

青春
上海

让孤灯黄卷的独守，化为云端星火的共耀

古籍数字化背后的年轻人与年轻技术

落日熔金，宿舍渐渐染上暖色。23岁的刘尔君点亮电脑屏幕，凝神审校《大般涅槃经》；21岁的彭显涵也在逐字校勘《永乐大典》的字句；刚结束一天工作的90后陆天丰正埋首于《周易》的文本推敲……在人工智能与大众校对的双重推动下，古籍数字化不再是冷门事业，它正吸引越来越多年轻人走进墨香与代码交织的世界。通过“识典古籍”这样的数字平台，他们既走向古籍，也让古籍走向更多人。

青年报记者 刘秦春



“曾经，古籍藏于深阁；如今，存于云端。我们每一个人，都可以成为它们的守护者。”

聆听古籍的心跳

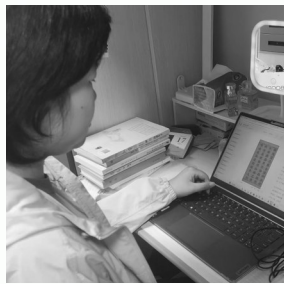
刘尔君注视着屏幕上泛黄书页的图像，逐行校对千年以前的文字。她所用的“识典古籍”平台，由字节跳动公益推出，已免费开放超3万部古籍，悄然搭建起一座连接往昔与当下的桥梁。

作为上海大学中国古代文学专业的研一升研二学生，刘尔君的初心早在童年时观看文学改编影视、翻阅明清小说时便已种下。“越长越大发现，文学不只是故事，更沉淀着历史，甚至映照着当代现实。透过它，我得以窥见古人在历史关头的抉择，从他们的生命轨迹中获得启迪。”她说。

今年四月，她偶然看到“识典古籍”招募“我用AI校古籍”志愿者的消息，毫不犹豫报了名，“这既是我的专业也是我的热爱。”顺利入选后，从“初阶组”做到“进阶组”，她累计校对近70卷古籍，其中包括7卷《永乐大典》。

“我其实挺佛系的”，她笑言，“有同学一人就校了200多卷，我目前才70卷，不算多。”她参与的首批任务以佛经为主，如《金刚经》《楞严经》《华法经》等佛教经典古籍。“佛经用字繁复，异体字多，初看如天书，但越难越有趣。”

有一次，她遇到一句诗：“小桃明淑？一树出低垣”。系统未能识别出问号处的字。“古人写诗讲究对仗，‘小桃’是植物，后面也应匹配名词。根据图片字形看明显是‘果’旁，我锁定了‘菓’与‘菓’。”她翻阅《本草纲目》，印证“菓”可表植物丛生，再援引《和子由记园中草木》中“苍苔菓”的用法来佐证，最终推定此处应为“菓”字。“那一刻，就像破案一样。”她把这次的校对心得写进了自己的笔记里，随着任务的增多，这份笔记内容也越来越多。



刘尔君

作为在上海求学的学子，刘尔君对这座城市的古籍资源如数家珍。“上海的古籍收藏非常丰富，有的资源只有在上海才能找到，这给我的学术研究带来了很大便利。”刘尔君说，更让她感动的是上海在古籍文献服务方面的用心。例如，上海图书馆的古籍文献传递系统允许读者在线申请，馆员帮忙复印后快递到家，“即使人不在上海，也能读到珍贵的版本。”

在她看来，“识典古籍”最大的意义是打破了古籍的“神秘感”。“从前，古籍像是学者的专属。《史记》不是想读就能读的，得去图书馆查目录、递申请。如今平台上有多种版本、原文注释、高清图像，一目了然，并且全部免费开放，人人可读。”她还特别提到平台的交互性：“你不仅能读，还能参与校对，指出错误，甚至和老

师、同学在群里讨论。这不再是单向的知识接收，而是一场跨越时空的对话。”

她困惑于《左传》注文与正文的界限限时，是群里北大数字人文研究中心余红雷老师和其他审核同学的耐心解答，帮她理清了文献体例的逻辑。“我们还定期开线上分享会，有人讲校对技巧，有人分享考证心得。”她说，“这种‘教学相长’的感觉，特别温暖。”

被问及“古籍对普通人有什么用”时，刘尔君的答案温柔而坚定：“它告诉我们从哪里来。那些诗文与思想，不是冰冷的文字，而是古人的呼吸与心跳。数字化，让这些心跳重新被听见。”

“很多古籍毁于战火与灾害，令人痛心。但数字化赋予它们另一种‘永生’。”她说，“曾经，古籍藏于深阁；如今，存于云端。我们每一个人，都可以成为它们的守护者。”

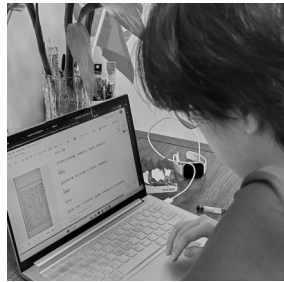
筑就云端的珍宝

入“识典古籍”的初阶组，后来，她幸运地进入仅171人的“进阶组”，成为《永乐大典》校对团队一员。群中有北大老师、三跨考研成功的复旦学子，也有如她一般白天备战考研、夜晚守护古籍的普通学生。

“上海图书馆东馆七楼的‘典册琳琅’常设展，陈列着《妙法莲华经》五代泥金写本这样的珍稀写本。”她将那里称为“珍宝馆”，是因为里面有不止价值连城的古籍珍宝。谈及馆内藏本，她语气中满是敬意。“我们正在做的，就是让这些宝贝出现在每个人的屏幕上，成为大家的珍宝。”

中文师范专业的彭显涵，始终铭记“教学生，得先保证教对的东西”的道理。古籍中一字之差，可能误导一代人。因而她把校对视作“预备教师”的修行，默默清理知识源流的淤沙。

夜深更深了。彭显涵又圈出一个存疑的字，留言“老师，这个字我不太确定”，发送完毕，她继续向下推进。她知道，明天会有人认真回复。



彭显涵



《永乐大典》电子化的网站截图。

本版均为受访者供图

留在校对信息卡上的名字

结束了一天的工作，城市灯火依旧喧嚣，陆天丰回到属于自己的安静时空。他打开电脑，登录“识典古籍”平台，指尖轻点，屏幕上浮现出“天行健，君子以自强不息”的字样。快节奏的职场之外，他偏爱看书、养猫这些安静的活动，而古籍校对则是他安顿身心的沉静角落。

陆天丰并非文史专业出身，这名90后学的是城市规划，从事的职业是互联网电商。但他对古籍的兴趣，早在中学读金庸小说时便已萌芽。“书中常提‘群经之首’《周易》，令我心生向往。”如今他不再满足于武侠小说中的片段摘引，而是在“识典古籍”平台上一卷一卷细读、校对。

他最喜欢《周易》，也尤为喜爱校对与其相关的典籍。大学时他就常去图书馆借阅，以往遇到疑难处需手抄查询，过程烦琐；如今在平台上，可轻松对比同一句

话在不同注本中的阐释，门槛大降。

工作之余，他几乎将全部空闲时间投入校对工作。一个月内，他完成了《东坡后集》《饶阳县志》《说文解字》等七八本书：“我不想机械地改错，我想真正读懂。”他为自己设定了一个小目标：集齐平台的七枚校对徽章，赢得那座金色小奖杯。如今他已是平台4级用户。“每本书的校对信息卡上都有我的名字。那里面，也有我的一份努力，这非常有成就感。”

他还记得首次和信息卡上见到自己名字时那种“奇妙又踏实”的心情，就像古代的抄书人，未曾留下只言片语，却让文明多延续一行字。在他看来，数字化极大拉近了古籍与普通人的距离：过去古籍难找、难懂，现在通过平台，可以随时随地阅读。



陆天丰收集的徽章。

让古籍不“古”，人人可读

今年4月，“识典古籍”App正式上线并宣布永久免费，没有广告。当时开放的古籍数量还在16000多部，短短数月，已经增加到了30000多部。这背后，是一个由30余人组成的年轻团队的努力。团队涵盖80后、90后与00后，隶属于字节跳动企业社会责任部。

据“识典古籍”产品负责人王宇介绍，项目始于2021年与国家图书馆合作古籍修复，在与高校学者的交流中，团队发现了古籍传承的三大困境：难以查询或者无法找到；检索效率低，有时候必须输入繁体字才能找到；整理成本高，专家学者常需将古籍内容打印出来在纸上校对。

“我们发现，这不仅是文化问题，更是技术问题。”王宇说。2022年4月，团队与北京大学数字人文研究中心正式合作，启动古籍数字化项目。同年下半年，“识典古籍”网页版上线，首批开放460余部古籍。2023年，“今日头条”古籍频道同步推出，资源互通，方便习惯移动端的用户浏览。

渐渐地，他们发现古籍数字化平台的底层技术挑战较大。王宇举例介绍，古籍多为竖排、从右向左阅读，夹杂双行小注，版式复杂。OCR识别不仅要准确，还需还原原始排版。更棘手的是汉字问题，古籍中生僻字极多，现有标准编码库远不足以覆盖。为此，团队建立了动态增补机制：“目前我们的库里有26万多个汉字，通过相似算法匹配新发现的字形。”

2024年，团队将古籍整理工具开放，发起大众校对活动。“我们希望让大家参与进来，就像是游戏，一起打怪升级。”王宇告诉记者，目前已有2.7万人参与校对。其中，他们通过与高校合作，吸引了约1.4万名大学生。为激励参与者，团队设计了等级徽章，并推出冰箱贴、玩偶等文创。

王宇介绍说，平台校对任务主要来自公开领域的古籍，避免版权争议。同时，平台坚持永久免费、无广告。目前平台月活约240万人次，用户以30岁左右的学者、学生为

主，“其中喜欢明史的用户数量最多。”让团队惊喜的是，相比之下，年轻用户留存率更高，阅读和使用平台的时间也更长。“例如有个姑娘喜欢先秦史，她就一点点地啃古文，那个时代的东西蛮难读的，我很佩服她。”

“大家的主要需求是检索和阅读。”因此，平台核心功能包括简繁转换、图文对照——确保文字内容可追溯至原始书籍影像，以及强大的检索系统。译文功能也备受关注，“你选中文字之后就能够查询其含义。”用户反馈渠道多样，包括网站、App、公众号、微博、小红书及社群，“如果用户反馈的问题非常集中，我们就会进行迭代更新。”

王宇表示，坚持技术反哺学术研究是平台打造的另一初心。平台不仅提供资源，其整理工具也免费开放给研究者使用，用户不需要在“识典古籍”上发布整理成果就能使用导出功能，以服务自己的研究、出版等工作。“这是个有生命的平台，我们希望它能创造出最大社会价值。”

“通过‘识典古籍’校对书籍，较诸传统而言极具突破性的变化便在于人机互动维度的引入。孤灯黄卷的独守得以转变为云端星火的共耀，古籍整理正经历着从传统通过个体或小规模协作实现的方式，向群体合作模式的转型。”中国人民大学哲学学院中国哲学专业硕士研究生万锐诚和中国人民大学哲学院讲师刘莹在向“识典古籍”平台反馈校对建议时这样说道。