FRTI64 · AI LEARNING PATH

AI 学习路径指南 · 资料汇编

144 条资源 · 14 个学习阶段 · 242 份 PDF 留档

指南原稿	《AI学习路径指南》PDF，32 页（Gemini 多轮问答打印稿）
资源条目	144 条（18 种类型，覆盖指南 14 个阶段）
已核验可访问	133 条；11 条出版社反爬但书目元数据已补齐
开放获取	46 条
PDF 留档	242 份 / 229.34 MiB（主库 32 + OCW 131 + 剑桥 72 + 其它 7）
整理时间	2026-11-22
网页检索版	https://frti64.top/ai-resources/

数据来源：《AI学习路径指南》PDF（32 页，Gemini 多轮问答打印稿）与其原始 HTML 导出；网页检索版：https://frti64.top/ai-resources/

目录 / Contents

一、概览 3
二、指南主线 4
三、分阶段资料清单 5
S1 一、核心知识底座 5
S2 二、主流细分方向 5
S3 三、渐进式学习路径 6
S4 顶会论文与高薪研发岗 6
S5 低算力顶会课题 7
S6 以研代学 / 30 天计划 8
S7 证明型选手的理论赛道 9
S8 数学诊断题 9
S9 AI 理论推荐教材 9
S10 数学基础教材 10
S11 开放获取教材 11
S12 配套习题集 12
S13 纯数学教材 13
S14 开源纯数学与题库 15
四、指南摘要 17
五、PDF 下载留档 20
六、缺口与已知问题 21
七、怎么用 22
八、数据来源与复现 23

一、概览

指标 / Metric	数值 / Value
资源总数 Total resources	144
已抓取 Crawled OK	133 / 144
抓取失败但元数据已补 Blocked, metadata recovered	11
开放获取 Open access	46
PDF 留档 Archived PDFs	242 份 / 229.34 MiB
覆盖阶段 Stages	S1-S14（14 个）
网页检索版 Web index	https://frti64.top/ai-resources/

二、指南主线

阶段 / Stage	名称 / Name	覆盖内容 / Focus
S1	一、核心知识底座	数理/编程/机器学习与深度学习核心
S2	二、主流细分方向	大模型算法、AI 应用工程、多模态具身、AI 基础设施、垂直行业
S3	三、渐进式学习路径	夯实基础→经典 ML→Transformer→实战项目→论文闭环
S4	顶会论文与高薪研发岗	选题/方法论/面试三维能力/12-18 个月节奏
S5	低算力顶会课题	五个低算力课题与饱和度分析
S6	以研代学 / 30 天计划	最小技术栈、代码驱动读论文、PyTorch+Transformer+LoRA 30 天
S7	证明型选手的理论赛道	优化几何/RLHF 理论/注意力理论/可信 AI
S8	数学诊断题	5 道顶会级数学推导诊断题
S9	AI 理论推荐教材	高维概率、矩阵分析、凸优化、信息论、前沿讲义
S10	数学基础教材	线代/分析/概率/泛函/数值优化
S11	开放获取教材	作者官网免费 PDF 的教材与讲义
S12	配套习题集	带官方解答的证明/放缩习题集
S13	纯数学教材	实分析/抽象代数/测度概率/微分几何与最优传输
S14	开源纯数学与题库	开放获取纯数学教材 + MIT OCW/剑桥 Tripos/arXiv 题库

说明：阶段名称取自《AI学习路径指南》原文（S1-S14），原文为中文；英文为对照翻译。

三、分阶段资料清单

按指南 14 个阶段编排；同一资料可能出现在多个阶段，编号唯一。

S1 一、核心知识底座

> 数理/编程/机器学习与深度学习核心

编号	资料	类型	获取方式	说明	链接
P001	Attention Is All You Need 留档 1 份	论文	开放获取	Transformer 与自注意力机制原始论文，现代生成式 AI 基石	https://arxiv.org/abs/1706.03762
T001	Python	平台	免费	编程基石：核心语法、面向对象，AI 全部生态的语言底座	https://www.python.org/
T002	NumPy	库	免费	脱离 for 循环手写向量化矩阵运算与损失函数	https://numpy.org/
T003	Pandas	库	免费	数据科学栈的数据清洗与分析入口	https://pandas.pydata.org/
T004	Scikit-learn	库	免费	跑通分类/回归/聚类/交叉验证的经典机器学习基线	https://scikit-learn.org/
T005	XGBoost	库	免费	表数据经典强基线，理解梯度提升	https://xgboost.readthedocs.io/
T006	PyTorch	框架	免费	深度学习主框架：不依赖高层封装手写训练循环	https://pytorch.org/
T027	Docker	工具	免费	容器化部署，复现实验环境	https://www.docker.com/
T028	Git	工具	免费	版本控制；开源贡献与技术信用的基础设施	https://git-scm.com/
T007	Hugging Face	平台	免费	模型/数据集/镜像源生态，单卡跑通开源模型的入口	https://huggingface.co/

S2 二、主流细分方向

> 大模型算法、AI 应用工程、多模态具身、AI 基础设施、垂直行业

编号	资料	类型	获取方式	说明	链接
T030	MCP 协议	网站	免费	Model Context Protocol：Agent 工具调用与上下文接入标准	https://modelcontextprotocol.io/
T014	Megatron-LM	框架	免费	分布式训练：TP/PP/EP 并行与大规模预训练工程	https://github.com/NVIDIA/Megatron-LM
T015	DeepSpeed	框架	免费	ZeRO 显存优化与分布式训练，面试高频考点	https://github.com/deepspeedai/DeepSpeed
T023	CUDA	平台	免费	GPU 计算平台；底层推理加速/算子开发的必备技能	https://developer.nvidia.com/cuda-toolkit
T024	Triton	框架	免费	Python 写 GPU 算子：融合算子实现真实吞吐提升	https://github.com/triton-lang/triton
T025	Isaac Sim	框架	免费	机器人仿真平台，具身智能/Sim2Real 迁移	https://developer.nvidia.com/isaac/sim
T026	MuJoCo	框架	免费	物理仿真引擎，VLA/机器人策略研究	https://github.com/google-deepmind/mujoco
T012	vLLM	框架	免费	高吞吐推理引擎：评测加速、KV Cache 压缩实验平台	https://github.com/vllm-project/vllm
T013	SGLang	框架	免费	推理加速框架，与 vLLM 并列用于采样加速与吞吐测试	https://github.com/sgl-project/sglang
T016	veRL	框架	免费	RL 训练框架：提交 PR 建立开源技术信用、申请研究型实习	https://github.com/volcengine/verl
T017	OpenRLHF	框架	免费	RLHF/RLVR 训练框架，后训练方向核心工具	https://github.com/OpenRLHF/OpenRLHF

编号	资料	类型	获取方式	说明	链接
T034	Qwen（Qwen2.5 / Qwen3）	模型	免费	开源模型家族：单卡实验与多模型家族验证的首选基座	https://github.com/QwenLM/Qwen3
T009	PEFT	库	免费	参数高效微调：LoRA/QLoRA 显存友好微调	https://github.com/huggingface/peft
T035	Llama（Llama-3）	模型	免费	开源模型家族，多家族对比实验	https://github.com/meta-llama/llama-models

S3 三、渐进式学习路径

> 夯实基础→经典 ML→Transformer→实战项目→论文闭环

编号	资料	类型	获取方式	说明	链接
P001	Attention Is All You Need 留档 1 份	论文	开放获取	Transformer 与自注意力机制原始论文，现代生成式 AI 基石	https://arxiv.org/abs/1706.03762
T001	Python	平台	免费	编程基石：核心语法、面向对象，AI 全部生态的语言底座	https://www.python.org/
T002	NumPy	库	免费	脱离 for 循环手写向量化矩阵运算与损失函数	https://numpy.org/
T004	Scikit-learn	库	免费	跑通分类/回归/聚类/交叉验证的经典机器学习基线	https://scikit-learn.org/
T006	PyTorch	框架	免费	深度学习主框架：不依赖高层封装手写训练循环	https://pytorch.org/
T007	Hugging Face	平台	免费	模型/数据集/镜像源生态，单卡跑通开源模型的入口	https://huggingface.co/
T031	GitHub	平台	免费	端到端项目产出与复现代码公开的载体	https://github.com/
C001	Machine Learning Roadmap: Beginner to Expert (2026) - Coursera	路线图	免费	指南 Sources：Coursera 机器学习路线图课程	https://www.coursera.org/resources/ml-learni…
C002	关于 AI 的学习路线图 - 凌杰 - 博客园	博客	免费	指南 Sources：中文 AI 学习路线图博客	https://www.cnblogs.com/owlman/p/19560606
C003	A Machine Learning Roadmap for 2026, from Python to Papers	博客	免费	指南 Sources：从 Python 到论文的 2026 机器学习路线图	https://www.tensortonic.com/guides/machine-l…
C004	Machine Learning Roadmap for Beginners: 5 Essential Skills - IAICB	博客	免费	指南 Sources：初学者五项必备技能	https://www.iaicb.org/blog/machine-learning-…
C005	查阅了十几篇学习资源后，我总结了这份AI学习路径 - mason's blog	博客	免费	指南 Sources：中文 AI 学习路径总结	https://mianshenglee.github.io/2019/09/06/ai…
C006	学习路径总览：从零基础到找到第一份相关工作 - GitHub	路线图	免费	指南 Sources：GitHub 上的从零到就业学习路径总览	https://github.com/datawhalechina/dive-into-…
T032	arXiv	网站	免费	每日刷 cs.CL/cs.LG/cs.CV，精读 1-2 篇最新高赞论文	https://arxiv.org/
T008	Transformers（Hugging Face 库）	库	免费	AutoModel/AutoTokenizer 三件套，提取中间层激活与注意力矩阵	https://github.com/huggingface/transformers

S4 顶会论文与高薪研发岗

> 选题/方法论/面试三维能力/12-18 个月节奏

编号	资料	类型	获取方式	说明	链接
T014	Megatron-LM	框架	免费	分布式训练：TP/PP/EP 并行与大规模预训练工程	https://github.com/NVIDIA/Megatron-LM
T015	DeepSpeed	框架	免费	ZeRO 显存优化与分布式训练，面试高频考点	https://github.com/deepspeedai/DeepSpeed

编号	资料	类型	获取方式	说明	链接
T023	CUDA	平台	免费	GPU 计算平台；底层推理加速/算子开发的必备技能	https://developer.nvidia.com/cuda-toolkit
T031	GitHub	平台	免费	端到端项目产出与复现代码公开的载体	https://github.com/
T032	arXiv	网站	免费	每日刷 cs.CL/cs.LG/cs.CV，精读 1-2 篇最新高赞论文	https://arxiv.org/
P006	DeepSeek-V3 Technical Report 留档 1 份	技术报告	开放获取	精读顶级技术报告的 Limitations，找可迁移到低资源场景的机制	https://arxiv.org/abs/2412.19437
P007	DeepSeek-R1 留档 1 份	技术报告	开放获取	RLVR/长思维链推理的标志性技术报告	https://arxiv.org/abs/2501.12948
P008	Qwen2.5 Technical Report 留档 1 份	技术报告	开放获取	开源模型技术报告：数据配比与后训练细节	https://arxiv.org/abs/2412.15115
P009	Qwen3 Technical Report 留档 1 份	技术报告	开放获取	最新开源模型技术报告，跟踪前沿训练配方	https://arxiv.org/abs/2505.09388
P010	GPT-4 Technical Report 留档 1 份	技术报告	开放获取	OpenAI 技术报告范式：能力评测与安全约束	https://arxiv.org/abs/2303.08774
P011	Anthropic Claude 3 Model Card 留档 1 份	技术报告	开放获取	Anthropic 技术报告/模型卡：可解释性与评测口径参考	https://docs.anthropic.com/en/docs/resources…
P012	ByteDance Seed 技术报告 留档 1 份（补档）	技术报告	开放获取	Seed（字节）技术报告：Seed 系列模型训练与评测	https://arxiv.org/abs/2505.07062
T012	vLLM	框架	免费	高吞吐推理引擎：评测加速、KV Cache 压缩实验平台	https://github.com/vllm-project/vllm
T013	SGLang	框架	免费	推理加速框架，与 vLLM 并列用于采样加速与吞吐测试	https://github.com/sgl-project/sglang
T016	veRL	框架	免费	RL 训练框架：提交 PR 建立开源技术信用、申请研究型实习	https://github.com/volcengine/verl
T017	OpenRLHF	框架	免费	RLHF/RLVR 训练框架，后训练方向核心工具	https://github.com/OpenRLHF/OpenRLHF
T029	LeetCode	网站	免费	算法题保持零 Bug 编写能力，面试硬性要求	https://leetcode.cn/
P002	Direct Preference Optimization (DPO) 留档 1 份	论文	开放获取	DPO 原始推导：RLHF/GRPO 理论分析基石，面试手推 KL 正则闭式解	https://arxiv.org/abs/2305.18290

S5 低算力顶会课题

> 五个低算力课题与饱和度分析

编号	资料	类型	获取方式	说明	链接
T024	Triton	框架	免费	Python 写 GPU 算子：融合算子实现真实吞吐提升	https://github.com/triton-lang/triton
T032	arXiv	网站	免费	每日刷 cs.CL/cs.LG/cs.CV，精读 1-2 篇最新高赞论文	https://arxiv.org/
T012	vLLM	框架	免费	高吞吐推理引擎：评测加速、KV Cache 压缩实验平台	https://github.com/vllm-project/vllm
T013	SGLang	框架	免费	推理加速框架，与 vLLM 并列用于采样加速与吞吐测试	https://github.com/sgl-project/sglang
T034	Qwen（Qwen2.5 / Qwen3）	模型	免费	开源模型家族：单卡实验与多模型家族验证的首选基座	https://github.com/QwenLM/Qwen3
T041	MATH-500	评测基准	免费	推理模型评测基准，课题一/二的评测集	https://huggingface.co/datasets/HuggingFaceH…
T042	AIME	评测基准	免费	数学竞赛评测基准，检验推理模型上限	https://huggingface.co/datasets/AI-MO/aimo-v…
T043	GPQA	评测基准	免费	研究生级问答评测基准	https://github.com/idavidrein/gpqa

编号	资料	类型	获取方式	说明	链接
T044	LongBench	评测基准	免费	长文本评测基准：KV Cache 压缩实验	https://github.com/THUDM/LongBench
T045	RULER	评测基准	免费	长上下文评测基准	https://github.com/NVIDIA/RULER
T046	Needle-in-a-Haystack	评测基准	免费	大海捞针测试，验证长上下文信息保留	https://github.com/gkamradt/LLMTest_NeedleIn…
T047	SWE-bench	评测基准	免费	软件工程 Agent 评测（SWE-bench Lite）	https://github.com/SWE-bench/SWE-bench
T048	OSWorld	评测基准	免费	计算机操作 Agent 评测环境	https://github.com/xlang-ai/OSWorld
T049	WebArena	评测基准	免费	网页智能体评测环境	https://github.com/web-arena-x/webarena
T018	Unsloth	库	免费	单卡微调加速，显存需求可降 60%	https://github.com/unslothai/unsloth
T035	Llama（Llama-3）	模型	免费	开源模型家族，多家族对比实验	https://github.com/meta-llama/llama-models
T036	Mistral	模型	免费	开源模型家族，多家族对比实验	https://github.com/mistralai/mistral-inference
T037	Gemma	模型	免费	开源模型家族，多家族对比实验	https://github.com/google-deepmind/gemma
P005	Teacher-Guided Curriculum for Data-Efficient RLVR 留档 1 份	论文	开放获取	低算力 RLVR 课程学习代表作（ACL Anthology），课题二的对标工作	https://aclanthology.org/2026.findings-emnlp…
T033	ACL Anthology	网站	免费	NLP 顶会论文库，检索低算力课题的代表作	https://aclanthology.org/

S6 以研代学 / 30 天计划

> 最小技术栈、代码驱动读论文、PyTorch+Transformer+LoRA 30 天

编号	资料	类型	获取方式	说明	链接
T006	PyTorch	框架	免费	深度学习主框架：不依赖高层封装手写训练循环	https://pytorch.org/
T007	Hugging Face	平台	免费	模型/数据集/镜像源生态，单卡跑通开源模型的入口	https://huggingface.co/
T012	vLLM	框架	免费	高吞吐推理引擎：评测加速、KV Cache 压缩实验平台	https://github.com/vllm-project/vllm
T034	Qwen（Qwen2.5 / Qwen3）	模型	免费	开源模型家族：单卡实验与多模型家族验证的首选基座	https://github.com/QwenLM/Qwen3
T009	PEFT	库	免费	参数高效微调：LoRA/QLoRA 显存友好微调	https://github.com/huggingface/peft
T035	Llama（Llama-3）	模型	免费	开源模型家族，多家族对比实验	https://github.com/meta-llama/llama-models
T008	Transformers（Hugging Face 库）	库	免费	AutoModel/AutoTokenizer 三件套，提取中间层激活与注意力矩阵	https://github.com/huggingface/transformers
C007	Let's build GPT: from scratch, in code — Andrej Karpathy	教程	免费	跟着视频用纯 PyTorch 手敲 Self-Attention/位置编码/自回归采样	https://www.youtube.com/watch?v=kCc8FmEb1nY
T019	einops	库	免费	rearrange 张量维度变换，养成标注 Shape 的习惯	https://github.com/arogozhnikov/einops
T038	nanoGPT	仓库	免费	跟随 Karpathy 从零拼装微型 GPT 的参考实现	https://github.com/karpathy/nanoGPT
T011	bitsandbytes	库	免费	4-bit/8-bit 量化加载，显存控制关键	https://github.com/bitsandbytes-foundation/b…

编号	资料	类型	获取方式	说明	链接
T039	Alpaca	数据集	免费	指令微调数据格式范本	https://github.com/tatsu-lab/stanford_alpaca
T040	ShareGPT	数据集	免费	对话式 SFT 数据格式（Alpaca/ShareGPT 两种格式对照）	https://sharegpt.com/
T010	trl	库	免费	SFTTrainer 与后训练工具链，30 天计划的 LoRA 实战	https://github.com/huggingface/trl
T020	SwanLab	工具	免费	实验记录与可视化，从第一天起记录学习率/Loss	https://github.com/SwanHubX/SwanLab
T021	TensorBoard	工具	免费	经典实验日志与曲线可视化	https://www.tensorflow.org/tensorboard

S7 证明型选手的理论赛道

> 优化几何/RLHF 理论/注意力理论/可信 AI

编号	资料	类型	获取方式	说明	链接
P004	Grokking: Generalization Beyond Overfitting 留档 1 份	论文	开放获取	顿悟现象（Grokking）与泛化相变，优化几何方向的经典论文	https://arxiv.org/abs/2201.02177
P002	Direct Preference Optimization (DPO) 留档 1 份	论文	开放获取	DPO 原始推导：RLHF/GRPO 理论分析基石，面试手推 KL 正则闭式解	https://arxiv.org/abs/2305.18290
P003	In-Context Learning as Gradient Descent 留档 1 份	论文	开放获取	证明 Transformer 上下文学习等价于隐式执行梯度下降，理论型顶会经典	https://arxiv.org/abs/2212.07677
T022	matplotlib	库	免费	理论验证曲线：双对数坐标斜率验证 O(1/sqrt(T)) 等定理预测	https://matplotlib.org/
B009	Conformal Prediction: A Gentle Introduction — Anastasios N. Angelopoulos & Stephen Bates 留档 1 份	教材	开放获取	约 50 页短篇专著：无分布依赖覆盖率保证，3 个月产出一篇不确定性/幻觉控制论文	https://arxiv.org/abs/2107.07511
C008	Stanford CS229M 讲义：Statistical Learning Theory — Tengyu Ma 留档 1 份	讲义	开放获取	为写 NeurIPS/ICML 理论论文设计：表达能力/隐式偏置/OOD 泛化	https://web.stanford.edu/class/stats214/

S8 数学诊断题

> 5 道顶会级数学推导诊断题

编号	资料	类型	获取方式	说明	链接
P004	Grokking: Generalization Beyond Overfitting 留档 1 份	论文	开放获取	顿悟现象（Grokking）与泛化相变，优化几何方向的经典论文	https://arxiv.org/abs/2201.02177
P003	In-Context Learning as Gradient Descent 留档 1 份	论文	开放获取	证明 Transformer 上下文学习等价于隐式执行梯度下降，理论型顶会经典	https://arxiv.org/abs/2212.07677

S9 AI 理论推荐教材

> 高维概率、矩阵分析、凸优化、信息论、前沿讲义

编号	资料	类型	获取方式	说明	链接
B001	High-Dimensional Probability: An Introduction with Applications in Data Science — Roman Vershynin 留档 1 份	教材	开放获取	AI 理论圣经（开源）：浓度不等式、sub-gaussian、覆盖数、随机矩阵奇异值界，精读 1-4 章	https://www.math.uci.edu/~rvershyn/papers/HD…
B006	The Matrix Cookbook 留档 1 份	教材	开放获取	矩阵求导/迹循环变换速查手册，打印放手边	https://www.math.uwaterloo.ca/~hwolkowi/matr…
B002	High-Dimensional Statistics: A Non-Asymptotic Viewpoint — Martin J. Wainwright	教材	付费	更严谨的进阶工具书：尾部界与 Rademacher 复杂度/度量熵	https://www.cambridge.org/core/books/highdim…
B003	Convex Optimization — Stephen Boyd & Lieven Vandenberghe 留档 1 份	教材	开放获取	变分推导与对偶理论标杆（免费）：凸函数/共轭/KKT，DPO 闭式解来源	https://web.stanford.edu/~boyd/cvxbook/
B004	Convex Optimization: Algorithms and Complexity — Sébastien Bubeck 留档 1 份	教材	开放获取	微软研究院短篇专著：梯度下降/镜像下降/SGD 收敛率证明模板	https://arxiv.org/abs/1405.4980
B005	Matrix Analysis — Roger A. Horn & Charles R. Johnson	教材	付费	当字典查：奇异值不等式、Weyl 定理、Davis-Kahan 特征子空间扰动	https://www.cambridge.org/core/books/matrix-…
B007	Elements of Information Theory — Thomas M. Cover & Joy A. Thomas	教材	付费	信息论经典：熵/互信息/KL/DPI、Fano 不等式与 Method of Types	https://www.wiley.com/en-us/Elements+of+Info…
B008	Patterns, Predictions, and Actions — Moritz Hardt & Benjamin Recht 留档 1 份	教材	开放获取	免费在线新书：用现代视角重写统计学习、优化动力学与因果推断	https://mlstory.org/
B009	Conformal Prediction: A Gentle Introduction — Anastasios N. Angelopoulos & Stephen Bates 留档 1 份	教材	开放获取	约 50 页短篇专著：无分布依赖覆盖率保证，3 个月产出一篇不确定性/幻觉控制论文	https://arxiv.org/abs/2107.07511
C008	Stanford CS229M 讲义：Statistical Learning Theory — Tengyu Ma 留档 1 份	讲义	开放获取	为写 NeurIPS/ICML 理论论文设计：表达能力/隐式偏置/OOD 泛化	https://web.stanford.edu/class/stats214/

S10 数学基础教材

> 线代/分析/概率/泛函/数值优化

编号	资料	类型	获取方式	说明	链接
B010	Linear Algebra Done Right — Sheldon Axler 留档 1 份	教材	开放获取	摒弃行列式直接建谱定理，官方开放获取 PDF + 作者习题讲解视频	https://linear.axler.net/
B011	Linear Algebra and Learning from Data — Gilbert Strang	教材	付费	面向 ML/数据科学：SVD、低秩近似、图拉普拉斯与梯度下降的代数本质	https://math.mit.edu/~gs/learningfromdata/
B012	Matrix Computations — Gene H. Golub & Charles F. Van Loan 元数据已补	教材	付费	数值线性代数权威经典：正交分解/条件数/特征值扰动/低秩逼近误差	https://www.press.jhu.edu/books/title/10678/…

编号	资料	类型	获取方式	说明	链接
B013	高等代数 — 丘维声	教材	付费	中文打底：多项式、线性变换不变子空间、二次型的严格证明训练	https://book.douban.com/subject/26739983/
B014	高等代数学 — 张贤科	教材	付费	中文打底：与丘维声并列的高等代数严格证明训练	https://book.douban.com/subject/1051081/
B015	Principles of Mathematical Analysis（Baby Rudin）— Walter Rudin	教材	付费	严格 ε-δ 语言、紧致性、一致收敛与函数项级数放缩试金石	https://archive.org/details/principlesofmath…
B016	Measure, Integration & Real Analysis — Sheldon Axler 留档 1 份	教材	开放获取	现代实分析/测度论：Lebesgue 积分、控制收敛、L^p 与 Hilbert 空间（开放获取）	https://measure.axler.net/
B017	Mathematical Analysis — V. A. Zorich（卓里奇《数学分析》中译本）	教材	付费	兼具物理几何直觉与严谨性：多元微分、隐函数定理、流形初步	https://link.springer.com/book/10.1007/978-3…
B018	Probability: Theory and Examples — Rick Durrett 留档 1 份	教材	开放获取	测度论概率标准教材：大数定律/CLT/条件期望/鞅收敛（作者官网免费）	https://services.math.duke.edu/~rtd/PTE/
B019	Probability and Random Processes — Geoffrey Grimmett & David Stirzaker 元数据已补	教材	付费	从初等概率过渡到马尔可夫链、更新过程与连续时间随机过程	https://academic.oup.com/book/52807
B020	概率论基础 — 李贤平	教材	付费	中文打底：培养古典与初等概率直觉	https://xuanshu.hep.com.cn/front/book/findBo…
B021	概率论基础 — 严士健、王隽骧	教材	付费	中文打底：测度论公理化体系的概率论	https://book.douban.com/subject/1553785/
B022	Introductory Functional Analysis with Applications — Erwin Kreyszig	教材	付费	证明极其详尽：Banach/Hilbert 空间、有界与紧算子谱理论，NTK/RKHS 敲门砖	https://www.wiley.com/en-us/Introductory+Fun…
B023	Numerical Optimization — Jorge Nocedal & Stephen Wright	教材	付费	连续优化迭代算法收敛性证明：线搜索/信赖域/共轭梯度/拟牛顿	https://link.springer.com/book/10.1007/978-0…

S11 开放获取教材

> 作者官网免费 PDF 的教材与讲义

编号	资料	类型	获取方式	说明	链接
B001	High-Dimensional Probability: An Introduction with Applications in Data Science — Roman Vershynin 留档 1 份	教材	开放获取	AI 理论圣经（开源）：浓度不等式、sub-gaussian、覆盖数、随机矩阵奇异值界，精读 1-4 章	https://www.math.uci.edu/~rvershyn/papers/HD…
B003	Convex Optimization — Stephen Boyd & Lieven Vandenberghe 留档 1 份	教材	开放获取	变分推导与对偶理论标杆（免费）：凸函数/共轭/KKT，DPO 闭式解来源	https://web.stanford.edu/~boyd/cvxbook/
B010	Linear Algebra Done Right — Sheldon Axler 留档 1 份	教材	开放获取	摒弃行列式直接建谱定理，官方开放获取 PDF + 作者习题讲解视频	https://linear.axler.net/

编号	资料	类型	获取方式	说明	链接
B016	Measure, Integration & Real Analysis — Sheldon Axler 留档 1 份	教材	开放获取	现代实分析/测度论：Lebesgue 积分、控制收敛、L^p 与 Hilbert 空间（开放获取）	https://measure.axler.net/
B018	Probability: Theory and Examples — Rick Durrett 留档 1 份	教材	开放获取	测度论概率标准教材：大数定律/CLT/条件期望/鞅收敛（作者官网免费）	https://services.math.duke.edu/~rtd/PTE/
B024	Linear Algebra — Jim Hefferon 留档 1 份	教材	开放获取	GFDL 开源教材 + 逐题完整解答手册（证明步骤无跳跃），官方习题全解	https://hefferon.net/linearalgebra/
B025	Introduction to Applied Linear Algebra — Stephen Boyd & Lieven Vandenberghe 留档 1 份	教材	开放获取	向量几何/正交投影/最小二乘/矩阵分解的直观推导，配套 Python/Julia 代码	https://vmls-book.stanford.edu/
B026	Basic Analysis I & II: Introduction to Real Analysis — Jiří Lebl 留档 1 份	教材	开放获取	CC BY-NC-SA 开源实分析教材，逻辑严谨度对标 Rudin 但语言更现代	https://jirka.org/ra/
B027	Introduction to Probability — Charles M. Grinstead & J. Laurie Snell 留档 1 份	教材	开放获取	AMS/达特茅斯免费开源：组合与初等概率过渡到马尔可夫链与随机游走	https://math.dartmouth.edu/~prob/prob/
B028	Functional Analysis — Gerald Teschl 留档 1 份（补档）	教材	开放获取	维也纳大学研究生教材：Banach 代数、紧算子谱理论、无界算子（作者主页免费）	https://www.mat.univie.ac.at/~gerald/ftp/boo…
B029	Foundations of Data Science — Avrim Blum, John Hopcroft & Ravindran Kannan 留档 1 份	教材	开放获取	面向计算机理论研究者：高维几何、SVD、随机游走与泛化界	https://www.cs.cornell.edu/jeh/book.pdf
B030	Mathematics for Machine Learning — Marc Peter Deisenroth et al. 留档 1 份	教材	开放获取	线代/矩阵微积分/概率/优化直接映射到 PCA、高斯过程、SVM（剑桥出版社授权免费）	https://mml-book.github.io/

S12 配套习题集

> 带官方解答的证明/放缩习题集

编号	资料	类型	获取方式	说明	链接
B010	Linear Algebra Done Right — Sheldon Axler 留档 1 份	教材	开放获取	摒弃行列式直接建谱定理，官方开放获取 PDF + 作者习题讲解视频	https://linear.axler.net/
B024	Linear Algebra — Jim Hefferon 留档 1 份	教材	开放获取	GFDL 开源教材 + 逐题完整解答手册（证明步骤无跳跃），官方习题全解	https://hefferon.net/linearalgebra/
B031	Convex Optimization 官方补充习题与全解（Additional Exercises）— Stephen Boyd 留档 1 份（补档）	仓库	开放获取	EE364a 历年作业/考试证明题与官方推导：对偶构造、KKT、Fenchel 共轭	https://github.com/cvxgrp/cvxbook_additional…
B032	Linear Algebra 官方全解手册 — Jim Hefferon（官方习题解答） 留档 1 份	习题集	开放获取	与教材同步的逐题证明全解，规范线性空间/正交投影/特征值分解书写	https://hefferon.net/linearalgebra/
B033	The Cauchy-Schwarz Master Class — J. Michael Steele	教材	付费	不等式放缩艺术神作：Cauchy-Schwarz/Hölder/Minkowski/Jensen/重排，书后完整解答	https://www.cambridge.org/core/books/cauchys…

编号	资料	类型	获取方式	说明	链接
B034	Problems in Mathematical Analysis（全三卷）— W. J. Kaczor & M. T. Nowak 元数据已补	习题集	付费	全是证明题且逐题完整解答：实数与序列/连续与微分/积分	https://www.ams.org/stml/004
B035	数学分析中的典型问题与方法 — 裴礼文	习题集	付费	中文互补：一致收敛、积分不等式与泰勒余项放缩	https://book.douban.com/subject/35339701/
B036	数学分析习题课讲义 — 谢惠民	习题集	付费	中文互补：一致收敛、积分不等式与泰勒余项放缩	https://book.douban.com/subject/30389024/
B037	Problems and Theorems in Linear Algebra — V. V. Prasolov 元数据已补	习题集	付费	AMS 矩阵证明宝库：秩不等式、迹不等式、特征值交错定理，均附解答	https://bookstore.ams.org/MMONO/134
B038	高等代数中的典型问题与方法 — 李志慧、李永明（指南原标樊恽、刘合国） 元数据已补	习题集	付费	中文互补：不变子空间、若尔当标准形与二次型正定性证明	https://www.ecsponline.com/goods.php?id=29259
B039	One Thousand Exercises in Probability — Geoffrey Grimmett & David Stirzaker 元数据已补	习题集	付费	1000 道概率/随机过程证明题完整推导：条件期望、马氏链收敛、鞅与停时	https://academic.oup.com/book/52798
B040	Exercises in Functional Analysis — Constantin Costara & Dumitru Popa	习题集	付费	Banach/Hilbert 空间、算子范数与紧算子谱理论证明习题集	https://link.springer.com/book/10.1007/978-9…

S13 纯数学教材

> 实分析/抽象代数/测度概率/微分几何与最优传输

编号	资料	类型	获取方式	说明	链接
B010	Linear Algebra Done Right — Sheldon Axler 留档 1 份	教材	开放获取	摒弃行列式直接建谱定理，官方开放获取 PDF + 作者习题讲解视频	https://linear.axler.net/
B016	Measure, Integration & Real Analysis — Sheldon Axler 留档 1 份	教材	开放获取	现代实分析/测度论：Lebesgue 积分、控制收敛、L^p 与 Hilbert 空间（开放获取）	https://measure.axler.net/
B018	Probability: Theory and Examples — Rick Durrett 留档 1 份	教材	开放获取	测度论概率标准教材：大数定律/CLT/条件期望/鞅收敛（作者官网免费）	https://services.math.duke.edu/~rtd/PTE/
B028	Functional Analysis — Gerald Teschl 留档 1 份（补档）	教材	开放获取	维也纳大学研究生教材：Banach 代数、紧算子谱理论、无界算子（作者主页免费）	https://www.mat.univie.ac.at/~gerald/ftp/boo…
B041	Real Analysis: Modern Techniques and Their Applications — Gerald B. Folland	教材	付费	测度论与实分析标杆：Radon 测度、Riesz 表示、L^p 插值、分布理论	https://www.wiley.com/en-us/Real+Analysis%3A…
B042	Princeton Lectures in Analysis（第 3 卷 Real Analysis & 第 4 卷 Functional Analysis）— Elias M. Stein & Rami Shakarchi 元数据已补	教材	付费	从 Lebesgue 微分定理、Hausdorff 维数到 Banach 三大定理、紧算子谱分解与无界自伴算子	https://www.degruyterbrill.com/document/doi/…
B043	Functional Analysis, Sobolev Spaces and Partial Differential Equations — Haïm Brezis	教材	付费	布尔巴基风格：弱拓扑、自反空间与变分不等式极通透	https://link.springer.com/book/10.1007/978-0…

编号	资料	类型	获取方式	说明	链接
B044	Problems and Theorems in Analysis（全两卷）— George Pólya & Gábor Szegő	习题集	付费	纯数学分析训练圣经：按逻辑链编排的纯证明题（附完整解答）	https://link.springer.com/book/10.1007/978-3…
B045	A Hilbert Space Problem Book — Paul R. Halmos	习题集	付费	250 个纯概念证明命题 + 提示 + 完整证明：有界/紧/正规算子与不变子空间	https://link.springer.com/book/10.1007/978-1…
B046	Linear Algebra — Kenneth Hoffman & Ray Kunze	教材	付费	纯数系线代试金石：商空间、对偶空间、若尔当/有理标准形、多线性代数与张量积	https://archive.org/details/linearalgebra00h…
B047	Advanced Linear Algebra — Steven Roman	教材	付费	研究生级抽象线代：自由模、半单算子、Hilbert 空间算子谱分解与仿射几何	https://link.springer.com/book/10.1007/978-0…
B048	Linear Representations of Finite Groups — Jean-Pierre Serre	教材	付费	Serre 薄册经典：特征标理论与不可约表示分解，等变网络权重约束的理论来源	https://link.springer.com/book/10.1007/978-1…
B049	Exercises in Algebra（《代数学题集》，莫斯科大学教材配套）— A. I. Kostrikin 等	习题集	付费	俄罗斯纯数学学派：群/环/域/模/表示论/多重线性，多为结构证明题	https://book.douban.com/subject/30323073/
B050	代数学引论 — A. I. Kostrikin（中译本）	教材	付费	Kostrikin 教材中译本，可与《代数学题集》配套使用	https://book.douban.com/subject/2036531/
B051	Linear Algebra Problem Book — Paul R. Halmos	习题集	付费	纯代数结构证明题库：向量空间公理推演与线性变换本质	https://www.cambridge.org/core/books/linear-…
B052	Probability with Martingales — David Williams	教材	付费	剑桥纯数学学派：σ-代数、单调类定理、鞅收敛、一致可积、Radon-Nikodym	https://www.cambridge.org/highereducation/bo…
B053	Probability and Measure — Patrick Billingsley	教材	付费	测度论与概率论完全融为一体的典范：弱收敛与遍历理论证明无可替代	https://www.wiley.com/en-us/Probability+and+…
B054	Brownian Motion, Martingales, and Stochastic Calculus — Jean-François Le Gall	教材	付费	沃夫奖得主：连续时间鞅、布朗运动构造、Itô 积分、Girsanov、SDE	https://link.springer.com/book/10.1007/978-3…
B055	Counterexamples in Probability — Jordan M. Stoyanov 元数据已补	教材	付费	用反例厘清几乎必然收敛/依概率收敛/L^p 收敛/独立与不相关的严格边界	https://store.doverpublications.com/products…
B056	Problems in Probability — Albert N. Shiryaev	习题集	付费	柯尔莫哥洛夫学派权威习题集：测度论概率、特征函数、鞅与马尔可夫过程（附解答）	https://link.springer.com/book/10.1007/978-1…
B057	Introduction to Smooth Manifolds — John M. Lee	教材	付费	光滑流形公理化黄金标准：切向量/余切丛/微分形式/Stokes/李群，自带证明习题	https://link.springer.com/book/10.1007/978-1…
B058	Riemannian Geometry — Manfredo P. do Carmo 元数据已补	教材	付费	黎曼度量、Levi-Civita 联络、测地线变分公式与截面曲率经典教材	https://web.archive.org/web/20180330191141/h…
B059	Topics in Optimal Transportation — Cédric Villani 元数据已补	教材	付费	Kantorovich 对偶、Brenier 定理与 Wasserstein 空间梯度流理论	https://bookstore.ams.org/gsm-58
B060	Functional Analysis — Theo Bühler & Dietmar A. Salamon 留档 1 份（补档） 元数据已补	教材	开放获取	ETH 研究生泛函分析：Hahn-Banach、一致有界、开映射/闭图像、弱拓扑、Fredholm 算子（作者主页免费）	https://people.math.ethz.ch/~salamon/PREPRIN…
B061	Abstract Algebra: Theory and Applications — Thomas W. Judson 留档 1 份（补档）	教材	开放获取	GNU FDL 开源抽象代数：群同态/Sylow/环与理想/PID/Galois，每章大量结构证明题	https://judsonbooks.org/

编号	资料	类型	获取方式	说明	链接
C015	Terence Tao UCLA Math 245A/245B/245C 讲义 留档 1 份	讲义	开放获取	陶哲轩实分析/泛函/调和分析讲义，正文内嵌数百道 Exercise	https://terrytao.wordpress.com/category/teac…
B062	Group Theory & Fields and Galois Theory — J. S. Milne 留档 1 份	讲义	开放获取	作者主页免费研究生代数讲义：证明言简意赅	https://www.jmilne.org/math/CourseNotes/
B063	Introduction to Representation Theory — Pavel Etingof 等（MIT 研究生讲义） 留档 1 份	讲义	开放获取	arXiv 免费讲义：Schur 引理/Maschke 定理/特征标/箭图与李代数表示，自带证明习题	https://arxiv.org/abs/0901.0827
B064	Stochastic Calculus（Princeton ORF 526 研究生讲义）— Ramon van Handel 留档 1 份（补档）	讲义	开放获取	从滤子/局部鞅/二次变差严格构建 Itô 积分、Girsanov 与 SDE 强弱解理论	https://web.math.princeton.edu/~rvan/
B065	Amir Dembo 斯坦福测度论概率讲义（Math 230A / Stat 310） 留档 1 份（补档）	讲义	开放获取	弱收敛（Portmanteau/Prokhorov）、大偏差原理与布朗运动构造，证明细节完整	https://adembo.su.domains/
B066	Lectures on the Geometry of Manifolds — Liviu I. Nicolaescu 留档 1 份	教材	开放获取	从光滑流形/微分形式/Stokes 到联络、曲率、Riemann 几何与 Hodge 理论（作者主页免费）	https://www3.nd.edu/~lnicolae/Lectures.pdf
B067	Optimal Transport: Old and New — Cédric Villani	教材	部分免费	最优传输百科全书：Kantorovich 对偶、Brenier 极分解、Wasserstein 完备性与 Otto 微积分	https://link.springer.com/book/10.1007/978-3…

S14 开源纯数学与题库

> 开放获取纯数学教材 + MIT OCW/剑桥 Tripos/arXiv 题库

编号	资料	类型	获取方式	说明	链接
C015	Terence Tao UCLA Math 245A/245B/245C 讲义 留档 1 份	讲义	开放获取	陶哲轩实分析/泛函/调和分析讲义，正文内嵌数百道 Exercise	https://terrytao.wordpress.com/category/teac…
B063	Introduction to Representation Theory — Pavel Etingof 等（MIT 研究生讲义） 留档 1 份	讲义	开放获取	arXiv 免费讲义：Schur 引理/Maschke 定理/特征标/箭图与李代数表示，自带证明习题	https://arxiv.org/abs/0901.0827
B068	18.100B / 18.102 / 18.701 / 18.702 / 18.965 题库（MIT OCW）	习题集	开放获取	MIT OCW 纯数学核心课程题库：作业（Problem Sets）与考试/复习材料；OCW 现行版本未公开官方解答，解答需自行推导或另寻	https://ocw.mit.edu/
C009	MIT OCW 18.100B（Real Analysis） 留档 36 份（OCW）	课程	开放获取	MIT 公开课：《数学分析原理》水准的度量空间/紧性/一致收敛证明题（OCW 现行版本只公开题单，未公开官方解答）	https://ocw.mit.edu/courses/18-100b-real-ana…
C010	MIT OCW 18.102（Introduction to Functional Analysis） 留档 37 份（OCW）	课程	开放获取	Banach/Hilbert 空间、有界/紧算子证明题（OCW 现行版本只公开题单与考试材料）	https://ocw.mit.edu/courses/18-102-introduct…
C011	MIT OCW 18.701（Algebra I） 留档 16 份（OCW）	课程	开放获取	群作用与线性代数纯证明题（OCW 只公开题单）	https://ocw.mit.edu/courses/18-701-algebra-i…
C012	MIT OCW 18.702（Algebra II） 留档 10 份（OCW）	课程	开放获取	表示论、模与域扩张纯证明题（OCW 只公开题单）	https://ocw.mit.edu/courses/18-702-algebra-i…

编号	资料	类型	获取方式	说明	链接
C013	MIT OCW 18.965（Geometry of Manifolds） 留档 32 份（OCW）	课程	开放获取	流形、张量场与微分形式证明题（OCW 只公开题单）	https://ocw.mit.edu/courses/18-965-geometry-…
C014	剑桥大学数学系 Tripos Part II / Part III Example Sheets 留档 72 份（补档）	习题集	开放获取	世界顶尖纯数研究生难度专题证明题单	https://www.maths.cam.ac.uk/undergrad/exampl…

三、指南摘要

五个低算力顶会课题

课题	核心痛点	研究切入点	实验与算力配置	目标顶会
一、推理模型"思维预算"动态控制与过度思考抑制	Long-CoT 推理模型面对简单问题易陷入"过度思考（Overthinking）"，生成大量冗余验证 Token 致延迟飙升；高难问题又可能过早停止探索（p5）	设计轻量甚至训练-free 的推理期早停（Early-Exiting）或动态路由：监测隐藏层状态、注意力熵值变化或思维链中自我否定词频，在模型锁定正确答案的瞬间动态截断思维链，或自适应分配推理计算预算（Test-time Scaling）（p5）	仅需 1 张 24GB/48GB 显卡；直接基于开源 7B/8B/14B 蒸馏推理模型，在 MATH-500、AIME、GPQA 等基准上跑推理评估，对比 Token 消耗量（降低 30%-50%）与准确率保持率（p5）	NeurIPS、ICLR、ACL、EMNLP
二、RLVR 高价值样本筛选与课程学习策略	用 GRPO 等算法做推理强化学习时，海量样本中存在大量"全对（模型早就会了，梯度为零）"和"全错（太难，采样不到正奖励）"的无效数据，严重浪费训练算力（p6）	探索数据高效（Data-Efficient）后训练配方：基于模型当前能力边界（Pass@k 分布方差、输出分布困惑度或思维链结构复杂度）做在线动态数据筛选（Online Data Selection）或课程学习，证明仅用 10%-20% 核心训练数据即可达到全量数据训练的 RL 效果，甚至缓解语言退化（p6）	2-4 张 4090 或单张 A100/H100（配合 LoRA + vLLM/SGLang 加速采样；3B/7B 模型在几千条高质量数学/代码数据上只需数小时至一两天即可跑完一轮完整实验）（p6）	ICML、NeurIPS、ICLR
三、无需训练的 KV Cache 极端压缩与长上下文注意力稀疏化	处理长文档、多轮 Agent 历史记录或长视频时，KV Cache 显存占用成为推理吞吐量最大瓶颈；现有静态驱逐策略（如只保留首尾 Token）经常误删关键"大海捞针"信息（p6）	结合注意力机制底层几何或频谱特征，提出自适应层级/头级（Layer-wise / Head-wise）KV Cache 压缩与低比特混合量化：区分"局部语法头"与"全局检索头"，按不同注意力头分配差异化缓存保留预算，把 KV Cache 显存压缩至 15%-20% 的同时保证长上下文推理与多跳问答几乎无损（p6）	1 张 24GB 显卡即可完成全部实验，完全不需要训练模型；编写自定义 Attention Hook 或基于 vLLM/HuggingFace 改造推理代码，在 LongBench、RULER、Needle-in-a-Haystack 上做充分消融与吞吐量测试（p6）	ICLR、NeurIPS、MLSys
四、多智能体协作中的上下文退化诊断与自适应记忆压缩	复杂软件工程（SWE）或深度研究智能体（Deep Research Agent）执行几十步工具调用后上下文迅速膨胀，出现"注意力迷失"、重复犯错以及工具调用格式崩溃（p6）	针对长程交互轨迹（Trajectory）设计基于因果归因或状态摘要的结构化记忆机制（Structured Memory / Context Folding）：自动折叠已完成子任务细节、保留关键环境状态与失败教训，并构建轻量重放/反思策略，提升 7B-14B 小模型作 Agent 基座的任务完成率（p6）	1-2 张 24GB 显卡（本地部署 7B/8B/14B 开源模型作 Agent 核心，配合局部轻量 LoRA 微调或纯推理策略优化），在 SWE-bench Lite、OSWorld 或 WebArena 子集上验证（p6）	ACL、ICLR、NeurIPS（Datasets & Benchmarks 或 Main Track）
五、大模型内部表征干预（Activation Steering）与知识边界校准	模型面对超出自身知识边界的问题时产生严重幻觉，或多模态场景下过度依赖语言先验而忽略视觉输入（视觉幻觉）；传统微调成本高且容易破坏通用能力（p7）	分析模型残差流（Residual Stream）或特定注意力头的激活方向，寻找控制"事实确定性""拒答倾向"或"视觉注意力权重"的低维线性子空间；设计推理期自适应激活转向（Adaptive Activation Steering）或轻量级探针干预，在不修改权重的前提下显著降低幻觉率并提升不确定性校准（Calibration）指标（p7）	单张 24GB 显卡极速出结果；单次前向传播提取特征与验证只需几分钟，非常适合快速迭代想法、绘制精美的表征可视化图（审稿人非常喜欢这类机理剖析工作）（p7）	NeurIPS、ICLR、EMNLP

五课题饱和度

课题方向	饱和度	已饱和的"红海做法"（审稿人易拒稿）	未饱和的"蓝海切入点"（高命中率）
课题三：KV Cache 极端压缩	极高 ★★★★★	针对"长输入、短输出"的常规 Token 驱逐（仅换个注意力打分公式在 LongBench 上刷分，ACL 2026 已有系统综述总结此类方法）（p8）	面向推理模型长输出（Long-CoT）的阶段感知异构压缩；结合 Triton 算子融合实现真实吞吐量（Throughput）提升，而非仅在 Transformers 里模拟稀疏率（p8）
课题一：推理模型"思维预算"控制	高 ★★★★☆	简单的基于 Token 熵值、全局置信度阈值早停，或单纯用 Prompt 限制思维链长度（相关综述与基准已密集涌现）（p8）	解决"伪收敛"与"过度思考导致对改错"问题：利用内部特定注意力头（如思维划分头）做细粒度步骤级干预；或转向多模态长思维链（Multimodal CoT）的跨模态冗余抑制（p8）
课题五：内部表征干预（Activation Steering）	中高 ★★★☆☆	全局静态、固定强度的线性转向向量（审稿人已熟知其会破坏通用能力、对超参敏感，且无法强行注入新事实知识）（p8）	动态闭环与条件门控干预（引入控制论反馈机制或 Token 级动态拒止，仅在越界风险出现时介入）；在 VLM 中抑制语言先验过强导致的视觉幻觉（p8）
课题二：RLVR 样本筛选与课程学习	中 ★★★☆☆	仅过滤掉"全对或全错（零梯度）"样本的静态筛选（已成为主流 RL 训练框架的默认工程技巧）（p8）	突破推理边界（Expanding Reasoning Boundary）：传统 GRPO 往往只提高 Pass@1 却导致大 k 下 Pass@k 探索覆盖率萎缩；研究如何用极少高难样本配合逆向课程解锁模型原本不会的难题（p8）
课题四：多智能体上下文退化与记忆压缩	中低 ★★☆☆☆	每隔固定轮次调用大模型做粗暴的纯文本摘要（Summarization），导致关键报错信息或约束丢失（p8）	可恢复的状态近似（Recoverable State Approximation）与潜在/KV 级跨步记忆复用：压缩 50+ 步长程工具调用轨迹时保留因果回溯指针，系统性诊断并修复智能体失败模式（p8）

30 天学习计划

周次	周目标	关键学习与动手内容	验收标准	关键工具
第 1 周（Day 1-7）	掌握 PyTorch 核心机制与训练闭环，彻底摆脱对封装训练框架的依赖	Day 1-2 张量内存布局、广播机制、view/reshape/permute/transpose；Day 3-4 autograd、detach、nn.Module、AdamW，手写含 LayerNorm 与残差的两层全连接小网络并打印 .grad；Day 5-7 完整训练循环（Dataset/DataLoader、交叉熵、梯度裁剪、torch.amp）（p11）	不查文档写出包含 optimizer.zero_grad()、loss.backward()、optimizer.step() 与 model.eval() 的标准 PyTorch 训练模板，并能独立排查 RuntimeError: shape mismatch（p11）	PyTorch、einops、nvidia-smi
第 2 周（Day 8-14）	手撕 Transformer 与现代大模型架构，从底层矩阵运算理解 Decoder-only 大模型的每个组件	Day 8-9 BPE Tokenizer、词嵌入、RoPE 旋转位置编码数学直觉与一维旋转矩阵手写；Day 10-11 用 torch.matmul 与 softmax 手写带因果掩码的 GQA 模块；Day 12-14 RMSNorm、SwiGLU、Pre-Norm 残差结构与 KV Cache 原理，跟 nanoGPT 拼装微型 GPT 完成预训练与逐 Token 自回归生成（p11-12）	能在白板或空白 .py 文件独立手写出 GQA 注意力核心前向代码，并用自己的话解释为什么推理时引入 KV Cache 能大幅降低时间复杂度（p12）	PyTorch、Karpathy nanoGPT
第 3 周（Day 15-21）	熟练驾驭 HuggingFace 工具链，学会加载开源大模型并提取内部表征（科研必备技能）	Day 15-16 AutoTokenizer / AutoModelForCausalLM / AutoConfig 三件套与 apply_chat_template，加载 Qwen2.5-7B-Instruct（或 3B）；Day 17-18 FP16/BF16、bitsandbytes 4-bit/8-bit 量化、temperature/top_p/do_sample，对比显存与生成速度；Day 19-21 用 output_hidden_states=True 与 register_forward_hook 提取激活张量，并部署 vLLM 跑通离线批量推理（p12）	能在单卡上流畅运行开源大模型的批量推理，且掌握用 PyTorch Hook 拦截并打印中间层激活值的能力（复现表征干预、注意力分析类顶会论文的核心基础）（p12）	transformers、bitsandbytes、vLLM、HuggingFace 镜像源
第 4 周（Day 22-30）	掌握 LoRA / QLoRA 原理，完成从数据清洗、模型微调到效果评测的完整闭环	Day 22-23 指令微调（SFT）数据格式 Alpaca / ShareGPT、Loss Masking，准备 500-1000 条垂直任务数据；Day 24-25 LoRA/QLoRA 数学原理（低秩旁路矩阵 W + BA）、秩 r / lora_alpha / target_modules 与 peft 库，打印可训练参数量占比（通常仅占总参数量的 0.5%–2%）；Day 26-28 BF16 + gradient_checkpointing + trl 的 SFTTrainer 跑通微调；Day 29-30 merge_and_unload、在独立 100 条验证集上对比原始基座与微调模型，代码与实验记录整理至 GitHub（p12-13）	拥有一个亲手微调、有明确量化对比指标（如格式准确率从 62% 提升至 96%）的 GitHub 仓库，具备正式进组或复现顶会论文基线代码的工程能力（p13）	peft、trl（SFTTrainer）、bitsandbytes、SwanLab/TensorBoard

四、PDF 下载留档

类别	文件数	大小	说明
主库 downloads/pdf/	32	147.16 MiB	教材/讲义/论文/技术报告（指南点名的开放获取单册）
MIT OCW 题单/考试/讲义	131	25.81 MiB	五门纯数学课程（18.100B/18.102/18.701/18.702/18.965）
剑桥 Tripos 题单	72	8.98 MiB	34 门课（原有 3 门 + 补档 31 门；补档经 Wayback 快照）
其它补档	7	47.38 MiB	Seed 报告、Teschl/Boyd/ETH/Judson/Dembo 等
合计	242	229.34 MiB	全部校验 %PDF + sha256

主库清单：`data/downloads.json`；OCW：`data/recover/ocw_manifest.json`；补档：`data/recover/misc_manifest.json`；逐文件校验：`data/recover/verify.json`；下载目录说明：`downloads/README.md`。

五、缺口与已知问题

1. 11 条书目因出版社反爬抓不到正文，但已用 Crossref / OpenLibrary / Wayback 补齐权威元数据（含 ISBN/DOI/年份）。
2. MIT OCW 现行版本未公开官方解答，留档为题单/考试/讲义；C011、C012 该版本无公开讲义。
3. 剑桥 Tripos 部分题单仅在受保护的校内 Moodle；部分课程当年 PDF 未被 Wayback 收录，改用同课程更早学年题单。
4. B064 van Handel ORF526 原讲义已 404 且无存档，收录作者主页同主题 ACM 217 讲义（已标注替代）。
5. B038 指南原标作者为樊恽、刘合国，权威书目记录为李志慧、李永明（科学出版社 2008），索引已按权威记录更正并保留原标。
6. 付费教材只记录出版社/作者官方页，不下载非授权全文；页面内容与链接可能随官网更新而变化。

六、怎么用

1. 按阶段用，不要按顺序刷：先读每个阶段开头的定位，再挑框架/论文跑通，教材留作系统阅读。
2. 开放获取优先：标“开放获取”的条目可直接下载；已留档的条目同时给出本地 PDF（合计 242 份 / 229.34 MiB）。
3. 多阶段资料是正常的：同一份资料可能在多个阶段出现（如 PyTorch 横跨 S1/S3/S6），以编号为准去重。
4. 付费书先找免费版：作者官网 → arXiv/机构 preprint → 开放课程页 → 图书馆数据库 → Internet Archive。
5. 想快速检索：使用网页版资料索引 frti64.top/ai-resources/（搜索 + 按方向/阶段/获取方式筛选）。

七、数据来源与复现

1. 原稿：《AI学习路径指南》PDF，32 页，Gemini 多轮问答打印稿（无文本层，逐页视觉识读转录）。
2. 补充来源：原始 HTML 导出（Google Docs），保留 PDF 打印时丢失的超链接，用于核对 Sources 与 OCR 忠实度。
3. 索引与数据：ai-learning/data/index.json（144 条资源，18 种类型，14 个阶段）；resources.jsonl / resources.csv 可程序读取。
4. 抓取与校验：scripts/ 目录可复现（build_resources → merge_resolve → apply_source_links → apply_recovery → crawl_group → fetch_pdfs → merge_crawl）。
5. PDF 版本：本文件由 scripts/make_pdf.py 生成（中文版 / 中英对照版），HTML 源见 pdf/html/。