

# 数据抓取与贩卖迎来强监管? 大数据风控公司遭遇“生死劫”

近日,有消息称,杭州魔蝎网络科技有限公司(简称“魔蝎科技”)和上海新颜人工智能科技有限公司(简称“新颜科技”)有高管被警方带走,几天后,公信宝的运营公司被杭州市公安局西湖分局古荡派出所查封。据知情人士透露,上述公司被查封,或与其爬虫业务非法获取用户信息、助力暴力催收等有关。

当数据成为一种资产,如何保护“资产安全”,如何实现金融创新与数据安全之间的平衡,在保护个人权利的同时如何激励社会更好地沉淀和使用数据,这都成为监管和行业需要深入探讨的问题。

■新快报记者 许莉芸



■廖木兴/图

“目前爬虫数据公司遭到监管,也是为了金融风险专项整治更深入化,整肃到数据源头。”

——有业内人士分析大数据风控公司被查封现象



## 被调查公司 或涉及运营商爬虫服务

据了解,“爬虫业务”是指平台方一种按照一定的规则,自动抓取互联网信息并存储到自身数据库的程序或者脚本。在用户授权后,风控数据提供商通过后台“爬虫”搜集信息,将通话信息、消费数据等互联网信息整合标准化,最终形成对借款人的综合评估,供金融机构做相应的后续决策。目前,网络爬虫存在着“是否经过用户授权,是否存在过度爬取信息,爬取到的信息用途不明等”诸多争议。

2017年,据一本财经报道,魔蝎科技就研发出一款被称为“同业爬虫”的产品,可以直接将其他现金贷平台的放款额和风控数据扒出来,相当于别家替你做了风控。据魔蝎科技的人员介绍,只需提供其他现金贷平台的用户名和密码,同业爬虫就可以爬取用户的基本信息、银行卡信息、职业、联系人、贷款记录、理财信息等,成功率在85%以上。

新快报记者拿到的一份“公信宝2018产品服务表”中,就介绍了其爬虫类数据类型涵盖社保、学信网、京东、电信、移动、联通、芝麻信用分、微信、支付宝,甚至是人行征信数据等——注意,目前人行征信数据,除了银行外,大数据风控公司并没有下载权限。

目前,魔蝎科技的官网也已经无法打开。新快报记者了解,多家与其有合作的网贷平台负责人表示,他们与魔蝎科技的服务已被终止,魔蝎科技提供的账号也已无法登录。目前,新颜科技的官网仍正常,对于高管被带走的消息,新颜科技客服对新快报记者表示:“没有接到通知,我司不提供爬虫业务。”

另据知情人士透露,锦程消费金融旗下两款产品“锦易贷”“收入贷”,疑似因供应商突然中止提供数据而暂停放款业务。对此,新快报记者向其客服求证,这两款产品确实已经暂停,至于暂停原因及何时恢复,客服表示不清楚。

9月16日,还有消息称同盾科技已解散其爬虫部门,该部门员工集体待岗,同盾科技实控人兼CEO蒋韬已出国避风头,并且魔蝎科技以及新颜科技被查封是同盾科技举报的。第二天,同盾科技官方声明指出,同盾科技创始人蒋韬一直在国内照常处理公司事务,数聚魔盒已经在2018年开始逐步调整业务,目前已经停止相关业务。数聚魔盒为同盾科技旗下的数据风控产品。针对大数据行业的严监管,是否影响其业务情况,同盾科技公关负责人对新快报记者表示“没有影响”。

“对小机构影响会比较大,如果风控模型极大程度依赖运营商,那么影响就很大,如果接的数据很多,影响就不大。”据业内风控人士分析。



## 灰色地带 贩卖数据产业化

事实上,从去年底现金贷整顿开始,不少大数据风控公司就陆续被调查。如去年底,多家大数据公司、征信公司如有脉金控、同牛科技、考拉征信爆出被监管调查。“目前爬虫数据公司遭到监管,也是为了金融风险专项整治更深入化,整肃到数据源头。”有业内人士分析称。

“魔蝎这类的大数据风控公司能够发展起来,无非是钻着中国数据隐私保护相关法律尚未完善这个空子。”有业内人士对新快报记者表示,风控算法业内无非

就是几种,然而决定风控能力的,最重要的就是原始数据的积累,“数据量越大越丰富,训练出的模型的效果自然也就越好。所以说这些游走在灰色地带的小公司肯定要被严监管。”

大数据风控公司被严监管的同时,也暴露出其背后用户信息隐私正在泄露的现状。

此前新快报记者调查时发现,在收集用户信息上,已然形成了贩卖数据的黑灰产业链,一些现金贷平台甚至与黑灰产业团伙“合谋”,寻求最大化的利益

空间。一家专门收集贷款数据的公司其负责人就对新快报记者称,一个贷款用户数据,根据数据维度不同,他们可以出价1元到10元不等。该人士对新快报记者表示,其数据一方面主要来源于固定合作渠道如现金贷平台、贷款超市等,另一方面则会通过爬虫等技术手段获取热门APP进行“反编译”(计算机术语,是指对他人软件的目标程序进行逆向研究分析,以推导出他人软件产品的源代码),并更改相应源代码。



## 迫在眉睫 上位法呼之欲出

有业内人士对新快报记者分析表示,爬虫并不犯法,而是爬出后的信息如何使用,存在隐私侵权、数据滥用等风险,特别是在数据的授权、来源、用途十分不透明的情况下。

西南财经大学普惠金融与智能金融研究中心副主任陈文也表示:“风控数据公司因为拥有用户的信息、关系图,在获客、营销、催收阶段都能助力现金贷机构。但数据来源大多处于灰色地域,对于个人隐私缺乏保护,存在广泛的数据滥用问题。”

对于金融风控公司在经用户授权后,获取通讯记录、支付记录等信息卖给信贷机构和催收机构的行为,上海九泽律师事务所高级合伙人朱敬律师表示:“根据《最高人民法院关于办理侵犯公民个人信息刑事案件适用法律若干问题的解释》的规定,未经被收集者同意,将合法收集的公民个人信息向他人提供的,属于

刑法第二百五十三条之一规定的‘提供公民个人信息’,该行为应当被追究刑事责任。”

针对这些问题,监管也在路上。今年5月28日,国家互联网信息办公室就《数据安全管理办法(征求意见稿)》(以下简称《办法》)公开征求意见。对于网络运营者超出运营需要收集个人信息的行为,《办法》作出了限制性规定,对“网络产品核心业务功能运行的个人信息”以外的信息,网络运营者不得因个人信息主体未同意收集而拒绝提供核心业务功能服务。

针对网络爬虫等抓取网页的自动化手段,《办法》明确应不妨碍网站正常运行,并列明具体的访问收集流量不得超过网站日均流量的1/3。针对手机APP过度获取权限的问题,要求“网络运营者不得以改善服务质量、提升用户体验、定向推送信息、研发新产品等为由,以默认授权、功能捆绑

等形式强迫、误导个人信息主体同意其收集个人信息”;对数据泄露才确定网络安全负责人的问题,其明确数据安全负责人的任职要求,突出网络运营者主要负责人、数据安全负责人的姓名及联系方式等。

更重要的是,《办法》还新增了两项配套的制度性规定,用以更好地落实政府部门对数据安全的监督。一是重要数据和个人敏感信息的备案,“如果网络运营者以经营为目的收集重要数据或个人敏感信息的,应向所在地网信部门进行备案。”二是数据安全认证和应用程序安全认证制度,“国家鼓励网络运营者自愿通过数据安全认证和应用程序安全认证,鼓励搜索引擎、应用商店等明确标识并优先推荐通过认证的应用程序。”但关于认证如何进行,办法除规定国家网信部门会同国务院市场监督管理部门进行指导外,未提供进一步的细则。