

讯飞人工智能和大数据实践

科大讯飞北京研究院院长
王士进 博士



- 成立于1999年，2008年上市
- 国家规划布局内重点软件企业
- 亚太地区最大的语音和人工智能上市公司
- 中国智能语音产业领导者，人工智能产业先行者
- 8000+员工，700亿市值

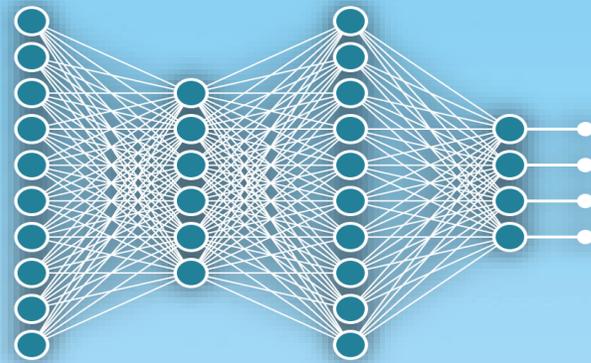
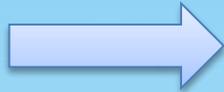


让机器能听会说，能理解会思考，用人工智能建设美好世界





数万小时有标注语音数据



数亿用户语音输入



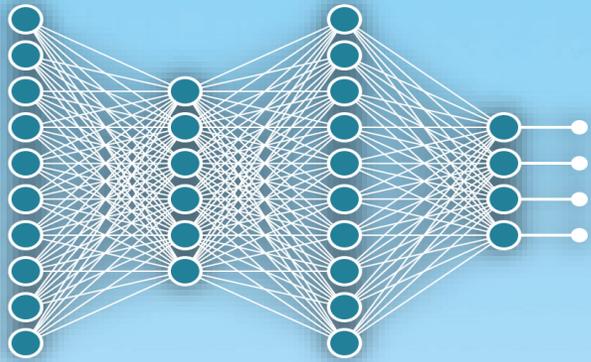
讯飞开放平台

- 处理能力：35亿次/天
- 准确率：超过97%



大数据让人工智能更加智能
人工智能让大数据更有价值

大数据是人工智能成功的三大法宝之一



深度神经网络



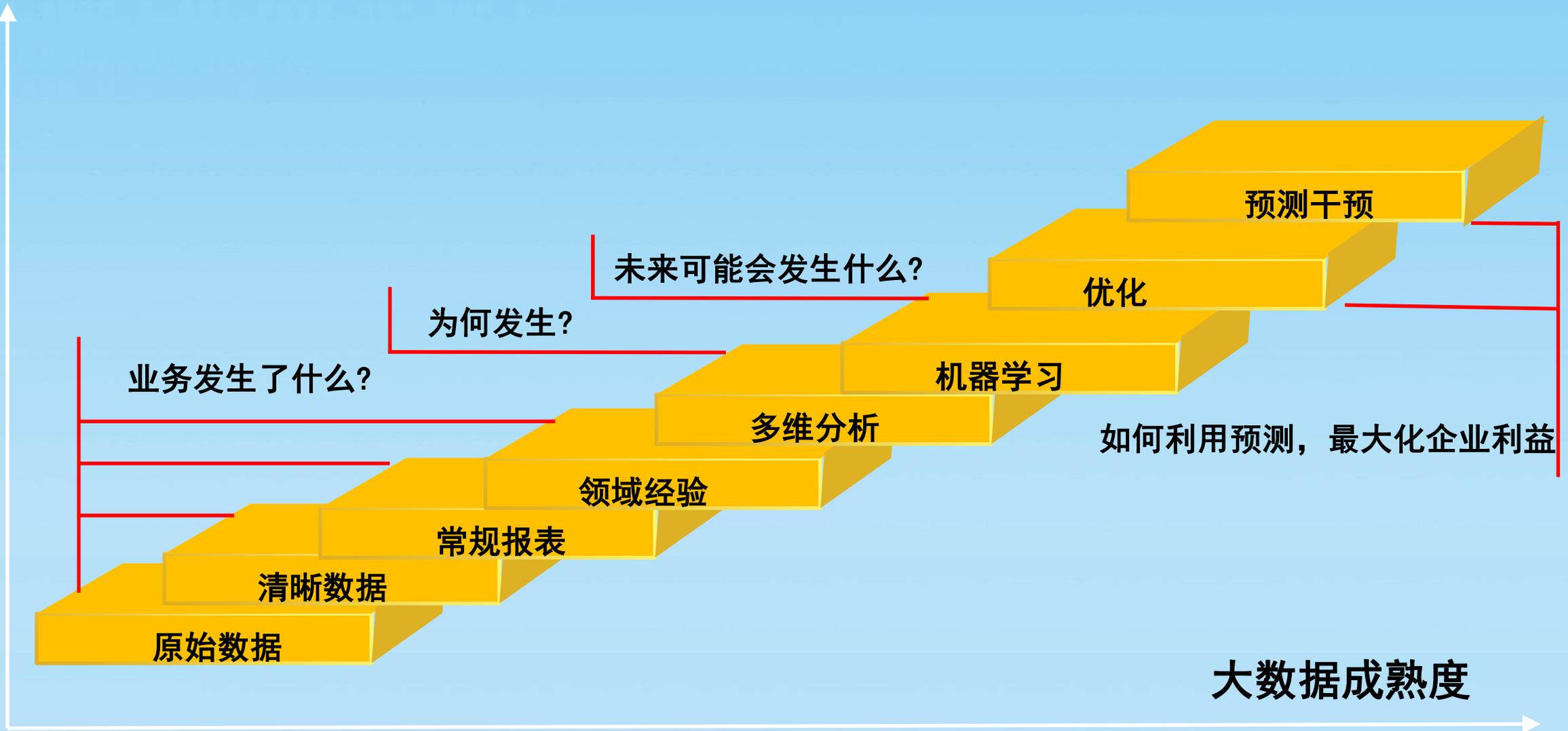
大数据



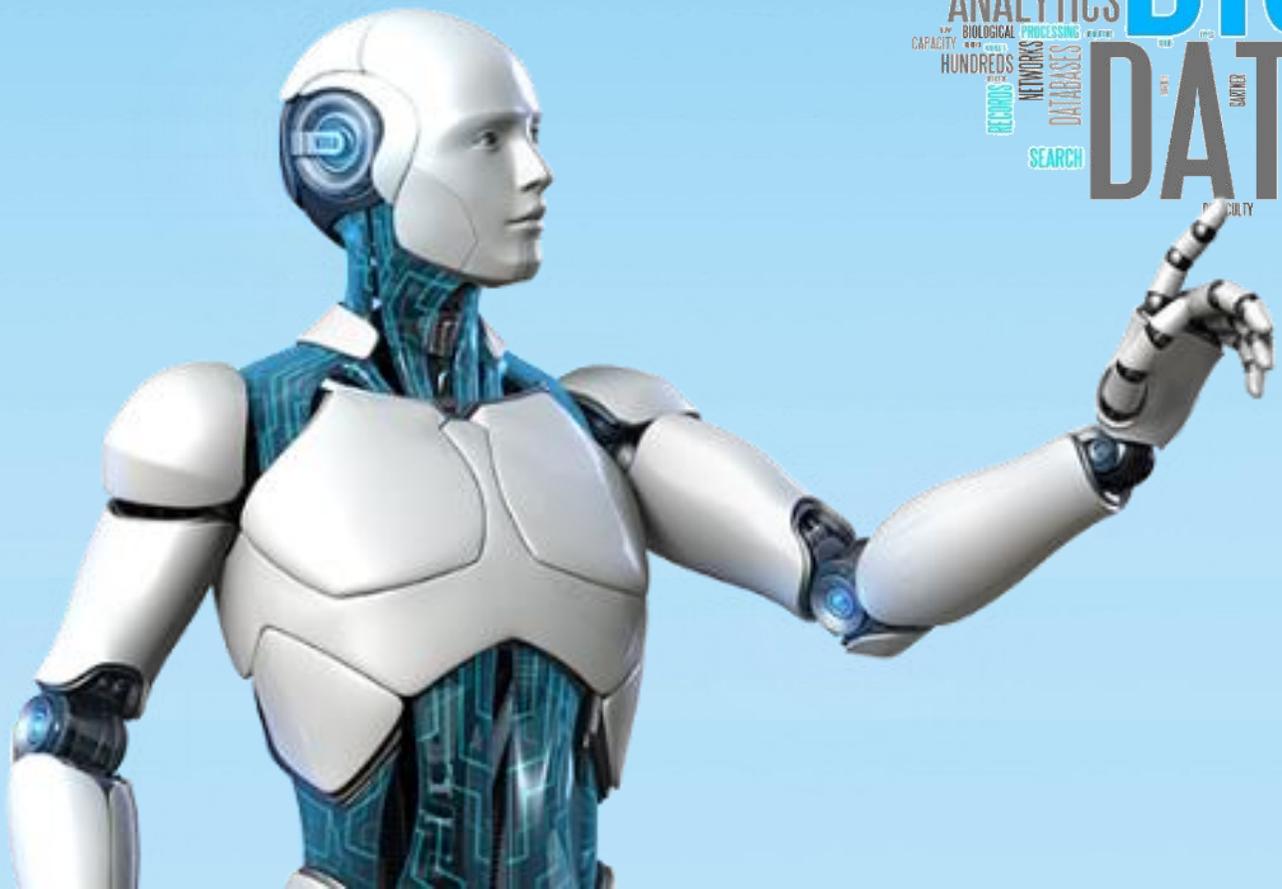
涟漪效应

- 深度神经网络与大数据的结合成为当前主流路径
- 基于互联网和移动互联网的“研究—工程—产品—用户”闭环优化加速了迭代优化进程

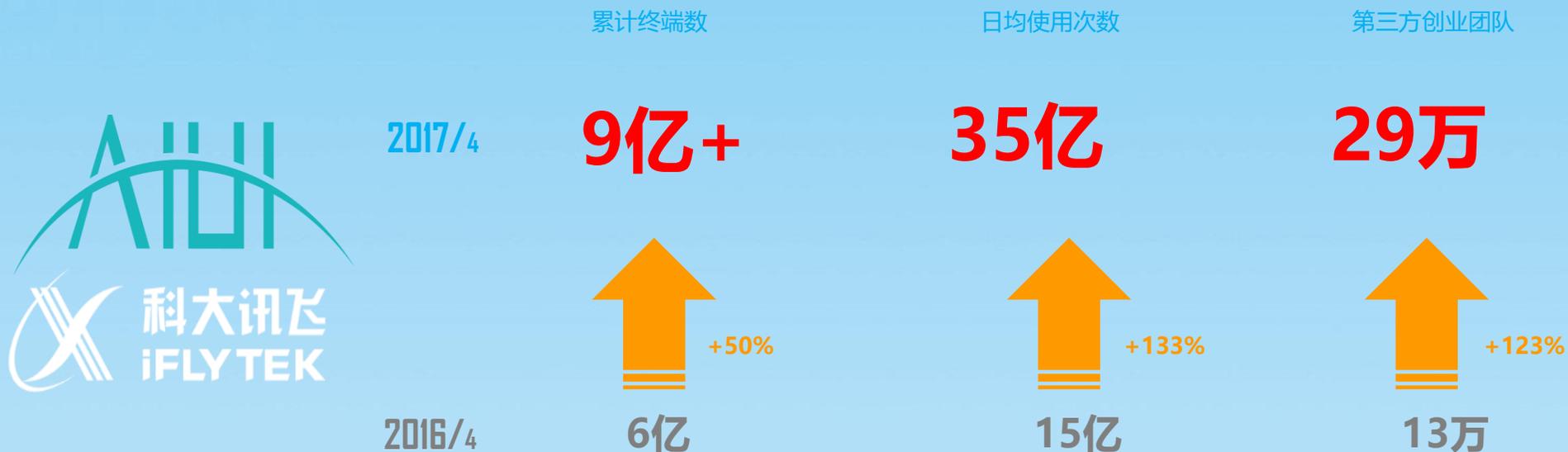
竞争优势



人工智能+大数据是讯飞发展的必然选择







亚太地区最大的人工智能开放平台 (AIUI)

数据来源：讯飞开放平台

讯飞拥有多源场景、真实可靠的大数据资源



汽车语音套件占自主品牌**95%**的市场份额，覆盖**30家**汽车厂商，**100余款**车型



讯飞输入法**4亿**用户，月活超**1亿**
灵犀语音助手国内市场占有率**第一**



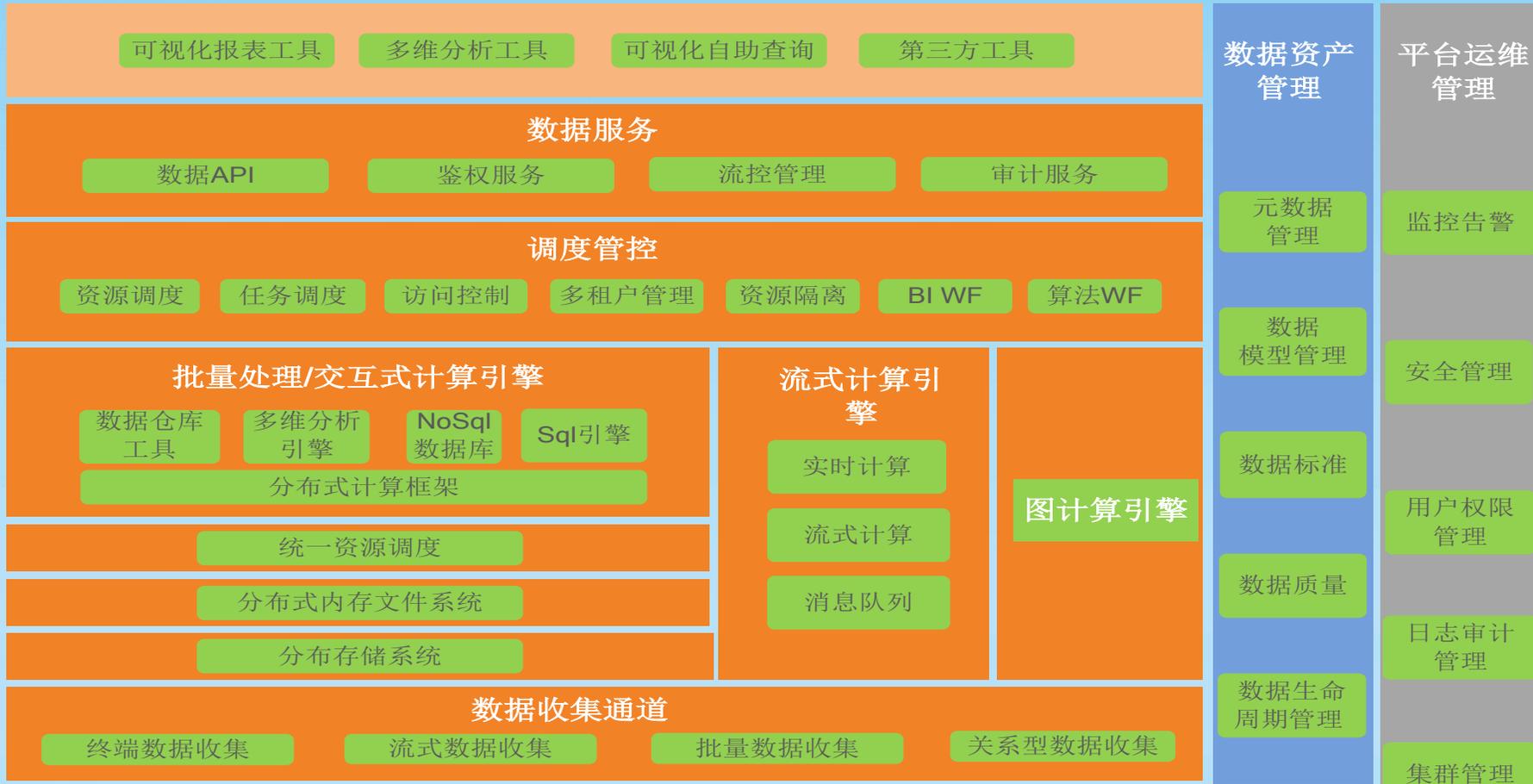
智能家居解决方案服务**3000+**合作伙伴，**1000+**智能产品，超**80%**市场占有率



讯飞教育产品服务**10000**所学校，**8000万**师生，包括**57所**全国百强校

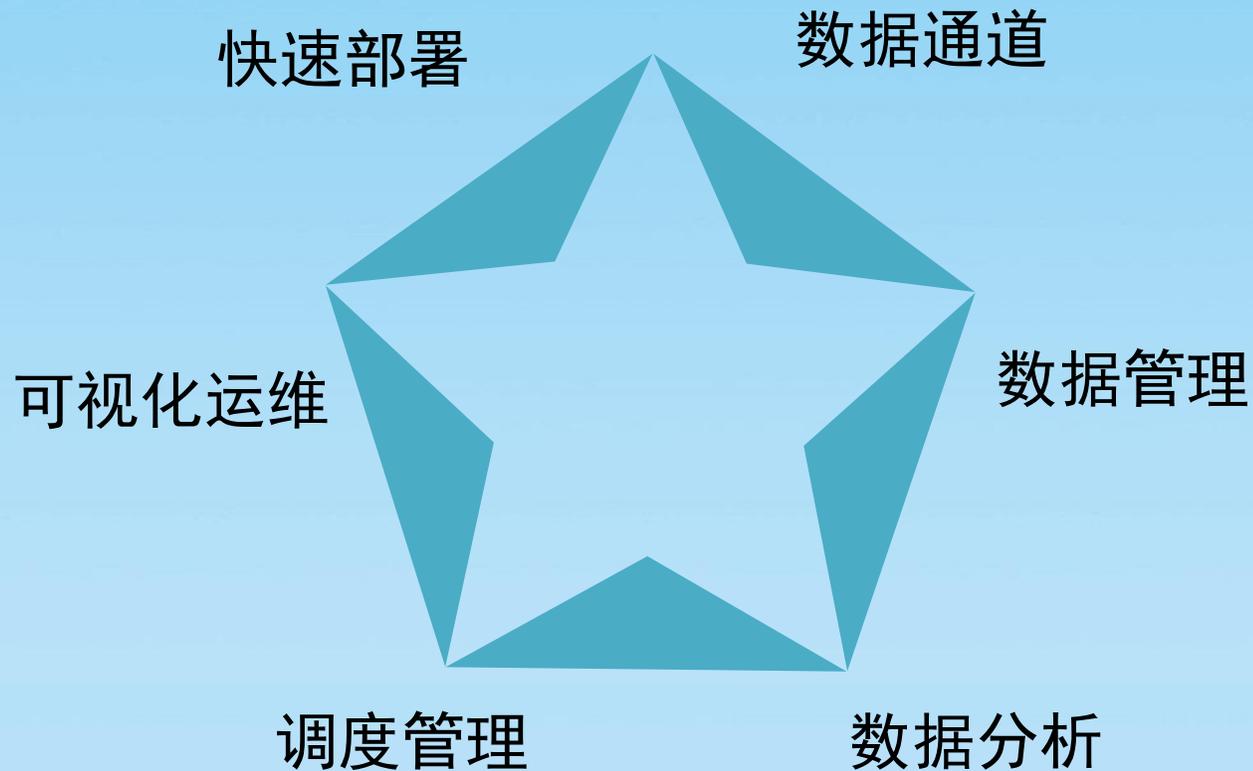


Odeon(奏乐堂)：日均15000任务、100TB数据，演奏出美妙的数据音乐



平台实现数据闭环完整能力：收集、存储、计算、数据服务

Odeon平台核心能力



数据通道：SQL、NoSql数据库及日志多源数据接入与导出

数据管理：完善的元数据管理、数据质量标准管理体系，具备完整的数据隔离及授权机制

数据分析：支持离线批处理、流式计算、OLAP、AdHoc等多种计算方式

调度管理：可视化 workflow 构建，基于Time、Event等触发机制，全程状态监控

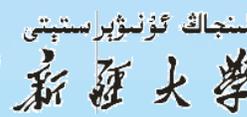
可视化运维：物理集群、分析作业、workflow、资源使用情况等全方位可视化监控及告警

快速部署：平台开发及部署采用业界主流DevOps模式，全组件镜像化，快速部署运维

• 高水平大数据领军人才

- 鄂维南院士（北京大数据研究院院长，北京大学和美国普林斯顿大学教授）
- 熊辉教授（美国罗格斯大学终身教授，长江学者，海外杰青）
- 陈恩红教授（中科大计算机学院副院长，杰青，CCF大数据专家委员会副主任）

• 广泛的产学研合作

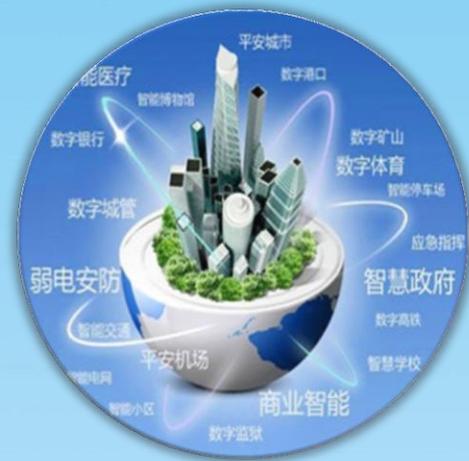




精准营销



个性化学习

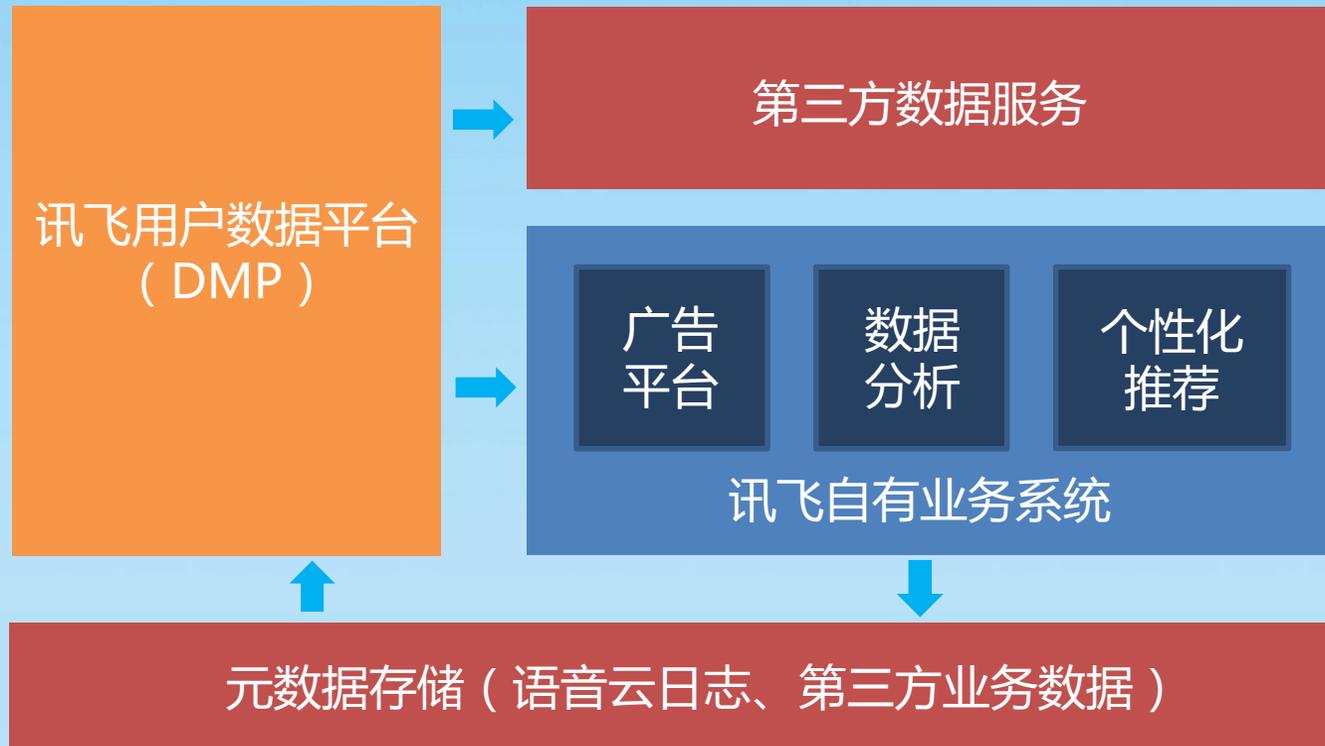


智慧城市

- 讯飞已经通过用户画像、计算广告和个性化推荐实现自有大数据价值挖掘闭环
- 讯飞大数据能力赋能不同行业，助力各行各业真正智慧化

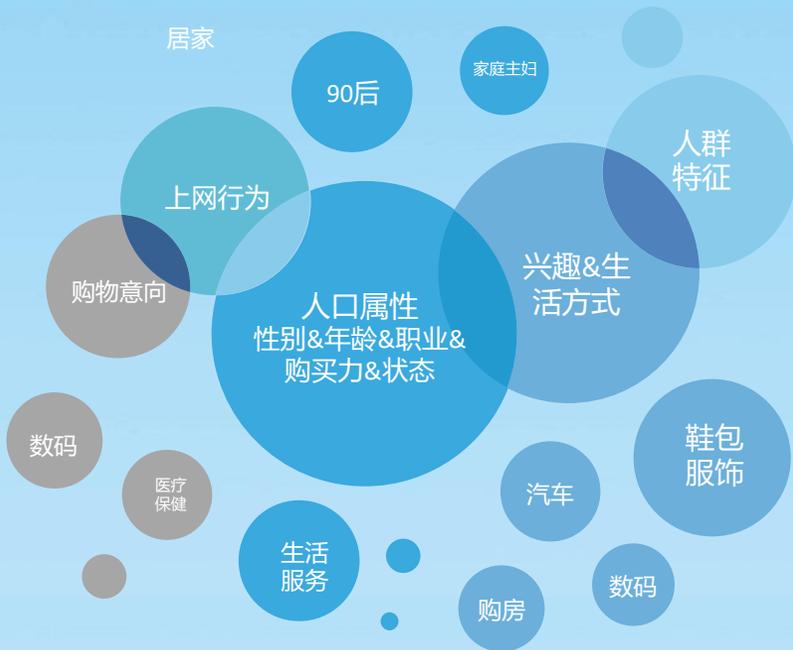
能力输出

核心价值



用户画像实现数据增值

结合讯飞大数据与人工智能技术构建的讯飞DMP平台，已经完成人生阶段、行业偏好、购物兴趣、媒介兴趣4个大类**1700个子标签**的用户填充工作。累计覆盖12亿终端设备。



✦ 数据管理平台

 人群洞察 | 标签视图 | 标签查询 | 标签管理 | 标签整合

 欢迎 bdwei | 注销

人口统计
人口统计数据
地域行为
APP行为
行为&兴趣

性别和年龄

年龄 不限 自定义

17岁以下

18-22岁

23-30岁

31-40岁

41-55岁

55岁以上

性别 不限 女性 男性

婚恋状态 不限 自定义

单身 有伴侣 已婚

生育状态 不限 自定义

孕期 育儿

男性:709514672人
所选人群:65.76%
DMP人群占比:65.76%

女性:369384390人
所选人群:34.24%
DMP人群占比:34.24%

年龄组	百分比
17岁以下	1.94%
18-22岁	5.26%
23-30岁	20.59%
31-40岁	26.29%
41-55岁	31.99%
55岁以上	13.93%

开始洞察
重置

快捷展示

版权所有 © 科大讯飞股份有限公司 皖ICP备05001217号

基于人工智能技术构造精细化用户标签

例：基于讯飞人工智能和自然语言处理技术，有效区分购车/有车语义，实现购车意向人群和有车人群的精细划分

购车意向



有车人群



计算广告帮助广告主实现受众精准营销



客户

某金融理财产品推广



人群定向

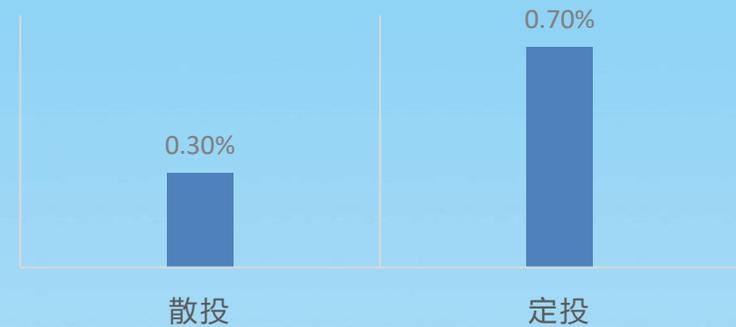
男性
广州地区
第三方网贷投资类产品用户群体



转化方式

有效注册 (预留电话)

页面点击率

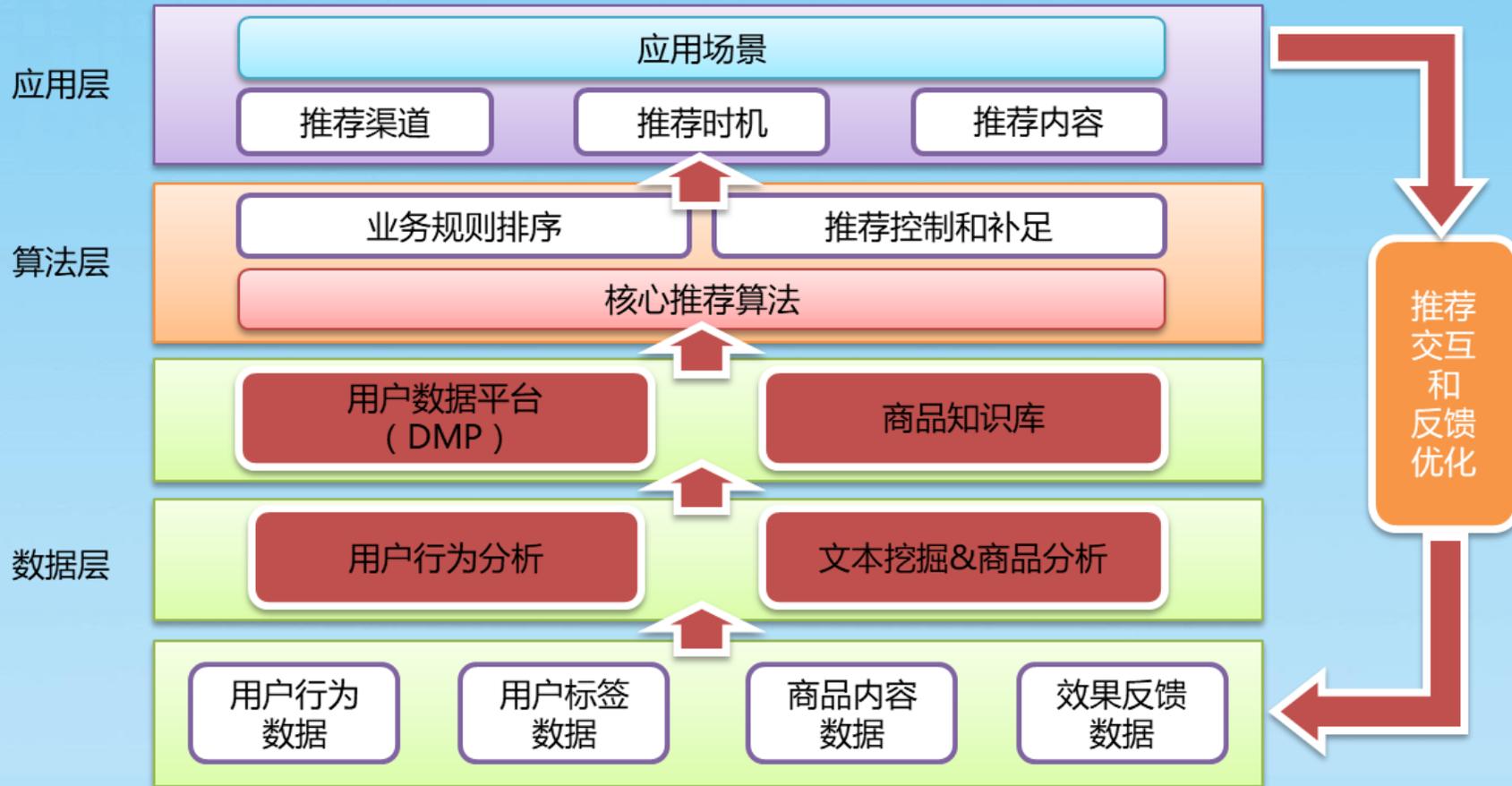


有效注册转化



点击率提升 **130%**，有效注册转化提升 **187%**

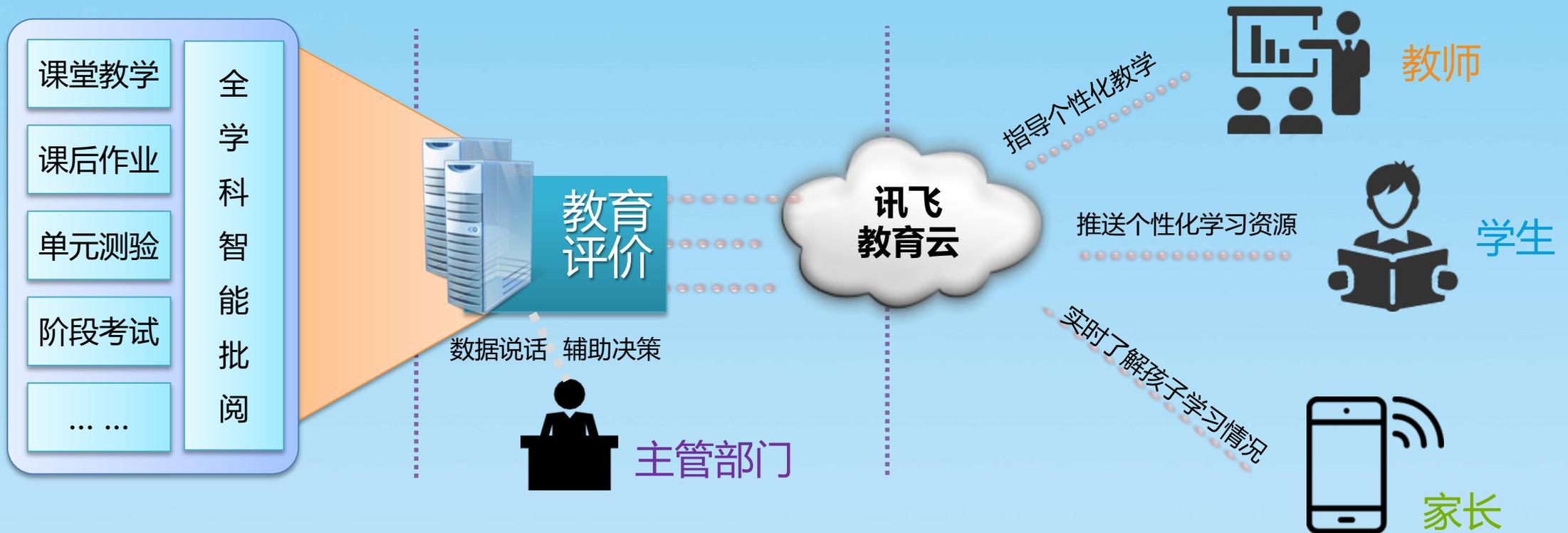
个性化推荐深度挖掘用户需求



讯飞酷音铃声：推荐场景覆盖多个交互场景，业务开通量提升**300%**

流量包推荐：下单率提升一倍

教育大数据可应用于教学过程各个环节



1 教学过程数据
采集与汇聚

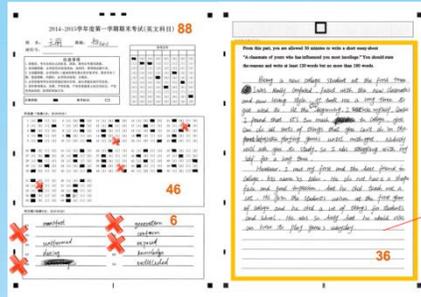
2 教育大数据
自动分析与评价

3 应用大数据实现
个性化教学

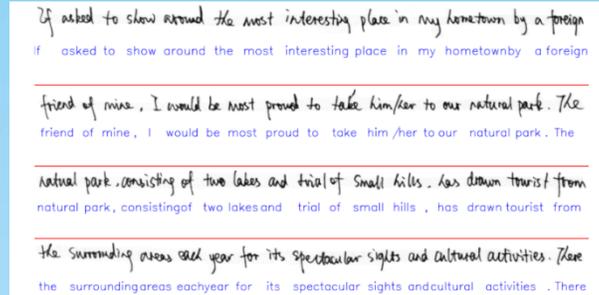
扫描试卷

识别结果

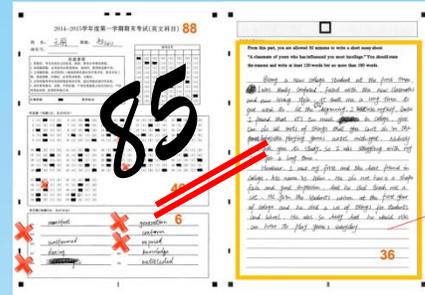
成绩



文字识别



自动评分



江苏省高考智能阅卷验证

	语文作文一致率 (分差 < 7分)	英语作文一致率 (分差小于5分)
机评-报道分	92.82%	94.57%
人1-人2	87.36%	82.85%

湖南省研究生智能阅卷验证

大作文一致率 (满分20)	分差 ≤ 2分	分差 ≤ 4分 (仲裁分差)	分差 ≤ 6分
人机一致率	92.78%	98.69%	99.83%
人人一致率	63.92%	86.85%	95.60%

教育部考试中心-科大讯飞人工智能联合实验室



全学科阅卷让学习过程数据全量采集成为现实

姓名：王小明
学号：121302

数据标签

作答情况
难度
知识点
解题方法
.....



过程化

粗心?
不熟?
不会?

北师大二附中

高三年级270名学生，语文作文使用一学期

40小时

3分钟

精批时间

周频

月频

使用频率

提升15%

分数提升



十万套密卷
400万道试题

习题精练



视频

“解析几何”
出问题需要从
“二元二次方程”补起

1、3、7、11、13

- 何>>圆锥曲线>>椭圆
- 何>>空间几何体的三视图与直观图
- 7、函数与导数>>导数的综合运用
- 11、统计与概率>>排列组合与二项式定理
- 13、不等式>>线性规划

教学产品覆盖师生数**8000万**

10000所学校应用，惠及**1500万**师生

68所全国百强校——（参考教育部发布“中国百强高中名单”）



人民大学附属中学



北师大二附中



杭州二中



江苏省启东中学



合肥市第一中学



河北衡水中学



呼和浩特市第二中学



华东师大二附中



合肥市第八中学



厦门外国语学校



深圳中学



马鞍山二中



邯郸市第一中学



华南师大附属中学



重庆育才中学



江苏省天一中学



山东省实验中学



武汉外国语学校



东北师范大学附属中学



福州三中

跨部门数据共享提升政府服务效率

便民、便企数据全部打通，政府权利清单全面覆盖

- 业务：
 - 打通公安、民政、教育、城管、市场监管等**45个部门715类数据**；建立完善的人口库和法人库。
 - 提供便民便企上网事项**1948项**；社区一站式服务事项**14个部门231项**。
- 效率：
 - 窗口由原来8-10个减少到现在2-3个，数量减少**70%**，前台办理人员**减少50%**
 - 事项办理由原先10-30天，缩减到现在1-5天，**办结率99.7%**，越来越多的事项当场办结



吴全荣是安徽省亳州市一位退休职工，两年前到北京和儿子住在一起，但每年要回亳州办理社保认证始终是她的心事。“不认证的话，社保就停发了。”吴全荣说，虽然车程比较远，她依然要专门跑一趟。今年老人不用再跑了。

原来，每年4至6月，全市的社会保障长期待遇领取人员都要集中进行资格认证。今年，老家的社区干部告诉吴全荣，通过手机网上办事大厅就可以进行社保认证。得知这个消息，吴全荣让儿子帮她在手机上下载并注册了亳州市网上办事大厅手机客户端。“我就对着手机摇摇头、眨眨眼，很快就认证好了。”吴全荣说，从下载到认证成功，也就几分钟的时间，“以后再也不用专门跑回家去认证了”。

现在在亳州，群众足不出户，只需要点点鼠标或手机，就可以办理就业失业登记、企业设立登记、公交卡等数千项业务。这些便利得益于亳州市政府建设的网上办事大厅，将政务服务由“实体大厅”向“网上大厅”迁移，从“办事窗口”向“口袋大厅”延伸，实现了“群众点键，数据跑腿”，让全市600多万群众享受到“互联网+政务”的便利。



过去

- 每年集中一次
- 本人必须到现场

办个事排了一上午的队

社会保险长期待遇人员资格认证



现在

- 随时随地
- 网上认证

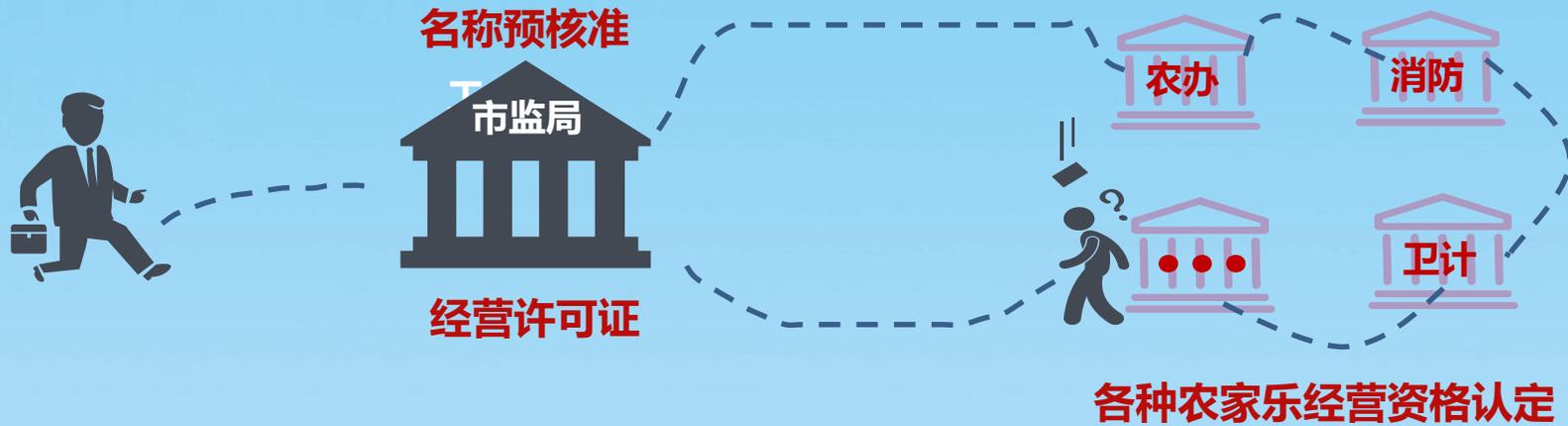
声纹+人脸

现在手机认证轻松搞定

政府数据共享让企业少跑腿

杭州市农家乐经营许可证申请

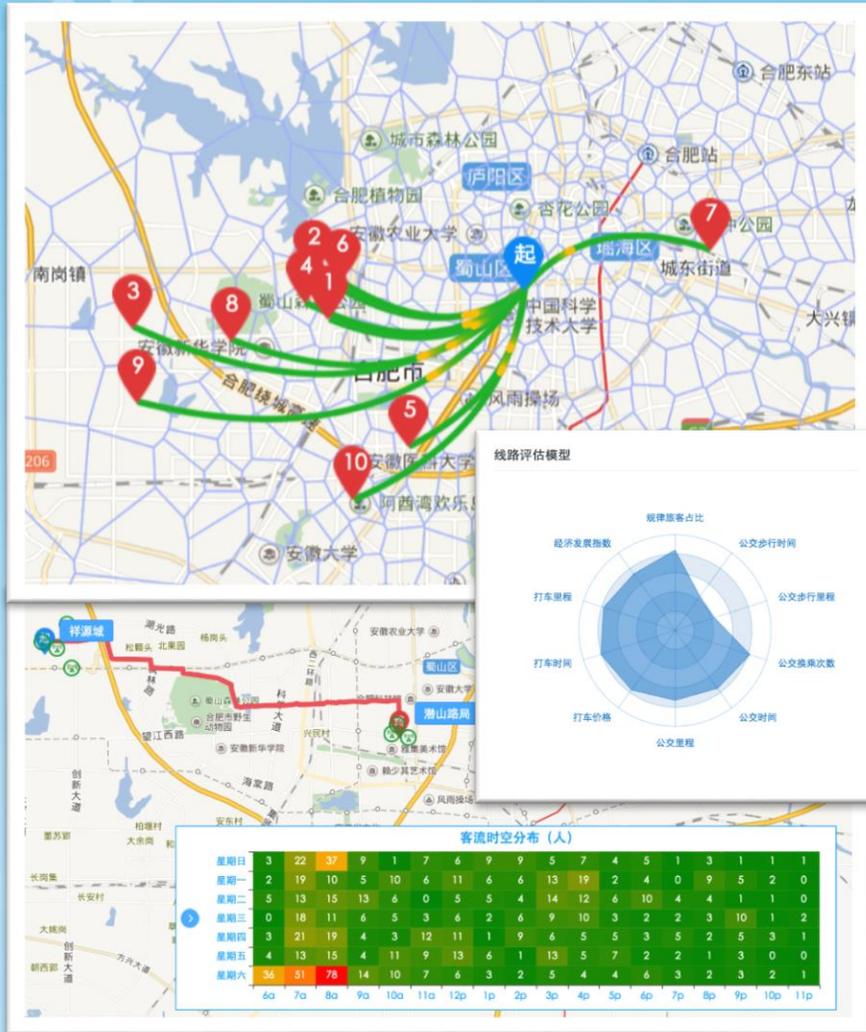
线下办
理模式



业务协同
办理模式



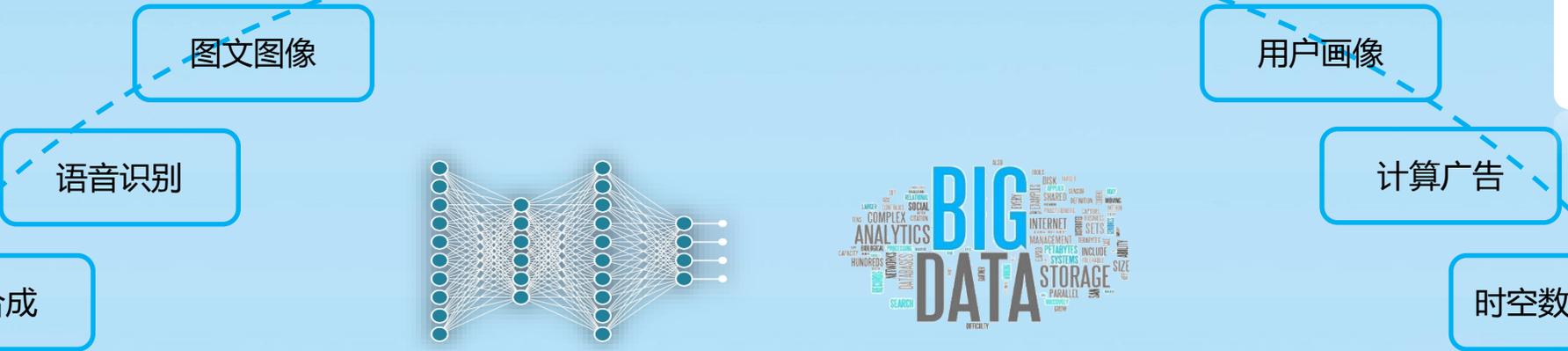
讯飞完成“数字杭州”顶层设计方案，用人工智能打造城市数据闭环



运营商、公交、交管数据融合，发现高频稳定客流，定制和优化公交线路



国土、规划、房地产数据融合，事先评估功能建筑建成后周边影响，实现科学规划



智能客服



智慧城市

AI改变世界，数据创造价值