

风险预测的制度化： 数智时代福利治理的多案例比较

朱 宁

【摘 要】 数智时代福利治理正从事后受理转向依据未来风险配置调查、服务与资源。本文通过四国案例比较，考察未然风险的制度化及预测结果向福利处置的转化。研究发现，风险经行政记录、代理标签与既往处置建构；预测输出须通过阈值、默认规则、资源配置与专业流程，转化为干预、排序或控制。四个案例分别呈现持续运行与校准、部署前取消、推广中受阻和运行后终止的轨迹，差异源于预测介入位置、预测－处置耦合、服务承接能力、专业人员的实质改变权限与外部问责时序，并受权利规范、政治争议与制度化否决渠道影响。本文认为预测性福利治理应从任务、福利、分配和程序四维检验绩效，并依处置后果、强制程度与错误可逆性配置“最小充分保障”，使被识别者获得可持续、可申诉和可纠错的支持。

【关键词】 预测性福利治理；福利对象识别；行政负担；福利权利；多案例比较

一、问题的提出

数智化进入公共服务，表层是纸质材料变为电子记录、分散事项进入整合平台、群众“多头跑”转向“一网通办”，深层是公共机构依据对未来风险的计算，在当下配置调查、服务与监管资源。在儿童保护、就业服务和福利合规等领域，公共机构开始利用跨系统数据识别儿童受伤害、持续失业和行政异常等风险。风险判断并非数智时代的新现象，新的变化在于，数智模型使风险判断成为可排序、可流转的行政结果。

这一变化与我国福利治理进程相交叠。数字政府、数字中国和“人工智能＋”等部署强调将数据融通、业务协同与公共服务创新纳入国家治理现代化布局。^① 社会救助改革推进低收入人口动态监测、主动发现和分类救助。^② 困境儿童政策强调精准识别、心理监测与早期

【作者简介】 朱宁，复旦大学社会发展与公共政策学院博士，东芬兰大学社会科学系博士、博士后，瑞典隆德大学、新加坡国立大学访问学者。主要研究方向：儿童与家庭福利、社会福利政策、数智化福利治理。

【基金项目】 芬兰社会事务与卫生部资助项目“儿童福利通报的作用和意义”。

① 国务院：《关于加强数字政府建设的指导意见》（国发〔2022〕14号），2022年6月23日；国务院：《关于深入实施“人工智能＋”行动的意见》（国发〔2025〕11号），2025年8月26日；中共中央、国务院：《数字中国建设整体布局规划》，2023年2月27日。

② 国务院办公厅：《关于加强低收入人口动态监测做好分层分类社会救助工作的意见》（国办发〔2023〕39号），2023年10月23日。

干预。^①2026年7月施行的《中华人民共和国社会救助法》将动态监测、主动发现、跨部门核对与信息保护、理由说明、听取意见和行政救济纳入同一框架。^②尽管数智风险预测尚未普及,但支撑制度在困难显现前行动的数据、组织与法律条件正在形成。

“政策找人”可弥补申请制对不熟悉政策、举证困难或因污名不愿求助者的遗漏,跨部门核对也可减轻材料负担。但风险预警增加的是行政注意,并非现金、住房、照护等资源;同一线索既可启动救助,也可引发复核调查。主动发现包括发现、核实与权利处置三个环节,数据预警主要作用于发现环节,不能替代后续的个案核实和权利处置。其有效性又受到认知偏差和资源约束的影响:前者可能使相关性线索被直接当作事实认定,绕过专业评估;后者则可能使预警止于工单,无法转化为实质行动。基于此,本文关注风险线索进入何种流程、由谁解释和对接资源,以及错误能否被当事人和专业人员纠正。

既有研究表明,技术只有嵌入权责结构和工作流程,才能产生治理效果;^③数据会与科层责任结合,形成留痕、考核与任务。^④公共服务依赖者往往留下更多记录,既可能早获帮助,也更易受监测。^⑤算法公平研究表明,删除性别、种族等敏感变量,亦不能消除地区、收入和教育经历等变量的代理作用。^⑥若仅从数据与模型界定公平,而忽视算法所嵌入的组织流程、专业裁量、服务能力与资源配置,便会落入“抽象陷阱”,将社会制度问题收缩为技术问题。^⑦

本文所称“福利治理”,是公共机构围绕生活风险,通过现金给付、社会服务、就业支持、儿童保护和资格监督配置权利、资源与责任的制度过程。它须受人民性、法治性、互助共济与促进公平等价值约束,否则易被收缩为一般风险管理和行政控制。^⑧四个案例领域虽异,但均涉及潜在对象识别、介入机会分配及个体的证明和配合义务。这一共同功能既避免将福利治理窄化为津贴发放,也解释了支持与规制何以共存。

“风险预测的制度化”,是指有关个体未来风险的计算,经由数据授权、分类规则、组织流程和资源配置,获得持续、可重复且能够影响权利与行政负担的制度效力。它并非福利制度的“制度风险”。其中,代理标签可能吸收既往偏差,影响风险知识形成;专业裁量决定预测如何转为个案处置;司法、数据保护和公共监督决定转换能否获得持续正当性。本文关注预测

① 国务院办公厅:《关于进一步加强困境儿童福利保障工作的意见》(国办发〔2025〕18号),2025年5月7日。

② 全国人民代表大会常务委员会:《中华人民共和国社会救助法》,2026年4月30日通过,2026年7月1日起施行。

③ 邱泽奇:《技术与组织的互构——以信息技术在制造企业的应用为例》,《社会学研究》2005年第2期;张燕、邱泽奇:《技术与组织关系的三个视角》,《社会学研究》2009年第2期。

④ 渠敬东等:《从总体支配到技术治理——基于中国30年改革经验的社会学分析》,《中国社会科学》2009年第6期;王雨磊:《数字下乡:农村精准扶贫中的技术治理》,《社会学研究》2016年第6期。

⑤ Paul Henman, Greg Marston, "The Social Division of Welfare Surveillance," *Journal of Social Policy*, 2008, 37(2); Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, St. Martin's Press, 2018; Philip Alston, *Report of the Special Rapporteur on Extreme Poverty and Human Rights*, A/74/493, United Nations, 2019.

⑥ Solon Barocas, Andrew D. Selbst, "Big Data's Disparate Impact," *California Law Review*, 2016, 104(3).

⑦ Andrew D. Selbst, et al., *Fairness and Abstraction in Sociotechnical Systems*, Proceedings of the Conference on Fairness, Accountability, and Transparency, 2019, 59–68.

⑧ 郑功成:《中国式现代化与社会保障新制度文明》,《社会保障评论》2023年第1期。

如何重新分配服务机会、公共资源、调查概率和证明负担。

据此，本文追问：个体或家庭未来风险的计算，如何经组织规则转化为福利处置，并产生差异化的资源与权利后果？考察内容包括：未然风险如何经由行政数据、代理标签和既往记录成为治理对象；预测结果如何转化为支持性干预、资源排序或合规控制；介入位置、处置耦合、服务能力和纠错权限，如何与权利规范、政治争议及制度化问责共同塑造制度轨迹。

本文比较美国阿勒格尼家庭筛查工具（AFST）、丹麦格拉德萨克塞市儿童风险预测项目（Gladaxe）、奥地利劳动力市场机会辅助系统（AMAS）和荷兰系统风险指示器（SyRI）。四案不代表各国全部数智福利实践，其可比性在于：公共机构均利用行政数据识别个体或家庭的未来风险，并将预测嵌入行政流程，影响服务进入、调查概率、资源机会或证明负担。比较单位为行政数据、风险标签、预测输出、组织流程、处置后果及问责机制构成的社会技术链条；政策领域和技术差异用于检验相同机制能否跨领域出现。

二、文献综述与分析框架

（一）福利对象识别、行政负担与条件化

福利对象识别并非将既有需要如实录入制度的中性过程。申请制和资格审查要求个体了解政策、表达需求并持续举证，由此产生学习、心理和合规成本，这些行政负担既可能源于资源与能力限制，也可能成为影响福利获得的隐性政策工具。^①中国社会救助研究同样表明，对象识别嵌入政策目标、基层执行和地方关系，“瞄准偏差”不能仅归因于信息不足。^②低收入人口动态监测和主动发现的制度意义，正在于将部分识别责任由个体转移至公共机构，降低未申请者 and 申请困难者的进入成本。

识别越严格并不意味着保障越精准。若保障标准未同步提高，严格审查可能增加漏保，并通过资格绑定影响其他救助权利。^③制度评价因此不能只考察错保率，还应关注漏保及其连带后果。福利分类还包含对应得性、行为责任和公共投资价值的判断。年龄、就业经历、家庭责任和既往制度接触，既用于描述需要，也可能区分谁值得支持、谁需重点核查。^④在资源约束和绩效管理下，一线组织往往将抽象政策转化为可操作的分类与处置规则。^⑤主动发现虽可减

① Donald P. Moynihan, et al., "Administrative Burden: Learning, Psychological, and Compliance Costs in Citizen-State Interactions," *Journal of Public Administration Research and Theory*, 2015, 25(1); Pamela Herd, Donald P. Moynihan, *Administrative Burden: Policymaking by Other Means*, Russell Sage Foundation, 2018.

② 李棉管：《技术难题、政治过程与文化结果——“瞄准偏差”的三种研究视角及其对中国“精准扶贫”的启示》，《社会学研究》2017年第1期；李迎生等：《福利治理、政策执行与社会政策目标定位——基于N村低保的考察》，《社会学研究》2017年第6期。

③ 关信平：《新时代中国城市最低生活保障制度优化路径：提升标准与精准识别》，《社会保障评论》2019年第1期。

④ Wim van Oorschot, "Who Should Get What, and Why? On Deservingness Criteria and the Conditionality of Solidarity among the Public," *Policy & Politics*, 2000, 28(1); Joe Soss, et al., *Disciplining the Poor: Neoliberal Paternalism and the Persistent Power of Race*, University of Chicago Press, 2011.

⑤ Evelyn Z. Brodtkin, "Policy Work: Street-Level Organizations under New Managerialism," *Journal of Public Administration Research and Theory*, 2011, 21(Suppl. 2).

少申请负担,也可能把监测、调查和反复自证提前至个人请求之前,使识别责任上移与处置负担前移同时发生。

这种转换具有明显的制度不对称性。被识别者通常无法退出基础公共制度,分类结果又可能影响基本生活资源、家庭关系和行政审查,错误难以通过更换平台消解。国家掌握的数据和解释能力远超个体,而低收入群体更依赖公共服务,也更难承担持续申诉成本。因此,数智福利治理的合法性不仅取决于效率增益,更取决于数据处理是否符合必要性、比例性和程序正义,以及能否提供充分的解释、救济与纠错。行政机关既不能以服务便利扩张数据权限,也不能以整体效率压缩个体权利。

(二) 技术嵌入、风险分类与差异化可见性

社会学并不将分类视为对客观差异的简单记录。量化、评分和排序在建立可比性、分配注意与机会的同时,也会反向塑造组织和行动者。^① 家外安置、再次转介、就业天数和行政异常虽可被记录,却分别受家庭处境、市场机会、服务供给和行政实践影响。模型可能准确预测机构的下一步行动,却被误读为识别了个体的固有风险。

风险模型的首要挑战因此是目标转译:复杂生活需求必须被压缩为可观察、可学习的标签。家外安置不等于虐待事实,就业天数不等于就业能力,行政记录不一致也不等于福利欺诈。标签越依赖机构自身生产的事件,模型越容易将既往组织行为写入个体未来;贫困、种族化制度接触和既往干预也可能通过代理变量进入风险画像。^②

国内研究已指出,数字技术在提高社会救助精准度和行政效率的同时,也可能造成资源与需求错配、信息共享不足、基层数字负担和群体数字不平等;^③ 自动化决策还可能在数据集构建、算法运行、透明度和可解释性方面产生偏差。^④ 既有研究主要关注技术赋能、算法偏差及其治理。本文则进一步考察风险标签如何由行政痕迹和既往处置生产,并经由组织流程转化为调查、服务、排序或排除。

这种可见性的分布并不均衡。低收入家庭频繁进入救助、医疗、住房和司法系统,留下密集行政痕迹;拥有私人资源者较少进入公共系统,数据画像反而稀疏。数智化由此同时生产“需要可见性”与“合规可见性”:前者使未满足需求进入服务视野,后者则将对象呈现为待核查、待证明或待矫正者。“照看社会”中照护与观察的共生关系,^⑤ 以及对象识别对政策目标、基层执行与地方关系的嵌入,^⑥ 均说明数智化并未消解制度生成过程,而是重塑了信息被记录、

① Wendy Nelson Espeland, Mitchell L. Stevens, "Commensuration as a Social Process," *Annual Review of Sociology*, 1998, 24; Marion Fourcade, Kieran Healy, "Classification Situations: Life-Chances in the Neoliberal Era," *Accounting, Organizations and Society*, 2013, 38(8).

② Emily Keddell, "The Ethics of Predictive Risk Modelling in the Aotearoa/New Zealand Child Welfare Context: Child Abuse Prevention or Neo-Liberal Tool?" *Critical Social Policy*, 2015, 35(1); Joanna Redden, et al., "Datafied Child Welfare Services: Unpacking Politics, Economics and Power," *Policy Studies*, 2020, 41(5).

③ 李卓:《社会救助数字化转型中的技术进步与价值重塑》,《社会保障评论》2024年第6期。

④ 匡亚林:《社会救助自动化决策的运行机理、算法偏差及治理路径》,《社会保障评论》2025年第2期。

⑤ 彭亚平:《照看社会:技术治理的思想素描》,《社会学研究》2020年第6期。

⑥ 李棉管:《技术难题、政治过程与文化结果——“瞄准偏差”的三种研究视角及其对中国“精准扶贫”的启示》,《社会学研究》2017年第1期;李迎生等:《福利治理、政策执行与社会政策目标定位——基于N村低保的考察》,《社会学研究》2017年第6期。

连接和信任的方式。

（三）从技术嵌入到福利后果：风险预测的制度化与“预测－处置转换”

风险预测的制度化包括三个递进层次：知识制度化，即以行政痕迹、代理标签和统计关联将复杂处境转化为可比较的风险对象；组织制度化，即使分数、档位和名单成为阈值设定、默认规则、资源门槛与专业流程的依据；后果制度化，即令预测持续影响服务入口、调查概率、资源机会与证明负担。但制度化并不意味着预测工具必然持续运行。Gladsaxe 部署前受阻、AMAS 推广中遇诉讼、SyRI 运行后被司法终止，表明数据授权、外部审查和退出机制也是制度化过程的一部分。

福利治理“前置化”，是公共机构依据尚未发生、未充分显现或未经个案核实的风险，提前配置注意、服务、调查与责任。它不同于以损害是否减少为标准的预防，首先指向治理时间与组织入口的变化，其结果既可能是支持，也可能是控制。“预测－处置转换”解释前置化如何实现：行政痕迹决定谁被纳入计算；风险客体化决定何种未来成为政策问题；预测输出将历史关联压缩为分数、档位或名单；组织处置借助阈值、默认规则、资源与专业流程赋予其行政效力；福利后果体现为服务进入、资源获得、调查概率、证明负担与救济机会。处置产生的新记录又反馈至下一轮识别，形成递归回路（见图 1）。

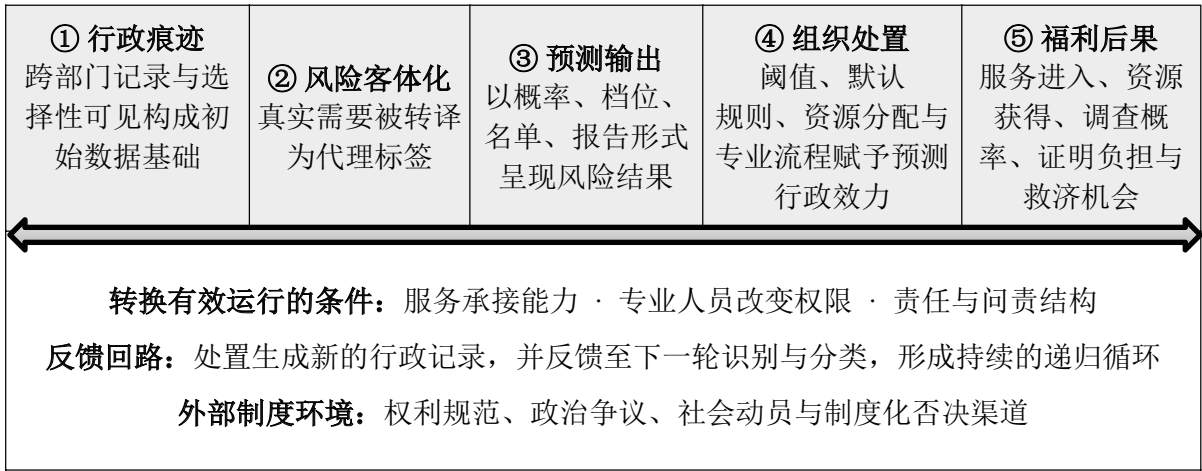


图 1 福利治理风险前置化中的“预测－处置转换”逻辑框架

资料来源：作者绘制。

本文从三个维度比较制度化程度：处置方向，即输出连接支持性干预还是限制性处置；耦合强度，即输出仅作为信息提示，还是事实上构成具有约束力的指令；实质纠错能力，即工作人员与服务对象能否在处置前补充情境、质疑数据并调整结果。街头官僚研究将一线裁量视为政策运行的核心，^① 信息系统却可能将部分规则上移至系统设计层。^② 保留人工最终决定权因而并不天然构成保障：若控制权内置于变量、阈值和界面，而责任外置于最终签字者，便会形成“道

① Michael Lipsky, *Street-Level Bureaucracy: Dilemmas of the Individual in Public Services*, Russell Sage Foundation, 1980.

② Mark Bovens, Stavros Zouridis, "From Street-Level to System-Level Bureaucracies: How Information and Communication Technology Is Transforming Administrative Discretion and Constitutional Control," *Public Administration Review*, 2002, 62(2).

德缓冲区”。^①关键在于人工环节是否具有信息、时间、资源与权限，以及个案错误能否回溯至模型与制度层面。

数字技术并非制度之外的中性工具，而是与官僚组织、权责结构和合法性机制相互塑造。本文将这一互动置于社会政策的风险共担、资源再分配和社会权利框架中，关注福利制度如何识别风险，如何在个人、家庭、市场与国家之间分配责任，以及依据需求补偿、预期回报还是合规程度配置资源。

风险预测由此带来三方面相互关联的变化。对象识别由申请后的资格判断前移至申请前的风险筛查；风险标签由描述需要转化为排序调查、服务和公共投资的依据；申请制中的学习和证明成本进一步转化为未申请者承担的监测、解释和自证负担。“预测—处置转换”所要检验的，正是新增识别能力究竟形成福利支持，还是更早的分类、调查和排除，并由此把技术评价重新置于补偿责任、分配正义和权利保障之中。

上述机制还受外部制度环境制约。本文将笼统的“国民文化”转化为三个可观察条件：隐私、平等、家庭自主和社会权利等规范如何划定技术应用边界；福利供给、公共支出和欺诈治理形成何种政治争议；服务对象、专业组织、数据保护机构、社会团体和法院能否经特定渠道介入。外部环境通过目标授权、争议动员和制度化否决发挥作用，决定何种问题成为公共争议、哪些主体能够提出异议，以及异议在部署前、推广中还是运行后进入决策。

三、研究设计

本文采用结构化、聚焦式多案例比较，并对关键制度节点进行案内时序重建。比较单位是风险预测从数据生成到行政处置的社会技术链条。各案例围绕相同问题展开：系统预测或提前识别何种风险，哪些行政记录被用作实际需要的代理标签，预测输出进入何种行政环节并如何连接调查、服务或资源配置，工作人员和服务对象能否改变分类，以及外部问责如何影响制度轨迹。

案例选择采取理论取样，须同时满足四项条件：第一，由公共机构或承担公共职能的组织开发或使用；第二，利用行政数据，对个体或家庭的福利风险、服务需要或合规风险进行预测性识别；第三，预测结果拟进入或已进入具体行政流程，并可能影响调查、服务、资源配置或证明负担；第四，具有足够的官方文件、技术报告、独立评估、裁判文书或同行评议研究，可重建从预测生成至制度处置的基本过程。

本文依据三个维度进行最大差异取样：处置方向涵盖主动支持、调查筛入、投资排序和合规审查；预测与处置的耦合程度涵盖拟议提示、默认建议、资源分类和风险报告；制度轨迹涵盖持续运行与校准、部署前取消、推广受阻和运行后司法终止。四个案例分别覆盖持续运行与

^① Madeleine Clare Elish, "Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction," *Engaging Science, Technology, and Society*, 2019, 5; Ben Green, "The Flaws of Policies Requiring Human Oversight of Government Algorithms," *Computer Law & Security Review*, 2022, 45.

校准、部署前取消、推广受阻后经司法复核和运行后司法终止四种制度轨迹。

本文以“制度事件时间”统一观察边界：以预测工具进入正式政策设计、行政试验或实际运行为起点，以持续校准、推广受阻、部署前取消或司法终止等当前可观察节点为终点，见表1。核心比较证据集中于2015—2025年。AMAS和SyRI在2000年代的制度沿革仅作为背景，不与2016年以后的数智预测项目并列比较。不同技术代际被视为分析变量，用于检验行政效力究竟来自技术复杂程度，还是预测输出与组织流程的结合。若不同领域和技术类型的系统均须通过风险标签、阈值规则、资源配置、专业流程和问责结构获得行政效力，跨领域差异反而有助于识别更具一般性的组织机制。

研究材料主要包括官方技术报告、项目文件、影响评估、裁判文书和同行评议研究。一手文件用于确认系统目标、变量、流程位置与法律状态，独立研究用于重建组织实践；另以议会讨论、政策咨询、数据保护文件、专业组织意见和诉讼材料识别政治议题、权利框架与问责渠道。鉴于证据不对称，本文不比较预测精度、不推断未公开状态，也不作跨国绩效排名，仅判断能够得到多源材料印证的制度转换。

表 1 四个案例的技术边界与制度轨迹

案例	核心观察阶段	领域与技术	输出进入的行政节点	制度轨迹	理论取样功能
美国 AFST	2016 年部署至 2023 年评估	儿童保护；机器学习评分	热线筛入决定	持续运行与校准	观察预测与调查入口的强耦合及人工纠错
丹麦 Gladsaxe	项目提出至部署前取消	儿童保护；拟议预测模型	举报前主动筛查	部署前取消	观察数据授权和正当性审查如何前置阻断
奥地利 AMAS	测试、推广受阻至 2025 年后续司法审查	就业服务；逻辑回归画像	咨询分类与培训资源配置	推广受阻	观察风险分类如何转化为公共投资排序
荷兰 SyRI	2015—2020 年	福利合规；规则型数据匹配	风险报告与后续调查	运行后司法终止	观察简单技术如何获得较强调查效力

资料来源：作者整理。

分析分为两步：首先按照政策目标、数据来源、代理标签、输出形式、流程位置、处置后果及纠错问责整理案内证据，区分制度文件规定的正式用途与研究材料揭示的实际运作；其次跨案例配对比较，考察相似技术为何产生不同处置、不同技术为何产生相似约束，以及问责介入如何在不同阶段改变项目轨迹。由此为中国主动发现机制提炼可检验的制度条件。

四、风险预测的四种制度轨迹

聚焦预测工具进入政策设计或行政实践后的关键阶段，依次考察风险标签如何形成、预测输出如何进入行政流程并获得效力、服务对象承担何种资源与权利后果，以及专业人员、服务对象和外部问责能否改变分类及制度轨迹。

（一）AFST：预测安置为何首先触发调查

美国阿勒格尼县于2016年部署家庭筛查工具（Allegheny Family Screening Tool，AFST），

2018年启用第二版,用于协助儿童保护热线决定是否将一般保护服务转介筛入调查程序。AFST预测热线转介所涉及儿童在未来两年内发生由法院裁定家外安置的可能性,并据此生成风险分数。^①家外安置虽与严重风险相关,却也受既往举报、调查方式、服务供给、家庭资源和法院判断影响。因此,模型学习的不只是家庭风险,也包括儿童保护系统以往识别、处置家庭的记录。

AFST输出1至20分的风险排序,并嵌入热线筛入决策。特定高分案件默认筛入,主管人员如决定筛出须书面说明;特定低分案件通常建议筛出,其余案件保留专业裁量。未来安置概率由此通过默认规则和说明义务获得流程效力。系统不直接决定安置,而是改变家庭进入强制调查的概率。筛入既可能使需要得到评估并连接服务,也可能带来入户调查、隐私暴露、家庭压力及解释和配合义务。调查记录还会用于后续风险判断。AFST的后果除预测准确率或安置率外,还应考察新增调查能否获得服务资源承接及哪些家庭承担更多程序负担。

AFST保留工作人员和主管人员的最终决定权,但实质纠错取决于其能否获得模型之外的情境信息并推翻分类。研究发现,工作人员结合举报叙述和家庭近期变化纠正不当评分,资深人员与模型建议的一致度低于缺乏算法经验的新手;^②人机共同决策还缩小了纯模型可能产生的部分种族差异。^③影响评估显示,高风险儿童进入调查的频率提高,部分短期指标发生变化,但同期政策调整使其难以完全归因于算法。^④AFST因而呈现“运行中制度化”轨迹,即预测通过默认筛入获得行政效力,又在专业复核、影响评估和持续校准中受到修正。

(二) Gladsaxe: 举报之前的预防何以被阻断

丹麦格拉德萨克塞市儿童风险预测项目将识别节点推进至正式举报或求助之前。地方政府拟整合就业、健康、教育和既往服务记录,以离家安置等制度事件为代理标签,识别未进入儿童保护系统的潜在脆弱儿童,通过主动联系提供自愿帮助。^⑤不同于处理既有热线转介的AFST,Gladsaxe拟在人口层面对尚未出现举报、申请或正式求助的家庭开展人口筛查。原用于其他公共服务的数据将被改用于判断儿童及家庭风险;即使不直接产生强制决定,也会改变公共机构观察和接触家庭的入口。

举报前识别可使困难家庭较早获得支持,减少政策不知晓、污名或求助能力不足造成的遗漏,也可能将风险尚未显现的家庭纳入监测,使父母承担解释情况、回应联系或质疑标签的压力。

① Rhema Vaithianathan, et al., *Developing Predictive Models to Support Child Maltreatment Hotline Screening Decisions: Allegheny County Methodology and Implementation*, Allegheny County Department of Human Services, 2017; Rhema Vaithianathan, et al., *Allegheny Family Screening Tool: Methodology, Version 2*, Allegheny County Department of Human Services, 2019.

② Maria De-Arteaga, et al., *A Case for Humans-in-the-Loop: Decisions in the Presence of Erroneous Algorithmic Scores*, Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 2020, 1-12; Lingwei Cheng, Alexandra Chouldechova, *Heterogeneity in Algorithm-Assisted Decision-Making: A Case Study in Child Abuse Hotline Screening*, Proceedings of the ACM on Human-Computer Interaction, 2022, 1-33.

③ Hao-Fei Cheng, et al., *How Child Welfare Workers Reduce Racial Disparities in Algorithmic Decisions*, CHI Conference on Human Factors in Computing Systems, 2022, 1-22.

④ Jeremy D. Goldhaber-Fiebert, Lea Prince, *Impact Evaluation of a Predictive Risk Modeling Tool for Allegheny County*, Phase 2, Allegheny County Department of Human Services, 2023.

⑤ Helene Friis Ratner, Kasper Trolle Elmholdt, "Algorithmic Constructions of Risk: Anticipating Uncertain Futures in Child Protection Services," *Big Data & Society*, 2023, 10(2).

主动联系能否保持自愿、拒绝接触是否会被解释为风险信号，是支持性预防与非自愿监测的界线。项目未实际运行，现有证据不能证明其曾对真实儿童评分，也不足以判断一线使用及治理绩效。

项目未获数据保护豁免，围绕数据用途改变、数据最小化、错误标签和儿童最大利益的争议使其在正式部署前取消。预防目标由此不能自动为跨部门数据使用提供正当性，问责无须等到系统运行后才发生；数据授权、目标设定和筛查范围即可接受审查。Gladaxe 呈现“部署前受阻”的制度轨迹，即预测尚未获得行政效力，数据保护和公共正当性审查已改变项目方向。

（三）AMAS：就业机会如何转化为投资价值

奥地利公共就业服务局的劳动力市场机会辅助系统（Arbeitsmarktchancen-Assistenzsystem, AMAS），以多元逻辑回归估计求职者未来 7 个月或 24 个月内达到规定无补贴就业天数的概率，并将其划分为高、中、低机会组。变量包括人口特征、健康与照护责任、就业经历和地区劳动力市场状况。^① 就业天数不同于抽象的就业能力，还受雇主选择、地区机会、经济环境和既往政策投入影响。这些结构性条件被压缩为个人就业概率，可能将女性、移民和残障者面临的制度性不利表述为个体“低机会”属性。

AMAS 输出进入就业咨询和培训资源配置。高机会组被认为可较快自行就业，低机会组被认为常规培训回报有限，积极就业项目主要投向中等机会组。分类由就业前景描述转化为公共机构“应当为谁投资”的判断，并未遵循“就业机会越低、支持越多”的补偿逻辑。低机会者可能因预期回报较低而较少获得培训；若资源减少又降低其就业可能，预测便通过行政处置参与塑造自身所预测的结果。“低机会”标签因而既描述就业前景，也影响实际支持。AMAS 的制度后果不能只以群体预测误差衡量，还应考察资源排序是否削弱福利制度对结构性不利的补偿责任。

就业顾问仍作最终决定，但人工终审不意味着实质纠错。诉讼争议之一，是顾问能否在有限时间、信息和组织压力下独立评估并改变分类。2023 年，奥地利最高行政法院撤销原判，要求查明 AMAS 评分是否实质决定求职者分组与咨询结果，未最终认定其违法；2025 年，联邦行政法院认定顾问可实质复核并改变评分，相关处理不属于完全自动化决定，遂撤销此前禁令。^② AMAS 由此呈现“推广中受阻后经司法复核”的制度轨迹，即预测已与资源配置强耦合，而工作人员能否真正改变结果，成为系统继续推广的关键问责问题。

（四）SyRI：简单技术何以形成较强调查效力

荷兰系统风险指示器（Systeem Risico Indicatie, SyRI）由跨部门数据匹配、风险指标、赋分和阈值构成，用于识别福利、税务和劳动法规的不合规风险。地方机构确定项目和风险模型，数据经匹配、复核后形成风险报告，再由法定机关决定是否调查。^③ SyRI 识别行政记

① Doris Allhutter, et al., "Algorithmic Profiling of Job Seekers in Austria: How Austerity Politics Are Made Effective," *Frontiers in Big Data*, 2020, 3.

② Verwaltungsgerichtshof, *Erkenntnis vom 21. Dezember 2023*, Ro 2021/04/0010-11; Bundesverwaltungsgericht, *Erkenntnis vom 1. September 2025*, W256 2235360-1/36E.

③ Maranke Wieringa, "'Hey SyRI, Tell Me about Algorithmic Accountability': Lessons from a Landmark Case," *Data & Policy*, 2023, 5.

录异常,但异常可能源于数据错误、更新滞后、复杂家庭处境或部门口径差异,不能直接等同于欺诈。

2015 至 2019 年间, SyRI 主要在少数城市项目中运行,部分集中于贫困和福利领取比例较高的社区。风险报告虽不直接改变待遇或法律身份,却会影响优先调查对象,并通过“风险报告—人工复核—后续调查”获得行政效力。人工决定不能消除上游指标、数据范围和阈值不透明的问题,因为复核发生在系统筛定重点对象之后。个人一旦被纳入风险报告,便可能承担材料提交、情况说明、入户核查和持续配合等义务。贫困地区居民在福利、住房、税务和就业系统中留下更多记录,也更易进入数据匹配和重点调查范围。SyRI 表明,即使不使用机器学习,只要输出进入调查入口,仍会重新分配调查概率、证明责任和隐私成本。

2020 年海牙地区法院依据《欧洲人权公约》第 8 条审查 SyRI 能否在反欺诈公共利益与私人生活权之间形成“公平平衡”。SyRI 可处理广泛的个人数据,但指标、赋分和阈值缺乏透明度,个人通常不知道自己是否进入风险报告,也难以在调查前理解和质疑分类依据。系统又主要用于贫困和福利领取比例较高的地区,增加了特定群体受到集中监测和差别对待的风险。^① 法院认为,现有框架的透明度、可预见性和程序保障不足以证明这种广泛数据干预具有相称性。

该案划定了风险分析进入行政调查的合法性边界。合法的反欺诈目标和下游人工复核本身,仍不足以证明这一制度安排具有正当性;风险报告一旦影响调查对象选择,法律框架应限定数据用途和处理范围,说明风险规则取得行政效力的方式,并提供告知、数据纠正、个案质疑和独立审查渠道。法院未以技术复杂程度为判断标准。SyRI 虽采用数据匹配和规则赋分,却因与调查流程紧密连接而产生较强权利影响。司法问责由此成为限制“预测—处置转换”、决定系统能否继续制度化的重要外部机制。

综上,四案呈现持续运行与校准、部署前取消、推广中受阻后经司法复核和运行后终止的制度轨迹。

五、相似预测逻辑何以走向不同制度轨迹：跨案例机制比较

四案分属儿童保护、就业服务和福利合规领域,技术形态涵盖规则型数据匹配、统计分类和机器学习评分,但处理的是同一制度问题:公共机构如何利用行政数据判断尚未发生的风险,并将预测输出转化为主动接触、调查启动、资源配置或证明负担。四案并非均以失败告终:AFST 仍在运行和校准,AMAS 在推广中受到质疑和限制后经司法复核,Gladsaxe 与 SyRI 则分别在部署前和运行后终止。本文旨在比较相似预测逻辑在不同组织机制和外部制度环境中形成的差异化轨迹,分析高风险标签为何分别通向更多支持、资源排序或合规负担(见表 2)。

^① Rechtbank Den Haag, NJCM c.s. tegen de Staat der Nederlanden (SyRI), ECLI:NL:RBDHA:2020:865, 5 February 2020; Marvin van Bekkum, Frederik Zuiderveen Borgesius, "Digital Welfare Fraud Detection and the Dutch SyRI Judgment," *European Journal of Social Security*, 2021, 23(4).

表 2 风险预测制度化的跨案例机制与轨迹分化

案例	风险标签的制度内生性	预测－处置耦合方式	主要资源或权利后果	纠错、问责与制度轨迹
AFST	家外安置是代理标签，同时受家庭处境、既往调查、服务供给和法院判断影响	较强：高分通过默认筛入和书面说明义务进入热线决策	提高调查概率；可能连接服务，也增加隐私与程序负担	工作人员和主管人员可修正；部署后持续评估，形成“持续运行与校准”轨迹
Gladsaxe	既往安置和服务接触被用于代理儿童脆弱性，包含机构识别与介入痕迹	尚未形成：拟在举报前筛查人口并主动联系家庭	拟提供自愿支持；未实际部署，真实后果不可观察	数据授权和公共正当性审查在部署前介入，形成“部署前取消”轨迹
AMAS	无补贴就业天数同时受个人处境、劳动力市场机会和既往政策投入影响	较强：机会分组直接连接就业咨询和培训资源配置	形成投资排序；低机会者可能获得较少培训和能力建设资源	顾问形式上终审，但实质改变权限受到司法质疑，形成“推广中受阻后经司法复核”轨迹
SyRI	行政异常由记录规则、数据质量、跨部门口径和核查重点共同生成	较强：风险报告经人工复核进入合规调查	增加材料提交、情况说明、核查配合、隐私和证明负担	因数据范围、透明度、可预见性和救济保障不足，形成“运行后终止”轨迹

资料来源：作者整理。

（一）共同起点：制度历史如何成为个体风险

四个案例均面临目标转译问题。公共机构必须将儿童安全、就业困难或福利违规等复杂问题压缩为可观察、可计算的标签，预测才能运行。AFST 和 Gladsaxe 以家外安置或既往儿童保护记录表征儿童未来风险，AMAS 以无补贴就业天数代表劳动力市场机会，SyRI 则根据行政记录的异常和不一致识别潜在违规。这些标签虽与治理问题相关，却不能完整等同于儿童伤害、个体就业能力或福利欺诈事实。

更重要的是，标签本身受到制度运行的塑造。AFST 与 Gladsaxe 表明，家外安置和服务接触除受家庭状况影响外，还取决于举报、调查、服务供给和法院裁判。AMAS 的就业结果受地区机会和政策投入影响，SyRI 识别的行政异常则受记录规则、数据质量和核查重点影响。当这些制度参与形成的结果被解释为个人固有属性时，过去的组织行为便可能被转化为未来的个体风险。

这种制度内生性还会通过反馈回路强化。高风险家庭接受更多调查，低机会求职者获得较少培训，重点监管地区留下更多核查记录；新记录又可能进入下一轮分类，使制度参与形成的差异被重新表述为对象之间自然存在的风险差异。因此，跨领域比较所揭示的是“制度记录－代理标签－个体风险”的共同转换机制。

由此形成第一个机制命题：风险标签越依赖既往行政处置，预测越可能延续已有的制度关注差异，将制度历史转化为个体风险属性，并通过后续处置强化可见性差异。模型能够预测某种机构事件再次发生，却不能据此证明个体具有相应的固有风险；预测结果反映的是个人处境与制度历史的共同作用。

（二）处置分化：流程位置、耦合方式与服务承接

风险标签只有进入具体行政节点，并与阈值、默认规则、资源门槛或专业流程相连接，才

会取得现实效力。AFST 与 Gladsaxe 同属儿童保护领域，却因介入位置不同走向不同轨迹。AFST 位于举报之后，处理已有人员表达担忧的热线转介，并通过高分默认筛入影响调查启动；Gladsaxe 则拟在没有举报、申请或正式求助的情况下筛查人口，位于既有服务入口之前。后者虽以早期支持为目标，却改变了原有数据用途和机构接触家庭的正当性基础，因而在形成稳定的预测—处置耦合前即被阻断。

耦合强度还取决于预测连接何种处置。AFST 与 AMAS 均形成较强耦合，但方向不同：AFST 高分主要提高家庭进入调查程序的概率，既可能连接服务，也可能增加隐私和程序负担；AMAS 则把机会分组转化为公共投资排序，使低机会者可能获得更少培训。SyRI 技术并不复杂，但风险报告进入调查入口后，仍会增加特定对象提交材料、解释异常和配合核查的义务。由此可见，技术复杂程度不能替代对处置方向和权利后果的分析。

服务承接能力进一步决定前置识别能否转化为预防性福利。预测系统可以扩张机构发现对象的能力，却不能自动增加社会工作者、培训项目、家庭支持或现金给付。新增识别缺乏资源承接时，组织容易将供给不足转化为对象排序：风险较高者被优先调查，预期回报较低者被减少投入，不能及时回应机构要求者还可能被解释为不合作。此时，数智化提升的是筛选和留痕能力，而非实际解决风险的能力。

因此，评价预测性福利治理需要考察预测首先改变服务入口还是调查入口，高风险能否带来可选择、可持续的资源，对象拒绝接触或无法配合是否会承受不利后果。只有预测与支持性资源相连接，且专业人员能够根据具体处境调整处置，前置化才更可能构成福利支持。

由此提出第二个机制命题：预测介入越靠前、与行政处置耦合越强而服务承接能力越弱，支持性识别越容易转化为对象排序和合规控制。制度结果不仅取决于预测进入哪个流程节点，还取决于组织如何使用预测以及预测最终连接何种资源。

（三）轨迹分化：实质改变权限与问责介入时序

四种制度轨迹的进一步分化，取决于谁能改变分类结果，以及外部问责何时介入。形式上保留人工最终决定并不等于有效纠错。实质改变权限至少需要四项条件：专业人员能够获得模型之外的情境信息，拥有重新判断所需的时间，能够调动与判断相匹配的服务资源，并可在不承受过高组织成本的情况下推翻系统建议。缺少其中任何一项，“人在回路中”都可能只是将系统后果的责任转移给一线人员。

AFST 与 AMAS 的对照说明了这一点。AFST 工作人员和主管人员能够结合举报叙述及家庭近期变化修正评分，系统运行后还通过影响评估和版本校准回应争议，因而形成持续运行与调整的轨迹。AMAS 也规定由就业顾问作出最终决定，但顾问能否在有限时间、信息条件和组织压力下独立改变分类，成为司法审查的焦点。两案均保留人工环节，却因人工权限的实际条件不同，呈现持续校准与推广受阻的差异。

问责时序同样影响制度轨迹。Gladsaxe 的争议发生在部署前，审查集中于数据用途、人口筛查范围和项目正当性，因而在个案评分产生前阻断制度化。SyRI 的问责发生在有限运行后，个人通常不知道自己是否被纳入风险报告，上游指标和阈值也缺乏透明度，最终只能借助信息

公开、社会动员和司法审查纠正制度框架。两案表明，问责越晚，纠正既成流程和累积后果的成本越高；但部署前审查也不能替代运行中的群体影响监测和个案救济。

完整的问责结构应覆盖制度化全过程：部署前审查预测目标、数据授权和用途边界，运行中监测人工修正、资源流向和群体差异，部署后提供告知、解释、申诉、暂停和退出机制。问责对象不仅包括单次误判，还应涵盖代理标签、阈值、默认规则和目标群体划定等上游选择。只有个案纠错能够反向推动指标和组织规则调整，人工复核才具有制度意义。

由此提出第三个机制命题：专业人员和 service 对象的实质改变权限越充分，外部问责越能在数据授权、流程耦合和群体后果形成的关键节点及时介入，风险预测越可能通过持续校准维持制度正当性；反之，若人工终审流于形式、上游规则不可见且救济滞后，系统越可能在推广或运行阶段遭遇阻断。

（四）外部制度环境：权利规范和否决渠道如何影响制度轨迹

外部制度环境通过问题界定和否决渠道调节制度轨迹。AFST 中的儿童安全、家庭自主和种族平等争议主要被吸收到持续评估与系统校准中；Gladsaxe 因将数据使用前移至举报之前，使隐私、知情和儿童最大利益成为部署前的授权条件；AMAS 的司法审查聚焦就业效率、平等对待及顾问能否实质改变分类；SyRI 则由社会团体将特定社区的监测负担转化为隐私、平等和比例性问题，并通过司法审查终止系统。

因此，权利规范和政治争议并不取代流程、资源和纠错机制，而是决定何种后果能够成为公共问题、谁可以提出挑战以及问责何时介入。问责越早，越可能在数据用途和目标形成阶段改变项目；问责越晚，越需要处理既成惯例和累积后果。系统最终终止还是校准，还取决于行政机构能否调整流程、补充资源并回应争议。

六、从域外机制到中国条件：预测性福利治理的适用边界

中国福利对象识别长期依赖个人申请、基层摸排、家庭经济状况调查和资格审核。精准扶贫时期建档立卡和数据比对，以及当前低收入人口动态监测、跨部门信息核对和基层主动发现，使另一种识别路径逐渐形成：个人即使尚未申请，也可能因收入下降、就业中断、医疗支出增加或家庭结构变化进入平台预警范围。“政策找人”由此将部分识别责任转移至公共机构，有助于减少政策不知晓、求助污名和证明能力不足造成的福利遗漏。^① 这一识别机制与域外预测性福利治理的共同点，在于利用行政记录提前识别尚未通过正式求助充分显现的生活风险。

但二者的制度结构并不相同。中国现阶段主要采用指标规则、信息比对和异常预警，尚未普遍形成以个体概率评分直接决定福利处置的模式；平台建设和指标设定具有较强的纵向统筹能力，而入户核实、需求评估和服务链接仍由街道、乡镇、社区及基层工作人员承担。属地管理、

^① 国务院办公厅：《关于加强低收入人口动态监测做好分层分类社会救助工作的意见》（国办发〔2023〕39号），2023年10月23日；Donald P. Moynihan, et al., "Administrative Burden: Learning, Psychological, and Compliance Costs in Citizen-State Interactions," *Journal of Public Administration Research and Theory*, 2015, 25(1); Pamela Herd, Donald P. Moynihan, *Administrative Burden: Policymaking by Other Means*, Russell Sage Foundation, 2018.

网格化发现、限时办结和工单闭环虽能提高响应速度，也可能促使基层优先完成容易留痕和考核的核查任务，而难以投入周期更长、成效不易量化的服务协调。^①此外，中国动态监测以主动发现和兜底保障为规范目标，其数据权限、行政复核和权利救济结构也不同于域外的儿童保护、就业分流和反欺诈系统。

域外案例表明，权利规范、政治争议和制度化否决渠道会影响数据授权、预测进入行政流程的条件和问责时序。其对中国的意义不在于移植具体制度，而在于检验行政数据如何形成标签、预测如何进入处置、识别能否获得服务承接、专业人员能否改变分类，以及对象能否提出异议。具体而言，应考察预警是否提高调查概率、风险分组是否影响资源机会、求助前的数据利用是否具有充分授权和比例性，以及规则预警进入调查后是否增加特定群体的证明负担（见表 3）。

表 3 域外机制与中国预测性福利治理的对应关系

域外机制	中国对应实践	中国特殊条件	需要检验的问题
行政记录转化为代理标签	低收入人口动态监测、跨部门信息核对	以规则预警和异常识别为主	预警是否被直接视为困难、失责或违规事实
预测输出进入组织处置	平台派单、网格发现、基层核实	纵向任务、属地责任和限时办结相结合	预警首先触发服务、联系还是核查
识别能力与服务能力分离	主动发现、需求转介和分类救助	地方服务供给与资源调动能力不均	风险被发现后是否有现实资源承接
人工复核可能形式化	基层入户调查和专业评估	了解家庭情境者未必掌握规则修改权和资源	基层能否实质改变分类及处置
外部问责改变制度轨迹	数据审查、行政监督、复议诉讼和社会监督	问责渠道不同于域外司法和社会参与结构	谁能在何时质疑数据用途、风险规则和处置结果
人口性影响缺乏正式救济	风险排序、材料复核和反复核查	正式救济多集中于最终行政决定	中间处置能否被记录、说明和纠正

资料来源：作者整理。

基于上述适用边界，中国预测性福利治理需要回答四个问题：如何根据处置后果配置程序保障，如何评价预测工具的实际绩效，如何保障基层专业人员实质改变分类的权限，以及如何将个案纠错转化为制度反馈。

（一）按照处置后果配置分层、相称的程序保障

风险预测的制度环节并非越多越好，而应根据输出用途、处置后果、错误可逆性和行政强制程度配置程序。对于政策提醒、申领信息推送和自愿服务转介等低影响用途，若数据可靠、对象可以拒绝或退出、错误能够及时撤回且服务资源充足，简单的“识别－提醒－服务”流程即可发挥作用。制度重点是保障自主选择，并防止拒绝服务被重新解释为风险信号；过多审批反而可能增加成本、延误服务。

当预测进入培训机会、救助顺序、服务强度或资源投入的排序环节时，保障程度应当提高。此类场景需要人工复核、理由记录和群体差异监测，并允许服务对象补充材料、说明具体处境，防止历史资源不足或结构性不利被转化为减少公共投入的理由。

① 渠敬东等：《从总体支配到技术治理——基于中国 30 年改革经验的社会学分析》，《中国社会科学》2009 年第 6 期；王雨磊：《数字下乡：农村精准扶贫中的技术治理》，《社会学研究》2016 年第 6 期。

当预测用于启动调查、暂停待遇、追缴资金或增加证明责任时，应设置更严格的数据核验、人工决定、理由告知、陈述申辩和复核救济程序。风险相关性只能说明需要进一步了解，不能直接证明资格不实、家庭存在危险或对象拒绝合作。用于主动联系、需求评估和服务转介的数据线索，不应直接成为拒绝、停发或追缴待遇的依据，也不能以形式上的“人工终审”掩盖工作人员无法改变系统建议的事实。^①

因此，应根据处置风险配置“最小充分保障”：预测对权益影响越大、错误越难逆转、行政措施强制性越强，保障程度就应越高。适度的核验、记录和纠错不仅保护权利，也能减少误判、重复调查、资源错配和信任损耗，是治理绩效持续改善的必要条件。

（二）以四维绩效检验预测—处置转换

预测性福利治理的绩效至少包括四个维度：任务绩效，即发现对象、完成核验和降低行政成本的效率；福利绩效，即识别能否转化为服务获得、资源改善和风险缓解；分配绩效，即不同群体承担的误判、调查和证明负担是否均衡；程序绩效，即服务对象能否获知分类及处置理由、提出异议并纠正错误。识别数量、核查速度和结案效率的提高，并不等于总体绩效改善。政策提醒、自愿服务和低影响转介可能同时改善任务绩效和福利绩效，而不明显损害其他维度。但预测一旦用于启动调查、降低资源优先级、暂停待遇或增加证明责任，效率提升可能伴随误判集中、服务不足和救济困难。此时，任务绩效不能抵消福利、分配和程序绩效的损失。

这一框架对中国尤为重要。数字政府具有较强的平台统筹和纵向整合能力，入户核实、个案服务和资源链接却主要由地方及基层承担。如果预警增长快于人员、服务名额和资源调动能力，风险便可能转化为基层必须消化的责任工单。在资源不足时，平台可能提高信息处理和任务结案效率，却未必改善困难群众获得的支持，甚至将技术效率转化为基层负担和个人证明责任。

因此，除风险发现数量、模型准确率、核验速度和按时结案率外，还应考察服务转介率、资源实际获得率、服务持续性、风险变化、群体误判比例、调查负担及申诉纠错情况，并区分“信息闭环”与“问题解决闭环”：前者仅指预警已经推送、核查和结案，后者要求对象实际获得服务、困难得到缓解。只有任务绩效转化为福利绩效，且不明显损害分配和程序绩效，才能认定风险预测改善了福利治理。

（三）赋予基层专业人员实质改变权限

数智工具可能同时扩大上级部门的识别能力和基层工作人员的解释负担。目标、数据和阈值的控制权通常集中于平台及管理部门，分类造成的冲突和责任却在一线显现。如果工作人员只能确认工单、补录信息或为系统结论签字，不能改变分类、调动资源或调整处置，人工复核便可能流于形式。^②

有效的人机协作至少需要四项条件：工作人员能够获得平台数据之外的家庭信息；能够根

① Danielle Keats Citron, Frank Pasquale, "The Scored Society: Due Process for Automated Predictions," *Washington Law Review*, 2014, 89(1).

② 章晓懿：《社区能力视角下的社会救助瞄准机制研究：转型国家的经验》，《社会保障评论》2017年第2期；Madeleine Clare Elish, "Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction," *Engaging Science, Technology, and Society*, 2019, 5; Ben Green, "The Flaws of Policies Requiring Human Oversight of Government Algorithms," *Computer Law & Security Review*, 2022, 45.

据个案情境推翻或调整系统分类；能够记录不同意见而不承担不合理的组织压力；能够启动跨部门服务，并将重复出现的错误反馈至平台规则。只有如此，专业裁量才是连接一般规则与个体处境的实质纠错机制。改变权限也须与责任记录相结合。制度应保存系统建议、工作人员判断、处置结果及其理由，以区分错误究竟来自数据、模型规则、组织要求、资源不足还是个案判断，使人工复核同时服务于个案纠错和系统改进。

（四）建立贯穿完整链条的可争议性与反馈机制

本文所称“可争议性”，是指服务对象能够知晓、质疑并促使数据、分类和处置结果得到纠正的制度能力。预测性福利治理的可争议性应覆盖风险预警、个案核实、救助决定、服务转介和复核申诉的完整链条。它是将社会保障利益确立为可救济权利、规范行政权运行的基本程序条件，非附加的技术伦理要求。风险预测一旦影响救助机会、调查概率、资源顺序或证明负担，就应进入理由说明、陈述申辩和权利救济范围。^①《社会救助法》将动态监测、主动发现、个人信息保护、理由说明、听取意见和行政救济置于同一制度框架，为此提供了制度基础。^②

具体而言，应明确数据使用目的、来源、共享范围和保存期限，记录指标或模型版本、风险阈值、人工推翻及最终处置，定期评估不同群体的误报、漏报和处置去向，并允许服务对象在受到不利影响前了解主要依据、补充情境和纠正数据。申诉、人工推翻和基层反馈还应返回指标、阈值和流程设计，使个案纠错转化为系统纠错。^③可争议性还应覆盖尚未形成正式决定的阶段。风险排序可能通过延后联系、降低资源优先级、增加访问频率或反复索取材料产生实际影响，却未必形成可供申诉的书面决定。因此，需要记录预警至处置的完整过程，明确哪些中间措施需要说明理由，并允许对象查询关键数据及其用途，防止权利保护停留在最终审批环节。

中国不必照搬以司法诉讼为中心的域外问责路径，但应建立分阶段监督机制：部署前审查数据用途、指标范围和预期后果；运行中通过行政监督、内部审计、群体影响评估和基层反馈发现持续性偏差；产生个案影响后，通过行政复议、行政诉讼及其他渠道纠正错误。问责不能等到系统造成严重后果之后才介入。中国主动发现机制应沿数据预警、基层核实、需求评估、救助决定、服务转介和复核申诉的完整链条接受评价，重点考察预警首先触发服务还是核查、工作人员能否改变分类并调动资源，以及对象能否纠正关于自身的制度判断。只有使风险预警成为通向权利和服务的线索，“政策找人”才会扩展福利可及性，而非仅扩展分类和监测能力。中央平台、地方政府与基层组织应分层负责：上级部门负责规则说明、群体影响评估和资源保障，地方政府负责服务网络与跨部门协调，基层负责情境核实和专业判断。平台不宜将全部预警直接转化为限时结案任务，基层也不应独自承担系统误报和资源不足的后果；个案经验应通过申诉、审计和运行评估反馈至上游。

① 李德健：《中国社会保障法现代化的路径选择》，《社会保障评论》2025年第2期。

② 全国人民代表大会常务委员会：《中华人民共和国社会救助法》，2026年4月30日通过，2026年7月1日起施行。

③ Danielle Keats Citron, "Technological Due Process," *Washington University Law Review*, 2008, 85(6); Jennifer Cobbe, et al., *Reviewable Automated Decision-Making: A Framework for Accountable Algorithmic Systems*, Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 2021, 598-609.

七、总结与讨论

四案表明，风险并非数据直接揭示的客观事实，而是由行政痕迹、代理标签和既往处置共同建构；预测输出也只有嵌入阈值、默认规则、资源配置和专业流程，才会改变调查概率、服务入口、资源机会与证明负担。四种制度轨迹应从预测介入位置、处置耦合、服务承接、实质纠错和外部问责时序的组合中理解。

本文将福利对象识别的分析单位从“对象—资格”的单次匹配扩展为“数据—标签—输出—流程—处置—反馈”的制度链条。主动发现虽可降低未申请者的进入成本，也可能将监测、调查、排序和自证负担提前配置给个人。识别责任上移至公共机构并不必然带来权利扩张，仍须结合资源投入、权利影响和负担分配判断其福利后果。^①

跨案例比较形成三个条件性命题：风险标签越依赖既往处置，越可能将制度历史转化为个体属性；预测介入越靠前、处置耦合越强而服务能力越弱，支持性识别越容易转为排序和控制；实质改变权限越有限、问责介入越晚，人工复核越可能退化为责任下沉。^② 预测治理因此应从任务、福利、分配和程序四维评价，并按照权利影响、强制程度和错误可逆性配置“最小充分保障”。^③ 只有识别能力转化为现实服务，且不以不均衡的行政负担和难以纠正的权利损失为代价，“政策找人”才真正构成福利保障能力的扩展。

本文的跨领域最大差异设计也有明确局限。儿童保护、就业服务和福利合规在政策目标、法定权限、错误成本及资源条件方面存在实质差异，本文不能直接比较四个系统的预测准确率、总体绩效或优劣，也不能将单一案例的后果直接外推至其他领域或中国。各案例公开资料亦不均衡：AFST 能够观察部署后的组织实践，Gladsaxe 只能分析部署前的授权和正当性争议；现有材料也较少呈现家庭、求职者和福利领取者如何理解主动联系、风险标签与申诉成本。因此，本文提出的是分析性机制命题，而非已经完成因果检验的普遍规律。

后续研究应进入预警生成后的组织过程，结合平台日志、个案记录、工作人员访谈和服务对象叙事，追踪预警如何经历派单、核实、推翻、转介与结案；比较高风险者实际获得的支持、

① 李棉管：《技术难题、政治过程与文化结果——“瞄准偏差”的三种研究视角及其对中国“精准扶贫”的启示》，《社会学研究》2017年第1期；Donald P. Moynihan, et al., "Administrative Burden: Learning, Psychological, and Compliance Costs in Citizen-State Interactions," *Journal of Public Administration Research and Theory*, 2015, 25(1); Pamela Herd, Donald P. Moynihan, *Administrative Burden: Policymaking by Other Means*, Russell Sage Foundation, 2018; Wim van Oorschot, "Who Should Get What, and Why? On Deservingness Criteria and the Conditionality of Solidarity among the Public," *Policy & Politics*, 2000, 28(1).

② Paul Henman, Greg Marston, "The Social Division of Welfare Surveillance," *Journal of Social Policy*, 2008, 37(2); Emily Keddell, "The Ethics of Predictive Risk Modelling in the Aotearoa/New Zealand Child Welfare Context: Child Abuse Prevention or Neo-Liberal Tool?" *Critical Social Policy*, 2015, 35(1); Joanna Redden, et al., "Datafied Child Welfare Services: Unpacking Politics, Economics and Power," *Policy Studies*, 2020, 41(5); Mark Bovens, Stavros Zouridis, "From Street-Level to System-Level Bureaucracies: How Information and Communication Technology Is Transforming Administrative Discretion and Constitutional Control," *Public Administration Review*, 2002, 62(2); Madeleine Clare El-ish, "Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction," *Engaging Science, Technology, and Society*, 2019, 5; Ben Green, "The Flaws of Policies Requiring Human Oversight of Government Algorithms," *Computer Law & Security Review*, 2022, 45.

③ 郑功成：《中国特色社会保障制度论纲》，《社会保障评论》2024年第1期；杨立雄：《从兜底保障到分配正义：面向共同富裕的社会救助改革研究》，《社会保障评论》2022年第4期。

低风险者可能错失的机会，并利用纵向资料检验调查、服务不足或资源退出形成的新记录是否反过来强化下一轮分类。只有将模型、组织过程、资源流向和权利后果置于同一经验链条，才能判断风险预测是在缓解生活风险，还是仅仅提前分配关注、资源与怀疑。数智时代福利治理的真正考验，不是公共机构能否看见更多未来风险，而是个体被提前看见之后，能否获得以权利为基础的持续支持，并能就相关决定提出申诉、促使错误得到纠正。^①

Institutionalising Risk Prediction: A Multi-Case Comparison of Welfare Governance in the Era of Digital Intelligence

Zhu Ning

(School of Social Development and Public Policy, Fudan University, Shanghai 200433, China; Department of Social Sciences, University of Eastern Finland, Kuopio 70150, Finland)

Abstract: In the era of digital intelligence, welfare governance is shifting from ex post case handling towards the allocation of investigations, services, and resources on the basis of future risk. Through a comparison of cases from four countries, this article examines how risks that have not yet materialised become institutionalised and how predictive outputs are translated into welfare action. The findings show that risk is constructed through administrative records, proxy labels, and prior interventions, while predictive outputs are translated into intervention, prioritisation, or control through thresholds, default rules, resource allocation, and professional procedures. The four cases respectively exhibit trajectories of continued operation and calibration, cancellation before deployment, obstruction during implementation, and termination after operation. These differences arise from where prediction enters the process, the coupling between prediction and welfare action, service delivery capacity, professionals' authority to make substantive changes, and the timing of external accountability, and are further shaped by rights norms, political contestation, and institutionalised veto channels. The article argues that the performance of predictive welfare governance should be assessed across four dimensions: task, welfare, distribution, and procedure. "Minimum sufficient safeguards" should be calibrated to the consequences of welfare action, the degree of coercion, and the reversibility of errors, so that those identified receive sustainable support with effective avenues for appeal and error correction.

Keywords: predictive welfare governance; welfare recipient identification; administrative burden; welfare rights; multi-case comparison

(责任编辑：李莹)

① Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, St. Martin's Press, 2018; Marion Fourcade, Kieran Healy, "Classification Situations: Life-Chances in the Neoliberal Era," *Accounting, Organizations and Society*, 2013, 38(8).