

**疫情防控中的数据与智能应用
研究报告
(1.0 版)**

中国信息通信研究院

2020 年 3 月

版 权 声 明

本报告版权属于中国信息通信研究院，并受法律保护。转载、摘编或利用其它方式使用本报告文字或者观点的，应注明“来源：中国信息通信研究院”。违反上述声明者，本院将追究其相关法律责任。

前 言

习近平总书记在 2020 年 2 月 14 日主持召开中央全面深化改革委员会第十二次会时强调，“要鼓励运用大数据、人工智能、云计算等数字技术，在疫情监测分析、病毒溯源、防控救治、资源调配等方面更好发挥支撑作用。”近期，在防控新型冠状病毒感染的肺炎疫情斗争中，各级政府、医疗机构、科研部门和科技企业迅速行动，把大数据等技术应用到疫情监测分析、人员管控、医疗救治、复工复产等各个方面，发挥了巨大作用，为疫情防控工作提供了强大支撑。

为了全景展现大数据与人工智能技术在此次疫情防控中的应用情况，中国信息通信研究院云计算与大数据研究所牵头联合多家单位，搜集分析了 200 多个案例，梳理了社会各界利用大数据及数据智能技术参与疫情防控的实践方向，分析了数据驱动在疫情防控关键环节，探讨了面临的问题与挑战，并对下一步通过大数据等新一代信息技术开展疫情防控进行了初步讨论。

由于时间仓促，编写组对案例搜集掌握还很不全面，对目前取得的效果和存在的问题分析还不够深入，再加上疫情防控工作尚未全面结束，一些应用的成效还有待进一步观察，很多应用还在陆续推出，得出确切结论还为时尚

早。有鉴于此，后续我们还可能根据需要对报告进行补充完善。欢迎大家和编写组联系（liuhan@caict.ac.cn），反馈您的宝贵意见，也欢迎各方面继续踊跃提供在疫情防控 and 复工复产中切实发挥成效的大数据应用案例。



目 录

一、综述	1
(一) 科技“战疫”时间线	1
(二) 抗疫技术“数据”轴	3
二、应用分析	6
(一) 疫情分析展现	6
(二) 疫情防范管制	9
(三) 医疗医治增效	13
(四) 生活便民举措	16
(五) 复工复产管理	21
三、深度洞察	25
(一) 数据能力是疫情防控的基础	25
(二) 疫情领域应用包含四种类型	32
(三) 企业技术“抗疫”硬实力展现	35
(四) “开源众包”成为独特风景线	38
四、问题与挑战	44
(一) 数据来源有限，采集手段落后	44
(二) 数据缺乏治理，数据质量堪忧	45
(三) 隐私保护不足，存在合规风险	46
(四) 数据流通不足，数据孤岛严重	48
五、总结与建议	50

（一）强积累，发挥数据应用价值	50
（二）提质量，加强数据治理能力	50
（三）保隐私，确保数据合法合规	51
（四）推技术，创新技术应用场景	51
（五）寻机遇，推动数字经济发展	52
结 语	54

一、综述

（一）科技“战疫”时间线

2020 年开年之际，新型冠状病毒带来的肺炎疫情汹涌而至。疫情突发性高、传染性强、扩散性广、风险性大，防控工作任务艰巨、时间紧迫、形势严峻。在这场疫情阻击战中，大数据、云计算、人工智能等快速发展的新一代信息通信技术加速与交通、医疗、教育等领域深度融合，让疫情防控的组织和执行更加高效，成为战“疫”的有力武器。

本报告梳理了疫情发展与数据和智能驱动相关案例的重要节点时间线。可以看出，随着疫情发展，数据驱动的疫情防控在迅速展开，各企业的疫情防控应用场景不断涌现，应用范围持续拓展。



图 1 科技“抗疫”行动时间轴

（二）抗疫技术“数据”轴

通过梳理总结这些应用场景，我们发现“数据”可以在如下方面助力疫情防控。

1、有力支持疫情防控知识传播

借助于移动互联网和智能手机，人们可以随时随地获取最新疫情动态、科学防疫知识等各种数据。各地政府通过电子政务平台、微博、公众号等定时发布最新疫情动态，各类新闻客户端、社交平台、搜索引擎、短视频平台等也积极配合疫情相关信息的发布和传播。此外，众多

“互联网+医疗”平台推出了在线问诊服务，方便网友向医生咨询新冠肺炎防治相关内容，有效缓解了因疫情期间医疗资源紧张导致的就医难等问题，避免了普通病症人群涌向医院、形成聚集性交叉感染。

2、迅速锁定“涉疫”人员流动轨迹

通过集成电信运营商、互联网公司、交通部门等单位的信息，大数据可以分析出人员流动轨迹。具体来说，利用数据分析、数据挖掘等技术，一方面可以通过手机信令等包含地理位置和时间戳信息的数据绘制病患的行动轨迹；另一方面，根据病患确诊日期前一段时间的行动轨迹和同行时间较长的伴随人员，基于大数据分析可以推断出病患密切接触者。综合分析确诊病患、疑似病患和相关接触者的行动轨迹，可以准确刻画跨地域漫入、漫出的不同类别人员的流动情况。

这既为精准施治提供有力指导，也为预测高危地区和潜在高危地区提供了精准依据。

3、开展疫情发展态势预测与溯源

基于疫情高危人群相关数据，结合疫情新增确诊、疑似、死亡、治愈病例数，借助传播动力学模型、动态感染模型、回归模型等大数据分析模型和实践技术，不仅可以分析展示发病热力分布和密切接触者的风险热力分布，还可以进行疫情峰值拐点等大态势研判。利用深度学习等新兴人工智能技术，联合出行轨迹流动信息、社交信息、消费数据、暴露接触史等大量数据进行科学建模，可以根据病患确诊顺序和密切接触人员等信息定位时空碰撞点，进而推算出疾病传播路径，为传染病溯源分析提供理论依据。

4、助力地方政府科学精准施策

运用大数据分析，结合算法模型对疫情的传播速度、传播趋势等进行预测，可为各地进行动态监测管理、统筹医疗物资储备、保障民生物资供应、制定交通管制政策等提供有效依据。例如，基于疫情高发地区人员在春运期间的交通出行数据进行疫情分析预警，能够通过追踪确诊患者、疑似患者和密切接触者的轨迹位置进行精准防控。同时，通过大数据分析还可以评估预测疫情对近期和远期社会经济运行带来的影响，建立快速、高效的经济应急反应机制，帮助政府适时出台减税、降费、专项补贴等各类措施，缓解中小企业

因疫情导致的资金链断裂风险及可能出现的连续经营困难，努力保持生产生活平稳有序¹。

5、推动病例诊断与疫情研究

运用大数据和人工智能等相关技术，可以有效加速新型冠状病毒宿主预测、药物筛选等数据分析和计算工作，极大提高病毒研究与攻克效率。

¹ <https://mp.weixin.qq.com/s/vuH0jg7L15rDyNNyKjrgvQ>, 《信息技术助力打赢疫情防控阻击战》

二、应用分析

本章梳理了数据在疫情分析展现、疫情防范管制、医疗医治增效、生活便民举措、复工复产管理等五个主要方面的应用案例。案例从互联网渠道和企业申报渠道获得，共搜集分析了 200 多个案例进行分析。入选的案例均已在实际应用中取得了良好的效果，在疫情防控过程中发挥了重要价值。

（一）疫情分析展现

我们每天打开手机各大新闻客户端看到的疫情数据就是疫情分析展现的典型应用。可以说，疫情相关的数据是开展疫情分析、管控等各方面的基础。因此，很多企业首先就针对各地提供的数据进行了多主体、多渠道、多维度的展现。

从本报告的调查结果看，有 39.6%的企业案例是对所采集数据进行直观的可视化展现，而超过 80%的企业案例在此基础上进行了更为深入的数据分析；数据来源渠道包括公开数据（8.1%）、人员上报数据（36.1%）、自有系统数据（41.3%）和其他系统对接及第三方数据（17.1%），各类案例数据来源渠道存在交叉情况；从服务主体角度看，有 52.2%的企业案例为政府提供数据采集和分析服务，有

30.4%为公众提供信息展示服务，有 17.4%是为其它企业提供数据分析和展示服务。

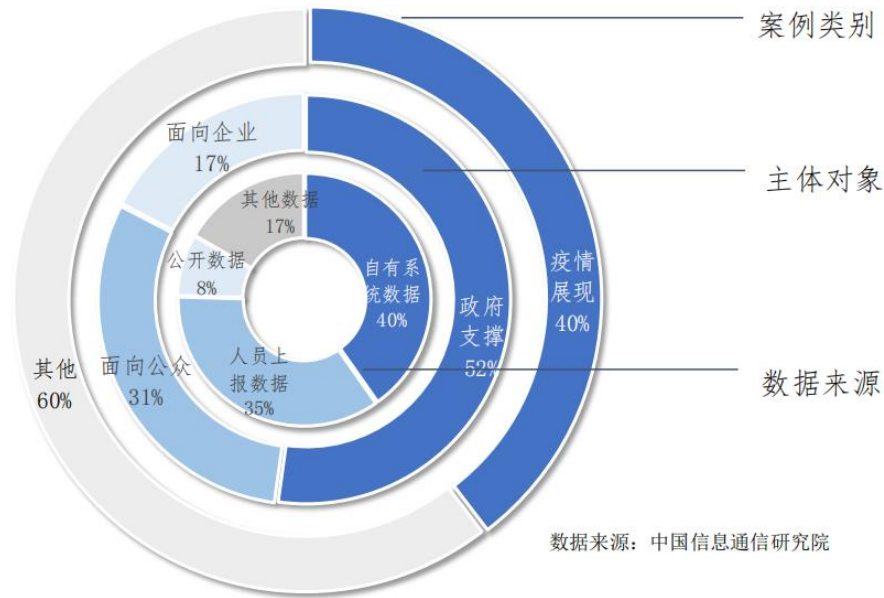


图 2 疫情分析展现

本节从三类服务对象角度出发，对疫情分析展现应用的价值进行探讨。

1、政府支撑

从政府支撑的角度看，疫情数据分析展现的应用主要包括政府管控范围内的疫情相关信息展示、人员流动情况展示、车辆流动情况展示、疫情相关资源情况展示、物流信息展示等核心功能。通过对这些重要信息进行全方位、多角度的实时展示，支撑了政府对于疫情的防范、管理和控制。

北京移动的“疫情防治人口大数据平台”自1月26日上线以来，为北京市委市政府、13个区县及相关委办局提供了包括疫情地进入用户、疫情地返回用户、外省进入用户、外省返回用户、疫情地未返回用户、非常驻用户的规模监测及分布，以及各类人群画像及分布热力图等疫情专项分析服务，对高危人群、潜在高危人群、潜在风险人群的精准疫情防御、排查、监测、宣贯全过程提供数据支撑。杭州数梦工场科技有限公司在浙江省衢州市落地了“城市大脑”项目，通过大数据分析驾驶舱，全局展现本地人员的网格分布情况，并对市内外地重点车辆进行监测、排查及监控，实现了漏报率小于1%。此外，东软、四方伟业、相数科技、和智信、朗新科技、美数信息等企业的疫情分析展现平台也为各地方的政府疫情防治和管控工作提供了有力支撑。

2、公众服务

从公众服务的角度看，疫情数据分析展现的应用主要包括疫情信息展示、人流迁徙呈现、疫情专题服务、舆论检测与评价、民众信息上报与展示等核心功能，及时为公众播报疫情信息动态，并提供有效疫情防控办法。

百度地图迁徙大数据平台开放查询的城市从100个扩展到了300多个，数据指标丰富，包含来源地、目的地、迁徙规模指数、迁徙规模趋势图等，甚至支持查询一个城

市自春运以来迁徙目的地或来源地的排行与比例，提供全面、立体的迁徙大数据服务支撑。同时，京东云的“市民疫情隐患上报系统”可实现市民随手拍随手上报，发布位置可自动实现地理位置定位，生成的数据安全可隔离，并有专属数据后台管理功能，可及时高效的支持民众疫情上报。此外，微信和支付宝的抗疫信息专题、百度推出的“社区防疫电子出入证”、太极集团的“全国一体化平台疫情防控专题服务”、中移雄安信息通信科技有限公司的“新冠病情动态展示系统”、北京百分点、智慧星光、云基华海信息的相关疫情分析与展示系统也纷纷上线，为公众防疫工作提供了有力支撑。

3、企业服务

从企业服务的角度看，多数企业通过自建或采购疫情分析与展示产品，实现企业内部疫情的有效防控和管理。

航天科工网信自主搭建了企业疫情管控平台，重点关注肺炎疫情的信息实时展示、高危人群筛选、疾病辅助诊断、爱心捐赠追踪等。

（二）疫情防范管制

大数据分析和展现已在很大程度上完成了数据采集和整理工作，通过更深层次的模型建立、分析挖掘等手段能够在原有基础上更大发掘数据价值。疫情防范类应用通过

数据来进行识别高危人群、开展区域检测、进行市场监管等功能，为政府部门进行决策提供了支撑。

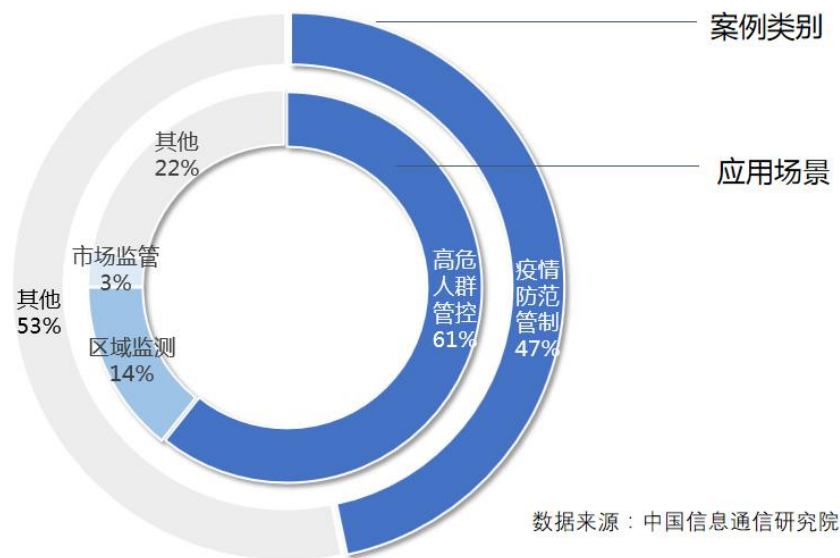


图 3 疫情防范控制

从本报告的调研结果看，能够有效支撑疫情管控的技术方案和场景应用案例占比达到 46.7%，可见通过各类技术手段提升疫情管控的手段和效率，已经成为此次疫情期间占比最高的应用场景。总体来说，疫情管控的各类应用场景当中，最为重要的是高危人群监测和管控，具有此类功能的案例占比达到 60.7%；其它管控手段，如区域监测（14.3%）、市场监管（3.1%）也都有一定程度应用。另一个重要结果显示，通过技术研发和应用适配形成有效的技术手段和技术方案，从而提升疫情管控的支持能力相关案例占比达到 32.1%。智能外呼、图像识别、微服务快速整

合、高维机器学习、知识图谱、时空数据分析、可视化展现等技术都成为有力手段。

1、高危人群管控是重中之重

通过位置数据和各类行为数据有效识别高危人员的行动轨迹和接触人群，能够从根本上降低疫情传播的程度，也是各级政府部门当前非常重要的工作。

中国联通大数据公司开发了传播风险分析、时空相关分析等一系列数据模型，通过多维数据融合分析，实现了对特定区域人群的扩散轨迹、已确诊人群的接触者范围等进行定位和分析，有效支撑政府部门区域化疫情防控工作。中国电信云计算公司开发的“翼知疫行”，通过电信的GIS系统数据分析，可提供高危人员近14天的行程，并进行密切接触风险判定，有效支撑了政府部门的疫情防控工作。此外，中航信的“新型冠状病毒确诊患者同航班自动通知系统”、美亚柏科的“新型冠状病毒传播监测系统”、曙光云计算集团有限公司的“疫情排查管理上报系统”等具有高危人群排查和监测功能的系统，都对各级政府针对高危人群防控的相关工作给予了大力的支持。

2、各类管控场景均有应用

除了高危人群的监测和管控以外，防疫产品的市场监管、区域人员的健康追踪也是重要的应用场景。

北京华宇软件公司的“网络交易监管”系统以网络交易信息智能采集和分析为基础，助力云南省市场监管局对疫情期间网络交易的价格波动实现有效掌控，精准开展特别是针对于防疫产品的市场监管工作，保障民生安全。中国联通大数据公司的“监测人员防疫合规情况监测系统”，可对口罩穿戴、人流聚集和体温异常等情况进行实时告警提示，帮助政府做到疫情的实时可知。京东云的“疫情通”产品，为社会管理机构提供“多维度”、“可视化”和“五位一体”（人、地、物、事、组织）的信息发布与疫情防控服务，为基层组织提供网格化疫区返工人员的健康动态评估，实现区域化精准防疫管理。

3、科技力量全面融合支持

科技企业的技术能力是防疫工作能够取得突破的核心推动力，AI 图像识别、智能外呼、知识图谱、安全多方计算、微服务等多项技术的广泛应用，有力推进了疫情防控工作高效安全开展。

百度基于 AI 图像识别技术和红外热成像技术，形成了 AI 测温系统，可对人流中多人额头温度同时进行快速筛选及预警，方便人流聚集处的快速筛选，有效减少公共场合人群聚集。同时，百度紧急推出了“疫情防控的免费智能外呼平台”，可提供流动人员排查、本地居民排查/回访、特定人群通知三大场景的外呼服务，可有效支撑各级政府

对于社区情况排查和通知回访等应用场景，比人工电话效率提高数百倍，目前已经在北京、上海等十几个地区投入使用。阿里云的“疫情信息采集系统”依托阿里云宜搭平台优势，通过可视化拖拽操作有效发挥后台微服务模块作用，快速支撑浙江省 11 个地市卫健委工作。此外，北京滴滴股份有限公司通过滴滴桔视（车载录像设备）采集的图片，经过人工智能识别算法来识别司机是否佩戴口罩；厦门渊亭科技公司基于知识图谱开发的“疫情智能作战平台”、四方伟业基于三维城市模型构建的“疫情防控分析系统”、洞见智慧科技有限公司基于时空大数据和多方安全计算技术开发的“疾控智能分析平台”都有效支撑了各级政府部门和企事业单位的疫情管控工作。

（三）医疗医治增效

在此次抗击疫情的过程中，大数据和智能技术被充分应用到病情诊断、医学科研、医疗辅助等与医护工作直接相关的场景中，是对大数据技术的最严苛的试炼。从本报告的调研结果看，有接近 17% 的应用在医疗医治增效中。医治增效应用的种类包括资源对接、辅助诊断、线上问诊、科研支撑和其他（包括基于图像分析的无接触体温监测应用以及时识别高风险人群等其他应用）。其中，辅助诊断指通过 AI 技术辅助或加速确诊病例的判断；线上问诊指通

过智能问诊服务，减轻医疗机构的诊疗压力；科研支撑指通过开放算法、模型或提供计算存储资源来提升科研效率，助力基因检测、疫苗研发等工作。具体占比如图 4 所示。

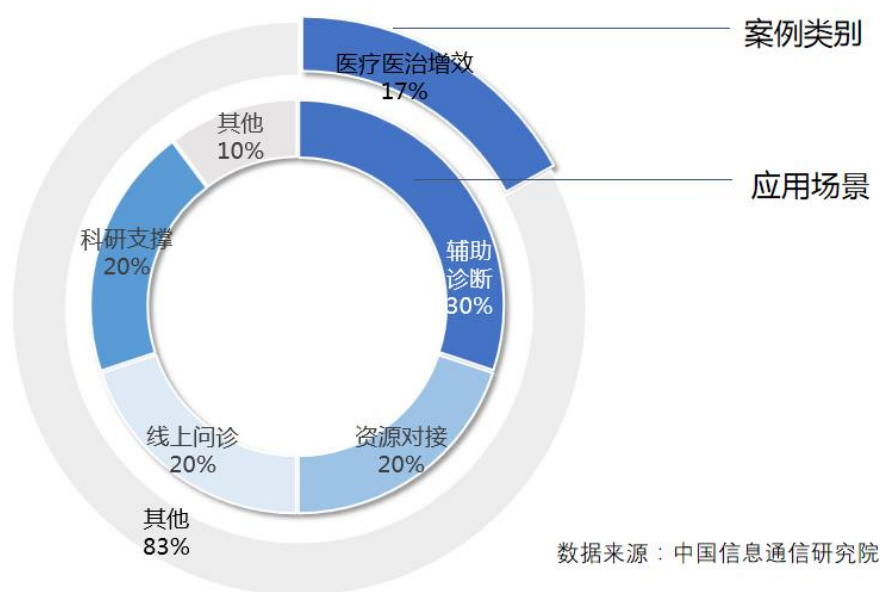


图 4 医疗医治应用

本节主要结合具体案例，介绍了算法算力、人工智能和智能问诊等技术在疾病研究、辅助诊断和线上问诊的具体应用。

1、算法算力辅助疾病研究

从科研的角度来看，人工智能、大数据等技术正在病毒结构分析、疫苗研发中崭露头角。1 月 30 日，百度研究院向各基因检测机构、防疫中心及全世界科学研究中心免费开放线性时间算法 LinearFold，以及世界上现有最快的 RNA 结构

预测网站。LinearFold 算法可将此次新型冠状病毒的全基因组二级结构预测从经典算法的 55 分钟缩短至 27 秒,提速 120 倍,能够节省两个数量级的等待时间,极大提升科研工作效率,助力疫情防控。此外,荣之联为中国疾控中心提供了急需的大数据计算和存储资源,以保证大规模并行样本分析、数据保存和管理,全面助力病毒测序工作;同时其也为中科院微生物研究所提供技术支持,搭建病毒基因组进化关系的分析流程,并提供进化树可视化的展示功能,以实时监控病毒的变异情况、追查病毒宿主来源。

2、人工智能加速疾病诊断

从诊断支持的角度来看,大量的人工智能技术被运用到病情诊断的过程中,以辅助判断病情、缩短确诊时间。2 月 1 日,浙江省疾控中心上线自动化了全基因组检测分析平台,该平台利用阿里达摩院研发的 AI 算法,有效缩短疑似病例基因分析时间,并能精准检测出病毒的变异情况。此外,科大讯飞的智医助理为其覆盖的省内基层门诊病历提供在线分析能力,以发热、咳嗽、呼吸困难、流行病学史(武汉相关史)、影像学、血常规六个维度进行病历内容挖掘分析,筛选出潜在患者,为安徽省卫健委提供决策参考。推想科技针对新冠肺炎推出的 AI 系统能够帮助影像科医生更高效地排查筛选疑似患者,减少患者在医院等待时的交叉感染风险,在缺少病毒检测试剂盒时,能够协助对早期患者进行排查。

该系统最早应用于武汉同济医院，并在全国各地陆续上线，包括深圳市第三人民医院（国家感染性疾病临床研究中心）等对抗肺炎疫情最前线的医疗机构。

3、智能问诊供需对接助力后勤保障

从医疗辅助的角度来看，大数据技术更是被运用到了减轻医院压力、减少人群聚集、整合医疗救护资源等多个方面。阿里和京东均推出了疫情服务机器人，能够向用户提供急需的线上问诊、疫情知识普及等服务，减少医护人员工作量，降低医院门诊压力。百度灵医智惠推出“智能咨询助手”，通过提供标准化预问诊路径提升医生诊治效率，并向在线健康咨询平台、政府疫情防控平台、互联网医院等提供平台免费开放 API 接口，全力支撑医疗医治领域智能应用。软通智慧推出的“新冠疫情防控指挥平台”通过整合医疗救护资源，优化防控物资保障，帮助各个城市提升基层防控能力。

（四）生活便民举措

生活服务类应用也是数据驱动疫情防控的重点突破口。诸多互联网企业采用 O2O 服务模式，形成线下活动到线上活动的映射，利用大数据技术实现海量生活数据的采集、分类和存储，为居民提供食品无接触外送、实时疫情

地图、互联网医疗等服务，在便利居民正常生活的同时，确保各类服务的健康安全。

1、电商平台保障饮食便捷安全

受新冠疫情的影响，线上买菜相较于线下买菜可以降低交叉传染的风险，因而更受欢迎。生鲜电商协同各方打通农产品上行通道，加大商品供给，让居民的“菜篮子”更稳当。

电商平台包括盒马生鲜、叮咚买菜、每日优鲜、美团买菜、天猫超市、永辉买菜、京东到家等，多数平台根据距离所在地区的远近优选超市门店，并提供全天配送服务。据统计，在商品丰富程度方面，半数以上平台在不同门店的商品丰富程度有所浮动；在配送时长方面，80%以上的平台可以满足2小时以内完成配送；在无接触配送方面，所有平台均提供无接触配送服务，降低了面对面接触带来的交叉感染风险；在预约配送方面，受限于物资紧缺和运力有限，所有平台在疫情前期提供预约配送服务，但随着物资和运力紧张程度逐渐缓解，30%左右的平台逐渐实现“即买即送”服务。

同时，为确保配送事物的安全，多家平台将“无接触配送”升级为“无接触安心送”。以美团外卖为例，在原有的基础上，将厨师、打包员、骑手的健康情况、餐具消毒情况等安全防护信息，以商家端和配送端的“电子卡+实

体卡”形式更透明的呈现给用户，确保无接触的同时实现全过程食品安全信息可视化、可追溯的安全配送模式。

相较于“无接触安全送”而言，“无人车配送”在此基础上，减少了骑手带来的潜在接触风险。百度公司针对抗疫情场景的自动驾驶作业车企业，免费提供百度 Apollo 平台的自动驾驶云服务矩阵支持，助力企业更好地开展疫情作业车研发工作。基于此，百度 Apollo 生态合作伙伴新石器在海淀医院投入一台无人车，2月14日起开始进行无接触送餐服务，最大限度的降低了工作人员的现场工作以及接触程度。中关村“白犀牛”科技公司提供的“无人车配送”已在北京市海淀区温泉镇的“佳苑便民外卖超市”落地，在线下单后，无人车可将商品配送至小区门口，居民确认身份后，打开相应车门即可取货。

2、地图应用平台保障出行安全

自新冠疫情发生以来，手握大数据资源的多地图应用平台都迅速推出利于疫情防控的出行指南，以满足用户特殊时期的出行需求。统计显示，80%以上的平台具备“疫情地图展示”、“发热门诊查询”、“同乘信息查询”等功能，部分平台提供“各国入境最新政策”等功能。百度地图于2月6日上线“疫情小区”专题地图，截止2月12日，覆盖200余个城市及乡镇。在“疫情小区”板块，输入所在小区名称，即可提供周边疫情提示，包括本市已公

布的疫情发病场所、距离最近的疫情发病场所，以及周围的人流聚集地。借助百度地图的“热力图图层”，还可查看到实时人口流量密度，以此来帮助居民决定出行路线，避开商圈、交通枢纽等人流密集场所，如图 5 所示。同时，“小度无接触式电梯服务”通过语音识别算法和自动控制技术，实现了语音呼叫大体按键，大大减少了电梯接触式病毒传播的风险，增强了便民出行手段。

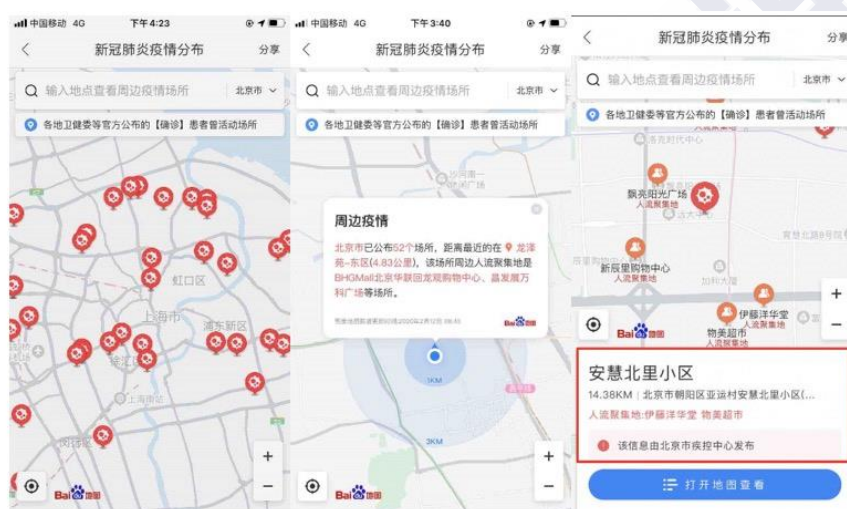


图 5 百度地图“疫情小区”功能示意图

随着复工人数的增加，作为主要通勤交通工具之一的地铁，在同行乘客过多的情况下，交叉传染的风险将会增加。百度地图也与北京市交通委取得合作，上线地铁客流量查询服务，方便用户通过百度地图随时查看地铁车厢的拥挤程度，为安全出行提供有效参考。高德地图于 2 月 12 日上线部分城市地图客流满载情况查询功能，有效区分拥挤程度，方便居民掌握地铁站线的实时客流满载情况，并选择合适的出行方式。

3、互联网医疗方便居民就医

在疫情就医领域，国内多家互联网医疗服务平台充分利用 5G 等信息通信技术，联动全国各地的医生、专家开展线上咨询、线上问诊、健康科普、心理援助辅导等，为百姓在家就医提供便利，让信息技术在疫情防控中“智慧相助”。工信部数据显示，目前全国 190 多家公立医疗机构及近 100 家企业互联网医院针对疫情提供在线义诊，缓解线下医院压力。好大夫在线、丁香园、春雨医生、平安好医生等企业集结了呼吸科、感染科、内科等领域过万人医疗专家资源，为患者提供免费问诊，同时为防疫一线的医生、护士等开通了热线服务，提供心理疏导。京东数科推出了疫情问诊机器人，借助人工智能，机器人通过学习与疫情相关的专业知识，为用户的提问进行基本的分析判断，同时，提供防疫知识科普等服务。目前，问询机器人已在近 20 家机构的微信公众号或 App 中接入使用，在杭州、苏州等 23 个智慧社区中完成落地。同时，百度灵医智惠、阿里健康等互联网平台提供疫情智能问答，北京、天津、浙江等地也积极搭建针对疫情的线上医生咨询平台。

为方便群众买药，多个平台提供发热门诊查询和配送药品等服务。百度地图上线“发热门诊地图”，可查看附近开设发热门诊的医疗机构名单以及营业时间，目前已覆

盖 300 多个城市。同时，高德地图在全国超过 200 个城市上线药品配送小程序，联合蜂鸟跑腿、叮当快药、UU 跑腿等合作伙伴，为广大用户提供快捷、安全的足不出户送药上门服务。阿里巴巴旗下阿里健康于 2 月 6 日在淘宝 App 上线“买药不出门”服务，通过线上问诊开方、药品配送到家的互联网就医方式，满足慢病患者需求。

（五）复工复产管理

受突发疫情波及，餐饮住宿、文化娱乐、交通运输等行业运行放缓甚至停顿，从业人员待岗时间拉长，相关上下游产业也相继受到波及。随着疫情防控取得积极进展，各地政府、企业也在逐步将精力放在复工复产上，尽快实现经济社会常态化运行。

从本报告的调查结果看，企业案例主要围绕政府、企业、学校的复产复工复学，以及协助政府开展复工后城市疫情防控各项工作，如图 6 所示：

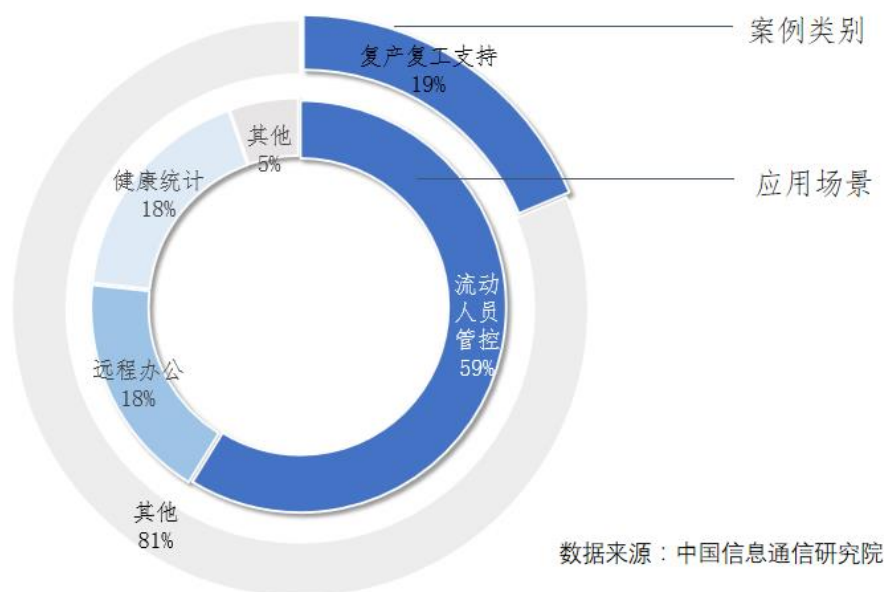


图6 复产复工应用

1、通信大数据行程卡提供地理位置查询

随着返程高峰来临之际，预防人员流动带来的交叉感染，是现阶段各地方、各单位防疫工作重点。为了简单、快速、权威、准确地证明自己过往 14 天内的行程信息，中国信息通信研究院联合三家基础电信企业利用电信大数据，推出“通信大数据通行卡”服务，为全国 16 亿手机用户免费提供其本人前 14 天内到访地服务。用户扫描二维码输入手机号即可实现跨运营商一站查询，操作方便快捷，无需安装软件。可帮助各地方、各单位及时识别外来及返工人员带来的风险，方便政府精准施策，帮助各地科学、高效安排复工复产工作。通信大数据行程卡已于 2 月 29 日下午上线，截至 3 月 1 日晚，已累计提供 53577 次查询。

2、健康码助力城市有序复工复产

当前，多地通过数字化管理措施研判疫情期间返工返学返岗形势，并针对重点区域、重点场所采取分级分类管控，从而助力分类有序复工复产。例如，杭州深圳相继上线“健康码”开展疫情期间的社区管理及交通出行等工作，市民或者返工返岗人员通过支付宝、微信平台自主申报获取健康码，作为个人在本地区出入通行的一个电子凭证，实现一次申报、全市通用，公共场所也针对健康码对人员进行分类分级管理。日前，在国务院办公厅电子政务办的指导下，腾讯和支付宝即将推出全国一体化的政务服务平台疫情防控健康信息码，有望解决多条线采集居民健康信息带来的数据标准不一致和跨地区不可用等问题，实现跨省跨地区的疫情服务互联互通，更好地支撑全国疫情防控工作。同时，三大运营商均推出了个人疫情期间行程查询的应用，例如中国联通大数据公司推出的“健康U码”，能方便用户查询自己14天行程，以及上报自己的健康状况，生成不同级别的健康码。

3、远程办公完成政企学异地协同运转

从复工复产的角度来看，远程办公和人员信息上报是主要的两个方向。百度公司宣布“百度Hi企业智能远程办公平台”于2月11日对外开放，并将免费为湖北等疫区企业提供高清音视频会议、企业云盘、企业IM和应用中心平

台等多项服务，满足疫情期间不断增长的远程办公需求，支持企业快速恢复生产能力，减少疫情对于企业和社会经济的影响。阿里“钉钉”面对本次疫情引发的延期开学、复工潮，针对各类企业运营真实场景定制每日健康打卡、百人高清视频会议、紧急通知发布等功能，同时也针对远程复学提供群直播等功能定制化服务。华宇软件则为法院、检察院、政法委、司法行政等法律业务场景提供云视频服务和法律视频业务相关服务的整体解决方案，包括法院互联网庭审、司法行政视频调解，检察院远程视频接访等。

4、大数据精准保障企业增产扩能

国家电网也通过电力大数据对城市复工复产情况进行分析预测。以国网浙江电力为例，其根据用电信息采集系统中企业历史用电量情况、当日用电量情况等数据，动态监测、精准分析各区域、各行业由点及面的复工复产情况。电力大数据助力增产扩能，快速确定供电方案，为政府部门指导疫情防控和复工复产提供有价值的数据支撑和决策参考。百度智能质检系统，助力常州一家工厂安全复工，通过无人值守的智能化检测设备 24 小时工作，比人工检测效率提升近 10 倍。该系统有效解决了疫情期多数工人无法返厂的难题，同时还减少了车间里疫情传染的风险。

三、深度洞察

（一）数据能力是疫情防控的基础

良好和丰富的数据是开展疫情防控应用的基础。本节从数据采集、数据互通和数据开放三个方面，对本报告收集到的相关案例中体现出来的数据能力进行分析。

1、数据采集

传统数据采集方式在大数据量面前暴露出了一些弊端。传统卫生数据的采集起点通常是基层的社区卫生服务中心，通过社区人员手工填报，经历区卫健委、市卫健委，最终汇集至省卫健委和国家卫健委。这一方面增加了基层数据采集工作人员的负担，降低了数据汇集的效率，另一方面难以在数据源头快速核验数据的正确性，增加了后期数据质量管理的成本。

借助信息化，目前已有多地采用在线信息填报系统和智能外呼平台代替手工填报，力求减少数据采集人力成本，缩短数据流转过程。北京于2月13日上线的“京心相助”服务，居民可在支付宝上登记个人信息，在不更换社区的情况下，市民无需反复进行个人信息申报，但需进行每日健康打卡²。同时，百度、阿里、北京来也网络等企业推出了针对疫情防控的智能外呼平台，在自动询问居民包

² <https://new.qq.com/omn/20200213/20200213A0K1Z200.html>

括离返京情况、身体状况等问题后，系统将根据关键词自行归纳信息档案，快速实现辖区内居民疫情数据的采集。除此之外，企业和个人也采用爬虫技术采集动态数据，以快速提供包括疫情数据可视化、疫情态势分析、疫情推演等服务。数据采集技术的主要使用对象和优劣下表所示。

表 1 数据采集技术对比

数据采集方式	主要使用对象	优势	劣势
纸质填报	政府	疫情初期和小数据量下可及时全面采集	● 数据采集效率低下 ● 数据汇集流程繁琐 ● 存储性可用性较差 ● 数据难以核实
线上填报系统	政府、企业	● 快速采集大量数据 ● 数据易于存储加工 ● 采集数据可靠性高	数据更新不及时
已有信息化系统提供	政府、企业	● 能够迅速获取数据 ● 数据的实时性较高 ● 数据的真实性较高	对接过程复杂
智能外呼	政府	快速采集大量数据	● 难以全面采集数据 ● 采集数据不准确 ● 开发成本较高
爬虫	企业、个人	● 实时获取大量数据 ● 可捕捉更多相关数据	● 技术门槛较高 ● 数据可靠性存在风险 ● 非公开数据无法获得

为保障数据采集的全面和准确性，应同步采用自动化采集工具和数据质量核验等手段，对于多源头数据应明确主要来源，从对应系统或平台中抓取，减少自报数据和重复采集。同时，充分利用大数据技术，实现各类疫情信息的快速实时采集。通过提供 ETL、API、消息队列、数据流等多种采集手段，提供移动端报表采集功能，确保实时准

确获取包括人口流动信息、交通实时信息、重点防控人员定位信息、物流信息、网络舆情信息等多类数据，满足疫情特殊时期下的数据资源需求。

2、数据开放

自新型冠状病毒肺炎爆发以来，面对公众对疫情数据的迫切需求，我国不少地方政府部门及时开放了疫情相关的数据。这不仅有利于消除公众的恐慌情绪，提高自我防护意识，配合政府防控措施，提升政府公信力，也为社会市场和社会上的专业组织和个人的疫情数据分析提供源源不断的“燃料”。

当前，各省市开放多种疫情数据，全面覆盖统计类疫情数据和医疗机构数据，为数据使用者提供了掌握疫情动态和查询诊疗信息的有效途径。各省市的统计类疫情信息基本覆盖日期、地段、时区、新增确诊病例、累计报告确诊病例、疑似病例、新增出院病例等数据项。同时，部分省市将统计类疫情信息分类，以方便用户查找数据，如深圳市开放 9 个数据集，将疫情信息分为每日确诊病例来源统计、每日确诊病例统计、每日诊疗情况等。

目前，许多省市开放结构化数据，并建立数据标准和元数据，提升疫情数据的可用性和易用性。例如，山东省建立了“山东省新型冠状病毒感染确诊病例信息”数据集的数据标准，规定了各数据项的英文名称、数据格式和中

文描述；深圳市针对各数据集建立元数据，包括关键字、发布时间、更新日期、资源格式、数据提供方、数据维护方等，实现数据快速检索。

同时，各省市已经基本做到每日上午前公布最新疫情，疫情的更新周期在 24h 以内，并提供多种数据下载方式，确保数据的时效性和可用性。各直辖市卫健委均从 2020 年 1 月 20 日或 1 月 21 日开始在官网上发布疫情数据，北京和天津每日的平均发布次数为 2 次，各地也根据疫情进展调整发布的时间和次数。北京、上海等地提供 API 接口调用和覆盖 xls、xml、json、csv 等主流格式的下

载。

但是，各省市的确诊病例数据以及地理空间数据开放不足，极大的限制了数据使用人员进一步分析挖掘疫情数据。同时，目前部分政府发布的疫情数据以碎片化的方式散落在不同时间发布的、位于不同页面上的、以不同名称出现的公告里³，增加了数据使用者提取和整理数据的困难。

医疗科研机构及时发布新冠肺炎最新的研究成果。2 月 28 日，钟南山院士团队论文在国际顶级医学期刊《新英格兰医学杂志》在线发表，分析了患者各年龄段人群分布特征、感染者症状、接触史、影像学表现、治疗手段、临床

³ 《哪些地方开放了真正“能用”和“好用”的疫情数据？这些数据能用来做什么？》

转归（包括病死率）等，帮助公众了解新冠肺炎诊疗动态和疫情发展细节⁴。科技企业由于掌握大量高精度用户行为数据，利用大数据技术，持续发布复工复产相关分析报告。高德地图于2月24日发布《驾车活力复工指数分析报告》，将实际导航的人数作为数据源，从驾车导航角度来反映20个主要城市复工热度⁵。百度发布的《新型冠状病毒肺炎搜索大数据报告-复工篇》通过使用“百度”搜索的关键词热度，为公众了解各行各业现状、员工及企业的真实需求、行业发展机遇等方面，提供了可靠参考⁶。

3、数据互通

数据互通将集合关键疫情信息，有效支撑政府疫情研判和复工复产。杭州在全国率先上线健康码，用红黄绿三色二维码作为数字化健康证明，由市民或返工返岗人员自行在线申报，提交包括目前所在地、出行和接触病患等信息，经政府后台系统审核后凭码同行。健康码实现了数据交叉可用，各地可根据健康码动态变化，科学评估复工复产比例，动态调整复工复产策略⁷。

在疫情防控工作中，部分政府与政府之间、政府各部门之间推进数据互通，在很大程度上简化数据采集流程，提高疫情防控效率。由中国电科与国务院办公厅电子政务

⁴ https://m.huxiu.com/moment/86665.html?f=app_24h_ios_friends

⁵ <https://www.xuexi.cn/lqpage/detail/index.html?id=12469248233056609514∓>

⁶ <http://cn.chinadaily.com.cn/a/202002/14/WS5e466b18a3107bb6b579fa63.html>

⁷ <https://baijiahao.baidu.com/s?id=1659038766757094374&wfr=spider&for=pc>

办公室、国家卫生健康委员会联合研制的“密切接触者测量仪”于2月8日上线，获得了国家卫健委、交通运输部、国家铁路集团公司和民航局等多方面的数据资源支持，从根本上保证数据来源权威性，普通公众通过输入个人信息就可以查询自己是否为新冠肺炎病人密切接触者⁸。工信部推进三家基础电信企业支撑疫情分析的大数据跨网统筹，搭建电信大数据平台，打通电信企业信令数据，同时，实现电信大数据与卫生疾控管理部门进行数据共享联动，通过结合卫生防疫等部门的数据，做进一步研判。济南市卫健委开发了发热就诊人群分析系统，汇集了全市86家主要医院和3000多家基层卫生机构，并结合健康医疗数据和出行数据，对重点人员进行全面跟踪和持续分析⁹。

服务型企业打通平台内企业和个人数据，企业之间逐步实现数据互通。航旅纵横汇聚了各航空公司的航班数据，于1月30日上线了“新型冠状病毒确诊患者同航班自动通知系统”，同步更新有确诊患者的航班记录，第一时间主动向该航班的所有旅客发送疫情通知¹⁰。国铁集团运营12306售票大数据资源，配合地方政府和卫生健康部门协查确诊病例乘坐火车上密切接触者，安排错峰返程运力¹¹。中

⁸ <http://www.sastind.gov.cn/n112/n117/c6808856/content.html>

⁹ <https://baijiahao.baidu.com/s?id=1659111197539074743&wfr=spider&for=pc>

¹⁰ <https://baijiahao.baidu.com/s?id=1657145992333744288&wfr=spider&for=pc>

¹¹

https://www.xuexi.cn/lgpage/detail/index.html?id=8145762554441060628&item_id=8145762554441060628

国铁塔股份有限公司联合 360 公司推出的“中国铁塔智慧疫情防控平台”，结合人员火车航班数据、计费数据，梳理出带有时间戳的人员出行记录和密切接触者，替代了传统流行病学的调查方式¹²，如图 7 所示。



图 7 中国铁塔智慧疫情防控平台

但是，数据互通仍面临诸多障碍，如尚未建立有效的数据互通机制，无法在较短的时间内明确数据互通的需求和范围，相关技术由于存在接口和规则壁垒，也难以支撑大数据量的快速联通。

因此，各地政府在未来应明确数据互通的需求，建立资源对接和调度机制，确定满足需求应配备的资源，集成已有 IT 系统的数据能力，消除数据互通各环节的技术壁垒。以工信部为例，在第一时间成立了电信大数据支撑服

¹² <https://mp.weixin.qq.com/s/fmFEc3N69GhBQwFPVzafKw>

务疫情防控领导小组，统筹协调相关工作，并组织各省市通信管理局、中国信息通信研究院、基础电信企业建立通信行业疫情防控联动工作机制，加强通信系统部省联动。同时，在国务院疫情联防联控机制下，与卫健委等部门建立疫情电信大数据共享联合工作机制，定时或按需向相关部门共享信息，有效的推进了部门间信息共享。

此外，加大政企间数据合作，使原本割裂分散在公共出行、社交媒体、电信服务等领域，与疫情防控相关的数据深度融合并有效使用，在复工复产后更加复杂的防控形势下，应用于减少二代感染者、阻断三代感染者等方面。同时，充分利用社交媒体和民间组织所拥有的大量疫情相关数据，通过与其他哨点监测数据、传统的公共卫生数据资源相结合，通过大数据技术和 AI，深入挖掘并识别出诱因。

（二）数据分析的深度逐渐递进加深

数据分析应用的深度一般会从其操作难度和产出价值等方面进行评估，在业界通常被分为四种类型：

描述性分析（Descriptive）：这类分析仅描述发生了什么。一般会通过可视化的方式呈现所传达的信息，比如大屏展示的疫情分布信息；

诊断性分析 (Diagnostic)：这类分析会追溯事件发生的原因。一般会基于描述性的信息进一步分析问题的本源，比如病毒传染的溯源；

预测性分析 (Predictive)：这类分析会预测可能发生的事件。需要对可变数据进行建模，并通过预测模型预测发生事件的可能性，比如对疫情发展态势的预测；

决策性分析 (Prescriptive)：这类分析能够指导下一步的决策。其复杂性和价值都相对较高，通常依赖描述、诊断、预测等多种分析行为的组合来判断最佳的方案，比如地方政府综合地方实情和周边疫情发展态势等多项因素后进行精准施策。

此次抗疫期间，数据应用主要呈现出以下特点：

疫情控制早期，是打好数据基础的关键时期。此时涌现的大部分应用多为信息收集和平台建设类项目，如各类自填报系统和大数据平台的建设。由于时间和数据量等方面的限制，2020年2月前的数据应用，60%以上属于初步的描述性应用，如疫情分布地图、人员流动展现等。出现较少部分诊断类的应用，多为病毒传染溯源分析以及结合人员流动的时空属性进行的高危人群判断等。例如航空和铁路部门在疫情初期发现确诊患者后会通知同航班和同车次的旅客进行重点关注和隔离。

随着平台的建设和数据的逐步积累，诊断类应用愈发成

熟，预测类应用逐步增加。至 2020 年 2 月初，描述类和诊断类应用的占比达到了 75%，数据粒度和分析粒度都更加细致，区县级甚至社区级的应用开始出现。例如百分点信息科技的“一区一码”系统，居民完成个人信息填报，进出小区时由系统自动分析辅助人工管理。预测类应用如潜在传染源和传染路径预测、传染人数趋势预测等开始出现，并逐步用于政府的决策参考。第四范式基于高维机器学习技术建立市区县级数字孪生系统，结合交通管制、复工时间、药物投放等多方因素对疫情发展的影响，对关键决策的实施影响进行仿真预判，为防控政策制定提供依据。

预测类和决策类应用将在疫情控制方面起到重要作用。预测和决策类应用复杂程度高，数量虽少，但具有更高的价值。此次疫情防控过程中，大数据和 AI 技术被用于病毒基因的检测分析中，基于病毒基因的特征，预测病毒蛋白结构辅助药物筛选，大大提高了诊断和疫情研究的效率。百度研究院此次向各大科研机构免费开放了线性时间算法 LinearFold，以及世界上现有最快的 RNA 结构预测网站，能显著提升新型冠状病毒 RNA 空间结构预测速度，以助力疫情防控。

前期各类描述、诊断、预测类应用的成熟为政府后续的精准确策打下基础。交通部门提供的乘客公路、铁路、航班出行分析与通信运营商提供的密切接触者轨迹分析能

够有效支撑政府下达针对性的防控措施；物资保障平台、“健康码”数据、企业用电量等分析应用也从生产的各个环节支撑着政府对于保障复工复产开展专项的应对措施。随着各环节数据的进一步开放以及疫情经验的不断积累，更多的预测决策性分析会应用会在疫情防控中发挥作用。

（三）科技企业展现技术“抗疫”硬实力

1、大数据价值尽显

“人传人”的疫情特征，将对于人类个体的监测与管控推到了核心地位，如何获取、描述和分析一个人的行为轨迹，及多人之间的位置重叠也就成为了最基础和最重要的分析数据。因此，无论是手机信令数据还是互联网 APP 所获取的位置数据都成为了重要的数据资源，众多电信企业和互联网企业也是基于此类数据开发了相应的产品。其中，由于实名制手机的改革，电信大数据的全面性、真实性和实时性就充分体现了其数据价值。

电信大数据来自于电信运营过程中的通信基础数据，运用电信大数据分析，统计人员流动情况，对支撑服务疫情态势研判、疫情防控部署以及对流动人员的疫情监测、精准施策有重要意义。电信数据具有以下特点：一是全面性。我国拥有 16 亿手机用户，电信用户规模大、覆盖面广、数据量大，如现有的公众通信网每日产生的电信数据

约数千亿条，在春运等节假日期间、人员流动性高的情况下，数据量会更大。二是真实性。电信大数据经过多年的实名制认证，已经基本可以达到一个号码对应一个人，通过基站信令数据对于人员位置的确定及身份的确定可靠性和真实性程度大大增加，为各类疫情数据分析奠定了有力的真实数据基础。三是实时性。可以实时采集、汇总和处理电信相关数据，及时提供各类数据分析结果，为疫情防控提供精细化数据支持。对于人员动态流动情况，分析预测确诊、疑似患者及密切接触人员等重点人群的动态流动情况，支撑疫情防控部署。

北京移动作为北京地区最大的移动通信用户运营商，拥有大量真实的用户数据，在遵照国家法律和行业规范的前提下，北京移动大数据中心上线了疫情防控整体解决方案。其中，防疫平台平台实现对信令数据的统一采集和实时处理，通过本地用户和外省地区间漫游数据的融合，更全面地了解全国涉疫人员流动信息，实现全轨迹链的还原分析。借助电信数据实名制并与自然人强关联特征，大大提高了防疫数据时效性。通过不间断手机数据的获取，建立自然人与地理空间及行为特征的深度联系，每日深入洞察重点区域人流情况，按日进行数据精细对比分析。

2、响应能力快速高效

新冠疫情如排山倒海般来势汹汹，面对如此紧急的突发事件，科技企业的快速应变能力成为了硬实力。从系统总线到面向服务思想，从模块化、松耦合到中台和微服务，技术的革新总是向着高效、敏捷的方向发展，这也正是解决快速应变需求的最重要手段。

无论是百度地图从 100 个城市到 300 个城市的迁徙大数据平台几乎无时延升级，还是阿里云宜搭平台将后台不可见的为微服务模块变成了前台的可拖拽操作，亦或是电信运营商依托长期的数据和服务积累迅速形成疫情大数据方案，这背后都是快速响应能力的重要体现。

3、智能化技术能力不断提升

大数据同人工智能技术相辅相成，共同提升人类智能化技术水平，智能化的方法和设备不断发展，辅助人类更多更好的完成任务已经成为了今后发展的重要方向。疫情期间，AI 智能化测温系设备、智能化机器人、智能无人车、智能化算法模型、智能化应用和智能化展示等技术帮助我们完成了更多人类不可能自己完成的事情。

智能化需求不断增加和智能化技术的不断革新也催生了智能化经济发展。数字经济在经历了 PC 的发明与普及、PC 互联网、移动互联网这三个阶段后，正在进化到以人工智能为核心驱动力的智能经济新阶段。智能经济将给

全球经济带来新的活力，是拉动全球经济重新向上的核心引擎¹³。

纵观此次疫情到目前进程，诸多科技企业通过充分激活 AI 技术潜力，在疫情分析与防范、医疗、复工复产等方面抗击疫情。其中 AI 测温系统、百度地图迁徙大数据平台等应用让人印象深刻。相关 AI 技术应用或在疫情结束后寻找到合适的商业落地场景，体现了智能化对于社会、经济发展的推动作用。

（四）“开源众包”成为独特风景线

在新型冠状病毒开始蔓延的初期，疫情相关的信息爆炸，发布渠道众多。医疗物资短缺的信息尚未有效统计和发布，给物资调度和捐赠带来巨大困难。此时，一群志愿者自发形成了研发团队，利用各自的专业优势，采用众包协作的方式构建了一个个疫情防护有关的信息化开源项目，致力于搭建可靠、高效的信息化渠道，成为了疫情防控科技力量中一道亮丽的风景线。

1、项目类型丰富

根据开源平台 OpenSourceWuhan 统计的 46 个开源项目来看，疫情信息类占到了 37%，新闻纪录类达到了 30%，其次是求助信息类 9%。其中疫情信息中一半以上的项目都是

¹³ 《智能经济带来新革命，产业智能化需修炼三重境界》，李彦宏，乌镇互联网大会

展示肺炎疫情实时的省市地图。新闻记录类的 14 个项目中，有记录各类媒体、非虚构与个人叙述的项目<2020 新冠肺炎记忆>，也有记录个体生活的<2020nCov-individual-archives>。

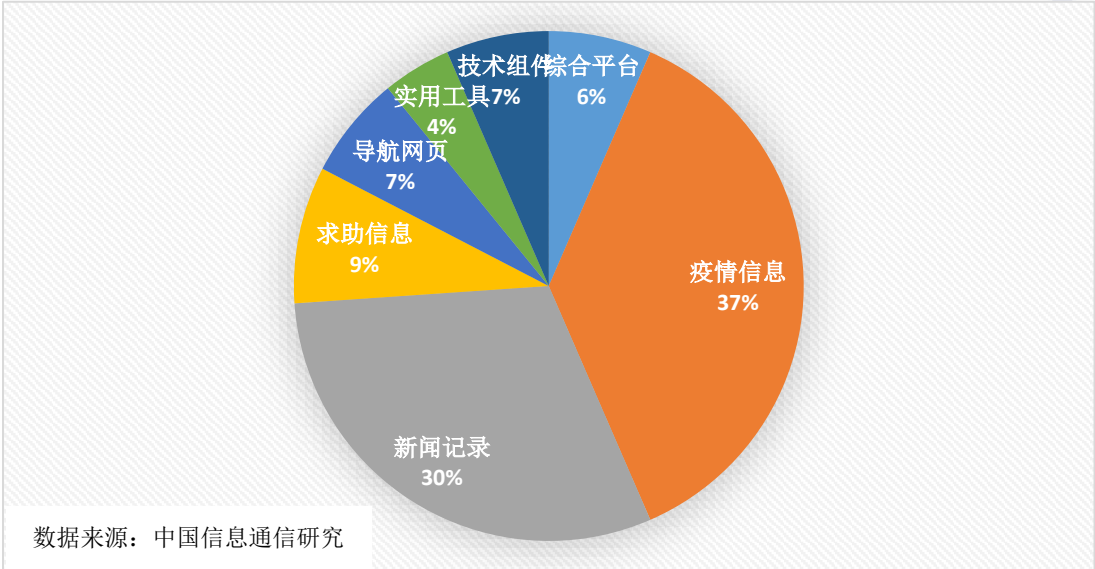


图 8 开源项目情况

2、运行快速高效

重大公共卫生事件在公共社会事件中属于较复杂的类型，调用资源类型多、统筹安排难度大、配置实施周期长。开源和众包的模式，可以让各供需方进行分布式自助对接，大幅提升效率。大多数疫情防控开源项目普遍在 2 到 3 天就能上线系统，项目参与的人数从几个人到上千人不等。以 wuhan2020 项目为例，该项目在 1 月 25 日发起，在两三天内便征集到几百人的参与，28 日上线系统，随后不断迭代系统。据统计，目前总参加的志愿者超过了 1000 人，项目在 Github 的贡献者达到了 39 人。而另外一个明星项目“2020 新冠肺炎记忆：各类报道、非虚构与个人叙

述”利用 Github 来存储各类报道，已经收录了 70 多家媒体的报道，参与者超过了上百人。2020 Wuhan Individual Stories 在短时间内收录了 300 多篇日志。相比而言，正常的委托类信息化项目通常的上线周期都在一周左右。

3、数据获取分为两个渠道

开源疫情防控项目的数据一般来源于两个渠道，一是公开数据，通过人工或者机器爬取，然后再整理同步。例如 wuhan2020 项目利用大量志愿者在网上人工爬取各种数据，经过审核人员审核后录入到 Github 中；而项目 DXY-COVID-19-Crawler 是一个爬取数据的工具，实时爬取网上的数据，也为其他项目调用提供了 API 接口；其他疫情地图的项目大多数都是通过爬虫来爬取各省市公布的疫情数据。二是来自于用户自己生成的信息和数据，直接汇总，例如 2020 Wuhan Individual Stories，号召武汉疫情中的普通人贡献自己的日记作为数据来源。

4、运行方式多样

涉及多人协作的项目需要制定一套分工协作的流程，包括协作规范、人员分工和沟通协作的工具。主要包括代码托管与协同，开发者之间的交流，数据和信息的收集、同步、审核，还有志愿者之间日常的沟通和正式的会议等。绝大部分的疫情防护开源项目都用到了 Github、Gitee 等代码托管平台来进行开发协作，其中 Github 使用频率较

高。Slack、微信、钉钉等工具作为志愿者日常沟通的工具，而很多开源项目都利用石墨文档等协作工具来搜集数据和文字，此外还有一些用于设计的图片处理工具等。同时，技术企业开源也成为重要的运行方式，百度飞桨免费开源业内首个口罩人脸检测及分类模型，该模型可以有效检测在密集人流区域中携带和未携戴口罩的所有人脸，同时判断该者是否佩戴口罩。百度借助国内最大的 AI 开放生态优势，推出“AI 开发者‘战疫’守护计划”，积极调动开发者、合作伙伴的力量，开发各类防疫控疫的产品，实现智能化能力全面输出。

5、影响力不断提升

开源项目在本次疫情防护过程中取得了巨大的社会效应，本报告统计了 Github 和 Gitee 上 Star 数前 7 款疫情防护开源项目。如表 2 所示，项目基本包含综合的信息展示、新闻记录、信息导航、疫情地图、信息爬虫等方面。截至 2 月 21 日，有两款项目在 Github 上 Star 数超过了 6000 个，Fork 数均超过了 800 次，均登顶 Github Trending。

表 2 开源项目功能及影响力

项目名称	分类	功能	Star 数	Fork 数
2019ncovmemory/nCovMemory	新闻记录	2020 新冠肺炎记忆：报道、非虚构与个人叙述	6600	880

wuhan2020 项目	综合信息平台	旨在收集各医院、酒店、工厂、物流、捐赠、捐款、预防、治疗、动态等信息，统一收集，统一发布，以便各方之间进行信息互通，有效调配社会资源。	6000	870
DXY-COVID-19-Crawler	信息爬虫	2019 新型冠状病毒疫情实时爬虫及 API	1000	207
wuhan2020-timeline	新闻记录	记录自 2019 年 12 月起武汉新冠肺炎疫情进展的时间线。	890	150
2019-ncov	疫情地图	展示	390	81
2020nCov_individual-archives	新闻记录	备份普通人在疫情期间的记录	340	45
xinguan2020	综合信息平台	外来人员登记系统、消毒检查登记系统、学生健康报备管理系统、违纪举报系统，物资管理系统，医院病患管理系统	206	59

综上所述，开源项目在这次疫情防护中表现出了独特的优势和特点。开源项目的运行管理使得大规模的社会化协作成为了可能，让普通民众和有专业技能的人员参与到疫情防护的工作中。开源项目可以在短时间内迅速的组织有效力量，快速搭建各类信息平台，其响应速度和迭代效率均高于常规的委托式的项目开发。开源项目的顺畅运行需要有一套协作流程和激励机制，通常涉及分工、协作规范和工具使用，其民主化的协调方式和决策机制调动了参与者的参与积极性。从协作的工具来看，主要分为代码托管、数据和信息采集汇聚、日常沟通等工具，这些分工协作的工具有效提升了分布式工作的效率，为远程在线合作

提供了有力的保障。从数据和信息来源来看，开源项目主要源依赖于网上的公开信息和志愿者的自发贡献，一般通过人工或机器来收集数据，面对人工收集和贡献的信息，还需要引入审核的机制来核验信息的有效性和真实性。通过实际项目的运行观察，这次开源项目发挥的作用和影响力都是巨大的，在疫情信息的展现和可视化、医院对物资的需求发布、记录媒体报道和个人生活等方面涌现出大量的应用，开源项目的信息和代码都遵循一定的开源协议开放给社会，大大提升利用率。

四、问题与挑战

虽然大数据和智能技术在疫情防控过程中可以发挥重大作用，但从此次疫情应对情况来看，特别是基于互联网企业在数据驱动的疫情防控过程中发挥的作用来看，仍有很大空间值得挖掘和提升。

（一）数据来源有限，采集手段落后

从各互联网企业的案例中不难看出，大部分企业在疫情防控中的数据基本来源于政府公开数据，而目前政府公开的数据维度不够丰富，难以满足公众需求。

在大数据时代，公众的信息需求发生了变化，面对疫情，公众关注的重点不仅仅是防控工作动态、自我防护知识、相关政策文件等信息，同时更加关注一些具体的、量化的疫情相关数据¹⁴。

例如，公众既希望获知一个地方总体概况的统计数据（各个省市或区县今天累计有多少疑似、确诊、危重、出院或死亡病例？今天又新增或减少了多少？疑似病例中有多少人被排除了？密切接触者中又有多少人解除了观察？），也希望获知相关病人个体的数据（这些病人分别从哪里来？去过哪里？哪天发病的？在哪家医院就诊？有什么病症？目前情况如何？）。而根据目前政府公开的信

¹⁴ 《哪个直辖市发布的疫情数据最全、最细、最标准？我们做了一个比较分析》，郑磊、陈瑶。

息和企业案例中提供的数据，绝大部分地区都还难以满足这些要求。

此外，一些政府部门和基层组织在采集疫情相关数据时仍然采用手工作坊式的人海战术。在采集时通常采用手工填表、电话问询等方式，这既给基层工作人员增加负担，也无法保证数据的真实性。例如，在采集患者个人信息时，由于没有区分所留手机号为病患或病患家属，可能导致运营商系统核验存在大量身份证号与手机号不一致的结果，这给后续数据分析的准确性和真实性造成不小困难。

（二）数据缺乏治理，数据质量堪忧

从疫情分析的数据来看，目前公开渠道获取的数据不规范，数据口径、数据统计时间等维度不统一，为分析带来很大困难。数据源混杂重复、数据质量不高的情况十分普遍，数据可信性有待提升。

疫情防控分析过程中，每时每刻都在产生大量的数据。一个疏忽错登的身份证号、一条记录不详的家庭住址、一次出行信息的瞒报、一例本人未知的密切接触，都可能成为新疫情触发的导火索。但对大数据的预处理阶段的不重视，数据处理的不规范，很可能导致真实数据获取

难度的增加。而从现有数据上报形式和数据上报的质量来看数据上报过程中还是存在很多不规范不合理的情况。

以上种种原因，导致数据的可用性差，数据质量差，数据不准确。而防控分析大数据的意义不仅仅是要收集规模庞大的数据信息，还有对收集到的数据进行很好的预处理，让数据分析和数据挖掘人员从可用性高的大数据中提取有价值的信息才是最终目标。

（三）隐私保护不足，存在合规风险

随着新冠肺炎疫情防控形势不断升级，各地纷纷采取多样的宣传预防方式，其中，排查上报重点地区返乡人员和确诊患者信息，是一项十分重要的举措，能够帮助卫生监督机构及时掌握情况，迅速切断传播。虽然有关部门第一时间制作了隐去个人隐私信息的确诊患者相同行程查询工具，但在原始数据上报过程中，数据还是遭到了不同程度泄露。不久前有许多载有个人信息文件，突然间在微信、微博等社交平台上疯狂转发，内容包括相关人员的姓名、照片、工作单位、就读学校、家庭住址、手机号码及身份证号等，给返乡人员及确诊患者的生活带来极大困扰，不少人接到骚扰电话和谩骂短信。更是暴露出个人数据保护的不足。

《中华人民共和国民法总则》第一百一十条、第一百一十一条规定：公民享有隐私权，个人信息受法律保护，任何组

织和个人不得非法收集、使用、加工、传输他人个人信息，不得非法买卖、提供或者公开他人个人信息。《中华人民共和国传染病防治法》第十二条也有相关规定，疾病预防控制机构、医疗机构不得泄露涉及个人隐私的有关信息、资料。

在疫情防控的特殊时期，公众知情权、公共安全与公民隐私权需要兼顾平衡。基于疫情防控这一公共利益的切实需要，可以对特殊人群的特定信息进行披露，但应当遵循合规原则，公开内容应确实必要，符合疫情防控目的，公开手段应合理适当，尽可能将负面影响降到最低。

为了为做好新型冠状病毒感染肺炎疫情联防联控中的个人信息保护，积极利用包括个人信息在内的大数据支撑联防联控工作，中央网络安全和信息化委员会办公室 在 2020 年 2 月 4 日发布《关于做好个人信息保护利用大数据支撑联防联控工作的通知》¹⁵，明确为疫情防控、疾病防治收集的个人信息，不得用于其他用途。任何单位和个人未经被收集者同意，不得公开姓名、年龄、身份证号码等个人信息。

通知要求，各地区各部门要高度重视个人信息保护工作，除国务院卫生健康部门依据《中华人民共和国网络安全法》《中华人民共和国传染病防治法》《突发公共卫生事件应急条例》授权的机构外，其他任何单位和个人不得以疫情防控、

¹⁵ http://www.cac.gov.cn/2020-02/09/c_1582791585580220.htm

疾病防治为由，未经被收集者同意收集使用个人信息。法律、行政法规另有规定的，按其规定执行。

通知还要求，收集必需的个人信息应参照国家标准个人信息安全规范，坚持最小范围原则，收集对象原则上限于确诊者、疑似者、密切接触者等重点人群，一般不针对特定地区的所有人群，防止形成对特定地域人群的事实上歧视。

通知明确，为疫情防控、疾病防治收集的个人信息，不得用于其他用途。任何单位和个人未经被收集者同意，不得公开姓名、年龄、身份证号码、电话号码、家庭住址等个人信息，因联防联控工作需要，且经过脱敏处理的除外。

（四）数据流通不足，数据孤岛严重

大数据的核心在于互联互通。只有融会了多源数据，才能发挥大数据的价值。在疫情防控中我们也看到，数据分散割裂、聚而不通、通而不用问题。这很大程度上是因为数据源与数据开发能力不匹配所导致的，即掌握数据的一方缺少足够的数据开发利用能力，而拥有数据开发能力的一方却没有足够的数据库，致使数据无法流通并发挥最大作用。

同时，不同企业和政府部门的数据呈现“孤岛”形态，数据亟需高效整合。全国各级地方政府部门、各大互联网企业利用大数据防疫的挑战之一就是数据的碎片化。

从多个互联网企业案例来看，数据常常散落在不同部门，而且这些数据存在不同的数据仓库中，不同部门的数据技术也有可能不一样，这导致地区内部、企业内部自己的数据都没法打通。如果不打通这些数据，大数据的价值将非常难挖掘，大数据需要不同数据的关联和整合才能更好的发挥优势。如何将数据打通，并且实现技术和工具共享，是更好的发挥大数据在防疫中的更大价值的关键。

五、总结与建议

习近平总书记指出，这次疫情是对我国治理体系和能力的一次大考。疫情防控既是对治理的重大挑战，也是优化治理体系、提升治理能力的重要契机。为加强大数据等新一代信息技术在疫情防控中发挥的作用，本报告提出如下建议。

（一）强积累，发挥数据应用价值

为什么在这次疫情防控中有些地方和部门会出现数据来源不足、采集数据困难重重等问题？可以说，这与平时相关单位工作中没有很好的采集和积累基础数据息息相关。例如，医疗机构如有相应的对患者病情进行实时的数据统计，在出现相似病症患者激增等数据异常之时就可以进行提前的数据预警和疫情的辅助判断。基层部门和社区组织也应当加强社区人员情况摸底调查，建设完善的大数据支撑机制，具备应对突发事件的基础数据能力，避免数据多元重复采集。

（二）提质量，加强数据治理能力

此次疫情过程中，各地公布的疫情数据格式不统一，甚至还出现了转换口径互相矛盾、重新修正等情况，给疫情数据分析带来了很大困难。因此，统一数据标准和口径非常关键。要明确定义数据的来源、数据的格式和数据管理的要求。

同时，不仅要规范数据来源，还要规范数据的处理过程。在保证数据真实性前提下，充分利用现有大数据平台、大数据建设成果，及时有效的实现数据的汇聚融合。

（三）保隐私，确保数据合法合规

疫情防控过程中，在保障公众知情权的同时，也要做好个人信息保护。一是明确收集和使用相关个人信息的执行机构，以及相关工作的流程规范，确保收集的信息仅用于疫情防控和疾病防治。二是在数据共享的过程中，明确各部门与相关单位的共享范围与使用权限，采取匿名化、去标识化、访问控制等措施对数据共享行为进行约束。三是在对社会发布疫情信息时，应以保护当事人个人信息为前提，适度公开相关脱敏信息。四是强化基层信息搜集工作人员的保密意识，从源头把控，杜绝个人信息泄露。

（四）推技术，创新技术应用场景

疫情防控为各种数字技术的应用提供了更多的场景。例如，探索利用机器人实现智能杀菌、发热初筛、隔离病房查房、药物配送等功能，通过智能传感检测技术实现对人体温度精准高效识别等功能。目前，企业在这方面的实践案例还不够多。因此，需要鼓励各方更加高效地利用信息通信技术创新应用，支撑疫情联防联控。一是各地政府应积极利用 5G、

云计算、大数据等技术，进一步完善疫情防控工作。通过建立大数据监测和分析平台，梳理确诊人员的生活轨迹，追踪人群接触史，定位病毒传播路径、预测疫情发展态势。二是鼓励医疗机构、医疗科研单位应用人工智能技术，优化算法和算力，助力在线诊断、病毒基因测序和疫苗/药物等研发攻关。三是鼓励云计算、大数据、人工智能等相关企业加强联合攻关，在疫情发现、分析、预警、防治等方面创新应用，提升疫情防控治疗成效。

（五）寻机遇，推动数字经济发展

疫情既是挑战也蕴含机遇。此次疫情中，无论是各类生鲜外卖、电子商务、在线娱乐及教育平台流量大增催生的“宅经济”，还是随着企业陆续复工复产，互联网和信息技术企业、工业互联网平台开始“大显身手”助力企业尽快恢复“马力”，无不凸显出数字经济的重要性。数字经济的快速发展将有利于保障当前生活生产的稳定有序进行，减轻疫情对我国实体经济带来的负面影响。

新模式新业态成为数字经济投资最活跃领域。疫情期间，居家隔离带动了在线娱乐、在线教育、在线办公、远程医疗等新模式新业态的发展，改变了社会运转模式和人们行为模式。随着疫情缓解，经济社会生产恢复正常，教育、医疗、办公等庞大的在线化市场需求，将吸引大量社

会资本涌入，有望催生出像百度 APP、微信、滴滴等一样的超级应用。

疫情导致大量企业复工延迟，面对招工难、用工贵的现状，很多制造企业不得不采取举措，通过引进生产设备、改革用工模式、提升管理效率、提高工人技能等解决难题。例如，苏泊尔应对招工难问题，通过自动化升级，建设人机混线工作环境，减少人工成本。此次疫情或成为传统企业进行数字化改造提升的助推器，企业受劳动力短缺限制被迫采用数字化手段维持生产，从而打破企业数字化转型“不敢转”、“不会转”、“不能转”的僵局，从而推动产业数字化转型投资实现跨越式发展。

结 语

2月23日，习近平总书记主持召开统筹推进新冠肺炎疫情防控和经济社会发展工作部署会议。总书记指出，当前疫情形势依然严峻复杂，防控正处在最吃劲的关键阶段。无论是分析研判疫情发展态势、精准施行疫情防控策略，还是提供便民应用服务、推动有序复产复工，我们都能看到数据与智能应用在这场没有硝烟的战争中正在发挥关键的作用。在此，我们号召全社会科技企业团结起来，共同发挥数据与智能在疫情防控中的重要作用，坚决打好、打赢这场疫情防控的人民战争、总体战、阻击战，确保全面建成小康社会和完成“十三五”规划，实现我国疫情防控和经济社会发展的双胜利！

致 谢

本报告由中国信息通信研究院云计算与大数据研究所牵头撰写。参与本报告编写的单位有：百度、阿里云、联通大数据、电信云计算公司、北京移动、京东云、滴滴、杭州数梦工场、成都四方伟业、东软、太极、浪潮、华为、华宇软件、航天科工、腾讯、科大讯飞、中航信、第四范式、曙光云计算公司、北京相数科技、厦门渊亭科技、北京相数科技、中移雄安信通、洞见智慧、和智信、朗新科技、美数信息，同事，本报告编写也得到了众多其他单位和专家的大力支持，在此一并对他们给与的支持表示感谢！