

**中国国际金融股份有限公司**

**关于上海合合信息科技股份有限公司**

**首次公开发行股票并在科创板上市的**

**上市保荐书**

**保荐机构**



（北京市朝阳区建国门外大街1号国贸大厦2座27层及28层）

**中国国际金融股份有限公司**  
**关于上海合合信息科技股份有限公司**  
**首次公开发行股票并在科创板上市的上市保荐书**

**上海证券交易所：**

上海合合信息科技股份有限公司（以下简称“合合信息”、“发行人”或“公司”）拟申请首次公开发行股票并在科创板上市（以下简称“本次证券发行”或“本次发行”），并已聘请中国国际金融股份有限公司（以下简称“中金公司”）作为本次发行的保荐人（以下简称“保荐机构”或“本机构”）。

根据《中华人民共和国公司法》（以下简称“《公司法》”）、《中华人民共和国证券法》（以下简称“《证券法》”）、《首次公开发行股票注册管理办法》（以下简称“《首发管理办法》”）、《上海证券交易所股票发行上市审核规则》（以下简称“《上交所发行上市审核规则》”）、《证券发行上市保荐业务管理办法》《保荐人尽职调查工作准则》等法律法规和中国证券监督管理委员会（以下简称“中国证监会”）、上海证券交易所（以下简称“上交所”）的有关规定，中金公司及其保荐代表人诚实守信，勤勉尽责，严格按照依法制订的业务规则、行业执业规范和道德准则出具本上市保荐书，并保证本上市保荐书的真实性、准确性和完整性。

本上市保荐书中如无特别说明，相关用语具有与《上海合合信息科技股份有限公司首次公开发行股票并在科创板上市招股说明书（上会稿）》中相同的含义。

## 目 录

目 录 .....	2
一、发行人概况 .....	3
二、本次发行情况 .....	20
三、负责本次推荐的项目组人员情况 .....	21
四、保荐机构及保荐代表人是否存在可能影响其公正履行保荐职责的情形的说明 .....	22
五、保荐机构按照有关规定应当承诺的事项 .....	23
六、发行人就本次证券发行上市履行的决策程序 .....	24
七、保荐机构关于发行人符合科创板定位及国家产业政策的核查情况 .....	25
八、保荐机构对发行人是否符合上市条件的说明 .....	67
九、保荐代表人对发行人持续督导工作的安排 .....	71
十、保荐机构和相关保荐代表人的联系地址、电话和其他通讯方式 .....	72
十一、保荐机构对本次股票上市的推荐结论 .....	72

## 一、发行人概况

### （一）发行人基本情况

发行人名称：	上海合合信息科技股份有限公司
英文名称：	Intsig Information Co., Ltd.
注册资本：	7,500 万元人民币
法定代表人：	镇立新
成立日期：	2006 年 8 月 8 日
整体变更设立日期：	2020 年 6 月 18 日
公司住所及办公地址：	上海市静安区万荣路 1256、1258 号 1105-1123 室
邮政编码：	200436
电话号码：	021-63061283
传真号码：	021-63061283
互联网网址：	<a href="https://www.intsig.com/">https://www.intsig.com/</a>
电子信箱：	ir@intsig.net

许可项目：第二类增值电信业务。（依法须经批准的项目，经相关部门批准后方可开展经营活动，具体经营项目以相关部门批准文件或许可证件为准）

一般项目：技术服务、技术开发、技术咨询、技术交流、技术转让、技术推广；软件开发；人工智能基础软件开发；人工智能应用软件开发；软件外包服务；基于云平台的业务外包服务；计算机软硬件及外围设备制造；机械设备研发；数据处理服务；数据处理和存储支持服务；广告设计、代理；广告发布；企业管理咨询；信息咨询服务（不含许可类信息咨询服务）；会议及展览服务；市场营销策划；

日用百货销售；电子产品销售；计算机软硬件及辅助设备零售；货物进出口；技术进出口。（除依法须经批准的项目外，凭营业执照依法自主开展经营活动）

## （二）发行人的主营业务

公司是一家人工智能及大数据科技企业，基于自主研发的领先的智能文字识别及商业大数据核心技术，为全球 C 端用户和多元行业 B 端客户提供数字化、智能化的产品及服务。

公司 C 端业务主要为面向全球个人用户的 APP 产品，包括扫描全能王（智能扫描及文字识别 APP）、名片全能王（智能名片及人脉管理 APP）、启信宝（企业商业信息查询 APP）3 款核心产品；公司 B 端业务为面向企业客户提供以智能文字识别、商业大数据为核心的服务，形成了包括基础技术服务、标准化服务和场景化解决方案的业务矩阵，满足客户降本增效、风险管理、智能营销等多元需求，助力客户实现数字化与智能化的转型升级。

通过 10 余年持续的自主研发与技术积累，公司在复杂场景文字识别、智能图像处理、NLP、知识图谱、大数据挖掘等核心技术领域均处于行业领先地位。2012 年，公司被谷歌评为“全球顶尖开发者（Google Play Top Developer）”。在 2019 年的国际文档分析识别大会（ICDAR）中，公司研发团队取得表格识别竞赛的冠军。截至 2022 年 12 月末，公司及其控股子公司已取得发明专利授权合计 117 项，其中中国境内发明专利授权 78 项，境外发明专利授权 39 项；若将基于同一优先权申请在不同国家/地区获批、内容基本相同或相似的发明专利的视为同族专利，且同族专利仅记为 1 项，则中国境内外专利合计为 79 项。

凭借领先的自主研发技术、成熟的产品落地能力、优质的用户体验及服务质量，公司的 C 端产品覆盖了全球百余个国家和地区的亿级用户，B 端服务覆盖了近 30 个行业的企业客户。在 C 端产品方面，截至 2022 年 12 月末，公司扫描全能王、名片全能王、启信宝 3 款 APP 在 App Store 与 Google Play 应用市场的全球用户累计首次下载量合计

超过 8 亿<sup>1</sup>，2022 年 12 月的各渠道月活合计超过 1.3 亿（各渠道月活为直接相加、无法去重，即在多个渠道活跃的同一用户可能被重复计算），其中 APP 渠道的月活合计超过 1.1 亿，小程序、Web 端等其他渠道月活合计约 0.2 亿。截至 2022 年 12 月 31 日，扫描全能王曾在 App Store 上 105 个国家和地区（含中国）的效率类免费应用下载量排行榜位列第一，名片全能王曾在 App Store 上 43 个国家和地区的商务类免费应用下载量排行榜位列第一。2021 年初，公司在 App Annie “2021 Top publisher Awards”中与 Google、Microsoft、Zoom Video Communications 及 Adobe 被评选为全球前五大商业与效率应用企业（Top 5 publishers of Business & Productivity Apps）<sup>2</sup>。在 B 端业务方面，公司智能文字识别与商业大数据服务已覆盖了银行、证券、保险、政府、物流、制造、地产、零售等行业的众多头部客户，《财富》杂志 2022 年发布的世界 500 强公司名单中，公司客户已覆盖超过 125 家。

报告期内，公司营业收入的构成情况如下表所示：

单位：万元

项目		2022 年度		2021 年度		2020 年度	
		金额	占比	金额	占比	金额	占比
智能文字识别	B 端服务	7,120.43	7.20%	6,782.50	8.42%	5,412.40	9.36%
	C 端 APP-扫描全能王	63,287.86	64.03%	49,325.74	61.21%	33,343.21	57.66%
	C 端 APP-名片全能王	2,080.95	2.11%	1,883.05	2.34%	1,667.80	2.88%
	小计	72,489.24	73.34%	57,991.29	71.97%	40,423.41	69.91%
商业大数据	B 端服务	9,759.56	9.87%	6,821.29	8.47%	5,028.77	8.70%
	C 端 APP-启信宝	7,635.28	7.72%	7,315.43	9.08%	6,758.80	11.69%
	小计	17,394.85	17.60%	14,136.72	17.54%	11,787.57	20.39%
互联网广告推广		8,224.26	8.32%	7,823.82	9.71%	4,155.71	7.19%
手机厂商技术授权		504.84	0.51%	517.29	0.64%	803.32	1.39%
其他业务		233.00	0.24%	109.05	0.14%	654.63	1.13%
合计		<b>98,846.18</b>	<b>100.00%</b>	<b>80,578.16</b>	<b>100.00%</b>	<b>57,824.64</b>	<b>100.00%</b>

<sup>1</sup> 注：截至 2022 年 12 月末，公司扫描全能王、名片全能王、启信宝 3 款 APP 在 App Store 与 Google Play 应用市场的全球用户累计首次下载量合计超过 8 亿，其中 2020 年初-2022 年 12 月末的累计新增首次下载量约 3.9 亿，由于部分安卓渠道历史数据不能区分首次与更新下载，所以未统计所有安卓渠道的首次下载量，下同。

<sup>2</sup> 注：该排名由 App Annie 根据各 APP 在全球的 iOS 及 Google Play 应用市场 2020 年整体下载量统计得出，下同。

### （三）核心技术及研发水平

公司旗下主要产品及服务皆由公司自主研发，公司核心技术包括智能文字识别及商业大数据技术。其中智能文字识别技术融合了智能图像处理、复杂场景文字识别、NLP 等 AI 技术，相比传统简单文字识别，具备更多认知与理解能力，可适应多语言、多版式、多样式等复杂场景，并可应用到多个商业化场景中并形成落地的产品或服务，例如票据分类、证照票据结构化、合同关键信息抽取、智能审核等。商业大数据技术包括大数据挖掘与知识图谱等技术，通过“数据→信息→知识→智能”4 个层次，挖掘商业数据背后蕴藏的价值，赋能各行各业。

报告期内核心技术相关业务收入占营业收入比例达 90% 以上。截至 2022 年 12 月 31 日，公司及其控股子公司已取得发明专利授权合计 117 项，其中中国境内发明专利授权 78 项，境外发明专利 39 项，若将基于同一优先权申请在不同国家/地区获批、内容基本相同或相似的发明专利的视为同族专利，且同族专利仅记为 1 项，则中国境内外专利合计为 79 项。

公司智能图像处理、复杂场景文字识别、NLP、知识图谱及大数据挖掘等核心技术均处于行业领先水平，具体如下：

#### 1、智能文字识别技术

智能文字识别可以助推企业与个人实现生产、工作或者生活方式的数字化转型，具有较高的实际应用价值，是 AI 应用中商业推广落地较快的领域，是 AI 领域的重要分支。公司成立至今在智能文字识别领域已拥有超过 10 年的研发经验及技术积累，智能文字识别技术水平行业领先，识别准确率高，且识别精度可随着数据的积累、算法模型的深度学习优化不断提升。公司在智能文字识别领域的核心技术包括智能图像处理、复杂场景文字识别、NLP 技术，不同于传统简单文字识别，技术创新性、先进性主要如下：

（1）智能图像处理技术：公司通过建模实现对复杂图像的精准预处理，例如对模糊图像进行去模糊化和图像增强，对形变图像进行还原矫正，对图像中因光照不均产生的阴影问题进行处理，对不同角度拍摄的图像进行定位和透视变换矫正。

（2）复杂场景的文字识别技术：公司利用多层神经网络从大量数据中进行持续数年的学习，对深度学习算法模型进行多次的迭代优化，有效数据量越大，模型的表现越

好，公司已在文字识别行业深耕超过 15 年，因此在数据积累、深度学习模型训练水平、各种复杂场景下的识别技术具有先发优势。

(3) NLP 技术：公司通过对文本的类型、文档的整体拓扑结构信息、多个文本条目关联前后语义的联想等进行理解与分析，实现关键信息的定位和抽取和结构化输出，同时对已识别的字符进行纠错，例如错字、前后文漏字和符号错误等情况，提升识别精准度。

公司智能文字识别技术所获得的主要奖项及成果如下：

公司早在 2012 年即被谷歌评为“全球顶尖开发者（Google Play Top Developer）”。

在 2019 年的国际文档分析识别大会（ICDAR）中，公司研发团队取得表格识别竞赛的冠军，并与华南理工大学团队联合取得票据识别竞赛的冠军，其中票据识别竞赛的识别对象是购物票据，包含票面质量模糊、格式不统一、票据局部损坏等真实商业场景会遇到的识别难题，表格识别竞赛是将工作中常用的表格作为识别对象，包含复杂的手写样本。ICDAR 票据识别及表格识别竞赛吸引了全球大量队伍参赛，既包括百度、华为、科大讯飞、平安科技、平安财险、华南理工大学、上海交大、华中科技大学等国内知名企业、科研机构，也包括其他众多全球知名团队，例如 Lenovo Research、韩国互联网集团 LINE 旗下的 Clova AI、美国财税软件集团 Intuit、德国罗斯托克大学的 CITlab、阿联酋的国家级人工智能研究院 IIAI（Inception Institute of Artificial Intelligence）、日本的人文开放数据中心、巴西的 Neural Mind 等多个国家团队。

在 2020 年国际模式识别旗舰会议（ICPR）中，公司研发团队与华南理工大学、联想研究院团队合作，在信息图表识别竞赛的 7 大任务 14 个子任务中获得 11 项第一。ICPR 是国际模式识别学会（International Association for Pattern Recognition, IAPR）召开的国际人工智能与模式识别领域的旗舰会议。

在 2023 年国际文档分析识别大会（ICDAR）的文本图像篡改检测算法竞赛（Competition On Detecting Tampered Text in Images）中，公司研发团队取得文本篡改检测任务（Text Manipulation Detection）的冠军。该项竞赛聚焦真实场景下的文本图像篡改检测问题，面临篡改手段多样、篡改痕迹微弱、篡改文本与邻近文本高度相似等技术难点，同时文档版面的复杂化、文本内容的多样化亦加大了篡改检测实施的难度。

公司已研发了基于深度学习的图像篡改检测技术及相应的系统，融合了 SRM、



BayarConv、ELA 等方法以提升 Tamper Detector 性能，综合 RGB 域、频域及噪声域存在痕迹的篡改，捕捉图像在篡改过程中留下的细微痕迹，可检测出擦除、重打印文本、复制、移动、拼接等多种篡改形式，并进行针对性的处理，在提高对篡改区域检测率的同时，防止误判，从而降低系统在实际落地应用中的成本。该技术有利于提升文档图像内容的安全性与可靠性，降低图像造假带来的潜在风险与损失，已在银行、保险和制造业等多种行业进行落地应用。

2023 年 5 月，公司加入中国信通院牵头的《文档图像篡改检测标准》制定工作，该项标准将确立文档图像篡改检测技术的行业标准、为 AI 视觉图像行业健康发展提供指引，由中国图象图形学学会、中国科学技术大学等业内知名机构联合编制。

根据中国科学院上海科技查新咨询中心 2020 年出具的《科技项目咨询报告》（编号 20201410），公司自主研发的“基于人工智能的复杂多场景文档图像识别与理解关键技术及应用”项目具有新颖性，国内外公开文献中未见与该项目技术特点完全相同的技术公开；经其分析，该项目综合技术达到国内领先水平，其中票据文字检测技术达到国际先进水平。

根据中国合格评定国家认可委员会（China National Accreditation Service for Conformity Assessment 英文缩写为：CNAS）认可的检测实验室（NO.CNASL8192）2020 年出具的一系列《软件测试报告》：（1）名片全能王 APP 针对常规的多语言名片的字符识别率平均值为 99.43%，名片识别对象包括简体中文、繁体中文、英文、日文、韩文及西欧语言；（2）扫描全能王 APP 针对常规的印刷体文档字符平均识别率为 99.77%，手写体文档字符平均识别率为 97.00%。

2023 年 5 月，公司在中国信通院旗下中国泰尔实验室的 OCR 智能化服务评估和智能文档处理系统技术能力模块评估中获得最高评级：（1）在识别性能方面，公司的智能文字识别产品在证件类和票据类性能测试中的字段召回率分别为 96.80% 和 97.83%，字符准确率分别为 99.21% 和 99.59%，字段准确率分别为 97.87% 和 98.42%，覆盖旋转、阴影、反光、褶皱、形变、模糊、多语言、低像素、光照不均等复杂场景，各项指标处于行业领先水平；（2）在 AI 核心能力方面，公司的智能文字识别产品在 NLP 领域的文本纠错、文本分类以及知识图谱领域的实体识别、关系抽取等方面获得最高评级；（3）在安全性、可靠性方面，公司亦符合测试标准。

## 2、商业大数据技术

2015 年，公司开始布局大数据产业的商业大数据细分领域，利用前期积累的深度学习、NLP 等 AI 技术的积淀，拓展知识图谱及大数据挖掘技术。公司内部的大数据底层平台快速成熟化，2018 年，公司自主研发的企业知识图谱解决方案开始进行商业落地；2019 年，商业大数据量已突破 700 亿条，汇集中国境内 2.1 亿家企业等组织机构；2020 年，商业大数据量已突破 1,000 亿条，汇集中国境内 2.3 亿家企业等组织机构。

公司在商业大数据领域的核心技术为知识图谱、大数据挖掘、NLP 技术，技术先进性主要体现在：

(1) 多源异构的超大规模动态知识图谱构建技术：公司通过知识图谱、NLP 技术，将企业的工商、司法、信用、舆情等多种多源异构信息中的实体进行对齐和融合，并进一步通过 NLP 技术挖掘出不同实体之间的显性和隐性关系，构建出上亿的节点、数十亿条关系的超大规模商业知识图谱，并在此基础上，利用相关技术将实时更新的多源异构数据动态的融合到知识图谱中，实现知识图谱的动态更新。该技术也同时解决了企业知识图谱构建过程中的同名节点的实体消歧问题。

(2) 超大规模知识图谱推理与挖掘技术：公司通过超大规模知识图谱推理与挖掘技术，对商业知识图谱进行推理与挖掘，获取更多新的商业关联关系。该技术还可以与客户的私有数据进行联合建模，进一步拓展业务场景的边界。例如对企业的股权信息与客户的担保信息等数据进行联合挖掘分析，自动挖掘出集团关系、一致行动人、实际控制人、受益人，集团内企业循环担保圈等企业数据产品。该技术在大规模知识图谱的循环圈检测任务中，将传统方法长达数月的计算时间和数 T（1T=1,024G）的存储量，缩减至 2 小时的计算量和 30G 的存储量。

(3) 大数据搜索与分析技术：可在启信宝平台海量商业大数据中进行毫秒级的实时搜索与分析，为用户提供企业信息画像、关联关系分析、风险查询、营销决策分析等大数据搜索与分析服务，是国内领先、实时、动态的商业大数据搜索与分析平台。

## 3、核心技术应用形成的产品/服务以及产业化情况

凭借领先的自主研发技术、成熟的产品落地能力，公司的 3 款 C 端 APP 产品覆盖了全球百余个国家和地区的亿级用户，品牌知名度与用户体验俱佳。截至 2022 年 12 月末，公司扫描全能王、名片全能王、启信宝 3 款 APP 在 App Store 与 Google Play 应

用市场的全球用户累计首次下载量合计超过 8 亿，2022 年 12 月的各渠道月活合计超过 1.3 亿（各渠道月活为直接相加、无法去重，即在多个渠道活跃的同一用户可能被重复计算），其中 APP 渠道的月活合计超过 1.1 亿，小程序、Web 端等其他渠道月活合计约 0.2 亿。

公司 B 端服务已在银行、证券、保险、政府、物流、制造、地产、零售等近 30 个行业实现成熟应用。公司的智能文字识别服务、商业大数据服务帮助客户切实解决了降本增效、改善风控、高效获客的业务痛点，提供了较高的技术附加值。如下列表格示例：

下游行业	客户痛点	公司智能文字识别技术的产业赋能
政府	<ul style="list-style-type: none"> <li>业务处理流程繁杂、所需时间较长，用户体验待优化</li> <li>机构间通过纸质文件办公，协同效率较低</li> </ul>	<ul style="list-style-type: none"> <li>自动识别、录入各类证件、申请表格等，并与政府机构的核查系统、申请系统等连接，赋能政务自动化转型</li> <li>无纸化办公，提升各机构间协同效率</li> <li>数字化用户体验，提升办事效率，提高用户满意度</li> </ul>
制造业	<ul style="list-style-type: none"> <li>信息化程度较低，大量文档数据未有效管理与利用</li> <li>上下游涉及企业较多，缺乏有效协同运作体系</li> </ul>	<ul style="list-style-type: none"> <li>自动识别、录入各类合同、票据等文件</li> <li>自动识别、录入各类识别编码、日期等信息，赋能设备、生产原材料管理等</li> <li>数字化企业运营，上下游协同工作、流程自动化</li> </ul>
供应链管理	<ul style="list-style-type: none"> <li>竞争加剧，利润承压，急需降低成本，提升效率</li> <li>运单版式多样、经手频次高、人工效率低且易出错</li> </ul>	<ul style="list-style-type: none"> <li>自动识别不同版式的快递单据，并将运单信息高效录入系统，完成核验</li> <li>自动识别快递信息，进行智能分拣，优化流程</li> </ul>
地产	<ul style="list-style-type: none"> <li>业务处理流程繁杂，用户体验较差</li> <li>上下游涉及企业较多，缺乏有效协同运作体系</li> </ul>	<ul style="list-style-type: none"> <li>自动识别、录入用户的各类证件，包括房产证、不动产权证等，并自动将信息录入至对应系统，完成核验</li> <li>数字化用户服务，提升办事效率，提高用户满意度</li> </ul>
银行、证券等金融业	<ul style="list-style-type: none"> <li>大量业务流程仍依赖员工或者用户的人工操作，效率较低，用户体验亟待优化</li> <li>存在海量未电子化的文档数据，未能打通各业务环节文档数据，实现数据价值</li> </ul>	<ul style="list-style-type: none"> <li>多终端文档、图像信息的自动采集，批量扫描识别多种非标准化文档</li> <li>各类表单、文档及工单等文档的智能录入、分类、核验、解析，简化业务流程，降本增效</li> <li>实现自动核保、远程开户等商业化应用,数字化用户体验，提高用户满意度</li> </ul>

下游行业	客户痛点	公司商业大数据技术的产业赋能
政府	<ul style="list-style-type: none"> <li>掌握多源社会数据，但政府各部门间数据未打通，数据价值亟待挖掘，同时决策依赖经验</li> <li>产业园区招商产业化难度大</li> </ul>	<ul style="list-style-type: none"> <li>打破数据孤岛，实现不同地域及部门数据融合，提升各机构间协同效率，赋能管理决策，助力城市管理</li> <li>为区域政府提供科学的经济形势预测、分析企业画像的能力，帮助区域留存优势企业、吸引优质企业入驻、优化营商环境和提前识别和预警企业异常和风险，实现政府对区域内企业的精准监管</li> </ul>
制造业	<ul style="list-style-type: none"> <li>数据量大，且数据源繁杂，但数据标准度不高、数据应用难、数据价值未完全挖掘</li> <li>供应链管理难度大</li> </ul>	<ul style="list-style-type: none"> <li>赋能供应链管理，优化供应商寻源及准入制度，进行关联关系排查，避免围标、串标、陪标等问题，对准入的供应商做全方位的资格审查，在招标后持续对供应商进行舆情、经营风险监控</li> </ul>

下游行业	客户痛点	公司商业大数据技术的产业赋能
供应链管理	<ul style="list-style-type: none"> <li>供应链复杂，但信息化程度较低，上下游信息不对称</li> <li>竞争加剧，利润承压，急需降低成本，提升效率</li> </ul>	<ul style="list-style-type: none"> <li>赋能供应链协同管理，通过上下游企业的信用大数据分析降低物流企业交易成本，降低经营风险</li> <li>优化区域产业布局，促进产业链供应链集群化发展，降低物流成本，提升产业链效率</li> </ul>
地产	<ul style="list-style-type: none"> <li>购地信息、楼盘信息、工程项目信息等非结构化数据繁杂，数据未有效管理，无法进行价值挖掘</li> <li>上下游信息不对称，风控难度大</li> </ul>	<ul style="list-style-type: none"> <li>对区域经济发展趋势、环境数据、市场供需等进行大数据分析，评估地产项目可开拓性，精准选址</li> <li>通过大数据技术协助地产供应链企业资质审核、存量供应商定期尽调，及时发现和控制各类风险，确保交付质量</li> </ul>
银行、证券等金融业	<ul style="list-style-type: none"> <li>上下游信息不对称，风控难度大，依赖风控人员的经验，尽调成本高</li> <li>数据量大，但数据未打通、数据质量低、数据价值未完全挖掘</li> </ul>	<ul style="list-style-type: none"> <li>通过大数据技术赋能银行等机构挖掘优质企业作为信贷支持对象，尤其是科创企业、绿色企业等，支持金融赋能政策鼓励的实体经济发展</li> <li>监控目标企业的各类动态信息，构建信息分析模型，实现风险实时监测和控制</li> <li>为企业资质审查、信贷准入等决策提供依据，提高风控的有效性</li> </ul>

#### （四）主要财务数据及财务指标

报告期内，发行人的主要经营和财务数据及指标情况如下：

项目	2022年 12月31日	2021年 12月31日	2020年 12月31日
资产总额（万元）	129,730.26	95,611.27	61,388.32
归属于母公司所有者权益（万元）	67,611.39	45,194.36	27,142.61
资产负债率（母公司）（%）	22.14	24.32	30.78
资产负债率（合并）（%）	47.88	52.73	55.79
<b>主要财务指标</b>	<b>2022年度</b>	<b>2021年度</b>	<b>2020年度</b>
营业收入（万元）	98,846.18	80,578.16	57,824.64
净利润（万元）	28,371.77	14,448.06	12,952.93
归属于母公司所有者的净利润（万元）	28,372.03	14,448.32	12,952.93
扣除非经常性损益后归属于母公司所有者的净利润（万元）	26,519.24	13,917.88	9,655.71
基本每股收益（元）	3.78	1.93	1.73
稀释每股收益（元）	3.78	1.93	1.73
加权平均净资产收益率（%）	50.30	39.95	66.69
经营活动产生的现金流量净额（万元）	37,676.03	28,768.63	24,973.29
现金分红（万元）	-	-	-
研发投入占营业收入的比例（%）	28.32	29.57	28.87

## （五）发行人存在的主要风险

### 1、与发行人相关的风险

#### （1）技术风险

##### 1) 技术升级、迭代和研发失败的风险

全球人工智能及大数据的技术和产品呈现日新月异的发展态势，新技术、新产品、新模式的更迭速度较快，公司产品及服务需在紧跟全球领先技术的基础上，结合客户的需求变化，进行技术和产品的持续创新更迭。在技术不断升级和迭代的过程中，若研发成果不达预期或不符合客户需求、研发进度落后于行业水平或短期之内无法转化为收入、研发产品出现替代性的创新技术等情况，将会对公司产品的销售和市场竞争力造成不利影响；同时，在新产品开发过程中，公司需要投入大量的人力和资金，若公司前期研发投入的成本无法收回，也会对公司经营业绩产生不利影响。

##### 2) 核心技术泄露及知识产权风险

公司核心技术包括但不限于复杂场景文字识别、智能图像处理、自然语言处理、知识图谱、大数据挖掘等技术，这些核心技术构成了公司的核心竞争优势，是公司在人工智能及大数据领域超过十年的研发经验及技术积累的结晶。如果公司的核心技术遭到泄露或盗用，则会导致公司竞争力下滑、客户流失等一系列不利后果。

公司拥有多项发明专利、计算机软件著作权等知识产权，该等知识产权的应用对公司运营具有极大推动力和商业价值。虽然公司制定了严格的知识产权保密制度，且在报告期内未发生严重的被侵犯知识产权事件，但公司的竞争对手以及其他人员未来可能侵犯公司的知识产权，或可能利用恶意的知识产权诉讼，影响公司的产品销售和推广，对公司正常经营造成不利影响。

#### （2）法律风险

##### 1) 数据安全及个人隐私保护相关风险

基于人工智能及大数据技术，公司面向个人用户及企业客户提供 C 端产品及 B 端服务。业务经营过程中，公司根据业务需要获取了用户的相关数据。对于获取的数据，公司建立了一系列的数据安全内控制度，采用防火墙、数据加密、权限管控、安全审计等技术方式，以保障数据资源存储、使用的安全性、可靠性。但如果受到恶意软件、计

算机病毒、黑客攻击的影响；或公司员工违反公司内部制度规定；数据合作方、客户违反协议约定以及其他原因造成了数据的不当泄露或使用，公司存储的信息数据资源可能被泄露或受到损失。公司还可能因侵犯个人隐私被投诉或受到主管部门处罚，或因侵犯个人隐私及个人信息相关权益导致诉讼或仲裁等纠纷，可能对公司市场声誉及经营业绩造成不利影响。

## 2) 经营资质瑕疵风险

合合信息、上海生腾、苏州贝尔塔报告期内存在从事相关业务未进行企业征信机构备案；上海生腾、上海找贝进行增值电信业务未及时取得《增值电信业务经营许可证》。上海生腾已于 2020 年 7 月 24 日取得企业征信机构备案；上海生腾、上海找贝分别于 2020 年 8 月 17 日、2019 年 12 月 16 日取得《增值电信业务经营许可证》。发行人已就上述瑕疵事宜进行整改。若未来主管部门就前述事项对发行人进行处罚，可能会对公司的生产经营造成不利影响。

## 3) 境外投资手续存在瑕疵风险

公司于 2017 年 3 月在美国设立子公司 CCI 用于对接境外用户的收款。2021 年 8 月，公司全资子公司上海临冠以并购方式受让了 CCI 的股权。公司投资设立 CCI 及转让 CCI 股权未履行境外投资备案手续，上海临冠向 CCI 实缴出资未履行外汇登记手续。为解决前述程序瑕疵，上海临冠新设子公司上海又冠，上海又冠在完成相关中国境内审批程序后设立了美国子公司 CamSoft，公司已将 CCI 的业务全部迁移至 CamSoft。2022 年 4 月 19 日，CCI 依法完成注销手续。

公司投资设立 CCI 及转让 CCI 股权未履行境外投资备案手续，上海临冠向 CCI 实缴出资未履行外汇登记手续，违反了当时适用的境外投资相关法律法规的规定，虽然公司已进行整改，但未来仍存在主管部门就前述事项对发行人进行处罚的风险。

## 4) 竞业禁止相关风险

发行人控股股东、实际控制人镇立新及陈青山、罗希平、龙腾、张彬、郭丰俊原为摩托罗拉中国研究院员工，其在摩托罗拉工作期间的主要工作内容是手写识别系统和语音识别技术的研发。镇立新及上述核心技术人员相继入股、入职发行人，在合合信息的主要工作内容为研究通过摄像头获得字符影像并进行识别，与在摩托罗拉的工作内容不存在重合或相关性。根据上述人员确认，上述人员不存在违反与原任职单位签署的保密

协议或竞业禁止协议的情况，与原单位不存在任何纠纷。

截至本上市保荐书签署日，发行人及上述人员未收到任何来自摩托罗拉提起的有关该等人员违反劳动合同或竞业限制义务的仲裁或诉讼，且至今已超过劳动争议或普通民事案件的诉讼时效，但不能排除如未来摩托罗拉就前述事项向上述人员提起仲裁或诉讼，则可能会对公司经营产生不利影响。

### 5) 数据类纠纷相关风险

上海生腾作为启信宝产品的运营主体已完成企业征信机构备案，根据《征信业管理条例（国务院令 第 631 号）》合法收集并展示相关企业的公开信息。虽然发行人已设置自动化访问定期检查机制，但由于被采集网站实时更新、变化，该种采集方式仍存在一定滞后性、以及个别数据采集与实际情况不匹配的情形。若未来有新增重大数据类纠纷案件，可能会对公司经营造成不利影响。

## (3) 财务及内控风险

### 1) 增值税税收优惠政策变化的风险

根据财税[2011]100 号《财政部 国家税务总局关于软件产品增值税政策的通知》：增值税一般纳税人销售其自行开发生产的软件产品，按 17% 税率征收增值税后，对其增值税实际税负超过 3% 的部分实行即征即退政策。

报告期内，公司享受的主要增值税税收优惠政策为：公司销售自行开发生产的软件产品，对其增值税实际税负超过 3% 的部分实行即征即退政策。报告期各期，公司获得的税收优惠分别为 686.73 万元、3,002.97 万元和 2,204.37 万元。若未来增值税税收优惠政策发生变化，或公司未能满足相关要求导致无法享受税收优惠，可能对公司的利润水平产生一定程度负面影响。

### 2) 政府补助政策变化的风险

报告期内，公司收到来自多个政府部门给予的补助收入。报告期各期，公司获得的政府补助对利润的影响金额分别为 749.73 万元、244.88 万元和 1,509.79 万元。若未来政府补助政策发生变化，或公司未能满足相关要求导致无法取得政府补助，可能对公司的利润水平产生一定程度负面影响。

### 3) 汇率波动的风险

发行人实行国际化发展战略，在北美、东亚、欧洲、南美洲等境外国家和地区均有业务收入。报告期内，发行人来源于境外地区的收入占比分别为 31.67%、38.03% 和 35.78%，主要通过美元进行结算。受汇率波动影响，公司报告期内的汇兑损益分别为 -1,310.73 万元、-641.27 万元和 -2,576.58 万元。随着公司海外业务规模的扩张，公司海外收入可能进一步增加，而人民币汇率受到国内外经济、政治等多种因素的影响存在波动，因此公司存在因汇率波动影响公司利润水平的风险。

#### **4) 内控风险**

公司不断加大研发投入，科技创新能力稳步提升，对公司的内控水平也提出了更高的要求。虽然公司已经建立起一系列完善的企业内控制度，但随着公司业务的发展，产品与服务新业态、新模式的拓展，如果内控制度不能及时完善，管理能力和管理团队无法及时得到提升和优化，将可能引发相应的内控风险，进而对公司整体经营产生不利影响。

#### **(4) 经营风险**

##### **1) 核心人才流失的风险**

公司人工智能及大数据业务的蓬勃发展与一批稳定的技术研发人员和管理人员紧密相关，公司目前设立了相对完善的公司技术人才培养体系，配套相应的培训、考核与激励相结合的机制，获得公司上下的高度认可；同时，为了更好地激励公司骨干技术成员与核心管理人员，公司实施了股权激励计划。但公司未来仍然可能存在核心人才流失的风险，从而对公司业务产生不利影响。

##### **2) 经营场所租赁的风险**

报告期内，公司经营场所是由租赁方式取得，若公司在租赁期间内因某些原因而无法继续租赁，则可能在短期内对公司的正常经营带来不利影响。截至本上市保荐书签署日，公司及其子公司对外承租的 11 处主要房产，除 2 处房产正在办理产权证书外，均具备产权证书，但均未完成租赁备案手续。根据住房和城乡建设部《商品房屋租赁管理办法》的规定，发行人存在被主管部门责令改正的风险，逾期不改正的，针对每一份房屋租赁合同存在被处以 1 千元以上、1 万元以下罚款的风险。

##### **3) 实际控制人不当控制的风险**



截至本上市保荐书签署日，公司实际控制人镇立新直接持有公司 32.2504% 的股权，通过上海狮吼间接控制上海目一然、上海端临、上海融梨然及上海顶螺持有的公司股权，合计控制公司 39.9331% 的股份。如果按本次预计发行数量 2,500 万股计算，本次发行后，镇立新合计控制公司股份比例近 30%，且同时担任公司董事长、总经理及法定代表人，为公司的控股股东、实际控制人。

控股股东、实际控制人可能通过股东大会和董事会对公司的重大经营决策以及业务、管理、人事安排等方面施加控制和影响，从而形成有利于实际控制人的决策并有可能损害公司及其他股东的利益。

#### 4) 管理能力不能适应公司业务发展的风险

本次发行后，随着募集资金的到位和募集资金投资项目的实施，公司总体经营规模将进一步扩大。这将对公司在战略规划、组织结构、内部控制、运营管理、财务管理等方面提出更高的要求。如果公司管理层不能持续有效地提升管理能力、优化管理体系，将导致公司管理体系不能完全适应公司业务的快速发展，对公司未来的经营和持续盈利能力造成不利影响。

#### 5) 公司 C 端产品被消费者投诉的风险

公司 C 端业务主要为扫描全能王、名片全能王、启信宝 APP 3 款核心产品。虽然公司已在相关 APP 支付界面明确提示用户自动订阅的相关内容，并且已根据《网络交易监督管理办法》对相关 APP 的自动订阅服务内容进行了调整，但未来若出现公司未能及时解决消费者投诉等事件的情况，将对公司产品的口碑和销售情况造成不利影响。

## 2、与行业相关的风险

### (1) 市场竞争加剧的风险

随着公司所在的人工智能及大数据行业技术日趋成熟，政府的多项利好政策加速行业的整体生态建设，未来的市场进入者预期将不断增加，行业整体竞争未来将呈现逐步加剧态势。

在商业大数据业务方面，启信宝所处的 C 端 APP 市场存在一定的同质化竞争，报告期内启信宝付费用户比例以及新增付费转化率持续下降主要系与主要竞争对手相比发行人营销推广投入较低，报告期内启信宝的广告宣传费分别为 7,095.48 万元、6,384.77

万元和 2,204.43 万元。主要竞争对手天眼查、企查查近几年通过聘请代言人、线下广告等方式在广告宣传方面投入了较大力度的资源。2020 年百度上线了自有的企业查询平台“爱企查”，主要面向 C 端用户提供企业信息免费查询服务，并于 2022 年 5 月上线付费会员，结合线下硬广等其他广告营销方式，凭借其互联网搜索平台的用户流量优势，进一步加剧了行业竞争。启信宝相比主要竞争对手的竞争劣势是：未聘请代言人，上线至今市场营销投入有限，在早期用户积累、后续用户流量获取方面存在劣势，市场占有率与行业前两名存在一定差距，考虑到市场的同质化竞争将短期内持续存在，未来启信宝 APP 存在品牌知名度下降、用户流失或转移至其他竞品、收入下降、短期内无法盈利的风险。

在智能文字识别业务方面，（1）扫描全能王面临来自多个包含扫描功能或文字识别功能的移动端 APP 的竞争，包括一些文档编辑类软件（WPS、福昕等）、云存储类软件（百度网盘、夸克等）、通讯协作类软件（钉钉、QQ、企业微信、微信等）均包含扫描与文字识别功能，未来竞争的加剧可能会导致扫描全能王 VIP 用户续费率及收入的下降。（2）名片全能王产品已进入产品发展的相对成熟阶段，并面临来自微信等社交软件的竞争，部分用户使用社交软件交换联系方式来替代交换纸质名片，报告期内名片全能王的 APP 月活出现一定下降，未来存在月活进一步下降及收入减少的风险。

（3）公司报告期内智能文字识别 B 端基础技术服务业务分别实现收入 3,013.02 万元、4,146.36 万元和 4,343.86 万元，占相应期间公司营业收入总额比例分别为 5.21%、5.15% 和 4.39%，主要是因为此领域产品偏向同质化，竞争较为激烈所致；百度、腾讯、阿里等综合型 AI 厂商均在其云平台提供智能文字识别标准化服务，加剧了公司智能文字识别 B 端业务所面临的行业竞争，未来存在收入下降的风险。（4）市场上部分主流手机自带的基础扫描功能和公司智能文字识别 C 端产品免费功能中的部分功能有所重叠，会分流一部分的免费用户，一定程度上加剧了公司所面临的行业竞争；另外，公司手机技术授权业务的主要客户为手机厂商，手机厂商拓展手机自带的扫描功能对公司手机技术授权业务带来一定不利影响，报告期内，手机厂商技术授权业务分别实现收入 803.32 万元、517.29 万元和 504.84 万元，占相应期间公司营业收入总额比例分别为 1.39%、0.64% 和 0.51%，2021 年度与 2022 年度分别同比减少 35.61% 和 2.41%，未来存在收入持续下降的风险。若公司未能保持自身在技术积累、服务模式以及项目经验上的优势，激烈的市场竞争将会对公司业务拓展及可持续经营带来不利影响。

## （2）海外经营风险

报告期内，公司在欧美、东亚等主要国家与地区，针对海外客户的不同偏好与需求进行 C 端产品及 B 端服务的销售。2020 年、2021 年和 2022 年，公司的境外及中国港澳台收入分别为 18,311.91 万元、30,641.36 万元和 35,364.76 万元，占公司总收入比重分别为 31.67%、38.03%和 35.78%，境外收入主要集中在亚洲（主要为日本与韩国）及北美洲（主要为美国），收入保持稳定增长趋势。

海外业务经营的合规风险方面，尽管公司自从事海外经营以来未受到过收入来源地的处罚，随着业务规模的进一步扩大，公司境外业务涉及的法律法规规制环境将会更加复杂，如果未来公司未能完全遵守产品销售地的法律或法规，则可能面临相应的处罚，进而影响其在当地的经营。

海外业务经营的政策风险方面：

（1）2020 年 6 月，印度政府以保护国家安全为由宣布封杀 59 款中国应用程序，涉及包括发行人、字节跳动、腾讯、百度在内的多家中国企业，发行人的扫描全能王（CamScanner）、名片全能王（CamCard）产品在印度主流应用市场被迫下架。2020 年、2021 年和 2022 年，扫描全能王、名片全能王在印度地区产生的收入分别为 716.58 万元、245.24 万元和 31.77 万元，占公司各期总收入比重分别为 1.24%、0.30%和 0.03%。

（2）2021 年 1 月 5 日，美国时任总统特朗普签署行政命令，以保护美国公民的隐私和数据安全为由，禁止美国司法管辖范围内的企业及个人与 8 家中国应用软件进行交易，其中包括扫描全能王（CamScanner）、支付宝、QQ 钱包、腾讯 QQ、微信支付和 WPS Office 等，禁令在 45 天后生效，2021 年 6 月 11 日，该项行政命令被新任美国总统拜登撤销（撤销前该禁令未实际执行），但美国联邦政府及各州政府仍在对有可能威胁美国国家安全、美国公民数据安全的中国应用软件进行评估分析，不排除未来有针对扫描全能王的新禁令颁发的可能性。2020 年、2021 年和 2022 年，扫描全能王通过中国境内外 App Store 及 Google Play 应用市场收款对应产生的收入以及在美国地区的互联网广告推广服务业务合计收入分别为 22,650.80 万元、37,787.30 万元和 46,014.07 万元，占公司各期总收入比重分别为 39.17%、46.90%和 46.55%。近期，中美关系正面临严峻局面，美国政府已经对一些中国科技公司施加威胁或实施出口管制、经济和贸易制裁措施，如果未来类似或更多的行政命令、出口管制或制裁措施出台，公司的业务可能受到

不利影响。

如公司境外业务所在国家和地区的产业政策或者政治经济环境发生重大变化、或因国际关系紧张、战争、贸易制裁等无法预知的因素或其他不可抗力而导致境外经营状况受到影响，将可能给公司境外业务的正常开展和持续发展带来潜在不利影响。

### **(3) 政策变化的风险**

公司所处行业具体细分领域为人工智能及大数据软件领域，国家颁布了《个人信息保护法》《数据安全法》《App 违法违规收集使用个人信息行为认定方法》《关于加强科技伦理治理的意见》等多项法律法规及产业政策，不断加强对数据安全、人工智能领域科技伦理治理的规范要求并提升对个人隐私的保护力度。虽然目前发行人的主营业务不涉及科技伦理敏感领域，若未来行业监管政策发生变化，或已有的行业政策及标准发生更新，而公司不能及时适应或满足未来可能更新或出台的监管政策标准，将对公司的持续经营及业务合规性产生不利影响。

截至本上市保荐书签署日，公司已取得开展现有业务所必须的资质。但如果各级行政主管部门提高或新增对公司目前经营业务的资质许可要求，或出台对数据安全、个人隐私保护方面更加严格的政策，或公司未来拓展出新的商业模式，而新模式可能需要新的资质许可。如果公司无法及时获得相关资质许可，或无法满足政策变化的要求，则有可能对公司业务经营造成不利影响。

### **(4) 人力成本上升的风险**

公司所处行业属于技术密集型行业，业务发展需要大量的高技能软件研发人员。报告期内，公司经营规模快速扩张，截至 2022 年 12 月 31 日，公司全职员工共 912 人，其中研发人员共 499 人，占总人数比重约为 55%。结合公司现有的商业模式，人力成本是公司成本费用的主要组成部分之一。随着公司所处行业人才竞争的加剧、社会整体薪酬水平的提升，公司的员工薪酬开支也呈现上升趋势，导致公司面临营业成本及费用持续上升的压力。

## **3、其他风险**

### **(1) 发行失败风险**

若本次发行时有效报价投资者或网下申购的投资者数量不足法律规定要求，本次发

行应当中止，若发行人中止发行上市审核程序超过交易所规定的时限，或者存在其他影响发行的不利情形，发行人将存在发行失败的风险。

## （2）募集资金相关风险

本次发行资金将用于人工智能 C 端产品及 B 端服务研发升级项目、商业大数据 C 端产品及 B 端服务研发升级项目、人工智能核心技术研发升级项目、商业大数据核心技术研发与数据中台建设项目。本次募集资金投资项目的投资金额较大，项目管理和组织实施是项目成功的关键，将直接影响到项目的进展和项目的质量。若投资项目不能按期完成，将对公司的盈利状况和未来发展产生不利影响。

本次发行完成后，公司净资产规模将有较大幅度提高，而募集资金经济效益的产生有一定的研发投入及市场扩展周期；同时，募集资金项目建成后，资产有较大幅度的增加，这将产生一定的折旧摊销。因此，短期内预计公司净利润增长幅度将小于净资产增长幅度，存在公司净资产收益率较以前年度有所降低的风险。

## （3）舆情相关风险

随着业务发展及产品影响力的提升，发行人也持续受到媒体关注和报道，主要关注事项包括：公司的经营资质及合规经营、海外经营情况、科创属性、数据安全与个人隐私保护、C 端业务发展可持续性 & 收入真实性、募集资金的必要性等。虽然公司已持续关注各类媒体报道并进行舆情管理工作，但未来若出现有关公司的重大负面不实报道而公司未及时核实说明，将对公司的经营发展产生不利影响。

## 二、本次发行情况

（一）本次发行的基本情况			
股票种类	人民币普通股（A 股）		
每股面值	1.00 元		
发行股数	不超过 2,500 万股（含 2,500 万股，且不低于本次发行后公司总股本的 25%），超额配售部分不超过本次公开发行股票数量的 15%	占发行后总股本比例	不低于 25%
其中：发行新股数量	不超过 2,500 万股（含 2,500 万股，且不低于本次发行后公司总股本的 25%），超额配售部分不超过本次公开发行股票数量的 15%	占发行后总股本比例	不低于 25%

股东公开发售股份数量	不适用	占发行后总股本比例	不适用
发行后总股本	不超过【】万股（未考虑公司本次发行的超额配售选择权）；不超过【】万股（若公司全额行使本次发行的超额配售选择权）		
每股发行价格	【】元		
发行市盈率	【】倍		
发行前每股净资产	【】元	发行前每股收益	【】元
发行后每股净资产	【】元	发行后每股收益	【】元
发行市净率	【】倍		
发行方式	采取网下向询价对象配售发行与网上按市值申购定价发行相结合的方式或中国证监会、上交所认可的其他发行方式		
发行对象	符合资格的询价对象和在上交所开立（A股）股票账户且符合相关法律法规关于科创板股票投资者适当性条件的中国境内自然人、法人等投资者以及符合中国证监会、上交所规定的其他投资者（国家法律、法规和规范性文件以及公司需遵守的其他监管要求所禁止购买者除外）		
承销方式	余额包销		
拟公开发售股份股东名称	不适用		
发行费用的分摊原则	【】		
募集资金总额	【】万元		
募集资金净额	【】万元		
募集资金投资项目	人工智能C端产品及B端服务研发升级项目		
	商业大数据C端产品及B端服务研发升级项目		
	人工智能核心技术研发升级项目		
	商业大数据核心技术研发与数据中台建设项目		
发行费用概算	本次发行费用总额为【】万元，承销及保荐费用【】万元，律师费用【】万元，审计及验资费用【】万元，发行手续费用及其他【】万元		
<b>（二）本次发行上市的重要日期</b>			
刊登发行公告日期	【】		
开始询价推介日期	【】		
刊登定价公告日期	【】		
申购日期和缴款日期	【】		
股票上市日期	【】		

### 三、负责本次推荐的项目组人员情况

#### （一）保荐代表人

冷小茂：于2020年取得保荐代表人资格，曾经参与/执行永安行科技股份有限公司

公开发行可转换公司债券项目、顾家家居股份有限公司主板上市项目、浙江康隆达特种防护科技股份有限公司主板上市项目，永和流体智控股份有限公司中小板上市项目、江苏新泉汽车饰件股份有限公司主板上市项目，在保荐业务执业过程中严格遵守《证券发行上市保荐业务管理办法》等相关规定，执业记录良好。

刘文博：于 2020 年取得保荐代表人资格，曾经参与/执行江苏国茂减速机股份有限公司主板 IPO、浙江天正电气股份有限公司主板 IPO 等项目，在保荐业务执业过程中严格遵守《证券发行上市保荐业务管理办法》等相关规定，执业记录良好。

## （二）项目协办人及其他项目组成员

项目协办人：杨雅菲，于 2017 年取得证券从业资格。

项目组其他成员：曹宇、孙远、高婧、程卓、方泽宇、赵越、何友。

## 四、保荐机构及保荐代表人是否存在可能影响其公正履行保荐职责的情形 的说明

1、中金公司或其控股股东、实际控制人、重要关联方不存在持有发行人或其控股股东、实际控制人、重要关联方股份的情况；中金公司将安排相关子公司参与本次发行战略配售，具体按照上交所相关规定执行。中金公司及其相关子公司后续将按要求进一步明确参与本次发行战略配售的具体方案，并按规定向上交所提交相关文件。

2、截至本上市保荐书签署之日，保荐机构或其控股股东、实际控制人、重要关联方不存在持有发行人或其控股股东、实际控制人、重要关联方股份的情况；

3、截至本上市保荐书出具之日，发行人或其控股股东、实际控制人、重要关联方不存在持有本机构及本机构 5% 以上股东股份及下属子公司股份的情况；

4、截至本上市保荐书出具之日，除本上市保荐书已说明情况外，中金公司的保荐代表人及其配偶，董事、监事、高级管理人员不存在持有发行人或其控股股东、实际控制人、重要关联方股份的情况，不在发行人或其控股股东、实际控制人、重要关联方任职；

5、中金公司控股股东为中央汇金投资有限责任公司（以下简称“中央汇金”或“上级股东单位”），截至 2022 年 6 月 30 日，中央汇金直接持有中金公司约 40.11% 的股

权,同时,中央汇金的下属子公司中国建银投资有限责任公司、建投投资有限责任公司、中国投资咨询有限责任公司合计持有中金公司约 0.06%的股权。中央汇金为中国投资有限责任公司的全资子公司,中央汇金根据国务院授权,对国有重点金融企业进行股权投资,以出资额为限代表国家依法对国有重点金融企业行使出资人权利和履行出资人义务,实现国有金融资产保值增值。中央汇金不开展其他任何商业性经营活动,不干预其控股的国有重点金融企业的日常经营活动。截至本上市保荐书出具之日,中金公司上级股东单位与发行人控股股东、实际控制人、重要关联方不存在相互提供担保或者融资的情况。

6、本机构与发行人之间不存在其他关联关系。

本机构依据相关法律法规和公司章程,独立公正地履行保荐职责。

## 五、保荐机构按照有关规定应当承诺的事项

(一)本机构已按照法律、行政法规和中国证监会、上交所的规定,对发行人及其发起人、控股股东、实际控制人进行了尽职调查和审慎核查,同意推荐发行人证券发行上市,并据此出具本上市保荐书。

(二)作为上海合合信息科技股份有限公司本次发行的保荐机构,本机构:

1、有充分理由确信发行人符合法律法规及中国证监会有关证券发行上市的相关规定;

2、有充分理由确信发行人申请文件和信息披露资料不存在虚假记载、误导性陈述或者重大遗漏;

3、有充分理由确信发行人及其董事在申请文件和信息披露资料中表达意见的依据充分合理;

4、有充分理由确信申请文件和信息披露资料与证券服务机构发表的意见不存在实质性差异;

5、保证所指定的保荐代表人及本保荐机构的相关人员已勤勉尽责,对发行人申请文件和信息披露资料进行了尽职调查、审慎核查;

6、保证本上市保荐书与履行保荐职责有关的其他文件不存在虚假记载、误导性陈述或者重大遗漏;



7、保证对发行人提供的专业服务和出具的专业意见符合法律、行政法规、中国证监会及上交所的规定和行业规范；

8、自愿接受中国证监会、上交所依照相关法律、行政法规采取的监管措施。

## 六、发行人就本次证券发行上市履行的决策程序

经核查，发行人已就本次证券发行履行了《公司法》《证券法》及中国证监会、上交所规定的决策程序，具体如下：

2020年12月22日、2021年7月26日、2021年12月20日，发行人召开第一届董事会第五次会议、第一届董事会第九次会议、第一届董事会第十一次会议，审议通过了《关于公司申请首次公开发行股票并在上海证券交易所科创板上市的议案》《关于公司首次公开发行股票募集资金投资项目及其可行性的议案》《关于制定<公司上市后三年稳定股价的预案>的议案》《关于公司就首次公开发行股票并上市事项出具有关承诺并提出相应约束措施的议案》《关于首次公开发行股票前滚存利润分配方案的议案》《关于制定<公司上市后三年股东分红回报规划>的议案》《关于制定<公司章程（草案）>的议案》《关于调整公司首次公开发行股票并在上海证券交易所科创板上市报告期的议案》等与本次发行及上市相关的议案，并同意将前述相关议案提交公司股东大会审议。前述董事会决策程序符合《公司法》和《公司章程》的规定，符合《首发管理办法》第十四条的规定。

2021年8月11日、2022年1月6日，发行人召开2021年第二次临时股东大会、2022年第一次临时股东大会，审议通过了本次发行相关的议案。

综上，保荐机构认为：发行人本次发行已获得董事会、股东大会的批准，发行人董事会已取得股东大会关于本次发行的授权，发行人就本次发行履行的决策程序符合《公司法》《证券法》《首发管理办法》等法律法规以及发行人公司章程的相关规定。

## 七、保荐机构关于发行人符合科创板定位及国家产业政策的核查情况

### （一）发行人符合科创板支持方向的核查情况

#### 1、发行人符合国家科技创新战略相关要求

发行人智能文字识别与商业大数据的产品及服务与国家科技创新战略和相关产业政策匹配，属于国家战略重点支持的领域。

##### （1）智能文字识别业务

公司的智能文字识别技术是人工智能的重要分支，该技术融合了智能图像处理、基于深度学习的复杂场景文字识别、自然语言处理（NLP）等多项 AI 技术。发行人智能文字识别业务与国家科技创新战略和相关产业政策匹配，具体如下：

**1) 人工智能已被纳入我国“十三五”与“十四五”纲要，复杂场景文字识别技术已被科技部列入“新一代人工智能共性关键技术”**

2016年11月国务院发布《“十三五”国家战略性新兴产业发展规划》（简称为“十三五纲要”），其中提出“培育人工智能产业生态，促进人工智能在经济社会重点领域推广应用，打造国际领先的技术体系，鼓励各行业加强与人工智能融合，逐步实现智能化升级”。

2021年3月十三届全国人大四次会议表决通过《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》（简称为“十四五纲要”），其中提出“加强关键数字技术创新应用、加快推动数字产业化、推进产业数字化转型”、“人工智能领域的‘智能识别系统’是数字经济重点产业。”

“复杂版面手写图文识别及理解关键技术研究”已被科技部列入“新一代人工智能共性关键技术”。公司的智能文字识别技术是人工智能的重要分支，是国家重点支持发展的技术。

**2) 公司智能文字识别业务响应了国家“促进人工智能和实体经济深度融合”的政策**

2017年12月，工信部发布《促进新一代人工智能产业发展三年行动计划（2018—2020年）》，提出“以市场需求为牵引，积极培育人工智能创新产品和服务，促进人工智能技术的产业化，推动智能产品在工业、医疗、交通、农业、金融、物流、

教育、文化、旅游等领域的集成应用”、“培育智能理解产品，加快模式识别、智能语义理解、智能分析决策等核心技术研发和产业化，支持设计一批智能化水平和可靠性较高的智能理解产品或模块，优化智能系统与服务的供给结构”。

2019年3月，中央全面深化改革委员会第七次会议审议通过《关于促进人工智能和实体经济深度融合的指导意见》，提出“促进人工智能和实体经济深度融合，要把握新一代人工智能发展的特点，坚持以市场需求为导向，以产业应用为目标，深化改革创新，优化制度环境，激发企业创新活力和内生动力”。

2020年9月，国务院办公厅发布《国务院办公厅关于全面加强新时代语言文字工作的意见》，强调“推动语言文字信息技术创新发展，发挥语言文字信息技术在国家信息化、智能化建设中的基础支撑作用；大力推动语言文字与人工智能、大数据、云计算等信息技术的深度融合，加强人工智能环境下自然语言处理等关键问题研究和原创技术研发，加强语言技术成果转化及推广应用，支持数字经济发展”。

2022年7月，科技部等六部门印发的《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》提出“围绕高端高效智能经济培育打造重大场景。鼓励在制造、农业、物流、金融、商务、家居等重点行业深入挖掘人工智能技术应用场景，促进智能经济高端高效发展”。

文档广泛地存在于各类实体经济行业的业务场景，从非标准化文档中提取结构化信息、进行数字化转型在多种行业中均存在需求。公司在政府、制造业、供应链管理、地产、金融等多个行业深入挖掘人工智能技术的应用场景，促进人工智能技术和实体经济深度融合：

下游行业	客户痛点	公司智能文字识别技术的产业赋能
政府	<ul style="list-style-type: none"> <li>•业务处理流程繁杂、所需时间较长，用户体验待优化</li> <li>•机构间通过纸质文件办公，协同效率较低</li> </ul>	<ul style="list-style-type: none"> <li>•自动识别、录入各类证件、申请表格等，并与政府机构的核查系统、申请系统等连接，赋能政务自动化转型</li> <li>•无纸化办公，提升各机构间协同效率</li> <li>•数字化用户体验，提升办事效率，提高用户满意度</li> </ul>
制造业	<ul style="list-style-type: none"> <li>•信息化程度较低，大量文档数据未有效管理与利用</li> <li>•上下游涉及企业较多，缺乏有效协同运作体系</li> </ul>	<ul style="list-style-type: none"> <li>•自动识别、录入各类合同、票据等文件</li> <li>•自动识别、录入各类识别编码、日期等信息，赋能设备、生产原材料管理等</li> <li>•数字化企业运营，上下游协同工作、流程自动化</li> </ul>
供应链管理	<ul style="list-style-type: none"> <li>•竞争加剧，利润承压，急需降低成本，提升效率</li> <li>•运单版式多样、经手频次高、人工效率低且易出错</li> </ul>	<ul style="list-style-type: none"> <li>•自动识别不同版式的快递单据，并将运单信息高效录入系统，完成核验</li> <li>•自动识别快递信息，进行智能分拣，优化流程</li> </ul>

下游行业	客户痛点	公司智能文字识别技术的产业赋能
地产	<ul style="list-style-type: none"> <li>•业务处理流程繁杂，用户体验较差</li> <li>•上下游涉及企业较多，缺乏有效协同运作体系</li> </ul>	<ul style="list-style-type: none"> <li>•自动识别、录入用户的各类证件，包括房产证、不动产权证等，并自动将信息录入至对应系统，完成核验</li> <li>•数字化用户服务，提升办事效率，提高用户满意度</li> </ul>
银行、证券等金融业	<ul style="list-style-type: none"> <li>•大量业务流程仍依赖员工或者用户的人工操作，效率较低，用户体验亟待优化</li> <li>•存在海量未电子化的文档数据，未能打通各业务环节文档数据，实现数据价值</li> </ul>	<ul style="list-style-type: none"> <li>•多终端文档、图像信息的自动采集，批量扫描识别多种非标准化文档</li> <li>•各类表单、文档及工单等文档的智能录入、分类、核验、解析，简化业务流程，降本增效</li> <li>•实现自动核保、远程开户等商业化应用,数字化用户体验，提高用户满意度</li> </ul>

### 3) 公司智能文字识别业务响应了国家“加快数字化发展、建设数字中国”的政策

2021年3月发布的十四五纲要专门设置“加快数字化发展、建设数字中国”章节，提出“加强关键数字技术创新应用、加快推动数字产业化、推进产业数字化转型”、“加快建设数字经济、数字社会、数字政府，以数字化转型整体驱动生产方式、生活方式和治理方式变革”的要求。

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》，强调“促进数字经济和实体经济深度融合，以数字化驱动生产生活和治理方式变革”，指出“整体提升应用基础设施水平，加强传统基础设施数字化、智能化改造”。

公司的智能文字识别技术可帮助政府及各行各业的企业客户及个人用户实现治理方式、生产方式、生活方式的数字化转型。公司的产品不仅仅提供扫描服务，也提供进一步的文档分类、解析、校对及审核等高技术增值功能，节省了大量人工的工作量。公司C端产品使得全球数亿用户的智能文字识别体验解除了时间地点的束缚，符合现代社会无纸化办公、移动办公的趋势，B端服务矩阵中的财报机器人、票据机器人、合同机器人等与客户业务场景紧密结合，助力客户降本增效，与国家“加快数字化发展、建设数字中国”的政策相匹配。

### 4) 智能文字识别技术为国家限制出口的关键技术之一，具有重大的国家战略意义

国家科技部、商务部发布的《中国禁止出口限制出口技术目录》已将“印刷体汉字识别技术、程序结构、主要算法和源程序”、“具有交互和自学习功能的脱机手写汉字识别系统及方法”、“汉字识别的特征抽取方法和实现文本切分技术的源程序”等纳入限制出口的信息处理技术范围。被列为国家限制出口的关键技术之一，意味着智能文字

识别技术是涉及我国核心利益、限制境外企业获取的先进技术。

2021年1月，时任美国总统特朗普签署行政命令，禁止美国司法管辖范围内的企业及个人与扫描全能王、WPS Office等8家中国应用软件进行交易，虽然该禁令没有实际执行、后被拜登撤销，但也代表公司的智能文字识别技术具有重大的国家战略意义。

### **5) 公司的智能文字识别技术促进无纸化办公，响应了节能减排的国家战略**

文档数字化不仅提高了用户的效率，无纸化办公也有助于减少碳排放量，响应了节能减排的国家战略。以扫描全能王为例，2022年扫描全能王中国地区用户总计将约65亿张纸质文档扫描并上传为电子化形式，假设用户后续使用主要以电子化形式分享、处理文档，不再纸质打印，一年65亿张电子化的文档后续可节约超过十万吨的碳排放量<sup>3</sup>，响应了国务院《2030年前碳达峰行动方案》中“加快实现生产生活方式绿色变革”的要求，符合国务院《“十四五”节能减排综合工作方案》中“营造绿色低碳社会风尚”的政策精神。

## **(2) 商业大数据业务**

根据工信部发布的《“十四五”大数据产业发展规划》，数据产业是以数据生成、采集、存储、加工、分析、服务为主的战略性新兴产业。商业大数据是大数据产业的重要领域，将企业的各类原始数据采集、清洗之后变为结构化信息，信息在挖掘后变为知识，知识再通过建模分析变为可用的数据资产，挖掘数据背后蕴藏的价值，提供风险预警、信用评级、供应链管理等附加价值。发行人商业大数据业务与国家科技创新战略和相关产业政策匹配，具体如下：

### **1) 国家政策重视大数据行业的技术创新，大数据已被纳入我国“十三五”与“十四五”纲要**

十三五纲要提出：“发展大数据新应用新业态；加强海量数据存储、数据清洗、数据分析挖掘、数据可视化等关键技术研发，形成一批具有国际竞争力的大数据处理、分析和可视化软硬件产品”。

十四五纲要提出：“推动大数据采集、清洗、存储、挖掘、分析、可视化算法等技术创新，培育数据采集、标注、存储、传输、管理、应用等全生命周期产业体系”。

<sup>3</sup> 按照中国绿色碳汇基金会官网的碳排放计算器，少打印一张纸可以节约17.5g碳排放量，65亿张纸质文档对应约113,750吨碳排放量

公司商业大数据业务利用大数据挖掘、知识图谱等技术，通过“数据→信息→知识→智能”4个层次，将企业多种实时动态的多源异构信息进行对齐、融合，对数据进行采集、清洗、标注与分析，构建了上亿的节点、数十亿条关系的知识图谱，属于十三五、十四五纲要中鼓励的大数据研发及发展方向。

## 2) 国家政策重视数据要素市场的培育、数据要素价值的利用及释放

2020年3月，中共中央、国务院发布《中共中央、国务院关于构建更加完善的要素市场化配置体制机制的意见》，将“数据”列入五大生产要素范围中，并提出从推进政府数据开放共享、提升社会数据资源价值、加强数据资源整合和安全保护三方面加快培育数据要素市场。

2022年12月，中共中央、国务院发布的《关于构建数据基础制度更好发挥数据要素作用的意见》中提出“加快发展数据要素市场，做大做强数据要素型企业；培育大数据交易市场，鼓励各类所有制企业参与要素交易平台建设，探索多种形式的数据交易模式”。

2023年2月，中共中央、国务院印发的《数字中国建设整体布局规划》提出“到2025年，数字基础设施高效联通，数据资源规模和质量加快提升，数据要素价值有效释放，数字经济发展质量效益大幅增强，政务数字化智能化水平明显提升，数字文化建设跃上新台阶”，强调“释放商业数据价值潜能”。

2023年3月，中共中央、国务院印发的《党和国家机构改革方案》提出“组建国家数据局。负责协调推进数据基础制度建设，统筹数据资源整合共享和开发利用，统筹推进数字中国、数字经济、数字社会规划和建设等，由国家发展和改革委员会管理”。

公司商业大数据业务利用大数据挖掘、知识图谱等技术，对数据要素的价值进行充分挖掘、有效释放。2021年，上海数据交易所正式成立，公司子公司上海生腾作为首批数商单位代表，与上海数据交易所完成签约，并逐步挂牌了多款数据产品，帮助建立高效便捷、合规安全的数据要素流通和交易体系。

2021年11月工信部发布的《“十四五”大数据产业发展规划》中强调“发挥大数据特性优势：1) 加快数据大体量汇聚；2) 强化数据多样性处理；3) 推动数据时效性流动；4) 加强数据高质量治理；5) 促进数据高价值转化”。公司的商业大数据业务积极响应前述政策要求：

“十四五”大数据产业发展规划的要求	公司的商业大数据业务积极响应政策要求
加快数据“大体量”汇聚	公司启信宝汇集中国境内 2.3 亿家企业等组织机构的超过 1,000 亿条实时动态商业大数据，提供包括工商、股权、司法涉诉、失信、舆情、资产等超过 1,000 个数据特征标签
强化数据“多样性”处理	公司自主研发了多源异构数据的融合与对齐技术、多源异构的超大规模动态知识图谱构建技术，将企业内部数据与外部的多源异构数据进行融合、对齐，构建出上亿的节点、数十亿条关系的超大规模的动态商业知识图谱，同时知识图谱具备自学习、自更新的功能
推动数据“时效性”流动	公司开发的 FISync 数据同步机制，在上游亿级数据发生更新时，能毫秒级匹配到需精准同步的数据；同时结合自研的数据顺序切片传输技术，保证在高并发高流量场景下具备较高的性能和稳定性，达成跨域异构数据库的近实时同步
加强数据“高质量”治理	公司通过优质合规的数据源的合作与筛选，通过数据清洗、NLP 结构化、算法校验等多个环节，围绕数据全生命周期，不断提升数据质量
促进数据“高价值”转化	公司将 NLP 技术与知识图谱技术相结合，挖掘出不同实体之间的显性和隐性关系，帮助客户发现隐藏在复杂网络之下的风险和隐性关联关系；同时针对银行、证券等行业客户痛点，融合行业特有数据，构建专有知识图谱，为金融风险管理、供应链企业全生命周期管控、政府园区企业监控等多种场景提供解决方案

### 3) 公司商业大数据业务响应了国家“促进大数据与各行业应用的深度融合”的政策

早在 2015 年 8 月，国务院发布《促进大数据发展行动纲要》中即提出“促进大数据与各行业应用的深度融合，形成一批代表性应用案例，以应用带动大数据技术和产品研发，形成面向各行业的成熟的大数据解决方案”。

2021 年 11 月，工信部发布的《“十四五”大数据产业发展规划》中强调“推动大数据与各行业各领域融合应用，加大对优秀应用解决方案的推广力度”。

2022 年 10 月，国务院办公厅印发的《全国一体化政务大数据体系建设指南》强调“推进政务数据开放共享、有效利用；促进数据高效流通使用，充分释放政务数据资源价值”。

2023 年 2 月，中共中央、国务院印发的《数字中国建设整体布局规划》强调“推动数字技术和实体经济深度融合，在农业、工业、金融、教育、医疗、交通、能源等重点领域，加快数字技术创新应用”。

公司商业大数据业务挖掘商业数据背后蕴藏的价值，为用户提供企业信息画像、关联关系查询、风险查询、营销决策分析等数据挖掘服务，赋能各行各业，已服务了政府、

产业园区、物流、制造、地产、零售、银行、证券、保险等近 30 个行业的众多客户。  
例如：

①公司为多个政府类客户开发了政务数字化解决方案。以上海市经信委一屏观产业决策分析系统项目为例，公司以政府部门内外部数据融合发掘为基础，进行辖区内产业链条颗粒化拆解、企业全景式画像描绘，以大数据动态展示辖区内产业健康度、政策落地效果等情况，通过构建企业异动预警模型等数据模型，对辖区内产业和企业的潜在风险进行动态监测，对辖区产业企业在区域贡献度、产业创新能力、企业迁入迁出分析等数据维度方面进行分析与追踪，助力政府进行精准施策、提升政务数字化智能化水平；

②制造型企业可利用启信宝商业大数据对上游供应商进行筛选，进行关联关系排查，避免围标、串标、陪标等问题，对准入的供应商做全方位的资格审查，在招标后持续对供应商进行舆情、经营风险监控。

因此，公司商业大数据业务充分响应了国家针对大数据与各行业应用的深度融合的号召。

**4) 公司子公司作为企业征信机构积极参与信用大数据的建设与服务，响应了国家“发展第三方大数据服务产业”、“推进社会信用体系建设”的号召**

2021 年 3 月发布的十四五纲要提出“加快推动数字产业化，发展第三方大数据服务产业”。

2021 年 11 月，工信部发布的《“十四五”大数据产业发展规划》中强调“发展第三方大数据服务产业；运用人工智能、自主学习等技术，构建信用大数据模型，提升信用风险智能识别、研判、分析和处理能力；深化信用信息在融资、授信、商务合作、公共服务等领域的应用”。

2022 年 3 月，中共中央、国务院发布的《关于推进社会信用体系建设高质量发展促进形成新发展格局的意见》提出“在确保安全前提下，各级有关部门以及公共信用服务机构依法开放数据，支持征信、评级、担保、保理、信用管理咨询等市场化信用服务机构发展”。

2022 年 12 月，中共中央、国务院发布的《关于构建数据基础制度更好发挥数据要素作用的意见》中提出“支持第三方机构、中介服务组织加强数据采集和质量评估标准制定，推动数据产品标准化，发展数据分析、数据服务等产业”。



公司商业大数据业务属于第三方大数据服务产业的一部分，且公司子公司上海生腾具备企业征信机构备案，上海生腾积极参与了信用大数据的建设与服务，自主开发出了一系列特色信用评价模型，包括启信分、合同履行指数、空壳指数等，响应了国家对加快发展第三方大数据服务产业、加快推进社会信用体系建设的要求。

### 5) 公司商业大数据业务响应了国家“筑牢数据安全保障防线”的要求

近年来，我国出台多项数据安全方面的法律法规及产业政策，提升了行业门槛，有利于头部规范企业。例如，2021年11月工信部发布的《“十四五”大数据产业发展规划》中强调“筑牢数据安全保障防线；鼓励行业、地方和企业推进数据分类分级管理、数据安全共享使用，开展数据安全能力成熟度评估、数据安全管理体系认证等”。

公司高度重视数据安全管理体系建设以及数据安全技术研发工作，采用智能数据分类分级、数据防泄漏（DLP）、网络安全态势感知等多种措施来保护系统和数据安全，坚持将数据安全治理作为一项体系化、常态化的工作持续开展。

2023年3月，公司子公司上海生腾成为首批中国信通院授予的“数据安全管理体系认证(DSMC)”企业。DSMC是由中国信通院面向全行业发起的数据安全领域的认证，根据数据安全制度落地要求、技术落地要求两大维度共计15项指标，全面评估企业的数据安全管理体系流程、组织机构、信息系统安全保障等多方面能力。该项认证体现了专业机构对合合信息在数据安全管理体系建设以及数据安全技术研发工作成果的认可。

针对上述事项，保荐机构采取的核查程序如下：

- (1) 检索了我国中央及各地政府对人工智能、大数据领域的政策文件；
- (2) 查阅了智能文字识别与商业大数据的相关行业研究报告；
- (3) 访谈公司核心技术人员，了解发行人是否符合科创板支持方向、科创行业领域以及科创属性要求，并获取了相关支持底稿。

综上，经核查，公司符合国家科技创新战略相关要求。

## 2、拥有关键核心技术等先进技术或产品情况

公司旗下主要产品及服务皆由公司自主研发，公司核心技术包括智能文字识别及商业大数据技术。其中智能文字识别技术融合了智能图像处理、复杂场景文字识别、NLP等AI技术，相比传统简单文字识别，具备更多认知与理解能力，可适应多语言、多版

式、多样式等复杂场景，并可应用到多个商业化场景中并形成落地的产品或服务，例如票据分类、证照票据结构化、合同关键信息抽取、智能审核等。商业大数据技术包括大数据挖掘与知识图谱等技术，通过“数据→信息→知识→智能”4个层次，挖掘商业数据背后蕴藏的价值，赋能各行各业。

报告期内核心技术相关业务收入占营业收入比例达 90%以上。截至 2022 年 12 月 31 日，公司及其控股子公司已取得发明专利授权合计 117 项，其中中国境内发明专利授权 78 项，境外发明专利 39 项；若将基于同一优先权申请在不同国家/地区获批、内容基本相同或相似的发明专利的视为同族专利，且同族专利仅记为 1 项，则中国境内外专利合计为 79 项。

公司智能图像处理、复杂场景文字识别、NLP、知识图谱及大数据挖掘等核心技术均处于行业领先水平，具体如下：

### **(1) 智能文字识别技术**

智能文字识别可以助推企业与个人实现生产、工作或者生活方式的数字化转型，具有较高的实际应用价值，是 AI 应用中商业推广落地较快的领域，是 AI 领域的重要分支。公司成立至今在智能文字识别领域已拥有超过 15 年的研发经验及技术积累，智能文字识别技术水平行业领先，识别准确率高，且识别精度可随着数据的积累、算法模型的深度学习优化不断提升。公司在智能文字识别领域的核心技术包括智能图像处理、复杂场景文字识别、NLP 技术，下图展示了三项核心技术之间的技术关系、解决的实际难题（部分示例）：



行业技术门槛、公司的技术创新如下表所示：

行业通用技术	含义	智能文字识别行业的技术门槛	公司的技术创新
机器学习与深度学习算法	机器学习是 AI 的一种底层算法,深度学习是机器学习的一个子集。机器学习与深度学习算法是 AI 领域应用最为广泛的算法之一,在文字识别、人脸识别、语音识别等多个 AI 领域均有广泛应用	技术门槛主要为: 1) 积累海量的有效场景数据以训练深度学习算法模型,从大量数据中进行多年的学习 2) 针对实际应用需求和各种不同的场景,设计和构建新的算法模型,并对算法模型参数进行不断调试,对模型架构进行不断优化,使得模型可在实际业务中的复杂场景下保持较高水平的性能,可部署在资源受限的移动端且保持较高的性能	<u>复杂文字识别场景的深度学习算法:</u> ✓ 公司在文字识别行业深耕超过15年,在深度学习模型的有效场景数据积累方面具有先发优势,并在数据的使用、生成、增广等方面进行创新 ✓ 公司针对智能文字识别中图像清晰度低、形变大、文字潦草、移动端部署难等难题,对算法模型进行多次迭代优化,主要包括:高效自定义算子实现;特定损失函数设计;在深度学习模型自主设计和构建等方面进行创新等。公司将前述技术创新应用到智能图像处理、复杂场景文字检测和识别、文档结构化理解等公司专有技术中,实现了业界领先的识别性能(准确率高、速度快、稳定性高、低存储)
OCR 技术	传统的 OCR 技术针对特定场景的文字图像进行建模,受限于手工设计特征的单一表达能力和复杂的处理流程,一旦跳出当前场景,遇到复杂场景,模型就会失效	技术门槛主要为: 1) 在各种复杂场景下(几何变形、曲面、背景干扰、角度倾斜、书写风格潦草多样的文字等)保持较高的文字检测与识别性能 2) 达到可商用的水平,与客户业务场景进行深度融合 3) 在资源受限的移动端的保持高准确率、高速、高稳定性、低存储的识别性能	<u>复杂场景的文字识别技术:</u> ✓ 针对文档图像中多种文字方向、文字重叠以及曲形印章中的文字识别等技术难题,公司研发了多项文字检测新技术,显著提升了复杂场景中任意四边形及不规则曲线文本的检测效果 ✓ 针对书写风格多样、笔迹潦草等手写体文字识别的技术难题,公司研发了基于边缘设备的深度学习量化加速压缩等技术,大幅提高了手写文本在复杂场景下的识别率,例如扫描全能王APP针对手写体文档字符平均识别率为97.00% <sup>4</sup> ✓ 公司复杂场景的文字识别技术成功应用在多个业务场景中,例如公司的身份证的条目识别率达99.6%,驾驶证的条目识别率达99.2%,增值税发票的条目识别率达97.7%,出租车票的条目识别率达98.5% <sup>5</sup> ,相关识别技术已成功应用在多个行业的证照审核、票据审核、发票报销等业务场景中 ✓ 公司解决了超大规模汉字识别模型的移动端应用技术难题,形成了高精度低存储的大类别(超过20,000类)文字识别方案,实现了在低资源、

<sup>4</sup>数据来源为中国合格评定国家认可委员会(China National Accreditation Service for Conformity Assessment 英文缩写为: CNAS)认可的检测实验室(NO.CNASL8192)2020年出具的一系列《软件测试报告》(以下简称为 CNAS 测试报告)

<sup>5</sup>数据来源为 CNAS 测试报告, 条目识别率=测试样本中识别正确的条目总数量/测试样本中的所有条目数量

行业通用技术	含义	智能文字识别行业的技术门槛	公司的技术创新
			性能低的移动设备下高准确率、高效率、高稳定性的识别体验，公司C端APP扫描全能王与名片全能王在全球拥有数亿用户
图像处理技术	用计算机对图像信息进行处理的技术，主要包括图像数字化、图像增强和复原、图像数据编码、图像分割和图像识别等。在文字识别、人脸识别、物体识别等多个 AI 领域也会使用到图像处理技术	技术门槛主要为： 1) 在实际业务中的复杂场景下保持较高的处理精度 2) 解决图像去模糊、形变图像矫正等技术难题	<u>智能图像处理技术：</u> ✓ 公司利用自主研发的文档图像的定位信息和透视变换技术、曲面文档图像校正技术，对形变图像进行还原矫正，使得用户在任何角度手机拍摄的文档图片都可以实现定位和校正，大幅提升复杂场景下文档的文字识别性能 ✓ 公司通过智能图像处理技术自动对文本图像实现去模糊并进行图像增强，使得文本图像的文字信息更加清晰和显著，大幅提升后续文字识别的准确率 ✓ 公司利用自主研发的文本图像拼接技术，在超大图像的识别和理解等场景下显著提升原有技术的性能
NLP 技术	是 AI 的一个子领域，用于分析、理解和生产自然语言，NLP 技术实现机器与人沟通的大前提，决定了机器对语言的理解能力。在文字识别、机器翻译、机器问答、内容推荐等应用场景及知识图谱领域均会使用到 NLP 技术	技术门槛主要为： 1) NLP 模型在各种复杂场景下（多语言、多样式、多版式、多行业的业务场景）保持较高的信息抽取与文字理解的准确率、速度及稳定性， 2) 解决文档关键信息抽取、文字纠错等技术难题	<u>复杂文字识别场景的 NLP 技术：</u> ✓ 公司通过对文本的类型、文档的整体拓扑结构信息、多个文本条目关联前后语义的联想等信息进行理解与分析，实现关键信息的定位、抽取和结构化输出，该技术已规模化运用在C端与B端不同场景下各类文档的关键信息抽取和结构化 ✓ 公司通过自主研发的基于NLP的文字纠错技术，对已识别的字符进行纠错，例如错字、前后文漏字和符号错误等情况，显著提升识别和理解的精准度

如何根据用户痛点及客户需求去分析处理文档面临大量的技术难点，基于复杂场景的客户需求，制定出适应性强的算法迭代框架和模型是极具壁垒的技术，算法工程师必须深刻理解场景的实质，根据需求调整模型的结构以及模型的训练和推理方式。公司在行业通用技术的基础上进行了大量的技术创新，包括自研算法模型的代码编写、模型设计与构建、模型训练、参数调节、场景验证等工作，智能图像处理、复杂场景文字识别、复杂文字识别场景的 NLP 技术是公司凭借多年技术研发活动中积累的自主研发的算法模型、C 端产品和 B 端服务开发中沉淀的专有技术，并在 C 端资源受限的移动端实现大规模应用，在 B 端不同行业的业务场景下实现商用级别的广泛应用。公司已在文字识别行业深耕超过 15 年，因此在数据积累、深度学习模型训练水平、各种复杂场景下的智能图像处理与文字识别技术具有先发优势和竞争壁垒。

公司智能文字识别技术所获得的主要奖项及成果如下：

公司早在 2012 年即被谷歌评为“全球顶尖开发者（Google Play Top Developer）”。

在 2019 年的国际文档分析识别大会（ICDAR）中，公司研发团队取得表格识别竞赛的冠军，并与华南理工大学团队联合取得票据识别竞赛的冠军，其中票据识别竞赛的识别对象是购物票据，包含票面质量模糊、格式不统一、票据局部损坏等真实商业场景会遇到的识别难题，表格识别竞赛是将工作中常用的表格作为识别对象，包含复杂的手写样本。ICDAR 票据识别及表格识别竞赛吸引了全球大量队伍参赛，既包括百度、华为、科大讯飞、平安科技、平安财险、华南理工大学、上海交大、华中科技大学等国内知名企业、科研机构，也包括其他众多全球知名团队，例如 Lenovo Research、韩国互联网集团 LINE 旗下的 Clova AI、美国财税软件集团 Intuit、德国罗斯托克大学的 CITlab、阿联酋的国家级人工智能研究院 IIAI（Inception Institute of Artificial Intelligence）、日本的人文开放数据中心、巴西的 Neural Mind 等多个国家团队。

在 2020 年国际模式识别旗舰会议（ICPR）中，公司研发团队与华南理工大学、联想研究院团队合作，在信息图表识别竞赛的 7 大任务 14 个子任务中获得 11 项第一。ICPR 是国际模式识别学会（International Association for Pattern Recognition, IAPR）召开的国际人工智能与模式识别领域的旗舰会议。

在 2023 年国际文档分析识别大会（ICDAR）的文本图像篡改检测算法竞赛（Competition On Detecting Tampered Text in Images）中，公司研发团队取得文本篡改

检测任务（Text Manipulation Detection）的冠军。该项竞赛聚焦真实场景下的文本图像篡改检测问题，面临篡改手段多样、篡改痕迹微弱、篡改文本与邻近文本高度相似等技术难点，同时文档版面的复杂化、文本内容的多样化亦加大了篡改检测实施的难度。

公司已研发了基于深度学习的图像篡改检测技术及相应的系统，融合了 SRM、BayarConv、ELA 等方法以提升 Tamper Detector 性能，综合 RGB 域、频域及噪声域存在痕迹的篡改，捕捉图像在篡改过程中留下的细微痕迹，可检测出擦除、重打印文本、复制、移动、拼接等多种篡改形式，并进行针对性的处理，在提高对篡改区域检测率的同时，防止误判，从而降低系统在实际落地应用中的成本。该技术有利于提升文档图像内容的安全性与可靠性，降低图像造假带来的潜在风险与损失，已在银行、保险和制造业等多种行业进行落地应用。

2023 年 5 月，公司加入中国信通院牵头的《文档图像篡改检测标准》制定工作，该项标准将确立文档图像篡改检测技术的行业标准、为 AI 视觉图像行业健康发展提供指引，由中国图象图形学学会、中国科学技术大学等业内知名机构联合编制。

根据中国科学院上海科技查新咨询中心 2020 年出具的《科技项目咨询报告》（编号 20201410），公司自主研发的“基于人工智能的复杂多场景文档图像识别与理解关键技术及应用”项目具有新颖性，国内外公开文献中未见与该项目技术特点完全相同的技术公开；经其分析，该项目综合技术达到国内领先水平，其中票据文字检测技术达到国际先进水平。

根据中国合格评定国家认可委员会（China National Accreditation Service for Conformity Assessment 英文缩写为：CNAS）认可的检测实验室（NO.CNASL8192）2020 年出具的一系列《软件测试报告》：（1）名片全能王 APP 针对常规的多语言名片的字符识别率平均值为 99.43%，名片识别对象包括简体中文、繁体中文、英文、日文、韩文及西欧语言；（2）扫描全能王 APP 针对常规的印刷体文档字符平均识别率为 99.77%，手写体文档字符平均识别率为 97.00%；（3）公司与其他文字识别企业识别率对比情况如下：

## 1) 以身份证识别为例, 国内几家头部文字识别企业对中文证照识别率的比较如下:

中文证照识别 条目识别率	合合 AI 开放 平台	百度 AI 开放 平台	腾讯 AI 开放 平台	阿里 AI 开放 平台	易道博识 <sup>6</sup> AI 开放平台
身份证	99.6%	98.7%	95.9%	82.1%	98.7%

注: 数据来源为前述 CNAS 测试报告; 条目识别率=测试样本中识别正确的条目总数量/测试样本中的所有条目数量

如上表可见, 公司智能文字识别技术在中文身份证照的条目识别率高于可比公司。条目识别率为文字识别领域常用的衡量识别效果的指标, 也是实际落地中客户会关注的指标之一。条目识别率可以理解为“行识别率”, 1 张文档包含多行条目, 1 个条目包含多个字符, 如 1 行字符中出现 1 个错误, 则该条目识别失败, 99.6%的条目识别率意味着对于 1000 行字符、有 996 行的识别完全正确, 仅 4 行出现识别错误。在客户的实际业务场景中, 容错率要求极高, 比如在财务/法务等业务场景中, 客户的预期是接近 100% 的识别率, 即使是 0.1%的条目错误率, 积累到整张文档的维度可能会导致几个百分点、甚至是十几个百分点的整张错误率, 会给客户的人工审核工作量带来成倍的提升。根据前述 CNAS 测试报告, 公司的身份证整张识别率为 97.5%, 处于较高水平 (即意味着对于 1000 张身份证样本, 有 975 张完全识别正确)。

## 2) 国际几家头部文字识别企业对复杂场景下文字识别率的比较如下:

复杂场景文字识别*	合合 AI 开放 平台	百度 AI 开放 平台	腾讯 AI 开放 平台	阿里 AI 开放 平台	Google Vision Cloud
字符识别率平均值	81.9%	70.0%	65.0%	66.9%	80.5%

注 1: 复杂场景文字识别指在多国语言、多版式、多曲面、包含干扰因素等情况下的文字识别

注 2: 数据来源为前述 CNAS 测试报告

注 3: 字符识别率的计算公式为  $F\text{-Score}=(2*Recall*Precision)/(Recall+Precision)$ , 其中  $Recall=$  识别正确的数量/真值的全部字符数量,  $Precision=$  识别正确的数量/识别出的总字符数

采用 F-Score 来计算的字符识别率指标在文字识别学术领域经常使用, 该指标既考虑了识别精度也兼顾了识别可信度, 在实际应用场景中, 客户对可信度的要求较高, 即要求 AI 产品可以自动判断识别率结果是否可以达到客户要求, 如不满足, 则自动返回提示, 可信度高的 AI 产品可以大幅减少客户的人工复核工作量。如上表可见, 公司智能文字识别技术在复杂场景下的字符识别率平均值领先国内外可比公司。由于实际商业

<sup>6</sup>北京易道博识科技有限公司, 以下简称为易道博识



化落地时业务场景均为包含多种干扰因素、个性化的复杂场景，客户在选择技术服务商时也会将识别精准度作为主要评选的标准之一，因此复杂场景下的识别精准度是决定文字识别企业市场竞争力的关键要素。

根据《科学技术成果鉴定证书（中图学鉴字[2021]第 001 号）》，中国图像图形学会综合考虑公司的扫描全能王、名片全能王、AI 开放平台的智能文字识别技术的技术创新点、技术水平、技术应用情况、社会认可程度，并对前述 CNAS 测试报告进行了复测，对测试报告中的手写和印刷体识别率、复杂场景下文字识别率等重点技术指标进行了鉴定，其认为，前述 CNAS 测试报告的测试结果客观真实，现场复测与测试报告结果一致，相关技术领先于国内外主流厂商的同类技术。

2022 年，CNAS 认可的检测实验室（NO.CNASL0286）上海计算机软件技术开发中心对公司开发的“扫描全能王”出具《软件测试报告》，报告对扫描全能王与 Adobe Scan、WPS、百度网盘、福昕 PDF、夸克 APP 的文字识别功能进行对比，具体情况如下：

不同场景样本的字符识别率	扫描全能王	Adobe Scan	WPS	夸克	百度网盘	福昕 PDF
普通文档数据集识别率	99%	89%	91%	98%	97%	95%
多语言数据集识别率	99%	79%	90%	93%	82%	80%
困难数据集识别率	95%	71%	79%	92%	93%	89%
手写数据集识别率	92%	12%	81%	87%	73%	37%
复杂场景数据集识别率	91%	21%	77%	83%	77%	65%

注 1：从大规模自行采集文字测试数据集中随机挑选部分办公文档场景下的图片样本

注 2：多语言数据集是指从大规模自行采集文字测试数据集中随机挑选英语、法语、德语、西班牙语等主流多语言场景下的图片样本

注 3：困难数据集指从大规模自行采集文字测试数据集中随机挑选出部分存在阴影、褶皱、模糊等识别比较困难的办公文档图片样本

注 4：手写数据集是指从大规模自行采集文字测试数据集中随机挑选中文手写场景下的图片样本

注 5：复杂场景数据集是指从大规模自行采集文字测试数据集中随机挑选复杂自然场景下的图片样本

注 6：字符识别率的计算公式为  $F\text{-Score}=(2*Recall*Precision)/Recall+Precision$ ，其中  $Recall=$  识别正确的数量/真值的全部字符数量， $Precision=$  识别正确的数量/识别出的总字符数。表格中所列的识别率为经过三轮测试后的平均值为最终识别率

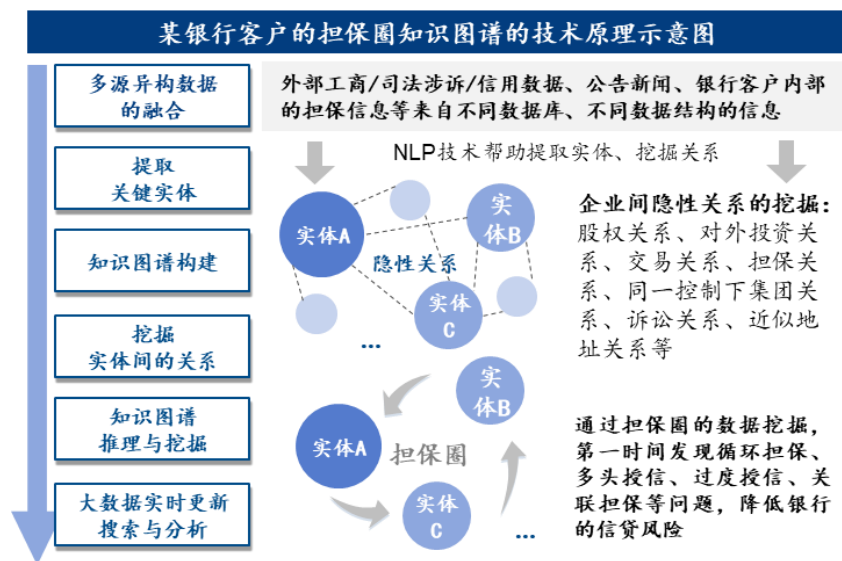
如上表所示，扫描全能王 APP 针对普通文档数据集和多语言数据集的识别率均为 99%，困难数据集识别率为 95%，手写数据集及复杂场景数据集识别率分别为 92% 和 91%，识别率水平均高于竞争性 APP。

2023 年 5 月，公司在中国信通院旗下中国泰尔实验室的 OCR 智能化服务评估和智能文档处理系统技术能力模块评估中获得最高评级：（1）在识别性能方面，公司的智能文字识别产品在证件类和票据类性能测试中的字段召回率分别为 96.80% 和 97.83%，字符准确率分别为 99.21% 和 99.59%，字段准确率分别为 97.87% 和 98.42%，覆盖旋转、阴影、反光、褶皱、形变、模糊、多语言、低像素、光照不均等复杂场景，各项指标处于行业领先水平；（2）在 AI 核心能力方面，公司的智能文字识别产品在 NLP 领域的文本纠错、文本分类以及知识图谱领域的实体识别、关系抽取等方面获得最高评级；（3）在安全性、可靠性方面，公司亦符合测试标准。

## （2）商业大数据技术

2015 年，公司开始布局大数据产业的商业大数据细分领域，利用前期积累的深度学习、NLP 等 AI 技术的积淀，拓展知识图谱及大数据挖掘技术。公司内部的大数据底层平台快速成熟化，2018 年，公司自主研发的企业知识图谱解决方案开始进行商业落地；2019 年，商业大数据量已突破 700 亿条，汇集中国境内 2.1 亿家企业等组织机构；2020 年，商业大数据量已突破 1,000 亿条，汇集中国境内 2.3 亿家企业等组织机构。

公司在商业大数据领域的核心技术为知识图谱、大数据挖掘、NLP 技术，下图以某银行客户的担保圈知识图谱为例，展示公司商业大数据几项核心技术的原理、技术关系、技术应用过程：



行业技术门槛、公司的技术创新如下表所示：

行业通用技术	含义	商业大数据行业的技术门槛	公司的技术创新
知识图谱	<p>知识图谱把所有不同种类的信息连接在一起而得到的一个关系网络，本质上是一种揭示实体之间关系的语义网络，提供了从“关系”的角度去分析问题的能力</p> <p>知识图谱是大数据行业的底层技术之一，同时也在数据治理、金融反欺诈、通用搜索与推荐等多个领域进行广泛应用，公司所聚焦的方向是基于商业大数据的知识图谱构建及推理</p>	<p>技术门槛主要为：</p>	<p><u>多源异构的超大规模动态知识图谱构建技术：</u></p> <ul style="list-style-type: none"> <li>✓ 公司将企业的工商、司法、信用、舆情等来自不同来源的非结构化数据中的实体进行对齐和融合，构建出上亿的节点、数十亿条关系的超大规模商业知识图谱</li> <li>✓ 公司针对公开的企业工商数据中的董监高人名信息进行消歧，解决企业知识图谱构建过程中的重名信息问题</li> <li>✓ 公司针对银行、证券等行业客户痛点，融合行业特有数据，构建专有知识图谱</li> </ul>
NLP 技术*	<p>如前所述，NLP 技术赋予了机器对语言的理解能力，技术难点主要在语言规律的不确定性、语言组合的多样性、语言对上下文的依赖性、一词多义等问题，在文字识别、机器翻译、机器问答、内容推荐等应用场景及知识图谱领域均有应用</p> <p>在商业大数据领域，知识图谱与 NLP 技术的区别是，知识图谱侧重于构建实体之间关系的语义网络，NLP 技术侧重于对语义的理解；二者经常结合运用，一方面，NLP 帮助知识图谱从非结构文本数据中抽取信息，实现文本理解、实体提取、关联分析以及数据标签化；另一方面，知识图谱提供的实体之间的关系网络信息，也可以帮助 NLP 模型在语义处理中得知此实体的额外语义信息，更精确的进行语义处理</p>	<p>1) 解决多源异构的数据对齐、消歧的技术难题，将海量的非结构化数据的结构化</p> <p>2) 建立包含海量实体、节点、关系的知识图谱</p> <p>3) 将外部数据与客户私有数据进行融合，构建针对于特定客户的专有知识图谱</p> <p>4) 根据客户需求深度挖掘数据背后的隐性关系与知识，为客户提供决策支持</p>	<p><u>超大规模知识图谱推理与挖掘技术：</u></p> <ul style="list-style-type: none"> <li>✓ 公司将NLP技术与知识图谱技术相结合，挖掘出不同实体之间的显性和隐性关系，例如企业间的十大隐性关系的挖掘：股权关系、对外投资关系、控制人关系、交易关系、担保关系、集团关系、诉讼关系、地址关系、事件关系、疑似关系</li> <li>✓ 公司将外部数据与客户私有数据进行联合建模，针对客户需求进行数据挖掘，例如对企业的股权信息与银行客户内部的担保数据进行联合分析，自动挖掘出集团关系、一致行动人、实际控制人、受益人等，解决潜在的循环担保风险</li> </ul>
大数据挖掘	<p>从海量数据中挖掘值得参考的样型或规则，转换成有价值的信息、洞察或知识，是大数据行业的底层技术之一</p>	<p>技术门槛主要为：</p> <p>对包含海量实体与关系的数据库保持较高的更新频率、搜索效率、数据挖掘的准确度及深度</p>	<p><u>商业大数据搜索与分析技术：</u></p> <ul style="list-style-type: none"> <li>✓ 公司利用相关技术将实时更新的多源异构数据动态的融合到已构建的知识图谱中，汇集中国境内2.3亿家的超过1,000亿条实时动态企业大数据，实现海量数据的动态更新</li> <li>✓ 可在海量商业大数据中进行毫秒级的实时搜索与分析</li> </ul>

\*注：智能文字识别、商业大数据领域均会使用 NLP 技术，原因是两个领域均会涉及大量的文字，而 NLP 技术是机器可以具有语言认知和理解能力的关键。智能文字识别领域对 NLP 技术的运用侧重于在文档图像转化成文本之后，对非结构化文本进行分类、信息抽取、语义分析等，进一步提高识别准确性与效率；商业大数据领域对 NLP 技术的运用侧重于对海量的多源异构数据进行融合与结构化处理，通过对文字信息的理解，更高效地提取关键实体、挖掘实体之间的关系。

公司在行业通用技术的基础上进行了大量的技术创新，包括自研算法模型的代码编写、模型训练、参数调节、场景验证等工作，商业大数据搜索与分析技术、多源异构的超大规模动态知识图谱构建技术、超大规模知识图谱推理与挖掘技术是公司凭借多年技术研发活动中积累的自主研发的算法模型、C 端产品和 B 端服务开发中沉淀的专有技术，在不同行业的业务场景下实现商用级别的广泛应用，在行业内具有领先性。

针对上述事项，保荐机构采取的核查程序如下：

(1) 访谈公司核心技术人员，了解公司核心技术及其功能性能、取得的研发进展及其成果、核心技术先进性与创新性，并获取了相关支持底稿；

(2) 核查了公司的主要奖项；

(3) 查阅了于国家知识产权局打印的分别以发行人及其控股子公司为权利人的专利清单；

(4) 查阅了中国信息通信研究院、中国人工智能产业发展联盟与腾讯集团联合发布的《智能文字识别（OCR）能力测评与应用白皮书》、灼识咨询出具的《中国商业大数据服务行业蓝皮书》等智能文字识别与商业大数据的相关行业研究报告。

综上，经核查，保荐机构认为发行人拥有和应用的关键核心技术具有先进性。

### **3、科技创新能力、科技成果转化能力情况**

#### **(1) 保持技术不断创新的机制安排**

##### **1) 公司建立了完善的研发管理体系**

公司建立了以产品为中心、以底层技术为支撑的研发机制。基于现有产品体系，公司的研发人员会不断进行产品性能的优化和更新，提升用户体验；同时，公司专注于智能文字识别、大数据挖掘、自然语言处理等技术的研发与创新，为上层产品提供底层技术支撑，针对各行业的多种应用场景，提供全方位的解决方案，并进行规模化示范应用。同时，公司也具有完善的知识产权的申请制度以保护研发成果。

## 2) 公司鼓励员工持续创新，并有完善的培训和激励机制

为鼓励员工持续创新，公司设立了完善的技术人才培养体系，以技术创新为目标，在公司内部跨部门实现技术方法论的沉淀及分享；配套相应的培训、考核与激励相结合的机制，获得公司技术人员的高度认可；并且为了更好的激励公司骨干技术成员，公司实施了员工股权激励。

公司每年会不定期对员工进行专利申请管理、产品研发、数据管理等方面的专业培训，启发员工不断地从研发过程中进行发明创造。公司现有研发体系、技术持续创新机制确保公司具备持续创新能力。

## 3) 公司重视研发方面的投入，为技术创新提供资金保障

2020年、2021年及2022年，公司研发费用分别为16,696.80万元、23,829.24万元和27,990.16万元。公司对研发人才的培养极为重视，截至2022年12月31日，公司全职工共912人，其中研发人员共499人，占总人数比重约为55%。

## 4) 公司的核心技术人员稳定，且具备较强的科研能力

报告期内，公司不存在核心技术人员变动情况。核心技术人员的研发能力及研发贡献情况如下：

镇立新，毕业于中国科学院自动化研究所，模式识别与智能系统博士。现任公司董事长、总经理及法定代表人。镇立新带领公司不断探索行业尖端技术与创新方向，其领导公司研发团队完成名片全能王、扫描全能王系列产品的研究开发，奠定了公司智能文字识别核心技术的基石，并组建了启信宝研发团队，开拓了公司在大数据领域的新技术研发方向与业务发展方向。镇立新深入参与技术创新，在2005年，获得华尔街日报亚洲年度发明金奖2020年，被亿欧智库入选《2020全球科学企业家研究报告暨G50&200人物榜单》中的“全球科学企业家200人物榜单”及2020世界创新者年会的“2020中国科学企业家30人”榜单。2021年，荣获中共上海市委组织部评选的“上海领军人才”。截至2022年底，其共发表10篇在智能文字识别领域的论文，其作为发明人授权的发明专利55项，其中国际发明专利23项，其作为第一发明人的授权发明专利33项。

龙腾，毕业于华南理工大学，通信与信息系统博士。现任合合信息董事、副总经理、AI技术负责人。龙腾领导公司研发团队完成了公司网络后台架构的建设，使之可支撑公司C端产品全球亿级月活用户；其领导研发了基于深度学习的AI底层技术平台，独

立完成行业领先的移动端图像扫描核心技术研发，为名片全能王、扫描全能王 APP 核心技术水平奠定行业领先的地位。截至 2022 年底，其作为发明人授权的发明专利 51 项，其中国际发明专利 22 项，其作为第一发明人的授权发明专利 19 项，并发表学术论文 9 篇，其中包括 1 篇发表于国际模式识别领域顶级期刊《Pattern Recognition》的论文。其于 2018 年参与起草了国家市场监督管理总局、中国国家标准化管理委员会发布的国家标准“信息技术—紧缩嵌入式摄像头通用规范”（GB/T 36480-2018）。

丁凯，毕业于华南理工大学，通信与信息系统博士。现任合合信息自然语言算法研发总监。丁凯作为核心研发人员主导和参与了公司智能文字识别、自然语言处理、知识图谱等核心技术的研发；其主导了服务于三星集团、招商银行等多个标杆 B 端客户服务项目的人工智能算法的研发；2019 年，其带领公司研发团队和华南理工大学机器视觉与深度学习实验室合作，取得了国际文档分析识别大会（ICDAR）票据检测竞赛的冠军。截至 2022 年底，其作为发明人授权的发明专利 8 项，发表学术论文 19 篇。

郭丰俊，毕业于上海交通大学，模式识别与智能系统博士。现任合合信息图像算法研发总监。郭丰俊领导公司研发团队完成了针对银行卡的智能文字识别系统的研发，作为第一代银行卡识别系统，被包括三星集团在内的多家跨国企业采购并应用于其主流产品；其领导研发了针对表格的智能文字识别系统，在 2019 年国际文档分析识别大会（ICDAR）的表格识别竞赛中获得冠军；其领导研发的基于深度学习的文字识别项目，完成了包括书籍扫描、图像质量提高在内的多项系统升级。截至 2022 年底，其作为发明人授权的发明专利 11 项，发表学术论文 10 篇。

陈青山，毕业于上海交通大学，控制理论与控制工程硕士。现任合合信息副总经理、董事、大数据技术负责人。陈青山领导创建了公司移动产品开发团队，带领移动开发团队获得“Google 顶尖开发者”称号；带领团队完成名片全能王、扫描全能王、启信宝 3 款 C 端产品的研发和运营工作；组建合合信息大数据资产和研发团队，创建了公司的数据资产中台、数据治理技术中台和数据安全体系；建立公司大数据企业级解决方案实施交付团队，将公司大数据企业解决方案在企业征信、金融风控、智能营销和政企数字化治理等领域实现落地。2020 年，陈青山入选 2020 世界创新者年会评选的“2020 中国人工智能企业家 30 人”。目前，陈青山是中国移动应用程序安全委员会的副主任委员。截至 2022 年底，其作为发明人授权的技术发明专利 27 项，其中作为第一发明人的授权技术发明专利 5 项。

张彬，毕业于华南理工大学，通信与信息系统硕士。现任合合信息工程算法研发总监。张彬负责公司名片全能王早期版本的研发，深入参与公司网络后台架构的建设，使之可支撑公司 C 端产品全球亿级月活用户；其长期参与基于深度学习的 OCR 引擎的研发，负责公司 B 端智能文字识别创新产品票据机器人的研发，推动基于深度学习的 OCR 引擎的性能升级。截至 2022 年底，其作为发明人授权的发明专利 12 项，其中国际发明专利 6 项。

## (2) 公司的技术储备情况

公司自设立以来即在智能文字识别、商业大数据领域不断深耕，目前已经形成全方位、多层次、相交融的技术储备格局。公司的技术储备情况详见公司进行中的研发项目情况。截至 2022 年 12 月 31 日，公司正在进行中的主要研发项目情况如下：

序号	项目名称	所处阶段及进展情况	项目负责人	项目整体预算(万元)	报告期研发投入(万元)	拟达到的目标	与行业技术水平的比较
1	基于深度学习的版面还原技术在移动设备的应用	技术/产品研发阶段	曹超阳	1,830	1,762	基于深度学习对移动设备扫描的图像内容进行识别，去除图像的背景干扰，保留文字内容和图像内容，并智能还原格式信息，让用户真正体验到所见即所识别。	处于同业领先水平
2	基于智能文字识别技术的数字化转型平台	技术/产品研发阶段	龙腾	6,500	5,825	升级对各种文档图像上的文字检测和识别的关键技术，涉及各类文档的拍照数字化及自动智能文字识别，重点提高复杂场景下的字符识别率，解决电子文档还原等复杂问题。	处于同业领先水平
3	基于社会化大数据的深度挖掘在赋能城市经济治理数字化转型的应用	技术/产品研发阶段	陈青山	1,300	524	通过对产业数据的分析挖掘，自动生成宏观产业地图、宏观经济运行分析系统等各种主题的业务分析系统，以可视化图形的形式，为政府、产业园区等 B 端客户提供多角度、可定制的数据洞察工具。	处于同业领先水平
4	建设电子信息领域标准大数据公共服务平台项目	技术/产品研发阶段	康龙	750	845	研发产业链应用情况分析与决策支撑系统，形成产业链基本数据库、产业链图谱工具及数据库、产业链标准应用情况分析工具及图谱展示工具、产业链标准应用情况数据库。	处于同业领先水平
5	扫描全能王的 PDF 文档服务扩展矩阵	技术/产品研发阶段	马小捷	580	408	开发扫描全能王的 PDF 文档服务扩展矩阵，为扫描全能王用户在扫描创建文档后的提供阅读、编辑文档等增值服务，打造优质的	处于同业领先水平



序号	项目名称	所处阶段及进展情况	项目负责人	项目整体预算(万元)	报告期研发投入(万元)	拟达到的目标	与行业技术水平的比较
						PDF 格式转换效果。	
6	基于标准图谱服务建设的基础项目-2 期	技术/产品研发阶段	谢小河	500	162	进一步升级标准知识图谱技术，进行通用基础功能设计开发，以便在未来进行知识图谱项目 B 端服务开发时能够高效、快速、准确得完成项目开发工作。	处于同业领先水平
7	区域产业大脑项目	技术/产品研发阶段	郝莉莉	1,500	237	利用大数据和知识图谱技术打造以数据驱动的“区域产业大脑”，包含区域经济监测模块、产业运行分析模块、区域产业招商模块，为区域政府提供科学的经济形势预测、分析企业画像的能力，帮助区域留存优势企业、吸引优质企业入驻、优化营商环境和提前识别和预警企业异常和风险，实现政府对区域内企业的精准监管	处于同业领先水平

### (3) 先进技术产品商业化情况、科技成果转化能力

凭借领先的自主研发技术、成熟的产品落地能力，公司的 3 款 C 端 APP 产品覆盖了全球百余个国家和地区的亿级用户，品牌知名度与用户体验俱佳。截至 2022 年 12 月末，公司扫描全能王、名片全能王、启信宝 3 款 APP 在 App Store 与 Google Play 应用市场的全球用户累计首次下载量合计超过 8 亿，2022 年 12 月的月活合计超过 1.3 亿。

公司 B 端服务已在银行、证券、保险、政府、物流、制造、地产、零售等近 30 个行业实现成熟应用。公司的智能文字识别服务、商业大数据服务帮助客户切实解决了降本增效、改善风控、高效获客的业务痛点，提供了较高的技术附加值。如下列表格示例：

下游行业	客户痛点	公司智能文字识别技术的产业赋能
政府	<ul style="list-style-type: none"> <li>业务处理流程繁杂、所需时间较长，用户体验待优化</li> <li>机构间通过纸质文件办公，协同效率较低</li> </ul>	<ul style="list-style-type: none"> <li>自动识别、录入各类证件、申请表格等，并与政府机构的核查系统、申请系统等连接，赋能政务自动化转型</li> <li>无纸化办公，提升各机构间协同效率</li> <li>数字化用户体验，提升办事效率，提高用户满意度</li> </ul>
制造业	<ul style="list-style-type: none"> <li>信息化程度较低，大量文档数据未有效管理与利用</li> <li>上下游涉及企业较多，缺乏有效协同运作体系</li> </ul>	<ul style="list-style-type: none"> <li>自动识别、录入各类合同、票据等文件</li> <li>自动识别、录入各类识别编码、日期等信息，赋能设备、生产原材料管理等</li> <li>数字化企业运营，上下游协同工作、流程自动化</li> </ul>

下游行业	客户痛点	公司智能文字识别技术的产业赋能
供应链管理	<ul style="list-style-type: none"> <li>•竞争加剧，利润承压，急需降低成本，提升效率</li> <li>•运单版式多样、经手频次高、人工效率低且易出错</li> </ul>	<ul style="list-style-type: none"> <li>•自动识别不同版式的快递单据，并将运单信息高效录入系统，完成核验</li> <li>•自动识别快递信息，进行智能分拣，优化流程</li> </ul>
地产	<ul style="list-style-type: none"> <li>•业务处理流程繁杂，用户体验较差</li> <li>•上下游涉及企业较多，缺乏有效协同运作体系</li> </ul>	<ul style="list-style-type: none"> <li>•自动识别、录入用户的各类证件，包括房产证、不动产权证等，并自动将信息录入至对应系统，完成核验</li> <li>•数字化用户服务，提升办事效率，提高用户满意度</li> </ul>
银行、证券等金融业	<ul style="list-style-type: none"> <li>•大量业务流程仍依赖员工或者用户的人工操作，效率较低，用户体验亟待优化</li> <li>•存在海量未电子化的文档数据，未能打通各业务环节文档数据，实现数据价值</li> </ul>	<ul style="list-style-type: none"> <li>•多终端文档、图像信息的自动采集，批量扫描识别多种非标准化文档</li> <li>•各类表单、文档及工单等文档的智能录入、分类、核验、解析，简化业务流程，降本增效</li> <li>•实现自动核保、远程开户等商业化应用,数字化用户体验，提高用户满意度</li> </ul>

下游行业	客户痛点	公司商业大数据技术的产业赋能
政府	<ul style="list-style-type: none"> <li>•掌握多源社会数据，但政府各部门间数据未打通，数据价值亟待挖掘，同时决策依赖经验</li> <li>•产业园区招商产业化难度大</li> </ul>	<ul style="list-style-type: none"> <li>•打破数据孤岛，实现不同地域及部门数据融合，提升各机构间协同效率，赋能管理决策，助力城市管理</li> <li>•为区域政府提供科学的经济形势预测、分析企业画像的能力，帮助区域留存优势企业、吸引优质企业入驻、优化营商环境和提前识别和预警企业异常和风险，实现政府对区域内企业的精准监管</li> </ul>
制造业	<ul style="list-style-type: none"> <li>•数据量大，且数据源繁杂，但数据标准度不高、数据应用难、数据价值未完全挖掘</li> <li>•供应链管理难度大</li> </ul>	<ul style="list-style-type: none"> <li>•赋能供应链管理，优化供应商寻源及准入制度，进行关联关系排查，避免围标、串标、陪标等问题，对准入的供应商做全方位的资格审查，在招标后持续对供应商进行舆情、经营风险监控</li> </ul>
供应链管理	<ul style="list-style-type: none"> <li>•供应链复杂，但信息化程度较低，上下游信息不对称</li> <li>•竞争加剧，利润承压，急需降低成本，提升效率</li> </ul>	<ul style="list-style-type: none"> <li>•赋能供应链协同管理，通过上下游企业的信用大数据分析降低物流企业交易成本，降低经营风险</li> <li>•优化区域产业布局，促进产业链供应链集群化发展，降低物流成本，提升产业链效率</li> </ul>
地产	<ul style="list-style-type: none"> <li>•购地信息、楼盘信息、工程项目信息等非结构化数据繁杂，数据未有效管理，无法进行价值挖掘</li> <li>•上下游信息不对称，风控难度大</li> </ul>	<ul style="list-style-type: none"> <li>•对区域经济发展趋势、环境数据、市场供需等进行大数据分析，评估地产项目可开拓性，精准选址</li> <li>•通过大数据技术协助地产供应链企业资质审核、存量供应商定期尽调，及时发现和控制各类风险，确保交付质量</li> </ul>
银行、证券等金融业	<ul style="list-style-type: none"> <li>•上下游信息不对称，风控难度大，依赖风控人员的经验，尽调成本高</li> <li>•数据量大，但数据未打通、数据质量低、数据价值未完全挖掘</li> </ul>	<ul style="list-style-type: none"> <li>•通过大数据技术赋能银行等机构挖掘优质企业作为信贷支持对象，尤其是科创企业、绿色企业等，支持金融赋能政策鼓励的实体经济发展</li> <li>•监控目标企业的各类动态信息，构建信息分析模型，实现风险实时监测和控制</li> <li>•为企业资质审查、信贷准入等决策提供依据，提高风控的有效性</li> </ul>

报告期内，公司核心技术相关收入占各期营业收入比例分别为 91.68%、90.16% 和 91.44%，核心技术相关收入的绝对值保持高速增长，2020-2022 年的年复合增速达

30.58%，具体情况如下表所示：

核心技术相关业务产生的收入金额（万元）	2022 年度	2021 年度	2020 年度	与核心技术的对应关系
智能文字识别 B 端服务	7,120.43	6,782.50	5,412.40	复杂场景文字识别、智能图像处理、NLP 技术
智能文字识别 C 端 APP（扫描全能王和名片全能王）	65,368.81	51,208.79	35,011.01	
商业大数据 B 端服务	9,759.56	6,821.29	5,028.77	大数据挖掘、知识图谱、NLP 技术
商业大数据 C 端 APP（启信宝）	7,635.28	7,315.43	6,758.80	
手机厂商技术授权	504.84	517.29	803.32	复杂场景文字识别、智能图像处理、NLP 技术
合计	<b>90,388.92</b>	<b>72,645.30</b>	<b>53,014.30</b>	-

针对上述事项，保荐机构采取的核查程序如下：

（1）访谈公司核心技术人员，了解公司核心技术应用形成产品的商业化情况、核心技术人员的科研能力、保持技术不断创新的机制安排和技术储备等，并获取了相关支持底稿；

（2）获取了公司的核心技术人员调查表、研发人员清单；

（3）获取了公司的在研项目清单，并了解各项目的研发所处阶段、目标应用领域；

（4）核查了公司的收入明细表及关键客户合同，访谈主要客户，了解公司核心技术相关业务取得的收入情况、客户对公司服务所提供技术价值的认可情况等。

综上，经核查，保荐机构认为发行人具有突出的科技创新能力及科技成果转化能力。

#### 4、行业地位或者市场认可度情况

##### （1）细分行业领域的排名情况

2021 年初，公司在 App Annie “2021 Top publisher Awards” 中与 Google、Microsoft、Zoom Video Communications 及 Adobe 被评选为全球前五大商业与效率应用企业（Top 5 publishers of Business & Productivity Apps）。

公司智能文字识别技术为核心的 C 端产品扫描全能王与名片全能王在全球具有一定的市场影响力。截至 2022 年 12 月 31 日，扫描全能王曾在 App Store 上 105 个国家和

地区（含中国）的效率类免费应用下载量排行榜位列第一，名片全能王曾在 App Store 上 43 个国家和地区商务类免费应用下载量排行榜位列第一。在 App Store 可以公开检索到的与公司扫描全能王与名片全能王属于同类别的 APP，包括 ABBYY FineScanner、ABBYY Business Card Reader、Google Lens 等，根据 App Annie 数据，2020 年、2021 年以及 2022 年，名片全能王与扫描全能王的活跃用户、用户下载量之和远大于同类 APP，iOS 评分也均高于同类 APP。

公司商业大数据技术为核心的 C 端产品启信宝在国内具有一定的市场影响力。根据灼识咨询《中国商业大数据服务行业蓝皮书》，商业大数据 C 端 APP 市场集中度较高，国内已形成三足鼎立的局面，头部三家分别为启信宝、企查查、天眼查，2022 年启信宝为中国商业信息查询市场排名第三位的手机 APP。

## （2）公司的市场认可程度

C 端业务方面，公司 APP 品牌知名度与用户口碑俱佳，尤其是名片全能王、扫描全能王在 2009 年、2010 年上线，在品牌影响力方面具有超过十年的深厚积累：早在 2010 年，名片全能王即被纽约时报（The New York Times）称为“为低效的名片管理带来有效的秩序”；在 2013 年，扫描全能王被时代周刊（Time）评选为“iPhone 必备的 50 个应用之一”、被华盛顿邮报（The Washington Post）报道为“手掌扫描仪”；在 2013 年，名片全能王被 BBC 报道为“更智能化的名片管理”；名片全能王在 2012-2016 年被上海市软件协会连续评为“年度优秀软件产品”；在 2017 年，扫描全能王被 App Annie 评为“一带一路最佳 APP”。根据 App Store 数据，扫描全能王付费版已被 App Store 精品推荐超过 660 次；名片全能王已被 App Store 精品推荐 440 多次。2020 年，启信宝获得上海大数据联盟颁发的“数字赋能先锋奖”；2022 年，启信宝获得上海大数据交易中心主办的 2022 全球数商大会授予的“年度数据产品奖”。从 App store 的搜索关键词来看，截至 2022 年 12 月，“扫描全能王”的搜索热门程度高于“扫描”或者“扫描文档”，说明在用户心智中扫描全能王一定程度上已成为扫描的代名词。

B 端服务方面，公司头部客户资源丰富，已积累了建设银行、中国银行、三星、联想、亚马逊、中国人寿保险、中国太平洋保险等超过 125 家世界 500 强客户。基于领先的核心技术、强大的技术落地能力以及深刻的行业理解，公司在下游多元化行业获得了多项关键成绩：

下游应用领域	时间	关键成绩
供应链管理	2022 年	荣获“最佳供应链（数智化）风控奖”（SZ&W Group 及 SSCL 主办的亚太智慧供应链与物流创新峰会&精品展）
	2021 年	荣获“最佳供应链金融实践案例”（由《贸易金融》、知钛财经与北京财资和供应链应用技术研究院，共同发起第 5 届中国供应链金融“行业标兵大奖”）
	2020 年	合合信息的“供应链大数据风控管理平台”获“金科杯”年度最佳风控平台（由上海市经信委软件和信息服务业处指导，Top 智汇主办的 2020 数字金融峰会）
零售	2019 年	2019 年度最佳新零售数据化解决方案（上海连锁经营协会与 WRE World Retail Elite 联合举办的 2019 第四届 WRE 零售数字转型峰会）
政务	2021 年	合合信息智慧招商解决方案荣获 2021 年中国数字化转型与创新评选之智慧城市创新解决方案（数字产业创新研究中心）
	2020 年	工信部官网、人民日报等多家主流网站/媒体报道了“启信宝抗疫物资查询平台”
制造	2021 年	合合信息的智能解决方案获得中国轻工业信息中心评选的“2021 年中国轻工业数字化杰出解决方案”
地产	2021 年	合合信息与时代地产的合作案例“合合信息智能 OCR 解决方案”被《互联网周刊》评选为“2020 人工智能案例 TOP100”

针对上述事项，保荐机构采取的核查程序如下：

（1）访谈公司核心技术人员，了解公司在细分行业领域的排名情况、市场认可程度等，并获取了相关支持底稿；

（2）核查了公司的主要奖项；

（3）查阅了中国信息通信研究院、中国人工智能产业发展联盟与腾讯集团联合发布的《智能文字识别（OCR）能力测评与应用白皮书》、灼识咨询出具的《中国商业大数据服务行业蓝皮书》等智能文字识别与商业大数据的相关行业研究报告。

经核查，保荐机构认为发行人在所属细分行业领域具有突出的行业地位，市场认可度高。

综上，经核查，保荐机构认为：发行人符合《上海证券交易所科创板企业发行上市申报及推荐暂行规定（2022 年 12 月修订）》（以下简称“《申报及推荐暂行规定》”）第三条规定的科创板支持方向。

## （二）发行人符合科创板行业领域的核查情况

根据《申报及推荐暂行规定》，公司的智能文字识别、图像处理、自然语言处理、知识图谱、大数据挖掘等人工智能及大数据核心技术是“新一代信息技术”，因此公司属于“新一代信息技术”领域。

所处领域	《战略性新兴产业重点产品和服务指导目录（2016）》	《战略性新兴产业分类（2018）》	公司核心技术	公司产品/服务	公司业务内容
人工智能(AI)	“1、新一代信息技术产业”之“1.5.2 人工智能软件”	“1、新一代信息技术产业”之“1.5.1 人工智能软件开发”	智能文字识别技术（复杂场景文字识别、智能图像处理 NLP 技术）	C 端 APP	扫描全能王 APP：智能文字扫描及识别软件，可将复杂场景下的文档转变为扫描仪效果的 PDF 文件或图片，并进一步识别为文本，实现文档资产的便捷管理 名片全能王 APP：智能名片及人脉管理软件，可将复杂场景下的纸质名片转变为电子联系人，实现名片资产的便捷管理
				B 端基础技术服务	证照、银行卡、名片、通用文本、表格等智能文字识别的基础技术服务
				B 端标准化服务	名片全能王企业版 SaaS 软件服务
				B 端场景化解决方案	财报机器人、票据机器人、合同机器人、智能配置机器人等与业务场景紧密结合的文档智能识别机器人 智能文字识别 AI 训练平台
大数据	“1、新一代信息技术产业”之“1.2.3 大数据服务”	“1、新一代信息技术产业”之“1.4.3 云计算与大数据服务”	商业大数据技术（大数据挖掘、知识图谱、NLP 技术）	C 端 APP	启信宝 APP：企业商业信息查询平台，汇集中国境内 2.3 亿家企业的超过 1,000 亿条实时动态数据，并提供多种数据挖掘分析报告
				B 端基础数据服务	数据包服务 数据 API 服务
				B 端标准化服务	启信宝企业版 SaaS 软件服务
				B 端场景化解决方案	企业知识图谱解决方案：包括金融风险知识图谱解决方案、供应链大数据风控平台、企业信息实时分析大屏 启信宝商业全景数据库

以公司主要产品作为行业分类依据，根据国家发改委《战略性新兴产业重点产品和服务指导目录（2016）》，公司的智能文字识别业务中扫描全能王 APP、名片全能王 APP 及智能文字识别 B 端企业服务属于“1、新一代信息技术产业”之“1.5.2 人工智能软件”；公司的商业大数据业务中启信宝 APP 及商业大数据 B 端企业服务属于“1、新一代信息技术产业”之“1.2.3 大数据服务”。

根据国家统计局《战略性新兴产业分类（2018）》，公司的智能文字识别业务中扫描全能王 APP、名片全能王 APP 及智能文字识别 B 端企业服务属于“1、新一代信息技术产业”之“1.5.1 人工智能软件开发”；公司的商业大数据业务中启信宝 APP 及商业大数据 B 端企业服务属于“1、新一代信息技术产业”之“1.4.3 云计算与大数据服务”。

根据中国证监会颁布的《上市公司行业分类指引（2012 年修订）》<sup>7</sup>，属于“I65 软件和信息技术服务业”；根据《国民经济行业分类（GBT 4754-2017）》，属于“I65 软件和信息技术服务业”。

公司的可比公司包括主营业务或底层技术相似的软件和信息技术服务企业，如金山办公、福昕软件等。前述可比公司均在招股书中披露了其根据中国证监会颁布的《上市公司行业分类指引（2012 年修订）》，属于“I65 软件和信息技术服务业”；根据《国民经济行业分类（GBT 4754-2017）》，属于“I65 软件和信息技术服务业”。其中金山办公、福昕软件在 2020 年 9 月之前完成上市，因此其招股书未披露归属于《上海证券交易所科创板企业上市推荐指引》、《战略性新兴产业重点产品和服务指导目录（2016）》、《战略性新兴产业分类（2018）》的具体分类。根据《上海证券交易所科创板企业发行上市申报及推荐暂行规定》第四条，公司属于“新一代信息技术”领域中的“人工智能、大数据、云计算、软件、互联网”领域，符合科创板行业定位。

上述可比公司与发行人的行业领域归类不存在显著差异。公司其他可比公司尚未在 A 股市场上市。

针对上述事项，保荐机构采取的核查程序如下：

（1）获取了可比公司的招股说明书等公开披露文件，分析其主营业务、核心技术及所处领域，并与公司进行比较；

（2）访谈公司核心技术人员，了解公司主营业务、核心技术及其应用领域。

因此，保荐机构认为，发行人属于《申报及推荐暂行规定》第四条“新一代信息技术领域，主要包括半导体和集成电路、电子信息、下一代信息网络、人工智能、大数据、云计算、软件、互联网、物联网和智能硬件等”重点推荐领域的企业。发行人主营业务与所属行业领域归类相匹配，与可比企业的行业领域归类不存在显著差异。发行人属于

<sup>7</sup>中国证监会于 2022 年 8 月 12 日发布了《关于废止部分证券期货规范性文件的决定（2022）》，其中现行《上市公司行业分类指引（2012 年修订）》已被废止，将由新的规则予以规范。截至本上市保荐书出具之日，中国证监会暂未发布新的相关指引。

科技创新行业领域，行业领域归类准确。

### （三）发行人符合科创属性相关指标要求的核查情况

科创属性评价标准一	是否符合	指标情况
最近三年累计研发投入占最近三年累计营业收入比例 $\geq 5\%$ ，或最近三年累计研发投入金额 $\geq 6,000$ 万元	是	公司最近三年累计研发投入（不含股份支付）占最近三年累计营业收入比例为 28.88%，最近三年累计研发投入金额为 68,516.20 万元。
研发人员占当年员工总数 $\geq 10\%$	是	截至 2022 年 12 月 31 日，公司全职员工共 912 人，其中研发人员共 499 人，占总人数比重约为 55%。
应用于公司主营业务的发明专利 $\geq 5$ 项	是	截至 2022 年 12 月 31 日，公司发明专利 117 项中应用于主营业务的为 89 项；若将基于同一优先权申请在不同国家/地区获批、内容基本相同或相似的发明专利的视为同族专利，且同族专利仅记为 1 项，则中国境内外专利合计为 79 项，其中应用于主营业务的为 55 项。
最近三年营业收入复合增长率 $\geq 20\%$ ，或最近一年营业收入金额 $\geq 3$ 亿	是	公司最近三年营业收入复合增长率为 30.74%，最近一年营业收入金额为 98,846.18 万元。

注：相关指标的计算基础和计算方法为：

最近 3 年累计研发投入占最近 3 年累计营业收入比例=最近 3 年累计研发投入/最近 3 年累计营业收入；

研发人员占当年员工总数的比例=研发人员数量/员工总数；

应用于公司主营业务的发明专利：该专利是否应用于公司主营业务；

最近 3 年营业收入复合增长率=（报告期最后一年营业收入/报告期第一年营业收入）<sup>^0.5-1</sup>。

#### 1、对发行人最近三年累计研发投入的核查情况

单位：万元

项目	2022 年度	2021 年度	2020 年度
研发费用	27,990.16	23,829.24	16,696.80
营业收入	98,846.18	80,578.16	57,824.64
研发费用占营业收入比例	28.32%	29.57%	28.87%

2020 年、2021 年和 2022 年，发行人研发投入分别为 16,696.80 万元和 23,829.24 万元和 27,990.16 万元，营业收入分别 57,824.64 万元和 80,578.16 万元和 98,846.18 万元，占营业收入比例分别为 28.87%、29.57%和 28.32%。

最近三年，发行人累计研发投入金额为 68,516.20 万元，占最近三年累计营业收入比例为 28.88%，且金额大于 6,000 万元。



公司主要依据部门职能，将承担研发工作的部门人员的薪酬归集为研发费用，报告期内由财务部、人事部以及研发部门三方来共同认定部门职能对应的费用类别，归集方法合理。

**研发内控制度：**公司制定了《研发项目管理制度》、《合合信息专利制度》、《产品开发与设计流程合规管理规范》、《差旅申请及报销制度》、《采购管理制度》、《会计核算制度》、《财务月结工作管理制度》等，对研发过程管理和研发费用归集核算等流程制度化、规范化，明确研发费用的归集范围及核算程序，以保证研发项目达到预定的开发目标及研发费用的准确归集与划定。报告期内，公司就研发项目设立台账进行核算，对研发项目进度和费用按照公司制度进行管理。

**研发支出开支范围和标准：**根据公司的研发相关制度，公司可以归集为研发费用的内容包括：研发活动直接消耗的材料，直接从事研发活动的本企业在职人员工资、薪金、津贴、补贴、奖金、五险一金，专门用于研发活动的有关折旧费、租赁费、相关固定资产的运行维护、维修等费用，用于研发活动的软件、专利权、非专利技术等无形资产的摊销费用，研发成果的论证、评审、验收、评估以及知识产权的申请费、注册费、代理费等费用，通过外包、合作研发等方式，委托其他单位、个人或与之合作进行研发而支付的费用，与研发活动直接相关的其他费用。

公司研发人员存在部分人员兼职参与其他工作的情况，报告期内，公司存在部分研发人员为客户定制化开发提供技术支持，主要集中在互联网广告推广服务、手机厂商技术授权、智能文字识别 B 端业务、商业大数据 B 端业务。报告期内公司将研发人员薪酬划分到成本的业务场景具体如下：

(1) 互联网广告推广服务：涉及到广告请求及展示逻辑、广告运营投放等业务内容，由研发人员给予成本部门支持，进行相关开发工作。

(2) 手机授权预装业务：根据各家手机厂商新上市机型和新版操作系统，由研发人员跟进适配及测试反馈 bug 的修复。

(3) 智能文字识别 B 端场景化解决方案：特定国家的身份证识别、特定单据的识别等业务，由研发人员给予成本部门支持，为该等客户进行定制化开发工作。

(4) 商业大数据 B 端场景化解决方案：数据同步软件的数据库落地交付及数据更新等业务的全景数据库实施工作，由研发人员给予成本部门支持，为该类客户提供数据

定制化服务。

上述情况下，兼职研发人员会将其工时打卡至成本项目工时，对应成本项目工时的费用会归集在营业成本中。同时也存在其他非研发人员与研发相关的部分工时费用计入研发费用的情形，公司根据实际工时投入情况来进行费用拆分。报告期内，公司兼职人员相关的研发费用情况如下：

单位：万元

项目	2022 年度	2021 年度	2020 年度
公司营业收入总金额	98,846.18	80,578.16	57,824.64
公司研发费用总金额	27,990.16	23,829.24	16,696.80
研发费用率	28.32%	29.57%	28.87%
兼职人员按工时归集至研发费用工资总额	5,983.37	3,458.77	523.47
扣除兼职人员归集至研发费用的工资总额之后的研发费用金额	22,006.79	20,370.48	16,173.33
扣除兼职人员归集至研发费用的工资总额之后的研发费用率	22.26%	25.28%	27.97%

注：以上兼职人员是指兼职研发人员、参与研发项目的非研发人员。

由上表可知，2020 年，兼职人员对研发费用率的影响比例较低，对研发费用的整体影响较小。2021 年至 2022 年兼职人员增加主要系：（1）兼职研发人员主要为推动 B 端各业务的产品研发迭代和参与 B 端场景化解决方案的项目落地；（2）随着定制化项目的增加，公司持续投入研发人员与实施人员共同参与客户开发项目，为了提升未来项目落地效率，公司从个性化项目中提取共性需求，以便未来对客户的类似需求进行及时响应。综上，公司的实际业务需求变化，使得研发人员和实施人员兼职的情况增加，对应的薪酬费用上升，因此变动合理。

研发支出的内控流程：公司的研发支出涉及的审批程序为向供应商采购的审批流程和研发费用报销流程。供应商采购流程中，由具有采购需求的人员提出申请并经研发部门负责人审批。审批通过后，提交采购负责人进行供应商筛选与评估，并由采购负责人审核商务条款。研发部门负责人确定最终供应商并签订合同。研发部门在供应商提供完产品或服务时，将结算单或验收单按研发项目进行归集后提交至财务部入账。研发部门在 OA 系统按研发项目提交付款申请，经研发部门负责人及财务部审批后，由财务部付款并入账。如果是预算外的采购或采购金额超过 3 万的，增加 CEO 复核流程。研发费

用报销流程中，由报销人在 OA 系统按研发项目提交报销申请，经由研发部门负责人及财务部审批，审批通过后，财务部付款并入账。

针对上述事项，保荐机构采取的核查程序如下：

(1) 取得并查阅公司研发相关内控制度，了解公司研发支出内部控制流程；

(2) 询问财务人员，了解研发支出归集和核算方法；取得并查阅研发费用科目设置情况，确认研发支出的归集和核算方法，判断其合理性，核查是否存在将研发不相关的支出计入研发费用的情况；

(3) 取得并查阅公司员工花名册及工资表、工时表，核查研发人员的划分口径是否合理、准确，核查存在兼职人员的情况，是否存在高级管理人员兼职研发的情况，研发人员工资如何分配到项目以及是否存在研发人员薪酬划分不准确的情况；获取报告期内各年度工时表、项目表的全量数据字段，统计分析研发项目工时的分布情况；

(4) 取得并查阅研发支出相关入账依据。对前二十大采购供应商进行抽凭核查，核查与研发支出相关的合同、发票、付款单据等原始凭证是否真实、完整，核查是否按照研发开支用途、性质列支，并计入准确的明细科目；核查是否存在将与研发无关的费用在研发支出中核算的情况；

(5) 对研发费用中的办公经费、交通差旅费、业务招待费等其他费用进行随机抽凭核查，核查相关款项是否均与研发活动直接相关；

(6) 取得并查阅公司各年研发费用加计扣除的鉴证报告、所得税汇算清缴报告，与账面研发投入进行核对分析。

经核查，保荐机构认为：发行人已建立与研发项目相对应的管理机制和研发支出审批程序；发行人已明确研发支出开支范围和标准，并得到有效执行；发行人严格按照研发开支用途、性质据实列支研发支出，不存在将研发无关的费用在研发支出中核算的情况。报告期内，发行人的研发投入真实、研发投入金额归集准确。

## 2、对发行人研发人员的核查情况

截至 2022 年 12 月 31 日，公司全职员工共 912 人，其中研发人员共 499 人，占总人数比重约为 55%。研发人员主要是算法研发、数据挖掘、产品开发、产品测试等方面的技术人才。

保荐机构对发行人的研发内控制度及其执行情况、研发人员的认定等进行了核查，核查手段包括查阅研发内控制度、公司员工花名册及研发人员名单，核查研发项目立项文件中的人员名单，检查研发人员岗位是否有变动，对研发人员进行访谈等。

经核查，保荐机构认为：报告期内，发行人研发人员划分准确，不存在将其他人员转划入研发人员的情况。

### 3、对发行人发明专利数量的核查情况

截至 2022 年 12 月末，公司及其控股子公司已取得发明专利授权合计 117 项，其中中国境内发明专利授权 78 项，境外发明专利授权 39 项；若将基于同一优先权申请在不同国家/地区获批、内容基本相同或相似的发明专利的视为同族专利，且同族专利仅记为 1 项，则中国境内外专利合计为 79 项。发行人的发明专利具体情况详见《招股说明书》“第五节 业务和技术”之“五、与发行人业务相关的资产情况”之“（二）主要无形资产”。截至 2022 年 12 月末，公司与核心技术相关的发明专利共 58 个，均运用在公司主营业务中，同时公司将非核心技术专利中一些信息传输、手写笔迹处理等领域的专利运用在主营业务中，共计 31 个，合计共 89 个专利应用于主营业务；若同族专利仅记为 1 项，则应用于主营业务的专利合计为 55 项。

与核心技术相关的 58 个发明专利具体情况详见下列表格：

核心技术	序号	所属国家	专利号（申请号）	状态	专利名称
用于嵌入式设备的高性能低存储的文字检测和识别技术	1*	中国	200910195252.3	授权	多触点字符输入方法及系统
	2*	美国	US8743058	授权	MULTI-CONTACT CHARACTER INPUT METHOD AND SYSTEM
	3*	美国	US10248878	授权	CHARACTER INPUT METHOD AND SYSTEM AS WELL AS ELECTRONIC DEVICE AND KEYBOARD THEREOF
	4*	欧洲	EP2306270	授权	CHARACTER INPUT METHOD AND SYSTEM, ELECTRONIC DEVICE AND KEYBOARD THEREOF
	5*	中国香港	HK1138921	授权	字符輸入方法及系統、電子設備及其鍵盤
	6*	韩国	10-1366723	授权	다중 점점 문자 입력 방법 및 시스템
	7*	韩国	10-1264897	授权	문자 입력 방법 및 시스템, 전자 장치 및 이의 키패드
	8*	中国	200910055150.1	授权	字符输入方法及系统、电子设备及其键盘

核心技术	序号	所属国家	专利号（申请号）	状态	专利名称
基于深度学习的多语言文字检测和识别技术	9	中国	200810042489.3	授权	一种通过名片识别技术在地址簿中自动生成并添加图标的方法
	10	中国	200610118618.3	授权	一种用带数码相机的手机实现名片扫描的方法
	11	中国	201210516663.X	授权	方便电子化的专业笔记本及其电子化文档的自动分类方法
	12	中国	201210517167.6	授权	一种自动切分电子化笔记本中手写条目的方法
手写文字检测和识别技术	13*	中国	201010583601.1	授权	一种重叠书写的手写输入方法
	14*	中国	201010583582.2	授权	多字符连续书写的手写输入方法
	15	中国	200910055524.X	授权	手写输入系统及方法、电子设备
	16*	美国	US9323456	授权	MULTI-CHARACTER CONTINUOUS HANDWRITING INPUT METHOD
	17*	欧洲	EP2650766	授权	MULTI-CHARACTER CONTINUOUS HANDWRITING INPUT METHOD
	18*	美国	US9052755	授权	OVERLAPPED HANDWRITING INPUT METHOD
	19*	日本	5854442	授权	複数キャラクターを連続筆記する手書き入力方法
	20*	日本	5784141	授权	重叠筆記による手書き入力方法
	21*	中国香港	HK1184571	授权	多字符連續書寫的手寫輸入方法
	22*	韩国	10-1802876	授权	다문자 연속 쓰기의 핸드라이팅 입력방법
23*	韩国	10-1825154	授权	중첩 쓰기 핸드라이팅 입력방법	
基于机器学习的名片图像匹配技术	24	中国	201010225606.7	授权	基于图像匹配和网络查询的识别方法及系统
	25*	美国	US9298708	授权	BUSINESS CARD INFORMATION EXCHANGE METHOD COMBINING CHARACTER RECOGNITION AND IMAGE MATCHING
	26	中国	201010225609.0	授权	通过网络查询自动校正识别结果的方法及系统
	27*	美国	US9147109	授权	METHOD FOR ADDING BUSINESS CARD INFORMATION INTO CONTACT LIST
	28*	中国	201110187278.0	授权	判断联系人列表中是否存在欲添加名片的方法

核心技术	序号	所属国家	专利号（申请号）	状态	专利名称
	29*	日本	5863962	授权	追加を希望する名刺が連絡先リストに存在しているか否かを判断する方法
	30*	日本	5757366	授权	文字認識と画像照合を組み合わせて名刺情報を交換する方法
	31*	韩国	10-1805731	授权	문자 식별과 이미지 매칭을 결합시켜 명함 정보를 교환하는 방법
文档图像的检测、定位与校正技术	32	中国	201310330784.X	授权	用于智能无线通信终端的四边形边框识别方法及装置
	33	中国	201310338246.5	授权	在图像中线段吸附的方法及装置，构造多边形的方法及装置
	34	中国	201110226374.1	授权	图像检测方法
	35*	中国	201310357997.1	授权	直线/线段吸附的方法及装置，构造多边形的方法及装置
	36	中国	201310462071.9	授权	用于智能无线通信终端的识别四边形边框的方法及装置
	37*	中国	201010583973.4	授权	一种文本图像的切边方法
	38	中国	201310746613.5	授权	一种图像方向校正方法及系统
	39	中国	201010129600.X	授权	调整图像四边形框检测结果的方法
	40	中国	200910054970.9	授权	图像旋转矫正方法及系统、电子设备
	41	中国	201210461924.2	授权	卡片扫描图像的切分方法
	42*	美国	US9082192	授权	TEXT IMAGE TRIMMING METHOD
	43*	欧洲	EP2650821	授权	TEXT IMAGE TRIMMING METHOD
	44*	日本	6193498	授权	直線 / 線分の吸着方法及び装置、多角形の構築方法及び装置
	45*	美国	US10332236	授权	METHOD AND APPARATUS FOR ADSORBING STRAIGHT LINE/LINE SEGMENT, METHOD AND APPARATUS FOR CONSTRUCTING POLYGON
	46*	欧洲	EP3035289	授权	METHOD AND DEVICE FOR ADSORBING STRAIGHT LINE/LINE SEGMENT, AND METHOD AND DEVICE FOR CONSTRUCTING POLYGON
	47*	日本	5896245	授权	テキスト画像のトリミング方法
48*	韩国	10-1784919	授权	텍스트 이미지의 트리밍 방법	
49*	韩国	10-1825200	授权	직선/선분의 흡착방법 및 장치, 다변형의 구축방법 및 장치	

核心技术	序号	所属国家	专利号（申请号）	状态	专利名称
	50*	中国香港	HK1184580	授权	一種文本圖像的切邊方法
文本图像拼接技术	51	中国	201010558888.2	授权	用于文本图像拼接的拍摄方法
	52	中国	201010558868.5	授权	拍摄多幅文本图像并拼接的方法
	53	中国	201010558966.9	授权	一种文本图像的拼接方法及系统
智能图像去模糊及增强处理技术	54	中国	201010558948.0	授权	提高文本图像清晰度的方法及系统
基于视频的文档图像自动检测技术	55	中国	200910056762.2	授权	相机自动聚焦方法及系统
	56	中国	202010526508.0	授权	一种基于光流法的视频关键帧提取方法及系统
基于 NLP 的文档关键信息抽取与结构化技术	57	中国	201010548188.5	授权	自动提取即时通讯账号并作出相应操作的方法
超大规模知识图谱推理与挖掘技术	58	中国	201910342464.3	授权	一种企业集团关系获取方法及系统

\*注：第 1/2/6 号专利为同族专利；第 3/4/5/7/8 号专利为同族专利；第 13/18/20/23 号专利为同族专利；第 14/16/17/19/21/22 号专利为同族专利；第 25/30/31 号专利为同族专利；第 27/28/29 号专利为同族专利；第 35/44/45/46/49 号专利为同族专利；第 37/42/43/47/48/50 号专利为同族专利

31 个属于非核心技术、但运用在公司主营业务中的专利如下表所示：

专利名称	专利号（申请号）	国家/地区	状态
BLUETOOTH COMMUNICATION METHOD AND BLUETOOTH COMMUNICATION DEVICE	US9848285	美国	授权
METHOD AND SYSTEM INTEGRATING GEOGRAPHICAL LOCATION INFORMATION AND BLUETOOTH TECHNOLOGY FOR RELAYING ELECTRONIC BUSINESS CARD	US9008652	美国	授权
METHOD FOR OBTAINING INFORMATION PROVIDED BY CONTACTS VIA NETWORK AFTER INPUTTING CONTACT INFORMATION	EP2613483	欧洲	授权
BLUETOOTH COMMUNICATION METHOD AND BLUETOOTH COMMUNICATION DEVICE	EP3016296	欧洲	授权
地理位置情報とブルートゥース技術を組み合わせた電子名刺伝達方法及びシステム	5830732	日本	授权
지리적 위치 정보와 블루투스 기술을 결합시킨 전자 명함 전달 방법 및 시스템	10-1886241	韩国	授权
輸入聯繫信息後通過網絡獲取聯繫人提供的信息的方法	HK1181567	中国香港	授权
手写区域显示信息的方法及系统	200910055919.X	中国境内	授权
输入联系信息后通过网络获取联系人提供的信息的方法	201010271280.1	中国境内	授权

专利名称	专利号（申请号）	国家/地区	状态
网络帐户建立连接的方法及其网络终端设备、云端设备	201410489050	中国境内	授权
一种利用电子邮件发送及接收名片信息的方法	201110376644.7	中国境内	授权
获取纸质笔记本中内容的方法及装置	201210496657.2	中国境内	授权
实时手写笔迹的美化方法及美化装置	201210513797.6	中国境内	授权
对手写笔迹进行编辑处理的方法及电子设备	201210512978.7	中国境内	授权
实时手写笔迹的美化方法及电子设备	201210513019.7	中国境内	授权
互换信息的方法及系统	201110422783.9	中国境内	授权
自动开关摄像单元的方法及系统	201010593232.4	中国境内	授权
一种利用网络查询名片传递历史信息的方法	201110433644.6	中国境内	授权
手写输入时自动滚动电子墨水的方法	201010146627.X	中国境内	授权
一种结合相机和光学定位的扫描方法	200810037226.3	中国境内	授权
移动设备之间的配对方法及系统	201210165031.3	中国境内	授权
基于网络交换个人相关信息的方法	201110143342.5	中国境内	授权
结合地理位置信息和蓝牙技术的电子名片传递方法及系统	201110214509.2	中国境内	授权
一种结合蓝牙无线技术的电子名片传递方法及系统	201110214846.1	中国境内	授权
结合蓝牙技术和蓝牙设备地址的电子名片传递方法及系统	201110214524.7	中国境内	授权
高速节点自组网中节点安全加入方法	201210374378.9	中国境内	授权
面向高速节点自组网的三级树形结构令牌双簇首分簇方法	201210374437.2	中国境内	授权
高速节点自组网中节点安全退出方法	201210404454.6	中国境内	授权
一种实现撤销重做的手势操作方法及装置	202110533007.X	中国境内	授权
一种用于撤销重做的手势操作方法及装置	202110533106.8	中国境内	授权
一种实现复制粘贴的手势操作方法及装置	202110532850.6	中国境内	授权

对于公司的发明专利具体情况，保荐机构执行了以下核查程序：

①查阅了于国家知识产权局打印的分别以发行人及其控股子公司为权利人的专利清单；

②查阅发行人及其控股子公司目前持有的专利证书，并检索国家知识产权局中国及多国专利审查信息查询系统（<http://cpquery.sipo.gov.cn/>），确认相关专利的权利归属及剩余期限；



③查询中国裁判文书网、全国法院失信被执行人名单信息公布与查询网、全国法院被执行人信息查询网公示信息，公开检索公司对外的诉讼、纠纷；

④访谈公司相关负责人员，了解公司在行业内的获得专利的情况；

⑤访谈公司相关负责人员，了解公司近年来的科研成果，对业务流程中核心技术的运用及业务收入与发明专利的匹配情况等方式进行核查；

⑥查阅了发行人境外专利代理机构出具的关于境外专利情况的确认函以及境外律师出具的法律意见书；

经核查，保荐机构认为：发行人应用于主营业务收入的发明专利数量真实、准确。

#### 4、对营业收入增长的核查情况

2020年度、2021年及2022年，发行人营业收入分别为57,824.64万元、80,578.16万元及98,846.18万元，2020-2022年年均复合增长率为30.74%，大于20%，且最近一年营业收入金额大于3亿元。报告期营业收入的构成如下：

单位：万元

项目		2022年度		2021年度		2020年度	
		金额	占比	金额	占比	金额	占比
智能文字识别	B端服务	7,120.43	7.20%	6,782.50	8.42%	5,412.40	9.36%
	C端APP-扫描全能王	63,287.86	64.03%	49,325.74	61.21%	33,343.21	57.66%
	C端APP-名片全能王	2,080.95	2.11%	1,883.05	2.34%	1,667.80	2.88%
	小计	72,489.24	73.34%	57,991.29	71.97%	40,423.41	69.91%
商业大数据	B端服务	9,759.56	9.87%	6,821.29	8.47%	5,028.77	8.70%
	C端APP-启信宝	7,635.28	7.72%	7,315.43	9.08%	6,758.80	11.69%
	小计	17,394.85	17.60%	14,136.72	17.54%	11,787.57	20.39%
互联网广告推广		8,224.26	8.32%	7,823.82	9.71%	4,155.71	7.19%
手机厂商技术授权		504.84	0.51%	517.29	0.64%	803.32	1.39%
其他业务		233.00	0.24%	109.05	0.14%	654.63	1.13%
合计		<b>98,846.18</b>	<b>100.00%</b>	<b>80,578.16</b>	<b>100.00%</b>	<b>57,824.64</b>	<b>100.00%</b>

公司报告期内收入的增长主要来源于智能文字识别C端APP业务。智能文字识别C端APP业务在报告期内分别实现收入35,011.01万元、51,208.79万元和65,368.81万元，占相应期间公司营业收入总额的比例分别为60.55%、63.55%和66.13%，2021年度、

2022 年度分别同比增长 46.26%、27.65%，占比持续提升。

从移动互联网行业、公司业务角度分析公司智能文字识别 C 端 APP 收入快速发展的原因如下：

### **(1) 移动互联网行业发展角度**

#### **1) 用户付费意愿的提升**

2015 年至今，用户对各类 APP 的付费转化率逐年提升，代表用户对可为其带来价值的 APP 会员付费产品接受度快速提升，尤其是在 2016-2017 年，娱乐类 APP 对用户付费方面的教育培养也推动了办公工具类型 APP 付费意愿的提升。

#### **2) 付费方式的多元化、便捷性提升**

随着网购的兴起推动了移动支付软件的崛起，Appstore 于 2016 年开始接入支付宝，2017 年开始接入微信支付，提升了用户付费的便捷度。

### **(2) 公司业务角度**

智能文字识别 C 端 APP 收入在报告期内高速增长，主要系：扫描全能王和名片全能王切实为用户带来无纸化移动办公的效率提升，由于产品本身的用户群体广泛，不仅财会、行政、销售等商务场景对其具有应用需求，教师、学生、家庭等生活场景也对各类文档具有电子化管理的需求，协助用户实现个人文档资产的数字化管理，其应用场景的可延展性强，用户群体不断拓宽，公司在庞大的用户群体中逐步变现。公司在积累用户数量的过程中，利用技术提升、服务升级等办法逐步引入增值服务，实现用户付费的自然转化，付费用户数量持续增加。报告期内，名片全能王和扫描全能王的新增付费用户数量合计分别为 100.86 万名、108.53 万名和 115.79 万名。

针对报告期内发行人营业收入增长情况，保荐机构主要履行了以下核查程序：

#### **1) 针对 C 端收入，执行的主要核查程序包括：**

①对业务系统执行一般控制和应用控制测试：执行一般控制测试主要包括系统开发变更管理、系统安全和访问控制管理、系统运维管理；执行应用控制测试范围主要包括数据输入、数据处理、数据输出、数据接口，确认用户注册信息被后台系统准确记录，用户购买商品信息被后台系统准确记录，用户退订信息被后台准确记录，测试用户订单能够有效记录；

②执行用户量核查：通过执行用户增量趋势分析、用户注册时间段分析、注册用户 IP 地址集中度分析、用户行为数据分析、IT 运维成本的匹配分析对用户量和用户行为进行核查，确认是否存在虚增用户、刷单等情形；

③执行订单数据一致性核查：检验业务后台与第三方支付账单（包含中国境内支付、Appstore 支付、Google pay 支付）一致性，确认公司业务系统记录的用户订单信息的真实性、准确性和完整性；

④执行第三方支付账号与用户 ID 关系核查：汇总报告期内各年度单支付账号的金额，查看支付最高金额是否合理，对用户支付金额进行分层，并检查合理性；统计报告期内各年度使用 1 个支付账号为 2 个及以上 user\_id 进行支付的账号数量并判断其合理性；统计给测试订单付款的支付账号在报告期内的总支付金额并判断合理性；

⑤执行付费用户及订单数据分析：从 C 端产品各类型付费订单数量、用户的地域分布、购买金额、购买频次、新增付费转化率、续费率、付费用户的活跃情况等多维度对付费用户及订单量的合理性进行分析；

⑥对用户采取电话、APP 推送等方式进行核查，确认用户购买产品/服务的真实性；

⑦对 C 端收入的增长驱动因素进行分析，查阅了移动互联网 APP 代表性企业公开披露的付费转化率、移动支付行业相关研究报告，从移动互联网整体行业发展、公司 APP 收入增长的驱动因素及商业合理性等维度进行论证。

**2) 针对 B 端收入，执行的主要核查程序包括大额核查、截止性测试、穿行测试、走访、函证和第三方回款等核查程序：**

①针对大额核查，获取销售合同、交付邮件、验收资料/对账单/结算单、销售发票、回款单据等，对收入的真实性进行核查；根据销售合同、发货邮件、验收资料、对账单等，重新计算收入确认金额与递延金额，对收入确认的准确性进行核查；

②针对截止性测试，获取销售合同、交付邮件、验收资料/对账单/结算单或回款单据、销售发票等，对收入确认时点进行核查，核实是否存在收入跨期。由于公司从交付产品到确认收入通常在在一个月內可以完成，因此截止性测试的样本范围为收入确认时间在报告期各资产负债表日前后 1 个月的单笔 5 万元以上的收入；

③针对穿行测试，了解各产品的具体服务内容及相关内控流程，获取合同审批 OA

流程、销售合同、交付邮件、验收资料、开票申请 OA 流程、销售发票、回款单据等各业务流程相关资料，检查 CRM 管理系统，核查企业销售与收款业务流程的内部控制设计是否有效并得到执行；

④针对走访，对报告期内各期前 20 大客户及随机抽取的小额客户进行了走访或电话访谈，取得了盖章签字的访谈记录、工商档案或互联网工商检索资料，与发行人无关联关系的承诺函等材料；

⑤针对函证，针对与客户在报告期各期内发生的项目销售额、项目收款金额、项目验收时间（如适用）、项目维保期间（如适用）等进行确认。对于回函不符或未回函的客户，通过与交易对方沟通，确认回函不符的具体原因后，进行再次检查销售合同、验收资料/对账单/结算单、回款单据等作为替代程序，以核实收入的发生和收入确认的准确性。

经核查，保荐机构认为：发行人营业收入快速增长真实且具有合理性。

#### （四）关于发行人符合科创板定位的结论性意见

经充分核查和综合判断，本保荐机构认为发行人出具的专项说明和披露的科创属性信息真实、准确、完整，发行人符合科创板支持方向、科技创新行业领域和相关指标等科创板定位要求。

## 八、保荐机构对发行人是否符合上市条件的说明

（一）发行人符合《首发管理办法》、《上海证券交易所科创板股票上市规则》规定的发行上市条件

### 1、本次证券发行符合《首发管理办法》第十条的规定

本机构核查了发行人设立至今相关的政府批准文件、营业执照、公司章程、发起人协议、创立大会文件、评估报告、审计报告、验资报告、工商设立及变更登记文件、股本变动涉及的增资协议、股权变动涉及的股权转让协议、主要资产权属证明、相关董事会和股东大会决议，董事会、监事会和股东大会议事规则和相关会议文件资料、董事会专门委员会议事规则、独立董事工作制度、董事会秘书工作制度、总经理工作细则等文件，向主要董事、监事、高级管理人员进行了访谈，并向发行人律师、审计师和评估师

进行了专项咨询和会议讨论。核查结论如下：

合合有限成立于 2006 年 8 月 8 日，2020 年 6 月 18 日，合合有限按照经审计的账面净资产进行股改，因此发行人持续经营时间超过三年以上。发行人自设立以来，未出现法律、法规及发行人《公司章程》规定可能导致发行人终止的情况。发行人按照相关要求建立了规范的法人治理结构，具备健全且运行良好的组织架构，相关机构和人员能够依法履行职责。

发行人符合《首发管理办法》第十条的规定：“发行人是依法设立且持续经营三年以上的股份有限公司，具备健全且运行良好的组织机构，相关机构和人员能够依法履行职责。有限责任公司按原账面净资产值折股整体变更为股份有限公司的，持续经营时间可以从有限责任公司成立之日起计算。”

## 2、本次证券发行符合《首发管理办法》第十一条的规定

本机构对经审计的财务报告和经审核的内部控制鉴证报告以及其他相关财务资料进行了审慎核查；就发行人报告期内收入构成变动、主要产品价格变动和销量变化、财务指标和比率变化，与同期相关行业、市场和可比公司情况进行了对比分析；查阅了报告期内重大购销合同、应收应付款项相关资料、长期待摊费用及构成情况、主要税种纳税资料以及税收优惠或财政补贴资料，就发行人财务会计问题，本机构与发行人财务人员和审计师进行密切沟通，并召开了多次专题会议等，核查结论如下：

众华对公司最近三年及一期的财务报告进行了审计，并出具了标准无保留意见的众会字（2023）第 02665 号《审计报告》，意见如下：“我们审计了上海合合信息科技股份有限公司（以下简称合合信息）财务报表，包括 2020 年 12 月 31 日、2021 年 12 月 31 日及 2022 年 12 月 31 日的合并及公司资产负债表，2020 年度、2021 年度及 2022 年度的合并及公司利润表、合并及公司所有者权益变动表和合并及公司现金流量表以及相关财务报表附注。我们认为，后附的财务报表在所有重大方面按照企业会计准则的规定编制，公允反映了合合信息 2020 年 12 月 31 日、2021 年 12 月 31 日及 2022 年 12 月 31 日的合并及公司财务状况以及 2020 年度、2021 年度及 2022 年度的合并及公司经营成果和现金流量。”

根据众华出具的《上海合合信息科技股份有限公司内部控制鉴证报告》（众会字（2023）第 02666 号）、发行人出具的《上海合合信息科技股份有限公司内部控制自我

评价报告》，并经本机构审慎核查，发行人的内部控制在所有重大方面是有效的，并由注册会计师出具了无保留结论的内部控制鉴证报告。

发行人符合《首发管理办法》第十一条的规定：“发行人会计基础工作规范，财务报表的编制和披露符合企业会计准则和相关信息披露规则的规定，在所有重大方面公允地反映了发行人的财务状况、经营成果和现金流量，最近三年财务会计报告由注册会计师出具无保留意见的审计报告。发行人内部控制制度健全且被有效执行，能够合理保证公司运行效率、合法合规和财务报告的可靠性，并由注册会计师出具无保留结论的内部控制鉴证报告。”

### 3、本次证券发行符合《首发管理办法》第十二条的规定

经审阅、分析发行人的《营业执照》《公司章程》、自设立以来的股东大会、董事会、监事会会议文件、与发行人生产经营相关注册商标、专利、发行人的重大业务合同、募集资金投资项目的可行性研究报告等文件、众华会计师出具的《上海合合信息科技股份有限公司内部控制鉴证报告》、发行人董事、监事、高级管理人员出具的书面声明等文件，保荐机构认为：

(1) 发行人资产完整，业务及人员、财务、机构独立，与控股股东、实际控制人及其控制的其他企业间不存在对发行人构成重大不利影响的同业竞争，不存在严重影响独立性或者显失公平的关联交易。

(2) 发行人主营业务、控制权、管理团队和核心技术人员稳定，最近 2 年内主营业务和董事、高级管理人员及核心技术人员均没有发生重大不利变化。发行人的股份权属清晰，最近 2 年实际控制人没有发生变更，不存在导致控制权可能变更的重大权属纠纷。

(3) 发行人不存在涉及主要资产、核心技术、商标等的重大权属纠纷，重大偿债风险，重大担保、诉讼、仲裁等或有事项，经营环境已经或将要发生的重大变化等对持续经营有重大影响的事项。

发行人符合《首发管理办法》第十二条的规定。

### 4、本次证券发行符合《首发管理办法》第十三条的规定

最近 3 年内，发行人及其控股股东、实际控制人不存在贪污、贿赂、侵占财产、挪

用财产或者破坏社会主义市场经济秩序的刑事犯罪，不存在欺诈发行、重大信息披露违法或者其他涉及国家安全、公共安全、生态安全、生产安全、公众健康安全等领域的重大违法行为。

发行人董事、监事、高级管理人员不存在最近 3 年内受到中国证监会行政处罚，或者因涉嫌犯罪被司法机关立案侦查或者涉嫌违法违规被中国证监会立案调查，尚未有明确结论意见等情形。

发行人符合《首发管理办法》第十三条的规定：“发行人生产经营符合法律、行政法规的规定，符合国家产业政策。最近三年内，发行人及其控股股东、实际控制人不存在贪污、贿赂、侵占财产、挪用财产或者破坏社会主义市场经济秩序的刑事犯罪，不存在欺诈发行、重大信息披露违法或者其他涉及国家安全、公共安全、生态安全、生产安全、公众健康安全等领域的重大违法行为。董事、监事和高级管理人员不存在最近三年内受到中国证监会行政处罚，或者因涉嫌犯罪正在被司法机关立案侦查或者涉嫌违法违规正在被中国证监会立案调查且尚未有明确结论意见等情形。”

## （二）市值及财务指标符合《上海证券交易所科创板股票上市规则》规定的标准

根据众华会计师出具的《审计报告》（众会字（2023）第 02665 号），发行人最近一年（2022 年）营业收入为人民币 9.88 亿元；根据本次发行上市方案，发行人本次发行股票的数量不低于 2,500 万股且不低于本次发行后公司总股本的 25%，参考发行人本次发行上市前最近一次的融资价格，发行人预计市值不低于 10 亿元；2020 年度、2021 年度及 2022 年度，公司研发费用分别为 16,696.80 万元、23,829.24 万元和 27,990.16 万元，2020-2022 年研发费用保持 29.48% 的年复合增长率；公司对研发人才的培养极为重视，截至 2022 年 12 月 31 日，公司全职员工共 912 人，其中研发人员共 499 人，占总人数比重约 55%。

因此，公司符合并适用《上海证券交易所科创板股票上市规则》第 2.1.2 条规定的上市标准：预计市值不低于人民币 10 亿元，最近一年净利润为正且营业收入不低于人民币 1 亿元。

## 九、保荐代表人对发行人持续督导工作的安排

事项	安排
(一) 持续督导事项	在本次发行上市结束当年的剩余时间及以后 3 个完整会计年度内对发行人进行持续督导
1、发行人建立和执行信息披露、规范运作、承诺履行、分红回报等制度的情况	1、督促发行人严格执行信息披露、规范运作、承诺履行、分红回报等相关制度； 2、督促发行人根据证监会、交易所的规定，及时更新完善相关制度； 3、与发行人建立经常性沟通机制，持续关注发行人上述制度的执行情况及履行信息披露义务的情况。
2、公司持续经营能力、核心竞争力或者控制权稳定有重大不利影响的风险或者负面事项	1、识别并督促发行人披露对公司持续经营能力、核心竞争力或者控制权稳定有重大不利影响的风险或者负面事项； 2、与发行人建立经常性沟通机制，持续关注发行人上述事项的信息披露情况，并发表意见。
3、发行人股票交易异常波动情况	1、关注上市公司股票交易异常波动情况，督促发行人按照上市规则规定履行核查、信息披露等义务；
4、督导发行人有效执行并完善防止控股股东、实际控制人、其他关联方违规占用发行人资源的制度	1、督导发行人有效执行并进一步完善已有的防止控股股东、实际控制人、其他关联方违规占用发行人资源的制度； 2、与发行人建立经常性沟通机制，持续关注发行人上述制度的执行情况及履行信息披露义务的情况。
5、督导发行人有效执行并完善防止其董事、监事、高级管理人员利用职务之便损害发行人利益的内控制度	1、督导发行人有效执行并进一步完善已有的防止董事、监事、高级管理人员利用职务之便损害发行人利益的内控制度； 2、与发行人建立经常性沟通机制，持续关注发行人上述制度的执行情况及履行信息披露义务的情况。
6、督导发行人有效执行并完善保障关联交易公允性和合规性的制度，并对关联交易发表意见	1、督导发行人有效执行并进一步完善《公司章程》、《关联交易管理制度》等保障关联交易公允性和合规性的制度，履行有关关联交易的信息披露制度； 2、督导发行人及时向保荐机构通报将进行的重大关联交易情况，并对关联交易发表意见。
7、持续关注发行人募集资金的专户存储、投资项目的实施等承诺事项	1、督导发行人执行已制定的《募集资金管理制度》等制度，保证募集资金的安全性和专用性； 2、持续关注发行人募集资金的专户储存、投资项目的实施等承诺事项； 3、如发行人拟变更募集资金及投资项目等承诺事项，保荐机构要求发行人通知或咨询保荐机构，并督导其履行相关信息披露义务。
8、持续关注发行人为他人提供担保等事项，并发表意见	1、督导发行人执行已制定的《对外担保管理制度》等制度，规范对外担保行为； 2、持续关注发行人为他人提供担保等事项； 3、如发行人拟为他人提供担保，保荐机构要求发行人通知或咨询保荐机构，并督导其履行相关信息披露义务。
9、现场检查	1、定期进行现场检查，出具并披露持续督导跟踪报告； 2、对发行人存在的可能严重影响公司或者投资者合法权益的事项开展专项核查，并出具现场核查报告。
(二) 保荐协议对保荐机构的权利、履行持续督导职责的其他主要约定	1、指派保荐代表人或其他保荐机构工作人员或保荐机构聘请的第三方机构列席发行人的股东大会、董事会和监事会会议，对上述会议的召开议程或会议议题发表独立的专业意见； 2、指派保荐代表人或保荐机构其他工作人员或聘请的第三方机构定期对发行人进行实地专项核查。



事项	安排
(三)发行人和其他中介机构配合保荐机构履行保荐职责的相关约定	1、发行人已在保荐协议中承诺全力支持、配合保荐机构做好持续督导工作，及时、全面提供保荐机构开展保荐工作、发表独立意见所需的文件和资料； 2、发行人应聘请律师事务所和其他证券服务机构并督促其协助保荐机构在持续督导期间做好保荐工作。
(四)其他安排	无

## 十、保荐机构和相关保荐代表人的联系地址、电话和其他通讯方式

保荐机构（主承销商）：中国国际金融股份有限公司

法定代表人：沈如军

保荐代表人：冷小茂、刘文博

联系地址：北京市朝阳区建国门外大街1号国贸大厦2座27层及28层

邮编：100004

电话：（010）6505 1166

传真：（010）6505 1156

## 十一、保荐机构对本次股票上市的推荐结论

保荐机构中国国际金融股份有限公司认为，发行人上海合合信息科技股份有限公司申请其股票上市符合《公司法》、《证券法》、《首发管理办法》及《上交所发行上市审核规则》、《上海证券交易所科创板股票上市规则》等法律、法规的规定，发行人股票具备在上海证券交易所科创板上市的条件，同意推荐发行人在上海证券交易所科创板上市。


(本页无正文,为《中国国际金融股份有限公司关于上海合合信息科技股份有限公司首次公开发行股票并在科创板上市的上市保荐书》签章页)

法定代表人:

  
沈如军


2023年7月24日

保荐业务负责人:

  
孙雷

2023年7月24日

内核负责人:

  
章志皓

2023年7月24日

保荐代表人:

  
冷小茂

  
刘文博

2023年7月24日

项目协办人:

  
杨雅菲

2023年7月24日

保荐人公章

中国国际金融股份有限公司



2023年7月24日