证券代码： 300846　　证券简称：首都在线

北京首都在线科技股份有限公司
投资者关系活动记录表

投资者关系活动类别	☑特定对象调研　　□ 分析师会议 □ 媒体采访　　□ 业绩说明会 □ 新闻发布会　　□ 路演活动 □ 现场参观 □ 其他 （请文字说明其他活动内容）
参与单位名称及人员姓名	兴业证券　中国中金财富证券　中银国际证券 诺安基金　平安基金　永赢基金　易米基金 融通基金　中加基金　国泰基金　前海联合基金 金信基金　长盛基金　西藏东财基金　工银瑞信基金 中科沃土基金　金元顺安基金　西部利得基金　大家资产管理 中再资产管理　阳光资产管理　太平资产管理　交银人寿保险 长江养老保险　中国人寿保险　长生人寿保险　上海人寿保险 中国国际金融　中银香港资管　新华资产管理　光大理财 上海磐厚投资　上海景领投资　北京远惟投资　北京东方睿石投资 等（排名不分先后）
时间	2024 年 2 月 7 日 (周五)
地点	线上会议
上市公司接待人员姓名	1、董事、执行总裁：姚巍 2、董事、副总经理、董事会秘书：杨丽萍
投资者关系活动主要内容介绍	**公司情况介绍** 随着 AI 大模型技术日臻成熟，其未来发展势必会对社会结构、经济形态、国家竞争力以及人类的生产生活方式产生深刻变革，未来势必会涌现出更多模型及丰富的 AI 应用，成为驱动社会发展不可或缺的关键力量。 在此背景下，公司的战略布局、技术路线与资源配置，始终

	围绕这一趋势展开，助力行业模型及垂类模型实现快速迭代升级，推动 AI 应用在各行业的深度渗透。公司深度践行“一体两翼”战略规划，以“一云多模”“一云多芯”“一云多池”为切入点，全力打造基于“M 种大模型” 与“N 种芯片”的首都在线智算云平台。 “一云多模”，公司大模型平台已成功将国内的 DeepSeek、智谱、千问、零一万物等国产大模型，以及国外的 Llama、Bloom 等主流大模型全面部署至云平台。后续，公司还计划将各类行业应用模型全部转化为云上应用，并将模型封装为云服务。这一举措使得用户能够在平台上便捷、快速地加载并切换不同模型，高效完成训练、部署及推理任务。 “一云多芯”着重凸显平台卓越的芯片兼容性，支持英伟达、华为、燧原等多种芯片类型，用户无需关切底层硬件的差异，即可稳定获取算力支持，极大地提升了使用便捷性与算力保障。 “一云多池”则充分展现公司算力资源的灵活调配能力。公司不仅拥有自主建设的算力池，还能够灵活整合调度第三方碎片化算力资源。基于此轻资产运营模式，公司可获取海量弹性算力，显著提高算力资源利用率，为公司业务拓展与高效运营提供坚实保障。 **二、投资者问答交流环节** **Q1：首都在线目前 GPU 芯片规模如何？推理芯片和训练芯片的种类有哪些？** A：目前，首都在线整体算力芯片规模已超过 2 万张。其中，90%的芯片为推理芯片，主要为英伟达主流推理芯片及少部分渲染推理芯片，还有部分燧原、海飞科的国产推理芯片。此外，10%为训练芯片，包括英伟达 H 系列芯片、华为昇腾芯片。 公司算力资源管理方面采用“一云多池”策略，2 万张芯片中，60%为纳管第三方算力资源，公司进行统一调度管理。 **Q2：公司和英伟达及国产芯片的合作情况如何？**

	A：在英伟达芯片方面，公司目前使用的芯片包括英伟达主流推理芯片及 H 系列芯片，其中 L 系列芯片主要通过纳管第三方资源的方式进行调度。未来，公司会保持对英伟达新品的关注，后续一旦有新芯片推出，将及时且积极地引入。 在华为芯片方面，重点聚焦于搭建“训推一体”平台，不仅支持 910B，还包括 800I 系列推理芯片。这些华为芯片的性能表现优秀，但由于生态体系不同，需要进行适配。公司在适配方面投入了大量资源。目前，与华为的合作项目主要落地在北京门头沟地区，并与门头沟政府合作推进。 在国产其他芯片方面，公司与燧原有深度合作，提供基于燧原 GPU 的 MaaS 服务，例如燧原的文图产品“燧图”已在游戏行业应用。此外，在庆阳，公司将燧原的芯片以云服务的形式，支持智谱 AI 的推理应用。这是国产燧原芯片首次被应用于智谱 AI 的推理任务，并实现商用落地。此外，公司与海飞科也合作建设了实验平台，正在调测紫东太初的大模型，有望在海飞科芯片上实现良好的应用。与此同时，公司还与其他几家国产芯片厂商进行合作，主要集中于协助其与几个大模型厂家完成适配与接入，不过，仍处于技术调试阶段，尚未形成商业化的闭环。 **Q3、公司为何采用“自建+纳管”的算力管理模式？** A：公司采用“自建+纳管”模式，主要原因有三点： 一是盘活市场存量算力。许多地方政府和机构投资了大量算力资源，但由于找不到足够的用户，导致算力资源闲置。通过纳管模式，首都在线能够激活这些资源，有助于解决地方算力闲置问题。 二是平抑短期算力需求波动影响。按照行业发展规律，算力需求必定呈现波动上行趋势，短期内会出现算力供给过剩，导致空置率上升，这是市场发展过程中的正常现象。公司采取了“自建+纳管”的策略，可以有效保持算力资源的灵活性。如果空置率上升，公司可以减少纳管算力的使用，以提高利润率和资源利

	用率；如果空置率下降，公司可以增加纳管资源的比例，以应对市场波动。 三是获得政府支持，降低算力成本。为支持地方算力产业发展，部分地区政府在算力建设方面提供补贴，使得公司可以降低运营成本，同时推动自建算力的部署。 **Q4：纳管与自建模式在收入上有何区别？对公司的利润率影响如何？** A: 从收入角度来看，纳管模式和自建模式的收入差别不大，但在利润方面有所不同。自建模式下，由于公司自行投资建设，资产折旧周期为五年，利润率表现良好。纳管模式下，公司需要向算力提供方支付一定费用，利润率相对较低，通常在 10%-20%之间。 但从风险控制的角度来看，纳管模式的优势在于公司不需要承担全部资产管理的风险。如果客户需求波动，公司可以更灵活地调整上游资源。而在自建模式下，公司需要直接管理大量算力资源，面临更高的运营风险。 **Q5：未来在庆阳新建算力，公司以自建为主还是纳管为主？** A：公司在庆阳新建算力，计划主要以自建为主。纳管模式虽然可以降低风险，但其资源并不完全可控，有时无法满足灵活调度的需求。而自建算力资源完全可控，可确保公司在业务运营中的自主权。公司在庆阳的战略是形成“固定+弹性”的算力组合，以维持稳定运营。 总体而言，公司计划采用 40%自建、60%纳管的模式。这种比例对于云计算企业而言相对合理，既能保证控制力，又能降低业务风险。 **Q6：从客户需求与租金层面综合分析，算力资源展现出何种发展趋势？**

	A：单 token 算力成本持续下降是行业的必然趋势，同时，短期内算力的供需不平衡，也会导致空置率上升，这是市场发展过程中的正常现象。许多企业的算力出租率低于预期，并非个别情况，而是行业阶段性的挑战。首都在线采取“自建+纳管”相结合的策略，以保持算力资源的灵活性。如果空置率上升，公司可以减少纳管算力的使用，以提高利润率和资源利用率；如果空置率下降，公司可以增加纳管资源的比例，以应对市场波动。 前期，公司空置率较高的问题主要集中在早期投入的 A5000 和 3090 芯片，但公司采取了两项措施进行优化： 1.通过“自建+纳管”模式保持资源的弹性，即便客户需求减少，公司仍可调整自有和纳管算力的比例，以维持高出租率。 2.随着 DeepSeek 的推出，公司将部分早期算力资源加载到大模型平台，以模型即服务（MaaS）的方式销售，大幅降低闲置算力，提升整体利用率。 目前，首都在线的利用率良好，运营策略较为稳定。 **Q7：国产芯片在性价比和稳定性方面如何？公司未来扩建计划中，国产芯片的比例会占多少？** A：目前国产芯片的发展存在两项挑战。一是性价比。英伟达的芯片生产规模大，采购成本相对较低，而国产芯片的生产规模较小，因此采购成本较高，这导致国产芯片整体性价比成为挑战。二是生态兼容。华为的计算架构与英伟达不同，需要进行额外的翻译适配。公司在这方面投入了大量资源，帮助国产芯片优化兼容性和运行效率。 对于未来的算力扩建，公司仍坚持“以客户需求”为核心。英伟达的高端芯片，公司会继续合法合规的采购和使用；如国产芯片可满足推理应用，公司将优先采用国产芯片。 **Q8：国产芯片和英伟达芯片在财务折旧年限与使用年限上是否相同？**

	A：在财务上，所有芯片的折旧年限都是一致的，通常为五年。在使用年限上，目前，公司最早一批国产芯片已经使用近两年，其长期表现仍需进一步观察。 **Q9：公司的智能算力中心是否由公司内部团队进行运维？** A：是的，公司本身就是云计算起步，因此智能算力中心的运维主要由公司内部团队负责。机房运维方面，如果是租赁的机房，则由机房方进行基础设施维护，而设备运维和云平台管理均由公司自主完成。此外，公司已实现 90%以上的远程运维和调试，减少了对现场运维人员的依赖。现场运维团队按照区域划分，采用“全球网格化管理”模式。例如，在海外，每个洲可能配备一至两名运维人员；在国内，则按区域划分，如华东区可能安排数名技术人员进行支持。 **Q10：随着推理需求的快速增长，市场普遍预期未来可能会出现算力资源紧缺和租金上涨的情况。从短期来看，未来半年到一年的价格趋势如何？** A：从长期来看，算力的需求一定会增长，但单算力成本下降是必然趋势。这一趋势主要体现在以下几个方面： 1.单算力成本持续下降：即使未来英伟达、华为推出更高性能的芯片，其设备单价可能会上升，但单位算力成本仍会下降。这是技术发展的必然结果，否则 AI 产业难以蓬勃发展。 2.算力供应充足，价格下降：目前，中国各地都在建设智能算力中心，包括很多非行业内企业也在投资算力，这导致市场短期内算力供应较为充足，短期价格下降。 3.AI 行业进入大规模应用阶段，算力需求上升：随着 DeepSeek 等大模型的推广，行业正从单纯的数据训练转向应用端的爆发。各行业都在开发 AI 应用，推理需求将持续增长。 从发展趋势来看，AI 行业正在经历类似云计算早期的发展路径。最初，银行等大型企业自建 IDC 机房，但随着云计算的发

	展，越来越多的企业选择租赁云服务器，而不是自建基础设施。AI 算力也将经历类似的演变，从“客户购买裸金属服务器自行部署”逐步转向“租用云端 AI 推理服务”。算力需求的形态正在改变，云平台和推理服务将成为 AI 创业者和企业的主流选择。 总而言之，短期内算力租赁价格仍会下降，因市场供应充足。而长期来看，随着 AI 推理需求持续增长，云端推理成为主流，将逐步替代算力租赁模式。算力需求形式将发生根本性变化，对云平台的技术要求提高，AI 行业将迎来更广泛的商业化落地。 **Q11：如何看待当前中国云计算行业的竞争格局？** A：中国云计算行业可以分为两类：算力租赁和云计算，两者是完全不同的概念。目前，国内云计算行业可以分为两个梯队： 近年来，国内主要的大云厂商逐渐向大模型生态闭环发展。其各自的大模型均由自己的云平台支持。而首都在线的核心竞争力在于 “中立云性”，不涉及自研大模型，专注于提供云计算和算力调度服务，成为各大模型企业的合作伙伴。 **Q12：公司提到推理价格下降，在此趋势下，会对公司的营收造成什么影响？** A：这里所说的推理价格下降，指的是单 Token 的成本下降，而不是服务器租赁价格的下降。这类似于计算机行业的发展逻辑，如今电脑相比上一代，单位处理能力的成本降低了，可电脑整体价格未必降低。同理，AI 算力的发展也是如此，设备的计算能力越来越强，单 Token 成本下降，但 AI 应用的需求量在大幅增长。因此总体仍呈现增长趋势。 总的来看，单 Token 成本下降并不会影响公司的营收，反而会促进 AI 行业的快速发展，有望对公司经营带来正向的影响。 **Q13：当前 AI 算力需求的增长速度如何？如果按季度或年度来看，大致的增长趋势是怎样的？**

	A：AI 算力需求的增长并不是线性增长，而是呈现脉冲式上升或阶梯式跃迁的模式。总体来看，AI 算力需求长期呈上升趋势，但在不同阶段增长速率不同，可以形象地称为“螺旋式上升”。当行业进入新的技术周期，例如 DeepSeek 的发布、GPT-5 等新模型推出，都会带来新一轮的算力需求爆发。而在市场调整期，需求增长可能会趋于平缓。因此，虽然具体的增长速率难以精准预测，但整体趋势是持续上升的，并且会在关键技术突破节点出现快速增长。 **Q14：目前，公司从客户需求及客户结构角度有哪些变化？** A：用户需求整体趋势发生了变化，过去用户主要租赁裸金属服务器，而现在越来越多用户希望直接使用带有模型的云计算服务，对云平台的要求越来越高。主要的变化体现在三个方面： 1.服务模式转变：从单纯租赁算力转变为租赁算力+模型的结合服务。例如，用户不再只是租用算力，而是直接使用 DeepSeek 等大模型进行推理。 2.产品需求转变：基座大模型客户需求从单纯算力租赁，变为如容器产品等更具针对性的服务。 3.客户结构调整：应用类客户和中小企业数量增长迅速，过去云计算业务集中在少数大客户，而现在越来越多的中小企业开始利用云平台进行 AI 应用开发。 **Q15.公司目前与哪些大模型厂商有合作？进展如何？** A：首都在线与多个大模型厂商保持深入合作。以智谱华章为例，作为核心战略合作伙伴，我们的平台承接了其大部分的推理算力需求。随着双方合作的持续深化，业务边界不断拓展，有望达成更大规模的商业价值与技术突破。同时，我们在 AIGC 领域的服务广泛覆盖智能驾驶、图形图像等行业头部企业，并建立了深度合作。 此外，公司会将智谱全系列模型、紫东太初部分模型以及其

	他行业模型逐步加载至我们的大模型平台，并持续吸引其他企业在平台发布模型，不断丰富平台模型资源，最终进行应用优化后推向最终客户。 从合作趋势而言，公司起初多与基础类大模型合作，主要是因为它们对算力需求旺盛。随着业务不断发展，正逐步从推理端向行业应用类模型拓展，如文生图、文生视频相关的模型应用，目前合作态势良好，业务量增长迅速，未来有望进一步深化合作，开拓更多业务可能性。 **Q16：首都在线目前是否与 DeepSeek 有合作？是否有相关接洽或合作准备？** A：公司与 DeepSeek 尚处于沟通过程中，DeepSeek 作为一个大模型，本质上是一个软件，需要与硬件相结合才能提供服务，其本身有算力需求。此外，DeepSeek 的使用并不仅限于中国市场，而是受到了全球范围的关注和应用。这表明，全球市场对高质量大模型的需求非常旺盛，也可能会有出海的需求。我们很希望和 DeepSeek 这样优秀的模型厂商有更加密切的合作。 **Q17：像 DeepSeek 这样的优质大模型出现，会给云计算行业带来哪些变革？** A：在此之前，市场更关注的是算力的硬件资源，比如企业拥有多少张卡、算力规模如何。这意味着企业只要拥有足够的服务器，就能提供算力服务。但 DeepSeek 大模型出现后，市场的需求开始转向软硬件一体化，客户不再单纯关注算力规模，而是更加注重服务方式。未来，行业将从单纯的算力租赁模式，逐步向算力即服务、模型即服务演进。这将对云平台的技术能力提出更高要求，例如： 1.算力调度能力：客户需要的不仅是裸金属服务器，而是能够动态匹配大模型需求的算力环境。 2.模型加载与优化能力：云平台需要具备快速加载、优化和

	管理不同大模型的能力，以支持企业高效运行 AI 应用。 3.数据存储与分发能力：模型训练和推理需要高性能的数据存储和分发系统，以确保高效的算力资源利用率。 因此，DeepSeek 等大模型的出现将推动行业从传统的基础设施提供商，向高附加值的智能算力平台发展。这也正是首都在线“一云多模”战略的核心目标，即构建一个能够灵活适配不同大模型需求的智能云平台。 **Q18：DeepSeek 近期因推理算力不足出现宕机，是否说明其算力需求增长迅速，当前推理能力无法满足？如果公司与其合作，业务量级是否可能大幅提升？** A：DeepSeek 的情况类似于一个非常受欢迎的 APP，由于过于火爆，导致其服务器负载过高。目前，DeepSeek 的模型已经可以在首都在线的平台上进行试用，欢迎大家体验。 同时，DeepSeek 的热度不仅仅局限于国内，在海外也同样火爆。这种趋势对于整个中国大模型产业出海的影响是积极乐观的。由于推理需求的增长，整体算力需求也会呈现扩张。从整体趋势来看，我们非常看好 AI 市场的发展潜力。 **Q19：DeepSeek 已经部署在首都在线的云平台上，是否可能有大量企业在这基础上训练自己的行业模型，并进一步在公司平台上发布，产生裂变效应，其规模和影响有多大？** A：DeepSeek 是第一个现象级的大模型，但绝不会是最后一个。当前的大模型发展趋势表明，这类模型的涌现速度会越来越快，并且很快会扩展到各个行业。基于这一判断，我认为未来类似 DeepSeek 的现象级模型将成为行业常态，而大模型的裂变效应也会随之加速。越来越多的企业会基于基础大模型，进一步开发行业专属模型，并将其部署在云平台上，以支持更广泛的应用场景。这意味着未来的 AI 产业不仅仅是大模型的竞争，更是大模型生态的竞争，而这一生态的发展将直接推动 AI 产业进入真

	正的大规模应用爆发。 公司的“一云多模”战略正是针对这一趋势而设定的，目标是让更多的行业模型能够基于云平台快速部署、迭代和应用。 **Q20：随着大模型行业的发展，尤其是 DeepSeek 等热门模型的快速增长，首都在线的收入是否会进入一个量级的增长阶段？** A：互联网行业的快速发展确实创造了许多行业巨头，但同时也有大量企业失败。市场将经历分化期。关于公司业绩情况，我们会根据相关规定和公司实际情况进行披露。 **Q21：目前用户主要使用大模型来做哪些具体应用？是程序员编程，还是金融分析，或者其他行业应用？** A：目前 AI 应用已经涉及多个行业，各种场景都在逐步涌现。 1.互联网行业：文生视频、文生图等 AI 生成内容（AIGC）应用广泛增长。 2.电商行业：AI 模特生成，优化产品展示效果。 3.无人机行业：利用 AI 进行自动化控制和数据分析。 4.社交应用：如某些 APP 可以基于大模型提供“高情商聊天助手”，指导用户如何优化社交对话。 过去，很多应用并未广泛推广，主要因为算力成本较高，影响了用户规模。随着算力成本下降，这类应用有望迎来更广泛的增长。此外，AI 的普及正在改变日常生活。例如，过去推出 4G 时，许多运营商认为市场需求有限，但随后移动支付、短视频等应用爆发，使得 4G 和 5G 网络需求大幅增长。同样，现在我们可能还未完全看到 AI 的最终应用形态，但随着大模型的发展，现象级 AI 应用的出现只是时间问题。 **Q22：DeepSeek 在海外市场的发展是否会带来新的算力需求？是否有初步的规模预估？** A：DeepSeek 不仅是中国的现象级大模型，也是全球范围内

	的现象级模型。它带来了一个重要趋势：全球市场对中国 AI 技术的认可度的提升，以往用户可能会选择 ChatGPT，现在 DeepSeek 也成了新的选项。这为中国模型企业进军海外市场提供了更大的信心，促进更多企业进入国际市场。 具体的算力需求规模，目前难以准确预测。市场的发展不是线性增长的，而是呈现阶段性爆发。例如，某些时间段需求会急剧上升，而在另一些时间段则会趋于平稳。因此，具体数据需要市场进一步验证。不过，从整体趋势来看，AI 算力需求的未来增长乐观，期待 AI 行业迎来更大的突破。 **Q23：首都在线最初是如何决定开展出海业务的？目前的进展如何？未来对此业务有哪些规划？** A：首都在线的出海业务既是有计划地布局，也是一定程度上的运气使然。出海业务的优势在于公司在云计算的 CPU 阶段就已经较早地进行全球布局，建立了覆盖全球的网络，目前拥有近 100 个数据中心的边缘节点。这些节点在后续算力发展进程中，成为模型企业出海的关键支撑点，不仅为后续合作打下坚实基础，还逐步演变成公司的核心竞争力之一。 随着模型企业的全球化发展，越来越多的公司，如 MiniMax，因在海外发展态势良好，需要依托支点来拓展业务，公司前期布局的节点恰好满足了这类企业的需求，使公司在算力出海领域占据了一定先机。因为如果现在才开始进行全球算力部署，可能已经赶不上市场需求的快速增长。 在此背景下，模型出海已然成为当下的又一主流趋势，公司凭借多年在海外的布局，将其转化为独特优势，通过全球网络覆盖和边缘节点的合理运用，有力地助力企业出海发展。
附件清单（如有）	
日期	2025-02-09