

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
2023 年度向特定对象发行股票
募集资金使用可行性分析报告**

二〇二三年八月

拓尔思信息技术股份有限公司（以下简称“拓尔思”或“公司”）拟向特定对象发行股票募集资金。公司董事会对本次向特定对象发行股票募集资金使用的可行性分析如下：

一、本次募集资金使用计划

本次向特定对象发行募集资金总额不超过人民币 184,481.67 万元(含本数)，扣除发行费用后的募集资金净额将用于以下项目：

单位：万元

序号	项目名称	投资总额	使用募集资金投入
1	拓天行业大模型研发及 AIGC 应用产业化项目	184,481.67	184,481.67
合计		184,481.67	184,481.67

在本次向特定对象发行募集资金到位之前，公司将根据募集资金投资项目进度的实际情况以自筹资金先行投入，并在募集资金到位后按照相关规定的程序予以置换。本次发行募集资金到位后，若扣除发行费用后的实际募集资金净额少于拟投入募集资金总额，不足部分由公司自筹资金解决。

二、本次募集资金投资项目基本情况

（一）项目基本情况

项目名称：拓天行业大模型研发及 AIGC 应用产业化项目

实施主体：拓尔思信息技术股份有限公司

项目投资总额：184,481.67 万元，拟使用募集资金投入 184,481.67 万元

项目建设期：36 个月

项目建设地点：北京市海淀区建枫路（南延）6 号院 3 号楼

项目建设内容：本项目拟基于公司的海量高质量行业数据、行业知识体系等知识型资产，在通用大模型的基础上构建拓天行业大模型，并实现生成式 AI

(AIGC)在媒体、政务、金融等领域的产业化应用。公司将充分发挥自身行业数据及行业专业知识优势,搭建多模态、高质量训练数据集协同构建系统,作为拓天行业大模型训练的多模态数据实时采集、治理和标注的数据资源底座,以保证拓天行业大模型的专业能力与可靠性,充分发挥公司在数据要素和人工智能领域成果的协同作用,进一步提升公司的核心竞争力和服务能力,从而助力数字强国、网络强国战略的落地实施。

(二) 项目建设的背景、必要性及可行性

1、项目建设背景

(1) 人工智能是助推数字中国建设的重要引擎

自党的十八大以来,国家高度重视发展数字经济,并逐步将数字经济上升为国家战略,成为数字中国建设的重要路径。为进一步加快落实发展战略,相关政府部门出台《网络强国战略实施纲要》《“十四五”数字经济发展规划》等文件,从数字基础设施、大数据、产业数字化转型、数字产业化等多方面规划了数字经济发展前景,并强调人工智能、大数据等关键技术创新能力的增强,为数字经济发展指明方向。党的二十大报告亦明确指出,加快发展数字经济,促进数字经济和实体经济深度融合,加快网络强国、数字中国建设。2023年2月,中共中央、国务院印发《数字中国建设整体布局规划》,数字中国整体布局规划落地,夯实数字基础设施和数据资源体系的战略意义,推动数字技术与经济、政治、文化等深度融合,强化数字技术创新体系,助推数字中国落地。

人工智能作为引领新一轮科技革命和产业变革的战略性数字化技术,能够通过大数据、云计算等信息技术的融合,带动各产业的数字化转型,在数字经济建设中发挥着重要作用。我国政府先后发布《新一代人工智能发展规划》《关于促进人工智能和实体经济深度融合的指导意见》《关于加快场景创新以人工智能高水平应用促进经济高质量发展的指导意见》《关于支持建设新一代人工智能示范应用场景的通知》等相关政策,大力促进人工智能与实体经济深度融合,推动

人工智能在数字经济中的应用，为全面建设数字中国提供动力。人工智能成为助推数字中国建设的重要引擎。

(2) 开发行业大模型是当前实现 AI 大模型落地应用的重要路径

大模型技术具备强大的泛化性、通用性和实用性，能够降低 AI 开发门槛，提高模型精度和内容生成质量，实现对传统 AI 技术的突破，有效解决在复杂业务场景下传统小模型无法解决和理解的问题，逐步成为人工智能领域的重要发展方向。自 2019 年 OpenAI 推出 GPT-2 大模型后，国内高科技企业也相继布局 AI 大模型领域。且随着大模型应用场景的爆发，AI 大模型逐渐由 NLP 领域向 CV 及多模态领域演进。2023 年 3 月，在 OpenAI 正式推出多模态大模型 GPT-4 之后，谷歌、微软、百度、阿里、腾讯、华为等海内外科技公司纷纷加速了多模态通用大模型的研发，并陆续推出各自的大模型。

通用大模型虽然具有良好的泛化能力，但因其主要应用大规模无标注数据进行训练，缺乏特定场景语料和数据集训练的模型调优过程，在特定场景下其准确性、真实性和专业程度等相对欠佳。且现有通用大模型的训练需要依赖高质量大规模的数据和高性能的算力资源，在模型的训练中还存在大模型无法访问外界知识，知识更新不及时、灾难性遗忘等大模型与知识库融合问题，训练成适用于高质量特定领域的大模型所需的参数更新成本高、时间周期长、经济性差，很难快速形成落地应用。而行业大模型在特定的行业和场景下，在通用大模型的算法和逻辑基础上进行高质量实时更新的数据和知识体系训练与调优后，具备生产全面、准确、可信具有正确价值观内容的能力，表现力、可控性和应用价值更好，能够快速实现应用落地，有效赋能各行业前端业务场景，而成为当前实现 AI 大模型落地应用的重要路径。腾讯、百度等行业内高科技企业也纷纷在通用大模型的基础上开展行业模型以期加快 AI 大模型的应用落地。

(3) AIGC 产业化应用将促进数据要素价值的进一步释放

在构建以数据为关键要素的数字经济过程中，数据已发展成国家进行数字化转型和实施数字强国战略的重要生产要素。为进一步加强数据要素应用价值和数据要素的高质量发展，《中共中央国务院关于加快建设全国统一大市场的意见》

《中共中央国务院关于构建数据基础制度更好发挥数据要素作用的意见》等文件中均强调加快培育数据要素市场，构建数据基础制度体系，为进一步释放数据要素价值、激发数据要素潜能指明了方向。

数据要素是指在大数据经历采集、加工与分析处理后，具备应用价值的数据资产。作为促进数字经济与实体经济融合的关键技术，人工智能技术在不断升级过程中，将大量高质量数据要素应用于大模型算法的升级迭代。同时，基于大模型的 AIGC 也主要通过人工智能算法将数据要素运用于各实体产业的数智化发展，推动各领域的创新。因此，基于大模型的 AIGC 能够实现数据要素与各产业的多维互动、融合渗透，充分挖掘与释放数据要素的价值。

2、项目必要性

(1) 强化公司数据要素与人工智能的协同作用，助力数字强国建设

党的二十大以来，国家不断夯实数字基础设施和数据资源体系，强化数字技术创新体系，加快网络强国、数字中国建设。人工智能作为引领新一轮科技革命和产业变革的战略性数字化技术，不断通过与大数据的融合，带动各产业的数字化转型。而数据要素作为人工智能模型训练的核心资源，是人工智能应用和数字中国建设的基础。

公司运用大数据、人工智能等赋能各行业的数智化，并在此基础上形成了以自研大数据底座和人工智能技术底座为基础的技术架构，为各行业数智化发展提供人工智能、大数据和数据安全产品。为进一步加强人工智能与大数据的融合，公司拟通过本项目的实施，充分利用公司积累的数据要素资源与行业知识体系、人工智能技术，构建多模态、高质量数据训练集协同构建系统，开展多模态行业大模型的训练和建设，以进一步增强公司数据要素与人工智能的协同作用，发挥公司高质量数据要素的价值，更好地赋能各行业的数智化发展，从而助力数字强国建设。

(2) 建设拓天行业大模型，不断提高公司竞争优势

大模型具备强大的泛化性、通用性和实用性，能够有效解决繁杂、碎片场景

下的多任务处理成本高、精度不佳的问题，逐步成为人工智能领域的重要研究方向。国内外高科技企业相继开展 AI 大模型的布局，自 2023 年 OpenAI 正式推出多模态大模型 GPT-4 以来，谷歌、微软、百度、阿里、腾讯等海内外科技公司纷纷加速了多模态通用大模型的研发，并逐步开始探索在通用大模型的基础上建设多模态行业大模型，以弥补通用大模型在特定行业应用的局限性，实现大模型的产业化落地应用。

公司作为以人工智能和大数据赋能行业数字化转型的高科技企业，在人工智能技术快速发展的浪潮下，积极拥抱 AI 大模型发展。在开源通用大模型的基础上，借助公司 NLP、知识图谱、OCR、图像视频结构化等领域自主可控的多模态内容处理底层技术、AI 工程化建设经验，以及公司积累的千亿级规模高质量经营性数据资产，建设政务、媒体和金融领域的多模态拓天行业大模型，从底层架构上升级公司的人工智能技术平台，抢占行业大模型商业化应用先机，从而不断提高公司的竞争优势，巩固公司的竞争地位。

(3) 有利于加速公司 AIGC 商用，满足下游客户智能化需求

在当前数字经济与实体经济加速融合的背景下，各行业对数字内容总量和丰富程度整体需求不断提高。随着人工智能技术的升级迭代和大模型架构的应用，AIGC 逐步打破原有模板化、公式化、小范围的限制，转变为具备真实性、多样性、可控性和组合性的多模态内容生成方式，可进一步助力各行业数智化转型和生产效率的提升。

本项目中，公司将借助历史积淀的海量优质数据资产与行业知识，着力解决目前大模型虚幻生成的问题，提升内容事实核查与修改的能力，构建生成内容全面、准确、可信且具备正确价值观的行业大模型及 AIGC 应用，实现从行业大模型搭建到落地到应用场景的闭环，加速公司 AIGC 应用的产业化落地，为政务、媒体和金融领域各细分场景的智能化赋能，满足其数智化需求。

3、项目可行性

(1) 相关国家政策支持 AI 产业发展为项目建设提供良好的宏观环境

人工智能是各行业推进数智化转型的关键性支撑技术，在数字中国建设中发挥着重要作用，是各行业的重点发展方向。我国政府先后发布《新一代人工智能发展规划》《关于促进人工智能和实体经济深度融合的指导意见》等相关政策大力发展人工智能先进技术，促进人工智能与实体经济深度融合。2021 年 3 月，《中华人民共和国国民经济和社会发展第十四个五年规划和 2035 年远景目标纲要》指出要强化国家战略科技力量，加强人工智能等前沿领域技术攻关，推动人工智能等数字化产业发展。2022 年 3 月，国家发改委发布《“十四五”数字经济发展规划》，强调人工智能、大数据等关键技术创新能力的增强，为数字经济发展指明方向。为进一步加强人工智能应用落地，政府部门又发布了《关于加快场景创新以人工智能高水平应用促进经济高质量发展的指导意见》《关于支持建设新一代人工智能示范应用场景的通知》加强人工智能场景创新，着力打造人工智能重大场景。随着人工智能技术的不断升级，大模型逐渐成为人工智能发展的重要方向。为抢抓大模型发展机遇，2023 年北京市政府部门发布《北京市促进通用人工智能创新发展的若干措施（2023-2025 年）（征求意见稿）》提出充分发挥政府引导作用和创新平台催化作用，整合创新资源，加强要素配置，推动通用人工智能领域实现创新引领。

本项目进行拓天行业大模型的研发和 AIGC 应用的产业化落地，主要针对人工智能技术升级和人工智能应用场景的创新发展，符合国家及地方相关政策的指引，相关政策亦能为本项目的开展提供良好的宏观环境。

(2) 通用大模型的不断成熟为行业大模型的快速落地奠定基础

开发通用大模型需要大量的算力、巨大的数据量以及高昂的推理训练成本，而应用开源的通用大模型开发行业大模型，则可以大幅提升行业大模型的开发与落地应用效率，从而更快速地实现使用行业高质量的数据解决特定领域或场景的问题，加速各行业的数字化转型与数智化发展。为加强我国大模型技术能力的提升，国内多所高校、科研机构及科技企业等陆续推出了多个开源大模型，如清华大学的 ChatGLM-6B、复旦大学的 MOSS 系列大模型等。

本项目拟基于开源通用大模型，结合公司积累的千亿级高质量行业数据、知识图谱等，研发通用大模型与行业知识库的融合技术、进一步解决大模型虚幻生成问题、持续提升行业大模型训练及调优、推理等能力，构建适用于政务、媒体和金融等领域的拓天行业大模型，并实现大模型的行业落地应用。前述不断成熟的开源通用大模型，为本项目的实施奠定了坚实基础。

（3）人工智能产业快速发展，生成式人工智能具备广阔的市场前景

近年来，人工智能作为赋能各领域产业升级的关键核心产业，规模呈现爆发式增长。根据商汤发布的《AI 大模型市场研究报告》，2022 年中国人工智能行业市场规模达到 3,716 亿元，预计在下游政务、媒体、金融等多领域数智化发展的驱动下，人工智能将实现大规模的落地应用，预计到 2027 年整体市场规模高达 15,372 亿元，2021 至 2027 年实现 34% 的年均复合增长。

而随着技术的积累和迭代，以及大模型应用逐步落地，人工智能也逐渐突破传统分析型，向生成式演化。生成式 AI 因具备数字内容孪生、数字内容的智能编辑、数字内容的智能创作而被广泛应用于实体经济的各领域，生成式 AI 商业应用规模迎来快速增长。根据中关村大数据产业联盟发布的《中国 AI 数字商业展望 2021-2025》报告分析，到 2025 年，我国生成式 AI 商业应用规模将达到 2,070 亿元，未来五年的年均增速 84%，未来市场空间广阔。

（4）公司千亿级高质量数据资产是构建高质量行业大模型的重要保障

公司 2010 年自建大数据中心，以长期服务多行业用户持续累积的公开信源数据为基础，拥有了规模及质量均位列业界前茅的公开信源大数据。公司数据采集站点覆盖 80 万余家新闻网站、3,000 余个资讯 APP、1,200 余家国内报刊、20 余个第三方平台、79 个语种和 50 万余个其他站点，7*24 小时不间断采集，并严格履行“三审三校”程序，数据来源多样、权威、可靠，数据量规模大、更新快。

在此基础上，公司基于自然语言处理技术，对采集的内容进行低噪、去重、数据结构化、数据归一化、内容标签化、属性知识化、安全合规核查等精加工后，

再将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体系。目前，公司已经具备的数据资源情况如下：

应用领域	数据资源积累
媒体领域	超 1,000 亿条互联网媒体资讯数据、超 100 亿条权威信源媒体数据、超 10 亿条政府机构数据、14 类知识标引规则、超 12,000 个知识标引规则和 5 亿条行业资讯数据。
金融领域	拥有超 110 亿条金融主题数据、百亿级产业指标数据、超 30 亿条产业要素明细数据、超 2 亿条产业动态本体、超过 10,000 个知识标引规则和超 10 万个产业标签。
政务领域	自有 10 亿政务数据文件、政策文件、政务办事指南等数据。

大规模高质量行业数据训练是大模型实现落地应用的基础和前提。公司积累的海量高质量行业数据资产可迅速应用于拓天行业大模型的训练，为拓天行业大模型的开发建设、以及在媒体、政务、金融等领域的落地应用提供重要保障。

(5) 公司丰富的客户资源为本项目提供良好的市场基础

公司已深耕政务、媒体、金融等领域多年，为 80% 的中央和国务院机构，60% 的省级政府，50% 的地市政府；72% 的中央媒体，60% 的省级媒体，40% 的行业媒体以及 5 大国有银行、3 大政策性银行、92% 的股份制商业银行等企业单位的数字化转型提供大数据、人工智能产品及相关服务。

本项目主要面向政务、媒体和金融等领域，开展行业大模型的研发，并在此基础上推出适用于各行业细分应用场景的 AIGC 应用。公司在政务、媒体和金融领域积累的丰富的客户资源为本项目的市场开拓提供良好的基础。

(6) 公司积累的 AI 技术、工程经验和 AIGC 应用实践有效保障本项目顺利实施

公司始终坚持核心技术自主研发，专注于人工智能“计算智能—感知智能—认知智能—创造智能”四个层级中的“认知智能”领域，拥有超过 20 年的 NLP 研发经验，在 NLP、知识图谱、OCR 图像视频结构化等领域掌握了自主可控的多模态内容处理底层技术，取得了拓尔思基于自然语言技术的句法分析软件 V1.0、TRS 分布式自然语言处理引擎 V1.0、TRS 基于深度学习的智能学习和分析系统

V1.0、拓尔思智能检索系统 V1.5、拓尔思麦文智创平台 V1.0、天行小样本人工智能系统 V1.0、知识图谱系统 V1.0 等丰富的知识产权成果，形成了一套完整的多模态人工智能技术与产品体系，能够为政务、媒体、金融领域广大用户提供文本、音视频、多模态等全栈 AI 服务。

经历多年业务发展，公司积累了丰富的 AI 工程经验。例如，在中国经济信息社（新华财经）区域产业分析项目中，公司利用 AI 智能技术，融合宏观、微观近百种异构数据，构建超过 8,000 个产业领域图谱，10 大产业类量化模型，上万个产业场景应用标签，赋能金融高质量营销获客与政府产业招商。在某金融监管部门项目中，公司构建了百亿级知识图谱，有效帮助该部门打击金融违法行为和执法专项调查。

同时，公司积极开展人工智能先进技术的布局，以数据智能应用为核心，积极拓展 AIGC 领域，在机器写作、对话式 AI、内容人机协同和自动报告生成等应用场景打造一批实践案例，并积极推出具有 AIGC 能力的虚拟人开放平台，与公司行业知识图谱相结合，赋予虚拟数字人知识储备、语义理解、推理分析、自主决策和交互表达的智慧能力。此外，公司积极尝试基于大模型的融合迭代，以进一步提升现有 AIGC 产品的内容生成质量，开启生成式大模型的创新应用。

综上，公司在人工智能领域所积累的扎实的技术能力、丰富的工程经验以及 AIGC 应用实践经验，均为本项目的顺利开展奠定良好基础。

（三）项目投资概算

本项目投资总额为 184,481.67 万元，具体投资情况如下：

单位：万元

序号	项目	投资金额	拟使用募集资金额
1	软硬件购置	84,068.00	84,068.00
2	租赁费	25,740.00	25,740.00
3	研发人员薪酬	51,028.80	51,028.80
4	数据资源购买	19,800.00	19,800.00
5	铺底流动资金	3,844.87	3,844.87

序号	项目	投资金额	拟使用募集资金金额
	合计	184,481.67	184,481.67

（四）经济效益分析

本项目内部投资收益率（税后）为 13.68%，税后投资回收期为 5.96 年（含建设期），项目具有良好的经济效益，对公司发展有较好的促进作用。

（五）项目涉及的审批事项

根据《中华人民共和国环境影响评价法》《建设项目环境影响评价分类管理名录》等相关法律法规的规定，本项目不属于环保法规规定的建设项目，不需要进行项目环境影响评价，亦不需要取得主管环保部门对上述项目的审批文件。

截至本报告公告日，本项目涉及的备案事项正在办理中。

三、本次向特定对象发行对公司经营管理、财务状况的影响

（一）本次向特定对象发行对公司经营管理的影响

公司本次向特定对象发行股票募集资金用途符合国家产业政策和公司的发展战略。本次向特定对象发行后，将有助于提升公司的资金实力和资产规模，募集资金投资项目具有良好的市场前景，有利于增加公司的业务收入和提高长期盈利能力，进一步增强公司的核心竞争力。

（二）本次向特定对象发行对公司财务状况的影响

本次发行完成后，公司的资金实力将得到有效提升，公司总资产和净资产规模将有所增加，资产负债率将有所下降，营运资金将得到进一步充实。同时，公司的财务结构将更加稳健合理，经营抗风险能力将进一步加强。

由于本次向特定对象发行募集资金投资项目从建设投入到产生经济效益需要一定时间，净利润短期内难以与净资产保持同步增长，公司的每股收益和净资产收益率预计在短期内将出现一定程度的下降。但从长远来看，随着募集资金投资项目预期效益的实现，公司的盈利能力将会进一步增强。

四、可行性分析结论

综上所述，公司本次向特定对象发行股票募集资金投向符合国家产业政策及行业发展趋势，与目前上市公司的主营业务紧密相关，符合公司未来发展的战略规划。本次向特定对象发行股票募集资金投资项目拟开展拓天行业大模型的构建和训练，并在此基础上实现媒体、金融和政府领域的 AIGC 应用产业化落地，具有良好的市场前景和经济效益，将进一步提升公司的核心竞争力和服务能力，巩固公司市场地位，增强公司的综合实力，符合公司及全体股东的利益。因此，本次向特定对象发行股票募集资金使用是必要的、可行的。

拓尔思信息技术股份有限公司董事会

2023 年 8 月 7 日