



清华大学深圳国际研究生院
Tsinghua Shenzhen International Graduate School

从0到1:面向初学者的科研经验分享

—

肖劲宇

数据科学与信息技术研究院

2021级计算机技术硕士生

导师：江勇教授

2024年6月14日

sigs.tsinghua.edu.cn

自我介绍

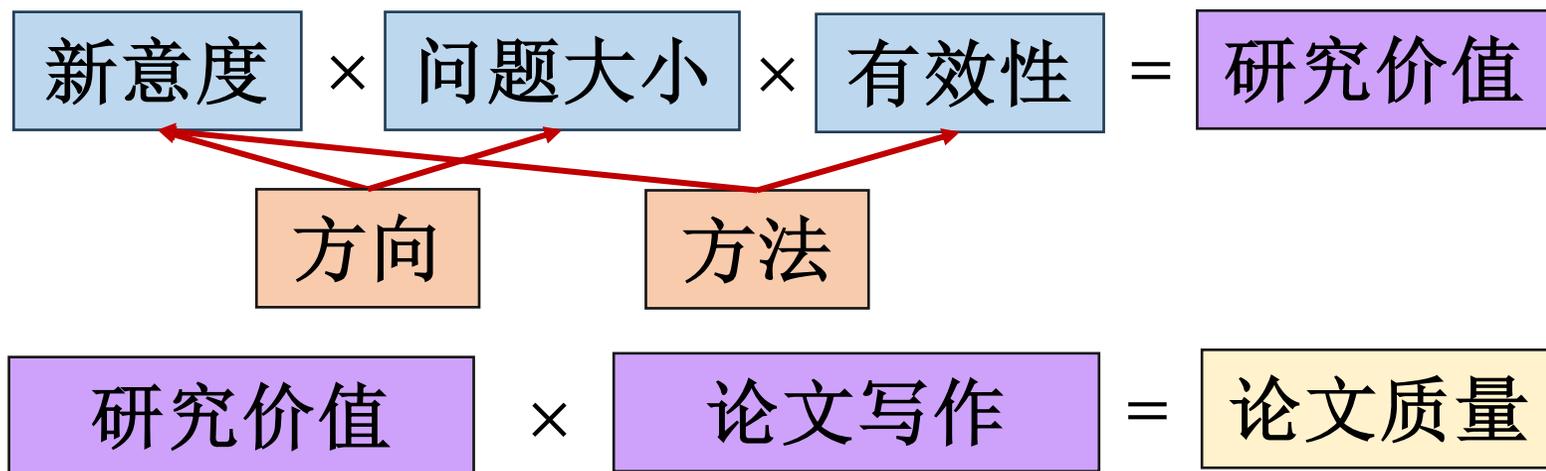


- 研究方向：计算机网络、物联网
- 第一作者论文发表：
 - SIGKDD 2024 (数据挖掘领域顶会, CCF-A)
 - ToN 2023 (计算机网络顶刊, CCF-A)
 - UbiComp 2023 (人机交互和普适计算顶会, CCF-A)
 - AAMAS 2023 (Oral, 多智能体领域顶会, CCF-B)
- 共同作者论文发表
 - SIGMOD 2023、UbiComp 2023、CoNEXT 2022、NSDI 2022、IPCCC 2021
- 其他
 - 清华大学优秀硕士毕业生 (2024)
 - 清华大学SIGS专业实践一等奖 (2024)
 - 清华大学一等奖学金 (2023)
 - 腾讯犀牛鸟精英人才计划 (2023)
 - AAMAS Student Scholarship (2023)
 - 清华大学博智林-惠研英才奖学金 (2022)



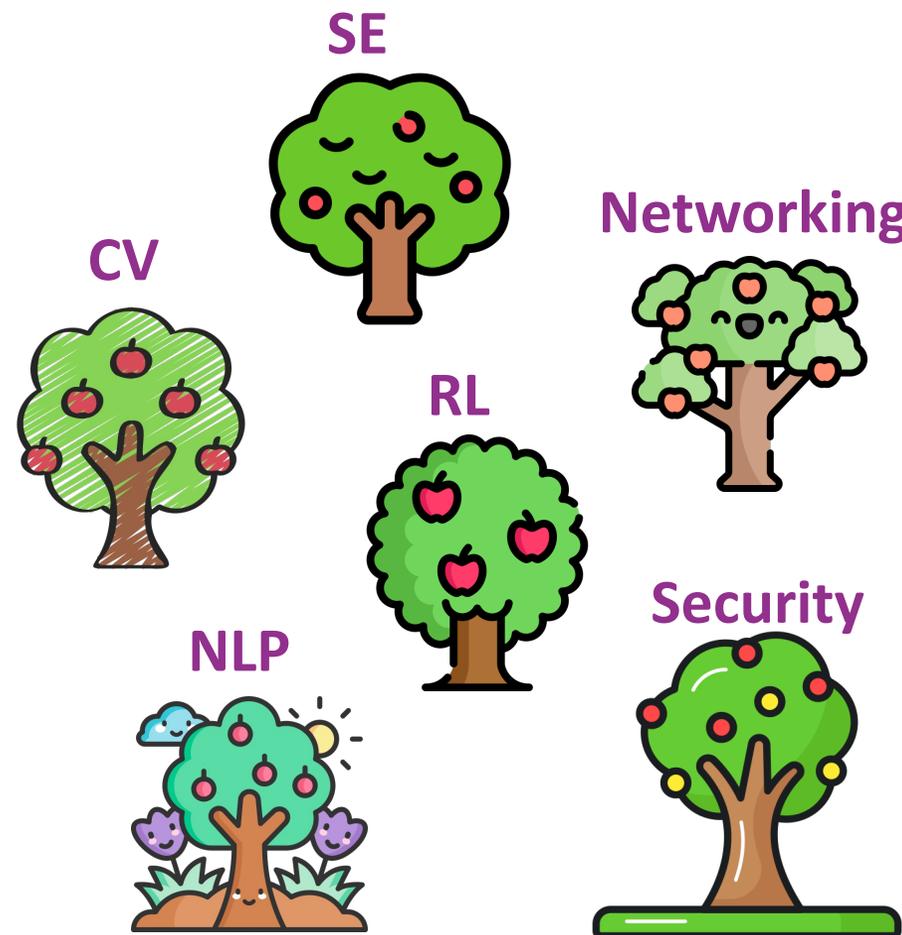
研究价值和质量

- 科研：用有**新意**的方法**有效**地解决一个**研究问题**
- 方向决定了新意度和问题大小，方法决定了新意度和有效性
- 研究价值和论文写作决定了最终的论文质量



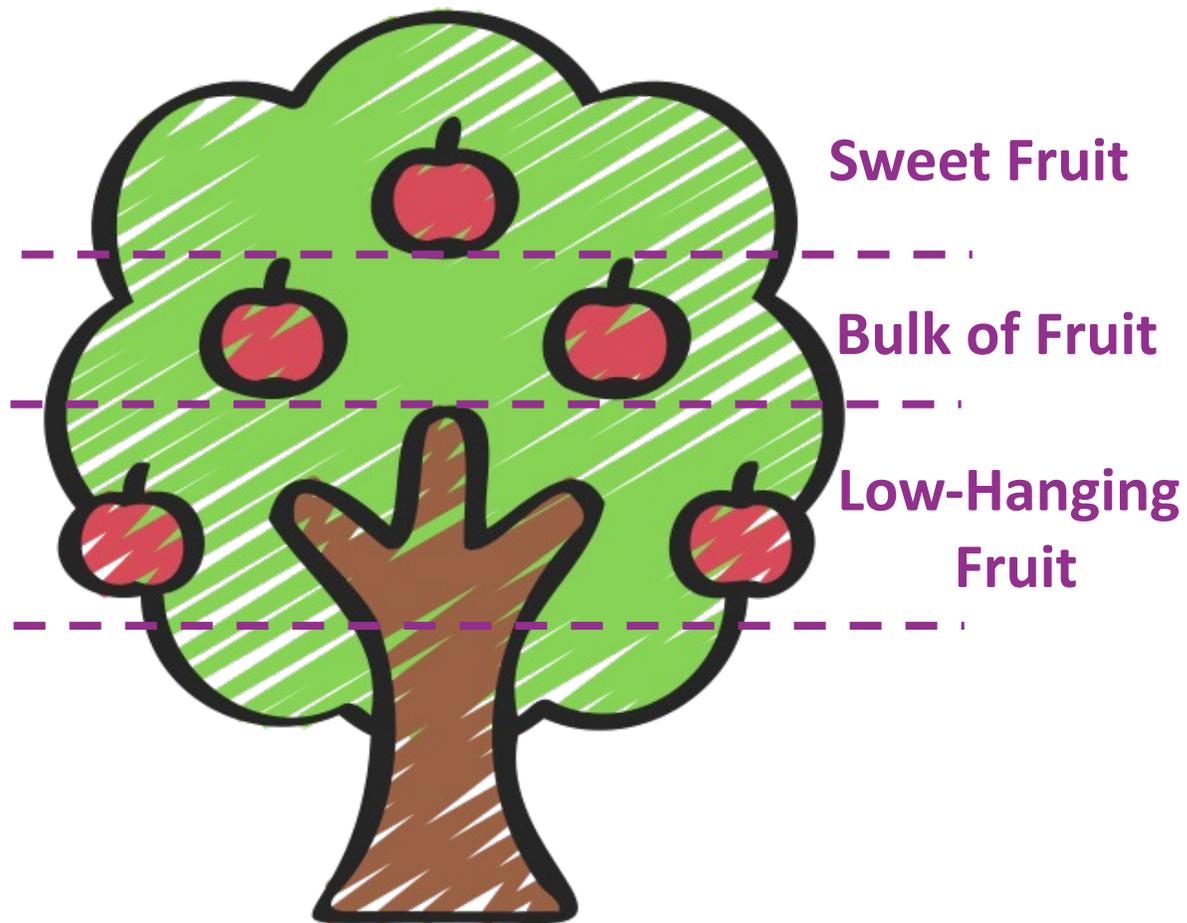
科研方向

- 计算机科研方向
 - AI: CV/NLP/Speech/RL/DM/Theory/Security...
 - 系统: Networking/OS/SE...
 - 交叉: AI for Science...
- 方向选择
 - 个人兴趣?
 - 资源评估?
 - 热门 vs 冷门?
 - ...



科研方向：科研难度 vs 问题难度

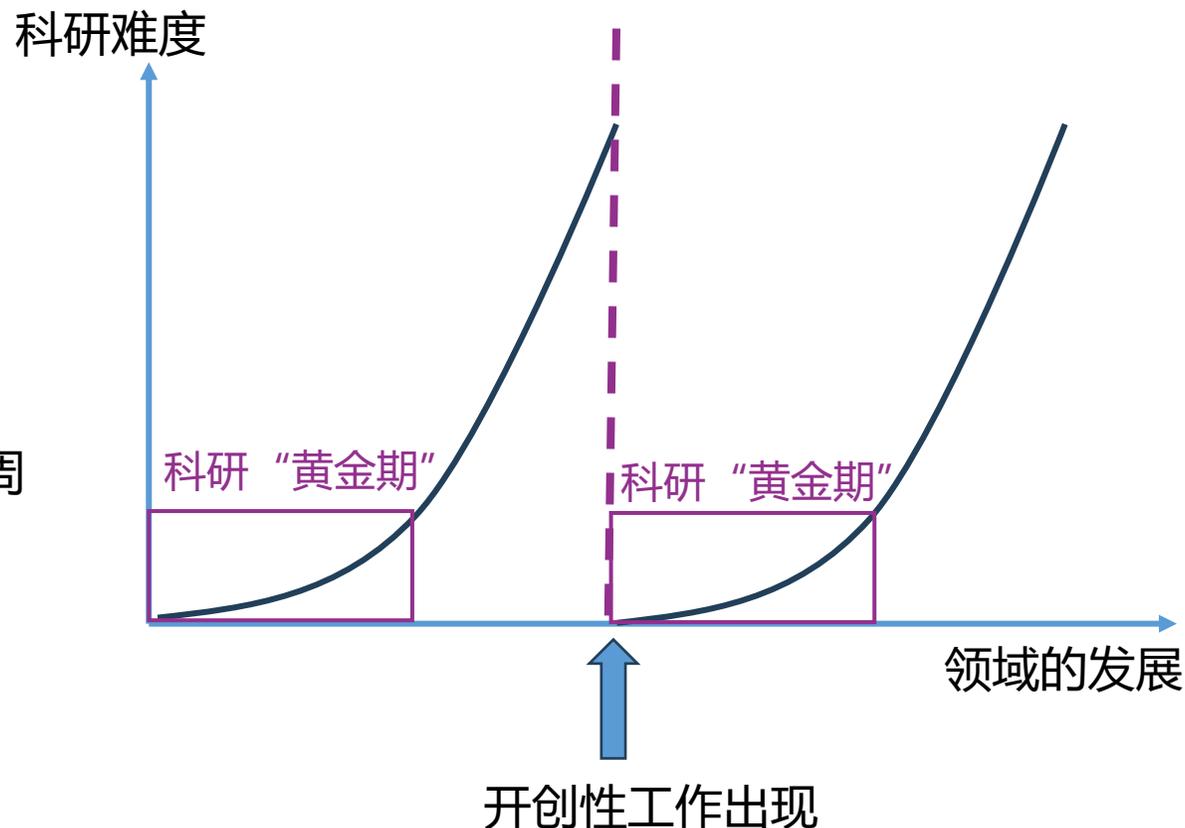
- Low-hanging Fruit、Bulk of Fruit、 Sweet Fruit
- 做科研问题的难度并非和影响力呈正比





科研方向：科研难度 vs 领域发展

- 科研“黄金期”更容易做出创新性的工作
 - ✓ 新的场景/应用
 - ✓ 新的思考角度
- 领域越发展，做科研的难度越大
 - ✓ Low-hanging fruits 所剩无几
- 开创性工作的出现会让领域的发展进入一个新的周期，新的科研“黄金期”涌现
 - ✓ NLP: LLM
 - ✓ Networking: P4 Switches





科研方法：如何定义创新

- Novelty = Complexity?
- Novelty = Difficulty?

Simple and Effective is also novelty

- Novelty = Surprise?

The novelty must be evaluated *before* the idea existed.

- Novelty = Technical Novelty?

New dataset/benchmark/experience

其他研究者未曾提过的新问题

已知问题的新视角新解决方案

更好的分析和理解

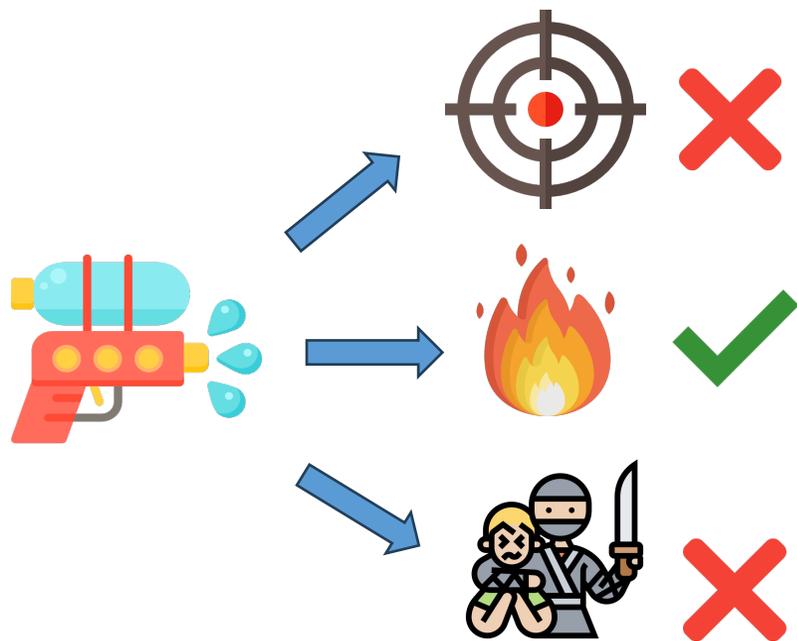
更好的实验结果

论文的创新性等级

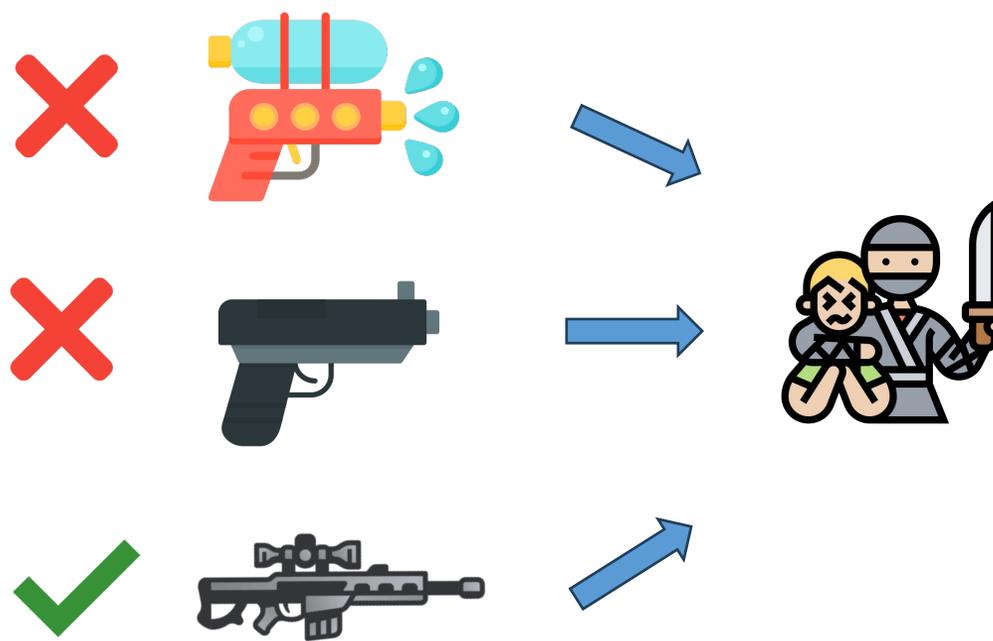


科研方法：方法驱动型 vs 问题驱动型

- 方法驱动型科研：用现有的技术去找一个问题场景
- 问题驱动型科研：针对现有的问题去做适配的技术



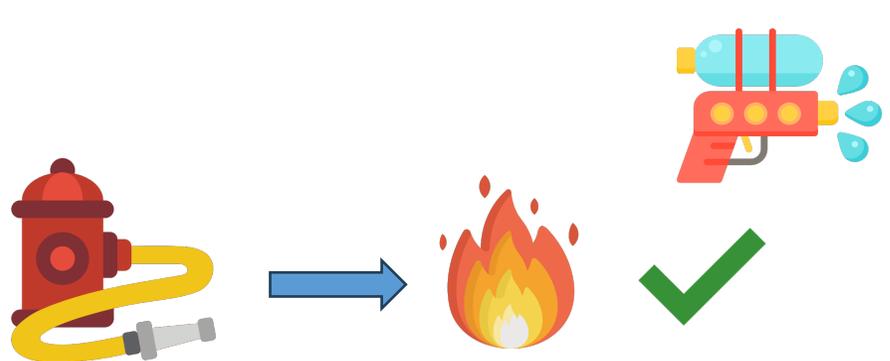
方法驱动/技术驱动



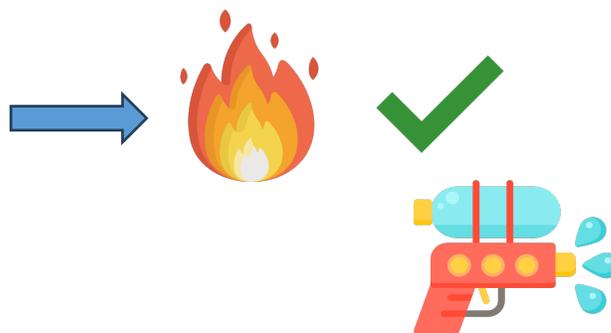
问题驱动/场景驱动

科研方法：如何做创新

- 对于科研初学者：新瓶装旧酒、旧瓶装新酒

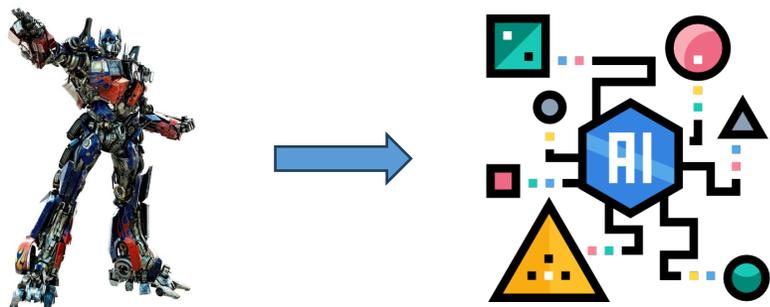


新的技术：更快、更高效



新的应用场景

- 对于成熟的科研工作者：追求新瓶装新酒



新的技术实现新的领域突破

论文写作



- **第一印象很重要**: Title、Abstract、Introduction、Figure/Table

你以为审稿人应该是这样审稿的:

审稿人一定是专家，无所不知。打印出来，仔细研读揣摩数天，对于看不懂的地方反复推敲。即使你的英文写得极其糟糕、即使你的文章组织很混乱、即使你的表述很难看懂，审稿人花费了大量的时间后终于看懂了，他认为你的工作是有意义的，决定给你个**border line**或以上的分数。

审稿人实际上往往是这样审稿的:

他不一定是专家，一直忙于其他事，在**deadline**到来之前一天要完成n篇。审稿时他往往先看题目、摘要，扫一下**introduction**（知道你做什么），然后直接翻到最后找核心实验结果（做得好不好），然后基本确定录还是不录（也许只用5分钟！）。如果决定录，剩下就是写些赞美的话，指出些次要的小毛病。如果决定拒，下面的过程就是细看中间部分找理由拒了。

第一印象定录拒，5分钟内打动审稿人

微博上的佐证

胡云华MSRA



最近有很多论文需要评审，跟同行聊天，得出一个有意思的结论：如果一篇论文在看完abstract和conclusion后还不能判断论文是否有价值的话，基本上这篇论文也就悲剧了。自己试了多次，屡试不爽。最极端的一篇我看了整整两天，全部搞懂作者在说什么后，仍然觉得应该拒掉，就跟只看5分钟得出的结论一致。



胡云华MSRA: 回复@shirlywang1983:我说的是“小论文”，毕业论文之类的评审得少，不好说。好的论文需要准确提炼观点，让读者在尽量短的时间内明白你做了什么，你的贡献是什么。如果自己没想清楚，肯定写不清楚的。当然这个过程很不容易，没有深厚积累谁做不到。(12月5日 09:01)



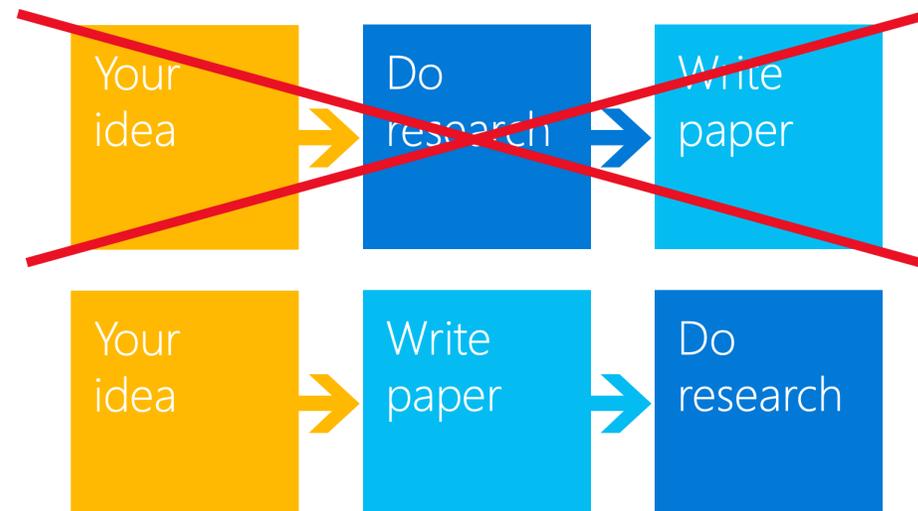
kingdy9: 说明第一印象很重要，也很准确。有了第一印象后再找找文章中值得批判的地方就好了。@朱小燕THU: 悲哀的是，已经感觉到了，但是为了写评语还是要看到底(12月5日 09:38)



王伟DL: 回复@胡云华MSRA:谢谢！我得修正我的观点，很同意“审论文时，abstract和conclusion写不好但内容好的情况少之又少。” (12月5日 14:22)

论文写作

- 论文写作流程
 - Idea → Write paper → Do research
- 故事逻辑清晰：
 - 背景→挑战→方法→实验结果支撑
- 突出贡献，重要的结论和发现
 - 加粗重要的结论
- 多举例
 - 举例比描述更容易让读者理解
- 多向他人寻求改进意见





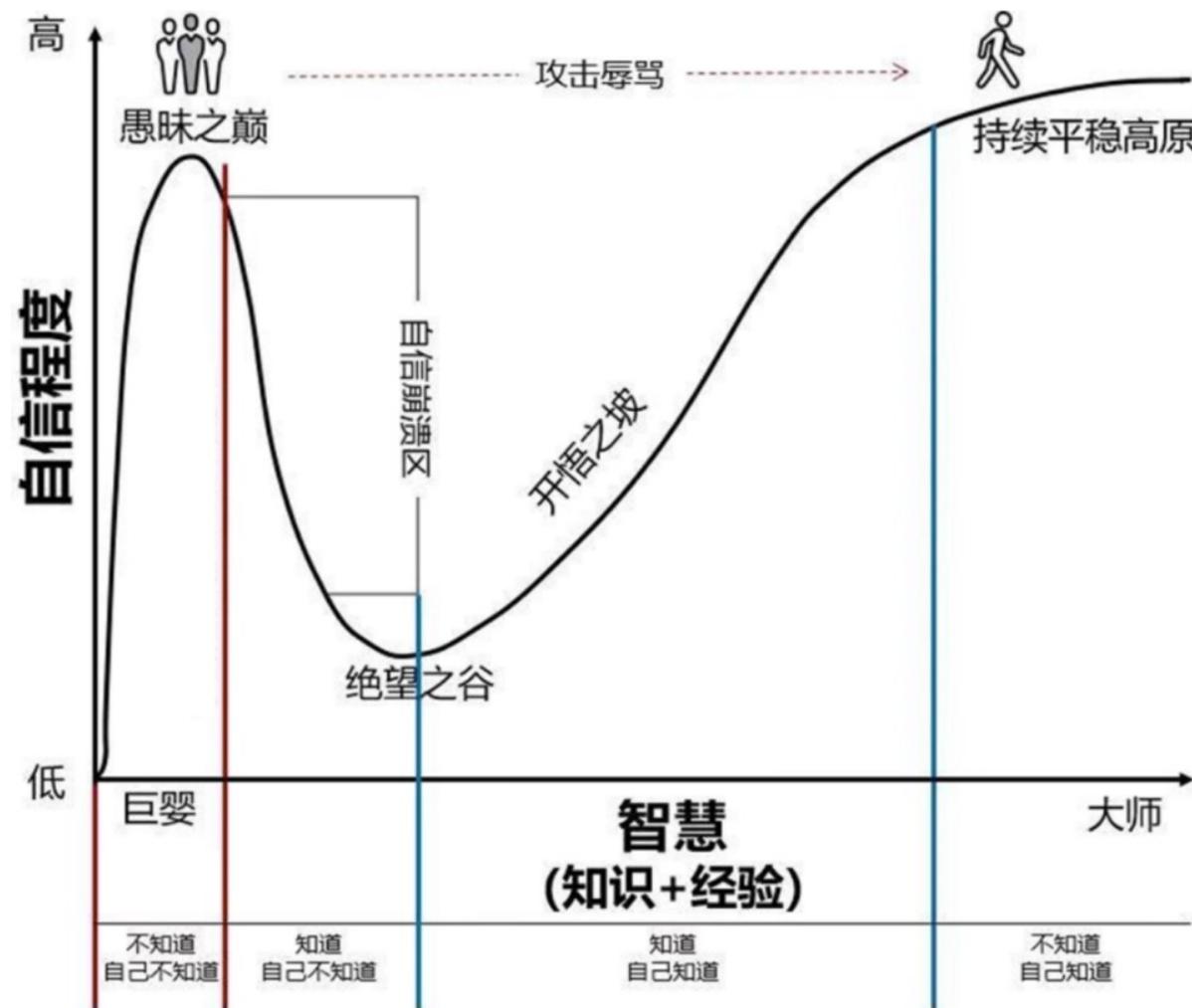
为什么研究生拥有自由的空间和时间，反而写不出论文呢？

- 1. 当你看了一篇文献时：我有一个完美的idea! 😄
- 2. 当你看了十篇文献时：我的idea虽然有瑕疵，但修改一下就完美了! 😊
- 3. 当你看了五十篇文献时：我的idea好像被人做过了，而且设计的也有漏洞，需要再加入新方案，多论证几个点，不过好在我满脑子都是idea。😬
- 4. 当你看了一百篇文献时：我已经把我的idea改的面目全非了，感觉这个想法糟糕透了，为什么不舍弃这个idea呢? 😓
- 5. 当你看了五百篇文献时：我已经有新的idea了，而且是很多个，但不知选哪个好。😎
- 6. 当你看了一千篇文献时：我的idea都是bullshit。😞
- 7. 当你看了三千篇文献时：我要换课题，我们课题组做的都是bullshit。😓
- 8. 当你已经不记得自己看过多少篇文献时：文献都是bullshit。😓



达克效应

- 攀登愚昧之巅：不知道自己不知道
 - ✓ 刚进入一个领域，阅读文献较少
- 陷入绝望之谷：知道自己不知道
 - ✓ 阅读一些文献，发现了知识盲区
- 攀爬开悟之坡：知道自己知道
 - ✓ 阅读大量文献，拓宽知识边界
- 进入持续平稳高原：不知道自己知道
 - ✓ 融会贯通，对领域有了自己的洞见和理解





从科研“小白”到科研“学者”的成长路径

- 阶段一：在引路人的带领下，把一件事情做好
- 阶段二：自己想一个idea，在被告知哪些不能做后，把能做的部分做到极致
- 阶段三：完全独立想一个idea，独立地做完、做好
- 阶段四：能够带领第一阶段的人，引导他把一件事情做好





其他的一些Tips

- 多和领域内、其他领域的学者交流，建立connection
- 重视领域和学科交叉，能带来思维新视角
- 永远不要觉得所有的坑都被填完了，好的技术一直在路上
- 阅读论文：Passive reading → Active reading → Critical reading → Creative reading [1]
- 保持良好的心态，对于初学者来说论文拒稿很正常，关键是如何针对性改进
- 科研工具：
 - 论文检索：google scholar、arxiv、公众号、小红书
 - 论文阅读：kimichat
 - 论文写作：overleaf/texpage, chatgpt
 - 绘图：PPT, icon网站 (flashicon)
 - DDL管理：ccfdeadline (<https://ccfddl.github.io/>)

[1] 沈向洋、华刚《读科研论文的三个层次、四个阶段与十个问题》



课题组介绍： Smart Internet Group (SIG)

- 智能网络组 (SIG) 介绍

- 智能互联网研究组科研平台强大，团队氛围轻松，补助充足。导师组包括：清华大学江勇教授、鹏城国家实验室李清副研究员等，在IEEE JSAC、IEEE/ACM ToN、IEEE TC、IEEE TMC、IEEE TPDS、IEEE TMM、USENIX Security、ACM SIGMOD、ACM SIGKDD、UbiComp、NeurIPS、ACM CoNext、ACM MM、ACM WWW、IEEE INFOCOM、IEEE ICNP等会议期刊发表论文100多篇。研究组联合鹏城国家实验室、南方科技大学未来网络研究院等相关团队，共同探索未来网络技术，并与字节、腾讯、华为等有良好的合作关系，研究组学生有大量机会参与企业的合作研究工作。近五年培养的学生拿到Offer王要包括华为（天才少年）、Google、Microsoft、字节跳动、腾讯、阿里等。

- SIG课题组公众号：



欢迎关注「智能互联网研究组」

Smart Internet Group (SIG)

- SIG课题组网站：

- <https://smartinternet.group/>



清华大学深圳国际研究生院
Tsinghua Shenzhen International Graduate School

Thanks! 祝各位同学夏令营顺利!

分享人：肖劲宇

邮箱：jy-xiao21@mails.tsinghua.edu.cn

个人主页：<https://whalexiao.github.io/>

微信：右下角二维码，欢迎大家批评指正

sigs.tsinghua.edu.cn

