开放。图书馆工作人员说，现行排班已饱和，延时意味着重新调配人手或增加加班成本，闭馆后还需清场、设备检查与安保交接，并非到点即可下班；若无数据证明延时能显著惠及多数学生，仅满足少数偏好便增派人力，将挤压日常维护资源。教务负责人说，学校决定在期末前两周试行，正是为了用真实数据验证需求，避免凭感觉决策，试行期间将同步记录入馆人数、座位使用率与反馈，部分同学更关注白天座位管理，这提醒要平衡资源，而非简单二选一。

三方都同意的一点是，优化图书馆服务，为学生创造更好的学习环境。真正分歧的一点是，资源有限时，优先延长夜间开放，还是优先改善白天座位管理。试行前必须弄清的两个问题是，白天座位紧张的具体表现与成因，以及真正会使用延时服务的学生规模与分布。

我们逐段读过。三段立场分明，没有编造已发生的数据，分歧点与两个问题和练习册设计的判断要点一致。

还有一件真实发生的事。发出提示词时，输入框把三处数字吞掉了，“开放到 23:00”“闭馆 22:00”“每段不超过 120 字”都变成了没有数字的句子。模型在思考过程里写，用户说每段不超过字，确实漏了数字，我按 150 字来控制。条件缺失时它自己补了一个，没有反问。它补的条件，得您自己看出来。

还有一句话必须写明。这是模拟观点，不是真人发言。哪些理由成立，还要找谁讨论，仍由人来选。它做的，是帮您多几个视角，先看见分歧在哪里，还有什么没弄清。

5.5 评价怎么改：先分级，再看过程

读一遍就懂

知识获取容易了，考知识的题就要少。改评价分两步，先给每次作业定一个允许级别，再把过程与判断变成可评的东西。

分级有现成框架。AI 评价量表，AIAS，Perkins、Furze、Roe、MacVaugh 四位研究者 2023 年 12 月提出，2024 年 4 月修订并发表。五个级别：

级	允许什么	交什么
1 不用 AI	全程不用	只靠自己的知识与技能
2 帮构思	头脑风暴、搭结构、提改进建议	终稿里不得有 AI 写的内容
3 帮修改	改清晰度与质量，不得新增内容	附上无 AI 内容的原稿
4 它做，学生评	AI 完成指定部分	学生评论它的输出；AI 内容须标注
5 全程用	作为副驾驶贯穿全程	支持自己的工作，说明用法

每次作业先写明级别，学生知道边界，老师知道该评什么。同一门课里可以不同作业不同级别，例如课堂小测定一级，文献综述定二级，项目报告定四级。

把过程变成可评的，这门课里用过五种做法。第一，交提示词与修改记录，3.4 节模板要求留下的那五件，原材料、任务说明、第一版、修改要求、第二版，齐不齐一看就知道。第二，口头解释一个关键选择，为什么这样改，换一个例子当场做。第三，AI 使用声明，软件、版本、用在哪一步，科技部对研究者的要求照搬

给学生，模板见工具箱 9.2。第四，先自己写，再让它批，7.3 节的做法。第五，按四条看结果，5.4 节演示里那四条，有没有编造已发生的数据，是否遗漏重要对象，行动是否具体，是否符合指定用途。

5.2 节哈佛那项研究里，效果来自教师预先写好的脚手架。评价也一样，来自您先定的级别，和先想好要评什么。

5.6 选工具与本地部署

查一下再用

选工具之前，先把“我要完成什么”放在前面。第三部分讲过三种交付形式，每一种要看的地方不一样。要回答。解释概念、改一段文字、比较两个想法，或者让它反问您缺少的条件，任何对话产品都行。看它能不能读您的文件、记您的习惯。

要作品。图片、视频、课件、表格，看产品出的东西您能不能改。可编辑的课件，原图的分辨率，视频的时长上限。第四部分那次做课件，导出的文件是否可编辑，下载打开看一眼。

各家模型的优缺点随版本变，这里不排名，按任务看谁接得上您的材料与工具。下表是 2026 年 9 月各官网的说明和这门课的实测，产品名与主打属于查一下再用。

家族	官网主打（2026-09）	本课实测	适合先试	注意
Kimi，月之暗面	长文本；K3；课件、深度研究、Work、Code	读长文档、数据、教案、综述、小工具	读长材料、备课、综述	页码靠估；引用要点开
DeepSeek	推理模型 R1；开放权重	未实测	推理题；可本地部署	许可以官方页为准
通义千问，Qwen，阿里	开放权重全系列；听悟做纪要	本地 Qwen3.8-27B 断网运行	敏感材料本地处理	机器要求见下文
豆包，即梦，字节	语音通话；即梦出图出视频	即梦出图、Seedance 十秒短片	示意图、短片、口语	语音本课未测
文心，百度；智谱 GLM	对话产品；GLM 有开放权重版本	未实测	备选入口	以官网当日为准
ChatGPT，OpenAI	GPT-6 Astra；agent；操作电脑	未实测；6.1 节读它的发布	任务型、跨软件的生活	分批开放
Claude，Anthropic	长文档；Claude Code	本课件的脚本与导出由 Claude Code 完成	编程、批量、长文档	
Gemini，Google	多模态；NotebookLM	未实测	把录音、文件、图片放在一起读	

表里写“未实测”的，是这门课没有跑过，不作评价。

要任务。看产品有没有工作台，能读规范、多步执行、交付文件、把发送留给您。截至 2026 年 9 月各官网的说明，Kimi 有课件、深度研究和工作台，ChatGPT 有 agent，Claude 有 Code，飞书有命令行工具，腾

讯 WorkBuddy 与 Qoder 属于桌面工作台一类。这些名字属于 [查一下再用](#)，用之前请看官网当日说明。

比较时不需要一次注册很多账号。选两个自己能正常使用的入口，给相同的一小份材料和要求，记录哪个结果更贴近用途，自己分别改了什么。对多数教师，更好的起点是先用熟一个，练出提要求、给反馈的手感，碰到它做不了的事再加第二个。

模型分两类拿到手，怎样拿到、在哪里运行、材料去哪里，下表并排看。开放权重这一类，常有人叫它“开源”，DeepSeek-R1、Qwen3、Llama 都属于此类；闭源服务这一类，GPT-6 Astra、Claude 属于此类。第四部分十件事用的 Kimi K3 网页版是闭源服务；那台断网的笔记本跑的是开放权重模型。两类各有代价，前者要自己的设备与维护，后者材料要出门。严格说，“开源”一词在 AI 里另有定义，要求训练数据、代码与权重都开放，多数“开源模型”只开放了权重，所以这里一律说“开放权重”。

	开放权重	闭源服务
怎样拿到	参数文件可下载	只通过网页、App 或接口使用
在哪里运行	自己的机器或租的服务器，可断网	厂商的服务器
材料去哪里	留在本机；检索、插件、同步仍可能联网	发给厂商，按其条款处理
能不能商用	看各自许可证，能下载不等于能商用	看服务条款
本课证据	第四部分那台笔记本，三句话十余秒	实测 01 到 09，Kimi K3 网页版

最基础的本地部署搭配，按机器分三档，这几档是经验之谈，以各软件官方说明为准。

您的机器	能跑的规模	够做什么
笔记本，16GB 内存	4 到 8B 参数，4-bit 量化	改写、摘要、翻译，速度可接受
笔记本，32 到 48GB 内存	14 到 32B，4 到 6-bit	本课这台：27B、6-bit，每秒 9.4 个 token
台式机加 24GB 显存显卡	30B 级 4-bit；70B 级要更多显存或很慢	本地处理整份文件与批量任务

软件用 LM Studio 或 Ollama，Mac 上可以用 MLX；模型从 Hugging Face 或魔搭 ModelScope 下载。第一次装，选一个 8B 左右、4-bit 的模型试，能跑起来再换大的。

本地部署要分开看三件事。开放权重，说明模型的参数文件可以下载，具备条件的人能部署到自己的机器上。许可证，说明下载之后允许怎样用，下载得到的未必能商用。部署条件，说明要什么设备、性能不够，商用许可有了，您的机器也未必跑得动。三件事分开核，以各模型官方页面的说明为准。

什么时候值得。学生数据、未发表的研究、内部文件，规则说“不能出门”的材料，就在门里处理。第四部分那台 48GB 内存的笔记本，断网运行一个 27B 参数的开放权重模型，三句话用了十几秒，峰值内存 23.12 GB，占整台机器一半。

代价也要写清。本地模型不知道今天的事，读不了您没给的文件，速度看机器。运行在本地，这次任务的输入就不必交给远程服务，可数据也不一定就不出门。检索、插件、同步仍可能联网；谁能访问这台电脑、数据怎样保存，仍要弄清楚。部署牵扯设备、维护和组织安排，通常要和专业同事一起定。

5.7 换一个学科：品牌出海怎么用它

动手前看

有教师问，怎样用 AI 辅助企业品牌出海。这不是这门课的主线，但正好看同一套做法换一个学科怎么用。情境为课堂构造，一家深圳做智能家居小家电的企业想进入德国市场。2026 年 9 月 11 日，Kimi K3。

交给它两件事。第一，列出进入市场前必须查清的 10 个问题，每个问题写明去哪类一手来源核实，不许编数字。第二，把一段中文品牌介绍改成面向德国消费者的英文版，标出三处要按当地文化调整的地方。它先联网查了一轮，再给清单。

# 要查清的问题	它说去哪核	我核的
1 适用哪些 CE 协调法规，评估路径是什么	EUR-Lex 指令原文；TÜV、VDE 等公告机构书面确认	对：LVD 2014/35/EU、EMC 2014/30/EU、RED 2014/53/EU 均为现行指令
2 是否指定了欧盟境内的 GPSR 责任人	Regulation (EU) 2023/988 原文	对，法规编号存在；适用日期本轮未打开原文，讲前核
3 是否完成 WEEE 注册 (ElektroG)	Stiftung EAR 官网	对：官网核，注册与境外厂商授权代表均为其职能
4 是否完成包装法注册 (LUCID)	ZSVR 包装中央登记处	对：官网原文"在德国销售带包装商品须在 LUCID 注册"
5 智能功能的用户数据是否合规	GDPR 原文；BfDI；各州数据保护局	对，机构存在；网络弹性法生效日期以官方为准
6 税务、担保与消费者法义务	BZSt；BGB、ProdHaftG、PAngV 原文	对，法条存在；2 年瑕疵担保为 BGB 通行规定

核法是，机构和法规逐个查官网是否存在、是否管这件事；不核数字，因为要求它不给数字。它改写的英文版：We are a smart home brand from Shenzhen, China. For over ten years, we have focused on small home appliances, and our products are now used in more than 30 countries. We build for reliability and durability — technology that makes everyday life at home simpler.

它标出的三处调整：第一处，"匠心品质"不直译，改为可靠与耐用，它的理由是德国消费者看检测报告、质保年限与材料说明，不看形容词；第二处，"让科技温暖每一个家"改为让生活更简单，理由是强情感诉求在德国广告文化里被视为营销话术；第三处，"30 多个国家""十年"保留，但去掉"每一个"这类绝对化表达，理由是德国反不正当竞争法对事实性宣称要求可举证。三条理由是它给的判断，可信度由懂市场的人定。

对商学院和外语系的教师，做法和 3.3 节那份教案完全一样。材料是中文原文和目标市场，任务是清单加改写，核对是查机构在不在、说法准不准，最后把这条提示词存进模板。它列的清单能用。

06

PART 06

下一段变化

6.1 怎样读一次发布：GPT-6 Astra

查一下再用

把第一部分那条时间线接到最新的节点。2026 年 9 月 3 日，OpenAI 发布 GPT-6 Astra。前面讲过的模型、工具和连续任务，到这里用到了更复杂的专业工作上。这一节分三块读，分数、定义、发言，三块要分开。先看分数。下面三组都是 OpenAI 自己报告的结果，从上一代 GPT-5.6 Sol 到 GPT-6 Astra。

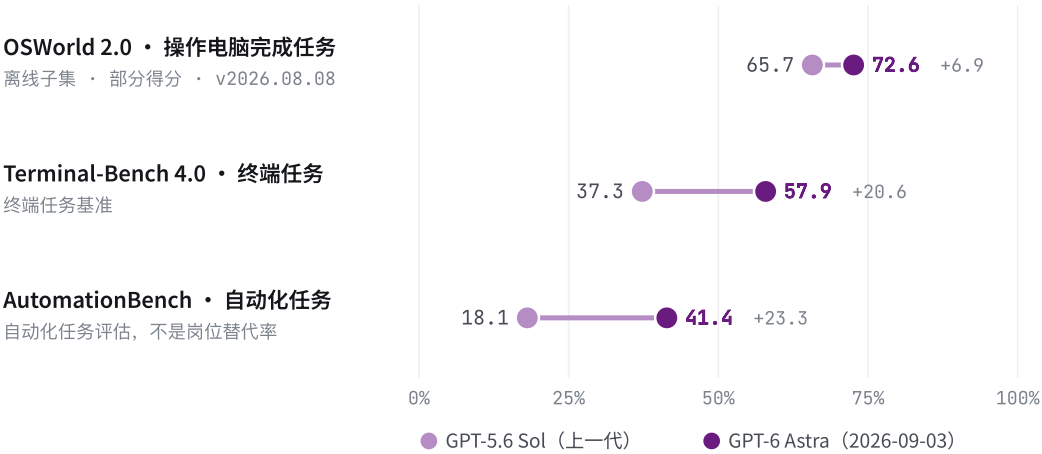


图 6 · OpenAI 报告的三项结果：GPT-5.6 Sol 到 GPT-6 Astra。取厂商在不同推理设置下报告的最高结果；测评环境与日常产品不同。来源：openai.com/index/gpt-6-astra, 2026-09-03。

第一组，OSWorld 2.0，操作电脑桌面的测评，65.7% 到 72.6%。第二组，Terminal-Bench 4.0，终端任务的测评，37.3% 到 57.9%。第三组，AutomationBench，自动化任务的评估，18.1% 到 41.4%。三组都在上升，第二组和第三组上升最多。

读这三组数要保留的条件。三组都是厂商在不同推理设置下报告的最高结果，测评环境和日常产品不是一回事。第一组是离线子集上的部分得分，版本号 v2026.08.08。第三组是自动化任务评估，不能读成“岗位替代率”；100 减去分数，也不是现实工作的失败率，我不做这种换算。分数是厂商报的，判断要自己下。

开放条件。这次发布先向企业客户分批开放。GPT-6 Astra 在 OpenAI 自家安全框架里达到了网络安全“关键”级别，这是特定类别的能力等级，不是总体危险等级，也不是任何形式的 AGI 认证。您手里的产品开放到哪一步，用之前再查。

读任何一次新发布，都可以照这一节的办法。先找厂商自己报告的数字，把测评的名字、版本和条件抄下来；再找定义，看它用的词是谁定义的；再找发言，看是谁在什么场合说的。三样东西分开放，判断才有依据。Astra 放在尾声讲，是因为读它要用到前面讲过的任务、工具和连续执行这几个概念。不知道任务是什么，就读不懂“自动化任务评估”在测什么。

教师可以把它换成一个更具体的问题。过去要自己在几个软件之间接起来的工作，系统现在能参与到哪一段。整理一组材料、比较观点、做成报告、按反馈修改，按现在的能力，这些可以交给系统。这一句是事实。岗位怎样重新组合，是推断。成本降下来以后出现的新需要，是价值选择。三句话要分开写，也要分开读。

6.2 AGI 是谁说的

查一下再用

这次发布又引出一轮 AGI 的讨论。AGI 通常译作通用人工智能。讨论这个词，先弄清三层。定义，讨论的人在这个词下具体指什么。判断，哪家机构、哪位人物下了什么判断。证据，哪些任务、环境和观察结果支持这个判断。

先看定义。OpenAI 在自己的章程里写的定义是，高度自主，并在多数有经济价值的工作上超过人类的系统。这是一家公司的定义，不是各方共用的标准。学术界另有讨论，例如 Morris 等 2023 年的 Levels of AGI，它分别处理表现水平、通用程度、自主性与部署问题。这个框架能把从哪几方面评讲清楚，但这里不用它给任何模型定级。

再看发言。据 Axios 2026 年 9 月 3 日的报道，OpenAI 的 Greg Brockman 在当天的媒体简报中说，“Welcome to the AGI era.”这句话来自发言人对技术节点的判断，不是模型自己宣布，也不是学界共识。引用时要保留人名、场合、媒体和日期。

机构定义与人物判断都不是共识，但都有价值。读到这类表述时，先问是谁在什么场合说的，依据的是哪些任务和结果。用 6.1 节的三组分数对照这条定义，可以看出两者之间的距离。分数测的是特定测评环境里的特定任务，定义说的是“多数有经济价值的工作”。从前者推到后者，中间缺的是任务分布、现实环境和长期观察，这些都不在一次发布里。

课堂上如果学生或同事问“AGI 是不是已经到了”，可以用这三层回答。先说定义是谁的，再说判断是谁做的，最后说证据覆盖到哪里。这样回答不需要预测未来，也不会把一家公司的发布会写成学界结论。能力再往前走，读懂材料、提出问题、和人合作、解释自己的选择，这几样照样用得上。教什么会跟着任务变，工具已经能顺手做完的机械步骤可以少教一些，多教真实情境里的理解和动手。

6.3 变的是任务

查一下再用

导言留下的问题到这里回答。AI 能参与这么多工作，我们为什么还在招人。先看一份数据，再回答。

Anthropic 在 2025 年 2 月 10 日发布了第一份经济指数报告，分析了自己平台上百万量级的匿名对话，看人们实际拿 AI 做什么。先说清这组数是怎么来的。这是一家公司一个平台的使用数据，映射到职业任务分类上，只反映该平台使用者，不是全社会统计。后续期数的数字有变动，这里只取首份报告。

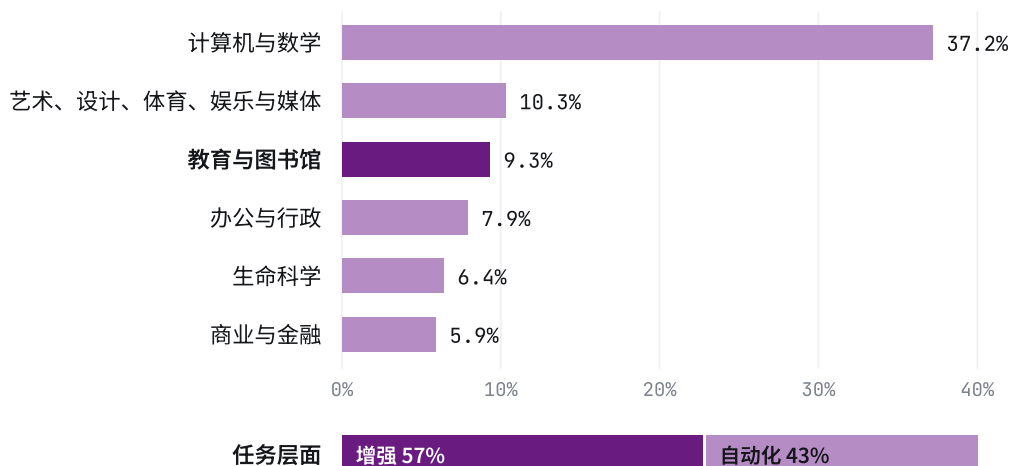


图 7 · 使用占比最高的六类职业任务，以及任务层面的增强与自动化。Anthropic Economic Index, 2025-02-10；一家公司平台的匿名使用数据，不是全社会统计。

按任务分。57% 的用法是增强，人在回路里来回改、学习、核对；43% 是自动化，整段交出去。增强多于自动化。

按职业类别的使用占比。计算机与数学 37.2%，最高。艺术、设计、体育、娱乐与媒体 10.3%。教育与图书馆 9.3%，排第三，高于办公与行政的 7.9%。生命科学 6.4%，商业与金融 5.9%。您的同行已经在用了，用在具体的任务上。

再看一组数。约 36% 的职业，至少四分之一的任务出现过 AI；约 4% 的职业，四分之三以上的任务出现过。两个数放在一起看，AI 进入的是任务，不是整个岗位。一个岗位是一堆任务，有的任务交出去了，有的任务多了人在旁边核。

我的回答从自己的经历说起。导言里写了四处变化，人才触达平台自己带 AI 团队做出来并上线，全球人才数据库由多个 AI 程序持续维护，对外材料与研究初稿当天出，敏感数据在本地断网处理。四件事的共同点是，准备和试错的成本下降了。成本下降以后，以前做不动、做不起而没有开始的事，现在有机会尝试。我们招人，招的是能理解真实需求、组织材料、看结果改做法的人。这是一家机构的经历，不推广到所有行业。

放回大学。以前只来得及给全班一份练习，现在有条件准备三种难度的材料；以前一次活动只能做一份通用说明，现在可以针对不同受众分别解释清楚；以前每位学生一份针对性反馈做不到，现在可以让它先出初稿。这些事要有人发现、组织、看结果。哪件事值得做，做给谁，怎样接着改，数据里没有答案。这是趋势判断，来自当前能力与一家平台的数据，具体到您的一周，看自己正在怎样变化。

6.4 知识、判断、行动

读一遍就懂

这一节不空讲标题里的三个词，拿第三部分那份教案的三处判断来对。每处判断背后有一块专业知识，每处判断之后有一个动作，每一行最后还能看出系统能替您做哪一段。

那一处判断	背后的专业知识	当时做的判断	接下来的行动	它能替您做哪一段
时间给了名词，没给学生	学生要自己完成一次解释才算学会；知道这批学生会卡在“AI 懂了等于 AI 有常识”	目标是解释，不是认识名词；时间要转给学生	写修改要求；第二稿把学生时间从 5 分钟改到 24 分钟	第二稿它改，目标您定
连线是示意，不是模型计算	2.1 节那张热力图，权重是算出来的，不是投出来的	活动保留，补一句话把概念说准	在活动后加一句“这是帮助理解的示意”	它不知道您的活动会误导什么
通用、正确、没有您的味道	上次课讲过搜索引擎；这批学生没有编程基础	给它材料，而不是要它更努力	把 2.3 节那份材料清单放到它面前	材料到位后初稿它给
课后：试教、看表现、修订	哪个活动让学生真的说出理由，哪个只是热闹	下次保留什么，换掉什么	把有效做法存成模板	模板它帮您整理，有效无效您看

四行合起来看。知识用来认出学生会卡在哪里，判断用来选哪项活动真能达到目标，行动是试教、看学生表现、照结果修订。看最后一列，这三件事里只有第一件是系统能大量补充的。它可以补充知识，可以按您的判断修改，可以帮您整理模板。它不知道您的活动会误导什么，也看不见哪个活动让学生真的说出了理由。AI 参与以后，教师的专业性具体就在这里。

回到课程的题目。论文里的注意力，是处理信息的一种机制。拿它做课程题目，问的是，能调用的能力越来越多，教师准备把自己的注意力放在哪里。从这份教案看，值得放的地方有两处，学生终于能自己解释清楚的那个时刻，和看出哪个活动让学生真说出了理由的那双眼睛。这两处它替代不了。技术还会往前走，不必等到全弄明白才开始。先做成一件真事，再照结果改自己的做法。第七部分把这门课最想说的话说完，第八部分给出具体的安排。

07

PART 07

认知与价值观

读一遍就懂

前面六个部分讲工具、方法和数据。这一部分不再给新工具、新名词和操作清单，只把这门课最想说的话说完。它回答三个问题。墙塌了以后，多出来的事该由谁来定；工具一样了，人为什么不一样；当技术什么都能实现的时候，什么决定世界往哪里走。这一部分的看法都是我自己的，引用的原文都注了出处。

7.1 那堵墙塌了

第四部分的十件事、第三部分的那份教案、第五部分那台断网的笔记本，如果只留一个结论，是这一句。过去挡在“想到”和“做出来”之间的那堵墙，塌了。

把这两天做成的七件事排成一张表，看过去需要什么，现在中间隔着什么。

这两天做成的事	本课记录	过去需要什么	现在中间隔着什么
一份 45 分钟教案，初稿与第二稿	实测 10 与 14，Kimi K3，第二稿把学生时间从 5 分钟改到 24 分钟	教师自己写两稿，没有人替您出第二稿	一份写清对象、目标与材料的任务说明
十页教学课件	实测 02，Kimi PPT，约 6 分钟	一位会排版的人，或自己一个晚上	一份大纲，和上台前逐页读一遍
一份带引用的研究综述	实测 06，Kimi 深度研究，约 20 分钟，抽查两条引用都真	研究助理数日检索与整理	一个问清范围的问题，和点开两条引用的耐心
45 个分数的分布与统计	实测 07，Kimi K3 自写 Python，本地复算一致	一位会用统计软件的人	知道均分会掩盖什么
一小时二十七分钟会议的纪要	实测 09，飞书妙记，2,536 字，待办 0 条	专人整理	自己把待办列一遍
270 亿参数模型在笔记本上断网运行	实测 11，Qwen3.8-27B，每秒 9.4 个 token	机房、运维与预算	一台 48GB 内存的笔记本
一个上线的人才触达平台	我的经历，见导言	外包团队，按月计	一个人加一支 AI 团队

“过去需要什么”一列按我的经历写，不含工期与金额，您可以按自己单位的情况改写。七件事背后是同一件事。有一个说法叫技术平权，技术从稀缺资源变成了基础设施，像水，像电。以前，“我有一个想法”和“这件东西真的存在”之间，隔着一个技术团队和一笔预算。现在中间隔着的，是您愿不愿意坐下来，把它讲清楚。

我是学计算机的，这堵墙塌得有多快，我比多数人清楚。我原来就住在墙的那一边，靠这堵墙工作。

第六部分用数据回答了“为什么还在招人”，这里补一个我自己的类比。从 A 地到 B 地，最早只能走路，那时的要求只有一个字，到。有了蒸汽机，坐火车，追求的是快。有了电，追求的是省、是舒服、是准点。今天订一张票，看的是几点到、要不要转、座位舒不舒服，到不到已经不是问题。工具每变一次，人追求的东西就往上走一层。路还是那条路，要求变了。AI 也一样。机器接过去的是“做”，多出来的是能做的事。这个类比是我的归纳，没有出处。

那多出来的这些事，该做哪些，谁来定。

7.2 认知成为分水岭

人人都拿得到同样的工具，工具就不再是差距。差距去了别处，去了您怎样理解这件事、怎样定义问题、看见了什么别人没看见的地方。本课里差距出现过五次，都有记录。

哪件事	用法一	用法二（本课记录）	差在哪里
写教案	只说“帮我写一个教案”（实测 12）：通用、正确、没有您的味道	给材料清单与五项条件（实测 10 与 14）：第二稿把学生时间从 5 分钟改到 24 分钟	知道这批学生是谁；知道目标是解释，不是认识名词
看成成绩	三个班均分 78.1、82.1、76.0，相差六分，看似接近	看分布（实测 07）：A 班一个 45 分把均分拉低，去掉后 80.5	知道均分会掩盖什么；比较之前先问是不是同一份试卷
用综述	二十分钟的报告拿来就用	点开两条引用（实测 06）：都真，其中一篇有正式更正	知道引用是最容易出错的地方
核页码	相信它标的页码	逐题回原文翻（实测 01）：五题一错，一题课程组判错后更正	知道位置靠估；也知道核的人自己会错
做课件	接受第 8 页的 41.8	回原文（实测 02）：正文 41.0，表二 41.8，论文自身不一致	知道来源冲突时要自己定以哪个为准

五行合起来是一句话。技术把人和人拉平了，认知把人和人拉开了。这里说的认知很窄，只指两件事，定义问题，选择值得做的事。同一堂课，同一个工具，两个人用出来能差很多。账号权限、会员等级拉不开这个差距，拉开的是对这件事该怎样做，两个人各理解到了哪一步。

教师本来就是靠认知工作的人。对一个学科的理解，对学生状态的判断，对什么值得研究的那点直觉，这些在 AI 时代没有贬值。过去它们要经过一层技术才能变成产品，现在这一层没有了。也要说清楚哪一部分会贬值，能写成说明书的那部分。流程、模板、标准答案，凡是能写清楚的，系统都能做。写不清楚的那部分，对学生的判断、对题目的直觉，还在您手里。

7.3 知识与认知

这两个词平时混着用，混着用就看不清楚。

知识，指事实、公式、条文、操作步骤，可以检索、复制、交给系统的东西。今天知识近乎免费，问一句就有。认知，上一节已经收窄，只指定义问题和选择值得做的事，也就是决定问什么问题、什么值得做、为谁负责。差别有两条。知识是答案，认知是问题，系统能给一万个答案，给不了一个值得做的问题。知识是存量，认知是算法，存量可以下载，算法只能长出来。

把这两个词放回 6.4 节那张表。第二列“背后的专业知识”是知识，第三列“当时做的判断”是认知。看最后一列，知识它能大量补给您，判断您定，行动您做，模板它帮您整理。知识可以外包给系统，认知不能。一外包，人就成了传话筒。也要说一句实话，认知里能写清楚的部分，将来机器也会接过去。

认知怎么长出来，我有三条做法。第一，下场做一件真事。您要承担后果，认知才真的更新；在旁边看一百场，不如自己下场输一次。第八部分的一个月计划，就是为这一条设计的。第二，定期把自己推翻一次。找出您三年前深信不疑、今天已经不信的一条判断，能说清它是怎么变的，说明您的认知还在长。

第三，先自己想，再问它。2025 年麻省理工学院媒体实验室一项研究，54 人分三组写作文，一组用大模型，一组用搜索引擎，一组只靠自己，全程戴脑电设备。用大模型的一组，脑内连接是三组里最弱的，研究者把这种现象叫认知负债，现在省下的脑力，以后要还。同年微软与卡内基梅隆大学对 319 名知识工作者的问卷显示，对生成式 AI 越信任的人，报告的批判性思考越少；对自己越有信心的人，思考越多。两项都是单

一研究，样本不大，但方向一致。所以我的做法是，问它之前先花几分钟写下自己的判断；拿到答案先问两句，怎么验证，有没有反面证据。把它当教练、当对手，用它检验和挑战自己的想法，不把它当答案。

7.4 行易知难

导言里埋下的四个字，这里说完整。四段原文，三段中文，一段英文，按时间排。

《尚书·说命中》，傅说对商王说，“非知之艰，行之惟艰。”知道不难，做到才难。这是中国人两千多年的常识。

1918 年 12 月 30 日，孙中山在上海写下《孙文学说》的自序，第二年刊行，收入《建国方略》，题为“心理建设”。自序里说，“此思想之错误为何？即‘知之非艰，行之惟艰’之说也。”他认为革命党人心理上的错误就在这句老话，于是把它倒了过来。第五章写道，“由是而知‘行易知难’，实为宇宙间之真理。”那时这四个字是勉励，劝人不要怕做。

再往前四百年，王阳明在《传习录》卷上里说，“未有知而不行者。知而不行，只是未知。”真知一定会发动为行动，没有发动，说明还没有真知。这句话容易听岔，它与“懂了就算做了”那种便宜话没有关系。

2000 年 8 月 25 日，Linux 的创始人 Linus Torvalds 在内核邮件列表里写下一句，“Talk is cheap. Show me the code.”空谈便宜，给我看代码。那一年代码稀缺，想法谁都有，能写成跑得起来的东西的人很少。二十六年过去，技术圈把这句话翻了过来，“Code is cheap. Show me your idea.”代码便宜了，让我看看你的想法。这一句是流行表述，没有单一出处；它说的事，这门课的十八项实测都能作证，稀缺的那一端换了位置。

四段放在一起，说的是同一件事。今天机器把“行”接了过去，孙中山那四个字成了事实。用我自己的话说，在 AI 时代，难的不是做到，难的是知道。读完这本手册，您知道了它能替您做的十件事，也知道了它哪里靠不住。按王阳明的标准，如果读完之后什么都没变，就还没有“知道”。

7.5 同一套架构，两种用途

这一节我在台上也犹豫过要不要讲，最后决定讲。

世界上所有找人、建立联系、持续沟通、推进转化的系统，技术上拆开，都是同样四步。下表中间一列是我的团队在做的事，右边一列是公开反诈宣传中对电信网络诈骗的通用描述，只写到这一层，不再细。

四步	全球人才引进（我的团队在做的事）	电信网络诈骗（公开反诈宣传中的通用描述）	技术侧差别
找到联系方式	公开的学术履历、会议名单、校友网络	泄露或买卖的个人信息	无
建立联系	邮件、社交平台、校友引荐	冒充身份加好友	无
持续沟通	了解研究方向、家庭考虑与回国意愿	长期取得信任	无
推进转化	对接高校与企业岗位，请人回中国发展	诱导转账	无

最后一列四行都是“无”。同一套架构，同一批工具，两种用途。技术这一侧没有区别，区别全在人这一侧。同一套 AI 能力，可以帮一个学生找到方向，也可以骗走一位老人的养老钱，中间隔着的不是技术水平。

教师是这个链条上最靠前的一环。您教出来的学生，会拿着这些工具去做事。他们拿去做什么，和您在课堂上怎样讲“该做什么”有关。这本手册讲工具讲了六个部分，这一节和下一节，才是我最想说的。

这一节不写具体客户、个人信息和金额，反例里也没有可以照着做的操作细节。

7.6 价值观决定世界走向

技术平权之后，工具已经拉不开人和人，拉开的是认知。可认知里能写清楚的那部分，前面说过，将来也留不住。到了那一步，人手里剩下的只有价值观，选什么值得做，为结果负责。技术什么都能实现的时候，世界往哪里走，就由这一样东西决定。

技术是不是中性的，学界有两派。一派说刀不杀人，人杀人；另一派说，工具一上手，局面就已经变了。我不裁定哪一对，只引后一派引用最多的一句。技术史学者 Melvin Kranzberg 1986 年在《Technology and Culture》期刊发表的六条定律，第一条是，“Technology is neither good nor bad; nor is it neutral.” 技术既不善，也不恶；但它也不是中性的。刀不杀人，对；但人人有刀和没人有刀，是两个世界。

放到学校里。学校开不开 AI 课，学生怎样用 AI，这两件事发生之后，学校已经是另一所学校了。往哪个方向变，和您怎样讲有关。AI 不会问您这件事该不该做，它只会问，您要我怎么做。

这门课的副题叫一篇论文之后的世界。论文是八位研究者写的，论文之后的世界往哪里走，是拿着这些工具的人定的。

7.7 三句话

两个半小时的课，压回三句话。

第一句，二十六年 before Linus 说，别空谈，给我看代码。今天代码不稀缺了，稀缺的是您想做什么。

第二句，行易知难。知道了，剩下的就是修行。知道了以后，该不该做，是您的选择。这个选择，会变成您学生的选择。

第三句，也是我最在意的一句。教师是教人的。技术把门槛拆掉以后，判断力和价值观，是老师能给学生的最后两样东西。这两样，AI 教不了。

08

PART 08

从今天开始

读一遍就懂

要让今天读到的东西进到工作里，接下来一个月不必急着追新产品。从第三部分您选出的那件事开始，分三段走。今晚一件事，一周三条，一个月一件真事。一个月后手中应当有三样东西，任务说明、好样例、核对清单。

修行清单：一周做三条

今晚一件事。把今天满意的那次对话存成模板，两分钟。它答得不满意的那一处，补进提问里。下次换个材料，改两个词就能用。

一周三条。下面十行是本课十类场景的第一步提示词，都是 2026 年 9 月 4 日至 5 日实际发出过的原文缩写，旁边是当时用的工具和核过以后的结论。从中勾三条，本周做完。完整提示词与补全条件的版本，见工具箱 9.1。

#	场景	第一步提示词（本课实测原文，缩写）	本课用的工具	结论
01	读长文档	请读这篇 PDF：列三级大纲；出 5 道选择题并标页码；指出两个易混概念	Kimi K3	需核对
02	做 PPT	按这份大纲生成 10 页教学 PPT，每页一句标题、三条要点，学术风格，白底	Kimi PPT	需核对
03	做图	一张《____》课的板书风格示意图：白底黑线，含中文标注，无装饰	即梦 5.0 Pro	需核对
04	做视频	生成一段 10 秒延时视频：一粒种子发芽，侧视剖面，固定机位，无文字	即梦 Seedance 2.5	试用
05	陪练	你是我的英语口语陪练，用大学面试场景问我三个问题，每答完一个纠正	Kimi K3	可用
06	深度研究	就“____”做一份研究综述：分歧点、近三年进展、研究空白，每条附来源	Kimi 深度研究	需核对
07	数据	期末成绩（已匿名）三个班：按班级画分布图；标异常值；给均分与标准差	Kimi K3, Python	可用
08	邮件	读最近一周收件箱，按紧急度分三档，每封一句建议回复，只存草稿	飞书邮箱 CLI	需核对
09	纪要	把这段录音转成纪要：决定了什么、谁负责、什么时间、还有什么没定	飞书妙记	需核对
10	做小工具	做一个给学生用的课堂练习网页：对象、主题、5 道单选题、答完显示得分与解析、单文件 HTML	Kimi K3	可用

“可用”指核对后可以直接进入工作，“需核对”指初稿可用但必须按第二部分讲的方法核，“试用”指能用但不稳。结论是 2026 年 9 月的记录，工具会换，这十件事和核法不换。

一个月一件真事。就是下面的四周计划。1950 年，图灵在《Computing Machinery and Intelligence》的最后一句写道，“We can only see a short distance ahead, but we can see plenty there that needs to be done.”我们只能看见前面很短的一段路，但看得见那里有很多事要做。七十多年了，这句话还是这样。师傅领进门，修行在个人。

一个月计划

周	做什么	留下什么
第 1 周	用一个顺手的工具，完成一件真会用到的小事，例如一段课程导入、一份通知、一组材料的整理	原材料与第一版结果，以及您改了哪一处
第 2 周	再做一次同类任务，把上次漏说的条件提前写进去；给这条提示词配一套小测试集，五份不同材料各跑一遍	第一稿是不是更接近用途；哪些问题来自材料，哪些来自要求；五次里几次要改
第 3 周	把每周重复的那件事拆成步骤，标出全交、协作、自己（3.5 节），给全交的步骤写专门的任务说明，存成模板，附一份您认可的样例	一张分工表，和您的第一份 Skill
第 4 周	找一位同事互相看一次结果，请对方指出一个您没看见的问题，用这条反馈再改一次	一条别人给的反馈，和改后的版本

四周只有一个原则，每一周都从上一周留下的东西开始。第 1 周留下的第一版结果，第 2 周拿来比较；第 2 周分清的问题类型，第 3 周写进模板“要准备什么材料”那一栏；第 3 周的模板和样例，第 4 周交给同事看。一个月结束时，三样东西是一路长出来的，不用最后一天补写。

第1周，挑一件真会用到的事。最好下周的工作里就要用，这样您会认真核，也会认真改。留下原材料和第一版结果，同时记录自己改了哪一处，这是后面三周的起点。

第2周的关键是把上次漏说的条件提前写进去。做完之后比较两次的第一稿，您会开始分清三类问题。哪些来自材料不够，哪些来自要求模糊，哪些需要换一种功能。这个区分比任何技巧都有用。

第2周还要做一件小事，给这条提示词配一套小测试集。选五份不同的材料，或者五个不同的问题，用同一条提示词各跑一遍，记下几次结果能直接用、几次要改、改在哪里。一次跑对说明不了什么，五次里四次能用才算这条提示词可靠。这门课的实测也是这样做的，读长文档那一次的页码核对重复了三遍才写进结论。模板见工具箱 9.4。

第3周，先拆再存。把那件事拆成步骤，标出哪些全交给它、哪些一起做、哪些自己来，只给全交的步骤定下固定做法。存下的是做法，不是答案。模板里不只放一句提示，还要写适合什么任务、要准备什么材料，以及一份您认可的样例。工具支持 Skill 的，就存成 Skill；只有普通对话框的，先存成文档。

第4周的关键是别人的眼睛。请同事指出一个您自己没有看见的问题，可能是一处编造，可能是一处遗漏，也可能是系统替您补的一个条件。用这条反馈再改一次，留下改后的版本和适用条件。

自评四问

一个月以后，不必用掌握了多少名字来评价自己。看四件事。

第一问，是否完成了一件以前想做但没有开始的工作。第二问，是否更会交代条件，对象、目标、材料、交付、做到哪一步，五项里有几项是自己主动写进去的。第三问，是否能说明自己为什么修改，能不能用一句话讲清哪条反馈改变了结果。第四问，是否留下了一套下次还能用的方法，任务说明、样例、核对清单是否都在。

四问里没有一问问工具。这是有意的。一个月后如果您换了工具，四问的答案不会变；如果四问里有一问答不上来，换再多工具也补不上。第二问和第三问尤其值得认真对待，一个对应第二部分的任务说明，一个对应第三部分的判断，是整门课最想留给您的两种能力。

自评时可以对照本课的教案案例。第一稿出来后，我们能说出三处判断的原因，能把它们写成一条修改要求，能核第二稿的数字，也留下了模板和核对清单。如果您的那件事也能走到这一步，这个月就没有白过。

记录时间

请记录每次任务实际花费的总时间，包括准备材料、等待生成、核对结果、修改要求四段。这样记有两个用处。一个是看它到底有没有帮上忙，有些任务算上核对时间并不比自己做快，这也是有价值的发现。另一个是看自己是不是更理解问题了，第二次同类任务的准备时间通常会缩短，因为您知道该给什么、该核什么。

本课的实测记录里有几个可以参照的时间。深度研究那次约 20 分钟，做课件那次约 6 分钟，做练习网页那次约 2 分钟，本地模型生成三句话十几秒。但这些都只是等待生成的时间。深度研究的两条引用各点开一次、做课件的分数回原文查三处、五道题逐题回 PDF 检索，这些核对时间没有单独计，可这件事值不值得交出去，就看这一段。您记录时请把四段分开写。

现在请在第三部分写下的那件任务旁边写一个具体时间，准备什么时候继续。可以就是本周某天备课前的十分钟。开始得小一些没有关系，让它进入真实工作，才会得到属于您自己的经验。

继续学的入口

想系统学习的教师，可以从面向公众的人工智能概论与生成式 AI 课程开始，例如吴恩达在 deeplearning.ai 上的 [AI Prompting for Everyone](#) 与 [Generative AI for Everyone](#)，或者芬兰赫尔辛基大学的 [Elements of AI](#)。想看公式的教师可以看 [Stanford CS224N](#)。讲教师能力的，有 UNESCO 的教师 AI 能力框架，和教育部等五部门 2026 年 4 月印发的《“人工智能+教育”行动计划》。看的时候手边留一件正在做的事，学到一段就试一段。完整的一手来源列在手册末尾。

09

PART 09

工具箱：拿来即用

这一部分全部可以直接复制。凡标注“本课实测原文”的提示词，都是 2026 年 9 月 4 日至 5 日在具体工具上实际发出并核对过结果的，按手册正文的标记属于 **动手前看**；其余是在实测基础上补全条件的版本，属于 **读一遍就懂**。提示词里方括号【】内的内容是填空，使用时换成自己的，删掉不需要的行。

发出之前检查三件事 对象写了没有；这一步做到哪里写了没有；结果怎样核对写了没有。三件都写了，再发。

9.1 十类场景的提示词

① 读长材料：出大纲与题目

什么时候用 一章教材、一篇论文或者几十页的文件摆在面前，要先理出结构，再拿它出题，或者找出学生最容易卡住的地方。

怎么填 直接能用的是第三条。【课程名】【X】填到具体章节；对象那一行写年级、专业和他们学过什么，它照这一行决定把话说到多深。例如：这是《材料力学》第 6 章"弯曲应力"的讲义，对象是大二土木专业学生，学过理论力学，没学过弹性力学。

本课实测原文（Kimi K3，2026 年 9 月 4 日，输入为一篇 15 页论文的连接）：

请读这篇 15 页的 PDF：【链接】然后：① 列出三级大纲；② 出 5 道选择题，每题标出对应页码；③ 指出初学者最容易混淆的两个概念。用中文回答。

课堂上用的教材版本（按实测改写）：

这是我的教材第【X】章。请：① 列三级大纲；② 出 5 道选择题，每题标页码；③ 指出学生最容易混淆的两个概念。

补全条件的版本：

这是【课程名】第【X】章的材料（已附上 / 链接如下）。对象是【年级、专业】学生，他们【已学过什么 / 没有什么基础】。

请：

- ① 列三级大纲，每级标注在原文的页码；
 - ② 出 5 道选择题，每题标页码，并引用原文中支持答案的那一句；
 - ③ 指出学生最容易混淆的两个概念，各用一句话说明为什么容易混。
- 先给大纲，我确认后再出题。

拿回来核什么 页码逐题翻回原文；它引的那一句是不是真在那一页；选项里有没有原文没说过的说法。实测那次五题中一题页码错了一页。

常见的坑 发过去的是链接不是文件时，页码是它估的，思考过程里就写着"需要估算页码"。能下载的先存成 PDF 再发。

② 做课件：大纲变成十页

什么时候用 手里已经有一份大纲或讲义，想先要一版能编辑的课件初稿，再自己调。

怎么填 对象和目标两行最要紧，不写它就照着写大纲的人的口气做页面；【做到什么】写学生下课当场能做的动作。例如：按这十条大纲做 10 页《西方经济学》课件，听的是大二会计专业的学生，我要他们下课能用机会成本解释一次自己最近的选择。

本课实测原文（Kimi PPT，2026 年 9 月 4 日）：

按这份大纲生成 10 页教学 PPT，每页一句标题、三条要点，学术风格，白底。大纲：【十条大纲】

补全条件的版本：

按下面的大纲生成【10】页教学课件。对象是【年级、专业】学生，目标是让他们能【做到什么】。

每页一句标题、三条要点；凡引用数字，请在页脚标注来源与页码；不用公式。

大纲：【逐条列出】

先给每页标题，我确认后再生成全部内容。

拿回来核什么 同一个数字在大纲和页面上一致不一致，实测那次大纲写 41.0、页面写 41.8；导出的文件能不能编辑；上台前每页读一遍，看是不是学生听得懂的话。

常见的坑 它把给研究者写的大纲直接切成十页，术语一个没换。先让它把要讲的那几条改成学生能进入的说法，再让它排页面。

③ 做图：一句话出示意图

什么时候用 课件缺一张帮着讨论的图，自己画太慢。要求数据准确的图表不用它画。

怎么填 先想清楚这张图要让学生看出什么，再写左中右各是什么，最后才是白底、黑线这些。例如：《高等数学》那一张，左侧是曲线上的两个点，中间是割线随两点靠近转成切线，右侧是这一点的导数值，白底黑线那几项照抄。

本课实测原文（即梦 AI 图片 5.0 Pro，2026 年 9 月 4 日）：

一张《人工智能通识》课的板书风格示意图：白底，黑色线条，左侧一句话进入模型，中间一层"注意力"把词与词连线，右侧输出下一个词，含中文标注，无装饰，教学插图

改成自己的课：

一张《【课程名】》课的板书风格示意图：白底，黑色线条，【左侧是什么，中间是什么，右侧是什么】，含中文标注，无装饰，教学插图。

拿回来核什么 图上每一个字读一遍。实测那次图里写出“预测结果：明天”，语义不通。错了就重出，不要在图上补字。

常见的坑 气氛插图它做得稳，一到每个字都要对的示意图就容易写错。先让它把图上会出现的文字按顺序列给您看，确认了再出图。

④ 做视频：十秒短片

什么时候用 课堂开场想放几秒画面，把一个只能靠说的过程变成看得见的东西。

怎么填 一句话说清主体、过程、机位、质感、有没有文字，少一样它就自己替您定。例如：生成一段 10 秒视频，一根低碳钢试件受拉，先均匀伸长，再出现颈缩直至断裂，侧视，固定机位，实拍质感，无文字。

本课实测原文（即梦 Seedance 2.5，2026 年 9 月 4 日）：

生成一段 10 秒延时视频：一粒种子在土壤中发芽，侧视剖面，先向下扎根再向上出芽，实拍质感，固定机位，无文字

拿回来核什么 看一遍原片，读一下分辨率和时长，看动作顺序对不对。这段画面是生成的，不是某次真实实验的记录，讲的时候别拿它当证据。

常见的坑 超过十五秒、要人物或者要连贯情节，动作就开始乱。拆成几段十秒的分别生成，每段只放一个主体。

⑤ 口语陪练

什么时候用 学生练口语、练面试问答，或者您想给一门课配一个一问一答、答完就纠错的练习。

怎么填 【课程】填课名，【场景】填学生真会碰到的场合，场景越具体它出的问题越像样。例如：你是《大学英语》的口语陪练，用留学面试场景向我提问，一次一个问题。

本课实测原文（Kimi K3，2026 年 9 月 4 日，文字对话）：

你是我的英语口语陪练，用大学面试场景问我三个问题，每答完一个纠正我的发音和语法。

改成学生用的版本：

你是【课程】的口语陪练。用【场景】向我提问，一次一个问题。我每答完一个，先指出语法与用词错误并给出更自然的说法，再给下一个问题。不要一次给出全部问题。

拿回来核什么 发给学生之前，自己拿一段故意写错几处的答案试一遍，看它是不是逐条指出来了、改写读着自不自然。实测那次是文字对话，四处纠正全对；实时语音、可打断那一路没有实测，课上不演示。

常见的坑 它一口气把三个问题全列出来，练习就成了做题。回一句“一次只问一个，我答完你再问下一个”。

⑥ 深度研究：一份带引用的综述

什么时候用 碰上一个不熟的问题，先要一张材料地图，知道有哪些文献、大家在争什么。进入材料还得自己走。

怎么填 【主题】写成一句能判对错的话；范围那一栏不填，它会给全领域。末尾三条是同一件事的三种问法，要比较两样东西用“对比”，只要最近的进展用“动态”。例如：就“翻转课堂对大一《高等数学》期末成绩的影响”做一份综

述，范围限定在国内高校、2021 年以后。

本课实测原文（Kimi 深度研究，2026 年 9 月 4 日）：

就"【主题】"做一份研究综述：分歧点、近三年进展、研究空白，每条结论附可点击来源。

补全条件的版本：

就"【主题】"做一份研究综述，范围限定在【学科 / 地区 / 年份】。分三部分：分歧点、近三年进展、研究空白。每条结论附可点击的一手来源（论文、官方文件），不用自媒体。最后单列一段：这些引用里有没有已被更正或撤稿的，若有请标出。

同一条提示词换三种问法：

综述：就"【主题】"做一份研究综述，范围限定在【学科 / 地区 / 年份】，分三部分：分歧点、近三年进展、研究空白，每条结论附可点击的一手来源。
对比：比较【A】与【B】在【场景】下的差异，按【我关心的三到五个方面】列表，每个判断附来源；最后一段单列"资料里没有说清楚的地方"。
动态：汇总【主题】近【时间段】的进展，只收官方发布、论文与署名报道，按时间排列，每条一句话加来源；不收自媒体转述。

拿回来核什么 至少点开两条引用，看那句话是不是真在链接里；再看有没有更正或撤稿。实测那次抽查两条都为真，其中一篇发表后有正式更正，只读摘要看不出来。

常见的坑 链接打得开，题名却对不上，打开是一篇转述。把它给的题名单独搜一次，搜不到就按"没找到"处理。

⑦ 数据：一张表看分布

什么时候用 手里一张成绩表或一份问卷结果，想先看清分布再定下一步怎么教。均分是起点，不是结论。

怎么填 【表格名】说明这份表是什么；【我关心的问题】写成一句具体的问法；【需要的统计量】写清要哪几个数。例如：这是《思想道德与法治》三个班的期末成绩（已匿名），我想看三个班的分布差在哪，要均分、标准差，并标出异常值。

本课实测原文（Kimi K3，2026 年 9 月 4 日）：

期末成绩（已匿名）。A 班：【15 个分数】。B 班：【15 个分数】。C 班：【15 个分数】。请：① 按班级画分布图；② 标出异常值并说明判断标准；③ 给出每个班的均分与标准差。

补全条件的版本：

这是【表格名】（已匿名）。先说明每一列代表什么，检查有没有明显缺项或录入错误，再回答：
① 用一张合适的图展示【我关心的问题】；
② 标出异常值并说明判断标准；
③ 给出【需要的统计量】。
不要推断表里没有的原因。

拿回来核什么 统计量自己用 Excel 或计算器复算一次，实测那次三个班六个数全部一致。比较之前先问三件事，同一份试卷吗，分制一样吗，缺考是空着还是按零分录的。

常见的坑 它标出一个异常值，顺口给一个表里没有的原因，比如说这名学生缺考。原因不采信，先回去看录入有没有错。

⑧ 邮件：分档与起草

什么时候用 一批邮件积着没处理，先分出哪些今天必须回，再逐封起草。发送那一步留给人点。

怎么填 【一周】按您要处理的范围改。最后那句关于姓名与单位的要求不要删，摘要里带着人名就不方便存档。例如：读我最近三天的收件箱，按紧急度分三档。

读我最近【一周】的收件箱，按紧急度分三档：截止今天、本周、可稍后。每封给一句建议回复，只保存为草稿，不要发送。姓名与单位不要写进摘要。

拿回来核什么 数两个数，草稿箱几封、已发送几封，实测那次是草稿 3、已发送 0。再逐封看，草稿有没有替您宣布已经做完某件事，承诺的时间跟您的日程对不对得上。

常见的坑 草稿里写着“材料已提交”，可您还没交。凡是出现完成时的句子，逐句对一遍再发。

⑨ 纪要：录音变纪要

什么时候用 一场会开完，录音转成了文字，您要的是决定和待办，不是一份长摘要。

怎么填 这条整段发，把转写稿或录音附上，四项一项都不要省。例如：把《材料力学》教研组这次备课会的录音转成纪要，只要这四项。

把这段录音转成纪要，只要四项：决定了什么、谁负责、什么时间、还有什么没定。每项用一行，没有的写“无”。

拿回来核什么 待办自己先列一遍，再拿它的摘要对照，不要反过来。实测那次一场 1 小时 27 分的会，摘要 2,536 字给得出，待办抽到 0 条。

常见的坑 录音里没点名，它补一个看着合理的负责人。没说清的写“未定”，或者单列成一条要确认的问题。

⑩ 做小工具：一句话做一个课堂练习网页

什么时候用 想给课堂做一个当场能答、能出分的小练习页，自己不会写网页。

怎么填 【年级、专业】【主题】【N】照实填；最后一句限定教材范围，它才不会出到学生还没学的地方。例如：做一个给大二机械专业学生用的课堂练习网页，主题是“应力与应变的关系”，5 道单选题，题目只用第 2 章讲过的内容。

本课实测原文（Kimi K3，2026 年 9 月 6 日，约两分钟交付，本地打开可用）：

做一个给大一非计算机专业学生用的课堂练习网页，主题是“AI 是怎么读懂一句话的”。要求：5 道单选题，每题四个选项，答完显示得分，并逐题给出解析；单文件 HTML，白底，中文，不用外部库，直接给完整代码。

改成自己的课：

做一个给【年级、专业】学生用的课堂练习网页，主题是“【主题】”。要求：【N】道单选题，每题四个选项，答完显示得分，并逐题给出解析；单文件 HTML，白底，中文，不用外部库，直接给完整代码。题目只用【教材范围 / 已学内容】。

拿回来核什么 双击打开能不能答、能不能出分；每道题和解析逐条读，实测那次五题都对；代码里有没有外部链接。发给学生前把题换成自己的。

常见的坑 它自己补条件，实测那次页面上多了一句“建议用时 8 分钟”，提示词里并没有。看到就删掉，或者改成自己的数。

⑩ 补 · 替您操作：批量处理文件

什么时候用 一学期攒下的作业或讲义文件名乱了，要按一个规则统一改名。只在可重复、可核对、可撤回的动作上用它，发送、付款、删除留给别人点。

怎么填 【文件夹】填您电脑上真实存在的那个目录，先复制一份副本再动手；【日期】【课程】按您以后想怎么找来定。例如：把"桌面/西方经济学 2026 秋/作业"里的文件按"20260915-西方经济学-原名"改名。

把【文件夹】里的文件按"【日期】-【课程】-【原名】"改名。先列出改名前后的对照表给我看，我确认后再执行；执行后再列一遍结果。

拿回来核什么 先看对照表，重名的、名字读不出的它有没有单独列出来；执行完打开目录看一眼，文件数对不对，有没有动到不该动的文件。

常见的坑 读不清的文件它猜着改了名。在要求里加一句"读不清的不要猜，单独列出来我来定"，改完再执行。

9.2 教学专用提示词

教案初稿（本课实测原文，2026 年 9 月 4 日）

什么时候用 一节课的教案从零开始，主题、对象、时长都已经定了，先要一份能拿来挑毛病的稿子。

怎么填 换掉分钟数、对象、主题和后面三个数量，其余照发。例如：为一节 90 分钟的《材料力学》习题课出教案，对象是大二机械专业学生，主题是"梁的弯曲正应力"，要 2 个目标、1 个互动、1 个作业。

你是课程设计师。请为一节 45 分钟的课出一份教案：对象是大一非计算机专业学生，主题是"AI 是怎么读懂一句话的"。要求：恰好 3 个学习目标、1 个课堂互动、1 个课后作业。直接给教案，不要先问我问题。

拿回来核什么 各环节分钟相加是不是等于总时长；学习目标里的动词当场能不能查（说出、列出、算出算，了解不算）；它选的例子跟您导言里已经用过的有没有撞车。

常见的坑 栏目齐全、时长也对，读着却像给任何一位教师都能发的那一份。2026 年 9 月 4 日那次交回约 2,400 字，开头的例子正好和这门课的导言重了。把上次课讲过什么、这批学生卡在哪里补进去再要一稿。

教案第二稿的修改要求（本课实测原文，2026 年 9 月 5 日）

什么时候用 初稿读完，您看出一处影响学习的问题，要它改而不是重写。在同一个会话里接着发这一条。

怎么填 先说保留什么，再说希望学生多做哪一步，最后给约束（时长不变）和只要什么（流程表）。不写"再好一点"。例如：保留课上那道悬臂梁例题，缩短公式推导，让学生两人一组各算一个截面的应力并写出依据，保持 90 分钟，只要新流程表和时间怎么挪的说明。

保留苹果的例子，缩短切块和数字表示的讲解。让学生两人一组，自己选一个可能有不同意思的词，各造一句话，圈出帮助判断含义的信息。最后换一组新句子，每个人独立说明理由。按这个目标重排活动，保持 45 分钟，并说明哪部分时间转给了学生。只给新的教学流程表和一段时间变化说明。

拿回来核什么 先核数字，新流程各段相加等于原定时长，转出的分钟与新增的分钟相等（9 月 5 日那次两边都是 28 分钟，学生动手的时间从 5 分钟变成 24 分钟）。数字对了再看改动是不是朝您写的那个目标去的。

常见的坑 说了只要流程表，它把整份教案重写一遍发回来。不要顺手读全文，回一句只要流程表 and 变化说明，让它重发，否则您又替它核了一遍不打算用的内容。

课堂活动设计

什么时候用 课上要插一个学生动手的环节，您知道要他们做到什么，还没想好怎么做。

怎么填 【做到什么】写成能当场检查的动作，写不出动作就先别发。例如：为大一思政课设计一个 8 分钟的活动，让学生能用一件自己经历过的事说明"实践是检验真理的唯一标准"，每人写一张便签交上来。

为【年级、专业】学生设计一个【5】分钟的课堂活动，目标是让学生能【做到什么】。要求：每个学生都要产出一个能检查的东西（一句话、一个选择、一张便签）；给出教师的三句引导语；说明这个活动可能在哪里失败。

拿回来核什么 每个学生手里是不是真有一件东西交上来，还是只有小组代表发言；"可能在哪里失败"说的是您这个班会出的事，还是泛泛的学生不积极。

常见的坑 它爱自己加分组人数、计分办法和奖励，这些您没要。删掉再看剩下的活动本身还成不成立。

给学生的针对性反馈

什么时候用 一叠作业里挑几份要写评语，自己先看过一遍，想要一份评语初稿。分数仍由您定，它只写话。

怎么填 姓名学号先去掉，评分标准附原件。例如：这是《大学英语》写作课一份匿名作业和评分表，评分表四栏，内容、结构、语言、规范。

这是【课程】的一份学生作业（已匿名）和评分标准。请写三条反馈：一条指出做对的地方并说明为什么对；两条指出最重要的问题，各附一个改法。不给分数，不改写学生的原文。

拿回来核什么 它指出的问题在原文里真能找到那一处；给的改法是这学期学过的方法，不是学生还没学到的做法；有没有背着您把学生的句子改掉。

常见的坑 换一份作业，三条评语几乎照搬上一份。发现两份撞车，把这份作业最特别的地方自己写一句发给它，再要一版。

试题变式与答案核对

什么时候用 期中期末要出 AB 卷，或者一道题讲完想再给学生一道同考点的练习。

怎么填 教材那几页附成文件，不要只写章节名，页码才靠得住。例如：这是《高等数学》一道求极限的题和标准答案，教材是同济第七版第 1 章，PDF 已附上。

这是一道【课程】的题目和标准答案。请：① 出两道同一考点、不同情境的变式；② 给出每道的答案与依据；③ 指出学生最可能犯的一个错误。答案依据必须能在【教材 / 材料】里找到，标出页码。

拿回来核什么 两道变式自己做一遍；标的页码逐个回原文翻，9.1 ① 那次实测就在五题里错了一处页码；它挑出的那个易错点，您在这个班上真见过没有。

常见的坑 情境换了，考点也悄悄换了，求极限换成了求导数。对不上考点的那一道直接删，不要在它的基础上修。

通知与公文草稿

什么时候用 调课、补考、教研活动、材料提交，要发给学生或同事的短通知。

怎么填 【事项】把已经定下的信息一次写全，没定的留着让它写【待填】。例如：写一则给《人工智能通识》选课学生的通知，第 6 周的课改到明理楼 302，上课时间不变。

写一则给【对象】的通知，说明【事项】。包含：时间、地点、要做什么、截止、联系人。缺少的信息留【待填】，不要编。语气正式，不超过【200】字。

拿回来核什么 时间、地点、截止、联系人四样在不在，【待填】还剩几处；字数；有没有替您定下一个您没说过的截止日期。

常见的坑 新教室学生多半没去过，它只写"教室变更"四个字。补上楼名、楼层和从原教室走过去大概几分钟，学生才不会走错。

答疑助教的手脚架（参考 Kestin 等 2025 的做法）

什么时候用 学生课后反复问同一类题，想给一门课配一个不直接给答案的答疑对话。Kestin 那项研究里效果好的那一组，用的也是教师事先写好的分步提示和护栏，不是把对话框直接交给学生。

怎么填 【粘贴材料】填学生真正在看的东西，讲义、例题和常错的解法，材料不给它就按自己的理解讲。例如：你是《西方经济学》的答疑助教，材料是第 4 章讲义和课后八道题的解析。

你是【课程】的答疑助教。学生提问时：先复述问题确认理解；不直接给答案，先问一个引导性问题；学生答后再给下一步提示；连续三次仍不会时，给出完整解答并标出关键步骤。不讨论与本课无关的话题。本课的材料如下：【粘贴材料】

拿回来核什么 自己扮学生跑两轮，其中一轮故意答错，看它是先问还是直接讲；连问三次之后它给的完整解答，对着您的讲义核一遍步骤。

常见的坑 学生说一句"直接给答案"，它就给了，前面的规则等于没写。发给学生之前在规则里加一条："我要求直接看答案时，先给一个提示，我第二次要求再给答案。"

评价分级说明与 AI 使用声明（给学生的两段话，可直接贴进作业要求）

什么时候用 布置一次作业，要在作业要求里写清这次允许用到哪一步。第一段定级别和要交的材料，第二段是学生自己填的声明。

怎么填 五个级别留一个，其余删掉；级别出自 AIAS 量表（Perkins 等四位研究者 2023 年 12 月提出，2024 年 4 月修订）。例如：《学术写作》的文献综述定 3 级，学生随稿附一份没有 AI 内容的原稿；期末项目报告定 4 级。

本次作业允许使用 AI 的级别：【1 不用 / 2 帮构思，终稿不得有 AI 内容 / 3 帮修改，附无 AI 原稿 / 4 它做你评，AI 内容须标注 / 5 全程用】。

请随作业提交：原材料、任务说明、第一版、你的修改要求、第二版，五样缺一不可。

AI 使用声明：本次作业我使用了【软件名称与版本】，用在【哪一步：构思 / 查资料 / 改句 / 生成图表】。以下内容由 AI 生成并经我核对：【列出】。其余为本人完成。

拿回来核什么 收上来的五样材料齐不齐，提示词和两版稿子能不能对上；声明里写的软件版本和用途，跟他交的修改记录对得上对不上。抽两份当面问一句为什么这样改，装不出来。

常见的坑 一门课所有作业都用同一个级别，学生读一遍就不再看那一行。课堂小测和项目报告分开定，级别才起作用。

9.3 三份岗位说明书（可粘贴到产品的"自定义指令 / 系统提示 / 项目说明"）

备课助手

什么时候用 备课、出题、改课件都用得上。这一份管长期不变的要求，每个新会话先带上它，省得每次重新交代自己是谁、学生是谁。

怎么填 改三处，学科与职称、学生是谁、交付格式；常用的格式写在最后一行，改完自己念一遍。例如：我是《材料力学》的讲师，学生是大二机械专业本科生，没有有限元基础；教案按目标、流程、活动、作业四段给，流程表带分

钟。

你是我的备课助手。我是【学科】教师，学生是【年级、专业】，没有【编程 / 数学】基础。
规则：概念先用例子解释再命名；不用公式和术语堆砌；每次先给流程或小样，我确认后再展开；引用数字必须标来源与页码；缺关键条件先问我，条件够就直接做。
交付格式：教案用"目标、流程（含分钟）、活动、作业"四段。

拿回来核什么 新会话里先问一句这次给谁上课，它答不上来，就是产品根本没读这份说明书；交回来的教案是不是按您定的那几段排，引的数字有没有标来源和页码。

常见的坑 规则写成十几条抽象要求，它一条也守不住。先留五六条最要紧的，用上两周，把您每次都得重说的那句再加进去。

科研助手

什么时候用 查文献、找一手来源、核一个数据的出处时换上这一份。它和备课助手分开放，两边的规矩凑在一起会打架。

怎么填 改两处，学科领域和您认的来源范围；职称在这一份里写不写都行，它看的是来源不是身份。例如：领域是高等教育学，只认中文核心期刊、教育部文件和能下载的 PDF，学位论文单独标出来。

你是我的科研助手，领域是【学科】。
规则：每条结论附一手来源（论文、官方文件），标题、作者、年份、可点击链接四项齐全；不引用自媒体；不确定的地方写"未核实"；主动标出可能已被更正或撤稿的引用。
结论少给，材料给全，每一条我都要能顺着来源追下去。

拿回来核什么 随手点开两个链接，看里面真有那句话，还是只是同一个主题；标了"未核实"的几条自己搜一次；题名和年份对不上的，当它没找到处理。

常见的坑 链接打得开，打开是一篇转述文章。让它回到原始出处重给一条，回溯不到就不引这一条。

行政助手

什么时候用 写通知、整理会议纪要这类事，凡是一不小心就会发出去的，先让它戴上这份规矩。

怎么填 这一份基本不用改，把您单位的固定格式加在末尾。例如：通知落款固定写"教务处 2026 年 9 月"，纪要按决定、负责人、时间、待定四项列。

你是我的行政助手。
规则：邮件与通知只写到草稿，不发送；涉及姓名、单位、金额的内容不进入摘要；缺少的日期、地点、联系人留【待填】；每次交付先列做了什么、没做什么。
发送、付款、删除三类动作，永远交回给我。

拿回来核什么 它列的"做了什么、没做什么"两栏，先看没做的那一栏；摘要里有没有混进学生姓名、学号、金额。

常见的坑 对话长了它会把该留空的【待填】自己补成一个具体日期。看到就把规则重发一次，并说清哪几处必须空着等我填。

第四份：教师个人 AI 档案（五个文件，写它在为谁工作）

岗位说明书写的是它扮演谁，个人档案写的是它在为谁工作。写一次，此后每个新会话都从这里开始。对话产品里把五个文件粘到"自定义指令"或项目说明，工作台与编程工具里放进工作目录，它会先读。五个文件各一页以内够用：

文件	写什么
01 关于我	身份与学科，这学期在做什么，AI 帮我时需要知道的基本背景
02 我的学生与课程	对象是谁，常见困难在哪里，上过什么、没上过什么
03 我的目标	这学期与这一年想做成的事，优先级与大致时间
04 我的表达偏好	喜欢的长度、语气、格式；不喜欢的词语和写法；什么时候要它主动提醒
05 我的原则与禁区	它不能替我定的事；不能编的内容；不能出门的材料；涉及学生数据时的要求

采访式填写的提示词（可直接复制）：

什么时候用 头一次给自己配 AI，拿一个下午一次做完。只有一小时，就先做 01 和 02 两个文件，其余下次补。

怎么填 这一条不用填，整段发出去，它会一个一个问您。答的时候用平时说话的样子，别自己先书面化。例如：问到表达偏好时直接说，我不要每段末尾都来一句总结的写法，教案给我流程表就行。

请用采访的方式，帮我完成"个人 AI 档案"的五个文件：01 关于我、02 我的学生与课程、03 我的目标、04 我的表达偏好、05 我的原则与禁区。

规则：

1. 每次只问一个问题，问题短而具体，按上面的顺序问。

2. 我答得模糊就追问，不要替我猜，也不要替我补条件。

3. 不收集证件号、账号、密码、住址、健康记录等敏感信息。

4. 保留我的原话和习惯说法，只做必要整理，不要包装成正式文书。

5. 每完成一个文件，先把准备写入的内容给我看，我确认后再写入，再进入下一个文件。

6. 全部完成后告诉我：哪些信息还薄弱，哪些需要定期更新，哪些前后有矛盾需要我确认。

现在从"01 关于我"的第一个问题开始。

拿回来核什么 写入之前它给您过目的那一版，逐句读一遍，看它有没有替您补条件，有没有把您的口语改成套话。五个文件都要您点过头才写入。

常见的坑 您含糊说一句学生基础一般，它写成"学生数学基础薄弱，畏难情绪明显"，还当成您说的。看到这种回一句我没这么说，让它按原话改。

写作说明（去 AI 味，贴进任何岗位说明书末尾）

什么时候用 它交回来的东西读着对，就是不像人写的，每段都收在一句漂亮话上。给学生看的、给同事看的、要印出来的，都贴上这一段。

怎么填 不用填，整段贴在上面那几份说明书的规则后面，另起一段。例如：贴在"备课助手"的交付格式那一行之后。

写作要求：先给例子，再给术语。句子要短。写得像一个人在说话，语气克制，不确定的地方直说。段末不加金句，判断也不用冒号引出来。不写三个一组的排比。"不是.....而是....."这个句式不用，"赋能""助力""全方位""重塑"这类词也不用。

拿回来核什么 把每段最后一句单独抄成一行念一遍，句式全一样就是没改到；再数一数那几个点名的词还剩几处。

常见的坑 头两轮它听话，第三轮又开始给每段末尾加总结陈词。发现一次就把这段原样再贴一遍，比在对话里责怪它管用。

9.4 两份模板与两张清单

两份模板在交代任务和提修改时用，两张清单在结果拿回来之后用。

任务说明模板

什么时候用 第一次把一件事交给它的时候。手头有材料、有对象、有期限，直接说一句"帮我写个教案"，它就按最常见的写法交一篇，返工花的时间比填这张模板长。

怎么填 【】里填的是交代任务的五个条件，对象、目标、材料、交付、这一步做到哪。最要紧的是【一个小样或第一部分】，第一次别让它写全文。例如：我想完成一节 45 分钟的《人工智能通识》导论课教案，给大一非计算机专业的学生用，目的是让他们能用两个例子解释"同一个词要结合上下文理解"；材料是教材第 1 章和上次课讨论过的搜索引擎排序；已有条件是这节课在机房上，没有分组桌；请先给教学流程表，不要全文。

我想完成【具体任务】，给【使用者】使用。目的是让对方能够【解释、判断或完成什么】。
这次依据的材料是【粘贴或附上材料】。已有条件是【背景、时间、限制】。
请先给出【一个小样或第一部分】。缺少影响方案的关键条件时先问我。条件充分时先做小样。交付前对照以上条件自查一遍，把没做到的列出来。

拿回来核什么 用下面那张交付前检查表逐条过，那张表就是核这一版结果用的。先看它有没有替您补条件，再看交回来的东西写给的是不是您说的那个对象。

常见的坑 条件缺了它不问，自己补一个接着做，比如您没写字数，它按 150 字写。"缺少影响方案的关键条件时先问我"这一句别删掉；结果回来先翻它的思考过程，看补了什么。

修改要求模板

什么时候用 读完第一版，看出某一处不对，想让它改而不是自己重写的时候。一次只提一处，三处一起发，它多半只认真改第一处。

怎么填 三个【】分别是哪里、为什么不行、往哪改，缺一样它就自己猜。例如：三层讲解占 32 分钟这一处不合适，因为目标要求学生自己完成一次解释；请改成两人一组造句并圈出判断依据，总时长仍是 45 分钟。

这一处【原句或具体位置】不适合，因为【对象、事实、目标或条件】。请按【明确方向】修改，并说明主要改动。

拿回来核什么 先核数字，再看它说的"主要改动"跟您要的是不是同一件事。您没点到的地方有没有跟着变，两版并排读一遍就看得出来。

常见的坑 它顺手把旁边也改了，改到第三轮已经不是您那份稿子。发现一次就在下一条修改要求后面加一句"只改我指到的地方，别处不动"。

五类错误与核对方法

错误类型	表现	本课实例	核对方法
猜位置	内容读到了，页码靠估	第 5 题标第 9 页，实为第 8 页	回原文按页翻，核不上时再翻一遍
来源冲突	两个数字都有据，没说选了哪个	大纲写 41.0，页面写 41.8	看它引的是哪一处，自己定口径
漏项	摘要给得出，待办抽不出	一小时二十七分钟会议待办 0 条	待办自己列，用摘要对照
似是而非	结构对、标注对，细节不通	示意图里"预测结果：明天"	每个字读一遍
自补条件	条件缺失时自行补上	字数限制丢失，它按 150 字写	看它的思考里替您补了什么

交付前检查表

什么时候用 每一版结果拿回来、决定用不用之前过一遍。要发给学生、要交上去的材料尤其别跳过这一步，发出去之后再发现数字不对，改起来比现在看一遍费事得多。

怎么用 六项按顺序对，全打上勾再用。打不上的那一项不要绕，回到任务说明把缺的那句补上重做一遍。第一行那五项，对的就是任务说明模板里填的那几个空。

检查项	打勾
对象、目标、材料、交付、这一步做到哪，五项都写了	
数字回到原文核过一次	
引用至少点开两条	
它替我补的条件我看到了	
发送、付款、删除留给自己	
原材料、任务说明、第一版、修改要求、第二版，五样存下了	

五条小测试集（第2周用）

什么时候用 一条提示词用着还行，打算存下来反复用之前。一次跑对说明不了什么，换五份不同的材料各跑一遍，才知道它稳不稳。

怎么用 五行填五份不同的材料，同一条提示词一字不改地发，每次记下结果能不能直接用、要改在哪里。五次里四次能直接用，这条提示词才进模板；不到四次，看"要改在哪里"那一系列反复出现的是哪一条，把它写进提示词再跑五次。

#	材料或问题	结果能否直接用	要改在哪里	备注
1				
2				
3				
4				
5				

9.5 三个 Skill 文件

支持 Skill 或自定义技能的产品，把下面的文件整段粘贴进去；只有普通对话框的，把"步骤"那一段放在提示词开头。每份文件里的"适用""输入""检查"是给它看的，前后的备注是给您看的。

Skill 一：备课三步走

什么时候用 每学期都要备、一节一节往下走的那门课，每节都得起初稿改到能上。第一次开一门全新的课不适合用它，那时候讲什么您自己还没定下来。

怎么填 "输入"那一行的六样在开头一次给全，材料给文件不给链接。例如：《无机化学》"化学平衡"那一节，大一化工专业，50 分钟，教材第 5 章的 PDF 已附上，上次课讲过反应速率，这节的目标是学生能判断给定条件下平衡往哪边移动。

名称：备课三步走

适用：为一节课准备教案，从初稿改到能用

输入：课程名、对象、时长、材料（教材片段或大纲）、上次课讲过什么、本节目标

步骤：

1. 先复述：这次课给谁上、重点是什么、不能偏到哪里。复述不对，先纠正。
2. 出初稿：目标、流程（含分钟，合计等于时长）、一个活动、一个作业。
3. 等我给出修改要求（保留什么、希望学生多做哪一步、约束）后，只交回新的流程表和一段时间变化说明。

检查：各环节分钟相加等于时长；学生自己动手的时间是否达到我的要求；活动是否每个学生都有产出。

输出：教案（四段）；第二稿流程表；时间变化说明。

常见的坑 第 3 步说好只交回流程表和时间变化说明，它常常连教案全文一起重写一遍。不要在那一版上接着改，只留流程表，否则前面确认过的目标和作业会被它顺手换掉。

Skill 二：题库出题与核页

什么时候用 从一章材料出一批题，而且题目要标页码，给学生或同事查得到。只要一两道题，直接问匹配这份文件快。

怎么填 "输入"里的材料用 PDF 文件，别给网页链接，链接进去它标的页码多半是估的。例如：《统计学》教材第 6 章的 PDF，大二会计专业，10 道题，单选和简答各一半。

名称：题库出题与核页

适用：从一章材料出题，并保证页码可核

输入：材料（文件或链接）、对象、题量、题型

步骤：

1. 列三级大纲，每级标页码。
2. 每题标页码，并引用原文中支持答案的那一句。
3. 自查：引用句是否在所标页码；选项是否有原文没有的说法。

检查：我抽查两题回原文；页码不对的题重做。

输出：大纲；题目、答案、页码、引用句。

常见的坑 它引的原句确实在书上，却不是支持答案的那一句，读着像回事，回原文才看得出来。抽查的时候连引用句一起对，别只对页码。

Skill 三：引用核查

什么时候用 一份综述、论文或申报书的参考文献，二三十条以上，自己一条条点开要花一下午。学生作业里的引用也可以这么核。

怎么填 "输入"按题名、作者、年份、链接四样给全，手头没有链接的条目照样列进去，让它去找。例如：把论文末尾的参考文献整段复制过来，一条一行，共 28 条。

名称：引用核查

适用：检查一份综述或论文初稿里的引用

输入：引用列表（题名、作者、年份、链接）

步骤：

1. 逐条确认题名、作者、年份、期刊或来源四项是否一致。

2. 标出无法打开、信息不全、可能已更正或撤稿的条目。

3. 输出一张表：条目、状态（一致 / 存疑 / 无法打开 / 有更正）、说明。

检查：我至少点开两条"一致"的条目。

输出：核查表。

常见的坑 找不到的条目，它有时不写"无法打开"，而是照着题名补一条看着齐整的信息。状态写"一致"的条目里，年份特别新的那几条把题名单独搜一次最稳。

9.6 一页纸速查卡

可以打印后放在电脑旁。交代任务之前对着第一栏点一遍五个条件写全没有；结果拿回来、觉得哪里不对又说不上来的时候，翻第二栏按错误类型找核法。

五个条件（交代任务时写全）	五类错误与核法	两条红线	四个动作（一轮的顺序）
对象	猜位置：回原文翻	学生数据、未发表研究不出门	材料与任务
目标	来源冲突：自己定口径	发送、付款、删除由人确认	初步产物
材料	漏项：自己列一遍		人的判断
交付（结果要什么形式）	似是而非：每个字读		下一次修改
这一步做到哪	自补条件：看它补了什么		

先试的一件事，看三条就够。材料在手，结果可核，出错代价小。

交不交，三问	铺轨道（把流程定好），三步	想清楚，三句
我自己做要多久	拆：把每周重复的一件事拆成小步骤	问它之前，先写下自己的判断
它做对的概率多高	分：每步标全交、协作、自己	拿到答案，先问怎么验证
从交代到核完总共多久	铺：先给全交的步骤写任务说明与核对清单	再问有没有反面证据

9.7 更多场景的模板，在工具箱文件里

还有十九条模板没有印在这里：先让它问够了再动手、改完给一张改动对照表、一篇文献的精读卡、一个概念五个学生手边的例子、课题计划分两次出，这一类。每条同样带四行备注。它们在随手册提供的工具箱文件包里，文件名是 07-更多场景的提示词与 workflow 模板，打开就能复制。手册不重印这一节，是为了让您能把它塞进包里。

修行清单：一周做三条

师傅领进门，修行在个人。今晚一件事：把今天满意的那次对话存成模板。一周三条：从下面十行里勾三条，本周做完。一个月一件真事：见第八部分的四周计划。

#	场景	第一步提示词（本课实测原文，缩写）	本课用的工具	结论
01	读长文档	请读这篇 PDF：列三级大纲；出 5 道选择题并标页码；指出两个易混概念	Kimi K3	需核对
02	做 PPT	按这份大纲生成 10 页教学 PPT，每页一句标题、三条要点，学术风格，白底	Kimi PPT	需核对
03	做图	一张《____》课的板书风格示意图：白底黑线，含中文标注，无装饰	即梦 5.0 Pro	需核对
04	做视频	生成一段 10 秒延时视频：一粒种子发芽，侧视剖面，固定机位，无文字	即梦 Seedance 2.5	试用
05	陪练	你是我的英语口语陪练，用大学面试场景问我三个问题，每答完一个纠正	Kimi K3	可用
06	深度研究	就"____"做一份研究综述：分歧点、近三年进展、研究空白，每条附来源	Kimi 深度研究	需核对
07	数据	期末成绩（已匿名）三个班：按班级画分布图；标异常值；给均分与标准差	Kimi K3, Python	可用
08	邮件	读最近一周收件箱，按紧急度分三档，每封一句建议回复，只存草稿	飞书邮箱 CLI	需核对
09	纪要	把这段录音转成纪要：决定了什么、谁负责、什么时间、还有什么没定	飞书妙记	需核对
10	做小工具	做一个给学生用的课堂练习网页：对象、主题、5 道单选题、答完显示得分与解析、单文件 HTML	Kimi K3	可用

附

APPENDIX

练习

练习一：把任务说清楚

读一遍就懂

题目。选一件自己熟悉的小任务，用 3.4 节的任务说明模板写一份任务说明，让系统完成一个小样，然后提出一次具体修改。

步骤。第一步，选任务。满足三个条件，材料在手，结果可核，出错代价小。第二步，填模板。把五项条件写进去，对象、目标、材料、交付、这一步做到哪。第三步，发出并读小样。先不改文字，看它是否抓住了对象和目标。第四步，用修改要求模板提一处修改。写明位置、原因、方向。第五步，比较两版，用一句话说清哪条反馈改变了结果。

示范答案。以一份课程通知为例。任务说明可以这样写。我想完成一份下周课程安排变更的通知，给选修《人工智能通识》的大一学生使用。目的是让他们能够知道时间地点的变化并按新安排到场。这次依据的材料是原通知全文和教务处发来的调课单，已有条件是变更只涉及第 6 周的一次课，新教室在另一栋楼，学生多数没有去过。请先给出通知正文，不超过 150 字。缺少影响方案的关键条件时先问我。条件充分时先做小样。拿到小样后，修改要求可以这样写。这一处“请同学们注意教室变更”不适合，因为学生多数没去过新教室，只说变更不说位置会走错。请按“写明楼名、楼层和从原教室步行的大致时间”修改，并说明主要改动。

常见遗漏。第一，没有写使用者，于是系统按通用读者写，术语和语气都不合适。第二，没有写交付的形式和长度，于是对方给了一整篇，而您只要一段。第三，没有写这一步做到哪，于是它先问了五个问题，而您本来希望先做。第四，修改要求只说“不好”，没有说原因和方向，于是第二版改在了别处。第五，没有核它替您补的条件，例如自己定了一个截止时间。

练习二：用学科材料完成一次协作

读一遍就懂

题目。选一段适合的学科材料，请系统生成一项讲解、课堂活动、问题或事务草稿。与一位同事互看，找出一个需要调整的地方，完成修改。结束时用一句话说清“我为什么这样改”。建议保留五项材料，原材料、任务说明、第

一版结果、您的反馈、修改后的结果。没有合适材料时，可以用下面的备用案例。

备用材料。以下情境和数据全部为课堂构造，不对应真实学校。

一所大学拟试行图书馆延长开放。学生代表提出，部分宿舍不便晚间讨论，希望自习区开放到 23:00。图书馆工作人员指出，现行闭馆时间为 22:00，延时需要另排人员。学校准备先在期末前两周试行，并记录入馆人数、座位使用、人员安排与学生反馈。当前没有延时试行的实际使用数据。部分学生希望每天延时，另一些学生更希望改善白天座位管理。

三选一。第一项，给学生设计一个 5 分钟讨论活动，区分材料中的事实、意见和待验证问题。第二项，给负责人写一份 200 字建议，比较两种可行安排，明确还需要什么信息。第三项，给学生写一则通知草稿，清楚说明试行范围和反馈方式；缺少的日期、地点或联系人保留待填项。

判断标准。看四条。有没有编造已发生的数据；是否遗漏重要对象；行动是否具体；是否符合指定用途。

示范答案要点，以第二项为例。一份合格的 200 字建议应当包含四部分。第一，两种可行安排各一句，例如“期末前两周每天延长至 23:00”与“先改善白天座位管理，延时只在考试周试行”。第二，比较的依据，写明各自需要增加的人员安排与预期惠及的学生范围，但因为材料里没有实际数据，这里只能写“需要试行数据支撑”，不能出现具体人数。第三，还需要的信息，至少包括白天座位紧张的具体表现与成因，以及真正会使用延时服务的学生规模与分布。第四，建议的下一步，例如试行期间记录哪几项数据、由谁汇总。核这份建议时，重点看它有没有为了让建议显得完整而补出材料里没有的数字，以及有没有漏掉图书馆工作人员这一方。5.4 节那次实测里，系统在字数条件丢失时自己补了“150 字”，这类自补条件在这个练习里同样可能出现。

附

APPENDIX

术语表

以下解释按这门课的法写，不是教科书定义。

Transformer。2017 年那篇论文提出的模型结构。它的做法是让句子里相隔很远的位置也能建立联系，并且训练可以并行。今天多数对话产品背后的模型都沿用了这一结构。名字听过即可，内部计算这里不展开。

GPT。Generative Pre-trained Transformer 的缩写，生成式、预训练的 Transformer。2018 年 OpenAI 用 Transformer 的解码器一半做成的模型，先读大量文本学会猜下一个字，再做具体的事。ChatGPT 到 GPT-6 沿用这个名字，见第一部分。

注意力。模型处理一个字时，对句子里其他字的参照权重。参照谁、参照多少，由整句话算出来，不是查词典。2.1 节那张热力图是一个真实模型算出的权重。用在课程题目上，它另指人把注意力放在哪里。

Token。模型处理文本的小块，也是很多产品计费与限长的单位。它不固定等于一个汉字或一个单词。本地模型那次生成了 121 个 token，用了十几秒。上传很多材料时，一次能放进去的 token 数有上限。

上下文。模型这一次能用到的全部信息，可以理解为摆在它面前的一张桌子。桌上有您的要求、历史对话、上传的材料，以及工具刚查到的结果。训练时学到的知识不在这张桌子上，两者要分开看。

记忆。产品可能记住您的偏好和历史会话，并在新任务里当作背景使用。本课有一次实测，系统在思考过程里引用了账号过去的会话。记忆不等于它保存了每一份教材，重要条件要直接写进任务。

提示词。您交代任务的那些话，英文是 Prompt。不需要特殊句式，有用的是把对象、目标、材料、交付、做到哪一步写清楚。“互动有趣”过宽，“学生必须先选一个解释，再用句子里的信息说明理由”才具体。

岗位说明书与任务单。提示词两种。前者是长期不变的要求，例如优先面向初学者、概念先用例子解释，很多产品可以单独设置。后者是这一次的对象、时长、材料、交付，每次都写，因为每次都不一样。

工具。模型本身不能读网页、计算或打开文件，平台接入的这些手段叫工具，英文是 Tool。本课三条轨迹分别用了打开网页、运行 Python 和联网搜索。看到系统说“我查到了”，可以看轨迹里是否真有那一步。

Agent。系统根据当前结果选择下一步，持续推进任务，直到交付。中文常译作智能体。它不一定是新的界面，更像一种组织工作的方式。缺材料、缺权限或需要教师判断时，它应当停下并交回。

Skill. 把任务说明、步骤、样例和相关资源打包保存，系统需要时加载，下次就有起点。常译作技能。只有普通对话框的产品，存成一份文档也一样。

开放权重。模型的参数文件公开可以下载，有条件的人能部署到自己的机器上。Llama 2、DeepSeek-R1、Qwen3 都属于这一类。能下载不等于能商用，商用条件看许可证。

许可证。开放权重模型附带的使用条款，规定下载后允许怎样使用，是否可以商用，有什么限制。以各模型官方页面的说明为准，具体条款这里不讲。

本地部署。把模型装在自己的电脑或学校的服务器上运行，输入不必交给远程服务。本课那台 48GB 内存的笔记本断网跑了一个 27B 参数的模型。代价是不知道今天的事、读不了没给的文件、速度看机器。

幻觉，本课称“似是而非”。模型生成的内容结构对、语气对，细节经不起读。本课的例子是示意图里的“预测结果：明天”。模型每一次都在算最可能的答案；答得合乎语境是一种能力，每句都经过事实确认是另一种。核对的方法是逐字读。

核对。把系统给的结果放回材料或事实里检查。不同错误类型核法不同，位置回原文翻，来源看它引哪一处，漏项自己列再对照，细节逐字读，自补条件看思考过程。核的人也会错，核不上时再翻一遍。

交付形式。您要对方给的是一个回答、一件作品，还是一段完成的任务。回答当场能核，作品核每个细节，任务核每一步的状态。交付形式定了，核对方式才定得下来。

初稿。系统第一次交回的结果。初稿通常栏目齐全，能不能用，看教师能不能看出哪一处不合适。本课那份教案初稿一次生成，约 2,400 字，三处判断之后才改到能用。

事实的可信程度。这本手册里的每条事实都按可信程度分了五档，说法随之不同。本机实测过或一手来源打开核过的，直接写结论；会随时间变的，提醒“用之前再查”；二手的，写“据公开报道”；我的经历和判断，标明是我的判断；没实测、没核的，只写“现场演示”或注明没核。课程组另有一份记录，逐条登记来源、核验日期和档次，随课程材料提供。

三种标记。手册每一节标题下的标记，按内容变化的速度分。[读一遍就懂](#) 是三五年不会变的道理，只修表述。

[查一下再用](#) 是模型、产品、价格、最新发布这类会变的信息，这本手册核验到 2026 年 9 月 5 日，使用时以官方页当日说明为准。[动手前看](#) 是本课实际运行并核对过的记录，每次开课重跑，出错的证据随之更新。

AGI。通常译作通用人工智能。OpenAI 章程的定义是高度自主、在多数有经济价值的工作上超过人类的系统，这是一家公司的定义。讨论时先分清定义、判断和证据，机构定义与人物判断都不是共识。

工作流。把一件事拆成步骤，规定每一步谁来做、给它什么、核什么。同一个模型，换个流程，结果可以差很多，见 3.5 节。

认知负债。2025 年 MIT 一项脑电研究提出的说法，指现在把思考交出去省下的脑力，以后要还，见 7.3 节。单一研究的用词，这里借用，不作定论。

AI 使用声明。学生或研究者在提交成果时写明用了什么软件、什么版本、用在哪一步；科技部 2023 年指引对研究者的要求，建议照搬给学生，模板见工具箱 9.2。

技术平权。技术从稀缺资源变成基础设施，像水，像电。7.1 节用它概括做成的七件事的共同点。通行表述，不引出处。

认知。与知识相对。知识是事实、公式、条文、步骤，可以检索、复制、交给系统；认知在这本手册里取窄义，只指定义问题与选择值得做的事。这本手册的说法是，知识可以外包，认知不能；认知里能写清楚的部分，将来也会交出去，见 7.3 与 7.6 节。

行易知难。孙中山《孙文学说》1919 年提出的主张，反《尚书》“非知之艰，行之惟艰”。这本手册借它说 AI 时代的处境，“行”这一半机器接了过去，剩下难的是“知”，见 7.4 节。

附

APPENDIX

常见问题

它会不会拿我的材料训练。 看产品条款。多数对话产品在设置里有是否允许用于训练的选项，用之前看当日的官方说明，不要凭印象。规则说不能出门的材料，学生数据、未发表研究、内部文件，不要贴到公网产品里，走本地模型或先脱敏。本地运行也要知道谁能访问那台电脑、检索和同步是否联网。

学生用它作弊怎么办。 先分清任务。练习时用与考试时用不同，允许怎么用，写进作业要求。让学生保留独立尝试的记录，解释一个关键选择，换一个例子独立完成，或者根据新条件修改自己的答案。5.2 节两项研究的提醒就在这里，练习做得好与离开帮助后学会要分别看。不能只凭文字风格认定学生有没有使用工具。

哪个产品最好。 先问要回答、作品还是任务，再看产品接不接得上您的材料与工具。任何对话产品都能给回答；要作品看能不能改；要任务看有没有工作台。选两个自己能正常使用的入口，给同一份小材料比较，先用熟一个，再按缺的功能加第二个。这本手册不给排名，产品名属于 [查一下再用](#)。

它说的对不对怎么知道。 先辨认错误类型。位置靠估，回原文翻；来源冲突，看它引哪一处，自己定以哪个为准；漏项，自己列再对照；似是而非，逐字读；自补条件，看思考过程。阅读与整理可以直接用，位置、来源、漏项、细节要核。核的人也会错，我们自己就判错过一题。

会不会很快过时。 模型会换，产品会改名，这一类信息用之前要查一下。材料、任务说明、核对、存模板这四步不换，读一遍就懂。您在一个月计划里留下的任务说明、样例和核对清单，换了工具还能用。看到新发布时只问一句，我原来自己接着做的哪一步，现在可以交给它参与。

我不会编程能用吗。 能。本课十八项实测里，多数是在网页对话框里完成的，包括读长文档、做课件、算成绩、深度研究。算成绩那次是系统自己写的 Python，您不需要读懂那段程序，需要的是本地复算或抽查。需要反复处理文件、做小工具时，学一点基本概念有帮助，但可以从真实需求开始。

要不要交钱。 截至 2026 年 9 月，多数对话产品有免费额度，具体数字这里不写，以官网当日说明为准。视频生成那次消耗了 200 积分，生成前有确认弹窗。建议先用免费额度完成第一周的任务，遇到确实缺的功能再考虑付费，不要为了“可能用到”先买。

本地模型要什么机器。 本课那次用的是一台 48GB 内存的 MacBook，Apple Silicon，运行 27B 参数的 6-bit 版本模型，峰值内存 23.12 GB，生成速度每秒 9.4 个 token。内存越小，能跑的模式越小，速度也越慢。开放权重、许可证、部署条件三件事分开看。要不要部署，和懂技术的同事一起定。

能不能替我批改。 不建议用它替您批改并直接把分数发给学生。批改是核不完的工作，逐份核的代价大。系统可以帮您起草反馈的初稿，可以按您给的标准先分类，分数由您定。学生作答属于学生数据，走本地或脱敏。

老师会被替代吗。可以先把工作拆成任务。第六部分那份数据显示，AI 进入的是任务，不是整个岗位。有的任务交出去了，有的任务多了人在旁边核。教案案例里，初稿它给，第二稿它改，中间那步判断不能交出。职业变化还受需求、组织和社会选择影响，单个演示回答不了一个职业会减少多少岗位。

从哪一件事开始。满足三个条件的那一件。材料在手，结果可核，出错代价小。把一章教材变成练习题、给一节新课出初稿、一张成绩表看分布，都适合先试。财务与报销表不要交。回复邮件可以起草，发送自己点。现在就写一句“我这周想让它帮我完成什么”，再写一个具体时间。

出了错谁负责。交出去的结果署名是您的，责任在使用者。这本手册因此反复讲核对，也因此把发送、付款、删除三件事留给人。用了 AI 的稿件，按期刊与学校的要求披露。把核对做成习惯，比讨论责任归属更实际。

附

APPENDIX

一手来源与延伸阅读

论文与研究。

- Vaswani 等，2017，Attention Is All You Need。arxiv.org/abs/1706.03762
- Clark 等，2019，What Does BERT Look At? An Analysis of BERT's Attention。arxiv.org/abs/1906.04341
- Liu 等，2024，Lost in the Middle，TACL。aclanthology.org/2024.tacl-1.9
- Yao 等，2023，ReAct。arxiv.org/abs/2210.03629
- DeepSeek-R1，2025。arxiv.org/abs/2501.12948
- Kestin 等，2025，Scientific Reports。doi 10.1038/s41598-025-97652-6
- Bastani 等，2025，PNAS。pmc.ncbi.nlm.nih.gov/articles/PMC12232635
- Anthropic Economic Index，2025-02-10。anthropic.com/news/the-anthropic-economic-index
- Morris 等，2023，Levels of AGI。arxiv.org/abs/2311.02462
- Perkins、Furze、Roe、MacVaugh，2024，The AI Assessment Scale (AIAS)，JUTLP 21(6)；arxiv.org/abs/2312.07086
- Nature 社论，2023-01-24，Tools such as ChatGPT threaten transparent science: here are our ground rules for their use，Nature 613, 612
- Kosmyrna 等，2025，Your Brain on ChatGPT。arxiv.org/abs/2506.08872
- Lee 等，2025，The Impact of Generative AI on Critical Thinking，CHI 2025。microsoft.com/en-us/research
- De Simone 等，2025，From chalkboards to chatbots，World Bank Blogs，2025-01-09。blogs.worldbank.org

思想来源，第七部分引文。

- 《尚书·说命中》，“非知之艰，行之惟艰”。zh.wikisource.org/wiki/尚書/說命中
- 孙中山，《孙文学说》（《建国方略》之一，心理建设），自序署 1918 年 12 月 30 日，1919 年刊行。zh.wikisource.org/wiki/心理建設
- 王阳明，《传习录》卷上，徐爱录。zh.wikisource.org/wiki/傳習錄/卷上

- Linus Torvalds, Linux 内核邮件列表, 2000-08-25。lkml.org/lkml/2000/8/25/132
- Melvin Kranzberg, 1986, Technology and History: “Kranzberg’s Laws”, Technology and Culture 27(3), 544–560。
- Alan Turing, 1950, Computing Machinery and Intelligence, Mind LIX(236), 433–460。
- “Code is cheap. Show me your idea.” 为流行表述, 不指定出处。

课程与官方页。

- 吴恩达 deeplearning.ai, AI Prompting for Everyone; Generative AI for Everyone。deeplearning.ai
- Andrew Ng, 2024-03-20, Agentic Design Patterns Part 1, The Batch。deeplearning.ai/the-batch
- Elements of AI, 赫尔辛基大学。course.elementsofai.com
- Stanford CS224N, 想看公式的教师。web.stanford.edu/class/cs224n
- UNESCO 教师 AI 能力框架。unesco.org/en/articles/ai-competency-framework-teachers
- 教育部等五部门, 《“人工智能+教育”行动计划》, 教科信〔2026〕1号。moe.gov.cn
- 科技部监督司, 《负责任研究行为规范指引(2023)》, 2023-12-21。most.gov.cn
- OpenAI, ChatGPT、GPT-4o、ChatGPT agent、GPT-6 Astra 发布页; OpenAI Charter。openai.com/index/…; openai.com/charter
- Dartmouth 1955 提案原文。www-formal.stanford.edu/jmc/history/dartmouth
- 本课材料与实测记录, 随课程材料提供, 可向课程组索取。

以上信息截至 2026 年 9 月, 以官方页面为准。每一条都是发布机构、课程提供者或原始研究, 没有二手转述。链接属于 [查一下再用](#), 引用之前逐条确认可访问。

附

APPENDIX

本手册的证据与更新

这本手册的事实分三种标记，课件用的是同一套分法。[读一遍就懂](#)的内容三到五年不动，只修表述。[查一下再用](#)的内容，产品名、开放条件、Astra 三组分数、经济指数数字、Kestin 研究数字，核验日期为 2026 年 9 月 11 日，每次开课前逐条重核。[动手前看](#)的内容，十八项实测全部在 2026 年 9 月 4 日至 11 日运行，每次开课前重跑一遍，出错的证据随之更新。

实测 01 到 16 的原始记录随课程材料提供。每条记录写明日期、工具、提示词、会话链接与人工核验结论；留有完整输出的是教案初稿、本地模型与教案第二稿三项，其余以会话链接为准。每条事实的来源、核验日期和可信程度，课程组另有一份记录，随课程材料一起提供。手册里的数字和那份记录对不上时，以记录为准，下次更新时改手册。

下次更新时会改三类内容。[查一下再用](#)的部分只换最新的发布，写法不动，Astra 之后有新发布就换上，经济指数用哪一期就注明哪一期。[动手前看](#)的部分按重跑结果更新数字，出错的案例若不再复现，标注日期并保留旧记录，不删除。还没实测的几项补做之后，相应内容从“现场演示”改为“实测”，例如实时语音陪练、桌面操作、本地模型的图片理解。[读一遍就懂](#)的讲法与练习不变。

一篇论文之后的世界 · 学员手册

Attention Is All You Need · Student Handbook · v4.0 · 2026-09

作者 李雨林

深圳欧美同学会副秘书长 · 世青峰理事长

计算机科班出身的工程师与创业者，长期在一线使用 AI 构建真实上线的产品与系统。

本手册配套课程《Attention Is All You Need | 一篇论文之后的世界》，主办：深圳欧美同学会 · 深圳留学人员联谊会。

手册中标注"本课实测"的内容，均为 2026 年 9 月 4 日至 11 日在所述工具上实际运行并经人工核对的结果，原始记录随课程材料提供；引用的论文、官方文件与公开报道在"一手来源与延伸阅读"中列出。标注"查一下再用"的内容以官方页当日说明为准。

课程材料按"长期有效的内容、会变的信息、实测记录"三类维护，每次开课前重核会变的信息、重跑实测记录。转载或改编请注明作者与出处。



课后交流，扫码进群。

联系 liyulin@globalyouth.cn · 世青峰 CGYS