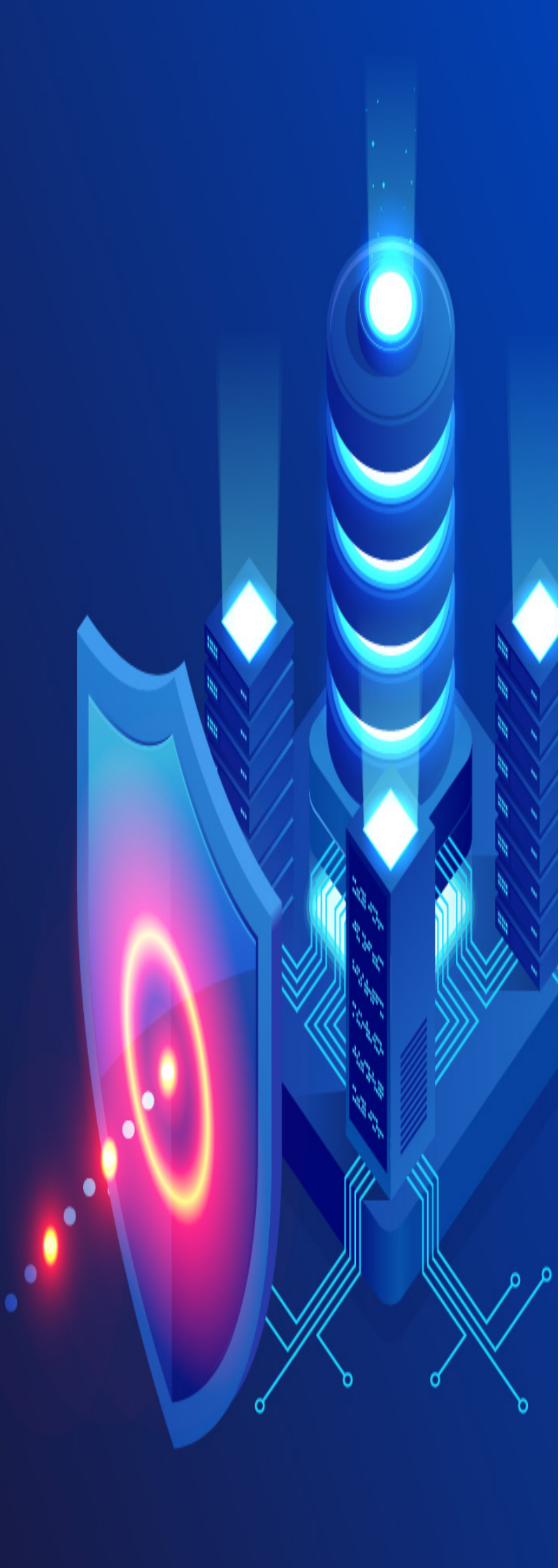


专家洞察

洞察先机
防患未然

基于人工智能的
电话反欺诈行动指南

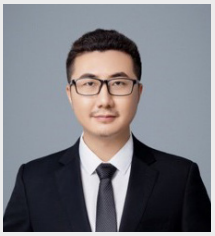
IBM 商业价值研究院

IBM

主题专家



吴大维
IBM GBS CBDS 团队
副合伙人
wudavid@cn.ibm.com



杨杭
IBM GBS CBDS 团队
人工智能解决方案负责人
首席业务咨询顾问
bjyhang@cn.ibm.com



钱雪莹
IBM GBS CBDS 团队
高级数据科学家
bjqianxy@cn.ibm.com



高康睿
IBM GBS CBDS 团队
高级数据科学家
gaokrui@cn.ibm.com



郝希蓓
IBM GBS CBDS 团队
资深数据科学家
haoxish@cn.ibm.com



何佳惠
IBM GBS CBDS 团队
高级数据科学家
hjhjhui@cn.ibm.com



王莉
IBM 商业价值研究院
高级咨询经理
gbswangl@cn.ibm.com

扫码关注 IBM 商业价值研究院

官网　微博　微信　微信小程序

谈话要点

数据技术联动，完善预警系统

数据驱动，技管结合，不断提升及时预警电信诈骗行为的能力。

明确建设方向，开展省级试点

对各省电信运营商而言，电信反欺诈发力的重点不同，应根据自己的需要确定实施路线。

挖掘欺诈套路，建立长效机制

挖掘欺诈分子常用的欺诈套路，洞察先机，防患于未然，将变化多端的欺诈套路扼杀在摇篮之中。

一

知己知彼，百战不殆

一通电话、一条短信，一个链接，一不小心，钱可能就没了；欠费了、中奖了、退税了、口罩来了，电信诈骗的骗术层出不穷。

尤其是在 2020 年上半年，受疫情影响，就业压力加大，有的资金短缺的用户，被贷款诈骗；有的找不到工作的用户，被兼职刷单诈骗；一些经常网购的用户，被冒充客服和虚假购物诈骗；一些有投资意愿的用户，被引诱参与虚假投资理财和网络赌博；冒充公检法实施的诈骗案件也时有发生。

公安部通报，2019 年全国共破获电信网络诈骗案件 20 万起、抓获犯罪嫌疑人 16.3 万人，同比分别上升 52.7%、123.3%。[1] 2020 年上半年全国共破获电信网络诈骗案件 10.1 万起，抓获犯罪嫌疑人 9.2 万名，同比分别上升 73.7%、78.4%。[2]

诈骗电话的治理已成为电信运营商不可忽视的工作重点。虽然三大运营商也都进行了诈骗活动识别的探索，但因其复杂性，仍是电信行业的顽疾之一：

诈骗内容花样繁多，严重影响用户满意度。疫情期间“新热点 + 旧诈骗手法”的出现也告诉我们，与诈骗内容的对抗将是一场持久而关键的战斗。如何在覆盖现有诈骗方式的基础上，利用数据与 AI 技术的力量，与时俱进，及时有效地识别新的诈骗内容？

现有的分析手段效果不佳。传统的黑名单模式的时效性和准确率不足，且需要持续运维投入，难以及时有效地识别并干预电话诈骗行为。如何加大对电话渠道的监控，实现联动分析，及早识别并预警劝阻？

整体解决方案不清晰，不知从何入手。反欺诈领域的技术能力已初步形成，如何综合成本、反欺诈准确率、法律风险、处理能效等关键要素，形成整体的解决方案？

2020 年 8 月 19 日工信部印发了《关于运用大数据推进防范治理电信网络诈骗长效机制建设工作方案》。《方案》明确，按照“整体推进、分步实施、数据驱动、技管结合、务求实效”的总体思路，坚持数据融合、数据驱动和数据共享，加快推进大数据反诈长效机制建设。[3]

电信运营商应充分利用手中的通信渠道和通信内容等数据资源，结合 AI 分析手段，建立和完善电信反欺诈预警系统，不断提升精准预警尽早干预的能力，以最大程度降低用户损失，提升用户满意度。

诈骗电话识别与干预

人工智能在反欺诈领域的技术能力已经初步形成并在生产应用中不断迭代更新。在电话反欺诈建设领域，需要考虑以下四点：

综合成本：包括软件成本（反欺诈平台）、硬件成本等

反欺诈准确率：要求在尽可能抓住坏人的同时尽量少的冤枉好人

法律风险：在法律支持的数据、操作范围内进行处理，不触犯红线

处理能效：满足一定量的并发能力支撑以及平台横向扩展要求以及满足单次判断处理时间要求

然而，目前不存在满足综合成本、反欺诈准确率、法律风险、处理能效四个关键要素的完美解决方案。业界方案普遍需要在关键要素中取舍平衡。

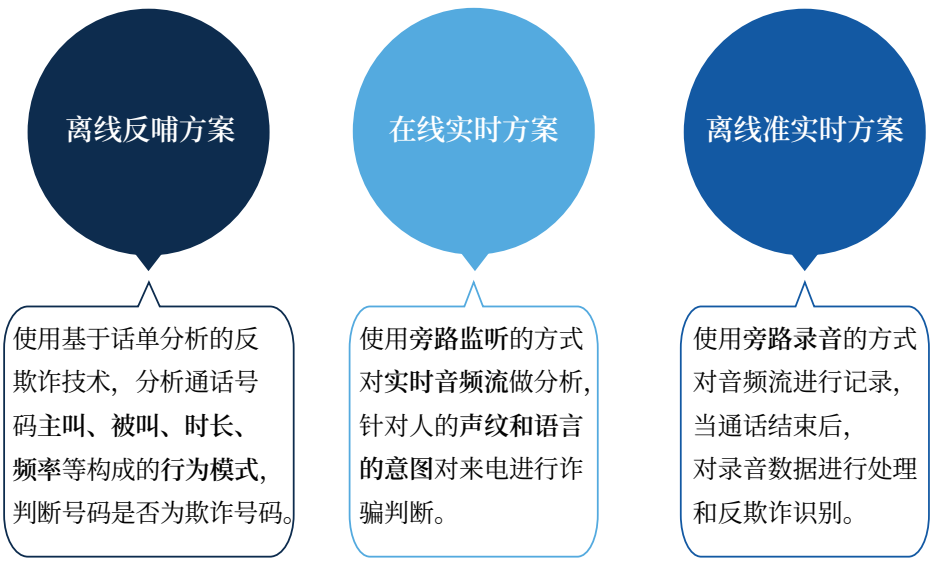
目前主流方案分为三种：离线反哺方案、在线实时方案、离线准实时方案（见图 1）。

这三种主流的反欺诈方案，各有优劣，电信运营商可以根据自身的需求进行选择：

	优势	劣势
离线反哺方案	- 充分利用电信大数据优势，在防止区域性欺诈风险上起到很好的作用	- 对诈骗行为的反馈较慢
在线实时方案	- 及时干预，及时止损	- 与其他两种方案相比，技术难度最大 - 需考虑系统的并发性和稳定性
离线准实时方案	- 与在线实时相比，该场景下，人工智能模型可根据完整对话内容进行判断，可提高识别准确率	- 录音文件的存储与处理需要大量的硬件资源

在与电信运营商的合作过程中，我们发现离线方案反应较慢，无法及时止损，事后欺诈分子转移较快；在线实时反欺诈可形成电信服务产品，供个人进行开通选择，避免法律风险。鉴于以上原因，电信运营商纷纷开始实时电话反欺诈建设。

图 1
三种主流反欺诈方案



欺诈话术套路分析与洞察

在电信反欺诈领域，与短信欺诈、网络欺诈相比，电话欺诈的核心在于复杂的欺诈话术套路能迷惑人心。所谓的欺诈话术套路，是指欺诈分子在电话中首先说了什么方面的内容，然后又说了什么方面的内容，接着又采取了什么样的话术技巧迷惑对方，直至最后成功拿到对方关键信息。正是套路的复杂性，容易让人上当受骗。

实际上，电话反欺诈并不是一个新鲜的课题，一些电信运营商在过去多年通过敏感关键词列表（如“我是公安局的”）等方式也进行过尝试。

根据 IBM 多年反欺诈的经验，电信欺诈分子常以团伙的形式出现，欺诈分子集中培训欺诈话术套路。故而，欺诈分子可在同一个欺诈话术套路下通过变换关键词等绕开此反诈机制。同时，因为语音转写质量的问题，往往个别错别字的出现，也会造成此反诈机制失效。电信运营商通过人力维护更新反诈关键词列表费事费力且反应滞后。

因此，对频繁出现的欺诈套路进行挖掘，不但能够揭露隐藏的欺诈团伙，同时，将这些典型的欺诈套路加入反欺诈机制中，也可以增加欺诈分子未来欺诈的难度。

除了已发现的复杂欺诈套路外，IBM 发现，欺诈分子也在原有的话术套路上迭代更新使套路更复杂。通过对欺诈套路的分析，还可以提前洞察未来欺诈套路的可能变化，提前预防。基于过去多年在电信行业与金融行业的反欺诈经验，IBM 建立起了一套独有的复杂欺诈套路识别资产。可从海量的欺诈电话记录中快速给出频繁出现的复杂欺诈套路。

某电信运营商识别发现最新欺诈套路 [4]

2020 年的疫情对经济造成了严重影响。在经济下行的同时，欺诈分子也蠢蠢欲动。某电信欺诈频发省份，据统计，在疫情最严重的 1-3 月，电话欺诈成功率高达 18%。

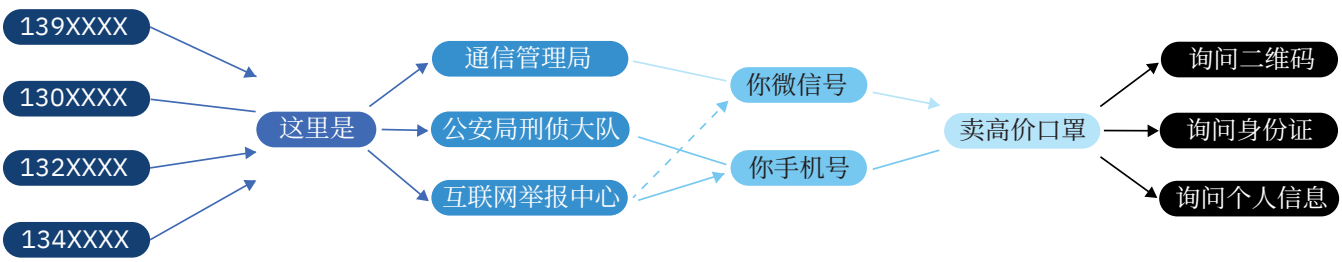
该省公安、电信运营商与 IBM 合作，利用 IBM 复杂欺诈套路识别资产，从 20 万余通疑似欺诈的电话中，半自动化挖掘出 10 种频繁出现的典型欺诈话术套路。典型欺诈套路示例如图 2 所示。

我们发现，该欺诈套路在某市频繁出现，有理由怀疑手机号 139XXXX、130XXXX、132XXXX、134XXXX 是同一个欺诈团伙，建议作进一步核实。

另外，我们发现欺诈分子往往冒充某国家机构（如通信管理局），且污蔑对方通过手机号或微信号倒卖高价口罩。故而可以预测，未来，欺诈分子可能会在更多的国家机构名以及倒卖其他某种紧缺商品上做文章。

该欺诈套路的挖掘，一方面，用于提高反欺诈模型对未出现的欺诈电话有更稳定的识别准确率；另一方面，补充到欺诈套路库，当类似欺诈正在进行时，匹配到该欺诈套路，推送至公安进行进一步审查核实。

图 2
典型欺诈套路示例



某电信运营商搭建电话实时反诈平台，识别诈骗电话，解决社会难题[5]

某省作为全国经济大省，饱受电信诈骗所扰。2019 年该省的诈骗案件数量、被骗金额总量均为全国第一。由于诈骗问题存在已久，该省电信运营商大数据中心已于 2016 年推出天盾服务用于诈骗电话识别，但诈骗问题仍然严峻。

电话欺诈分子极其狡猾，骗取钱财后迅速转出，给公安等机关追查造成了很大的困难，该电信运营商与 IBM 合作，建立在线实时智能反欺诈平台，利用人工智能技术，实时识别欺诈分子；同时，为了避免法律风险，平台将电话内容遮掩，只将欺诈摘要信息实时反馈给公安机关单位，助其提升诈骗案件监管力度与执行效率。

该项目旨在降低诈骗成功率，重点进行实时诈骗行为拦截，确保在诈骗分子成功实施诈骗之前，完成诈骗判断、通话拦截、受骗劝阻的过程。利用 AI 技术解决社会难题，实现企业社会责任感。

目前平台在试运行阶段，对欺诈电话的识别准确率达到 95% 以上，对欺诈通话与欺诈手机号判断准确性显著提升，并通过 24 小时监控干预，自动挂断与警示功能直达潜在受害人，成功降低区域诈骗成功率 50%，提升了社会公信力。

电话反欺诈行动指南

结合 IBM 在反欺诈的经验，我们总结出四条关键的建设指南，供电信运营商参考。

1. 明确反欺诈建设方向

在电话欺诈高发的省份，建议电信运营商与公安、通信管理局充分做好反欺诈建设的沟通。根据公安和通信管理局的要求进行建设。同时，对反欺诈模型的要求应遵循“尽量在不冤枉好人的情况下抓住坏人“这一原则，减少对正常电话的干扰。

若公安选择电话欺诈监测但事后核实查证再采取行动，对反欺诈模型的要求应遵循“可以暂时冤枉一批好人，但应尽量覆盖更多的坏人”的原则。此原则保证了在事后的检查证核实中，“好人”会被排除掉，而“坏人”可以一网打尽。

另外，欺诈分子往往不会只尝试一次就停止欺诈行为。而且电信欺诈往往以团伙的形式存在。离线的大数据结合人工智能可以对实时反欺诈起到很好的反哺作用。通过电话主叫被叫大数据记录，可以构建人际关联网络，挖掘出欺诈分子、骚扰者等。

2. 保护个人隐私是应尽的义务和责任

反欺诈的建设，往往绕不开对个人隐私保护这一大前提。保护个人隐私不但要满足监管机构的要求，而且，从架构设计的角度，也不允许设计任何数据库来留存数据。当通话内容经过 AI 模型判断是否欺诈后就会自动丢失，不会保留下来，从而实现对个人隐私的保护。技术测试成熟后，反欺诈业务可作为一项订购业务，由用户自己进行选择是否打开反欺诈功能。

3. 高质量的语音转写是高效识别欺诈的必要条件之一

当前，主流的语音转写提供商多是在极其理想的环境（无噪声、标准普通话等）下测试语音转写质量。但这样的假设往往不现实。当在真实外部世界中，并且有较严重口音的情况下，语音转写质量就会大打折扣，甚至出现转写出来的绝大多数内容都是错别字，人类无法理解。

IBM 建议，在当地口音较严重的省份，语音转写提供商应根据运营商的要求，在真实的环境中完成测试，优化语音转写质量。甚至在有的省份，各地级市各区县也有不同的方言口音。即使在同一个省，也因根据地区的不同，进行语音转写质量的优化工作。

4. 丰富的反欺诈模型构建经验是项目成功的必要保证

在电话反欺诈领域，往往极少量欺诈样本隐藏在海量普通的通话样本中，利用人工逐条听录音并不现实，这就对建模效果造成了很大的挑战。对此，数据科学家应结合自然语言处理手段与外部公开数据，将那些极少量的欺诈样本挖掘出来，形成建模基础。

另外，数据科学家应对电信反欺诈有深刻的理解和认识，利用 AI 技术更好地识别出欺诈分子的复杂欺诈套路。所谓“魔高一尺，道高一丈”，欺诈与反欺诈永远是在相互博弈的过程中进化。故而，反欺诈系统应形成定期迭代更新机制，能够识别最新的欺诈套路。

需要思考的重要问题

— 在电话反欺诈领域，贵企业遇到了哪些问题或挑战?

— 您打算如何构建反欺诈平台?

— 您是否打算联手经验丰富的合作伙伴，加速向前推进?

备注和参考资料

1. 熊丰，“2019 年全国共破获电信网络诈骗案件 20 万起，抓获犯罪嫌疑人 16.3 万人”，新华网，http://www.xinhuanet.com/legal/2020-01/21/c_1125491558.htm
2. 朱紫阳，“公安部：上半年全国共破获电信网络诈骗案件 10.1 万起”，人民网，http://legal.people.com.cn/n1/2020/0728/c42510-31800812.html
3. “关于运用大数据推进防范治理电信网络诈骗长效机制建设工作方案”，工业和信息化部，http://www.miit.gov.cn/n1146285/n1146352/n3054355/n3057724/n3057728/c8056526/content.html
4. IBM 案例研究
5. IBM 案例研究

选对合作伙伴，驾驭多变的世界

在 IBM，我们积极与客户协作，运用业务洞察和先进的研究方法与技术，帮助他们在瞬息万变的商业环境中保持独特的竞争优势。

IBM 商业价值研究院

IBM 商业价值研究院 (IBV) 站在技术与商业的交汇点，将行业智库、主要学者和主题专家的专业知识与全球研究和绩效数据相结合，提供可信的业务洞察。IBV 思想领导力组合包括深度研究、专家洞察、对标分析、绩效比较以及数据可视化，支持各地区、各行业以及采用各种技术的企业做出明智的业务决策。

访问 IBM 商业价值研究院中国网站，免费下载研究报告 https://www.ibm.com/ibv/cn

© Copyright IBM Corporation 2020
IBM Corporation

国际商业机器中国有限公司
北京市朝阳区北四环中路 27 号
盘古大观写字楼 25 层
邮编：100101

美国出品
2020 年 10 月

IBM、IBM 徽标、ibm.com 和 Watson 是 International Business Machines Corp. 在世界各地司法辖区的注册商标。其他产品和服务名称可能是 IBM 或其他公司的商标 。以下 Web 站点上的“Copyright and trademark information”部分中包含了 IBM 商标的最新列表：ibm.com/legal/copytrade.shtml。

本文档为自最初公布日期起的最新版本，IBM 可随时对其进行修改。IBM 并不一定在开展业务的所有国家或地区提供所有产品或服务。

本文档内的信息“按现状”提供，不附有任何种类（无论是明示还是默示）的保证，包括不附有关于适销性、适用于某种特定特定用途的任何保证以及非侵权的任何保证或条件。IBM 产品根据其提供时所依 据协议条款和条件获得保证。

本报告的目的仅为提供通用指南。它并不旨在代替详尽的研究或专业判断依据。由于使用本出版物对任何组织或个人所造成的损失，IBM 概不负责。

本报告中使用的数据可能源自第三方，IBM 并不独立核实、验证或审计此类数据。此类数据使用的结果均为“按现状”提供，IBM 不作出任何明示或默示的声明或保证。