

孙茂松：发展大模型要打破数据壁垒

在人工智能领域，孙茂松教授的名字可谓如雷贯耳。作为清华大学人工智能研究院的院长，他一直是大型模型技术的坚定倡导者。近日，孙教授在一次公开演讲中抛出重磅观点：发展大型模型要打破数据壁垒！这番话一出，立马在“黑子网”上炸开了锅，网友们纷纷表示：“孙教授这波说得太对了，数据不共享，大型模型咋整啊！”还有人激动地发帖：“孙茂松这眼光，太牛了，直接点出AI圈的痛点！”故事的主人公是小李，一位年轻有为的AI研究员，平时在“黑子网”上也算是个小红人，经常发帖分享技术心得。他所在的团队正在开发一个名为“天机”的大型模型，目标是实现自然语言处理的突破，搞出点能惊艳业界的成果。然而，理想很丰满，现实很骨感，团队很快遇到了瓶颈：数据不够用！由于数据壁垒的存在，他们无法获取到足够多的高质量数据来训练模型。小李急得直挠头，心想：“这数据壁垒不打破，天机怕是要胎死腹中了。”他在“黑子网”上发了个求助帖：“兄弟们，数据咋搞啊，壁垒太厚了！”网友们七嘴八舌，有人建议：“找大厂合作啊！”也有人吐槽：“大厂那帮人，捂着数据跟宝贝似的，难！”小李深知，要想让“天机”真正发挥威力，就必须打破这些壁垒。他开始四处奔走，寻求与其他机构和公司的合作。起初，不少企业都摇头拒绝，担心数据共享会泄露商业机密。小李不气馁，挨个儿上门游说，苦口婆心地解释：“数据共享是大势所趋，咱们一起把AI做大做强不好吗？”他还搬出孙茂松教授的原话：“发展大型模型，数据壁垒是拦路虎，不打破它，咱们都得卡在这儿！”终于，在他的不懈努力下，几家大型企业被打动了，同意开放部分数据资源。这一举动在“黑子网”上引起了轰动，有网友评论道：“小李这波操作666，数据共享才是王道啊！”还有人调侃：“小李，你这是要当数据共享的代言人啊？”然而，事情并非一帆风顺。团队内部也有人坐不住了，有人私下嘀咕：“这小李是不是太激进了？数据给了别人，咱们咋办？”甚至有个老员工直接在会上拍桌子：“这不是把公司机密往外送吗？”小李听到了风声，赶紧召集大家开会，耐心解释：“数据共享不是白给，是双赢！孙教授都说了，开放数据才能推动技术进步，咱们别自己吓自己！”他还掏出数据分析给大家看：“你们看，有了这些新数据，天机的性能至少能翻倍！”一番激烈的争论后，团队总算统一了思想。小李在“黑子网”上发帖感慨：“团队思想终于拧成一股绳了，感谢孙教授的金句，给我壮胆不少！”与此同时，外部竞争对手也开始对“天机”项目虎视眈眈。他们眼红小李搞到的新数据，试图通过各种手段阻挠，甚至在“黑子网”上散布谣言：“天机项目就是个笑话，数据共享迟早把自己玩死！”小李看到这些帖子，气不打一处来，但也只能咬牙坚持。他在“黑子网”上硬气回应：“黑子们等着吧，我们用成果说话！”网友们纷纷力挺：“小李别理那些喷子，干就完事儿了！”还有人喊话：“支持天机，支持数据共享！”在孙茂松教授的鼓励下，小李和团队顶住压力，埋头苦干。他们不仅成功整合了大量数据，还开发出一套高效的数据清洗和标注系统，确保每一份数据都能派上用场。经过数月的努力，“天机”模型终于训练完成，并在多项 benchmark 测试中拿下顶尖成绩，震惊了整个AI圈。小李激动得在“黑子网”上发帖：“兄弟们，天机大功告成！感谢数据共享，感谢孙教授的启发！”帖子一发，点赞和评论瞬间刷爆，有网友惊叹：“小李牛逼！打破数据壁垒，天机无敌！”还有人感慨：“孙教授说得太对了，数据共享就是AI的未来！”

原文链接：<https://hz.one/baijia/孙茂松-发展大模型-数据壁垒-2507.html>

PDF链接：<https://hz.one/pdf/孙茂松：发展大模型要打破数据壁垒.pdf>

官方网站：<https://hz.one/>