隐私合规风险知几何？数据合规商用需过九重关

原创 严强 微众银行区块链 3月11日

来自专辑

WeDPR隐私保护周三见

严强，SMU信息安全方向博士，信息安全顶级国际学术会议最佳论文奖获得者；曾作为Google隐私保护基础技术架构部门唯一来自中国的早期核心成员，领导研发的技术方案在Android和Google Play生态各大门户产品中全面集成投产。

在现代商业发展进程中，数据驱动的业务创新正起着至关重要的推动作用，隐私数据的引入，让企业能够更精准地发现潜在客户，更好地服务目标人群，甚至开辟全新的市场空间。

然，福兮祸之所伏。每一次创新，都有可能打破隐私数据使用的常规范式，带来诸如对个人隐私空间侵犯、企业敏感信息泄露等额外的隐私风险和不良社会影响。政府相关部门作为市场秩序和社会秩序的守护者，就会为此设定对应的法律法规，来规范企业在进行依赖隐私数据的业务创新时，应遵循的必要标准。

近年来，以欧盟《通用数据保护方案》（简称GDPR）为代表，世界各国政府对于隐私的立法保护不断细化，惩罚力度大大加强。在我国，以《中华人民共和国网络安全法》、《信息安全技术 个人信息安全规范》、《个人金融信息保护技术规范》为代表的现有法律框架内，对于侵犯隐私的不法行为也有判罚，轻则罚款，重则锒铛入狱。

就在上周五（3月6日），个人隐私安全在国家层面得到更细粒度的保护，2020年新版国家标准《信息安全技术 个人信息安全规范》正式发布，对个人信息收集、储存、使用做出了明确规定，并规定了个人信息主体具有查询、更正、删除、撤回授权、注销账户、获取个人信息副本等权力，同时新增「多项业务功能的自主选择」「用户画像的使用限制」「个性化展示的使用」「第三方接入管理」等内容。

如何才能让业务创新有效地满足隐私合规的严格要求？这里，我们将从控制合规成本的角度，分享关于平衡隐私合规风险和现代商业发展的一些思考：如何识别隐私合规风险，理解不同层面的合规需求，通过技术手段控制合规成本，并应对企业发展业务扩张过程中可能出现新的隐私合规挑战。

明确隐私合规的目标

由于存在企业发展阶段和区域市场法律法规的差异性，有效应对隐私风险的首要任务在于明确隐私合规的目标。

我们可以观察到，伴随着立法的细化深入，近年来关于「什么数据才算是隐私数据」的争议在不断减少。尽管每个区域法律法规对于隐私数据的定义不尽相同，但都提供了具体的类型定义和敏感性分级，例如，位于最高敏感级的KYC身份数据、金融数据等。这使得我们现在能够避免以往权利边界不清的问题，从而明晰隐私合规的目标。

对于在某一区域开展的业务，隐私合规的目标可以归结为：

保护当前区域市场法律法规中定义的隐私数据，并在产品设计中提供相应的特性，以此保障客户的法定权利。

这里提炼出的两组关键词——“数据内容保护”和“数据权利保障”，代表了隐私合规的两条主线。

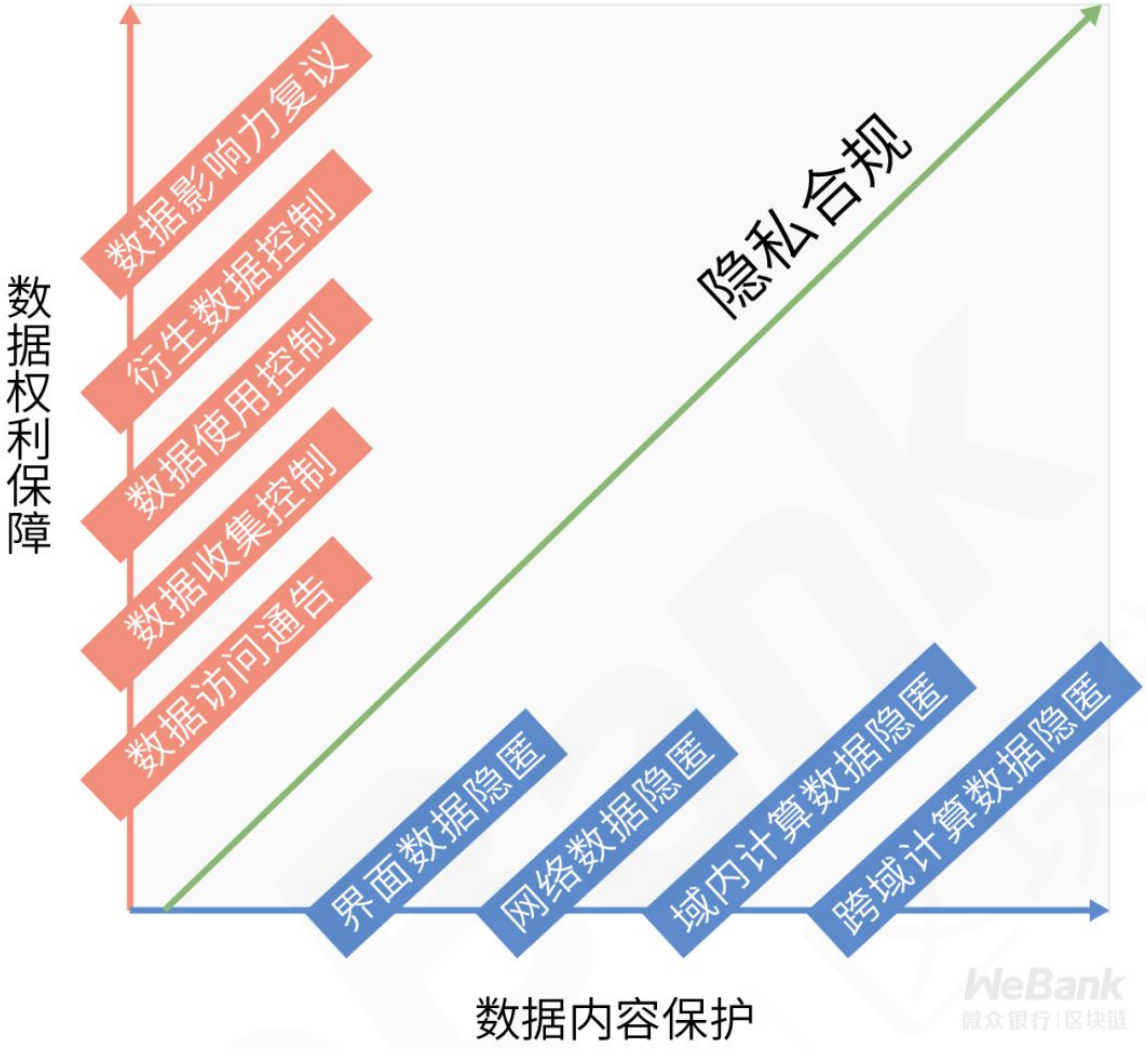
接下来，我们将围绕两条主线相关的九个维度，具体描述其对应的合规需求。

从九个维度拆解合规需求

鉴于信息化社会中绝大部分成功的商业模式，难免会依赖源自海量客户的海量隐私数据，传统的人为治理手段效率十分有限，而潜在的违规惩罚金额相当“可观”。所以，我们需要引入技术手段来满足来自各个维度的合规需求。

九个维度的合规需求犹如隐私合规的九重关卡。对于普通企业而言，重点关注最基本的维度便可满足合规需求。但对于处于强监管行业中的企业，如金融科技公司，或者运营跨国信息业务的企业，如在线社交网络、跨境电商等，则可能需要满足所有维度的合规需求。

如何才能过关斩将，最终在合法合规的框架下，实现业务稳健发展？这里，我们将一一阐述相关要点。

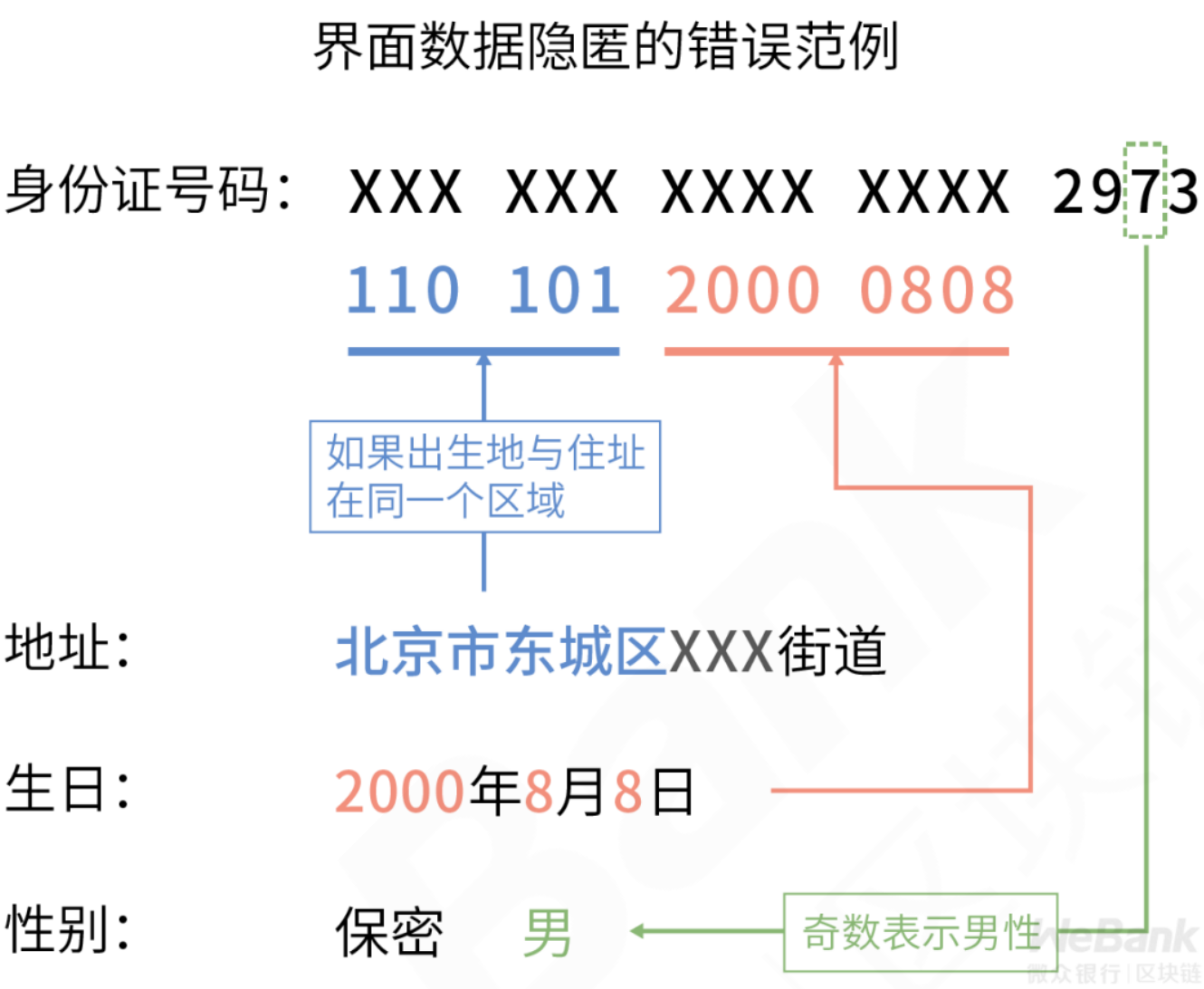


第一维度：界面数据隐匿

在用户界面中隐匿数据，使得客户在使用产品时，其隐私数据无法被附近位置的恶意第三方所窥视。

作为数据内容保护合规中最容易满足的一个需求，直接的界面渲染操作，如简单的显示打码、数据截断等都是有效的技术手段。

然而，它往往也是最容易因为忽视而出现隐私事故的一个维度。尤其是在多个敏感数据字段同时显示的前提下，若隐匿技术使用不当，可能等同于没有任何隐匿效果。



第二维度：网络数据隐匿

在网络维度上隐匿数据，使得隐私数据在传输过程中，无法被恶意第三方截获明文。

经典的传输层安全TLS/SSL系列协议，都可以满足这一需求。但需要注意，这类协议的安全性，依赖可信公钥数字证书服务的正常运行，一旦该服务受到攻击，可能会导致证书造假、证书过期等，最终影响到现有业务的安全性和可用性。

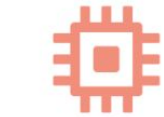
第三维度：域内计算数据隐匿

在同一个计算域内，如由企业完全掌控和部署的云计算环境，任何隐私数据的明文，在计算和存储过程中，都不离开安全隔离环境，防止企业存在内鬼进行未授权的隐私数据访问。

隐私数据只有在安全隔离计算环境中，才会被解密成明文，在安全隔离计算环境之外，只能进行密文运算，并以密文形式存储在介质中。这里需要用到可信硬件或者软件隔离来构建安全隔离计算环境，它们分别依赖不同的安全假设，需要根据业务的特性来进行选择。

验证硬件物理隔离方案的安全假设

需要**访问物理机房**



硬件物理隔离

安全假设：

- 硬件厂商没有植入后门
- 平台服务商在配置硬件系统时没有植入后门
- 平台服务商在运行服务时正确启用了硬件隔离特性
- 应用开发者在开发应用时正确调用了硬件隔离特性

验证软件协议隔离方案的安全假设

需要**访问软件源码**

软件协议隔离

安全假设：

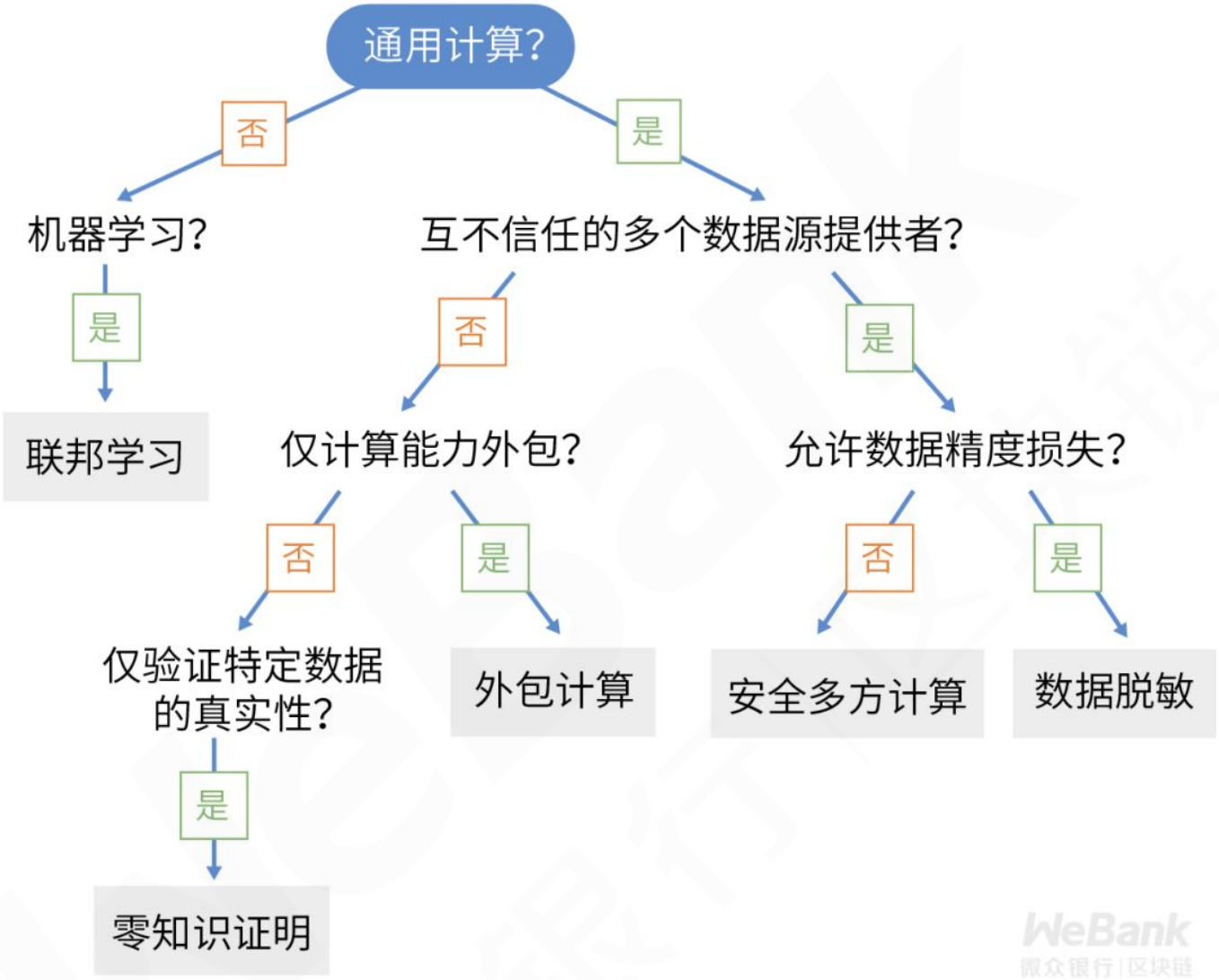
- 软件厂商没有植入后门
- 应用开发者在开发应用时正确调用了软件隔离特性

第四维度：跨域计算数据隐匿

隐私数据的明文只在同一个计算域内出现，在与其他计算域进行联合计算时，其他计算域的控制方无法直接访问或间接推测出隐私数据的明文，防止其他合作方获得合作协议授权之外的敏感隐私数据。

作为数据内容保护合规中最具挑战性的需求，其对于以医疗数据、金融数据等高度敏感数据业务尤为重要。如果无法满足合规要求，通常意味着业务无法开展或者面临巨额罚金。而且可能会出现双向判罚，即企业不仅会因为自身方案漏洞导致隐私数据泄露而受罚，还会因利用合作方企业的方案漏洞违规获取未授权的敏感隐私数据而受罚。

为了避免相关隐私合规事故，可采用的通用技术方案包括数据脱敏、安全多方计算、数据外包计算、零知识证明等。在涉及机器学习的特定场景下，联邦计算等新兴技术可以提供更为有效的方案效果。



第五维度：数据访问通告

数据访问通告是指，让客户了解当前业务会收集哪些隐私数据、为什么需要这些数据、将会如何使用这些数据、将以什么方式保存这些数据、保存期多久等隐私数据流通生命周期的细节。作为数据权利保障合规中最基础的需求，它保障了客户的知情权。

满足该需求的难点在于，如何让客户理解晦涩的技术语言、理解相关隐私风险的后果，避免相关监管机构以混淆客户理解为由，判定企业违规。

进行用户体验和人机交互技术的研究，是处理好这一需求的关键。适度采用基于机器学习的自动风险匹配是近年来业界比较推崇的技术，简化客户理解成本，帮助其更理性地评估对应业务的潜在风险。

机器可读

源代码

```
Result predict_user_interest(
 EventCollection user_event,
 Metadata1 data1, …, MetadataN dataN) {
 … (细节太多不再详述)
}
```

技术人员可读

技术需求

数据输入：

用户与本服务交互时产生多种事件数据。

逻辑步骤：

1. 将事件数据汇总并清理。
2. 清理后转化为高维向量。
3. 经过一系列XXX模型变换，获得有效表征用户兴趣的YYY类型元数据。
4. YYY类型元数据与现有Metadata1数据进行整合。

… (细节太多不再详述)

普通用户可读

隐私政策

本服务将基于用户的观看历史进行个性化的内容推荐。

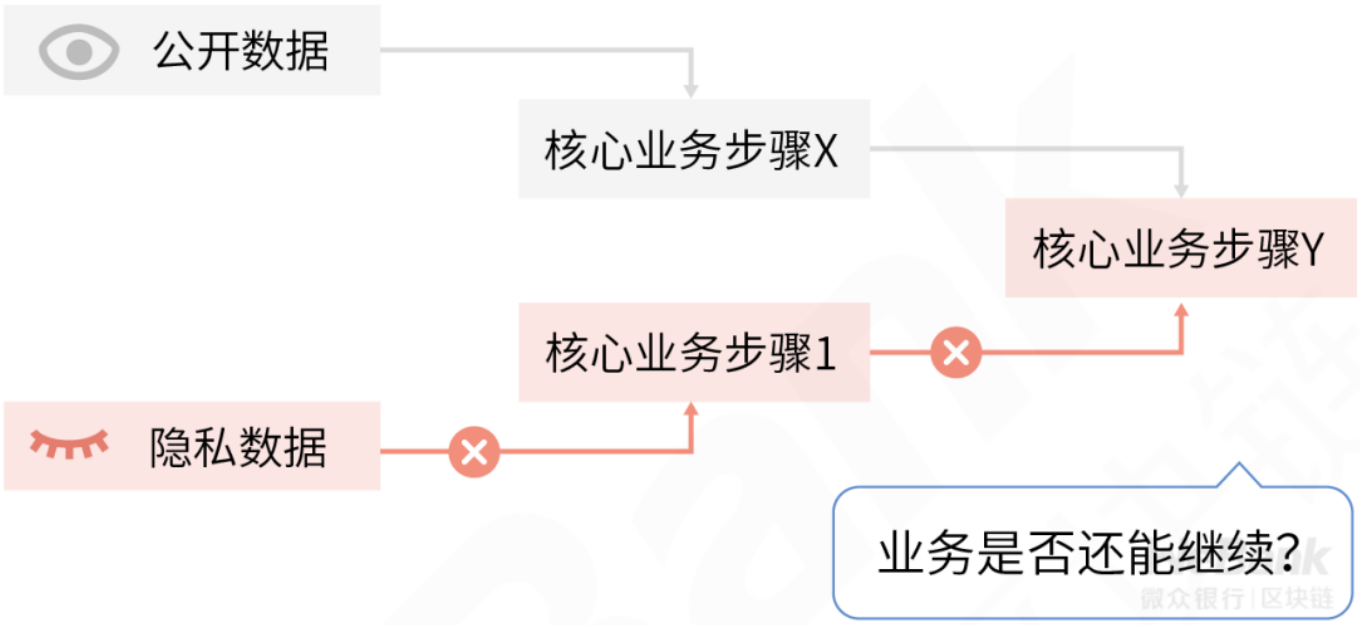
具体收集的信息可能包括：

- 内容分类
- 观看时长

第六维度：数据收集控制

数据收集控制是指，允许客户选择哪些隐私数据会被业务系统所收集，并在初始选择之后，允许对未来的数据收集选择进行调整。由于数据收集作为隐私数据流通生命周期的起始点，该需求能够赋予客户对于自身隐私数据流通的全局控制权。

对于客户不愿意分享的隐私数据，在数据收集控制机制的作用下，无法以未授权的方式进入业务系统，以此营造客户的心理安全感。传统的访问控制技术可以很好地实现这一需求，但若原系统架构设计扩展性不佳，相关历史系统改造将是一项巨大的工程挑战。

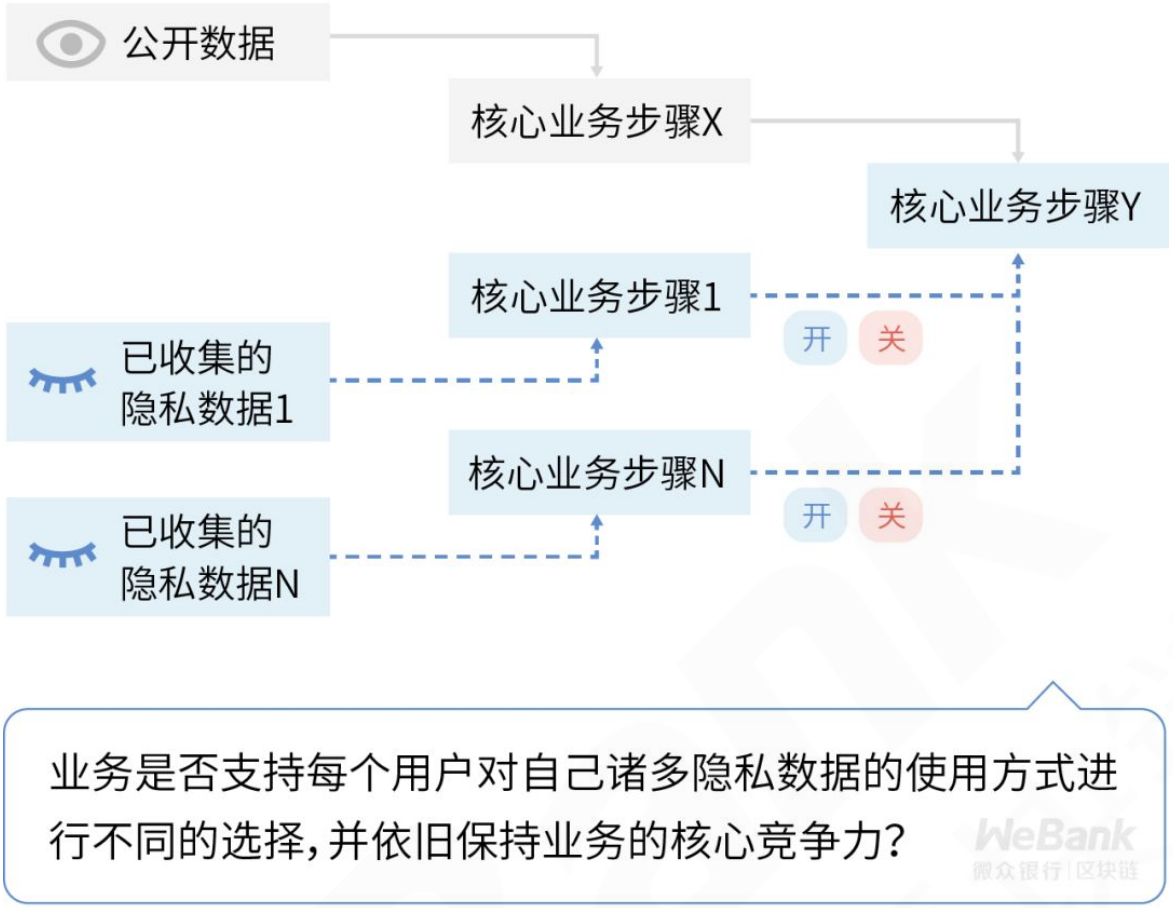


第七维度：数据使用控制

数据使用控制是指，允许客户对于特定的业务系统中使用隐私数据的方式进行调整或限制。

原先是GDPR特有的合规需求之一，称之为限制处理权，最新版的《信息安全技术 个人信息安全规范》中也有相关规定，仅对部分业务类型有效。目前主要针对在线广告投放相关的个性化推荐业务，设立的初衷是避免过于个性化推荐引发的个人隐私空间强烈侵入感。

鉴于GDPR巨额罚款机制，这对于相关业务在注重个人隐私的区域的稳健发展十分重要。有效应对该项需求的关键，在于企业能否在系统架构设计早期，为相关隐私数据变动预留空间，减少后期系统改造的代价。



第八维度：衍生数据控制

衍生数据使用控制是指，允许客户对其原始隐私数据在经过变换、聚合后产生的衍生数据有一定的控制权。

这也是GDPR特有的合规需求之一，目前主要表现在两方面：

1. 数据被遗忘权：在客户删除账户之后，清理对应个体历史数据和包含该客户的聚合数据；

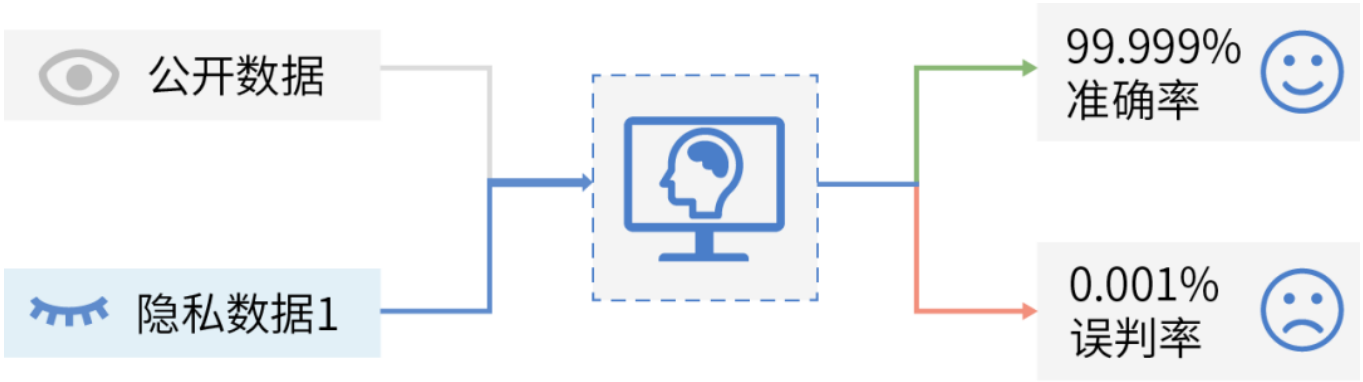
2. 数据携带权：客户有离开当前业务平台的意向，打包提取之前所有的相关历史数据，如电子邮件、评论留言、云主机数据等。

该项需求的实现，通常也面临着高昂的系统改造代价。建议企业在系统架构设计早期，务必考虑完备的隐私数据溯源机制，为后期改造减少合规成本。

第九维度：数据影响力复议

数据影响力复议是指，允许客户对基于其隐私数据产生的业务决策进行复议，由此更正自动化决策系统可能做出的不公平判断，消除数据歧视等负面影响。

这可能是权利数据保障合规中最具挑战性的需求，其关注点在于数据驱动决策系统设计的可解释性，并限制难以解释的机器学习模型在民生、医疗等关键领域的应用。这就要求企业在研发自动化决策系统设计时，研发具备较高解释能力的决策模型，或者提供备选技术方案，减少误判导致的合规成本。



业务需要为 自动化决策 中可能出现的误报提供**人工复议**渠道

正是：**强立法监管严管控，兴科技企业巧合规！**

日益细化的隐私保护法律法规，对于海量隐私数据提出了不断量化的合规需求。只有借助科技的力量，才有可能有效实现隐私合规，以此扫清企业持续业务创新的阻碍，并为企业开拓国际市场做好准备。

因此，自本系列的下一篇推文开始，我们将以隐私保护的核心技术领域「密码学」为起点，逐步与大家分享关于关键技术的深入解析和理论分析，欲知详情，敬请关注下文分解。

---END---

《隐私保护周三见》

“科技聚焦人性，隐私回归属主”，这是微众银行区块链团队推出《隐私保护周三见》深度栏目的愿景与初衷。每周三晚8点，专家团队将透过栏目和各位一起探寻隐私保护的发展之道。

栏目内容含括以下五大模块：关键概念、法律法规、理论基础、技术剖析和案例分享，如您有好的建议或者想学习的内容，欢迎随时提出。

栏目支持单位：零壹财经、陀螺财经、巴比特、火讯财经、火星财经、价值在线、链客社区

往期集锦

第1论 | 隐私和效用不可兼得？隐私保护开辟商业新境地

长按二维码关注

微众银行区块链

白皮书下载 | 订阅干货 | 进群交流 | 合作联系