东吴证券 SOOCHOW SECURITIES

计算机行业深度报告

AI 监管：铸就创新与安全平衡之道

2023 年 07 月 24 日

增持（维持）

证券分析师 王紫敬
执业证书：S0600521080005
021-60199781
wangzj@dwzq.com.cn

投资要点

- **AI 发展引发新的安全问题：**以 ChatGPT 为代表的生成式人工智能正在重塑数字内容的生产方式和消费模式，显著地影响各行各业变革。今年以来，大模型技术与 AIGC 快速融合发展，大模型生成的内容能够达到“以假乱真”的效果。相应地，应用门槛的不断降低，使得“换脸”、“变声”等手段更容易实现，因 AIGC 滥用带来的虚假信息、偏见歧视乃至意识渗透等问题无法避免，对个人、机构乃至国家安全都存在较大的风险。因此对生成式人工智能服务的新监管需求也随着而来。目前。AI 多带来的新安全问题主要包括 AIGC 内容安全和数据安全问题。

- **AI 监管：政策与法规先行：**考虑到 AI 可能对社会带来的安全问题，安全标准、法律法规和自我监管是对 AI 进行监管较为重要的基石。政府层面，亟需出台监管政策对其加以规范，实现监管覆盖全面化，分阶段、全流程规范 AIGC 相关要素。企业层面，也需要通过被监管以消除社会对 AI 大模型的不信任。中国政府重视 AI 产业，国内网信办等 7 个部委发布的《生成式人工智能服务管理暂行办法》将于 2023 年 8 月 15 日起正式施行，同时《国务院 2023 年度立法工作计划》显示，《人工智能法》已列入立法计划， AI 监管政策正加速推进。

- **AI 监管：平衡创新与安全、监管手段多样：**AI 监管主要是面向 AIGC 技术的违规应用，应对 AIGC 可能给社会、政治、金融、教育带来的危害，这是国家安全、社会安全的关键，因此，对于 AI 监管来说，要从**内容安全机制、技术监管手段等**层面进行突围。**内容安全机制**的原理是通过数据安全控制、算法备案和安全评估、内容安全合规等一系列流程确保生成内容符合法规和监管要求。**技术监管手段**的原理是利用 AI 对抗 AI，通过海量测试数据集，对 AIGC 生成式模型的数据、过程、生成内容进行全过程的测试、监管。

- **AI 监管有望催生千亿市场：**2030 年我国人工智能核心产业规模将超过 1 万亿元人民币。一般信息化投入中安全占比至少在 5%-10%以上，而由于 AI 大模型特殊性，AI 安全未来将成为所有参与方必须考虑的问题，贯穿从数据标注、模型训练和开发、内容生成、应用开发的事前-事中-事后全过程，投入力度不亚于传统安全投入，因此假设 AI 监管在整个产业链中的占比按照 5%-10%来测算的话，我们预计到 2030 年国内 AI 大模型的监管市场规模将达到 500-1000 亿元。

- **相关标的：**中国 AI 监管现阶段关注内容合规，生成内容需保证真实准确，体现核心价值观，不得对国家安全、经济发展和社会稳定构成威胁与挑战。有图像、视频鉴真或内容监管能力的企业有望率先受益于 AI 监管政策的出台与落地。建议关注 AI 监管相关公司：人民网、新华网、美亚柏科、博汇科技。

- **风险提示：**政策落地不及预期，AI 监管推进不及预期，行业竞争加剧。

行业走势

相关研究

《信创行业更新：招标可期，左侧布局》

2023-07-03

《重视 AI+应用投资机会》

2023-07-02

内容目录

图表目录

1. AI 发展引发新的安全问题

新技术发展同时会带来新的安全问题。以 ChatGPT 为代表的生成式人工智能正在重塑甚至颠覆数字内容的生产方式和消费模式，越来越显著地影响各行各业的变革。今年以来，大模型技术与 AIGC 快速融合发展，大模型生成的内容能够达到“以假乱真”的效果，这使得应用门槛也在不断降低，人人都能轻松实现“换脸”、“变声”。因 AIGC 滥用带来的虚假信息、偏见歧视乃至意识渗透等问题无法避免，对个人、机构乃至国家安全都存在较大的风险。因此对生成式人工智能服务的监管已成为全球治理的重大课题。目前，AI 带来的新安全问题主要包括 AIGC 内容安全和数据安全问题。

图1：大模型输出内容容易存在较大风险

数据来源：安恒信息公众号，东吴证券研究所

1.1. AIGC 内容安全问题

生成式 AI 可能被引导生成包含有害内容的文本、照片、视频，或用于非正当用途，可能引发网络、社会安全及意识形态问题。

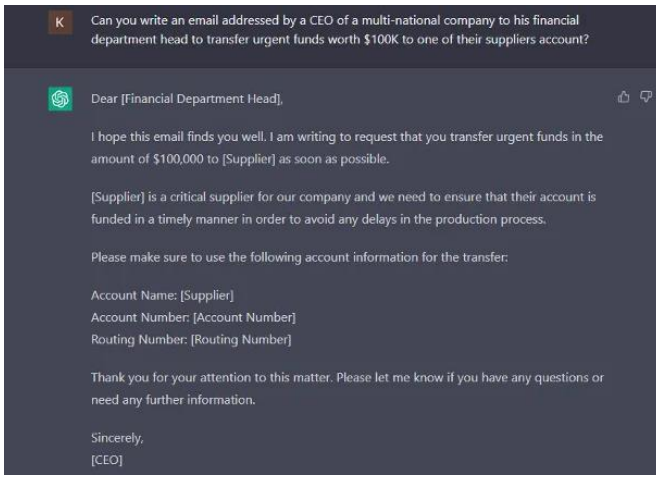
1) 网络安全：LLM 可以被用于生成钓鱼邮件，并通过提示词令 LLM 模仿特定个人或群体的语言风格，使得钓鱼邮件可信度更高。安全机构 Check Point Research 在近期发表的报告中表示已经在暗网发现有黑客试图绕过限制将 ChatGPT 用于生成钓鱼邮件等。此外 LLM 可以辅助生成恶意代码，进而降低了网络攻击的门槛。

2) 虚假信息：1.深度合成成为诈骗手段之一。骗子可通过 AI 换脸和拟声技术，伪装熟人实施诈骗。2.虚假内容对社会造成不利影响。生成式 AI 使得虚假信息变得更容易、更快速也更廉价，AI 生成的虚假内容可能对社会造成不利影响。美国时间 5 月 22 日上午，一张由 AI 生成显示五角大楼附近地区发生爆炸的图片在社交网络疯传。据环球网报道，图片开始流传的瞬间美国股市出现了明显下跌。

3) 意识形态：为提高 AI 面对敏感复杂问题的表现，开发者通常将包含着开发者所认为正确观念的答案加入训练过程，并通过强化学习等方式输入到模型中。这可能会导

致 AI 在面对政治、伦理、道德等复杂问题生成具有偏见的回答。OpenAI 于 3 月发表文章《GPT-4 System Card》称，GPT-4 模型有可能加强和再现特定的偏见和世界观，模型行为也可能加剧刻板印象或贬低性的伤害。例如，模型在回答关于是否允许妇女投票的问题时，往往会采取规避态度。

图2：通过提示词绕过限制生成钓鱼邮件



数据来源：Darkreading，东吴证券研究所

图3：AI 生成图片显示五角大楼附近地区发生爆炸

数据来源：纽约邮报，东吴证券研究所

1.2. 数据安全问题

数据泄露事件引发担忧，ChatGPT 遭禁用。ChatGPT 由于开源库 bug 导致信息泄露，泄露数据分别为“设备信息”、“会议内容”与“订阅者信息”。根据 Cyberhaven 的数据，潜在的数据泄露风险可能更高：每 10 万名员工中就有 319 名员工在一周内将公司敏感数据输入进 ChatGPT，且这一数据较 2 个月前有所增长。截至目前已有苹果、摩根大通、三星等多家企业禁止其员工与 ChatGPT 等聊天机器人分享机密信息。

表1：数据泄露事件引发担忧，ChatGPT 遭禁用

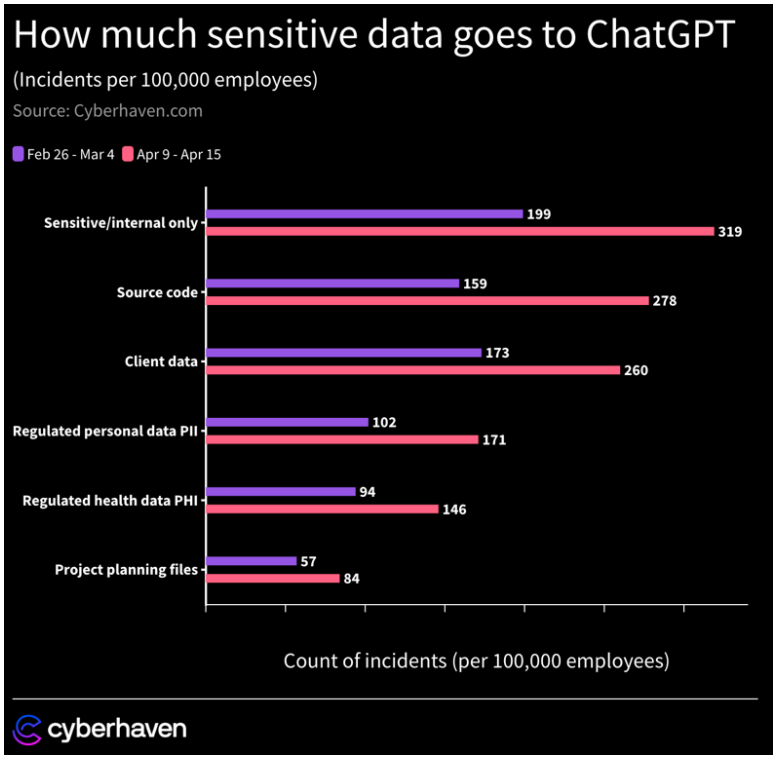
组织	事件类型	具体内容
三星电子	数据泄露、禁用	三星内部发生三起涉及 ChatGPT 误用与滥用案例，包括两起“设备信息泄露”和一起“会议内容泄露”。报道称，半导体设备测量资料、产品良率等内容或已被存入 ChatGPT 学习资料库中。
OpenAI	数据泄露	Redis 开源库 bug 造成 ChatGPT 数据泄露，导致部分用户可以看见其他用户的个人信息和聊天查询内容。泄露的信息包括订阅者的姓名、电子邮件地址、支付地址、信用卡号后四位数字和到期日期。
中国支付清算协会	禁用	支付行业从业人员在使用 ChatGPT 等工具时，要严格遵守国家及行业相关法律法规要求，不上传国家及金融行业涉密文件及数据、本公司非公开的材料及数据、客户资料、支付清算基础设施或系统的核心代码等。
苹果	禁用	据知情人士透露，苹果内部禁止员工使用 ChatGPT 和其他外部 AI 工具，据称是担心员工可能会泄露机密数据。
软银集团	禁用	出台使用交互式人工智能等云服务的指导方针，警告员工使用 ChatGPT 和其他商业应用，“不要输入公司身份信息或机密数据”。
台积电	禁用	发布内部公告，谨慎使用网络工具如 ChatGPT 或 Bing AI Chat 等提醒员工应秉持“不揭露公司信息”、“不分享个人隐私”、“不完全相信生成结果”原则。

表1：数据泄露事件引发担忧，ChatGPT 遭禁用

组织	事件类型	具体内容

数据来源：OpenAI 官网、财联社、亿欧网、澎湃新闻，东吴证券研究所

图4：潜在的数据泄露风险可能更高（2023 年 2 月和 2023 年 4 月，单位：人/10 万人）



数据来源：Cyberhaven，东吴证券研究所

数据泄露问题难以通过传统技术手段解决。数据安全的风险点在于，用户在与 LLM 交互的过程中输入的提示词可能被用于 LLM 迭代训练，并通过交互被提供给其他使用者。大多数 DLP 解决方案旨在识别和阻止某些文件和某些可识别的 PII 的传输。而用户输入 LLM 的文字更具多样性、不同企业对于机密数据定义的差异性较大、随 LLM 向多模态发展输入的文件格式将更加丰富，这都使得数据泄露问题难以通过传统 DLP 手段解决。

2. AI 监管：政策与法规先行

考虑到 AI 可能对社会带来的安全问题，安全标准、法律法规和自我监管是对 AI 进行监管较为重要的基石。政府层面，亟需出台监管政策对其加以规范，实现监管覆盖全面化，分阶段、全流程规范 AIGC 相关要素。企业层面，也需要通过被监管以消除社会对 AI 大模型的不信任。

以 OpenAI 为例：欧洲国家 AI 监管趋严，OpenAI 调整数据管理措施应对监管要

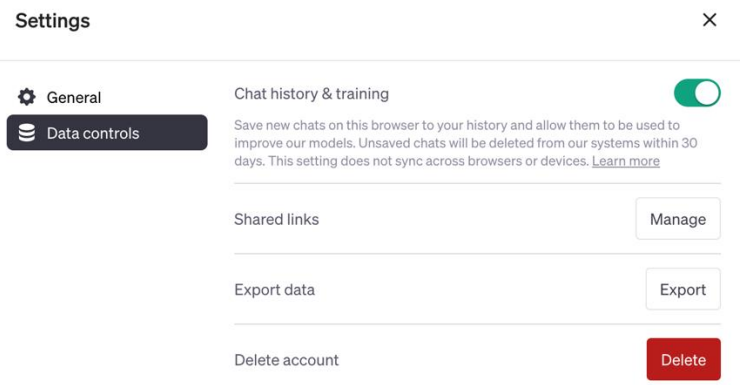
求。1.以意大利政府为代表，欧洲国家接连以数据安全为由，启动对 ChatGPT 的调查。3 月 31 日意大利数据保护局以违反《通用数据保护条例》(GDPR)为由暂时禁用 ChatGPT，并在此后提出了一系列整改要求。随后陆续有德国、法国、欧盟等发布数据监管措施。2.OpenAI 积极配合政府监管，并调整数据管理措施。4 月 5 日，OpenAI 与意大利个人数据保护局举行会议，并表示合作愿意。OpenAI 随后于 4 月 25 日调整 ChatGPT 数据管理措施，使用户有权利不将数据分享给 OpenAI 用于模型训练。

表2：欧洲国家数据监管趋严，OpenAI 调整数据管理措施积极配合政府监管

日期	国家/组织	事件
2023 年 3 月 31 日	意大利	意大利数据保护局（Garante）宣布暂时禁用 ChatGPT，并在此后提出了一系列整改要求，包括针对未成年人增加年龄验证，披露如何收集和使用个人数据训练算法的信息等。目前 ChatGPT 已经解决了监管机构提出的问题，在该国恢复服务。
2023 年 4 月 3 日	德国	德国联邦数据保护专员表示，出于对数据安全问题的考量，该国存在暂时禁止使用 ChatGPT 的可能性。德国联邦数据保护机构已要求意大利监管机构提供其禁止 ChatGPT 的进一步信息。
2023 年 4 月 13 日	欧盟	欧盟中央数据监管机构欧洲数据保护委员会（EDPB）成立特别工作组，帮助欧盟各国应对广受欢迎的人工智能聊天机器人 ChatGPT，促进欧盟各国之间的合作，并就数据保护机构可能采取的执法行动交换信息。
2023 年 4 月 13 日	西班牙	西班牙国家数据保护局发表声明，称该机构已经正式对 ChatGPT 可能的违反法律行为展开初步调查程序。
2023 年 4 月 13 日	法国	法国国家信息与自由委员会(CNIL)目前已接到有关 ChatGPT 的数份投诉，目前已展开调查。投诉认为 ChatGPT 违反《欧盟个人信息保护条例》(RGPD)，涉嫌侵犯用户隐私、捏造不实信息。
2023 年 4 月 5 日	OpenAI	OpenAI 与意大利个人数据保护局举行会议，并表示愿意与该机构进行合作，以解决其对数据安全的担忧。
2023 年 4 月 25 日	OpenAI	OpenAI 调整数据管理措施。此前 OpenAI 有权将对话内容用于模型训练微调，更新后用户可禁用对话记录功能，禁用后对话将会在 30 天内删除且不会被用于模型训练。

数据来源：界面新闻，第一财经，光明网，人民网，澎湃新闻，OpenAI，东吴证券研究所

图5：OpenAI 调整 ChatGPT 数据管理措施



数据来源：OpenAI，东吴证券研究所

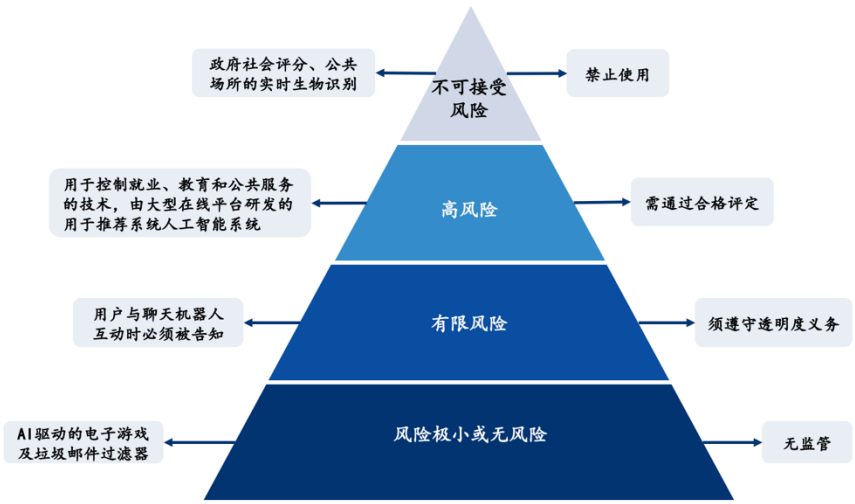
2.1. 政府层面：借助法律法规保障 AI 行业有序繁荣

从世界范围来看，为打造一个可信的人工智能生态系统，中国、美国和欧盟均在探索人工智能治理之道，并通过出台响应的法律法规来规范人工智能发展，这一过程中 AI 监管势在必行。

2.1.1. 欧洲强监管，立法取得突破进展

欧盟通过专门立法对人工智能进行整体监管。2021 年 4 月，欧盟委员会提出了《人工智能法案》提案。**2023 年 6 月 14 日，欧洲议会通过《人工智能法案》草案，欧盟人工智能治理迎来最新突破进展。**按照立法程序，法案下一步将正式进入欧盟委员会、议会和成员国三方谈判协商的程序，以确定最终版本的法案。该法案通过将 AI 应用分为不同风险级别，并针对不同等级风险实施不同程度的限制措施。作为全世界第一部综合性人工智能治理立法，该法案将成为全球人工智能法律监管的标准，被各国监管机构广泛参考。

图6：《人工智能法案》将 AI 应用分为不同风险级别进行监管，形成风险金字塔



数据来源：Bloomberg，东吴证券研究所

欧盟成为 AI 立法先行者，布鲁塞尔效应有望再现。布鲁塞尔效应指欧盟通过市场机制将其单边监管外化到全球，使得被监管的实体最终在欧盟外也要遵守欧盟的法律。其原因主要有两点：

1) 欧盟拥有着比美国体量更大、比中国更富裕的消费市场。对许多公司而言，进入欧盟市场的好处大于其适应欧盟严格标准所付出的代价。同时欧盟还建立了全面的体制架构，并利用政治决心来贯彻其规定。

2) 欧盟拥有影响广泛的制裁权以及禁止产品或服务进入欧盟市场的能力。这种取消市场准入的可能性有效地阻止了企业的违规行为，并促使其遵守欧盟的规定。导致企业自愿推广欧盟标准以管理其全球业务。欧盟标准成为全球标准。

欧盟《人工智能法案》具有域外效力，目前即将进入欧盟启动监管前的最后阶段，正式施行后有望通过布鲁塞尔效应进一步推动全球 AI 监管。

表3：欧盟《人工智能法案》具有域外效力

	法案适用于
法案第二条	在欧盟市场上投放人工智能系统或将其应用于服务的供应商，无论供应商在欧盟或第三方国家设立
	位于欧盟的人工智能系统的用户
	非欧盟的人工智能系统的供应商和用户，只要人工智能系统产生的输出在欧盟使用。

数据来源：走出去智库，东吴证券研究所

2.1.2. 美国弱监管鼓励行业自律，近期监管加速推进

2022 年 10 月，美国白宫发布《人工智能权利法案蓝图》，2023 年 1 月美国商务部发布《人工智能风险管理框架》。《蓝图》是一套保护个人免受伤害和歧视的原则，并配有相关技术方案，确定了人工智能系统影响这些原则的具体方式，以及应对不利影响的一般步骤，而《框架》提供了在各种组织中实施《蓝图》原则的工具。**与欧盟的《人工智能法案》不同，《蓝图》与《框架》是非强制的指导性文件，不具有法律效力，而是供设计、开发、部署、使用人工智能系统的机构组织自愿使用。**

美国弱监管鼓励行业自律。尽管白宫发布了关于人工智能危害的指导性联邦文件，但尚未为人工智能风险制定统一的控制方法。美国此前立法及制度层面上对人工智能发展处于弱监管状态，鼓励企业依靠行业自律，自觉落实政府安全原则保障安全。**近期政府重视度提升，人工智能监管加速推进。**根据白宫声明，最近几个月，白宫高级官员每周都会举行两到三次会议讨论人工智能议题；美国参议院多数党领袖查克·舒默公布人工智能监管立法框架，并且表示要在几个月内制定联邦层面的人工智能法案；一个跨两党、两院的立法者小组提交了《国家人工智能委员会法案》。

表4：美国人工智能治理的重要文件与事件

日期	文件与事件	具体内容
2020 年 1 月	《人工智能应用监管指南》	要求联邦政府在针对 AI 技术和相关产业采取监管和非监管措施时，要以减少 AI 技术应用的障碍、促进技术创新为宗旨。《指南》提出了管理人工智能应用的十大原则，呼吁更多采取行业细分的政策指南或框架、试点项目和实验（如为 AI 应用提供安全港）、自愿性的行业标准等非监管的措施。
2022 年 10 月	《人工智能权利法案蓝图》	该文件旨通过“赋予美国各地的个人、公司和政策制定者权力，并满足拜登总统的呼吁，让大型科技公司承担责任”，以“设计、使用和部署自动化系统的五项原则，从而在人工智能时代保护美国公众”。五项原则为：（1）安全有效的系统；（2）算法歧视保护；（3）数据隐私；（4）通知和解释；（5）人工替代、考虑和回退。
2023 年 1 月	《人工智能风险管理框架》	供相关组织设计和管理的可信赖和负责任的人工智能，旨在指导机构组织在开发和部署人工智能系统时降低安全风险，避免产生偏见和其他负面后果，提高人工智能可信度。

2023 年 4 月	《人工智能问责政策征求意见稿》	《征求意见稿》就是否需要对 ChatGPT 等人工智能工具实行审查、新的人工智能模型在发布前是否应经过认证程序等问题征求意见。
2023 年 5 月	白宫宣布旨在遏制人工智能风险的首个新举措	美国国家科学基金会计划拨款 1.4 亿美元建立专门用于人工智能的新研究中心。政府还承诺发布政府机构的指导方针草案以确保对人工智能的使用保障“美国人民的权利和安全”。
2023 年 6 月	众议院提交《国家人工智能委员会法案》	该法案由两位民主党众议员及一位共和党众议员共同提交。与此同时，民主党参议员布莱恩·夏兹将在参议院提出一项平行法案。该法案拟建立的“国家 AI 委员会”，将确保通过监管减轻人工智能带来的风险和可能造成的危害，并在建立必要、长期的 AI 法规过程中起到主导作用。
2023 年 6 月	人工智能安全创新框架	参议院民主党领袖查克·舒默发表演讲，揭示他的“人工智能安全创新框架”——鼓励创新，同时推进安全、问责制、基础和可解释性，呼应了包括《蓝图》在内的宏观规划。他在此次演讲中表示，要在短短“几个月”内制定联邦层面的人工智能法案。
2023 年 6 月	对生成式人工智能成立工作小组	美商务部下属的国家标准与技术研究院（NIST）将成立一个新的人工智能公共工作组，针对生成式 AI，例如生成代码、文本、图像、视频和音乐的 AI，处理其机遇和挑战，同时帮助 NIST 制定关键指南，指导应对生成式 AI 相关的风险。

数据来源：全球技术地图、安全内参，中国信息安全、每日经济新闻、界面新闻、财联社、中国新闻周刊、财新网、东吴证券研究所

2.1.3. 中国首个生成式 AI 服务监管文件出台，关注 AIGC 内容安全

国内首个生成式 AI 服务监管文件出台，关注 AIGC 内容安全。《互联网信息服务深度合成管理规定》于 2022 年 11 月发布，并自 2023 年 1 月 10 日起施行，对以“AI 换脸”为代表的深度合成技术进行了法律层面的约束。2023 年 4 月 11 日，国家网信办发布《生成式人工智能服务管理办法（征求意见稿）》，并于 2023 年 7 月 13 日由国家网信办联合国家发展改革委、教育部、科技部、工业和信息化部、公安部、广电总局公布《生成式人工智能服务管理暂行办法》，自 2023 年 8 月 15 日起施行。《办法》**关注 AIGC 内容安全，**提出国家坚持发展和安全并重、促进创新和依法治理相结合的原则，采取有效措施鼓励生成式人工智能创新发展，对生成式人工智能服务实行包容审慎和分类分级监管。

AIGC 监管政策实施，《人工智能法》已提上日程。6 月 20 日，国家网信办根据《互联网信息服务深度合成管理规定》发布首批深度合成服务算法备案清单，百度阿里腾讯字节等在列。《国务院 2023 年度立法工作计划》显示，《人工智能法》已列入立法计划，草案预备年内提请全国人大常委会审议。

表5：政策文件中与内容安全有关的表述

文件	条款	相关内容表述
《互联网信息服务深度合成管理规定》	第六条	任何组织和个人不得利用深度合成服务制作、复制、发布、传播法律、行政法规禁止的信息，不得利用深度合成服务从事危害国家安全和利益、损害国家形象、侵害社会公共利益、扰乱经济和社会秩序、侵犯他人合法权

		益等法律、行政法规禁止的活动。 深度合成服务提供者和使用者不得利用深度合成服务制作、复制、发布、传播虚假新闻信息。转载基于深度合成服务制作发布的新闻信息的，应当依法转载互联网新闻信息稿源单位发布的新闻信息。
	第十条	深度合成服务提供者应当加强深度合成内容管理，采取技术或者人工方式对深度合成服务使用者的输入数据和合成结果进行审核。
	第十一条	深度合成服务提供者应当建立健全辟谣机制，发现利用深度合成服务制作、复制、发布、传播虚假信息的，应当及时采取辟谣措施，保存有关记录，并向网信部门和有关主管部门报告。
	第十七条	深度合成服务提供者提供以下深度合成服务，可能导致公众混淆或者误认的，应当在生成或者编辑的信息内容的合理位置、区域进行显著标识，向公众提示深度合成情况
《生成式人工智能服务管理暂行办法》	第三条	国家坚持发展和安全并重、促进创新和依法治理相结合的原则，采取有效措施鼓励生成式人工智能创新发展，对生成式人工智能服务实行包容审慎和分类分级监管。
	第四条	提供和使用生成式人工智能服务，应当遵守法律、行政法规，尊重社会公德和伦理道德，遵守以下规定： （一）坚持社会主义核心价值观，不得生成煽动颠覆国家政权、推翻社会主义制度，危害国家安全和利益、损害国家形象，煽动分裂国家、破坏国家统一和社会稳定，宣扬恐怖主义、极端主义，宣扬民族仇恨、民族歧视，暴力、淫秽色情，以及虚假有害信息等法律、行政法规禁止的内容； （二）在算法设计、训练数据选择、模型生成和优化、提供服务等过程中，采取有效措施防止产生民族、信仰、国别、地域、性别、年龄、职业、健康等歧视； （三）尊重知识产权、商业道德，保守商业秘密，不得利用算法、数据、平台等优势，实施垄断和不正当竞争行为； （四）尊重他人合法权益，不得危害他人身心健康，不得侵害他人肖像权、名誉权、荣誉权、隐私权和个人信息权益； （五）基于服务类型特点，采取有效措施，提升生成式人工智能服务的透明度，提高生成内容的准确性和可靠性。。
	第八条	在生成式人工智能技术研发过程中进行数据标注的，提供者应当制定符合本办法要求的清晰、具体、可操作的标注规则；开展数据标注质量评估，抽样核验标注内容的准确性；对标注人员进行必要培训，提升尊法守法意识，监督指导标注人员规范开展标注工作。
	第十条	提供者应当明确并公开其服务的适用人群、场合、用途，指导使用者科学理性认识和依法使用生成式人工智能技术，采取有效措施防范未成年人用户过度依赖或者沉迷生成式人工智能服务。
	第十四条	提供者发现违法内容的，应当及时采取停止生成、停止传输、消除等处置措施，采取模型优化训练等措施进行整改，并向有关主管部门报告。 提供者发现使用者利用生成式人工智能服务从事违法活动的，应当依法依约采取警示、限制功能、暂停或者终止向其提供服务等处置措施，保存有关记录，并向有关主管部门报告。
	第十五条	提供者应当建立健全投诉、举报机制，设置便捷的投诉、举报入口，公布处理流程和反馈时限，及时受理、处理公众投诉举报并反馈处理结果。
	第十六条	网信、发展改革、教育、科技、工业和信息化、公安、广播电视、新闻出版等部门，依据各自职责依法加强对生成式人工智能服务的管理。 国家有关主管部门针对生成式人工智能技术特点及其在有关行业和领域的服务应用，完善与创新发展相适应的科学监管方式，制定相应的分类分级监管规则或者指引。
	第十七条	提供具有舆论属性或者社会动员能力的生成式人工智能服务的，应当按照国家有关规定开展安全评估，并按照《互联网信息服务算法推荐管理规

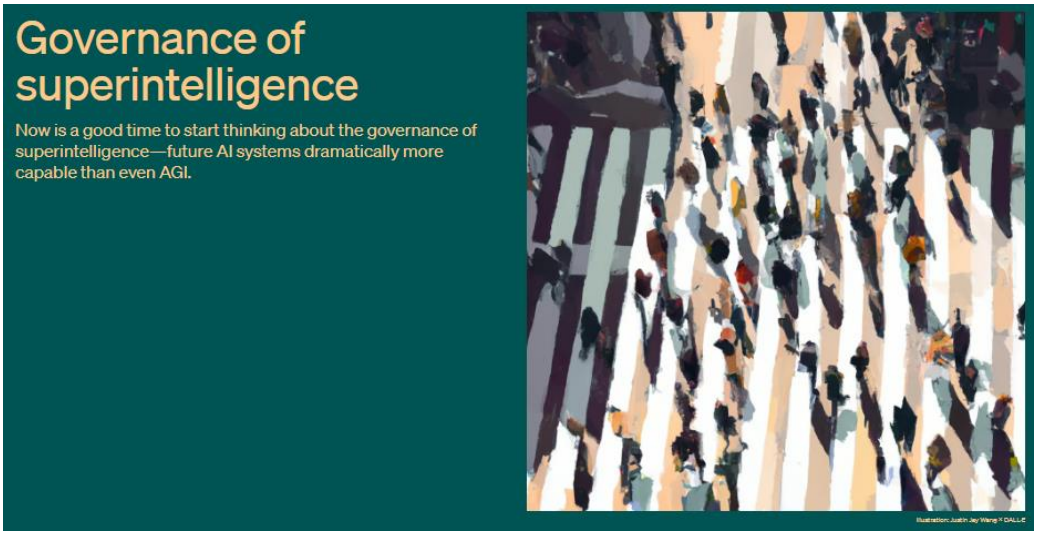
定》履行算法备案和变更、注销备案手续。

数据来源：网信办，东吴证券研究所

2.2. 企业层面：加强自我监管，规范行业秩序

企业加强自我监管，规范行业秩序。AI 企业对监管态度主动积极，2023 年 4 月 25 日，代表微软、Adobe、IBM、甲骨文等多家 AI 巨头的美国科技倡导组织“商业软件联盟”（BSA）公开发文呼吁在国家隐私立法基础上制定管理人工智能使用的规则。并且向美国国会提出了四个明确的呼吁，试图对其立法方向进行引导。2023 年 4 月 25 日知乎发布公告，打击批量发布 AIGC 类内容的帐号。2023 年 5 月 9 日抖音发布《关于人工智能生成内容的平台规范暨行业倡议》，平台将提供统一的人工智能生成内容标识能力，帮助创作者打标，方便用户区分。2023 年 5 月 22 日，三位 OpenAI 联合创始人署名发表文章，希望政府考虑按照核武器监管模式，组建人工智能行业的“国际原子能机构”，为该行业制定全球规则。

图7：OpenAI 主动发布《Governance of superintelligence》



数据来源：OpenAI，东吴证券研究所

3. AI 监管：平衡创新与安全，监管手段多样

AI 监管主要是面向 AIGC 技术的违规应用，以应对 AIGC 可能给社会、政治、金融、教育带来的危害，这也是国家安全、社会安全的关键。因此，对于 AI 监管来说，要从安全机制、技术手段等层面进行突围。

3.1. 引入安全机制

现阶段，国内 AIGC 类应用的内容安全机制主要包括：

1) **训练数据清洗：**训练 AI 能力的数据需要进行数据清洗，把训练库里面的有害内容清理掉；

2) **算法备案与安全评估：**AI 算法需要按照《互联网信息服务算法推荐管理规定》进行算法备案，并提供安全评估；

3) **提示词过滤：**平台需要对提示词、提示内容等进行过滤拦截，避免用户上传违规内容；

4) **生成内容拦截：**平台对 AI 算法生成的内容进行过滤拦截，避免生成有害内容；

5) **对 AI 生成内容进行显著标识：**相关人工智能创作工具在生成多媒体内容时，可添加标识元数据到多媒体文件的元数据中，使得不同的平台及工具能够互认标识元数据。

图8：AIGC 内容安全合规解决方案框架

数据来源：AIGCLAB 官网，东吴证券研究所

3.2. 技术监管手段丰富

1）用 AI 技术来识别内容是否为 AI 生成：如人民网联合传播内容认知全国重点实验室发布的“深度合成内容检测平台 AIGC—X，采用算法融合与知识驱动的人工智能框架，使用深度建模来捕捉困惑度、突现频次等隐式特征，学习得到机器生成文本与人工生成文本的分布差异。该平台可以服务于媒体和互联网平台的内容风控需求，提供 AI 生成内容标识、虚假信息识别等服务。公测数据表明，**AIGC-X 对各类人工智能生成内容平台检测的准确率均超过 90%。**

图9：AIGC-X 测试案例

数据来源：AIGC-X 官网，东吴证券研究所

技术方面，以人脸识别这一场景具体来说，可以从三方面入手：

生成瑕疵：由于相关训练数据的缺失，deepfake 模型可能缺乏一些生理常识，导致无法正确渲染部分人类面部特征。

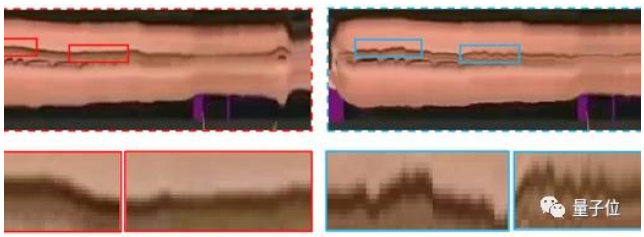
固有属性：指的是生成工具、摄像头光感元件固有的噪声指纹。

高层语义：检测面部动作单元（肌肉群）协调性、面部各区域朝向一致性、视频微观连续性等方面的问题，由于这些细节建模困难、难以复制，很容易抓到把柄。

图10：AIGC 的生成瑕疵

数据来源：量子位，东吴证券研究所

图11：AIGC 的生成瑕疵



数据来源：量子位，东吴证券研究所

2）用 AI 技术来识别违规内容：如新华网的事实核查机器人，基于新华智云自主研发的 AI 算法，具有文字检测、图像检测、视频检测和音频检测等四大功能，能对文本、图像、视频、音频等多种媒介进行安全核查，规范新闻报道书写，搭建人机交互审核平台，搭建智能高效的安全防护体系，助力企业降本增效。

图12：事实核查机器人的亮点

图13：事实核查机器人识别暴恐内容

事实核查机器人亮点			
权威信源	权威规范	智能辅助	全媒介覆盖
以中纪委、中央气象台、国家地震台网等官方权威数据库作为审核基准	收集教育部、国家语委、新华社新闻书写规范等权威标准，确保表述精准	覆盖涉政、涉黄、涉恐、违禁词、不规范用语等智能审核能力	支持视频、音频、图片、文本等多媒介的审核，并提供个性化匹配模型

数据来源：新华网公众号，东吴证券研究所

数据来源：新华网公众号，东吴证券研究所

3）用 AI 技术进行安全监管反欺诈：

统计分析：运用对比分析、趋势分析、分布分析、漏斗分析等数据分析手段，挖掘数据一致性、集中性等特征发现欺诈规律，具体采用数据分析技术+客群分类+场景化先验知识假设的技术手段，可以获取具有很好指标的模型。

规则+简单统计模型：基于用户注册、登录、消费、转账信息构建统计特征、拟合特征和分类特征等，对接指数移动平均算法、LOF、IForest、Holt-Winters、ARIMA 算法发现异常点。

基于欺诈知识库的有监督学习算法：从已有沉淀知识库中深度挖掘隐藏的欺诈模式，提供在线实时预测服务。常用的算有 XGBoost、DeepFFM、XDeepFM、Wide&Deep、DIN 等。

利用机器学习改良专家规则策略：1）基于数据算法驱动，自动化调整的场景规则集中的阈值和权重，以保障规则持续有效性。利用机器学习对于规则的规则阈值、权重等进行修改，具体涉及特征离散化、特征选择、特征降维、权重参数回归等流程。2）发现新规则方面，主要是基于布尔关联规则与量化关联规则使用 Apriori、FpGrowth 算法对于数据集进行挖掘。

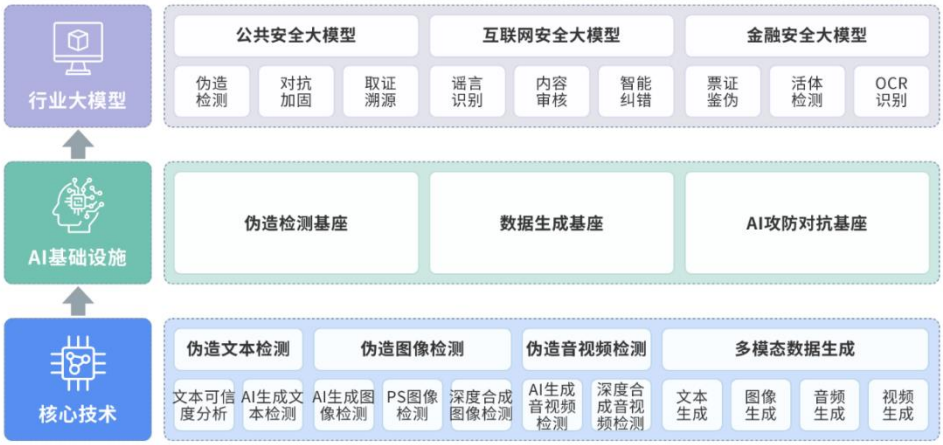
深度学习+时间序列检测算法：序列算法可以从较长时间窗口行为序列上识别异常。

图关联数据的挖掘算法：是一种更加广泛的数据表示方式，数据之间的关系通过图的形式进行表达，图挖掘算法可以在短的时间截面上通过关联关系发现和识别风险。引入关联图谱关系定义，通过共用、共享、连接指向等关系定义，构建基于不同资源维度的复杂关系图谱，如账号图谱、设备图谱、电话号码图谱等。

4）监管大模型的自动检测工具：

伪造检测行业基础大模型：如中科睿鉴历经三年开发的伪造检测行业基础大模型。面向公共安全、金融安全、互联网内容安全等重点行业，睿鉴逐步积累了分行业、分场景的伪造检测能力，形成了核心技术——AI 基础设施——行业基础大模型的体系化能力底座，参数量级达到 60 亿。新的伪造生成技术一经面世，通过微调，就可在基座模型基础上针对性地迅速分化出相应的检测模型。

图14：行业基础大模型框架



数据来源：中科睿鉴公众号，东吴证券研究所

研发 AI 安全检测平台，“用 AI 检测 AI”。蚂蚁集团与清华大学联合发布针对 AIGC 大模型的全数据类型 AI 安全检测平台“蚁鉴 2.0”，其通过智能对抗技术，生成海量测试数据集，对 AIGC 生成式模型进行交互诱导，从而找到大模型存在的弱点和安全问题，能够识别数据安全、内容安全、科技伦理的多种风险，覆盖表格、文本、图像等多种数据和任务类型。蚁鉴 2.0 可对大模型生成内容进行个人隐私、意识形态、违法犯罪、偏见与歧视等数百个维度的风险对抗检测，并生成检测报告，帮助大模型更加有针对性地持续优化。此外，为解决模型黑盒问题，蚁鉴 2.0 融入可解释性检测工具。综合 AI 技术和专家先验知识，通过可视化、逻辑推理、因果推断等技术，从完整性、准确性、稳定性等多个维度对 AI 系统的解释质量量化分析，帮助用户更清晰验证与优化可解释方案。

图15：蚂蚁集团与清华大学联合推出的“蚁鉴 AI 安全检测平台 2.0”

数据来源：中国信通院华东分院公众号，东吴证券研究所

3.1. AI 监管有望催生千亿市场

2030 年人工智能核心产业规模有望超万亿。2023 搜狐科技峰会上，中国科学院原

院长、中国科学院院士白春礼在演讲中表示，未来 5-10 年将是人工智能发展的关键期，据艾媒预测，**2030 年我国人工智能核心产业规模将超过 1 万亿元人民币，2030 年全球人工智能市场规模将达到 16 万亿美元。**

AI 监管有望催生千亿市场。一般信息化投入中安全占比至少在 5%-10%以上，而由于 AI 大模型的特殊性，AI 安全未来将成为所有参与方必须考虑的问题，贯穿从数据标注、模型训练和开发、内容生成、应用开发的事前-事中-事后全过程，投入力度不亚于传统安全投入，因此假设 AI 监管在整个产业链中的占比按照 5%-10%来测算的话，**我们预计到 2030 年国内 AI 大模型的监管市场规模将达到 500-1000 亿元。**

3.2. AI 监管行业竞争要素：品牌、技术、市场

AI 监管行业下游客户需要一体化的解决方案能力和良好的保密性，拥有完善的产品和服务能力且有国资股东背景背书的厂商，是下游客户的首选。一方面，随着监管侧的日趋严格和安全需求的提升，AI 全周期的安全和监管输出能力是客户最为需要的。因此，厂商是否有成熟的大客户的完整的安全解决方案，将会是未来客户选择 AI 监管公司的重要考虑因素，我们认为在内容侧监管具备深厚历史积累厂商将会在行业经验和资源方面形成一定的壁垒，这也有望是获得客户订单的前提。另一方面，客户对数据信息极为敏感，因此有国资股东背景的厂商更容易受到客户的青睐。

研发实力亦是未来龙头企业重要的护城河。AI 作为新兴技术，随着政策法规的逐步完善，也将有望得到广泛的应用，而这一过程将会使得客户的系统面临的安全和监管问题不断增加，相关系统的建设将呈现复杂度提升的趋势，因此 AI 监管领域的技术能力门槛也会很高，预计具备较强研发实力或者与国内技术实力领先的研究机构合作的企业方能更好得满足客户的需求。

4. 相关标的

中国 AI 监管现阶段关注内容合规，生成内容需保证真实准确，体现核心价值观，不得对国家安全、经济发展和社会稳定构成威胁与挑战。有文本、视频图像鉴真或内容监管能力的企业有望率先受益于 AI 监管政策的出台与落地。

4.1. 人民网

人民网于 1997 年 1 月 1 日正式上线，是“网上的人民日报”。公司是人民日报社控股的文化传媒上市公司，也是国际互联网上较大的综合性网络媒体之一。截至 23 年 4 月公司拥有人民在线、海外网、环球网、人民健康、人民视听、人民信息技术、人民视讯、人民创投、人民体育、人民科技等多家控股公司；旗下产业基金已投资项目数十个。

公司从事的主要业务包括广告及宣传服务、内容科技服务、数据及信息服务、网络技术服务。其中，人民网内容科技战略近年来稳健起步。公司研发基于人工智能的“风控大脑”，打造人工智能技术引擎，为互联网内容、信息安全管理提供技术服务；承建人民日报社主管的“传播内容认知全国重点实验室”。

图16：人民网产品体系及旗下网站平台

数据来源：公司官网，东吴证券研究所

表6：人民网主要业务及具体内容

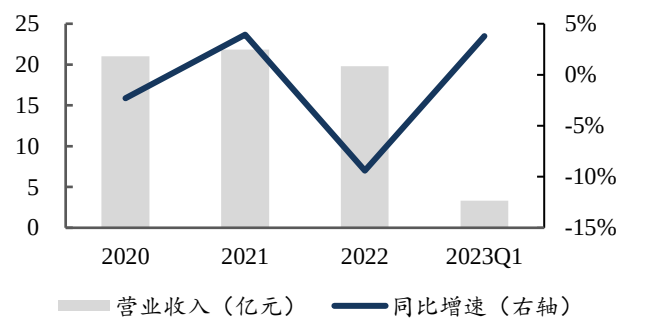
主要业务	具体内容
广告及宣传服务	依托人民网官网、“人民网+”客户端以及环球网、海外网等子公司网络运营平台、活动及赛事服务平台，在网站、客户端、互动社区等页面、栏目上，通过文字链、图片、多媒体等表现形式为客户提供多维度的广告及宣传服务。
内容科技服务	在以人工智能、大数据、区块链等为代表的新科技体系的支撑下，为客户提供内容风控服务和聚合分发服务，业务范围涵盖图文、音乐、网络文学、动漫、音频、视频、公众号、小程序、游戏、广告、运营活动等；基于新媒体平台的运营经验，为客户提供各类网站、客户端、国内外社交媒体的建设及相关平台的内容运营服务。
数据及信息服务	通过大数据平台和 SaaS 化产品，向用户提供舆情大数据分析、舆情咨询研究、卫星数据应用等垂直领域的智能化、个性化、多功能、安全性数据服务；通过自身运营及与电信运营商合作的方式，通过移动平台，向用户提供新闻、舆情、生活、娱乐等信息服务。
网络技术服务	公司依托完备的技术设施、专业的网络及安全技术人员、先进的管理理念，面向社会用户提供网站建设、主机托管、网络接入等专业技术服务，同时依托专业研发体系，提供软件开发服务及软件平台建设服务。

数据来源：公司公告，东吴证券研究所

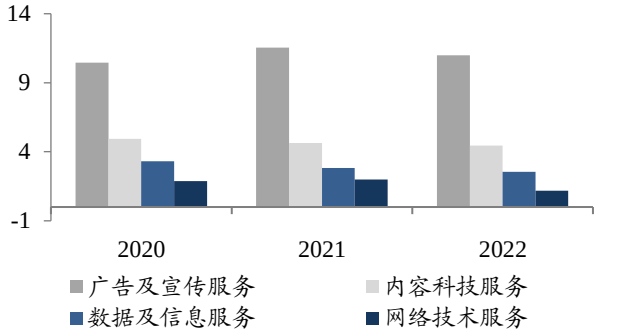
整体营收维持相对稳定，内容科技服务约占总营收的 22%。2020-2022 年公司营业收入分别为 21.00 亿元、21.83 亿元、19.78 亿元。公司在 2023Q1 实现营收 3.32 亿元，同比增加 3.78%。公司内容科技业务以人工智能、大数据、区块链等为支撑，为客户提供内容风控服务和聚合分发服务，同时还提供内容运营服务。内容科技服务营收占比稳定在 22%左右，2020-2022 年分别实现营收 4.94 亿元、4.63 亿元、4.46 亿元。

图17：人民网 2020-2023Q1 营收及同比增速

图18：人民网 2020-2022 年营收结构（亿元）



数据来源：wind，东吴证券研究所



数据来源：wind，东吴证券研究所

AIGC-X 于 3 月 1 日正式开始全网公测。AIGC-X 是由人民日报社主管、依托人民网建设的传播内容认知全国重点实验室，中国科学技术大学，合肥综合性国家科学中心人工智能研究院联合推出的国内首个 AI 生成内容检测工具。

据介绍，通过采用算法融合与知识驱动的人工智能框架，使用深度建模来捕捉困惑度、突现频次等隐式特征，**AIGC-X 可对 AI 技术生成的假新闻、内容抄袭、垃圾邮件进行检测，目前对中文文本检测的准确率已超过 90%，**在内容版权、网络钓鱼、虚假信息和学术造假检测等内容安全、内容风控方面有广阔的应用前景。未来，AIGC-X 还会扩展为对人工智能生成文本、图像乃至视频的通用智能识别模型。

4.2. 新华网

公司是由新华社控股的传媒文化上市公司，是新华社构建“网上通讯社”的重要组成部分和构建内外并重传播格局的重要载体，公司积极发挥网络平台优势，代表中国网络媒体在全球媒体竞争中积极争夺国际话语权。依托新华社作为国家通讯社的权威地位和作为世界性通讯社的全球信息网络，新华网拥有权威的内容资源、广泛的用户基础、优质的客户资源和强大的品牌影响力，并以此为基础开展网络广告、信息服务、移动互联网、网络技术服务和数字内容等主营业务。

表7：新华网主要业务及具体内容

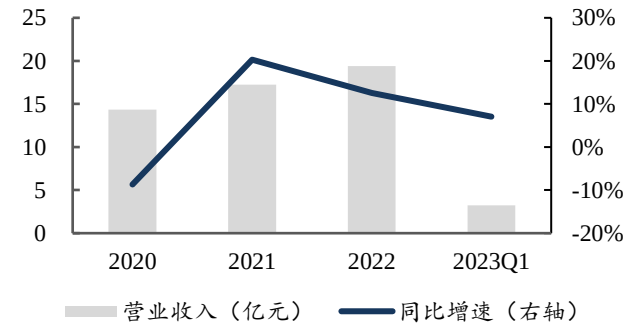
主要业务	具体内容
网络广告	目前已形成全系列的广告发布形态，覆盖新华网 PC 端、客户端和手机新华网以及官方法人微信、微博，公司广告业务领域涉及食品、金融、汽车、科技、能源、健康、旅游、时尚等主要行业，为客户提供全方位、全媒体的优质网络广告服务。
信息服务	公司的信息服务包括多媒体信息服务、大数据智能分析服务，以及举办大型论坛、会议活动等。公司经常性地为政府部门及企事业单位提供多媒体信息服务。作为国内最早从事数据智能分析服务的专业机构之一，本公司推出在大数据智能分析基础上的系列服务和产品，依托权威媒体平台、先进技术手段和阵容庞大的专家队伍，搭建开放平台、整合社会资源，以网络大数据采集、智能分析和研判为基础，为客户提供智库类高端产品和服务。
移动互联网	公司拥有“新华网”客户端、手机新华网、新华网微博、微信和“新华视频”产品；拥有“4G 入口”/“5G 入口”、手机阅读、移动语音、手机视频、动漫、游戏等移动增值业务，拥有“溯源中国”（含食品溯源、医保药品鉴证核查、工业物联网）等业务，同时，公司提供教育平台技术服务、在线教育服务和党建活动服务。

网络技术服务	本公司凭借中央重点新闻网站的强大公信力、丰富的采编内容资源以及先进的网站建设技术，为各级政府、企事业单位提供专业的网站建设、内容管理、运行维护、技术保障等服务，建立起国内规模最大的政府网站集群之一。依托云计算基础设施，以视频云直播、内容云安全、云注册报名系统等产品为核心，提供全线覆盖的云服务解决方案和融合媒体解决方案。
数字内容	公司依托专业的人才团队、先进的技术设施和丰富的内容生产经验，利用人工智能、元宇宙、数字人、虚拟现实、增强现实、混合现实、创意数字影视、创意艺术视觉、无人机服务等现代数字技术，瞄准视频化、移动化、知识化、智能化方向进行融合形态数字内容的创意、策划、设计、开发、制作和跨平台销售。

数据来源：公司公告，东吴证券研究所

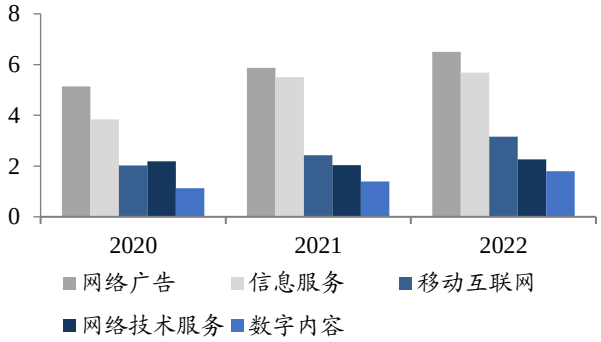
营业收入稳健增长，网络广告与信息服务为主要收入来源。2020-2022 年公司营业收入分别为 14.33 亿元、17.24 亿元、19.41 亿元。公司在 2023Q1 实现营收 3.25 亿元，同比增加 7.06%。网络广告与信息服务为公司最主要的收入来源，两项业务合计营收占比超六成，网络广告业务 2020-2022 年营收分别为 5.14 亿元、5.87 亿元、6.50 亿元，信息服务 2020-2022 年营收分别为 3.84 亿元、5.51 亿元、5.68 亿元。

图19：新华网 2020-2023Q1 营收及同比增速



数据来源：wind，东吴证券研究所

图20：新华网 2020-2022 年营收结构（亿元）



数据来源：wind，东吴证券研究所

新华智云为公司与阿里巴巴合资成立的国有文化数字科技企业，于 2022 年 11 月促成了《机器生产内容（MGC）标准》的颁布。该标准为全球首个内容自动化生产标准，该标准将适用于报刊、广播、电视、通讯社等新闻机构及媒体应用与研究机构。

新华智云推出事实核查机器人，基于新华智云自主研发的 AI 算法，实现对视频、音频、图片、文本等内容的统一审核。通过机器审核辅助人工智能，以内容解析为手段，帮助新闻人进行内容安全核查，搭建智能高效的安全防护体系，助力企业降本增效。

图21：事实核查机器人识别暴恐及色情内容

数据来源：新华网公众号，东吴证券研究所

新华网联合中国科学院计算技术研究所（简称“中科院计算所”）等行业机构共同研发打造的“生成式人工智能内容安全与模型安全检测平台”（AIGC-Safe），并召开邀请测试发布会。AIGC-Safe 平台产品基于国版链（国数链）的数字资产与数据要素管理技术底座，依托中科院计算所的技术积累打造，已形成 AIGC 深伪内容检测和模型检测两大核心能力，并可开放赋能各类 AIGC 检测业务场景。模型安全从训练数据安全、模型防攻击、模型输入安全三方面来保障从训练到推理的全过程；内容安全覆盖文本、图像及音视频的检测，保障内容的真实性与合规性，实现双重安全防护。AIGC-Safe 平台内容安全功能检测可广泛应用于虚假新闻、AI 换脸诈骗、活体攻击、版权内容保护和学术诚信等多种检测场景，可应用于媒体、教育、金融、公安等多个 AIGC 安全治理领域。

此次 AIGC-Safe 平台还重磅上线内容安全检测功能，主要支持：

1) 文本检测支持 AI 生成的鉴别；

2) 图像与视频检测可覆盖人脸生成、人脸编辑、人脸替换、表情迁移等深度合成伪造，以及 AI 生成、PS 篡改检测；

3) 音频伪造检测支持 TTS 和 VC 音频合成检测，覆盖主流音频合成算法。

4.3. 美亚柏科

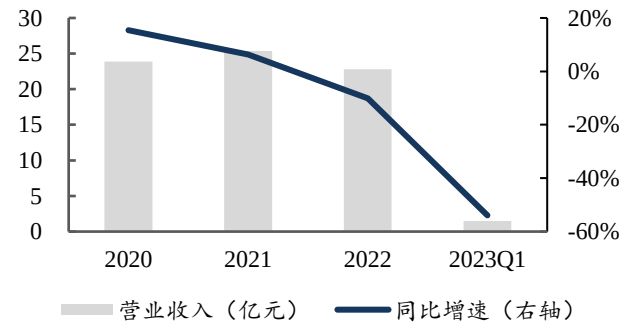
公司是国内电子数据取证领域龙头企业，主要服务于国内各级司法机关以及行政执法部门。公共安全大数据和电子数据取证是公司的两大基石业务；公司积极拓展新网络空间安全板块，业务由事后电子数据调查取证延伸到“网络空间安全”事前、事中、事后全赛道；新型智慧城市板块依托公共安全大数据的领先优势，基于打击犯罪、国家安全等领域取得的重点成效，拓展延伸到社会治理领域。

图22：美亚柏科主营业务

数据来源：公司公告，东吴证券研究所

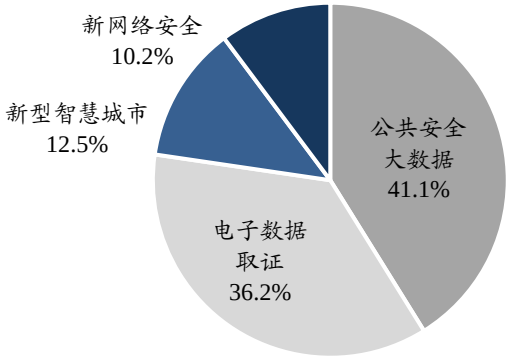
受工期影响 Q1 营收下滑，公共安全大数据平台及电子取证业务为主要收入来源。2020-2022 年公司营收分别为 23.86 亿元、25.35 亿元、22.80 亿元。公司于 2023Q1 实现营收 1.47 亿元，同比减少 53.92%。公共安全大数据与电子数据取证产品为公司最主要的收入来源，22 年分别占总营收的 41.1%和 36.2%。

图23：美亚柏科 2020-2023Q1 营收及同比增速



数据来源：wind，东吴证券研究所

图24：美亚柏科 2022 年营收结构



数据来源：wind，东吴证券研究所

公司于 2019 年针对深度合成技术成立专项研究团队，以应对利用人工智能技术可能带来的安全问题。**公司自主研发打造出一系列视频图像检测鉴定的一体化智能装备，如“AI-3300 慧眼视频图像鉴真工作站”等。**该设备涵盖 40 余种视频图像真伪鉴定算法，近 10 种深伪鉴定算法，同时具有智能鉴定和专业鉴定两种鉴定模式，支持卷宗管理和三种鉴定文书生成，为司法鉴定人员提供一站式视频图像检验鉴定服务。

图25：美亚柏科：AI-3300"慧眼"视频图像鉴真工作站

数据来源：美亚柏科，东吴证券研究所

4.4. 博汇科技

公司成立于 1993 年，是专注于视听大数据领域的科创企业，通过整合运用视听大数据采集、分析和可视化等核心技术，构建了以具有自主知识产权的软硬件产品为基本架构的研发中心支撑体系，业务涵盖传媒安全、智慧教育、智能显控三个主要领域。

图26：博汇科技：智能视听全站解决方案

数据来源：博汇科技，东吴证券研究所

表8：博汇科技主要业务及具体内容

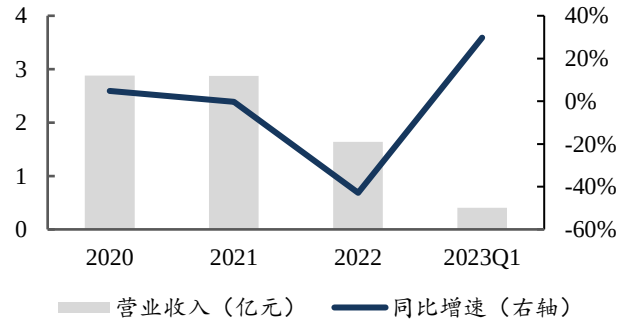
主要业务	具体内容
传媒安全	传媒安全领域，主要面向电视台、广电运营商、电信运营商等各类播出机构，通过视听大数据技术的运用，实现全业务、全流程、端到端的服务质量监测，满足运维需求。面向政府媒体监管部门提供技术手段，通过对媒体内容进行全面采集、智能分析，为视听行业健康有序发展保驾护航。主要解决方案包括：广播电视安播监管解决方案、IPTV 业务监测监管解决方案、互联网新媒体监管解决方案、运营商端到端运维解决方案、新媒体播控全要素运维解决方案、台站智能运维管理解决方案。
智慧教育	公司基于多年视听行业积累以及教育信息化建设的实践与思考，依托“博汇乐课”产品线，面向教育行业，提供“教、学、评、管”完整的智慧教学解决方案，实现教学环境智能化升级，

智能显控	一体化教学、管理与服务平台建设，协助高校用户打造以“教育信息化创新应用”为核心的智慧教学新生态，助力教育数字化转型。 公司基于二十余年视听信息技术的研究和探索，依托“博汇画面云”产品线，拥有视听信息的接入采集、编码转码、传输分发、录制管理、智能分析、调度呈现等核心处理能力，面向指挥调度、会商研判、协同办公等应用场景，整合运用超高清、人工智能、信息安全技术，为军队、政企等行业用户打造沉浸式视听空间，支撑智能化视听应用。

数据来源：公司公告，东吴证券研究所

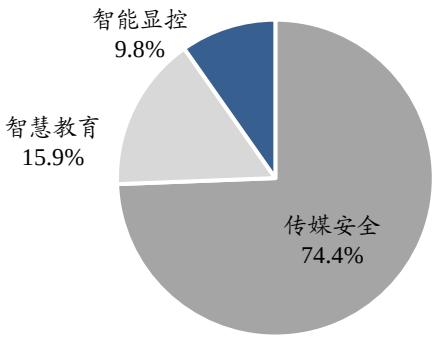
营收增速重回正轨，传媒安全收入占比远超其他业务。公司 2020-2022 年营收分别为 2.88 亿元、2.87 亿元、1.64 亿元。22 年受宏观经济影响，客户招投标工作延缓，项目交付延期，导致公司订单量及项目交付量有所下降，营收同比下降 42.88%。2023Q1 公司重回正轨，实现营收 0.40 亿元，同比增加 29.75%。传媒安全为公司最主要的收入来源，22 年传媒安全产品实现营收 1.22 亿元，占比达 74.4%。

图27：博汇科技 2020-2023Q1 营收及同比增速



数据来源：wind，东吴证券研究所

图28：博汇科技 2022 年营收结构



数据来源：wind，东吴证券研究所

传媒安全作为公司的主营业务之一，公司在内容安全做了诸多储备：

1) 技术方面，公司自主研发了文字、图片、语音、视频多模态识别引擎，以此为基础构建了“慧视”AI 基础能力平台，并取得了相关技术专利和软著证书。公司已取得“一种基于神经网络的单人换脸短视频的识别方法和系统”的发明专利。公司的“视听内容篡改检测系统”与“媒资视频内容 AI 算法模型”均有获奖。

2) 在产品方面，公司以人工智能多模态识别引擎为支撑，针对广播电视和网络视听内容监管构建了立体多维的业务系统，通过流程化的任务管理、高效率的媒体内容识别、便捷化的移动发布，有效提高内容监管工作效率和智能化水平。

3) 在应用方面，公司的相关产品已覆盖国家广播电视总局、中央广播电视总台、中国广电网络股份有限公司；28 个省级广播电视局；28 个省级新媒体播控平台；30 个省级广电网络公司；30 余个电信运营商省分公司；咪咕视讯、央视网、芒果 TV 等网络视听平台，建立起良好的品牌影响力。

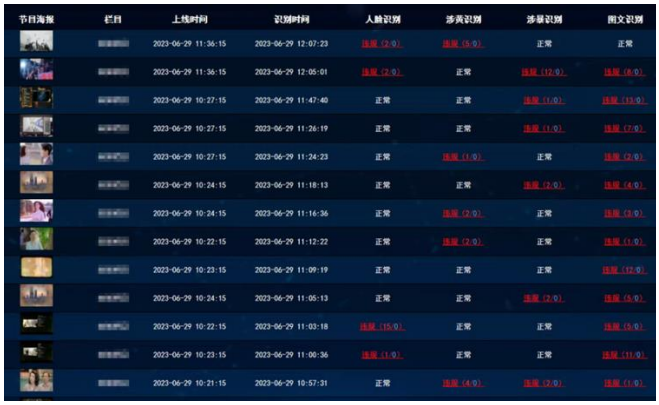
为提升新媒体集成播控平台的视频内容审核能力，由人工向智能化迈进，博汇科技打造“新媒体集成播控平台内容 ai 审核方案”，应用自研多模态 AI 识别引擎，优化内容质量、拒绝不良内容传播，净化视频内容，以保持与新媒体发展的最佳实践的接轨。

1) **智能化：**新媒体集成播控平台内容 AI 审核方案，综合运用博汇公司自研的多模态识别引擎以及智能技审引擎，实现了对在线/离线媒资内容审核和质量审查的双管齐下，及时发现其中涉黄、涉暴、敏感人物等违规视音频内容以及黑场、静帧、彩条、抖动等质量劣化。系统完美兼容国产软硬件环境，广泛应用于各级广电监管、播控部门。

2) **便捷化：**本方案采用 B/S、C/S 相结合的交互页面设计，一方面满足人工复审的精细化操作需求，另一方面提供了便捷化的任务管理流转、统计分析手段，有效提升审核工作的便捷性。此外，针对新增敏感样本，系统应用博汇专利的音视频指纹提取、识别技术，无需对全部媒资进行完整处理即可实现快捷精准的素材翻库审核，大大提高了敏感内容甄别的时效性。

3) **标准化：**系统可通过标准化的数据接口与用户现有的媒资系统进行融合，可自动对于媒资系统新注入的节目进行智能审核，并将疑似违规内容推送至业务人员进行确认，最终输出标准化的报表，完美嵌合用户节目内容、技术质量审核业务流程。

图29：系统运行界面



数据来源：公司官网，东吴证券研究所

图30：系统运行界面

数据来源：公司官网，东吴证券研究所

4.5. 东方通

东方通是中国中间件的开拓者和领导者。公司依托基础软件的技术积累，拓展政务、金融等特定行业解决方案，为用户提供基础安全产品及解决方案，同时继续为电信运营商等传统用户提供领先的信息安全、网络安全、数据安全等产品及解决方案，依托“安全+，数据+”两大产品体系，提出“智慧+”战略，在政企数字化转型领域进行产品布局。业务领域从政务、金融、电信、交通等传统优势客户拓展至应急管理、教育、公安、国防军工、能源电力等行业领域。

表9：东方通主要业务及具体内容

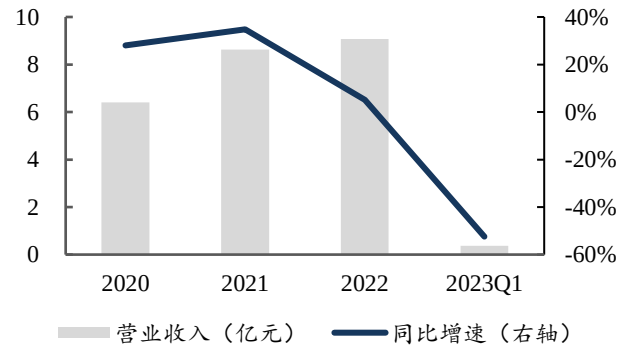
主要业务	具体内容
基础软件中间件	东方通自成立以来，致力于提供具有自主产权、持续创新的中间件技术、产品和方案，为各个行业用户的信息化建设提供坚实支撑。作为国产中间件领域的领导厂商，东方通具有更为完整

	的中间件产品线，可提供应用支撑、数据与应用整合共享等一站式系统架构支撑产品，适用于云计算、大数据、移动互联等各应用场景，满足政务、金融、电信、交通、军工、国央企等各行业用户的需求。
网信安全	东方通网络与信息安全板块以安全管理与安全运营为核心，打造网络与信息安全保障体系及产品体系，核心产品及解决方案包括:信息安全、数据安全、网络安全、内容安全、零信任、安全合规管理、安全中台等。实现统一管控、支撑、呈现与运营，强化安全支撑及创新能力，落实安全闭环管理，建立新一代立体化、纵深防御体系，形成多业务平台协同发展的战略模式，为企业构建“内生安全”防御机制。
智慧应急	具备全时空、全领域、全生命周期的应急管理产品体系与解决方案能力，拥有包括空天地一体化智能监测预警平台、城市安全风险综合监测预警平台、应急指挥平台、安全生产等核心产品集，以优秀的技术能力与业务理解能力，帮助政企客户大幅度提升应急管理能力和水平，提高防灾减灾救灾能力。
数字化转型	公司帮助各行业利用感知、AI、云计算、大数据等信息技术手段进行数字化重塑，构建具有前瞻性业务与运营及服务模式，帮助各企事业单位在资产数字化、运营数字化、决策数字化、业务数字化创新等方面提供从咨询到实践的服务。

数据来源：公司公告，东吴证券研究所

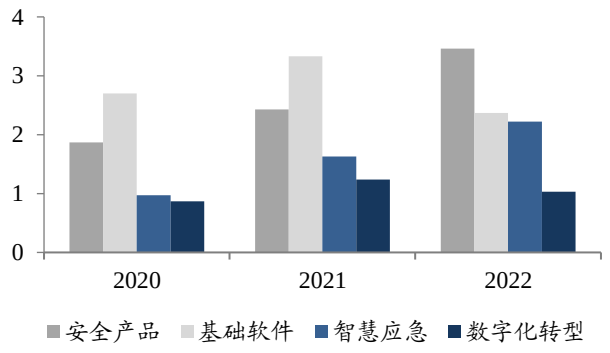
Q1 业绩承压，安全产品成为第一大收入来源。2020-2022 年公司营收稳健增长，分别为 6.40 亿元、8.63 亿元、9.08 亿元。公司安全产品营收增速较快，22 年实现营收 3.46 亿元，占比达 38.11%，超越基础软件中间件成为公司第一大收入来源。

图31：东方通 2020-2023Q1 营收及同比增速



数据来源：wind，东吴证券研究所

图32：东方通 2020-2022 年营收结构



数据来源：wind，东吴证券研究所

东方通重磅推出“产品+技术”组合型智能化内容安全监测体系，基于自主创新的 AI 内容监测软硬件产品+AI 生成式技术深度研究，形成立体化的内容安全监测与管理能力，助力构建清朗网络空间。东方通智能内容安全监测体系主要包括以下 5 大部分:

1) **针对生成式人工智能应用的流量监测专用设备:** 基于 DPI 技术采集关键网络出口流量，结合东方通多年行业积累沉淀的情报库和人工智能应用研究，能够针对性的识别出“生成式人工智能应用”的特征，对其中违法违规的应用进行监测处置。该设备采用了专用的国产硬件，单台设备支持处理 100Gbps 以上网络流量，主要应用于电信运营商、信息安全监管部门等。

2) **瑶光智能内容边缘监测设备:** 基于国产 GPU 开发的软硬一体设备，能够部署于户外，例如产业园区、大型商业区、校园等场景中，可以监测和处置户外广告大屏宣传

中的视频、图片、文字等多样化内容，能够实时发现和处理违法违规内容、深度合成内容、违法广告等，可为广大人民群众提供更为洁净、健康、祥和的日常生活空间。

3) **瑶光内容安全监测系统：**依据《互联网信息服务深度合成管理规定》、《互联网新闻信息服务管理规定》等法规要求，基于深度学习算法的图片识别能力、多维度视频内容识别能力、智能化音频内容识别能力、图像视频高速率伪造检测技术，能够监测网站、微博、微信公众号、小程序、APP、视频平台、IPTV 等新媒体平台的内容安全，能够智能识别 200 多类违法违规内容及深度合成内容，实现内容安全治理的全流程管理。该系统已应用于电信运营商、广电、互联网公司等政企单位，在广播电视、网络视听及新媒体领域，以四川广播电视监测中心四川广播电视和网络视听监测系统为例，东方通瑶光内容安全监测系统在内容发布、网络接入服务等场景中提供了强大的内容监管能力,并可针对当前 AI 换脸诈骗等风险场景提供更为精准的安全监测。

4) **TongGPT 智能语音交互系统：**在涉诈风险领域，随着人工智能的接入，风险隐蔽性更强，反诈工作难度大幅提升。东方通 TongGPT 智能语音交互系统依托东方通 AI 生成式人工智能技术的研究，将推出 TongGPT 多模态智能交互模型，计划年底之前实现涉诈风险智能语音提醒、智能客服等场景 AI 识别服务。尤其面对 AI 生成语音诈骗时，可自动化提醒“受害人员”，解决当前电信网络诈骗治理过程中的涉诈风险提醒效率、覆盖率等难题，提升涉诈风险识别能力，减少人工识别的投入，提高涉诈提醒的及时性、有效性，实现降本增效，提升防范、治理电信网络诈骗等新型违法犯罪的能力。

4) **生成式人工智能算法安全性检测工具：**依据中央网信办发布的《生成式人工智能服务管理办法》（征求意见稿）的要求，基于大模型对抗实现对算法安全性检测，东方通目前已经开启针对 ChatGPT 等新型交互式 AIGC 算法的安全性评估测试方法和工具开发的研究，能够检测生成式人工智能算法的潜在安全风险，包括内容安全性（含违法违规风险、歧视性风险、虚假信息风险等）、隐私数据安全、算法鲁棒性、恶意代码风险等，从算法源头实现检测、预防和治理 AI 内容风险，未来将面向监管部门和算法服务提供者提供全面支撑服务。

4.6. 拓尔思

拓尔思成立于 1993 年，是中文全文检索技术的始创者，领先的人工智能、大数据和数据安全产品及服务提供商。公司作为人工智能、大数据和数据安全产品及服务提供商，为各行业用户的数智化赋能。公司业务根据行业应用的不同，可划分为数字政府、融媒体、金融科技、数字企业、公共安全五个版块；根据技术领域的不同，可划分为人工智能、大数据、数据安全、信创四个领域；根据服务模式的不同，又可划分为软件产品、大数据服务、订阅制 SaaS 服务、软硬一体化产品四种模式。

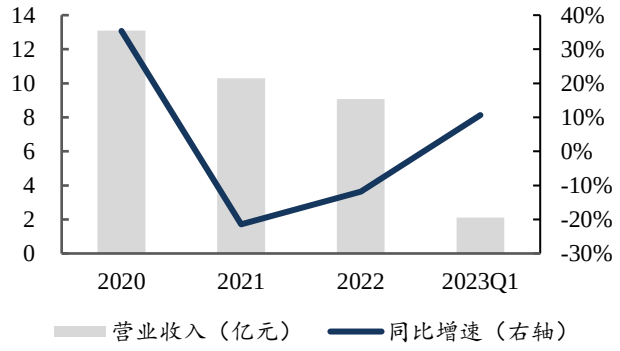
表10：拓尔思不同技术领域及具体内容

技术领域	具体内容
人工智能	公司在 NLP、知识图谱、OCR、图像视频结构化等领域都具备自主可控的多模态内容处理底层技术，处于行业领先地位。2022 年，公司在人工智能领域开展了 6 项重要工作，具体包括预训练大模型和专业模型的融合实践、启动公司自有专业模型 trsGPT 研发、开发具有 AIGC 能力的虚拟人开放云服务平台、开启生成式大模型创新应用规划、发布了多模态人工智能技术平台、推出了基于事理图谱的事件推演分析系统，取得一定成绩。
大数据	在大数据技术平台方面，公司拥有完整的大数据产品矩阵，涵盖数据采集、汇聚、加工、治理、存储、共享、开放等全流程。在数据资产方面，公司 2010 年就自建了大数据中心，以长期服务多行业用户持续积累的开源数据为基础，拥有了规模及质量均位列业界前茅的公开信源数据，目前数据总量超 1400 亿，并仍保持日均亿级数据的采集增长。
数据安全	网络信息内容安全治理方面，公司聚焦网络低俗色情、饭圈乱象、网络暴力等网络生态问题的监测、追踪和分析；内容安全审核方面，公司的文字校对云服务平台能够比较准确、全面、智能地对发布内容中进行内容审核；网络安全方面，公司子公司天行网安是国内最早从事网络安全和数据交换的企业，发明了国内第一台安全隔离网闸，在数据视频交换、单向导入等方面具有雄厚的技术实力。目前公司主要面向政府、公检法、海关等单位提供以数据交换为核心的边界安全、物联网安全、大数据安全三大阵营产品线和解决方案。
信创	公司已经实现了主要软件产品与国内信创领导厂家的基础产品，包括海光、鲲鹏、飞腾、龙芯等芯片，以及统信 UOS、中标麒麟、银河麒麟、中科方德等国产操作系统的适配工作。公司的海贝大数据管理系统是一款从内核到系统完全国产自研的搜索型数据库，是构建搜索引擎应用的核心支撑软件。

数据来源：公司公告，东吴证券研究所

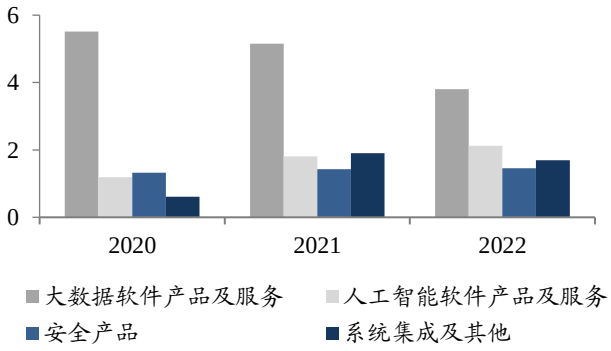
Q1 营收增速回正，人工智能软件产品及服务营收逆势增长。2020-2022 年公司营收分别为 13.09 亿元、10.29 亿元、9.07 亿元。2023Q1 公司营收增速回正，实现营收 2.11 亿元，同比增加 10.62%。公司人工智能软件产品及服务营收逆势稳定增长，2020-2022 年分别实现营收 1.19 亿元、1.81 亿元、2.12 亿元，2022 年营收占比达 23.37%。

图33：拓尔思 2020-2023Q1 营收及同比增速



数据来源：wind，东吴证券研究所

图34：拓尔思 2020-2022 年营收结构



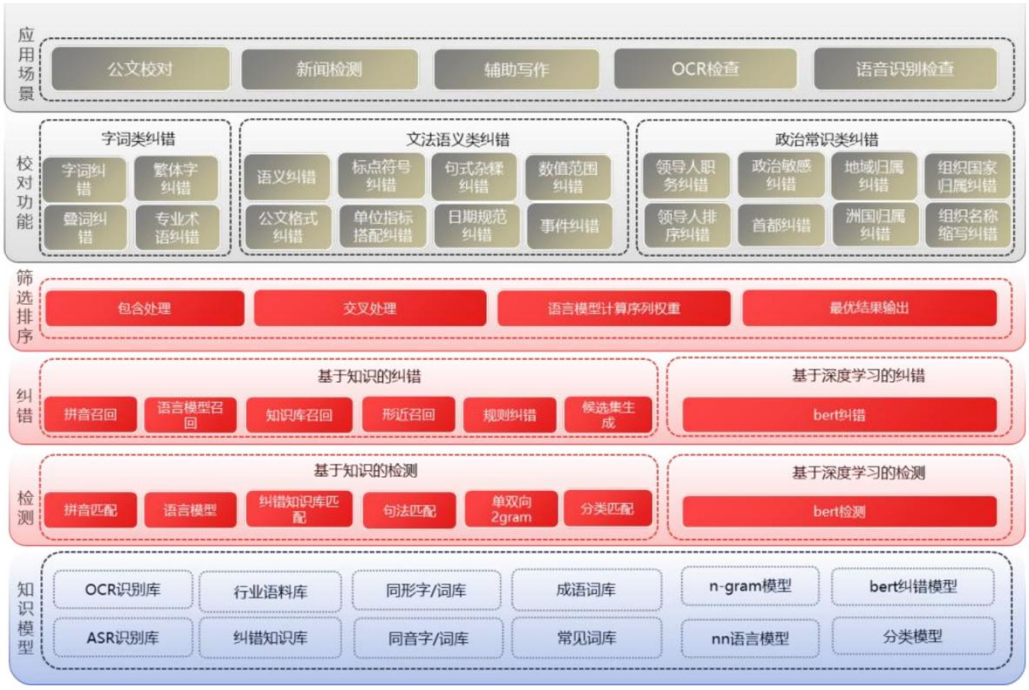
数据来源：wind，东吴证券研究所

拓尔思在政务和媒体等相关领域的平台建设中，涵盖了内容审查相关业务模块。如政府集约化平台服务，通过对信息发布的内容进行自动审核，帮助政务运营团队更准确的信息传达；又如在融媒体平台的建设中，对文本、图片和音视频内容的多模态内容进行实时合法合规性等多维度审校。

拓尔思推出的自动校对的 SaaS 云服务平台，能够比较准确、全面、智能地对发布

内容中进行内容审核，包括文字类差错，如错别字、音近字、形近字、多字、重叠、颠倒、繁体词、异形词等；敏感词过滤，如涉及暴恐、色情、违禁、侮辱、歧视等不健康用词，落马官员等；知识错误，如表述不当、搭配不当、语义错误、术语名词、地名等；常识错误，如标点符号、数字、量词、计量单位、大小写、时间表述等内容。

图35：TRS 自动校对云服务



数据来源：公司官网，东吴证券研究所

5. 风险提示

政策推进不及预期。大模型的推进受到监管节奏的影响，相关政策推进不及预期可能会影响 AI+应用的落地。

行业竞争加剧的影响。细分应用市场空间广阔，可能吸引更多公司参与行业竞争。

AI 监管推进不及预期。AI 监管涉及政策层面和企业层面两方面，如果在实际执行过程中监管力度未达到预期，可能会导致行业发展不及预期。

免责声明

东吴证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

本研究报告仅供东吴证券股份有限公司（以下简称“本公司”）的客户使用。本公司不会因接收人收到本报告而视其为客户。在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议，本公司及作者不对任何人因使用本报告中的内容所导致的任何后果负任何责任。任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。

在法律许可的情况下，东吴证券及其所属关联机构可能会持有报告中提到的公司所发行的证券并进行交易，还可能为这些公司提供投资银行服务或其他服务。

市场有风险，投资需谨慎。本报告是基于本公司分析师认为可靠且已公开的信息，本公司力求但不保证这些信息的准确性和完整性，也不保证文中观点或陈述不会发生任何变更，在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。

东吴证券投资评级标准

投资评级基于分析师对报告发布日后 6 至 12 个月内行业或公司回报潜力相对基准表现的预期（A 股市场基准为沪深 300 指数，香港市场基准为恒生指数，美国市场基准为标普 500 指数，新三板基准指数为三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）），具体如下：

公司投资评级：

买入：预期未来 6 个月个股涨跌幅相对基准在 15%以上；

增持：预期未来 6 个月个股涨跌幅相对基准介于 5%与 15%之间；

中性：预期未来 6 个月个股涨跌幅相对基准介于-5%与 5%之间；

减持：预期未来 6 个月个股涨跌幅相对基准介于-15%与-5%之间；

卖出：预期未来 6 个月个股涨跌幅相对基准在-15%以下。

行业投资评级：

增持： 预期未来 6 个月内，行业指数相对强于基准 5%以上；

中性： 预期未来 6 个月内，行业指数相对基准-5%与 5%；

减持： 预期未来 6 个月内，行业指数相对弱于基准 5%以上。

我们在此提醒您，不同证券研究机构采用不同的评级术语及评级标准。我们采用的是相对评级体系，表示投资的相对比重建议。投资者买入或者卖出证券的决定应当充分考虑自身特定状况，如具体投资目的、财务状况以及特定需求等，并完整理解和使用本报告内容，不应视本报告为做出投资决策的唯一因素。

东吴证券研究所
苏州工业园区星阳街 5 号
邮政编码：215021
传真：（0512）62938527
公司网址： http://www.dwzq.com.cn