

证券代码：300229

证券简称：拓尔思

拓尔思信息技术股份有限公司  
投资者关系活动记录表

编号：2023-003

|               |                                                                                                                                                                                                                                                                                            |
|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 投资者关系活动类别     | <input checked="" type="checkbox"/> 特定对象调研 <input type="checkbox"/> 分析师会议<br><input type="checkbox"/> 媒体采访 <input type="checkbox"/> 业绩说明会<br><input type="checkbox"/> 新闻发布会 <input type="checkbox"/> 路演活动<br><input type="checkbox"/> 现场参观<br><input checked="" type="checkbox"/> 其他（电话会议） |
| 参与单位名称及人员姓名   | 银河证券：吴砚靖、赵巧敏；银河创新资本：肖挺；申万宏源证券自营：杨晓彤；本炎投资：李刚；东方证券：浦俊懿、谢忱、宋鑫宇；万家基金：耿嘉洲、胡文超、陈飞达、况晓、郎彬、刘洋、汪洋、武玉迪；鑫元基金：龙凌波；兴业证券：桂杨、唐梓娟、杨本鸿；建信基金：程广飞、邱宇航、孙晟、李登虎、张文浩、黄子凌、左远明、张湘龙、刘洋、徐文琪；人保资产：李京洋。                                                                                                                 |
| 时间            | 2023年3月6日 14:00-15:00；<br>2023年3月7日 16:00-17:30；<br>2023年3月8日 15:00-16:00；<br>2023年3月9日 16:00-17:00；<br>2023年3月10日 10:30-12:00。                                                                                                                                                            |
| 地点            | 拓尔思信息技术股份有限公司会议室                                                                                                                                                                                                                                                                           |
| 上市公司接待人员姓名    | 副董事长兼总经理施水才先生；<br>副总经理林松涛先生；<br>副总经理、董事会秘书李党生先生。                                                                                                                                                                                                                                           |
| 投资者关系活动主要内容介绍 | 一、副总经理、董事会秘书李党生先生就公司主营业务、核心产品及优势、未来战略规划进行介绍。<br>二、参会机构提问及回答。                                                                                                                                                                                                                               |

**（一）在数据要素市场发展势头强劲的大背景下，请问拓尔思有哪些发展规划？**

回答：拓尔思在经营过程中形成了一套 1+1=N 的数据要素商业模式。“1+1”是指一个大数据底座和一个人工智能技术平台。“N”是指行业 SaaS 服务集群，未来可实现 N 个云服务产品的拓展。

“1+1=N”的商业模式主要依托拓尔思自建的大数据中心，基于自研的大数据底座和人工智能技术平台，面向优势行业的业务应用，根据不同场景封装产品，基于公司成熟营销体系迅速推向市场。这种模式充分体现了拓尔思的数据要素具有可复制性强、迭代速度快、复用价值高以及无限供给等特点，实现了统一数据在产品化方面的裂变增长。

目前，拓尔思拥有完整的大数据基础建设产品线、丰富的行业 SaaS 服务集群和可靠的安全一体化平台。公司将以完善自身数据要素资源体系、壮大数据要素应用体系及建立数据要素安全体系为发展目标，从资源、应用、安全三个方面全面布局数据要素市场。

基于对 1300 亿以上资源性数据资产，通用、行业/领域知识库大类 30 余种，以及 350 种以上深度学习算法模型的融合治理，公司目前已形成了产业大脑、媒体资讯、网络舆情三大数据资产平台，为公司在网络信息内容治理、金融监管、政务资讯服务、金融机构风控管理、智能媒体服务、开源情报等领域的云和数据服务提供数智赋能。

**（二）公司积累了海量数据资产，请问这些数据来源于哪里？**

回答：目前公司积累了 1300 亿条基础数据，日均增加 1 亿条，以上数据全部来源于互联网公开数据。拓尔思数据中心目前拥有 3 大专业 IDC 机房，上千台高端机架式服务器完成对互联网公开数据的 7\*24 小时不间断采集，同时租用了阿里云及微软云的服务器实现全球数据采集。数据中心监测的采集站点包括 80 万余家新闻网站、3000 余个资讯 APP、1200 余家国内报刊、20 余个第三方平台、79 语种 50 万余个其他站点。每日数据量高达 1 亿多条，几乎涵盖了全量的微博新增数据，网站覆盖度及数据存储量均明显高于业内水平。

**（三）公司积累了丰富的数据处理经验，请问将服务客户积累的数**

**据用于变现是否存在数据纠纷风险？**

回答：拓尔思基于自然语言处理技术、积累的各场景行业术语及主题数据形成的各行业知识图谱,对海量多源异构数据进行融合、关联、标注以及知识化处理,构建形成大规模领域知识图谱。目前,公司拥有通用、行业/领域知识库 30+大类,标签规则 30000+。每类知识库具有规模大、语义关系丰富、结构友好、质量精良等特点。公司向客户提供的是以自然语言处理技术为核心的软件产品和数据中台产品,并不接触客户的内部数据,因此不存在数据纠纷风险,且从未产生数据纠纷。

**（四）请问公司的数据采集方式是自动采集还是人工采集？**

回答：拓尔思大数据中心的数据是基于自研的互联网数据采集系统实现全网数据全自动采集的。早在 2002 年,拓尔思就推出了第一代网络信息采集系统——TRS 网络信息雷达,已成功应用于数千家媒体、企业、金融和政府部门的数据中心或信息中心。2020 年,拓尔思推出了 TRS 海蜘蛛大数据采集平台(以下简称“海蜘蛛”)。海蜘蛛是一款专业的网络数据深度采集产品,具有智能采集、全网适用、简单易用、高效稳定的特点,可提供多种网页采集策略与配套资源,帮助整个采集过程实现数据的完整性与稳定性。“海蜘蛛”根据互联网不同的数据类型与网站结构,采用分布式抓取技术,对指定网站定向数据采集、存储,然后进行分类、分词、查重、打标等数据加工,在专业知识库建立、企业竞争情报分析、报社媒体资讯获取、舆情监测等领域应用广泛。同时,公司掌握高质量、具有公信力背书的数据采集点,在数据的质量和行业关联度方面具备优势。

**（五）请问公司在数据标注方面是人工标注还是机器辅助？**

回答：在大数据领域,目前多数公司在数据处理方面以人工为主。拓尔思作为以自然语言处理为核心技术的公司,公司核心技术的对内优势主要体现为提升数据处理效率。公司利用深度模型和算法已成功打造多款数据中台,大力拓展数据治理和自动化清洗功能,以机器代替人工的方式实现数据自动打标,已累计智能标签模型 120 类,共计 2800 余个。目前公司采取机器+人工的数据标注方式,且机器标注率已经达到 80%,降本增效成果明显。

|          |                 |
|----------|-----------------|
| 附有清单(如有) | 无               |
| 日期       | 2023 年 3 月 10 日 |