

《AI 会不会让人类智力趋同？》

“人脑智能”课程研讨汇报 PPT 方案与演讲逐字稿

- 课程：人脑智能 (Human Brain Intelligence)
- 主题：AI 会不会让人类智力趋同？
- 立场：正方观点（会趋同）
- 汇报时长：约 8~10 分钟
- 基调：自然、平实、基于实验数据与神经科学机制，去除夸张抒情与宣讲口号，呈现真实高校学术研讨状态。

目录与幻灯片规划

- Slide 01 · 引入：从一个反直觉的现象说起
- Slide 02 · 核心分歧：我们在讨论“及格线提升”，还是“思考方差”？
- Slide 03 · 机制一（神经生理）：自由能假说与认知卸载
- Slide 04 · 机制二（计算语言）：88 万篇文本实证——写作复杂度的方差减半
- Slide 05 · 机制三（行为实验）：UCL 实验——个人得分变高，群体重合度暴增
- Slide 06 · 提示词迷思：0.92 的高相似度与“预训练凸包”
- Slide 07 · 机制四（认识论）：Nature 封面警告——科研视角的“单一种植”
- Slide 08 · 机制五（信息论）：合成数据循环与模型坍缩定理
- Slide 09 · 常见反驳回应：算盘、印刷术与大模型的本质区别
- Slide 10 · 总结：对课程与未来人机协作的启发

Slide 01 · 引入：从一个反直觉的现象说起

- 【页面要点】
- 课题：AI 会不会让人类智力趋同？
- 直觉预判：工具普及 $\rightarrow$ 门槛降低 $\rightarrow$ 表达更多元？
- 实证研究的反转：个体质量提升的同时，群体多样性在加速收敛
- 本次汇报主线：结合神经科学机制与近三年的顶刊实验数据，探讨这一收敛背后的必然性
- 【排版与视觉】
- 左侧：常识假设 vs 实证结果的对比列表。
- 右侧：简明的概念对照图（发散点集 vs 聚拢点集）。

- **【演讲逐字稿】**“老师好，各位同学好。今天我们组汇报的课题是‘AI 会不会让人类智力趋同’。

在刚拿到这个题目的时候，我们组里其实有过激烈的讨论。大家的初始直觉普遍觉得：‘怎么可能趋同？现在每个人都可以用不同的 Prompt，AI 降低了画画、写代码、查文献的门槛，不是应该千人千面、各显神通吗？’

但当我们去检索了近两年《Nature》、《Science Advances》以及认知神经科学的实证研究之后，发现实验数据和这个直觉正好相反：普通人使用 AI 之后，个人交出来的作业质量确实上去了，但大家做出来的东西，在语义空间和逻辑链条上的重叠率却大幅升高了。

今天我们不想用宏大的口号来争辩，而是想从人脑神经机制、计算语言学测试和信息论的角度，跟老师和大家一起梳理：为什么 AI 作为一种中心化的概率工具，在客观上正在把人类的思考方式往同一个模子里拉。”

Slide 02 · 核心分歧：我们在讨论“及格线提升”，还是“思考方差”？

- **【页面要点】**

- 混淆点：把“任务完成效率”当成“智力多样性”

- 统计学区隔：

- 均值 (Mean) 抬升：所有人都能写出合格的代码或论文（下限提高）
- 方差 (Variance) 收缩：解题思路、概念关联和假设发散度降低（多样性下降）

- 智力的核心特征：面对开放问题时的异质性认知路径 (Cognitive Diversity)

- **【排版与视觉】**

- 两个正态分布图对比：一个宽而平（原始人类分布，方差大），一个窄而高（AI 辅助后分布，均值右移但方差极小）。

- **【演讲逐字稿】**“首先必须澄清一个容易被偷换的概念：很多人说 AI 让人变聪明了，依据是以前不会写代码的人现在能写了，以前不及格的文章现在能打80分了。这在统计学上叫‘均值抬升’。

但这等于智力更多元了吗？举个很现实的例子：期末考试，如果全班同学都参考了同一份高质量参考答案，平均分从60分涨到了85分，但所有人答题的逻辑结构完全一样。这时候全班的思维是更多元了，还是高度趋同了？

在心理学和人脑科学里，智力从来不等于背诵或者套用标答的速度，而是指面对未知的复杂问题时，不同的大脑根据自己的生活经历、突触连接和直觉，能够产生完全不同的假设和思考进路。

所以今天讨论‘趋同’的核心指标，不是看平均分有没有涨，而是看人类思考的‘方差’是在扩大还是在收窄。”

Slide 03 · 机制一（神经生理）：自由能假说与认知卸载

- **【页面要点】**

- 大脑的基本运行规律：卡尔·弗里斯顿 (Karl Friston) 自由能原理——大脑是能耗最小化的预测机器

- 认知吝啬 (Cognitive Miser)：人类大脑天然倾向于选择能耗最低的认知路径

- 认知卸载 (Cognitive Offloading) 的升级：
 - 搜索时代：外包陈述性事实记忆 (Sparrow et al., Science 2011)
 - 生成式 AI 时代：外包逻辑推演、提纲构建与假设生成 (Jaan Aru, 2025)
- 神经可塑性与用进废退：缺乏高负荷思维磨砺，负责发散思考的额顶叶神经网络活跃度下降
- **【排版与视觉】**
- 上方：搜索时代（外包记忆）与生成式 AI 时代（外包推理）的对比流程图。
- 下方：Jaan Aru 论文中的脑网络机制简图（DMN 联想网络与 ECN 控制网络的协同受阻）。
- **【演讲逐字稿】**“结合我们课程学到的人脑机制，为什么人类会不由自主地走向趋同？根源在大脑的生物学本能。
弗里斯顿的‘自由能原理’大家都很熟悉：大脑为了维持代谢平衡，本质上是一个极其‘吝啬’的系统。能不费力就不费力。过去我们写一篇文章，需要在脑内反复权衡、碰壁、检索记忆，经历漫长的神经元突触重塑；现在大模型一秒钟就给出了一套结构工整的提纲。大脑基于最小能耗本能，必然会顺着这套现成的框架走。
2011年《Science》发表过著名的‘谷歌效应’：人一旦知道网上能查到，大脑就不会费力去记这个知识本身。到了今天的大模型时代，我们外包的已经不是‘知识本身’，而是把‘怎么归纳、怎么联想、怎么提出假说’这个最消耗脑力的核心思考过程，全部外包了出去。
爱沙尼亚神经科学家 Jaan Aru 在今年发表的论文里明确指出：缺乏持续的认知负荷刺激，大脑负责发散性假设生成的神经回路就会逐步钝化。这就像长期坐电梯的人，腿部肌肉自然会退化一样。”

Slide 04 · 机制二（计算语言）：88 万篇文本实证——方差断崖式减半

- **【页面要点】**
- 核心研究：Sourati et al., Nature Human Behaviour (2026) / USC 团队
- 样本量：涵盖 Reddit、本地社区新闻、arXiv 学术摘要在内超过 880,000 篇真实文本
- 关键量化数据：
 - 经大模型润色或辅助后，文本的句法与文风复杂度方差下降了 21% 至 50%
 - 原本能反映作者年龄、性别、文化背景的“社会语言学指纹”被显著清洗
 - 表达无意识地向 WHELM 风格收敛（西方、受过高等教育、温和精英语调）
- **【排版与视觉】**
- 左侧：88 万文本测试的核心数据卡片（-21%~-50% 方差，880k 样本）。
- 右侧：重写前后文本特征分布收缩的小提琴图（Violin Plot）。
- **【演讲逐字稿】**“接下来看实证。如果语言是思维的外壳，那使用大模型之后，人类的表达真的变单调了吗？
南加州大学团队刚在《Nature Human Behaviour》发了一篇规模很大的实证研究，他们抓取了 88 万篇真实场景下的文本，包括论坛帖子、新闻报道和学术摘要。他们测试了使用 GPT 和其他大模型润色之后文本的变化。
结果非常惊人：不管是哪个领域，只要经过大模型辅助，语言复杂度的方差直接掉了 21% 到 50%。”

更重要的是，本来每个人写东西都有独特的习惯、语气、甚至是方言倒装，研究人员原本可以通过机器学习模型识别出作者的大致年龄和文化背景。但在 AI 处理过之后，这些个人特质全被抹平了，文本统统被拉平到了那种典型的、极其客气、中规中矩的‘机器翻译腔’或者说高知中产文风。

当大家写邮件、写报告、甚至写学术论文都用同一种调和逻辑连接词的时候，我们实际能运用的概念工具和表达方式，其实已经被严重缩紧了。”

Slide 05 · 机制三（行为实验）：UCL 实验——个人得分变高，群体重合度暴增

- 【页面要点】
- 核心研究：Doshi & Hauser, Science Advances (2024)
- 实验设计：几百名作者进行短篇故事创作的随机双盲实验（纯人类组 vs 接入大模型灵感组）
- 实证结果：
 - 个人层面：故事评分明显提高（语言更流畅、逻辑更通顺）
 - 群体层面：通过语义嵌入（Embeddings）计算，作品之间的余弦相似度极显著上升（ $p < 0.001$ ）
- 结论：个人的理性选择（省力且拿高分），在群体宏观层面上演化成了多样性坍缩
- 【排版与视觉】
- 图表：条形图对比——个体创意评分（上升） vs 群体语义距离（下降）。
- 引用框：原论文中关于“生成式 AI 的社会两难（Social Dilemma）”的直接论述。
- 【演讲逐字稿】“光看文字表面可能还不够，那在深层的‘创意’上呢？
2024 年《Science Advances》发表了一篇很有代表性的行为学双盲实验。研究人员找了几百个人写故事，对照组纯靠自己想，实验组可以看 GPT-4 给出的几个情节大纲。
评审打分的结果是：看了 AI 灵感的人，写出来的故事普遍得分更高，尤其是原本不太会编故事的人，进步最明显。
但是，当研究人员把所有写出来的故事转化成高维向量、计算它们之间的语义相似度时，发现了一个很反讽的现象：AI 辅助组的所有故事，彼此之间长得极其相似，重合度显著高于纯人类组。
作者把这个现象叫做‘生成式 AI 的社会困境’：站在你个人的角度，用 AI 绝对划算，又快又好；但当所有人都这么干的时候，整个人类群体生产出来的内容，全部挤在了少数几个被算法判定为概率最高的情节套路里。原本那些虽然不成熟、但千奇百怪的离群点，彻底消失了。”

Slide 06 · 提示词迷思：0.92 的高相似度与“预训练凸包”

- 【页面要点】
- 反方常见假设：“只要 Prompt 写得个性化，AI 就会千人千面”
- 现实数据打脸：Wan & Kalman (2026, Computers in Human Behavior)
 - 普通人类在日常自然交互下，大模型输出的构思相似度高达 0.92
 - 只有在极其严苛的实验条件下、人工设定 10 个跨文化复杂 Persona，才能勉强缓解
- 技术底层逻辑：Prompt 只是微调，底层权重与 RLHF 价值观早已划定概率凸包（Convex Hull）

- **【排版与视觉】**
- 左侧：0.92 相似度柱状图。
- 右侧：示意图——千亿参数模型的概率凸包，以及用户 Prompt 只是在凸包内小幅晃动。
- **【演讲逐字稿】**“这时候很多技术流的同学可能会说：‘那是因为他们 Prompt 不会写。如果我加上特定的角色设定，比如让它扮演某个小众流派的作家，不就解决了吗？’
今年《Computers in Human Behavior》刚刊发了一篇复现实验，专门测了这个假设。结论很现实：在绝大多数人习惯的日常提问方式下，AI 给不同人生成的创意构思，相似度高达 0.92。
实验确实证明，如果你在实验室里极其费劲地设计 10 个横跨不同文化、极其刁钻的 Prompt 矩阵，确实能稍微拉开差距。但请大家想想现实：在座的各位平常用 ChatGPT、Claude 或者 Copilot，有谁每次干活前先写两千字的角色设定？
工程上讲究‘默认配置主导一切’。大家都是输入一两句话，或者顺着默认设置走。大模型是在相同的公网语料上预训练出来的，又经过了统一的人类反馈强化学习对齐。它给你框定的概率范围就那么大，无论你的 Prompt 怎么变，其实都是在同一个由算法划好的圈子里原地打转。”

Slide 07 · 机制四（认识论）：Nature 封面警告——科研视角的“单一种植”

- **【页面要点】**
- 核心文献：Messeri & Crockett, Nature 封面综述 (2024, Vol. 627)
- 科学领域的“单一种植”（Scientific Monoculture）：
 - 课题选择收窄：科学家优先研究“AI 容易计算和出结果的问题”，忽视难以量化、直觉性强的深刻难题
 - 方法论均质化：各学科方法论统统被降维成“特征提取 + 向量比对 + 模式匹配”
- 理解的三重幻觉（Illusions of Understanding）：把算力输出误当成人类自身的因果理解
- 核心警示：产出的论文越来越多，但真正具有范式突破性的认知却在变窄
- **【排版与视觉】**
- 树状对比图：
 - 原本的科学树（物理学、生物学、社会学各具独特方法论枝桠）；
 - AI 化后的科学树（所有枝桠被修剪成相同粗细的算法管道）。
- **【演讲逐字稿】**“不仅是普通写作者，哪怕是代表人类智力最高水平的科学家群体，也逃不过这种规律。
2024 年 3 月的《Nature》封面发了一篇非常深刻的综述，题目叫《AI 与科学研究中理解的幻觉》。作者指出了一个正在科研界悄悄蔓延的危险趋势：科学研究的‘单一种植’（Scientific Monoculture）。
这是什么意思？现在很多科研团队都在‘AI+X’。大家越来越倾向于去挑那些‘AI 擅长跑出数据、容易发文章’的课题；而那些需要长期实地考察、需要哲学反思、或者直觉性很强但很难被向量化的问题，被大家不自觉地冷落了。
而且，生物学、社会学、心理学原本都有自己独特的研究范式，现在全变成了‘调提示词、做文本嵌入、跑聚类分析’。方法论高度统一。
作者最后给出了一个很扎心的评价：我们现在的论文数据越来越多，但全人类对世界深层规律的真正理解，反而可能越来越窄了。”

Slide 08 · 机制五（信息论）：合成数据循环与模型坍缩定理

- 【页面要点】
- 核心文献：Shumailov et al., Nature (2024) 纯数学推导与实验
- 模型坍缩定理（Model Collapse）：
 - 递归使用前代生成的数据训练系统，导致概率分布长尾（低频离群值）不可逆地丢失
- 闭环生态风险：
 - AI 生成海量网络内容 $\rightarrow$ 人类大脑吸收同质化内容 $\rightarrow$ 人类产出更趋同的内容 $\rightarrow$ 回流训练下一代 AI
 - 认知生态的“自体吞噬”（Cognitive Autophagy）
- 【排版与视觉】
- 动态闭环流程图：从人类写作者到公网数据，再到模型微调。
- 右侧：高斯混合模型在多代递归后分布退化为单一尖峰的曲线图。
- 【演讲逐字稿】“从信息论的角度看，如果放眼未来几代人，这种趋同会不会自己改善？牛津大学团队在《Nature》上发表的一篇数学论文给出了否定的答案，这就是著名的‘模型坍缩’（Model Collapse）定理。他们用数学公式和实验严格证明：任何一个生成模型，如果反复使用它自己生成的数据来训练下一代，数据分布两端的长尾（也就是那些低频的、奇特的离群样本）就会迅速消失，模型最终会退化，只能输出毫无多样性的单调内容。
而现实中，我们正在形成这样一个闭环：互联网上现在有大量的文章、代码已经是 AI 生成的了。学生用 AI 写作业，学者用 AI 润色论文，程序员用 AI 写代码。我们的大脑天天在阅读这些高度标准化的产物，我们的思维习惯被塑形；然后我们再产出类似的内容发到网上，下一代模型又爬取这些内容去训练。
人类历史上那些真正颠覆性的天才，比如爱因斯坦、梵高，在他们那个时代在统计学上全都是绝对的‘离群点’。当全社会的知识生产都建立在‘消除离群噪声、寻找概率均值’的系统上时，这些宝贵的长尾方差正在被系统性地清洗。”

Slide 09 · 常见反驳回应：算盘、印刷术与大模型的本质区别

- 【页面要点】
- 反方王牌论据：“当年发明算盘、印刷术、计算器时，大家也担心过退化，但人类文明最终更繁荣了”
- 为什么不能做简单历史类比？
 - 物理工具（算盘/计算器/相机）：替代的是身体与机械执行（Execution），不碰问题定义与意义生成（Epistemic Framing）
 - 生成式 AI：人类历史上第一个认知生成型中介，直接代劳概念组合、假说构思与逻辑论证
- 类比谬误分析：把“大脑中枢的外包”误当成了“手指算盘珠的延伸”
- 【排版与视觉】
- 表格对比：
 - 工具类型 | 替代的环节 | 人类大脑核心角色 | 是否导致思维结构收敛
 - 算盘/计算器 | 算术法则运算 | 提出数学问题、建立公理体系 | 否

- 搜索引擎 | 事实检索与索引 | 综合判断、提炼因果 | 否（仅外包记忆）
- 生成式 AI | 概念抽象、逻辑生成 | 给出提示词并确认 | 是（让渡了思维过程本身）

- **【演讲逐字稿】**“在准备这次汇报时，我们猜到很多同学可能会用历史来反驳：‘算盘出现的时候大家也担心退化，计算器出来的时候老师也怕学生不会算数，但事实证明人类不仅没退化，反而探索了更高深的科学，这次不也一样吗？’

这个反驳听起来很顺理成章，但它存在严重的范畴混淆。

算盘和计算器替代的是什么？是确定规则下的机械运算。算盘从来不会替老账房决定今年这笔账该不该借，计算器也不会替物理学家构思广义相对论的场方程。相机的出现替代了写实技法，但它没有替画家构思什么是立体主义。

但生成式 AI 替代的是什么？是‘意义生成’本身。它直接帮你写提纲、帮你找论点、帮你提炼假说、帮你决定因果关系。

把一个替你‘思考和表达’的认知中介，简单等同于算盘这种‘四肢延伸’的工具，完全低估了生成式 AI 对人类心智结构的介入深度。历史经验在这里是无法无条件直接套用的。”

Slide 10 · 总结：对课程与未来人机协作的启发

- **【页面要点】**
- 核心结论重申：
 - 均值提升是技术红利，方差收窄是结构性代价
 - AI 不是让人变傻，而是通过低能耗的算法引力，让人类心智向模式中位数聚集
- 对我们学习“人脑智能”的启发：
 - 保持神经负荷：不要让前额叶放弃痛苦但必要的深层思考
 - 警惕算法中位数：在科研与工程中主动寻找非共识与异常样本
 - 人机共生的底线：把 AI 当作镜子或对手，而不是直接接受它给出的最终构架
- **【排版与视觉】**
- 左侧：核心结论三句话卡片。
- 右侧：供课堂讨论的开放性思考题：“如果未来考试所有人都能用 AI，高校应当如何考核人类智力真正的‘离群价值’？”
- **【演讲逐字稿】**“最后总结一下我们这组的观点。

我们论证‘AI 会让人类智力趋同’，并不是在鼓吹技术的‘末日论’，更不是建议大家明天就把 ChatGPT 删掉。

在现代工程和产业实践中，均值的提高是实打实的技术红利，它让社会生产力大幅提升。但作为学习‘人脑智能’的学生，我们必须保持足够的清醒：效率的代价，往往就是方差的萎缩。

AI 并不会让人类退化成笨蛋，但它确实像一股强大的引力场，用最低的认知成本，诱导我们所有人都去采纳那个经过统计对齐的最优解。久而久之，人类思维这片原本物种丰富、偶尔会长出奇特怪木的热带雨林，很容易被修剪成整齐划一、但抗风险能力极差的单一种植农田。

对我们而言，最重要的是认清这种趋同的机制。在日常使用 AI 的时候，有意识地保留自己的批判性思

维，敢于提出算法不推荐的冷门构思，守护好大脑深处那份不可预测的探索方差。
以上就是我们小组的汇报，谢谢老师和各位同学！”