

证券代码：300166

证券简称：东方国信

北京东方国信科技股份有限公司投资者关系活动记录表

编号：2023-001

投资者关系 活动类别	<input type="checkbox"/> 特定对象调研 <input type="checkbox"/> 分析师会议 <input type="checkbox"/> 媒体采访 <input type="checkbox"/> 业绩说明会 <input type="checkbox"/> 新闻发布会 <input type="checkbox"/> 路演活动 <input type="checkbox"/> 现场参观 <input checked="" type="checkbox"/> 其他（电话会议）
参与单位名称及 人员姓名	具体情况详见附件。 因本次投资者调研活动采取线上电话会议形式，线上参会者无法签署调研承诺函。但在交流活动中，我公司严格遵守相关规定，保证信息披露真实、准确、及时、公平，没有发生未公开重大信息泄露等情况。
时间	2023年2月21日、22日
地点	公司会议室
上市公司接待 人员姓名	副总经理、董事会秘书：刘彦斐 CTO：查礼
投资者关系活动 主要内容介绍	交流的主要内容： 1、问：东方国信一直是一家大数据的公司，公司大数据技术和ChatGPT能产生哪些关联？ 答：东方国信是一家大数据公司，多年的发展中沉淀了很多大数据的相关技术能力和产品。我想从两个方面来说明。首先，我认为1-2年之内国内会有类似ChatGPT模型出现。昨天我看到信息说复旦大学已经开放了一个类ChatGPT的系统对外开始公测。现在主要解决的问题是，像我们做To B业务的公司最想要看到的进步是，ChatGPT这样的模型或者系统小型化、轻量化。这个路径从技术角度来说应该是可行的，说白了不需要那么大量的知识或者那么大量的参数来存储知识关联或者知识本身。可能在某些垂直领域，比如像手机银行的智能助手或者

智能客服，它只需要知道我这个垂直行业里的专业知识就够了。还是类比人脑神经元，实际不需要那么多的参数，只需要提供1%的参数或者神经元就能去做这个事情。如果把ChatGPT小型化、轻量化到一台服务器就可以承载几千或者几万人的使用，这个系统在我们东方国信客户场景下就会很快的落地，能够去适配和使用。这个工作实际是比较有挑战性的，怎么样去收集垂直领域的知识，并且把这些知识能够和现在ChatGPT里面的预训练模型、奖赏模型，能够去做对接，能够去做训练，这是比较有挑战性的事。包括复旦大学他们公测出来的系统，自己也说了英文回答要比中文要好，因为英文训练的词量比中文多很多，所以回答中文问题时没有像英文问题那么更像是人说话的样子。这是ChatGPT小型化和轻量化以后，我认为它就可以和我们现有的很多业务对接，像智能客服这样的业务，包括像一些我们在To B领域做的平台，在大数据平台里面比如做数据分析，可能我不知道数据分析应该用什么样的算法或者路径去做，可能就嵌一个类轻量化的ChatGPT来回答我的问题，甚至于让它能够去帮我把相关的代码片段，或者直接把结果告诉我，这都是非常有帮助的。我们有很多的应用场景实际是在做数据分析，目前数据分析已经演化到了第二代，第一代是纯手工的，第二代是用工具做数据分析，第三代智能化数据分析是什么？想象一下，以后直接跟智能助手说你帮分析一下上个月财务情况，上个月用户增长情况和预测，你就跟它说一句话或者写一段问题，它就直接把结果给你，并且告诉这个结果是怎么得来的。目前数据分析还需要工程师操作工具完成，这个结合点是非常明确的，ChatGPT如果小型化了以后，和我们的业务结合点是非常明确的。

另外，本身大数据技术这个层面我认为也是有很强的关联。ChatGPT它的训练过程，一个是依赖大量的语料，训练出大型的语言模型，在大型语言模型里面做微调，它叫基于人类反馈的强化学习。用这项技术，实际就是用人回答加深它对知识的理解和正确性的判断。准备这些语料，实际这是庞大的数据集。我们知道ChatGPT3.5训练模型用到的数据是上百TB，用了哪些呢？有百科，有电子数据，还有网上公

开的数据集，那个数据集非常的庞大。这些数据本身要去做一些预处理的话必须要用到大数据技术。再举个例子，如果说我们把ChatGPT模型小型化以后，放到垂直领域用的时候，我也是需要对这个模型进行微调的。这个模型本身是通用模型，如果想要通用模型在特定领域完成特定任务就需要微调，所谓微调实际就是对它进行知识强化。告诉它我这个垂直领域有哪些知识需要去记住，并且对知识相关问题有答案的关联关系你要记住。这些知识的整理，尤其是垂直领域知识整理，不一定是问答方式，有可能是知识图谱形式构建的。全世界已知蛋白质有几百万量级，它就构成了特别强大的知识图谱，如果你想要ChatGPT回答蛋白质领域问题的时候，就需要把这个知识图谱作为微调的训练语料做再训练。做了再训练以后，相当于说教会了模型蛋白质领域，某一个蛋白质由哪个蛋白质衍生而来，而它自己本身又能衍生出几种蛋白质，这些都记入在模型里。训练好的、微调好的模型就可以用来回答专业领域的知识，而且非常精准。准备语料的时候，知识图谱就是大数据的应用，相关的语料收集和整理占到了整个模型训练80%的时间和成本。一般来讲现在深度学习模型训练，做数据整理、清洗、加工、标注，占到过程里面的80%的时间和成本。这样来看，我们现有的大数据技术在整个产业链当中是必备的一个环节。我们能够估计到的垂直领域里面，比如像目前东方国信的客户，运营商、金融、工业都是非常典型的垂直领域，在这些领域的知识我们沉淀的是非常多。如果未来有人把ChatGPT做小型化、轻量化以后，我们就可以把以前积累的、沉淀的这些知识作为微调的输入，形成我们自己能够使用的ChatGPT模型。

**2、看到公司也有一些AI相关技术，公司现在AI有哪些技术储备？
包括公司未来可预期的相关应用场景有哪些？**

答：我们的理解人工智能它的范畴比较广，原来东方国信承接的基于关系数据的挖掘和分析业务使用到的人工智能技术就是我们所说的传统的机器学习，比如说分类聚类、有监督无监督的机器学习等。深度学习使用的方式和传统机器学习不太一样，深度学习技术兴起以

来，我们也一直在关注和跟进。同时公司也在储备深度学习技术，这两块形成了公司最近三年时间形成的产品“数据科学云”，对外的产品名称叫“图灵引擎”。这个技术储备覆盖了之前已知的传统机器学习的算法和现在常用的深度学习算法，公司都有技术储备。

人工智能已在客户中有成熟的应用，主要应用场景现在一个是精准营销。我们用模型的办法来判别哪些是高价值用户，比如说运营商或者银行的高价值用户。在精准营销里面要对人群进行分类，没有人工智能之前是通过规则的方式，打标签的方式划定人群分类。目前通过深度学习算法可以更精准地去对人群进行分类，从而使得我们的活动投放或者广告投放成本降低，效率提升，这是明显的变化。我们主要的客户是运营商、金融，精准营销是一个很好的场景。

另外一个场景是工业领域，公司在工业互联网领域也有非常深厚的积累。东方国信已连续多年保持国家级工业互联网双跨平台前列，在业内处于领先地位。在工业领域除了机器学习以外，用的比较多的是工业视觉，比如在工厂对产品进行质检，以前的方式都是用人眼，老师傅拿个放大镜看这个东西有没有瑕疵。现在采用人工智能对图像物体的识别，我们管它叫“工业质检”或者“瑕疵检测”，效率和准确度都有非常明显的提高。还有场景比如智慧煤矿，在煤矿的场景底下，不光是视觉领域，我们还接入了非常多的传感器，通过这些传感器采集到的设备的时序数据，通过这些时序数据我们能够用机器学习的方式对设备进行预测性的维护，能够提前预知这个设备到底是有问题还是没问题，短期内会不会有问题，何时维修是最合理的。智慧煤矿也是主要的AI落地大场景。

3、公司自研的CirroData数据库具有哪些优势和应用？

答：CirroData分布式数据库自2011年研发之初就始终坚持走纯自研路线。十余年中，CirroData聚焦核心技术持续创新，从单一产品走向全栈服务，终端用户体验不断提升。CirroData是一款面向海量数据分析型应用领域的国产数据库。采用了先进的计算与存储分离技术架构，融合了分布式存储和MPP并行计算的各自优势，不但可以轻松实现

云平台上的伸缩扩展，而且可以提供按需部署的能力，是新一代云数据仓库的典型代表。2022年，CirroData数据库家族六款产品扎实根基，提升性能，完善功能；向上拓展生态体系，运维、数据治理和开发三大平台形成合力；顶端整合数据云平台、BEH大数据平台、数据科学平台、工业互联网平台、云计算与知识图谱平台，多场景应用。

2022年，CirroData共新增发明专利授权5项，新增软著17项；CirroData-TimeS 时序数据库通过信通院基础能力测试；北京银行“京Lake”数据湖平台通过信通院“云原生数据湖基础能力专项测评”；“中国联通可视化大数据洞察自助服务平台”、中国核电的“核电行业数据资产管理及应用”荣获信通院“星河”优秀案例。

在应用方面，目前CirroData数据库已经在百余家客户的生产环境上线，服务数百个项目，在通信、金融、政府、公共安全、智慧城市、能源、互联网等十余个行业落地。例如，中国联通“可视化大数据洞察自助服务平台”对标行业顶尖BI（商业智能）工具，基于中国联通B/M/O/D全域数据资产，为多层次用户提供千人千面的自助化、智能化数据分析服务能力的全流程企业级商业智能平台，覆盖中国联通31个省分公司、集团多部门和子公司，访问数据规模达3PB。“可视化大数据洞察自助服务平台”充分体现了联通能力、联通智慧。客户的成功是对底层支撑平台最大的肯定。东方国信数据科学云平台和CirroData-OLAP作为基础软件产品，实现了支持交互式探索的海量数据加工，和多用户高并发场景下的自助洞察分析探索操作，经受住了巨大访问量和复杂业务数据分析的考验，获得客户好评。国家开发银行大数据平台是以CirroData分布式数据库为底座的国产化自主可控的数据平台，2022年在原有建设基础上进行了扩容、升级和数据迁移，全面建成数据湖区、数据仓库集市区、应用区等，目前平台上运行有几十个系统。柳工数据湖建设项目是柳工打造基于全数字化驱动的智能制造战略中的关键一步。数据湖采用CirroData-OLAP数据库作为底层数据存储与分析计算平台；采用CirroData-TimeS时序数据库采集和计算车辆实时数据，打造智能车联网平台，目前已接入测点数超过六

	<p>千万。大核源平台是中国核电集团的重点数字化建设项目，用以满足核电厂运行海量数据采集、汇聚、分析的工业互联网平台。CirroData-TimeS时序数据库承担对生产中产生的测量数据进行采集和处理的工作：支持对不同时间戳序列对齐存储和查询，增强多时间序列分析的能力；提供多种UDF算子，计算下沉到数据库，提升计算速度，减轻业务负担。应用案例较多，不一一列举了。</p> <p>4、公司展望下今年的订单和业绩情况？</p> <p>答：随着疫情的缓解，公司经营也逐步恢复正常。随着企业数字化转型的推进和深入，未来将有更多的市场机会。国家相关政策的逐步出台和落地，也有利于公司开拓更大的市场。公司对未来的发展充满信心，相比于 2022 年，今年的业绩情况也会有大幅好转。</p> <p>5、公司今年的人员招聘计划怎样，人数是否会继续增长？</p> <p>答：公司将根据业务发展规划、企业经营需要以及市场环境来合理匹配人员配置，以经济效益为导向，优先从内部协调人员匹配需求。同时继续加强精细化管理，在保证企业良性发展的同时严格控制人力成本。总体上讲，公司人数保持平稳，未计划大幅人员扩张。</p> <p>6、业绩预告提到的几个主要子公司上海屹通、北科亿力等，2022 年业绩情况怎样，今年经营情况如何？</p> <p>答：2022 年子公司也是主要受疫情影响，业绩出现了下滑。这几个子公司的经营本身都稳定，在各自的细分领域都保持领先地位，随着疫情的缓解，经营将逐步恢复正常。</p>
附件清单(如有)	参会名单
日期	2023 年 2 月 21、22 日

附件：参会名单（排名不分先后）

姓名	公司	姓名	公司	姓名	公司
王益聪	名禹资产	卢杰	厦门财富	张鸿运	仁桥资产
王晓卓	易鑫安	卜乐	东方马拉松	曹国军	天猊投资
孙磊	中融基金	闫鸣	宝新资产	董丰侨	方圆基金
俞科进	中邮基金	程漫漫	太平洋证券	杨佳	友邦保险
Wise LYU	King Tower	曹佩	太平洋证券	陈勇	兴证全球基金
朱凌昊	东证自营	薛楠	明曜投资	李亚军	民森投资
何恒	汐泰	刘光谱	中国国际金融	田林矗	嘉实基金
朱义	富安达	朱伟华	鸿运私募基金	芮宁	安粮集团
杨晨	建信保险	阳桦	中银国际资管	郭琦	财通证券
马泽儒	泽铭投资	董涛	诚实资产	黎焜	力元股权投资
朱程辉	华富基金	杨嘉	国信弘盛	简桢	诺安基金
钟吉芸	泰康养老保险	李凯	旭日利邦	梁斌	中信建投
郭帅彤	建信基金	甘静芸	创金合信	黄伟	诺德基金
杨杰	长江证券	许耀文	诺昌投资	黄伟	诺德基金
马翔	汇添富	沙小川	聆泽投资	李昕蔚	太平洋证券
朱战宇	国寿资产	郝彪	誉辉资本	李昕蔚	太平洋证券
韩旭	珺则投资	程超	招商基金	任兆凯	柏治投资
崔帅	人寿资产	姚千程	恒复投资	祝天骄	盟洋投资
房晓	华安财保	简桢	诺安基金	麦浩明	航长投资
车克	泰石投资	丁俊	展博投资	王灿	誉辉资本
张晓	拓璞基金	周宁康	利位投资	田发祥	太平基金
王龙	名禹资产	韩帆	野村东方	莫启杰	惠正私募基金
傅晟	华富基金	陈秋芝	领久基金	柴志华	谦心投资
张鸿运	仁桥资产	李梦莹	鼎锋资产	王晓卓	易鑫安

李煜	信达新兴	单开佳	平安理财	高松	中欧基金
杨君桂	中财均策	胡建芳	磐厚动量	吴亚雯	工银国际
龙文利	正德泰投资	戴俊清	厦门财富	许高飞	中天汇富
吴桐	西部利得	赵朝侠	星壤资产	江昕	德睿恒丰
李晗	创金合信	盛 豪	华泰柏瑞	李迪心	汇丰晋信
黄伟	诺德基金				