计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.201000011

面向跨模态隐私保护的AI治理法律技术化框架

雷羽潇　段玉聪
海南大学计算机科学与技术学院　海口 570228
(229597800@qq.com)

摘　要　随着虚拟社区在网络用户中的普及,虚拟社区群已经成为一个小型社会,可通过用户浏览所留下的"虚拟痕迹"和发布的用户生成内容提炼出与用户相关的隐私类型资源。根据隐私类型资源自身的特性,可将其分类为数据资源、信息资源和知识资源,三者构成了用户的数据信息知识与智慧图谱(DIKW图谱)。虚拟社区中的隐私类型资源有4个流通过程,即隐私资源的感知、存储、传输和处理;4个过程分别由3个参与方(用户、AI系统和访问者)单独或合作完成,3个参与方所拥有的隐私权包括知情权、参与权、遗忘权和监督权。通过明确3个参与方在4个流通过程中的隐私权范围,结合隐私价值保护,设计了匿名保护机制/风险评估机制和监督机制,用于构建一个虚拟社区隐私保护的AI治理法律框架。

关键词:数据、信息、知识与智慧图谱;虚拟社区;隐私保护;隐私的价值;隐私权
中图法分类号　TP309

AI Governance Oriented Legal to Technology Bridging Framework for Cross-modal Privacy Protection

LEI Yu-xiao and DUAN Yu-cong
School of Computer Science and Technology, Hainan University, Haikou 570228, China

Abstract　With the popularity of virtual communities among network users, virtual community groups have become a small society, which can extract user-related privacy resources through the "virtual traces" left by users' browsing and user-generated content user published. Privacy resources can be classified into data resources, information resources and knowledge resources according to their characteristics, which constitute the data, information, knowledge, and wisdom graph (DIKW graph). There are four circulation processes for privacy resources in virtual communities, namely, the sensing, storage, transfern, and processing of privacy resources. The four processes are respectively completed by the three participants, the user, the AI system, and the visitor individually or in cooperation. The right to privacy includes the right to know, the right to participate, the right to forget, and the right to supervise. By clarifying the scope of privacy rights of the three participants in the four circulation processes, and combining the protection of privacy values, an anonymous protection mechanism, risk assessment mechanism and supervision mechanism are designed to build an AI governance legal framework for privacy protection of virtual communities.

Keywords　Data\Information\Knowledge and Wisdom graph, Virtual community, Privacy protection, Value of privacy, Right to privacy

1　引言

随着信息技术的进步,虚拟社区开发出了许多新的功能,用户可以在虚拟社区上完成交友、购物、宣传、求助等许多事务。用户在虚拟社区中的各种交互行为将产生数字类型的隐私资源,包括虚拟痕迹(Virtual Trace,记为 $T_{virtual}$),以及由交互行为产生的用户生成内容(User-generated Content,UGC)。$T_{virtual}$能够反映用户自身的性格与行为习惯[1],可用于观察、记录、分析用户在虚拟社区中的行为[2],达到一定数量级的 $T_{virtual}$甚至能够产生争执影响,引导网络舆论[3]。UGC则能够最直观地反映用户对事物的评价[4],可用于分析虚拟社区中用户的兴趣和需求[5]。随着大数据技术的进步,$T_{virtual}$和UGC中蕴含的巨大的经济价值得到了开发和利用,包括类型广告的精准投放和流行趋势的分析等。在经济利益的驱使

到稿日期:2020-10-03　返修日期:2021-04-04　　本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。
基金项目:国家自然科学基金(61662021,72062015);海南省自然科学基金项目(620RC561);海南省教育厅项目(Hnky2019-13);海南大学教育教学改革研究项目(HDJY2102,HDJWJG03)
This work was supported by the National Natural Science Foundation of China(61662021,72062015),Natural Science Foundation of Hainan Province,China(620RC561),Program of Education Bureau of Hainan Province,China(Hnky2019-13) and Educational Reform Research Program of Hainan University,China(HDJY2102,HDJWJG03).
通信作者:段玉聪(duanyucong@hotmail.com)

下，虚拟社区中数字资源的非法收集、未经允许的使用和超出允许范围外使用的事情时有发生。数字类型隐私资源属于用户隐私的一部分，但传统的隐私保护方法对数字隐私资源的保护作用有限[6]。

如今，人工智能系统的自动化决策在越来越多的领域代替了传统人工决策。虽然 AI 系统的自动化决策具有许多优点，如快速、便捷、准确度高、在一定程度上排除了人为因素的干扰，但是 AI 系统并没有解决数字隐私资源的非法开发和使用问题，甚至进一步导致针对人类个体的偏见与歧视[7]和有关算法道德的争议[8]。

本文搭建了一个 AI 治理虚拟社区的法律框架，利用 DIKW 图谱技术对虚拟社区中用户的数字隐私资源进行分类建模。DIKW 图谱是以“Human”和“Existence”两个概念为核心的数据、信息、知识和智慧框架 UML 元模型[9]，根据隐私资源对于“实体-关系”这一结构的不同表达将其分为数据、信息、知识 3 种不同的类型，构成以用户为中心的数据、信息和知识图谱[10]，三者共同构成用户的隐私资源 DIKW 图谱，可用于优化存储、传输与计算一体化的处理效率[11]。

隐私资源流通的过程中共有 3 个参与方，分别是作为资源产生方的用户、作为资源存储方的 AI 系统，以及作为资源获取方的访问者。3 个参与方单独或合作参与了隐私资源流通的感知、存储、传输和处理 4 个环节。隐私资源保护的标准是在满足各方在不同环节的隐私权需求的同时保护隐私资源的价值。

在隐私资源完成从用户到访问者的转移过程中，本文还设立了匿名保护机制、风险评估机制和监督机制等隐私保护机制，对流通环节中的隐私权和隐私价值进行监督保护和管理，搭建基于 DIKW 图谱的隐私资源法律保护框架，对自动化决策中的用户隐私资源进行系统性的保护，其属于隐私工程的一部分[12]。

本文第 2 节介绍了不同类型数字隐私资源的属性，包括隐私资源之间的互斥和隐私资源的价值，还说明了虚拟社区对隐私资源处理和 DIKW 图谱建立的方法；第 3 节介绍了 AI 系统决策行为中隐私资源在 3 个参与方之间的流通过程，以及各个流通环节中的参与权；第 4 节介绍了针对隐私资源设立的保护机制，包括匿名保护机制、风险评估机制和监督机制。

2 虚拟社区中的隐私资源

虚拟社区中的数字类型隐私资源（$Privacy_{DIK}$，记为 P_{DIK}）根据所表达的关键元素的语义形式化[13]，可分为数据类型资源（$Data_{DIK}$，记为 D_{DIK}）、信息类型资源（$Information_{DIK}$，记为 I_{DIK}）和知识类型资源（$Knowledge_{DIK}$，记为 K_{DIK}）3 种类型，如式(1)所示。D_{DIK}，I_{DIK} 和 K_{DIK} 分别表达了实体（$Entity$，E）之间 3 种不同类型的关系（$Relation$，R），关系能够定义一切的事物[14]。以 E 为结点、R 为边，可构建以用户为核心的数据图谱（$DataGraph_{DIK}$，记为 D_{Graph}）、信息图谱（$Information\text{-}Graph_{DIK}$，记为 I_{Graph}）和知识图谱（$KnowledgeGraph_{DIK}$，记为 K_{Graph}），三者共同构成了用户的 DIKW 图谱 $DIKW_{Graph}$，如式(2)所示。

$$P_{DIK}=\{D_{DIK},I_{DIK},K_{DIK}\} \tag{1}$$

$$DIKW_{Graph}=\{D_{Graph},I_{Graph},K_{Graph}\} \tag{2}$$

传统的数据挖掘注重数据类型资源本身，忽略了信息和知识类型的资源在分析用户行为与习惯中起到的作用。关系 R 中包含的资源量并不亚于实体 E 本身。通常仅基于数据类型资源 D_{DIK} 进行数据挖掘和分析是对虚拟社区中资源 P_{DIK} 的不完整利用或者浪费。基于 D_{IKW} 图谱的建模与分析与传统数据挖掘[15-16]方法相比，对资源关系的采集和利用更为全面与精细。

2.1 DIKW 图谱上的隐私资源

$DIKW_{Graph}$ 上的 P_{DIK} 经过提取和转换两个过程被分为 D_{DIK}，I_{DIK} 和 K_{DIK}，以图谱的形式被存储加载。P_{DIK} 除了表达实体之间的关系之外，其本身也相互关联，多个同类型或不同类型的 P_{DIK} 组合可生成新的隐私资源概念，例如互斥隐私资源组和群体隐私。

2.1.1 数据、信息、知识类型的隐私资源

$DIKW_{Graph}$ 由实体 E 和关系 R 构成，如式(3)所示。D_{Graph}，I_{Graph} 和 K_{Graph} 3 种图谱中关于实体 E 之间的关系 R 的描述不同，如图 1 所示，I_{Graph} 是 D_{Graph} 的实体化，K_{Graph} 是 I_{Graph} 的概念化。隐私资源结合与转换的过程即实体化和概念化的过程。

$$DIKW_{Graph}=(E,R) \tag{3}$$

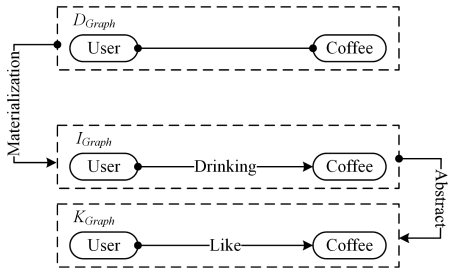


图 1 数据、信息、知识图谱示意图

Fig. 1 Diagram of data, information and knowledge

根据自身的属性和表达的内容，P_{DIK} 可按照主题 $Topic$ 归属分类。例如，D_{DIK1}，D_{DIK2} 和 D_{DIK3} 所表达的实体为相机、胶卷和照片，共同归属于摄影主题 Topic(Photography)。三者中的“Film”还能够表达电影的摄制工作，同时归属于电影主题 Topic(Movie)。

$D_{DIK1}=$“Camera”

$D_{DIK2}=$“Film”

$D_{DIK3}=$“Photo”

$\{D_{DIK1},D_{DIK2},D_{DIK3}\}\in Topic(Photography)$

$D_{DIK2}\in Topic(Movie)$

(1) D_{Graph}：数据类型资源 D_{DIK} 指可直接观察到的离散元素，表示单个实体 E 或是 E 的属性，在没有上下文联系的情况下不具有现实意义。如图 1 所示，构建完成的 D_{Graph} 是以 D_{DIK} 为节点、R 为边的无向图。如图 1 所示，从 D_{Graph} 中可知“$Coffee$”与“$User$”相关，但不可知两个实体之间的具体关系。

例如，D_{DIK1} 来自用户 U_a 的数据图谱 $D_{Graph}(U_a)$，表示照片实体 $Photo_A$，D_{DIK2} 来自 $Photo_A$ 的数据图谱 $D_{Graph}(Photo_A)$，表示时间 $Time_A$。$Photo_A$ 与 U_a 之间存在关系 R_1，

$Time_A$ 与 $Photo_A$ 之间存在关系 R_2。R_1 和 R_2 的内容存在多种可能性,U_a 可能是 $Photo_A$ 的拍摄者,也可能是被拍摄者,$Time_A$ 可能是 $Photo_A$ 的被拍摄时间,也可能是被发布的时间。

D_{DIK1} =“$Photo_A$” From $D_{Graph}(U_a)$

R_1 ExistIn (U_a, D_{DIK1})

D_{DIK2} =“$Time_A$” From $D_{Graph}(Photo_A)$

R_2 ExistIn (D_{DIK1}, D_{DIK2})

(2)I_{Graph}:信息类型资源 I_{DIK} 表达两个实体 E 之间客观存在的交互关系 R,以 $R(E_1, E_2)$的形式表示。

I_{DIK1} From $I_{Graph}(U_a)$

I_{DIK2} From $I_{Graph}(Photo_1)$

$I_{DIK1} = R(U_a, Photo_1)$

=“U_a 拍摄了照片 $Photo_1$”

$I_{DIK2} = R(Photo_1, Building_1)$

=“$Photo_1$ 中的内容包括 $Building_1$”

在 I_{Graph} 中,同一实体 E 可能存在多个关系边 R,I_{DIK} 由 D_{DIK} 与意图(Purpose)结合而产生,Purpose 驱动 I_{DIK} 的表达,如式(4)所示。I_{Graph} 在 D_{Graph} 的基础上明确了结点之间的边所表达的关系语义。

$$I_{DIK} = D_{DIK} + Purpose \tag{4}$$

例如,基于资源处理者不同的 $Purpose$,结点 D_{DIK1} 与其他结点之间存在多条关系边 R。

$D_{DIK1} = E_1$ =“$bread$”

$Purpose_1$ =“补充身体能量”

$I_{DIK1} = D_{DIK1} + Purpose_1 = R(D_{DIK1}, Food)$

=“$bread$ 可作为食物补充身体能量”

$Purpose_2$ =“保障人身安全”

$I_{DIK2} = D_{DIK1} + Purpose_2 = R(D_{DIK1}, Weapon)$

=“$bread$ 可作为武器随身携带”

$Purpose_3$ =“清洁油画画布”

$I_{DIK3} = D_{DIK1} + Purpose_3 = R(D_{DIK1}, Paint)$

=“$bread$ 可作为油画画布的清洁工具”

(3)K_{Graph}:知识类型资源 K_{DIK} 是对实体之间关系的归纳总结和推导,以 $K(E_1, E_2)$的形式表示。K_{DIK} 是对结点之间关系边 R 的进一步推导和总结。K_{DIK} 可能存在内容不完整、不确定、与 $DIKW_{Graph}(U_a)$上其他 P_{DIK} 所表达的内容不一致的问题,可通过建模进行偏好推理来解决[17]。偏好推理基于与 P_{DIK} 属于同一 $DIKW_{Graph}$ 或是同一 $Topic$ 的 P_{DIK} 进行,此类 P_{DIK} 属于 K_{DIK} 的关联隐私资源($Privacy_{DIK}^{associated}$,记为 $P_{DIK^{associated}}$)。

例如,K_{DIK1} 源自于对 U_a 所发布的一段文本内容 UGC(text)的提取。I_{DIK1} 与 I_{DIK2} 属于 K_{DIK1} 的 $P_{DIK^{associated}}$,可作为 K_{DIK1} 中内容准确性的认定。

I_{DIK1}, I_{DIK2}, K_{DIK1} From $DIKW_{Graph}(U_a)$

From $Topic(Photography)$

UGC(text)=“I like photography”

I_{DIK1} =“U_a 发布摄影作品的频率”

I_{DIK2} =“U_a 关注的摄影类博主数量”

$\{I_{DIK1}, I_{DIK2}\} \in P_{DIK^{associated}}(K_{DIK1})$

$K_{DIK1} = K(U_a, Photography)$

=“用户 U_a 喜欢摄影”

K_{DIK1} = Derive(UGC(text))

K_{DIK1} VerifiedBy$\{I_{DIK1}, I_{DIK2}\}$

2.1.2 隐私资源的属性

D_{DIK},I_{DIK} 和 K_{DIK} 分别表达了实体 E 之间不同类型的关系 R。D_{DIK} 所表达的 R 值为是与否(True/Flase),即两个实体 E 之间是否存在关系。I_{DIK} 所表达的 R 值则包括肯定(positive)和否定(negative)两种类型。

K_{DIK} 具有两个可通过计算得到的基本属性,即准确性($K_{DIK}(Validity)$,记为 $K_{DIK}(Val)$)和精确性($K_{DIK}(Precision)$,记为 $K_{DIK}(Pre)$)。$K_{DIK}(Val)$指 K_{DIK} 对 R 表述的正确程度,$K_{DIK}(Pre)$指 K_{DIK} 所表达的内容的精细程度。

如式(5)所示,$K_{DIK}(Val)$由函数 $Validity$ 根据 K_{DIK} 的基本属性来计算,例如来源 $K_{DIK}(source)$,$K_{DIK}(source)$的权威性越高,$K_{DIK}(Val)$值就越高。

$$K_{DIK}(Val) = Validity(K_{DIK}, P_{DIK^{associated}}) \tag{5}$$

同时,$K_{DIK}(Val)$值还会受到 $P_{DIK^{associated}}(K_{DIK})$的影响。语义相同或相近的 $P_{DIK^{associated}}$ 将导致 $K_{DIK}(Val)$值上升;语义相斥的 $P_{DIK^{associated}}$ 将导致 $K_{DIK}(Val)$值下降;K_{DIK} 与 $P_{DIK^{associated}}$ 的关系 R 存储在 $I_{Graph}(K_{DIK})$上。

$K_{DIK}(Pre)$是表达相同内容的不同 K_{DIK} 的内容精确度。例如,K_{DIK1} 和 K_{DIK2} 属于摄影主题 $Topic(Photography)$中的内容,K_{DIK2} 的内容精确度更高,$K_{DIK2}(Pre)$大于 $K_{DIK1}(Pre)$。

K_{DIK1}, K_{DIK2} From $K_{Graph}(U_a)$

$\{K_{DIK1}, K_{DIK2}\} \in Topic(Photography)$

$K_{DIK1} = K(U_a, Photography)$=“$U_a$ 喜欢摄影”

$K_{DIK2} = K(U_a, Photography)$=“$U_a$ 喜欢自然主义摄影”

$K_{DIK2}(Pre) > K_{DIK1}(Pre)$

2.1.3 互斥隐私资源组

互斥隐私资源组($P_{DIK(G)^{inconsistent}}$,记为 $P_{DIK(G)^{in}}$)指存在逻辑冲突的两个及以上 P_{DIK} 的集合,$P_{DIK(G)^{in}}$ 存在彼此之间存在否定关系 R 的 P_{DIK}。举例如下:

I_{DIK1}, I_{DIK2} From $I_{Graph}(U_a)$

D_{DIK1} =“虚拟社区注册资料”

D_{DIK2} = UGC(text)=“I am a boy”

I_{DIK1} = Derive(D_{DIK1}) = $R(U_a, Gender)$ =“U_a 是一名女性”

I_{DIK2} = Derive(D_{DIK2}) = $R(U_a, Gender)$ =“U_a 是一名男性”

$\{I_{DIK1}, I_{DIK2}\} \in P_{DIK(G)^{in}}$

其中,I_{DIK1} 提取自 U_a 注册时所填写的个人资料 D_{DIK1},I_{DIK2} 提取自 Ua 在虚拟社区中发布的文字类 UGC(text),记为 D_{DIK2}。I_{DIK1} 与 I_{DIK2} 的集合属于 $P_{DIK(G)^{in}}$。

2.1.4 群体隐私资源

现实中人类实体之间的交互会导致 P_{DIK} 的流通与共享。若从人类实体 E 的某个 P_{DIK} 入手,能够挖掘到实体 E 的家人、朋友、邻居等关联人群的 P_{DIK},则此项 P_{DIK} 属于群体隐私。

群体隐私(GroupPrivacy)[18] 存在于多个实体 $E_1, E_2, \cdots, E_n$ 之

间，根据隐私资源的性质可分为群体关系隐私($GroupPrivacy_{relation}$，记为 $GP_{relation}$)和群体内容隐私($GroupPrivacy_{content}$，记为 $GP_{content}$)。群体隐私所牵涉的人类实体大于或等于2，构成共享群体隐私的亲密关系群体($Group_{Intimacy}$，$G_{Intimacy}$)。$G_{Intimacy}$ 不仅限于存在关联的多个人类实体集合，也可以表示被按照种族、性别、年龄、职业等特定标签分类的群体。

$GP_{relation}$ 表达的是两个及以上的 E 之间的关系 $R(E_1, \cdots, E_n)$，同时存储在 $G_{Intimacy}$ 中所有实体的 I_{Graph} 上。$GP_{content}$ 指会对 $G_{Intimacy}$ 中所有实体 E 造成影响的 P_{DIK} 内容，通常情况下只储存在 $G_{Intimacy}$ 中某一个实体的 $DIKW_{Graph}$ 上。$G_{Intimacy}$ 中不同 E 对 $GP_{content}$ 的保留度由于个体的差异而有所不同，而各个实体之间彼此联系，任何一个实体的行为都将影响整个群体的隐私保护，因此群体隐私保护的难度高于个体隐私。

以 $G_{Intimacy}$ 中实体 E 的 $DIKW_{Graph}$ 为基础，构建整个 $G_{Intimacy}$ 的群体 DIKW 图谱($DIKW_{Graph^{Group}}$，记为 $DIKW_{Graph^G}$)，包括群体数据图谱($DataGraph^G$，记为 D_{Graph^G})、群体信息图谱($InformationGraph^G$，记为 I_{Graph^G})和群体知识图谱($KnowledgeGraph^G$，记为 K_{Graph^G})，如式(6)所示。

$$DIKW_{Graph^G} = \{DataGraph^G, InformationGraph^G, KnowledgeGraph^G\} \tag{6}$$

2.2 隐私资源的处理

虚拟社区中的 P_{DIK} 以杂乱无序的方式存在，AI 系统作为 P_{DIK} 的存储处理者和传播的中枢，在 $DIKW_{Graph}$ 的建模过程中需要对 P_{DIK} 进行资源整理，排除无效的资源，精简重复的资源，资源整理过程包括提取、转换和加载(Extract - Transform - Load，ETL)[19]，P_{DIK} 将在资源整理过程完成之后从来源端虚拟社区转移到目的端 $DIKW_{Graph}$。

2.2.1 隐私资源的提取

隐私资源的提取是从同质或异类源中采集得到 P_{DIK} 的过程。隐私具有自我主观性，P_{DIK} 的采集标准将根据用户对 P_{DIK} 的保留程度而确定。

用户对隐私资源的保留程度($Degree_{Reserve}$，记为 D_{Res})是 P_{DIK} 的一项基本属性，以 $P_{DIK}(D_{Res})$ 的形式表示。$P_{DIK}(D_{Res})$ 的值越大，保留程度就越高。$P_{DIK}(D_{Res})$ 大于保留阈值($Degree^{Weight}$，记为 D^W)的 P_{DIK} 属于秘密类型资源 $Secret_{DIK}$，即 P_{DIK} 采集标准所规定的不能提取的资源范围。

$P_{DIK}(D_{Res})$ 取决于 P_{DIK} 的自身属性以及用户与 P_{DIK} 相关行为记录组 $Inter(P_{DIK})$，如式(7)所示，AI 系统构造函数 $Reserve$ 计算 $P_{DIK}(D_{Res})$ 值。P_{DIK} 的自身属性决定 $P_{DIK}(D_{Res})$ 的基础数值，$Inter(P_{DIK})$ 决定了 $P_{DIK}(D_{Res})$ 的动态数值。

$$P_{DIK}(D_{Res}) = Reserve(P_{DIK}, Inter(P_{DIK})) \tag{7}$$

例如，P_{DIK} 的来源 $P_{DIK}(source)$ 包括 $T_{virtual}$ 和 UGC，前者属于被动类资源，后者属于主动类资源。在 $P_{DIK}(D_{Res})$ 的计算中，来自 $T_{virtual}$ 的 P_{DIK} 的基础数值大于来自 UGC 的 P_{DIK}。

用户在虚拟社区中的行为能够反映用户心理层面对该项资源的保留程度。与 P_{DIK} 相关的用户行为记录组 $Inter(P_{DIK})$ 包括了隐私保护正相关行为($Inter^{positive}$，记为 $Inter^{pos}$)和负相关行为($Inter^{negative}$，记为 $Inter^{neg}$)。例如：

$Inter(P_{DIK1})^{pos}$ ="关闭虚拟社区读取移动设备麦克风的权限"

$Inter(P_{DIK2})^{neg}$ ="公布自己的年收入"

$Inter(P_{DIK})^{pos}$ 中所记录的行为越多，表示用户对 P_{DIK} 的保留度就越高，$P_{DIK}(D^{Res})$ 值将上升。$Inter(P_{DIK})^{neg}$ 则会导致 $P_{DIK}(D^{Res})$ 值下降。$Inter(P_{DIK})^{pos}$ 与式(7)中的计算结果正相关，$Inter(P_{DIK})^{neg}$ 与计算结果负相关。

2.2.2 隐私资源的转换

虚拟社区中内容不完整、不一致的 P_{DIK} 具有不确定性，将导致人们对资源分析的过程缺乏信心，自动化决策结果缺乏面向众人的说服力[20]。P_{DIK} 的不确定性可通过资源转换来弥补，如图2所示，通过 Transformation 模块可将 P_{DIK} 转换成为新的隐私资源 $P_{DIK^{new}}$。Transformation 包括基本转换、组合转换和技术转换。P_{DIK} 有两个基本属性 $P_{DIK}(in)$ 和 $P_{DIK}(out)$，表示 P_{DIK} 可转换生成 $P_{DIK}(out)$ 个 $P_{DIK^{new}}$，以及 P_{DIK} 自身作为 $P_{DIK^{new}}$ 时，可由其他 $P_{DIK}(in)$ 个 P_{DIK} 转换生成。

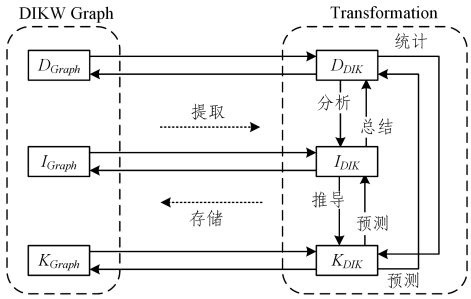


图2 隐私类型资源的转换

Fig. 2 Transformation of $Privacy_{DIK}$

(1)基本转换

基本转换指从单一 P_{DIK} 推导出由多个新资源组成的集合 $P_{DIK(G)^{new}}$ 的转换模式，$P_{DIK(G)^{new}}$ 中 P_{DIK} 的类型和数量不限。基本转换包括 D_{DIK}，I_{DIK}，K_{DIK} 的同类型转换和跨类型转换。例如：

I_{DIK1} From $I_{Graph}(U_a)$

I_{DIK1} ="用户 U_a 出生于 xx 年 xx 月 xx 日"

$I_{DIK2^{new}}$ = Transformation(I_{DIK1}) ="用户 U_a 今年 xx 岁"

(2)组合转换

组合转换指由 $P_{DIK}(G)$ 结合生成 $P_{DIK(G)^{new}}$ 的多对多的转换模式，$P_{DIK(G)}$ 和 $P_{DIK(G)^{new}}$ 中 P_{DIK} 的类型和数量都不限。例如：

I_{DIK1}，I_{DIK2} From $I_{Graph}(A)$

I_{DIK1} ="产品 A 第一季度的销量"

I_{DIK2} ="产品 A 第二季度的销量"

$I_{DIK3^{new}}$ = Transformation(I_{DIK1}，I_{DIK2})

="产品 A 上半年销量增长率"

(3)技术转换

技术转换指需要技术手段和其他资源内容的辅助来完成的 $P_{DIK^{new}}$ 生成过程。与基本转换和组合转换不同的是，技术转换存在转换难度 $T_{Difficulty}$，$T_{Difficulty}$ 的值取决于参与转换的实体 E。如式(8)所示，构造函数 $Difficulty$ 计算 P_{DIK} 到 $P_{DIK^{new}}$ 的技术转换难度 $T_{Difficulty}(P_{DIK}, P_{DIK^{new}})$，式(8)的计算结果小于实体 E 的转换能力阈值 $T_{Difficulty^W}(E)$ 时，表示

E 有能力完成技术转换。

$$T_{Difficulty}(P_{DIK}, P_{DIK^{new}}) = Difficulty(P_{DIK}, P_{DIK^{new}}, E) \tag{8}$$

其中，E 包括技术 $E(Tech)$ 和资源 $E(Resource)$。若 $E(Tech)$ 和 $E(Resource)$ 不能满足技术转换的需求，则在仅有实体 E 参与的情况下无法完成转换，$T_{Difficulty}$ 小于 $T_{Difficulty^W}(E)$。例如：

I_{DIK1} = “$Photo_1$ 由 U_a 所拍摄”

I_{DIK2} = “$Photo_1$ 中的内容包括 $Building_1$”

$I_{DIK3^{new}}$ = Transformation(I_{DIK1}, I_{DIK2})

= “$Photo_1$ 由 U_a 在 $Place_1$ 处拍摄”

I_{DIK4} = “$Building_1$ 位于 $Place_1$”

If $I_{DIK4} \in E(Resource)$：

$T_{Difficulty}((I_{IDIK1}, I_{DIK2}), I_{DIK3^{new}}) = Difficulty((I_{DIK1}, I_{DIK2}), I_{DIK3^{new}}, E)$

$< T_{Difficulty^W}$

Else：

$T_{Difficulty}((I_{DIK1}, I_{DIK2}), I_{DIK3^{new}}) > T_{Difficulty^W}$

2.2.3 隐私资源的加载

隐私资源的加载指将提取得到的 P_{DIK} 插入到储存介质中的过程，P_{DIK} 将以 $DIKW_{Graph}$ 的形式被储存，包括个体 DIKW 图谱 $DIKW_{Graph}$ 和群体 DIKW 图谱 $DIKW_{Graph^G}$，它们共同存放在一个可供访问和增删查改的介质中。

$DIKW_{Graph}$ 与 E 对应，$DIKW_{Graph^G}$ 与 $G_{Intimacy}$ 对应。$DIKW_{Graph^G}$ 与 $G_{Intimacy}$ 中所有 E 的 $DIKW_{Graph}$ 都存在关联，同时一个实体 E 可以根据不同的属性划分而属于多个 $G_{Intimacy}$，单个 $DIKW_{Graph}$ 也同时与多个 $DIKW_{Graph^G}$ 存在关联。

2.3 隐私资源的价值

隐私是一个大的范畴，狭义上指个体对自我资源的控制权，广义上则代表着许多不同的利益和价值，它们彼此不同甚至存在冲突。隐私的价值包括公平性（Fairness）、人身安全（Personal Security）、财产安全（Financial Security）、安宁（Peace and Quiet）、自主（Autonomy）、反商品化道德（Integrity Against Commodification）和声誉（Reputation）[21]，是一个多元化的概念。隐私价值保护是隐私保护工程的一部分。

隐私价值保护标准 $P_{DIK}(value)^W$ 如式(9)所示，公平和反商品化道德的计算结果为二元布尔值（True/False），其余项根据决策事件的不同性质被设置为不同的阈值。在任意一个决策行为中，当且仅当公平性指数和反商品化道德指数为真，其余项的隐私价值指数大于各自的阈值时，表示决策行为没有违背隐私价值，决策行为合法。

$$P_{DIK}(value)^W = \{\text{True}, V_{PS^W}, V_{FS^W}, V_{PQ^W}, V_{autonomy^W}, True, V_{reputation^W}\} \tag{9}$$

2.3.1 公平性

对于人工智能系统而言，对事件 $Event$ 的自动化决策应该严格遵循事件的决策规则 $Event(rule)$，以及保证受 $Event$ 影响的不同个体受到公平的对待[22]。如式(10)所示，构造函数 $F_{fairness}$ 计算过程的公平性指数 $V_{fairness}$。

$$V_{fairness} = F_{fairness}(Event, P_{DIK(G)}, U_{price}) \tag{10}$$

其中，$P_{DIK(G)}$ 表示 AI 系统在决策中所依据的隐私资源组；U_{price} 表示用户在事件中需要付出的代价，包括时间代价（$Price_{time}$，P_{time}）和物质代价（$Price_{material}$，$P_{material}$）。

例如，在 $Event_1$：“奖学金评比”的决策中，$P_{DIK(G)}$ 应包含所有候选者同样的申请资料和证明材料，不得有多余也不得有缺漏。而关于申请材料最晚提交时间 $Deadline$ 的设置，则需要考虑用户所要付出的不同的时间代价 P_{time}，否则 $V_{fairness}$ 为假，决策行为不合法，具体如下：

Decision($Event_1$)：

K_{DIK1} From $K_{Graph}(U_a)$，K_{DIK2} from $K_{Graph}(U_b)$

K_{DIK1} = “U_a 需要两个小时收集申请资料”

K_{DIK2} = “U_b 需要两天的时间从家中邮寄申请资料”

If $Deadline$ < “Two day”：

$V_{fairness} = F_{fairness}(Deadline, P_{DIK(G)}, K_{DIK1}, K_{DIK2})$

= False

除了 P_{time}，AI 系统还需要考虑用户在决策事件中需要付出的物质代价 $P_{material}$。例如 $Event_2$：“学生考试座位安排”的决策过程如下：

Decision($Event_2$)：

I_{DIK1} From $I_{Graph}(U_a)$，I_{DIK2} from $I_{Graph}(U_b)$

I_{DIK1} = “U_a 是正常考生”

I_{DIK2} = “U_b 是需要拄拐行走的残疾考生”

$U_a(P_{material}) < U_b(P_{material})$

$V_{fairness}$ = False

在这种情况下，AI 系统需要采取措施来解决 $U_a(P_{material})$ 和 $U_b(P_{material})$ 之间的差距问题，例如为 U_b 安排位于一楼的考场，或是赋予 U_b 提前进考场的权力。

由于 $DIKW_{Graph}$ 的两次更新之间存在一定间隔，AI 系统对隐私资源的提取存在一定的延迟，$DIKW_{Graph}$ 并不包含用户当前的所有隐私资源，导致自动化决策过程忽略了特殊情况下用户的需求。

例如，在 $Event_1$ 的决策中，用户 U_c 因病住院而错过了奖学金申请的通知，而 $DIKW_{Graph}(U_c)$ 因为延迟性未收录这一资源，导致 AI 系统在决策的过程中忽略了 $U_c(P_{material})$。针对此类情况，AI 系统应配备“申请—校验—批准”机制来处理用户关于意外情况的申请。

2.3.2 人身安全

虚拟社区中，如果 P_{DIK} 包含了家庭住址、出行方式、行程计划等内容，则 P_{DIK} 的泄露会对用户个体的人身安全（Personal Security）造成威胁。如式(11)所示，构造函数 F_{PS} 计算人身安全指数 V_{PS}。

$$V_{PS} = F_{PS}(E, P_{DIK(G)^T}) \tag{11}$$

其中，E 表示对 P_{DIK} 提出访问的实体，包括实体身份 $E(ID)$ 和访问意图 $E(Purpose)$。AI 系统将对访问者提供的身份证明进行真实性检验，确认访问者是否具有访问资格。检验完成之后再根据 $E(Purpose)$ 决定传输给访问者的隐私资源组（$P_{DIK(G)^{Trasfer}}$，$P_{DIK(G)^T}$）。$P_{DIK(G)^T}$ 需要经过安全性验证，确认传输完成后用户的人身安全指数 V_{PS} 是否大于设定的标准阈值 V_{PS^W}。

$F_{PS}(E, P_{DIK(G)^T})$：

ID = Vertiy($E(ID)$)

While(ID):

$P_{DIK(G)^T}$ = Analyse($E(Purpose)$)

V_{PS} = Check($P_{DIK(G)^T}$)

Return($V_{PS} > V_{PS^W}$)

2.3.3 财产安全

虚拟社区中的存在着许多可反映用户经济状况的 P_{DIK} 以($Finance_{DIK}$, F_{DIK})的形式表示，包含 D_{DIK}, I_{DIK}, K_{DIK} 3 种类型，具体如下。

$F_{DIK^{(D)}}$ =“用户 U_a 在虚拟社区中发布的一张园林别墅的照片”

$F_{DIK^{(I)}}$ =“用户 U_a 拥有一座园林别墅”

$F_{DIK^{(K)}}$ =“用户 U_a 属于高收入人群”

F_{DIK} 的泄露或是不正当使用会对用户财产安全(Financial Security)造成威胁。如式(12)所示，构造函数 F_{FS} 计算 AI 系统决策过程中用户的安财产全指数 V_{FS}。财产安全保护同人身安全保护一样，需要进行 $E(ID)$ 真实性验证和 $P_{DIK(G)^T}$ 安全性验证。

$$V_{FS} = F_{FS}(E, P_{DIK(G)^T}, GP_{content}) \tag{12}$$

在自动化决策过程中，AI 系统需要考虑部分 F_{DIK} 具有群体隐私 $GP_{content}$ 的特性，财产安全攻击者 $Hacker$ 的目标可分为 3 类：1) F_{DIK} 所反映的富有用户群体；2)亲密群体 $G_{Intimacy}$ 中资产排名在高位的用户；3)资产排名高位的用户所在的以财富为划分标准的另一个亲密关系群体 $G_{Intimacy}$ 中的其他用户。

例如，$Hacker$ 已知 $GP_{content1}$，那么 I_{DIK1}, I_{DIK2} 和 K_{DIK1} 的结合可能会对 $G_{Intimacy2}$ 中的用户个体产生影响，具体如下：

$G_{Intimacy1}$ =“$Hacker$ 经过推测已得知经济状况的用户群体”

$GP_{content1}$ =“$G_{Intimacy1}$ 中用户的资产排名”

I_{DIK1} = Transformation($GP_{content1}$)

=“U_a 是 $G_{Intimacy1}$ 中财产最多的人”

K_{DIK1} =“用户 Ua 属于高收入人群”

I_{DIK2} =“用户 Ua 的家庭住址 $Address_1$”

$G_{Intimacy2}$ =“用户 Ua 和家庭住址 $Address_1$ 附近的邻居”

$GP_{content2^{(K)}}$ = Transformation(K_{DIK1}, I_{DIK1})

=“$Address_1$ 附近属于高收入人群居住区”

$GP_{content1}$ 有利于 $Hacker$ 锁定犯罪目标范围，隐私泄露将威胁到 $G_{Intimacy2}$ 中个体的财产安全。因此，在式(12)的计算中，需要评估 AI 系统是否针对群体隐私的特性采取了保护措施。

$F_{FS}(E, P_{DIK(G)^T})$:

ID = Vertiy($E(ID)$)

While(ID):

$P_{DIK(G)^T}$ = Analyse($E(Purpose)$)

For P_{DIK} in $P_{DIK(G)^T}$:

If $P_{DIK} \in GP_{content}$:

$GP_{(G)} = GP_{(G)}$.Append(P_{DIK})

V_{FS} = Check($P_{DIK(G)^T}$, $GP_{(G)}$)

Else:

V_{FS} = Check($P_{DIK(G)^T}$)

Return($V_{FS} > V_{FS^W}$)

2.3.4 安宁

虚拟社区的存在为很多用户提供了一个逃脱现实世界的“窗口”，在虚拟社区中，用户在心理上保持着一个与现实世界中的真实形象不一样的虚拟形象，并且想要保持这种状态不被打扰。许多用户不希望虚拟社区中的其他用户知道自己在真实世界中的身份，也不希望现实生活中的联系人知道自己在虚拟世界中的账户。AI 系统如果对虚拟社区中的 P_{DIK} 使用得当，则能够在达到决策目的的同时给用户一种不被打扰的安宁(Peace and Quiet)的状态。

用户对现实世界和虚拟世界双重身份的追求中的一部分能够在互斥隐私资源组 $P_{DIK(G)^{in}}$ 上得到体现，具体如下：

I_{DIK1}, I_{DIK2} From $DIKW_{Graph}(U_a)$

I_{DIK1} =“用户 U_a 是一名小学生”

I_{DIK2} =“用户 U_a 是一名大学生”

(I_{DIK1}, I_{DIK2}) $\in P_{DIK(G)^{in}}$

如果 U_a 在现实生活中是一名小学生，则 I_{DIK1} 属于现实类隐私资源 $Reality_{DIK}$，I_{DIK2} 属于虚拟类隐私资源 $Virtual_{DIK}$。$P_{DIK(G)^{in}}$ 不一定同时拥有 $Reality_{DIK}$ 和 $Virtual_{DIK}$，也存在由于时间变动而导致 P_{DIK} 发生改变的可能。

如式(13)所示，构造函数 $Differ$，根据 $Event$ 的性质来判断 $P_{DIK(G)^T}$ 中的元素是属于 $Virtual_{DIK}$ 类型还是 $Reality_{DIK}$ 类型。式(14)构造函数 F_{PQ} 计算决策过程中用户的安宁指数 V_{PQ}。

$$P_{DIK(G)^T} = Differ(Event) \tag{13}$$

$$V_{PQ} = F_{PQ}(Event, P_{DIK(G)^{in}}, P_{DIK(G)^T}) \tag{14}$$

例如，AI 系统对 $Event_1$:“虚拟社区好友推荐”和 $Event_2$:“法律证据收集”两种不同类型的事件的决策过程如下：

Decision($Event_1$):

$P_{DIK(G)^T} = Differ(Event_1) = Virtual_{DIK}$

$V_{PQ} = F_{PQ}(Event, P_{DIK(G)^{in}}, P_{DIK(G)^T})$

Decision($Event_2$):

$P_{DIK(G)^T} = Differ(Event_2) = Reality_{DIK}$

$V_{PQ} = F_{PQ}(Event, P_{DIK(G)^{in}}, P_{DIK(G)^T})$

2.3.5 自主

自主指个人能够自由且有能力地按照自我真正的意愿行事，拥有自己选择自己想要的东西、做自己想做的事的权利，对自主权的保证有利于增进虚拟社区和用户双方的信任[23]。随着大数据技术的发展，隐私保护技术却停滞不前，AI 系统在用户个体隐私资源的收集和使用上产生了“技术越界”问题，代替用户个体作出决定，侵犯了用户的自主权，导致了用户对虚拟社区营销推荐方案的厌恶和不信任。大数据技术对隐私资源的分类和使用决定了用户将看到的内容、广告和享用的社区服务，可能会造成对用户的歧视、偏差或者机会限制等。

为了保证用户在虚拟社区中的自主权，AI 系统基于大数据技术对 P_{DIK} 的决策需要在取得用户同意的情况下进行，并根据实际情况进行适当的调整。例如，在 $Event_1$:“商品推荐”的决策中，AI 系统将根据的商品需求 $Topic(Commodity\ Demand)$ 这一主题下的 K_{DIK} 来进行精准推送，以提高商品推荐的成功率。这一过程首先要取得用户对商品推荐服务的同

意；然后，设定商品推荐门槛值 $K_{DIK}(Val)^W$，$K_{DIK}(Val)$ 值高于 $K_{DIK}(Val)^W$ 的商品属于推荐的范围；最后，授予用户无条件关闭推荐服务的权利。具体如下：

Decision($Event_1$)：

If $User(agree)$=False：

break；

Else：

K_{DIK1} From $K_{Graph}(U_a)$

K_{DIK1}="U_a 需要商品 $Commodity_A$"

$K_{DIK1}(Val)=Validity(K_{DIK1}, P_{DIK}^{associated})$

If $K_{DIK1}(Val) > K_{DIK}(Val)^W$：

$K_{DIK(recommend)}$. Append(K_{DIK1})

商品推荐转化成功率（$K_{DIK(recommend)}$，$R_{recommend}$）和用户对推荐内容的点击关闭的次数 $Time_{cancel}$ 可以反映用户对 AI 推荐系统的接受程度。如式(15)所示，构造函数 $F_{autonomy}$ 计算用户对商品推荐中自我自主权的满意指数 $V_{autonomy}$。若 $V_{autonomy}$ 小于阈值 $V_{autonomy}{}^W$，表示用户对推荐系统的认可程度低，*AI* 系统应降低商品推荐的频率。相反，若 $V_{autonomy}$ 大于 $V_{autonomy}{}^W$，则 AI 系统可适当地增加对用户商品推荐的频率和次数。

$$V_{autonomy}=F_{autonomy}(R_{recommend}, Time_{cancel}) \tag{15}$$

2.3.6 反商品化道德

商品化指将个人、生命或者人性等都视为单纯的商品，以金钱来衡量价值行为。商品化并不适用于所有的领域，在医疗、法律等系统中，如果将人类个体按照年龄、性别、种族或教育经济水平进行分类，提供有区别的服务，则这是对社会道德的一种损害。

例如，在 $Event_1$："法律知识普及宣传方案制定"的决策过程中，若 AI 系统将种族作为一个分类标签，认为黑人的犯罪率高、教育程度低，从而做出对所有黑人加大普法宣传力度的决策，则会违反商品化道德这一基本原则。

如式(16)所示，构造函数 F_{IAC} 对 AI 系统决策过程中的反商品化道德指数 V_{IAC} 进行计算，AI 系统将根据 *Event* 的性质和 $P_{DIK(G)^T}$ 的内容来判断决策过程是否违反了商品化道德的基本原则。

$$V_{IAC}=F_{IAC}(Event, P_{DIK(G)^T}) \tag{16}$$

例如，AI 系统对事件 $Event_1$："法律知识普及宣传"和 $Event_2$："个人信息登记"的决策过程如下：

Decision($Event_1$)：

If $P_{DIK}(race) \in P_{DIK(G)^T}$：

V_{IAC}=False

Decision($Event_2$)：

If $P_DIK(race) \in P_{DIK(G)^T}$：

V_{IAC}=True

2.3.7 声誉

P_{DIK} 与个人声誉密切相关[24]，*Hacker* 通过不正当手段获取 P_{DIK} 或是不正当地使用已获取的 P_{DIK}，对某人进行不正确的刻画或者联想，使其名誉或心理、情感的健康等受到影响的行为，与诽谤和造谣一样，这些行为都属于侵犯隐私行为。如式(17)所示，构造函数 $F_{reputation}$ 计算决策过程中对用户声誉造成影响的指数 $V_{reputation}$。

$$V_{reputation}=F_{reputation}(Event, E, P_{DIK(G)^T}) \tag{17}$$

其中，AI 系统将根据 *Event* 的性质和 P_{DIK} 中的内容评估决策行为是否存在对用户声誉造成影响的可能，若计算结果 $V_{reputation}$ 小于 $V_{reputation}{}^W$，则 AI 系统需要对 $P_{DIK(G)^T}$ 进行声誉审查和不合格 P_{DIK} 的替换。声誉审查包括多项内容，例如 P_{DIK} 的来源审查。一般而言，在会对用户声誉造成影响的 *Event* 决策中，$P_{DIK(G)^T}$ 应来源于用户自身，而非来源于他人对用户的评价。

3 隐私资源的流通

在 AI 系统的自动化决策中，隐私资源 P_{DIK} 的流通过程包括 4 个环节：感知(Sensing)、存储(Storage)、传输(Trasfer)和处理(Processing)。4 个环节共有 3 个参与方：用户(User)、AI 系统和访问者(Visitor)中的一个或多个。User 是 P_{DIK} 的产生者，AI 系统是 P_{DIK} 的存储者和传输中枢，Visitor 是 P_{DIK} 的使用处理者。流通过程涉及的隐私权利包括知情权(Know)、参与权(Participate)、监督权(Supervise)和遗忘权(Forget)，如图 3 所示。

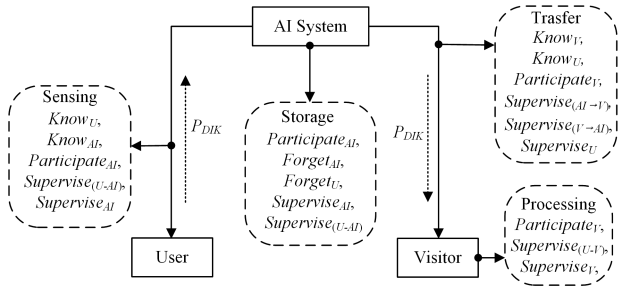


图 3 隐私类型资源的流通过程

Fig. 3 The circulation of $Privacy_{DIK}$

3.1 隐私资源流通过程建模

P_{DIK} 由 *User* 在虚拟社区中的各种交互行为产生，当 *Visitor* 出于自身意图 *Purpose* 向管理虚拟社区的 AI 系统提出访问申请时，隐私资源流通过程启动。各个参与方在不同的隐私流通环节中具有不同的隐私权，隐私权保证是隐私安全保证的必要条件。在 P_{DIK} 的流通过程中，*User* 的隐私权属于固有权利，不与隐私义务捆绑，即使 *User* 不行使自己的隐私权，*User* 的隐私权利仍然保留。

3.1.1 感知过程建模

感知过程指 AI 系统从 *User* 在虚拟社区中的 $T_{virtual}$ 和 UGC 中提取出 P_{DIK}，属于 P_{DIK} 的提取过程。感知过程中涉及到的参与方隐私权如下。

$Know_U$：*User* 对自身所有的 P_{DIK} 被 AI 系统收集感知的知情权。

$Know_{AI}$：AI 系统被允许从用户处感知收集到的 P_{DIK}，仅限于用户知情同意的情况下来自 $T_{virtual}$ 和 UGC 的 P_{DIK}，不包括 AI 系统从其他非法渠道获得的 P_{DIK}。

$Participate_{AI}$：AI 系统在感知过程中对 P_{DIK} 的采集提取方式。$Participation_{AI}$ 包括对 AI 系统感知时间、感知内容、感知方式的规定和限制。

$Supervise_{(U\to AI)}$：*User* 对 AI 系统在感知过程中的行为是

否符合规则的监督权。

$Supervise_{AI}$：AI 系统在感知过程中对自身行为的自我监督。

3.1.2 存储过程建模

存储过程由 AI 系统单独完成，包含了 P_{DIK} 的转换和加载。AI 系统将感知过程中采集得到的 P_{DIK} 代入 Transformation 模块中进行转换，并以 $DIKW_{Graph}$ 和 $DIKW_{Graph^G}$ 的形式将 P_{DIK} 和 $P_{DIK^{new}}$ 存储在可以访问和恢复的介质中。存储过程同时涉及的隐私权如下。

$Participate_{AI}$：AI 系统对 P_{DIK} 的参与处理权，包括对 P_{DIK} 进行转换和存储。

$Forget_{AI}$：AI 系统对于 $DIKW_{Graph}$ 上已失去存储价值的 P_{DIK} 进行遗忘的权利。

$Forget_U$：*User* 要求 AI 系统对 $DIKW_{Graph}$ 上已被更新和表达错误的 P_{DIK} 进行遗忘的权利。

$Supervise_{AI}$：AI 系统在存储过程中的自我监督，包括对 $Participate_{AI}$ 和 $Forget_{AI}$ 的监督。

$Supervise_{(U\to AI)}$：*User* 对 AI 系统在存储过程中是否以合适的方式对 P_{DIK} 进行存储以及是否对过时 P_{DIK} 进行系统遗忘的监督。

3.1.3 传输过程建模

传输过程由 AI 系统和 *Visitor* 合作完成，AI 系统基于 *Visitor* 的访问意图，在隐私价值 $P_{DIK}(value)$ 得到满足的情况下将 $P_{DIK(G)^T}$ 的形式向 *Visitor* 转移，传输过程涉及的隐私权如下。

$Know_V$：*Visitor* 可知情的 P_{DIK} 内容，由 AI 系统根据 *Visitor* 的身份和意图计算分析得到。

$Know_U$：*User* 对自身 P_{DIK} 由 AI 系统向 *Visitor* 传输这一过程的知情权。

$Participate_V$：在传输过程中，AI 系统应明确 *Visitor* 对 $P_{DIK(G)^T}$ 的处理参与权。

$Supervise_{(AI\to V)}$：AI 系统监督 *Visitor* 是否具有访问权、访问意图是否合理的权利。

$Supervise_{(V\to AI)}$：*Visitor* 对 AI 系统传输的 $P_{DIK(G)^T}$ 是否真实可信、是否符合自身合理访问要求的监督权。

$Supervise_U$：*User* 对 P_{DIK} 传输过程的监督权，包括 $Supervise_{(U\to AI)}$，即对 AI 系统输出的 $P_{DIK(G)^T}$ 的监督，以及 $Supervise_{(U\to V)}$，即对 *Visitor* 身份和访问意图的监督。

3.1.4 处理过程建模

处理过程是 *Visitor* 利用和开发在传输过程中得到的 $P_{DIK(G)^T}$ 的过程，涉及的隐私权如下。

$Participate_V$：*Visitor* 处理 P_{DIK} 的权利，其内容与传输过程中 $Participate_V$ 的内容一致。

$Supervise_{(U\to V)}$：*User* 对 *Visitor* 使用自身 P_{DIK} 的方式进行监督的权利。

$Supervise_V$：*Visitor* 对自身如何处理和利用 P_{DIK} 进行自我监督的权利。

3.2 隐私权定义

隐私资源流通过程的 4 个环节中的隐私权包括知情权(Know)、参与权(Participate)、遗忘权(Forget)和监督权(Supervise)。

3.2.1 知情权的定义

知情权[25]指个体知悉、获取隐私资源的自由与权利，在隐私资源流通过程中，知情权根据参与主体的不同有所区分。

知情权 *Know* 包括过程知情权 $Know(course)$ 和内容知情权 $Know(content)$，$Know(course)$ 指对 P_{DIK} 流通过程的知情权，属于二元布尔值属性，是 *User* 作为 P_{DIK} 所有者独有的权利 $Know_U(course)$。如式(18)所示，$Know_U(course)$ 包括对感知、存储、传输、处理 4 个环节的知情。$Know_U(course)$ 为 AI 系统决策行为合法的必要条件。

$$Know_U(course)=\{Sensing^K, Storage^K, Transfer^K, Processing^K\} \tag{18}$$

内容知情权 $Know(content)$ 表示参与实体可知情的隐私资源($Privacy_{DIK^{Know}}$，$P_{DIK^{Know}}$)。如式(19)所示，构造函数 *Know* 根据参与实体 E 的身份意图和参与过程 *process* 计算 E 的可知情隐私资源组 $P_{DIK(G)^{Know}}$。

$$P_{DIK(G)^{Know}}=\mathrm{Know}(E, process) \tag{19}$$

3.2.2 参与权定义

参与权(Participate)指参与实体 E 依照规定对 P_{DIK} 进行决策和使用的权利。参与权是一个大的概念范畴，用参与权元组 *Participation* 表示参与权的内容。如式(20)所示，*Participation* 包括实体 E 参与的形式、参与的次数、参与的截止时间等，是对参与实体 E 的行为的规则限制。*Participation* 由式(21)的函数 *Participate* 根据参与实体 E 和参与的流通环节 *process* 计算得到。

$$\mathrm{Participation}=\{\mathrm{Time}, \mathrm{Form}, \mathrm{Deadline}, \cdots\} \tag{20}$$

$$\mathrm{Participation}=\mathrm{Paticipate}(E, process) \tag{21}$$

User 的参与权 $Paticipate_U$ 主要表现为一票否决权。*User* 有权利对任意过程或任意一项 P_{DIK} 行使一票否决权，阻止 P_{DIK} 的流通。*User* 行使一票否决权后，$Know_U$ 为假，AI 系统的决策行为将被判定为不合法而中止。

AI 系统的参与权 $Paticipate_{AI}$ 包括感知过程中对 P_{DIK} 的收集、存储过程中对 P_{DIK} 的转换、DIKW 图谱的建立和更新以及在传输过程中对 $P_{DIK(G)^T}$ 的决策等。

Visitor 的参与权 $Paticipate_V$ 则包括传输过程中对 AI 系统的访问、P_{DIK} 的收集和处理过程中对 P_{DIK} 的开发利用。

3.2.3 遗忘权定义

遗忘权(Forget)[26]指参与方对 DIKW 图谱上的过时隐私资源($Privacy_{DIK^{old}}$，$P_{DIK^{old}}$)和无价值隐私资源($Privacy_{DIK^{unvalue}}$，$P_{DIK^{unvalue}}$)进行系统性清除遗忘的权利。

$P_{DIK^{old}}$ 指随着时间的推移而被其他语义表达不同的 P_{DIK} 所代替的隐私资源，*User* 拥有对 $P_{DIK^{old}}$ 的遗忘权 $Forget_U$。例如，在 $Event_1$："奖学金评选"的决策如下。已知是否通过英语等级考试是 $Event_1$ 的评判规则之一，而 I_{DIK2} 在感知过程的时间线上晚于 I_{DIK1}。

Decision($Event_1$)：

I_{DIK1}，I_{DIK2} From $I_{Graph}(U_a)$

I_{DIK1}="用户 U_a 未通过英语等级考试"

I_{DIK2}="用户 U_a 通过了英语等级考试"

$\{I_{DIK1}, I_{DIK2}\}\in P_{DIK(G)^{in}}$

$I_{DIK2}()$ is new the I_{DIK1}

$V_{fairness}=F_{f}airness(Event,I_{DIK1},U_{price})=\text{False}$

$V_{fairness}=F_{fairness}(Event,I_{DIK2},U_{price})=\text{True}$

I_{DIK1} 和 I_{DIK2} 彼此之间存在语义冲突,二者的集合属于 $P_{DIK(G)^{in}}$。而 I_{DIK1} 属于 $P_{DIK^{old}}$,在决策中其内容应被 I_{DIK2} 替代。出于保障自身权益的目的,*User* 有权要求系统对 I_{DIK1} 进行遗忘。

AI 系统也有义务定时对 $DIKW_{Graph}(U_a)$ 进行 $P_{DIK^{old}}$ 的系统清除遗忘,以避免 AI 系统在 $Event_1$ 决策上违背保障隐私的公平价值这一原则。

除了 *User* 之外,AI 系统拥有对 $P_{DIK^{unvalue}}$ 的遗忘权 $Forget_{AI}$。随着时间的推移,$DIKW_{Graph}$ 上所存储的 P_{DIK} 数量会呈几何式增长,AI 系统需要负担越来越大的存储成本,同时 $DIKW_{Graph}$ 上过多的 P_{DIK} 也会导致自动化决策时搜索成本上升。经过时间变化后一部分 P_{DIK} 将发生价值下降,成为无价值的隐私资源 $P_{DIK^{unvalue}}$,AI 系统对 $P_{DIK^{unvalue}}$ 的遗忘是降低运营成本和保证决策效率的必要之举。如式(22)所示,当存储 P_{DIK} 所要付出的成本 $P_{DIK}(cost)$ 高于阈值 $P_{DIK}(cost)^{W}$ 时,P_{DIK} 属于 $P_{DIK^{unvalue}}$。

$$P_{DIK^{unvalue}}=\{P_{DIK}\mid P_{DIK}(cost)>P_{DIK}(cost)^{W}\} \tag{22}$$

为了防止 $P_{DIK^{old}}$ 和 $P_{DIK^{unvalue}}$ 对决策产生影响,AI 系统将设定遗忘周期 $(Cycle_{forget},C_{forget})$,每经过一个 C_{forget} 周期,$DIKW_{Graph}$ 需要进行一次系统性遗忘。

如式(23)所示,由构造函数 Forget 将 $DIKW_{Graph}$ 上的 P_{DIK} 分类,$P_{DIK(G)^{F}}$ 表示经过计算后被遗忘的 P_{DIK} 集合,$P_{DIK(G)^{R}}$ 表示被保留的 P_{DIK} 集合。

$$\{P_{DIK(G)^{F}},P_{DIK(G)^{R}}\}=\text{Forget}(DIKW_{Graph}) \tag{23}$$

3.2.4 监督权定义

监督权(Supervise)是 3 个参与方都拥有的权利,包括对知情权的监督($Supervise_{know}$,记为 S_{know})、参与权的监督($Supervise_{participate}$,记为 $S_{paticipate}$)和遗忘权的监督($Supervise_{forget}$,记为 S_{forget})。S_{know},$S_{paticipate}$ 和 S_{forget} 的系统初始默认值为真,当参与方提出合理异议时转为假,AI 系统的决策行为不合法。

(1)监督知情权

对知情权的监督 S_{know} 与知情权内容对应,包括对知情过程的监督 $S_{know}(course)$ 和对知情内容的监督 $S_{know}(content)$。$S_{know}(course)$ 和 $S_{know}(content)$ 皆为真时,S_{know} 为真,如式(24)所示:

$$S_{know}=S_{know}(course)\&\&S_{know}(content) \tag{24}$$

$S_{know}(course)$ 是隐私价值保护的一部分。例如,在 $Event_1$:"药品 $Drug_1$ 的人体实验志愿者报名审核"的决策中,已知 $Event_1(rule)$ 如下:

$Event_1(rule_1)$:"参与人体实验的志愿者可获得 XX 元的补贴"

$Event_1(rule_2)$:"参与者需为自愿报名"

Decision($Event_1$):

K_{DIK1} From $K_{Graph}(U_a)$,K_{DIK2} From $K_{Graph}(U_b)$

$K_{DIK1}=K(U_a,Event_1)=$"用户 U_a 具有一定的医学知识,能够了解 $Event_1$ 背后的风险"

$K_{DIK2}=K(U_a,Event_1)=$"用户 U_b 需要用钱,且不具有了解 $Event_1$ 风险的条件"

$S_{know}(course)=\text{Supervise}(Event_1(rule),K_{DIK1},K_{DIK2})=\text{False}$

$U_{price}(U_a)<U_{price}(U_b)$

$V_{fairness}=\text{F}_{\text{fairness}}(Event_1,P_{DIK(G)},U_{price})=\text{False}$

出于对 $Event_1$ 决策过程中公平性的保证,AI 系统应在志愿者报名页面中设置风险提醒通知书,保证所有参与用户对 $Event_1$ 存在的风险的知情权。

知情内容监督 $S_{know}(content)$ 如式(25)所示,将传输隐私资源组 $P_{DIK(G)^{T}}$ 与经过式(19)计算得到的 $P_{DIK^{Know}}$ 代入监督函数 *Supervise* 进行比较计算。比较结果一致时 $S_{know}(content)$ 为真,结果不一致时表示参与方存在知情越权或是知情权未得到满足的情况,$S_{know}(content)$ 为假。

$$S_{know}(content)=Supervise(P_{DIK(G)^{T}},P_{DIK^{Know}}) \tag{25}$$

(2)监督参与权

对参与权的监督 $S_{paticipate}$ 指监督参与实体是否按照 *Participation* 中的规则对 P_{DIK} 进行采集提取和使用处理。监督的内容包括参与处理的次数是否超标、参与的形式是否在允许范围内、是否在规定的时间内完成等。

将决策过程中参与实体 E 的实际参与内容($Participation^{Real}$)与 *Participation* 代入监督函数 *Supervise* 进行比较计算,如式(26)所示。若 $Participation^{Real}$ 未超出 *Participation* 规定的范围,则 $S_{paticipate}$ 为真。

$$S_{paticipate}=Supervise(Participation^{Real},Participation) \tag{26}$$

(3)监督遗忘权

对遗忘权的监督 S_{forget} 包括对遗忘权使用的监督 $S_{forget}(course)$ 和遗忘的内容 $S_{forget}(content)$ 的监督,如式(27)所示:

$$S_{forget}=S_{forget}(course)\&\&S_{forget}(content) \tag{27}$$

$S_{forget}(course)$ 指监督 AI 系统是否根据 C_{forget} 周期定时对 $DIKW_{Graph}$ 进行更新优化,每一个 C_{forget} 开始时 $S_{forget}(course)$ 为假,C_{forget} 结束时若遗忘清除完成,则 $S_{forget}(course)$ 为真。

$S_{forget}(content)$ 指将 $P_{DIK^{old}}$ 和 $P_{DIK^{unvalue}}$ 相加,与经过式(23)计算得到的 $P_{DIK(G)^{F}}$ 一起代入函数 *Supervise* 进行比较计算,如式(28)所示。比较结果一致时,$S_{forget}(content)$ 为真。

$$S_{forget}(content)=\text{Supervise}(P_{DIK(G)^{F}},P_{DIK^{old}}+P_{DIK^{unvalue}}) \tag{28}$$

4 隐私资源的保护

出于对隐私价值和隐私权的保护,AI 系统设计了匿名保护机制、风险评估机制和监督机制,目的在于促使 AI 系统的决策符合社会道德和法律的要求。

4.1 匿名保护机制

匿名是不可识别性的一种形式,是对资源属性的一种隐藏,而不仅限于名称的改变[27]。AI 系统匿名保护机制指基于 P_{DIK} 的性质,将 P_{DIK} 的某一部分内容隐藏或是以符号代替表达,从而达到隐私保护的目的。匿名保护机制可分为数据匿名保护、信息匿名保护、知识匿名保护和群体匿名保护 4 种类型。

4.1.1 数据匿名保护

数据匿名保护指对于内容敏感或容易造成歧视和伤害的数据类型隐私资源，采用代参数加密的方法来降低隐私泄露的风险。$P_{DIK(G)^T}$ 中的 D_{DIK} 在传输过程开始之前由 AI 系统按照一定的规则进行加密，输出完成后由 *Visitor* 进行解密，目的在于降低传输过程中隐私泄露的风险。

例如，针对 HIV 检测结果 $Test_{HIV}$ 这一内容敏感 D_{DIK}，用代参数 a 进行匿名加密。用 D_{DIK^A} 来表示匿名加密后的结果，“$a=0$”表示 HIV 检测结果为阴性，“$a=1$”表示 HIV 检测结果为阳性。

D_{DIK1} =“$Test_{HIV}=Negative$”

D_{DIK2} =“$Test_{HIV}=Positive$”

D_{DIK1^A} =“$a=0$”

D_{DIK2^A} =“$a=1$”

在传输过程中，$P_{DIK(G)^T}$ 中原有的 D_{DIK} 将被 D_{DIK^A} 替换。具有访问资格的专业访问者 $Visitor_{pro}$ 拥有将 D_{DIK^A} 还原回 D_{DIK} 的能力，而 *Hacker* 无论是通过非正当的途径逃过了 AI 系统的身份审查，还是由于传输过程中的意外而得到了 D_{DIK^A}，其自身不具有将匿名数据还原的能力，从而在一定程度上达到隐私保护的目的。

4.1.2 信息匿名保护

信息匿名保护指通过隐藏 I_{DIK} 中的某一实体 E 或关系 R 来达到隐私保护目的的方法。I_{DIK} 中被隐藏的内容根据 *Visitor* 的访问意图和 I_{DIK} 内容的敏感度而定。例如，I_{DIK1} 表示的是病人实体是否患有艾滋病，出于 *Visitor* 不同的 *Purpose*，I_{DIK1} 经过匿名化处理后得到的 I_{DIK^A} 内容不同：

I_{DIK1} =“用户 U_a 的 HIV 检测结果为阳性”

$Purpose_1$ =“艾滋病例研究”

I_{DIK1^A} =“用户 XX 的 HIV 检测结果为阳性”

$Purpose_2$ =“检测结果统计”

I_{DIK1^A} =“用户 U_a 的 XX 检测结果为阳性”

$Purpose_3$ =“HIV 检测用户统计”

I_{DIK1^A} =“用户 U_a 的 HIV 检测结果为 XX”

根据 $Purpose_1$，$Purpose_2$ 和 $Purpose_3$ 3 种不同的访问意图，AI 系统分别隐去了 I_{DIK1} 中的两个实体 E 和关系 R，以减少在 $P_{DIK(G)^T}$ 转移过程中无关隐私内容的泄露。

4.1.3 知识匿名保护

知识类型资源 K_{DIK} 是对实体 E 的概括以及对未来行为的预测，同一个主题 *Topic* 下可能包含多个具有不同值 $K_{DIK}(Val)$ 的 K_{DIK}。

例如，K_{DIK1} 和 K_{DIK2} 同属于 $K_{Graph}(U_a)$ 下的药物需求主题 $Topic(Medicine\ Demand)$。

K_{DIK1}，K_{DIK2} Form $K_{Graph}(U_a)$

$\{K_{DIK1}, K_{DIK2}\} \in Topic(Medicine\ Demand)$

$K_{DIK1}=K(U_a, Medicine)$

=“用户 U_a 需要购买药物(M_1：齐多夫定)”

$K_{DIK1}(Val)=75\%$

$K_{DIK2}=K(U_a, medicine)$

=“用户 U_a 需要购买药物(M_2：去羟基苷)”

$K_{DIK2}(Val)=25\%$

M_1 和 M_2 都是治疗艾滋病的常规药物，患者对药物的适用性和需求视自身身体状况和过往病史而定，K_{DIK1} 和 K_{DIK2} 的准确属性值 $K_{DIK1}(Val)$ 和 $K_{DIK2}(Val)$ 也因此有所区分。由 $K_{DIK}(Val)$ 值可知，U_a 对 M_1 的购买需求大于 M_2。K_{DIK1} 和 K_{DIK2} 经过匿名化处理后得到的 K_{DIK1^A} 和 K_{DIK2^A} 为：

$K_{DIK1^A}=K(U_a, Medicine)$，$K_{DIK1^A}(Val)=$XX

$K_{DIK2^A}=K(U_a, medicine)$，$K_{DIK2^A}(Val)=$XX

当处理过程中 *Visitor* 向 U_a 提供推荐服务时，将根据被隐去 $K_{DIK}(Val)$ 的属性值的 K_{DIK1^A} 和 K_{DIK2^A} 对 M_1 和 M_2 提供同等推荐服务，U_a 能够根据自己的实际情况决定购买的药物种类，保证了用户拥有根据自身需求来选择购买药品的自主性。

4.1.4 群体匿名保护

匿名保护机制不仅可以作用于单个 P_{DIK}，也可以作用于群组 $P_{DIK(G)}$。例如，在评选类决策中，$P_{DIK(G)^T}$ 需要包含有关参选者身份的资源 $P_{DIK}(ID)$，用于后续评选结果的通知或公示，而出于公平性的目的，$P_{DIK}(ID)$ 又需要排除在评比决策所参考的 P_{DIK} 之外以防止舞弊行为的发生。此时，AI 系统需启动匿名保护机制。

例如，在 $Event_1$：“奖学金评比”的决策中，$Event_1(rule)$ 如下：

$Event_1(rule_1)$ =“以成绩排名选前 n 名”

$Event_1(rule_2)$ =“不同专业分开排名”

$Event_1(rule_3)$ =“按照公平公正的原则评比”

$Event_1$ 决策所依据的 $P_{DIK(G)^T}$ 包括学生的姓名、年龄、性别、籍贯、专业、成绩等，存在无关 P_{DIK} 妨碍决策公平性的可能。而经过匿名化方法处理后的 $P_{DIK(G)^{T(A)}}$ 可将无关 P_{DIK} 对公平性的影响降到最低。

Decision($Event_1$)：

If $P_{DIK(G)^T}$ = {Name, Age, Gender, Address, Major, Grade, …}：

$V_{fairness}=F_{fairness}(Event_1, P_{DIK(G)^T}, Uprice)$ =False

$P_{DIK(G)^{T(A)}}$ =Anonymity($P_{DIK(G)^T}$)

={XX, XX, XX, XX, Major, Grade, …}

If $P_{DIK(G)^T}==P_{DIK(G)^{T(A)}}$：

$V_{fairness}=F_{fairness}(Event_1, P_{DIK(G)^T}, Uprice)$=True

经过匿名化处理后的 $P_{DIK(G)^{T(A)}}$ 最大限度地排除了无关 P_{DIK}，无论对 $Event_1$ 的决策是由 AI 系统完成还是由工作人员人工完成，其对评选公平性的影响都将被降至最低。

4.2 风险评估机制

风险评估机制的作用是测试传输过程中的 $P_{DIK(G)^T}$ 是否存在因为 Transformation 而造成 P_{DIK} 泄露的风险。该机制基于 P_{DIK} 的转换难度 $T_{Difficulty}$ 和自身属性计算风险值 V_{Risk}。当计算结果大于设定的阈值 V_{Risk^W} 时，表示隐私泄露的风险过大，AI 系统需要删除或是替换 $P_{DIK(G)^T}$ 中导致 V_{Risk} 值增大的 P_{DIK} 来保证隐私安全。

风险评估机制主要基于 P_{DIK} 自身的属性警醒评估计算，包括 P_{DIK} 的转换难度 $T_{difficulty}$、转换入度 $P_{DIK}(in)$ 和出度 $P_{DIK}(out)$，以及知识类型资源 K_{DIK} 的精确度 $K_{DIK}(Pre)$。

4.2.1 关于转换难度的风险评估

关于转换难度的风险评估是评估 *Visitor* 是否具有将 P_{DIK} 转换为其他与访问意图无关且会对用户造成隐私侵犯的 $P_{DIK^{new}}$ 的能力。

若根据式(8)可计算得到 $T_{difficulty}$ 为无穷大,则表示 *Visitor* 不具有将 P_{DIK} 转换为 $P_{DIK^{new}}$ 的能力,P_{DIK} 泄露的概率低,V_{Risk} 值小于 V_{Risk^W},$P_{DIK(G)^T}$ 允许传输;若 *Visitor* 具有将 P_{DIK} 转换为 $P_{DIK^{new}}$ 的能力,则无关 P_{DIK} 有一定被泄露的概率,该 P_{DIK} 应被替换或删除。具体如下:

D_{DIK1}, I_{DIK1} From $DIKW_{Graph}(U_a)$

D_{DIK1} =“虚拟社区中的一张照片 $Photo_1$”

I_{DIK1} =“照片 $Photo_1$ 中由 U_a 所拍摄”

$I_{DIK2^{new}}$ =“照片 $Photo_1$ 的拍摄地点”

If $E==Visitor_{pro}$:

$T_{Difficulty}(I_{DIK1}, I_{DIK2^{new}}) = Difficulty(I_{DIK1}, I_{DIK2^{new}}, E) < T_{Difficulty^W}$

$V_{Risk} < V_{Risk^W}$

If $E==Visito_{rcommon}$:

$T_{Difficulty}(I_{DIK1}, I_{DIK2^{new}}) = Difficulty(I_{DIK1}, I_{DIK2^{new}}, E) > T_{Difficulty^W}$

$V_{Risk} > V_{Risk^W}$

4.2.2 关于隐私资源属性的风险评估

$T_{difficulty}$ 对计算条件的要求高,在一般的决策中,AI 系统无法做到对 *Visitor* 的能力完全了解和计算出精确的 $T_{difficulty}$ 数值。仅基于 $T_{Difficulty}$ 的风险评估很难完全规避 P_{DIK} 泄露的风险。

P_{DIK} 的转换入度 $P_{DIK}(in)$ 和出度 $P_{DIK}(out)$ 可表示 P_{DIK} 在 $DIKW_{Graph}$ 中的连通性。当同时有多个 P_{DIK} 可以满足 *Visitor* 的访问意图时,AI 系统应该选择在 $P_{DIK}(in)$ 和 $P_{DIK}(out)$ 两项属性上综合数值最小、连通性最差的 P_{DIK} 加入 $P_{DIK(G)^T}$,这将减小 P_{DIK} 可能的被转换次数和 P_{DIK} 泄露的风险。

除此之外,AI 系统还可以根据 K_{DIK} 的精确度 $K_{DIK}(Pre)$ 进行选择,在满足 *Visitor* 访问意图的前提下,$K_{DIK}(Pre)$ 值越小的 K_{DIK} 所包含的资源内容量越少,隐私泄露的风险也越小。

4.3 监督机制

隐私流通过程中的监督机制除各个参与方所拥有的监督权外,还包括 AI 系统对决策逻辑的监督和隐私价值的监督。如式(29)所示,当逻辑监督值($Supervise_{logic}$,S_{logic})、价值监督值($Supervise_{value}$,S_{value})和权利监督($Supervise_{right}$,S_{right})皆为真时,监督过程完成,决策行为合法。

$$Supervise = S_{logic} \&\& S_{value} \&\& S_{right} \tag{29}$$

4.3.1 逻辑监督

逻辑监督指对于事件决策过程中是否出现了逻辑错误的监督,将决策规则 $Event(rule)$ 和决策结果 $Event(result)$ 代入函数 *Supervise* 进行比较计算,如式(30)所示。S_{logic} 为真表示逻辑监督结果合格。

$$S_{logic} = Supervise(Event(rule), Event(result)) \tag{30}$$

例如,在 $Event_1$:“工作量分配”的决策中,规则 $Event_1(rule)$ 如下:

$Event_1(rule_1)$:“工作量总计为 100 份”

$Event_1(rule_2)$:“工作人员人数为三人”

$Event_1(rule_3)$:“按照工作人员的处理能力效率进行工作量分配”

Decision($Event_1$):

$Event_1(result) = \{U_a:20, U_b:40, U_c:50\}$

$S_{logic} = Supervise(Event_1(rule), Event_1(result))$ =False

4.3.2 价值监督

价值监督指 AI 系统对决策行为是否违背了隐私价值的监督,如式(31)所示,将决策过程中隐私价值的计算结果 $P_{DIK}(value)$ 和式(7)中的隐私价值标准 $P_{DIK}(value)^W$ 两项内容代入函数 *Supervise* 进行比较计算。当 $P_{DIK}(value)$ 满足 $P_{DIK}(value)^W$ 的要求,即 $V_{Fairness}$ 和 V_{IAC} 为真,V_{PS},V_{FS},V_{PQ},$V_{Autonomy}$ 和 $V_{Reputation}$ 大于各自对应的阈值时,S_{value} 为真,价值监督结果合格。

$$S_{value} = Supervise(P_{DIK}(value), P_{DIK}(value)^W) \tag{31}$$

4.3.3 权利监督

权利监督是对隐私流通过程中 4 种参与权的监督,包括知情权监督 S_{know}、参与权监督 $S_{participate}$、遗忘权监督 S_{forget}。如式(32)所示,当 S_{know},$S_{participate}$ 和 S_{forget} 三者为真时,权利监督值 S_{right} 为真,权利监督结果合格。

$$S_{right} = S_{know} \&\& S_{participate} \&\& S_{forget} \tag{32}$$

结束语 AI 系统对于隐私资源的法律保护框架按照以下的步骤来搭建。首先,明确隐私资源流通过程中各个参与方的隐私权,确保 AI 系统从合法的渠道提取到隐私资源,并以合法的方式存储和更新 DIKW 图谱。然后,当访问者提出访问需求时,AI 系统需要验证访问者身份的真实性,并根据访问者的身份和意图,以及决策事件的性质来决定传输隐私资源组的内容。最后,传输隐私资源组在输出之前需要经过匿名保护机制、风险评估机制和监督机制的检验,满足所有条件时允许输出。未来的工作,计划对隐私权进行进一步的定细分定义。

参 考 文 献

[1] O'NEILL E, KOSTAKOS V, KINDBERG T, et al. Instrumenting the City: Developing Methods for Observing and Understanding the Digital Cityscape [C]// Lecture Notes in Computer Science. 2006, 4206: 315-322.

[2] GIRARDIN F, CALABRESE F, FIORE F D, et al. Digital Footprinting: uncovering tourists with user-generated content [J]. IEEE Pervasive Computing, 2008, 7(4): 36-43.

[3] HU H, GE Y J, HOU D Y. Using Web Crawler Technology for Geo-Events Analysis: A Case Study of theHuangyan Island Incident[J]. Sustainability, 2014, 6(4): 1896-1912.

[4] KRUMM J, DAVIES N, NARAYANASWAMI C. User-generated content[J]. IEEE Pervasive Computing, 2008, 7(4): 10-11.

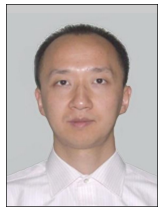
[5] LEI Y, DUAN Y. Personality Classification and Conversion Method of Virtual Community Personnel Based on DIKW Graph [J]. Chinese Journal of Applied Sciences, 2020, 38(5): 803-824.

[6] ULRIKE H. Reviewing person's value of privacy of online social networking[J]. Internet Research, 2011, 21(4): 384-407.

[7] BOZDA E. Bias in algorithmic filtering and personalization[J].

Ethics & Information Technology, 2013, 15(3): 209-227.
[8] MITTELSTADT B D, ALLO P, TADDEO M, et al. The ethics of algorithms: mapping the debate[J]. Big Data and Society, 2016, 3(2): 1-21.
[9] DUAN Y C, ZHAN L G, ZHANG XY, et al. Formalizing DIKW Architecture for Modeling Security and Privacy as Typed Resources[C]// Testbeds and Research Infrastructures for the Development of Networks and Communities. 2018: 157-168.
[10] DUAN Y C, SHAO L X, HU G Z, et al. Specifying architecture of knowledge graph with data graph, information graph, knowledge graph and wisdom graph[C]// IEEE International Conference on Software Engineering Research. IEEE, 2017.
[11] SONG Z Y, DUAN Y C, WAN S X, et al. Processing Optimization of Typed Resources with Synchronized Storage and Computation Adaptation in Fog Computing[J]. Wireless Communications and Mobile Computing, 2018, 2018: 1-13.
[12] GURSES S, ALAMO J M D. Privacy Engineering: Shaping an Emerging Field of Research and Practice[J]. IEEE Security & Privacy, 2016, 14(2): 40-46.
[13] DUAN Y C, SUN X B, CHE H Y, et al. Modeling Data, Information and Knowledge for Security Protection of Hybrid IoT and Edge Resources[J]. IEEE Access, 2019, 7: 99161-99176.
[14] DUAN Y C. Existence Computation : Revelation on Entity vs. Relationship for Relationship Defined Everything of Semantics [C]// 2019 20th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD). 2019: 139-144.
[15] OLSON D L. Data mining in business services[J]. Service Business, 2007, 1(3): 181-193.
[16] KURGAN L A, MUSILEK P. A survey of Knowledge Discovery and Data Mining process models[J]. The Knowledge Engineering Review, 2006, 21(1): 1-24.
[17] VISSER W, HINDRIKS K V, JONKER C M. Argumentation-Based Qualitative Preference Modelling with Incomplete and Uncertain Information[J]. Group Decision & Negotiation, 2012, 21(1): 99-127.
[18] MITTELSTAD B. From Individual to Group Privacy in Big Data Analytics[J]. Philosophy & Technology, 2017, 30(4): 475-494.
[19] VASSILIADIS P. A Survey of Extract-Transform-Load Technology[J]. International Journal of Data Warehousing & Mining, 2009, 5(3): 1-27.
[20] HARIRI R H, FREDERICKS E M, BOWERS K M. Uncertainty in big data analytics: survey, opportunities, and challenges[J]. Journal of Big Data, 2019, 6(1): 44.
[21] HARPER J. Privacy and the Four Categories of Information Technology[R]. American Enterprise Institute Report, 2020.
[22] EKSTRAND M D, JOSHAGHANI R, MEHRPOUYAN H. Privacy for All: Ensuring Fair and Equitable Privacy Protections [C]// Proceedings of the 1st Conference on Fairness, Accountability and Transparency. 2018: 35-47.
[23] HOFFMAN S, PODGURSKI A. Balancing Privacy, Autonomy, and Scientific Needs In Electronic Health Records Research[J]. SMU law review: a publication of Southern Methodist University School of Law, 2012, 65(1): 85-144.
[24] SOLOVE D J. The Future of Reputation: Gossip, Rumor, and Privacy on the Internet[M]// Yale University Press, 2007.
[25] GORMAN D. Rights in Collision: The Individual Right of Privacy and the Public Right To Know[J]. Asahi Law Review Asahi University School of Law, 1978, 35(2): 249-257.
[26] MANTELERO A. The EU Proposal for a General Data Protection Regulation and the roots of the right to be forgotten[J]. Computer Law & Security Review, 2013, 29(3), 229-235.
[27] WALLACE K A. Anonymity[J]. Ethics and Information Technology, 1999, 1(1): 21-31.

LEI Yu-xiao, born in 1997, postgraduate. Her main research interests include privacy protection, knowledge graphs and big data.



DUAN Yu-cong, born in 1977, Ph. D, professor, Ph. D. supervisor, is a senior member of China Computer Federation. His main research interests include service computing, artificial intelligence, knowledge graph and big data.