

自带“冠军基因”的团队想让国产芯片“跑起来”

■本报记者 田瑞颖

这是一支自带“冠军基因”的团队。翟季冬是带领清华大学学生超算队十余次站上世界冠军领奖台的“总教练”，但他心里一直有个声音：能否改变所有参赛队伍都使用英伟达图形处理器(GPU)的现状，让国产算力走到“台前”。

2023年12月，一群来自清华大学计算机系高性能计算研究所的年轻人联手创立了北京清程极智科技有限公司(以下简称清程极智)，清华大学教授翟季冬担任首席科学家。他们选择从软件层面切入创新，为国产芯片量身打造系统软件，让国产算力也能高效运行最前沿的大模型。

随后两年半的时间里，这支团队便打造出覆盖“训练—推理—流通”全链路的产品矩阵：“八卦炉”智能软件栈、“赤兔”推理引擎、“AI Ping”评测路由平台。他们还完成了三轮融资，获得中科创业、中金资本、中关村科学城公司等多家机构投资。

清程极智联合创始人师天鹰告诉《中国科学报》，公司高水平研发人员占比超过80%，平均年龄不到30岁。

“八卦炉”内炼“真金”

“清华的程序员很聪明！”这是“清程极智”名称的由来，也是创始团队的自勉。

随着全球人工智能(AI)大模型参数持续突破、应用场景全面爆发，国内高端算力供给受限、算力结构性短缺、芯片利用率低、软硬件适配壁垒高等四大瓶颈，已成为制约大模型训练与产业规模化落地的核心堵点。

2025年初，随着深度求索(DeepSeek)公司的出圈，清程极智接到了大批为企业做DeepSeek私有化部署的订单。这背后，是团队十余年在智能算力优化方向的积累。

从2014年起，翟季冬就带领清华学生参加全球超算比赛，虽然拿到过十几次世界冠军，但使用的都是英伟达最好的GPU。他一直在思考，如何把实验室的技术搬到货架，助力国产芯片发展。

在超算上验证过的“八卦炉”系统就是研究团队的核心技术之一。该系统提供针对芯片体系架构的算子编译优化，以及面向大规模集群的并行计算解决方案，在国际上首次开源了支持深度学习框架PyTorch的大规模混合专家模型训练，在国产超算系统上实现了百万亿参数量级的大模型训练加速。



受访者供图

“八卦炉”系统的研发可以追溯到2021年。彼时大语言模型ChatGPT还没有引爆全球，翟季冬带领的清华学生团队便开始在国产超算上进行大规模训练探索，使用了超过10万台纯国产服务器。这种超大规模实践，练就了这支国内极少数真正具备国产算力集群使用和调优经验的团队。

这支团队中不乏全国奥林匹克竞赛(奥数)金牌得主、世界大学生超算比赛冠军、中高考状元。清程极智联合创始人唐适之曾是清华超算队的主力队员；师天鹰曾获全国信息学奥赛金牌；首席执行官汤雄超毕业于清华大学计算机系，博士毕业后先后供职于深信服和字节跳动。

从小喜欢玩游戏的师天鹰对编程和算法有着天然的敏锐。2014年，他凭借全国信息学奥赛金牌保送清华大学计算机系，读博期间跟随翟季冬研究高性能计算。当多家大厂向他伸出高薪橄榄枝时，他却选择踏上一条最不好走的路——创业。

“创业是我见过最难的事。”师天鹰坦言，研发、产品、管理、融资……每一个环节都要做好。但比起赚钱，他更想做些有价值的事——让代码落地成产品，对国产算力起到实实在在的作用。

师天鹰解释，大模型的训练就像炼丹，大

量语料投入训练系统，要炼得有模有样。“在大规模算力集群上，‘八卦炉’已实现比肩DeepSeek的混合专家模型训练效率，为国产大模型研发提供了有力的系统支撑。”

让国产芯片“跑起来”

训练之后是推理。

DeepSeek“火”了，同时伴随着尴尬的市场局面：大多数国产芯片和英伟达旧卡无法原生支持DeepSeek采用的FP8数据格式。

“如果在国产卡上运行，传统方案要么把FP8转成BF16，占用两倍显存，要么转成INT8，牺牲模型精度。”唐适之解释。

随着AI的发展进入产业“深水区”，自主可控成为国内AI基础设施建设的重点方向。与此同时，海外厂商正加速通过资本并购和技术整合布局推理市场，但一些国内企业却陷入了“重融资、轻研发”的发展误区，过度追求融资规模与短期商业回报，忽视了推理引擎核心技术的自主研发，普遍采用国外开源软件搭建推理能力。

在唐适之看来，过度依赖国外开源软件不仅会导致国内推理技术发展陷入被动，更会挤压国产推理引擎的生存空间，破坏国内AI产业自主可控的生态根基。

于是，他们选择自主研发赤兔推理引擎，打破海外技术垄断，构建“国产算力+国产引擎+国产模型”全国产化闭环。“从最开始的行代码起，都是我们自己写的，完全自主可控。”唐适之表示，赤兔不仅做到了适配，而且针对昇腾、海光、沐曦、摩尔线程等国产芯片的标量向量计算、数据表达、硬件调度、卡间互联等硬件特性做了深度优化，以解决国产算力“能用不好用”的难题。

在性能与成本上，赤兔推理引擎实现突破性提升：通过软件方案在不具备FP4/FP8硬件能力的算力设备上支持FP4/FP8低位宽量化，使用FP8量化可降低50%算力门槛，FP4量化可降低75%算力门槛。

唐适之介绍，部署DeepSeek-V3满血版模型时，仅需1台8卡910B服务器即可替代传统4台8卡卡机，硬件成本从600万元降至150万元。

目前，赤兔已在券商、能源企业等场景落地应用。某能源央企实现超3500名员工全国产化算力稳定使用，算力效率提升近40%，相关案例获2025人工智能应用大赛铜奖。

评测倒逼行业“向上卷”

如果说赤兔解决的是“产得出”的问题，AI Ping解决的则是“选得准、省得好”的问题。

当词元(Token)成为AI时代的“水电煤”，这群年轻人还要解决一个根本问题：让Token像电流一样稳定、高效、低成本地流入千行百业。

师天鹰直言，Token犹如一个黑盒，你放进去问题，它吐出答案，中间发生了什么却不知道。同样的模型、同样的价格，不同服务商的性能可以差四五倍。

AI Ping平台正是为此而生。目前，该平台已接入30家主流服务商，600余个大规模服务，通过7×24小时多地分布式监测，从真实用户视角进行端到端评测，实时输出延迟、吞吐、可靠性等指标。

“AI Ping的核心功能是智能路由。”师天鹰介绍，实测显示，使用AI Ping后用户平均成本降低超37%，吞吐提升超90%，延迟降低超20%。

2025年9月，清华大学与中国软件评测中心联合发布《2025大模型服务性能排行榜》，AI Ping提供了评测工具和数据展示平台。这也让不少业内的用户将AI Ping比作大模型API(应用程序编程接口)服务的“大众点评”，这种榜单还成为倒逼这些榜单上企业持续优化的动力。随着产品矩阵成型、商业化落地加速，资本的目光也悄然聚焦。从创业初期，需要花费大量精力向投资者和客户解释什么是AI Infra(人工智能基础设施)，到Token经济爆发后，投资人主动咨询并表示投资意愿，这种变化让师天鹰感触颇深。

在融资策略上，汤雄超解释，AI Infra行业发展变化极快，“小步快跑”的融资节奏是为了持续储备现金流，加大产品研发投入，同时也是因为行业热度较高，持续有投资方表达投资意愿。

作为清程极智的投资方之一，绩优创投的相关负责人认为，这支团队有着构建高性能计算系统的罕见基因，能深入国产芯片的“毛细血管”进行优化，真正推动了国产算力从“能用”到“好用”。

在最近一轮融资中，清华大学资产管理公司战略入股了清程极智，这也标志着产学研协同进入新阶段。

面对日益激烈的国际竞争，翟季冬提出了一个更为高远的目标：“中国电力系统在全世界是第一的。我们希望把Token服务的质量和性能也做到世界第一。”

他要把“人造太阳”变成现实

■本报记者 沈春蕾

把“人造太阳”变成现实，这是中国科学技术大学(以下简称中国科大)教授孙玄多年来的心愿。

孙玄从事受控核聚变和基础等离子体物理研究。他希望将聚变能应用于建造巨大、复杂的工程装置，但他也清楚，可控核聚变商业化不是短期能实现的事情。“这是科学和工程的巅峰挑战，几十年来都没做成，不能指望一夜实现。”

2024年，孙玄牵头成立合肥星能玄光科技有限责任公司(以下简称星能玄光)，公司成立一年已完成了数亿元融资。不久前，星能玄光团队自主研发的先进场反磁镜聚变装置FLAME实现首次等离子体放电。

近日，孙玄在接受《中国科学报》记者采访时表示，当前，全球聚变研发进展超乎预期，一些国外同行已经宣布了建造示范电站的计划，虽然是否成功还未可知，但发展趋势是加速的。在中国，聚变商业化的实现或许会比预想的来得更早。

聚变领域的“黑马”

孙玄自1997年在中国科大读研时就开始做聚变研究，主要依托中国科大的KT-5托卡马克装置。随后，他来到美国西弗吉尼亚大学读博，曾在暑假期间前往普林斯顿等离子体物理实验室(PPPL)交流，并了解到场反磁镜(FRC)这种聚变研究方法。

博士毕业，孙玄加入美国洛斯阿拉莫斯国家实验室，在脉冲场反磁镜领域做了一些研究。他告诉记者：“我真正开始研究FRC是在进入美国核聚变能源企业TAE(Tri Alpha Energy)后，也是在那时候意识到FRC的物理研究很有趣，从实现聚变的角度来看也很有市场潜力。”

当时，美籍等离子体物理学家陈凤翔在其所著的科普书《一个不可缺少的真相——聚变能源如何拯救地球》中提到，FRC是聚变领域的“黑马”。孙玄对此非常赞同：“现在看来，FRC可能已成为‘白马’了。”

目前，实现可控核聚变主要有三种技术路线：磁约束核聚变、惯性约束核聚变和磁惯性约束核聚变。一般认为，FRC属于磁惯性约束核聚变，托卡马克属于磁约束核聚变。

FRC和托卡马克的本质区别在于拓扑结构不同。孙玄介绍，FRC属于紧凑型磁约束聚变装置，没有中心螺线管，物理挑战比托卡

马克更大，但装置结构更简单。FRC是高β值的等离子体，磁场利用率很高，所以建造成本相对较低，非常适合做先进聚变反应堆，如氦氖-3、氦硼等燃料的聚变反应堆。FRC可以稳态运行或绝热压缩，是非常高效和简单的聚变实现途径。

“聚变某种意义上像‘暴力美学’，需要输入大量能量，而FRC正好适合这种简单、粗暴但有效的路径。”孙玄认为，从聚变能源最终的经济性目标来看，根据现有的估算，FRC在经济效益上的优势比较明显。

美国的求学和工作经历让孙玄收获满满。他接受了系统的学术训练，学习了如何做研究。“在TAE公司的4年时间里，公司提供了适合科研人员发展的平台和氛围，这改变了我对企业的印象和认知，也为我创立星能玄光提供了很好的参考样本。”孙玄说。

回国潜心做科研

2012年，孙玄回国加入中国科大。“回母校工作，是一件不假思索的事。”孙玄感叹道，这里是国内等离子体物理研究实力最强的高校，也是做物理、做科研的理想场所。

“当时国内的核聚变研究领域，磁约束基本都采用托卡马克，其他路线很少，或者说基本没有。”孙玄却选择了FRC聚变这条路，他觉得某种意义上这是填补国内研究方向的一个空白。“因为中国已加入国际热核聚变实验堆计划(ITER)，国家对聚变的投入在加大，我们新建的仪器设备甚至比国外还好。”

尽管一切向好，但孙玄也发现，就市场前景而言，当年还没有社会资本关注聚变，更谈不上商业化。

孙玄决定从基础研究做起。2012年，他在中国科大设计建造国内最大的串列磁镜装置KMAX，该装置于2013年开始加工组装，2014年初实现放电。随后十年，孙玄和团队将场反磁镜两种磁约束位形有机结合，开创出先进场反磁镜聚变路径。

其间，孙玄团队在KMAX上观测、测量了两个FRC的碰撞融合过程，并且分析了其中的能量转换机制，给出了对磁融合的实验证据。“这实际上也是在国际上首次给出的直接实验证据，证明了FRC可以对磁融合。”孙玄说。

不过现实的困难也摆在眼前。孙玄还记得，当时托卡马克是主流，FRC这条路径想直



孙玄与先进场反磁镜聚变装置FLAME。

受访者供图

接申请项目还比较困难。“我们的解决方案是采取‘迂回’策略，不直接强调聚变能源目标，而是侧重研究FRC所涉及的基础科学问题，例如与空间物理现象相关的共线性物理，或者应用于推进器等方向。”

就这样，孙玄团队通过强调FRC基础研究和交叉应用潜力来争取支持，从而为聚变研究积累了经费。

创业赶上好时机

2024年，孙玄认为创业的好时机来了。资本市场对聚变的关注度显著提升，能量奇点、星环聚能等聚变创业公司都融资成功了。这让他意识到，随着国家对于聚变的关注和重视，国内资本也愿意投资聚变这个长期赛道了。

与此同时，孙玄团队在中国科大的基础研究也到了一个阶段性节点。“我们已经完成了一系列原理验证，接下来需要建造更大规模、更高参数的工程化装置向聚变目标迈进，这显然超出了高校常规科研的范畴，需要以公司化的形式来运营。”

在资本环境转暖 and FRC技术发展的双重需求下，星能玄光成立了。

从高校教授到创业者，孙玄发现最大的变化是目标导向的转变。在高校，一个项目结题的标准往往是发表论文、申请专利，以科研成果为导向；在企业，唯一的目标就是把聚变这

件事做成，是产品、结果导向，一切行动都要以最终实现聚变能源的商业化为导向。

创业初期，孙玄最难适应的是各种路演。在学校，面对的听众主要是学生和同行；路演面对的是投资人，目的是获得资金支持。“这种有目的的沟通方式和氛围在初期让我非常不适应，不过现在参加路演我已经驾轻就熟了。”

成立一年的时间，星能玄光就完成了数亿元的融资。“投我们的资本方都是‘耐心资本’，他们对聚变的长期性有深刻认知，愿意陪伴我们长跑。”孙玄说，“光是讲‘长期’也不行，我们会在内部设立一些短期的，甚至有点疯狂的目标，看看能否集中资源快速突破，也让大家在努力中不断获得成就感，这样才能把一个长期的事业坚持下去。”

虽然星能玄光目前与国际顶尖的团队在装置建设、迭代方面还有差距，但在科学和物理设计上，孙玄认为：“我们和他们处于‘并跑’状态，或者说各有特色。我们起源于中国科大，在基础研究阶段就形成了自己的独特路径和技术特色。”

如今，孙玄带领的星能玄光目标很明确：一方面要加强工程团队建设，另一方面要依托国家的工业制造能力和与高校院所的合作。“如果可以将物理需求与工业制造能力深度融合，在关键工程问题上实现突破，我们完全有可能从‘并跑’走向‘领跑’。”

资讯

2026广州“科创下午茶”香港专场活动举行

本报讯(记者朱汉斌 通讯员郑碧施)近日，2026广州“科创下午茶”香港专场活动举行。本次活动由中国科协企业创新服务中心等单位指导，广州市科协联合香港机电工程署、大湾区青年促进局、广州市科技局、“科创中国”大湾区青年百人会等共同主办。

主题分享环节，香港科技大学土木及环境工程学系副教授、智慧城市研究院副主任王君围绕智慧城市建设实践、人工智能(AI)技术在楼宇运维与智能建造领域的前沿应用以及德港产学研协同攻关方向展开剖析，为在场企业及科研院所提供产学研合作新思路。

路演对接环节，来自香港、广州两地的8个“AI+建筑”项目登台路演交流。其中，天空社公司展示了人工智能物联网(AIoT)智慧安全与健康管理体系的整体解决方案，香港机电工程署介绍了机电装备合成法在维修、保养、改建及加建工程中的应用，玄建科技介绍了智慧建筑数字孪生平台COSMOS，香港发展局介绍了自主研发并已获得专利的“智眼 Smart Eye”项目。

此外，华南理工大学亚热带建筑与城市科学全国重点实验室介绍了保温与遮阳性能独立调节窗膜系统，广东省建筑工程集团控股有限公司展示了广东建安安管平台，广州市三川田文化科技股份有限公司介绍了“AI牧场”实体空间智能觉醒项目，广东中建普联科技股份有限公司推出AI+绿色建材供应链——城市级建设投资全生命周期数智化解决方案。

2024年，广州市科协与香港机电工程署签署合作备忘录，在科技、学术、科普、人才等方面建立长期协作机制。备忘录签署以来，双方持续深入交流合作，为粤港澳大湾区科技创新与融合发展提供了有力支撑。下一步，广州市科协将继续深化与香港机电工程署等单位的合作，依托“科创下午茶”等平台，助力科技成果转化落地，持续推动大湾区高质量发展。

47项青少年创新成果亮相上交会

本报讯(见习记者江庆龄)“我的作品叫智能衣橱，能够结合语音识别、触摸屏交互、图像识别和互联网技术，为用户提供穿搭建议和衣物信息。”6月11日，在第十二届中国(上海)国际技术进出口交易会(以下简称上交会)青少年创新发展展区，来自上海市青浦区华新中学的刘家玮同学正在介绍自己的发明成果。

包括智能衣橱在内，全上海市47项青少年创新成果同步进入上交会线上展示，项目覆盖绿色环保、生活服务、科技体育、文化传承等领域。其中，21项成果在本届上交会的上海市青少年科技发明成果展示活动现场集中亮相。

今年恰逢“上海市青少年创新创业(模拟)团队行动”举办十周年。十年来，该行动累计孵化近400个学生创新项目，353项成果登上了上交会舞台，146个项目获得或正在申请专利，20余项通过专利银行、技术交易所和企业合作等方式实现成果转化，展现了青少年的创意从“做得出”走向“用得上”的潜力。

“青少年模拟创业营”在现场启动。依托专业商业模式系统，学生可以沉浸式体验公司设立、市场调研、定价策略、财务决策、路演汇报等完整创业流程，系统学习如何将技术创意转化为商业价值。

此外，上海市静安区教育学院附属学校“八爪鱼”项目与同济大学教授团队、冠研(上海)专利技术有限公司达成合作，开展技术验证和产品开发；“笔芯回收再利用的画笔”“视力检测仪(熊猫便携)”“刻灵一代——辅助篆刻学习智能系统”3个项目签署专利技术托管协议。

中科闻歌通过香港交易所聆讯

本报讯(记者沈春蕾)近日，香港交易所官网显示，北京中科闻歌科技股份有限公司(以下简称中科闻歌)已刊发聆讯后资料集，这意味着中科闻歌正式通过上市聆讯。招股书显示，中科闻歌成立于2017年，由中国科学院自动化研究所研发团队创立，专注于研发企业级大模型驱动的决策智能操作系统及服务。

从招股书披露的信息来看，中科闻歌的核心业务长期围绕“决策智能”展开。公司通过自主研发的DIOS(决策智能操作系统)，为政企客户提供决策支持服务。截至2025年12月31日，中科闻歌研发团队共有205名员工，占全职员工总数的43.9%。

财务数据显示，中科闻歌的收入从2023年的2.497亿元增长至2025年的4.053亿元，收入增长超过60%。毛利率由2023年的44.0%提升至2025年的51.2%，并稳定维持在50%。此外，中科闻歌产品部署项目的平均交付周期从2023年的185.0天缩短至2025年的80.2天，已累计为超过650家企业及政府客户提供专业AI服务。

招股书显示，中科闻歌计划进一步将AI应用拓展至商业、工业等领域，并持续孵化科学与教育、能源与可持续发展、医疗健康等新场景。

在首次公开募股(IPO)前，中科闻歌已完成10轮融资，投资者包括国开制造业转型升级基金、中国互联网投资基金、央视融媒体产业投资基金等国家基金，以及北京市人工智能产业投资基金、中科创业、深创投、盈富泰克等投资机构。