AI加剧网络可靠性风险

——访鸿雪科技董事长兼CEO郑乃东

本报记者 宋婧

近期，全球宕机事件频发，引发了用户和业内人士对网络稳定性与安全性的担忧。近日，鸿雪科技董事长兼CEO郑乃东在接受《中国电子报》记者独家专访时表示，随着大量的传统应用变成了互联网应用，且已深入到工作和生活场景中，宕机事件带来的影响越来越大。推动SRE（网站可靠性工程）是提高可靠性、避免各种宕机的重要路径。

宕机事件频发 加剧系统可靠性担忧

4月8日，“腾讯云崩了”冲上热搜。大量网友反馈，称腾讯云出现服务故障，接口响应报错、网页显示504错误，范围覆盖全国各地。6月4日，ChatGPT遭遇近8小时的大规模宕机，包括其网站和应用程序在内都无法访问，全球数百万用户受到影响。7月2日，阿里云发生宕机事件，虽说从发现故障到解决用时31分钟，从发现故障到影响恢复用时38分钟，但B站、小红书、恋与深空、酷安等多家大厂APP均受到波及。

“大家之所以感觉到宕机事件变多了，主要原因是互联网应用的数量变多了，像以前的Office和WPS这种单机软件现在也都连接了云服务，大量的传统应用都变成了互联网应用，仅苹果商店的互联网应用数量就超过200万个，而且这类互联网应用已经深入到我们的工作生活当中，比如微信、美团、抖音、腾讯会议等，因此我们对宕机的感知也会越来越明显。”郑乃东分析道。

实际上，随着互联网软件功能越来越多，结构越来越复杂，在日常运营过程中出现问题的概率也会越来越大。一些大型科技公司，如亚马逊、微软和谷歌等，每年在系统可靠性上的投入占其整体技术预算的15%~20%。而在国内，很多公司依然存在不重视可靠性、可靠性人才奇缺、没有可靠性管理、对可靠性认识模糊等问题。

“最要紧的是主观上的重视程度，不管是云供应商、软件开发商，还是运维环节的服务厂商等，各方都需要重视线上事故的预防、发现、定位、处理、复盘的全链条保障。比如投入专门的资金，设置专门的岗位来从事可靠性管控方面的工作。”郑乃东表示。

SRE有望在国内市场 快速推广应用

SRE全称是Site Reliability Engineering，指网站可靠性工程，最早由Google提出，旨在提高软件系统的可用性、低时延、性能、效率、变更管理、监控、应急响应和容量管理等方面的能力。

“SRE主要通过自动化、监控、预防性措施和持续改进来减少故障发生的概率，并且降低故障所造成的影响。”郑乃东向记者介绍道。首先，SRE能通过监控和告警系统提前发现潜在的问题，并快速响应和恢复系统服务；其次，SRE采用软件工程的方法，与开发人员紧密合作，倡导构建业务系统内置的可靠性，并在运维过程中使用自动化和标准化流程，减少人为错误，从而进一步提升系统的稳定性；最后，通过降低琐事和持续优化的运营流程，SRE可以减少运维人员的工作负担，提高运营效率。

近年来，随着中国互联网产业、云计算快速发展，系统复杂性和对可靠性的需求大幅增加，SRE的价值逐渐被广泛传播和认可。国内互联网大厂如百度、阿里巴巴、腾讯、京东等大型互联网公司由于业务规模庞大、用户数量众多、系统复杂度高，率先认识到SRE的重要性，并积极推动SRE的实施。

郑乃东分析道：“这些公司需要确保其平台能够在高流量情况下稳定运行，避免服务中断对用户体验、公司收入及声誉造成的负面影响。因此，SRE成为它们提升系统可靠性、提高服务质量的重要手段。”

传统型企业乃至中小企业也开始主动关注，并在生产环境中应用SRE。据权威调研机构统计，2022年，中国约有40%的大型企业和20%的中小型企业正在推行SRE实践，并且这样的企业在逐年递增。“未来，随着技术的不断发展和企业需求的增加，SRE在全国的应用将更加广泛和深入。”郑乃东判断道。

AI给SRE 带来挑战与机遇

尽管SRE可以显著提升系统的可靠性和稳定性，但郑乃东同时也指出，SRE存在局限性，并不能完全消除宕机现象和所有的技术问题。比如，复杂的业务逻辑问题，SRE主要关注系统可靠性层面的问题，复杂的业务逻辑错误仍需依赖开发团队解决。再比如，基础设施故障，硬件故障、网络中断等基础设施问题可能超出SRE的控制范围。另外，自然灾害、突发事件等不可预见的灾难，SRE无法完全避免，但可以通过灾备方案减小影响。

在郑乃东看来，AI的到来既为SRE带来了新的挑战，也带来了新的机会。一方面，AI系统本身就具有很高的复杂性，复杂的架构与当前系统的依赖关系使得企业需要花费很高成本学习和驾驭AI技术。据调研，超过60%的企业认为AI系统的复杂性是实施过程中最大的挑战之一。

另一方面，AI系统依赖大量的数据，数据质量和完整性问题可能导致模型误差和系统故障，SRE需要通过软件工程的方法，配合统一的数据模型确保所有管理数据管道的稳定性和可靠性。只有利用高质量的数据，才能使AI算法和大模型更加精确和高效。

此外，许多AI应用，特别是大语言模型相关的会话交互型场景，都需要实时处理和及时响应。在2023年的一项研究中，85%的AI应用对响应时间的要求在毫秒级以内。SRE需要确保相关系统具有足够的性能和低延迟，以满足这些实时性要求。

“AI模型管理、数据工程、安全性这三点非常重要。”郑乃东表示。他认为，SRE需要掌握AI模型的部署、监控和管理技能，确保模型在生产环境中的稳定性和性能，同时也要增加数据工程的能力，确保数据管道的可靠性和数据质量，以支持AI系统的正常运行。此外，AI系统可能面临新的安全威胁，SRE还应该关注AI模型和数据的安全性，防止内部敏感数据泄露和受到攻击。

国内SRE产业生态建设 亟须提速

随着新质生产力发展步伐加快，企业数字化转型逐渐走向深水区。在这一过程中，构建稳定、可靠且高性能的基础设施至关重要。SRE作为基础设施战略的关键组成部分，为业界提供了实现先进基础设施策略的关键思路。

然而，业内人士普遍认为，国内SRE产业生态建设仍然面临着人才短缺、技术积累不足、文化转型难、工具和平台集成难等多重挑战。以人才建设为例，SRE是一个相对较新的领域，具备相关技能和经验的人才供不应求。2023年的一项数据显示，中国SRE工程师的供需缺口超过30%。这导致企业在招聘和培养SRE工程师时面临困难。

“我国在SRE实践上的技术积累相对较少，很多企业缺乏成熟的SRE实施经验和最佳实践指导，而在国外，约60%的大型企业已经实施了成熟的SRE实践。”郑乃东坦言。

为缓解人才短缺的问题，越来越多的培训机构正在开设SRE相关课程。鸿雪科技便是其中之一。其培训涵盖了SRE的各个维度，包括自动化、可观测性、AIOps、平台工程、高可用、灾备等，确保学员能够全面掌握SRE所需的各项技能。讲师团队皆为行业内资深SRE专家，不仅具备丰富的SRE实践经验，还参与过许多大型项目的实施，能够提供深刻的洞见和实用的建议。据统计，2023年参加专业SRE培训的企业系统可靠性提升了20%，运维效率提升了15%。

“SRE人才保障了AI及所有业务系统生产环境的可靠性和性能，通过自动化和可观测性减少了宕机风险。他们确保所有服务在高并发情况下稳定运行，是AI系统和所有其他业务的‘守护者’。SRE团队的存在能够将系统宕机时间减少50%以上。”郑乃东强调。他指出，算法和数据提供智能和支持，SRE确保系统可靠运行，三者共同协作才能实现AI系统的全面成功。

编者按：今年是中国全功能接入国际互联网30周年。这30年来，中国互联网实现了“从无到有，从小到大，从弱到强”的大跨越式发展，探索出了一条具有中国特色的互联网发展道路。7月9日，“2024（第二十三届）中国互联网大会”在北京开幕。在会上，行业专家对中国互联网下一步发展提出了建议。

中国互联网下一步该怎么走？

本报记者 路轶晨 徐恒

中国工程院院士邬贺铨：

工业内网连接数空间巨大

中国工程院院士、中国互联网协会专家咨询委员会主任邬贺铨表示，随着互联网普及率的稳步提升和移动用户数接近饱和，连接数的增长将逐渐从公众用户转向各类专网，特别是算力网、政企专网、工业内网、车联网、物联网等领域，其中工业内网的连接数空间巨大，亟待开发。

到2023年年底，中国互联网普及率已达到78%，年增长率达2.6%，用户数接近饱和。在IPv6的普及方面，中国取得了显著进展，国内互联网企业在线统计并去重汇总后的数据显示，IPv6渗透率已达72.7%。

在固定宽带领域，邬贺铨透露，2024年第一季度，我国固定宽带用户中百兆和千兆接入分别占94.5%和27.4%。尽管千兆接入用户数增加显著，但对实际平均下载速率的提升拉动并不明显，且千兆与百兆接入的上行能力相当，均在30Mbps左右，这在一定程度上限制了入算应用的发展。

在移动宽带方面，邬贺铨指出，在Sub6GHz频段下，5G设计峰值可达Gbps级，而5G-Advanced（5G-A）更是达到了10Gbps级。目前实测5G下行峰值与均值约为4G的7倍，即百兆量级，但上行速率与4G相差无几，仅为30Mbps级。截至2024年4月，我国5G用户已占移动用户总数的一半以上，户均移动互联网接入流量（DOU）占比与4G相当。邬贺铨表示，网络宽带化已具备支持视频与AI场景的潜力，但仍需开发更多增强用户体验和体现网络价值的应用。

展望未来，邬贺铨认为，5G-A与IPv6的兴起将引领互联网进入新的发展阶段。特别是AI大模型的智能涌现，不仅使互联网原有业态焕新，还催生了众多新业态。互联网平台作为AI赋能消费与行业应用的重要模式，将在新引擎的加持下实现再出发。工业互联网是互联网的下半场，虽然开局不尽如人意，但是现在AI发力数字世界与物理世界的结合，提升了互联网接入物理实体服务垂直行业的能力，在促进产业数字化的同时也带动了数字产业化及互联网产业的发展。AI将为平台经济增添新动能，互联网平台也是AI赋能消费与行业应用的重要模式。此外，对于公众比较关注的大模型，邬贺铨表示，除了算力，在算法方面，目前中国的大模型的数量在全球排名还是比较高的，并且在大模型上差别不是很大。

中国互联网协会理事长尚冰：

构筑以大模型为代表的自主生态

中国互联网协会理事长尚冰表示，当前，全球科技创新进入了空前的密集活跃期，以通用人工智能为代表的新一代信息技术正在深刻重构数字世界和物理世界，加速经济社会数字化、网络化和智能化转型，为互联网行业发展注入更强劲的动力，提供更加广阔的发展空间。

对于互联网行业下一步发展，他提出四点建议。

一要巩固基础网络的领先优势。持续深化5G、千兆光网、IPv6、移动互联网等规模部署，提升网络能力和覆盖率，加快算力资源的多元化发展，提升算力网络的整体能力。优化国际通信出入口和通信网络节点布局。

二要强化互联网技术创新能力。加快人工智能芯片、算法框架等软硬件进行创新，构筑以大模型为代表的自主生态，加快形成新质生产力。系统开展先进无线通信、新型网络架构、空天地一体等下一代互联网的前瞻性布局，把握发展主动权。

三要提升融合创新的发展格局。加快云计算、大数据、人工智能等互联网技术与实体经济的深度融合，加快传统产业数字化转型步伐，持续壮大共享出行、电商直播、无人配送等新业态、新模式，提升经济社会的服务效率。

四要深化高水平的开放工作。持续提高国际化运营水平，加速跨境电商、社交娱乐、移动支付等优势，应用于出海发展，深化多领域深度对接交流，积极参与人工智能、数字贸易、数据安全等多双边数字治理工作。

奇安信集团董事长齐向东：

互联网需以AI驱动安全

全国政协委员、全国工商联副主席、奇安信集团董事长齐向东在主题演讲中表示，在人工智能新时代，网络安全面临着全方位的挑战，要以AI驱动安全，应对三大安全威胁、补齐三大薄弱环节，全方位提升新时代安全能力，护航互联网新质变，共同打造安全、稳定、繁荣的网络空间。

在人工智能新时代下，生产力得到了极大地提高，同时，安全威胁也正由点及面全方位扩散。一方面，网络安全攻击造成严重的经济损失；另一方面，网络安全事故还会严重威胁国家安全。

“想要应对三大安全威胁、补齐三大薄弱环节，就要用AI驱动安全，为安全能力带来指数级跃升。”齐向东介绍，在单点设备检测方面，AI可以对过去人工漏掉的告警进行全量研判，实现安全能力十倍级提升；在体系化防御方面，通过AI赋能的综合分析和全局联动，可以实现安全能力百倍级提升；在溯源和反制方面，从威胁发现到攻击溯源环节，依托AI的智能化、自动化，可实现响应能力的千倍级提升。

360集团创始人周鸿祎：

没有互联网就没有AI的发展

360集团创始人周鸿祎表示，几年前，Web3.0、区块链、元宇宙都曾昙花一现，正当很多人质疑互联网发展是否到了瓶颈期时，大模型的突破为互联网注入了新的活力。周鸿祎认为，大模型将掀起新一轮工业革命，所有的互联网应用都会被重塑，同时也给互联网产业带来了新的革命性机会，强化了互联网创业和投资的火爆需求。但是，没有互联网的发展就没有人工智能发展。正是得益于互联网的算力、数据、算法等积累，才为人工智能的发展奠定了基础。

周鸿祎回忆，过去30年内，以QQ、免费杀毒、门户、电商等为代表的互联网商业模式创新，不仅使中国老百姓的生活实现了数字化改变，也推动了中国数字经济的发展。中国的创业者们赶上了时代机会和有利环境，打造了互联网蓬勃向上的产业生态。“大模型的发展使互联网又迎来了新的春天。”周鸿祎希望，无论是传统的创业者还是人工智能时代涌现出的新创业者，都应该继承互联网创业精神，像鲶鱼一样把互联网产业的“水”搅动得更“活”。

周鸿祎认为，未来10到20年间，大模型都会深刻改变世界。“大模型不是操作系统，会成为未来整个社会、整个数字化业务中的重要组件。”他表示，人工智能的赛道非常宽，不仅互联网公司有机会，对于传统企业和创业者而言也有机会。他举例道，大模型的能力就像是电动机，虽不能直接使用，但装上轮子就可以变成汽车，装上叶片就可以变成风扇，因此必须与工作生活的场景相结合变成产品，才能走入百行千业、千家万户。传统企业和政府也有机会结合业务场景定制专业大模型，所以未来大模型会无处不在。

搜狐CEO张朝阳：

未来将是人脑与AI相结合

搜狐CEO张朝阳表示，过去十年互联网的发展基本上有两个方向，一个是人工智能的发展，另一个是长短视频等影像化内容的崛起。而这些内容的分发一是靠AI算法，二是靠社交分发。

在展望未来10年时，张朝阳表达了对人工智能改变世界的预期。

张朝阳以物理学为例，认为尽管大模型AI在处理大量数据和提供重复性答案方面非常强大，但在原创性思考和复杂问题求解方面，比如AI能不能理解量子力学、能不能原创思考一些物理问题等方面，还有待突破。他预计，未来将是人脑与AI相结合，可能会产生很多新的科研成果。他提醒，人们也要关注AI可能带来不好的一方面，比如虚假信息、身份认证、个人隐私问题等，这些都需要相应的法律来跟进。

对于10年以后互联网的机会在哪儿，张朝阳认为未来社交仍将是主流。人们以后的生活可能都是永远在线，即使人们未曾见面，也可能成为最熟悉的陌生人，实现24小时的无缝交流。