



人工智能应用于人类辅助生殖胚胎选择的伦理挑战与治理

唐健^{1*}, 令狐甫宜¹, 茫烈²

1. 天津医科大学 医学人文学院, 天津 300070;

2. 天津和睦家医院 辅助生殖中心, 天津 300074

摘要: 人类辅助生殖中的体外受精 (IVF) 技术依赖于胚胎选择来提高成功率。传统方法主要基于形态学评估, 虽透明度高且患者参与度强, 但存在主观性和低效率问题。近年来, 人工智能 (AI), 特别是深度学习技术, 为胚胎选择提供了新的可能性。然而, 该技术目前尚处于早期阶段, 面临透明性、患者自主性、责任归属和公平性等伦理问题的评估。AI 的“黑盒”特性可能影响临床可解释性以及患者知情同意机制。而 AI 的应用可能改变责任界定, 并对辅助生殖的公平性和安全性带来潜在影响。在当前阶段, 应审慎推进 AI 辅助胚胎选择, 避免盲目依赖技术; 要强化伦理治理, 推动透明度提升和长期跟踪研究, 以确保其健康发展。

关键词: 人类辅助生殖技术; 人工智能; 人类胚胎选择; 生命伦理学; 科技伦理治理

中图分类号: R-052; B823.4; TP2

文献标识码: A

文章编号: 1674-4969(2025)06-0672-11

引言

人类辅助生殖中的体外受精 (*in vitro* fertilization, IVF) 是一项革命性的临床技术, 为不孕不育患者实现生育愿望提供了新的选择。自 1988 年中国大陆地区首例试管婴儿诞生以来, 人类辅助生殖技术在国内迅速发展, 并在 1996 年成功应用卵胞质内单精子注射

(ICSI), 进一步提高了男性因素不孕的治疗可行性^[1]。这些技术的进步使在实验室内受精与胚胎培养成为可能, 之后将胚胎移植到子宫内, 从而提高了妊娠率。然而, IVF 的成功率在很大程度上取决于胚胎的选择和移植过程。优化胚胎选择可以帮助接受治疗的患者以最少的移植次数实现受孕。传统的胚胎选择方法多基于显微镜下的形态学观察, 胚胎学家会在特定的时间点对胚

收稿日期: 2025-03-16; **修回日期:** 2025-04-23

基金项目: 天津市教委社会科学重大项目: 高风险、高价值生物医学创新重点领域的伦理治理研究 (2024JWZD28)

作者简介: *唐健 (1983—), 男, 博士, 副教授, 研究方向为医学伦理学。E-mail: dmhjt@tmu.edu.cn (通讯作者)

令狐甫宜 (2002—), 女, 硕士研究生, 研究方向为医学伦理学。E-mail: 13908312435@163.com

茫烈 (1981—), 男, 硕士, 实验师, 研究方向为人类辅助生殖技术。E-mail: manglie@163.com

引用格式: 唐健, 令狐甫宜, 茫烈. 人工智能应用于人类辅助生殖胚胎选择的伦理挑战与治理[J]. 工程研究——跨学科视野中的工程, 2025, 17(6): 672-682. DOI: 10.3724/j.issn.1674-4969.20250037. CSTR: 32282.14.JES.20250037

TANG J, LINGHU F Y, MANG L. Ethical Challenges and Governance of Artificial Intelligence in Embryo Selection for Human Assisted Reproduction[J]. Journal of Engineering Studies, 2025, 17(6): 672-682. DOI: 10.3724/j.issn.1674-4969.20250037. CSTR: 32282.14.JES.20250037

胎进行静态评估。然而, 胚胎发育是一个复杂的动态过程, 仅通过静态观察可能无法准确反映胚胎的潜在发育能力^[2]。此外, 胚胎质量评估还会受到操作员间差异和主观因素的影响^[3]。在 IVF 中, 时间延迟成像技术 (time-lapse videography, TLV) 被用来动态地观察胚胎发育, 从而识别出用于胚胎选择的发育参数。近年来, 随着人工智能 (artificial intelligence, AI) 的迅速发展, 客观地解读和处理大量的视觉、表格数据成为可能。将人工智能的深度学习 (deep learning, DL) 技术与时间延迟成像技术相结合, 理论上可以极大地提升胚胎选择的准确性和效率^[4-5]。然而, 技术的飞速发展也带来了一系列伦理问题, 如透明性、患者自主性和责任归属等。目前, 对于使用 AI 进行胚胎选择所引发的伦理、社会和监管等问题的研究还十分有限。在 AI 应用于胚胎选择的早期发展阶段, 我们需要充分评估其可能存在的伦理问题, 以确保技术的正向发展。因此, 本文将从胚胎选择技术出发, 探讨 AI 在人类辅助生殖中的伦理问题, 包括透明度、患者自主性、责任归属和公平性, 以期提出较为系统的伦理治理路径, 为相关研究与实践提供参考。

1 胚胎选择方法的演变

每年, 有数百万夫妇因不孕问题而寄望于 IVF 技术。自 1978 年首例试管婴儿诞生以来, 全球已有超过 900 万名试管婴儿诞生^[6]。然而, 对于每对尝试的夫妇来说, IVF 并非总是能带来成功。在英国, 所有年龄组的平均单次治疗周期活产率仅为 26.1%^[7]。在 IVF 过程中, 女性会接受激素治疗, 以刺激多枚卵子的发育。这些卵子随后会被取出、受精, 并在实验室中培养 3~6 d。通常, 胚胎会在第 5 或第 6 天达到囊胚阶段。此时, 胚胎学家会选择一个囊胚进行移植到子宫中。任何被认为具有潜力但未被使用的胚胎都将被冷冻保存, 以备未来使用。其中, 胚胎的选择过程对 IVF 的成功率具有重要影响。

1.1 传统的胚胎选择方法

传统胚胎选择通常包含两个连续的步骤: 首先, 对胚胎的原始图像进行手动标注, 随后在胚胎中进行

排序^[4]。这种选择的标注依赖于在特定时间点通过显微镜对胚胎进行数次快照观察。基于形态学的胚胎选择是传统胚胎选择的主要方法。在卵裂期胚胎中, 可以评估细胞数量、碎片化程度、均一程度以及多核现象的存在^[9]。然而, 静态观察无法标注出如首次细胞分裂到两细胞阶段的时间、细胞周期的持续时间以及囊胚形成开始时间等参数^[10]。为了克服静态观察的局限性, 临床引入了时间延迟成像技术 (TLV), 它能够识别出新的胚胎发育参数^[12-13]。与静态形态学方法相比, 使用时间延迟成像技术的传统胚胎选择在排序环节中, 可以通过具有规则集的算法确定最具潜能的胚胎, 例如, 可以手动标注胚胎特征以生成简单的手动决策树^[11]。

传统的胚胎选择方法具有明显的优缺点。从优势方面来说, 胚胎选择过程中, 胚胎学家能够向患者夫妻详细解释胚胎选择的具体过程、形态评分结果以及胚胎移植的计划。所有这些评估步骤和结果都是透明的, 易于监管、可溯源, 患者有机会就最终移植哪个胚胎发表自己的意见, 通过与患者夫妻的共同决策完成最后的最优胚胎选择, 患者参与度高。此外, 由于该方法作为全球公认的协议, 对新晋胚胎学家的培训和投资要求较低, 因此许多医院仍然选择继续使用仅基于形态学的胚胎选择方法。然而, 在传统胚胎选择中, 无论是使用静态图像的传统胚胎选择还是使用 TLV 图像的传统胚胎选择, 都需要手动标注胚胎的特征, 排序也主要依据胚胎学家的经验判断, 缺乏统一的标准, 容易产生主观性的偏差, 耗时较长, 效率低下。

1.2 AI 辅助胚胎选择的技术革新

随着人工智能技术的不断进步, 人类辅助生殖领域也开始探讨将 AI 应用于胚胎选择的可行性。作为 AI 的核心之一, 机器学习 (machine learning, ML) 在处理复杂数据以及识别胚胎学家难以察觉的细节方面, 展现了其独特的优势^[14]。借助生成模型, 机器学习可以自动从图像或其他数据中学习并预测结果。同时, 计算机视觉 (computer vision, CV) 技术的引入, 使得分析和处理大量胚胎图像数据变得更加高效和自动化。

AI 在胚胎选择中的应用主要以决策支持系统 (decision support systems, DSSs) 的形式呈现, 这些系

统主要依赖于三种模型类型：透明盒、遮光盒和黑盒。其中，透明盒和遮光盒模型在标注步骤中既可以选择手动标注胚胎特征，也可以选择自动标注已知胚胎特征。在胚胎排序过程中，透明盒模型主要运用可解释的ML方法，主要是基于统计学的方法，如逻辑回归、决策树、随机森林和贝叶斯网络进行排序。而遮光盒则使用黑盒式方法对胚胎进行排序。黑盒模型则主要利用深度学习技术，尤其是神经网络，例如卷积神经网络^[15]，直接从原始数据中预测结果。作为机器学习的一个子类别，深度学习依赖于具有多个隐藏层的人工神经网络。虽然它在胚胎选择中的临床应用近年来取得了迅猛发展^[16]，但这些隐藏层的复杂性使得其推理过程难以解释，因此被形象地称为“黑盒”^[17]。

AI辅助胚胎选择相较于传统方法具有显著的优势。在面对大量数据时，AI的机器学习技术尤为有用，尤其是当胚胎学家们无法立即找到利用这些数据以提高预测能力的方法，或者无法手动处理所有数据以生成有意义的结果时。此外，AI可以让许多流程实现标准化和自动化，从而减少观察者之间的差异性，提供更加客观的胚胎评估^[18]。

2 AI应用于胚胎选择的伦理挑战

尽管AI在胚胎选择中的应用具有巨大的潜力，但在其早期发展阶段仍需谨慎对待。其标准化和自动化流程的特性在提升效率的同时，也可能带来一系列伦理挑战，例如决策透明度不足、威胁患者自主性、技术后果责任归属不清，以及削弱社会公平等。因此，这些问题需要全面评估，以确保该技术的发展能够造福更多的患者。

2.1 决策透明度不足

随着对AI在胚胎选择中的应用期望不断升高，黑盒决策支持系统（black-box DSS）逐渐成为了该领域的研究重点^[15]。这类系统通常基于不可解释的黑盒模型，这些模型要么因其复杂性难以被人类理解，要么因专有技术的限制而无法揭示其工作原理^[19]。在胚胎选择过程中，胚胎学家既无法理解黑盒模型的预测结果背后的逻辑，也无法明确黑盒模型在决策中依据了哪些具体

指标，整个过程都显得“不透明”。

不同AI模型之间的“不透明”性体现在它们缺乏统一的选择标准和互通性。尽管AI在胚胎选择中显示出一定的潜力，但不同模型在选择胚胎时使用的标准和算法往往不一致，这使得它们的结果难以比较或整合。例如，某些AI模型可能更注重胚胎的形态学特征，而其他模型则可能更多依赖胚胎的发育参数或培养液中的化学成分。这种差异使得不同AI系统之间的选择结果既不可互通，也无法在不同环境或设备条件下统一应用。实际上，AI模型的评估标准往往与临床实践存在脱节，数据和代码并未公开，因此，在不同机构之间共享或跨系统使用这些模型时，可能会面临显著的挑战。当前，评估AI胚胎选择模型效力的研究主要集中在以下两个关键标准：首先，是考虑对患者有意义的结果，即AI选择的胚胎是否能提高生育成功率，如活产率或临床妊娠率。这一标准能够直接反映AI的临床价值，但由于影响生育结局的因素众多，如母体健康状况、移植技术等，仅依赖该标准来评估AI可能会受到较大干扰。其次，是与现有标准的一致性，即AI对胚胎质量的评估是否与胚胎学家的判断相符。该标准相对容易操作，可快速验证AI的可靠性，但它依赖于现有的人类经验，而人工评估本身也可能存在主观性和局限性。这两个标准在衡量AI的透明度时各有优势，但也存在一定的矛盾。以“活产率”为核心的评估方式虽能直接反映AI的实际效果，但它受到非胚胎因素的影响，导致AI训练难以依赖此标准。而以“与胚胎学家评估一致性”为标准，虽然能够提供一个更直接的验证方法，但它可能限制AI发展出更优的判断标准，而仅仅是模仿人类专家的现有评估体系。当前，AI在胚胎选择中的透明性问题，正是由于这两个标准的局限性，使得不同AI模型之间难以建立统一的选择依据，也限制了AI评估结果的可解释性和推广应用。

以Tran、Cooke和Illingworth的研究为例^[20]，他们开发并评估了一个名为IVY（information value model）的模型，该模型通过一个置信度范围（0表示“绝对不会植入”，1表示“绝对会植入”）预测胚胎是否会发展为临床妊娠。研究显示，IVY模型的ROC（receiver operating characteristic curve）曲线AUC（area under

the curve) 达到了0.93 (AUC为机器学习中用于评估分类模型性能的重要指标, 越接近1说明性能越好)。然而, 该算法在训练和测试过程中使用的大多数胚胎质量极差, 这些胚胎即使不经筛选也会被淘汰, 从而导致AUC值被人为抬高。他们进一步强调, 临床实际需求在于从一组可能具有活力的胚胎中, 识别出成功概率最高的胚胎, 而非从那些质量明显较差的胚胎中进行筛选。另一方面, Khosravi等^[21]在研究中更侧重于与现有标准的一致性。他们依据多位胚胎学家的共识, 将胚胎划分为优质、中等和劣质3个组别。接下来, 他们评估了AI模型在识别优质和劣质胚胎方面的能力 (不包括中等胚胎), 并发现该模型在此任务中的准确率高达96.94%, 超过了某些胚胎学家的水平。然而, 在实际临床中, 当试图从一组质量相似的胚胎中选择最佳胚胎时, 这种将胚胎宽泛地划分为“优质”或“劣质”的方法, 其效用有限。

2.2 威胁患者自主性

AI的应用可能在一定程度上误导患者的价值偏好, 进而影响患者的自主性。研究表明, 男性和女性胚胎在早期形态动力学特征上存在差异^[22-25], 其他特征在这一阶段也可能表现出类似的差异。如果胚胎选择模型将这些特征视为与植入成功率相关, 就可能倾向于系统性地选择具有特定特征的胚胎。然而, 当患者希望随机选择胚胎性别时, 这种模型的运行方式可能会违背其价值观和自主选择权。在过去的几十年中, 临床实践逐渐从传统的父权主义模式转向以促进患者自主性为核心的共同决策模式^[26]。在我国, 辅助生殖技术的实施要求医疗机构严格管理患者全过程, 确保治疗方案的合规性。患者在接受辅助生殖治疗前, 需提供结婚证、准生证等证件, 并完成规范的知情同意程序^[27]。然而, 不透明的AI模型可能削弱这种共同决策模式, 因为临床医生和患者难以理解模型的决策依据^[28]。例如, 关于某个胚胎为何被选择或未被选择的关键信息 (如细胞数量和对称性、是否存在碎片化、植入成功的概率以及失败的可能原因) 通常不明晰, 这可能会降低医患在胚胎选择过程中的信任与协作。这一问题进一步加剧了临床实践中可能出现的“机器父权主义”现象。

尽管黑盒AI系统被宣传为临床医生的辅助决策工具, 但实际上, 这些系统可能使医患共同决策模式变得更为脆弱。研究表明, 人们往往容易对自动化系统产生过度依赖^[29], 并倾向于无条件接受那些熟悉或被认为表现优越的AI系统^[30]。在这种情况下, 临床医生很难质疑黑盒AI的判断结果。由于算法信息的不透明性, 患者也很难充分了解AI的决策依据, 这无疑增加了患者做出有效知情同意的难度。知情同意是患者自主性的重要基础, 要求患者能够理解与其医疗决策相关的信息。然而, 当AI和临床医生都未能清晰理解选择某个胚胎的理由时, 患者便更难以知晓不同选项的风险与益处。这种信息不对称可能使患者在决策过程中被动依赖临床医生或模型推荐, 从而削弱了其自主性和主动性。此外, 在某些情况下, 患者甚至可能因为缺乏清晰的信息而对整个医疗过程产生不信任, 这进一步挑战了医患关系的稳定性。

2.3 技术后果责任归属不清

在人工智能应用于胚胎选择时, 尤其是不透明的AI模型, 可能会使责任归属变得复杂, 甚至产生“责任空缺”现象。在传统医疗实践中, 临床胚胎学家对其决策负责, 包括胚胎的选择、培养过程的监测以及最终妊娠结局的评估。然而, 当AI介入甚至主导胚胎选择决策时, 如果出现不良后果, 如胚胎移植失败、低成功率, 甚至母体或婴儿健康受损, 责任如何划分便成为一个棘手的问题。

在传统归责体系中, 医生对患者的治疗决策负有直接责任。若患者因医疗失误受损, 他们有权通过法律途径追究医生的过失责任。但在AI辅助甚至主导决策的情境下, 医生的责任是否依然适用则成了一个问题。如果医生只是依据AI的推荐作出决定, 那么医生是否仍然应该为结果承担责任? 另一方面, 医院是否因为引入AI系统而需对患者负责? AI模型的开发是否应承担部分责任? 如果用于胚胎选择的人工智能模型推理方式不可解释, 那么法院如何评估医生的决策就变得不明确, 这也使得伤害的责任如何判定变得不清晰^[31-32]。这种责任空缺现象, 意味着缺乏对AI开发者等相关主体的追责机制。由于多主体间的责任不确定性, 进一步加剧了责任归属的模糊性。责任空缺

问题主要源于两个方面。首先, AI的决策过程可能是黑盒式的, 临床医生无法完全理解其推理逻辑, 因此当患者质疑治疗方案时, 医生可能无法提供足够的解释和合理的医学依据。其次, 现行医疗责任体系是基于人的决策而构建的, 而AI的引入打破了传统的责任分配模式, 导致责任主体难以明确。如果不能确定责任人, 患者的合法权益可能难以得到有效保护。

这种责任归属的不确定性可能带来严重的伦理和法律后果。首先, 医生可能会因为担忧无法控制AI的决策风险, 对AI辅助决策持怀疑态度, 进而阻碍其在临床中的应用。其次, 如果责任无法明确归属, 患者在遭受损害后可能面临求偿无门的困境, 这不仅会损害医患之间的信任, 还可能对医疗系统的公正性构成威胁。再者, 如果AI开发者无需承担责任, 那么技术监管的缺失可能导致潜在风险被忽视, 进而影响到人工智能在医疗领域的健康发展。

2.4 削弱社会公平

AI在辅助生殖领域的应用带来了胚胎选择的技术革命, 但同时也引发了公平性的问题。首先, AI可能对胚胎的道德地位产生进一步冲击。胚胎不仅被视为潜在的生命, 还承载着伦理和社会层面的价值。然而, 随着AI的介入, 胚胎可能被简化为基于算法评分的“对象”, 其内在的生命价值和伦理考量可能因此被忽视。这种算法驱动的选择可能将胚胎等同于一组数据或特征的集合, 而不再被看作拥有潜在生命权利的个体。这种技术工具化的倾向带来了问题: 一方面, AI通过客观算法优化选择, 减少了人为偏差; 但另一方面, 它也强化了工具主义视角, 仅将胚胎视为可供评估的“资源”。例如, 如果AI过度关注基因或形态特征, 可能导致某些被算法认为“不优”的胚胎被系统性排除, 但这些胚胎可能具有存活甚至发育为健康个体的潜力。

其次, AI在胚胎选择中的应用可能对人类遗传多样性和群体公平性产生影响。如果AI模型存在特定特征的偏向性, 可能会导致性别比例失衡或某些种族群体在选择过程中被系统性忽视。此外, AI可能基于早期形态动力学特征作出判断, 而这些特征可能偏向特定人群, 从而引发社会不公。残疾活动家提出

的“表现论者的反对意见”(expressivist objection)^[33]认为指出, 基于预测健康的选择可能构成对残障群体的歧视, 并传达负面社会信号。例如, 唐氏综合征筛查^[34]。AI的介入可能使这一争议进一步扩大, 使得任何可能降低着床机会的特征都成为潜在的筛选标准。

此外, AI的引入可能会加剧辅助生殖治疗的经济不平等。由于黑盒模型通常依赖于特定的训练系统, 医院需要购买相应的配套设备, 这使得AI公司在市场中拥有更多的经济控制力, 从而增加了医疗成本, 限制了IVF服务的可及性。同时, AI公司与临床医生之间的信息不对称问题也变得更加严重。胚胎学家可能需要依赖他们难以理解的决策程序, 这影响了对模型可靠性的评估。例如, 在放射学领域, 尽管黑盒模型已广泛应用于医学影像分析, 但其临床表现仍存在不确定性^[35], 引发了对其普适性的担忧^[36]。

3 对AI应用于胚胎选择的伦理治理

近年来, 我国逐步建立了医疗人工智能的治理框架, 自2015年起陆续出台相关政策和法律, 以规范技术应用的安全性与伦理性。特别是2018年的《国家健康医疗大数据标准、安全和服务管理办法(试行)》^[37]、2020年的《信息安全技术个人信息安全规范》^[38]以及2021年颁布的《中华人民共和国数据安全法》^[39]等法律法规, 为AI在医疗领域的应用提供了法律依据和伦理指引^[40]。在当前阶段, AI在胚胎选择中的临床应用需以审慎为原则。以2022年中共中央办公厅、国务院办公厅印发的《关于加强科技伦理治理的意见》^[41]为出发点, 结合生命伦理学的观点^[42-43], 短期内应在强化监管、建立长期跟踪机制、推动数据与代码公开透明、优先发展可解释的AI模型以及推动跨学科合作5个方面采取具体措施, 以强化伦理治理。

3.1 强化监管

当前的医学人工智能模型大多以“良好实践”为监管标准, 违反规定并不会受到法律惩罚。然而, 这种方式缺乏法律约束力, 难以应对AI在临床环境中的特有风险。因此, 推动模型的“严格”监管立场将是更

为可取的方法。目前的监管方法试图将医学人工智能模型作为一种医疗设备进行监管, 但监管应进一步要求, 如果可解释模型与黑盒模型表现相当, 应禁止使用黑盒模型, 或者要求黑盒模型提供优化后的可解释版本, 以提升模型的透明性和安全性。此外, 对于2.4节提到的公平性问题, 最好的应对方式也是持续监控这些影响, 并确保社会反应有助于促进所有人的平等, 包括残障人群。因此, 与其放弃选择, 不如确保提供足够的社会资源和支持, 让所有人都有合理的机会过上美好生活^[35]。可解释AI可能有助于更早、更有效地识别偏见问题。具体来说, 如果能够发现与各种结果相关的可解释特征, 就更容易监控可能的偏见并避免。

3.2 建立长期跟踪机制

许多国家的研究人员、专业学会和监管机构日益认识到, 在辅助生殖领域引入新技术后, 对出生儿童的健康与安全进行持续监测至关重要。这一原则适用于所有新型干预措施的应用, 如卵胞质内单精子注射 (ICSI)、胚胎植入前遗传学检测 (preimplantation genetic testing, PGT) 以及AI的辅助技术。尽管这种监测过程可能给家庭带来一定负担, 但由于临床试验的规模和样本量有限, 无法全面评估新技术的长期安全性与有效性。因此, 后续监测作为临床试验的延续是不可或缺的, 直至技术的安全性和有效性得到充分验证。

在此基础上, 研究人员需通过更为严格的评估手段, 为长期跟踪机制提供更可靠的依据。例如, 随机对照试验 (RCTs) 是评估人工智能用于胚胎选择模型的黄金标准。研究应将这些模型与最佳临床判断或已投入实践并表现良好的黑盒式人工智能模型进行对比, 重点评估多个关键指标, 包括从治疗到活产所需的时间、每次活产所需的胚胎移植次数、相关成本分析、每次取卵的活产率以及婴儿的健康状况。此外, 进行RCTs之前, 必须通过使用不同医院数据的验证研究, 以提高模型的广泛适用性。这种从临床试验到后续监测的全周期评估体系, 是保障辅助生殖技术健康发展和患者利益的关键举措。

3.3 推动数据与代码公开透明

为提高技术的可重复性和改进潜力, 用于创建机器

学习模型的数据和代码应当公开获取。开放数据集不仅能促进学术界对模型的验证, 还能增强公众对技术的信任。例如, 乳腺癌筛查相关的大规模数据集已经由美国国立卫生研究院 (National Institutes of Health) 资助并公开提供^[44], 胚胎选择领域也可以借鉴类似做法。然而, 在数据开放的同时, 需解决隐私保护问题。应制定严格的程序, 以在共享数据时保护患者隐私, 可通过数据匿名化实现这一目标。当前, 临床医生通常借助决策辅助工具如Gardner评分向患者解释胚胎选择的依据, 这些工具基于生物学上有意义的参数进行评估。理想情况下, 患者应全面了解胚胎选择的优势 (如提高受孕概率)、局限性 (如不能完全保证胎儿的健康) 以及潜在的未知风险 (可能的产科问题)。此外, 患者还应清楚模型给出建议的依据, 像是细胞数量、对称性和发育规律等因素。在此基础上, 应尽可能将患者的个人价值观纳入决策过程。然而, 如果使用黑盒模型, 解释模型得出建议的具体依据以及如何融入患者价值观将面临较大挑战。因此, 在黑盒模型的使用场景中, 确保患者能够理解并参与决策过程, 是需要特别关注的重要问题。

从长远视角出发, 将AI应用于胚胎选择, 必须正视其可能带来的深远社会和伦理影响, 同时积极推动技术发展方向优化。

3.4 优先发展可解释的AI模型

当前, 黑盒人工智能因其高效性和便捷性在医学领域得到了广泛推广, 但在胚胎选择方面, 其不透明性却引发了关于安全性和伦理的争议。因此, 应该优先发展可解释的AI模型, 以提高决策的透明度和可控性。只要合理设计出符合领域需求的可解释性指标, 此类模型的预测能力通常并不会逊色于黑盒模型^[45]。

首先, AI的可解释性有助于发现混杂因素并提高模型的纠错能力。黑盒模型可能因依赖隐含变量而产生偏差, 而可解释模型由于其工作机制透明, 使得临床医生能够识别并调整不合理的预测因素。例如, 在某些情况下, 预测医院治疗效果的AI可能依赖于X射线设备的类型, 而非影像内容本身, 因为特定设备的使用频率可能反映患者病情的严重程度。通过可解释模型, 可以避免这类问题, 从而提高AI决策的可靠性。

其次, AI能够增强适应性, 使模型在不同环境、设备或培养基条件下保持稳定。AI常常面临“领域适配”的挑战, 即模型可能无法直接适用于不同的临床环境。然而, 如果AI的预测依据是透明的, 那么临床医生可以在新环境下调整决策逻辑, 从而减少模型迁移时的性能损失, 并降低医院对特定设备供应商的依赖。这样的策略有助于增强AI在不同环境中的适用性和稳定性。

此外, 可解释模型能够明确责任归属, 从而提高医患之间的信任度。胚胎学家的核心职责涵盖了监测胚胎形态、培养过程中的动态变化, 以及妊娠和活产率。当遇到异常情况时, 透明的AI模型有助于医生识别影响因素, 并结合专业知识作出相应的调整。相比之下, 黑盒模型的不可解释性可能导致医生难以评估模型决策的合理性, 从而影响医疗责任的明确划分。在实际应用中, 患者对AI辅助决策的信任度与其可理解性息息相关。透明的模型更有助于患者参与决策, 使得他们的价值观(如对残障风险、性别偏好或胚胎移植数量的考量)得到充分的考虑。

推动可解释AI的发展, 从长远来看, 不仅能提升胚胎选择技术的公平性与可靠性, 还能增强临床实践的透明度, 从而为辅助生殖领域的人工智能应用构建更为坚实的伦理与社会基础。这一发展对相关领域具有深远意义。

3.5 推动跨学科合作

当前, 胚胎选择相关数据集尚未得到广泛开放。轻率地公开这些数据可能会产生更多的负面影响, 甚至可能使机器学习技术的评估过于片面。因此, 发布数据集的同时, 必须提出与实际应用相符且充分考虑伦理问题的评估标准, 这需要胚胎学家、伦理学家以及人工智能和统计学领域的研究者进行深入的讨论, 并妥善解决隐私问题。这一讨论应在数据发布后持续进行。尽管如此, 让更广泛的人工智能社区能够查看数据并参与分析, 将有助于确保存在缺陷和偏见的评估不会被轻易忽视, 也有可能揭示我们尚未意识到的其他重要问题。

4 结论

人工智能, 特别是深度学习技术, 在人类辅助生殖领域的应用, 为提升体外受精(IVF)中胚胎选择的效率与准确性提供了新的技术路径。传统形态学评估虽具备过程透明、患者参与度高等优势, 但其主观性强、观察者间差异显著及评估效率低下等问题, 制约了IVF成功率的进一步提升。时间延迟成像技术(TLV)与AI的结合, 理论上可通过自动化、标准化分析海量动态图像数据, 实现对胚胎发育潜能更为客观、精准的预测。然而, 本文的系统性分析表明, 该项前沿技术在向临床常规转化过程中, 正面临一系列深刻且复杂的伦理挑战, 其潜在风险亟需审慎评估。

具体而言, 其一, 决策透明度不足: “黑盒”模型的不可解释性削弱了临床决策的可追溯性, 亦为监管与验证带来困难; 其二, 患者自主性受威胁: 算法不透明导致知情同意难以真正落实, 存在“机器父权主义”风险, 患者可能被动接受系统推荐; 其三, 责任归属模糊: AI介入后, 医生、机构与开发者之间的责任边界不清, 易形成“责任空缺”, 影响患者权益保障; 其四, 社会公平性堪忧: 技术可能加剧辅助生殖服务的不平等获取, 算法偏见或对胚胎的工具化倾向还可能冲击遗传多样性与生命尊严。

鉴于当前技术尚不成熟、评估标准缺位, AI不应被仓促引入临床常规实践, 而应在严谨的伦理评估与制度化监督框架下审慎推进。未来应构建以伦理治理为核心的健康发展路径: 短期内, 须强化监管, 优先发展可解释性强的“透明盒”模型, 建立以随机对照试验和长期跟踪研究为基础的科学评估体系, 并在保护隐私前提下推动数据与代码的公开; 长远而言, 应深化跨学科合作, 持续关注技术普适性、医患互动模式演变及社会文化影响, 不断完善动态治理机制。

本研究提出的治理思路, 与我国现行科技伦理治理体系相衔接, 旨在为相关政策制定提供理论支持与实践参考。尤其在AI模型持续演进的背景下, 强调加强监管、长期跟踪、信息透明、临床可解释性与跨学科协同, 有助于提升公众信任与技术可接受度。未来研究可进一步探讨AI模型在不同人群与机构中的

适用性差异, 分析患者价值取向与决策行为, 以及临床团队如何整合专业判断与算法建议, 同时推动可解释性模型的规范应用, 平衡数据开放、跨机构验证与

隐私保护之间的关系, 促进伦理治理体系的持续完善。

致谢

感谢张新庆教授和李振良教授对本文写作所提出的宝贵修改建议。

参考文献

- [1] 胡琳莉, 孙莹璞. 中国辅助生殖技术发展历程和现状[J]. 生殖医学杂志, 2019, 28(10): 1113-1114.
Hu L L, Sun Y P. Development course and present situation of assisted reproductive technology in China[J]. Journal of Reproductive Medicine, 2019, 28(10): 1113-1114.
- [2] Gardner D K, Meseguer M, Rubio C, et al. Diagnosis of human preimplantation embryo viability[J]. Human Reproduction Update, 2015, 21(6): 727-747.
- [3] Liu Y H, Feenan K, Chapple V, et al. Assessing efficacy of day 3 embryo time-lapse algorithms retrospectively: Impacts of dataset type and confounding factors[J]. Human Fertility, 2019, 22(3): 182-190.
- [4] Liu Y H, Qi F, Matson P, et al. Between-laboratory reproducibility of time-lapse embryo selection using qualitative and quantitative parameters: A systematic review and meta-analysis[J]. Journal of Assisted Reproduction and Genetics, 2020, 37(6): 1295-1302.
- [5] Riegler M A, Stensen M H, Witzak O, et al. Artificial intelligence in the fertility clinic: Status, pitfalls and possibilities[J]. Human Reproduction, 2021, 36(9): 2429-2442.
- [6] European society for human reproduction and embryology resources[EB/OL]. (2021) [2025-03-12]. <https://www.eshre.eu/Press-Room/Resources>.
- [7] Human Fertilisation and Embryology Authority. Fertility treatment 2018: Trends and figures quality and methodology report[EB/OL]. (2020-06-23) [2025-03-12]. <https://www.hfea.gov.uk/about-us/publications/research-and-data/fertility-treatment-2018-trends-and-figures/fertility-treatment-2018-quality-and-methodology-report/>.
- [8] US Department of Health and Human Services, Centers for Disease Control and Prevention. 2018 assisted reproductive technology fertility clinic success rates report[EB/OL]. (2018)[2025-03-12]. <https://www.cdc.gov/art/pdf/2018-report/ART-2018-Clinic-Report-Full.Pdf>.
- [9] Alpha Scientists in Reproductive Medicine and ESHRE Special Interest Group of Embryology. The Istanbul consensus workshop on embryo assessment: Proceedings of an expert meeting[J]. Human Reproduction, 2011, 26(6): 1270-1283.
- [10] Nadir Ciray H, Campbell A, Agerholm I E, et al. Proposed guidelines on the nomenclature and annotation of dynamic human embryo monitoring by a time-lapse user group[J]. Human Reproduction, 2014, 29(12): 2650-2660.
- [11] Meseguer M, Herrero J, Tejera A, et al. The use of morphokinetics as a predictor of embryo implantation[J]. Human Reproduction, 2011, 26(10): 2658-2671.
- [12] Liu Y H, Sakkas D, Afnan M, et al. Time-lapse videography for embryo selection/de-selection: A bright future or fading star?[J]. Human Fertility, 2020, 23(2): 76-82.
- [13] Bickendorf K, Qi F, Peirce K, et al. Spontaneous collapse as a prognostic marker for human blastocysts: A systematic review and meta-analysis[J]. Human Reproduction, 2023, 38(10): 1891-1900.
- [14] Fernandez E I, Ferreira A S, Cecilio M H M, et al. Artificial intelligence in the IVF laboratory: Overview through the application of different types of algorithms for the classification of reproductive data[J]. Journal of Assisted Reproduction and Genetics, 2020, 37(10): 2359-2376.
- [15] Dimitriadis I, Zaninovic N, Badiola A C, et al. Artificial intelligence in the embryology laboratory: A review[J]. Reproductive Biomedicine Online, 2022, 44(3): 435-448.
- [16] Curchoe C L, Bormann C L. Artificial intelligence and machine learning for human reproduction and embryology presented at ASRM and ESHRE 2018[J]. Journal of Assisted Reproduction and Genetics, 2019, 36(4): 591-600.
- [17] Tu J V. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes[J]. Journal of Clinical Epidemiology, 1996, 49(11): 1225-1231.
- [18] Rosenwaks Z. Artificial intelligence in reproductive medicine: A fleeting concept or the wave of the future?[J]. Fertility and Sterility, 2020, 114(5): 905-907.

- [19] Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead[J]. *Nature Machine Intelligence*, 2019, 1(5): 206-215.
- [20] Tran D, Cooke S, Illingworth P J, et al. Deep learning as a predictive tool for fetal heart pregnancy following time-lapse incubation and blastocyst transfer[J]. *Human Reproduction*, 2019, 34(6): 1011-1018.
- [21] Khosravi P, Kazemi E, Zhan Q S, et al. Deep learning enables robust assessment and selection of human blastocysts after *in vitro* fertilization[J]. *NPJ Digital Medicine*, 2019, 2: 21.
- [22] Charles C, Gafni A, Whelan T. Shared decision-making in the medical encounter: What does it mean?(or it takes at least two to tango)[J]. *Social Science & Medicine*, 1997, 44(5): 681-692.
- [23] Tarin J J, García-Pérez M A, Hermenegildo C, et al. Changes in sex ratio from fertilization to birth in assisted-reproductive-treatment cycles [J]. *Reproductive Biology and Endocrinology*, 2014, 12: 56.
- [24] Bronet F, Nogales M C, Martínez E, et al. Is there a relationship between time-lapse parameters and embryo sex?[J]. *Fertility and Sterility*, 2015, 103(2): 396-401.e2.
- [25] Wang A G, Kort J, Behr B, et al. Euploidy in relation to blastocyst sex and morphology[J]. *Journal of Assisted Reproduction and Genetics*, 2018, 35(9): 1565-1572.
- [26] Huang B, Ren X L, Zhu L X, et al. Is differences in embryo morphokinetic development significantly associated with human embryo sex? [J]. *Biology of Reproduction*, 2019, 100(3): 618-623.
- [27] 杨一华, 张涛, 王瑶, 等. 辅助生殖技术领域的知情同意[J]. *中国医学伦理学*, 2008, 21(4): 41-43.
Yang Y H, Zhang T, Wang Y, et al. Informed consent in assisted reproductive technology services[J]. *Chinese Medical Ethics*, 2008, 21(4): 41-43.
- [28] Bjerring J C, Busch J. Artificial intelligence and patient-centered decision-making[J]. *Philosophy & Technology*, 2021, 34(2): 349-371.
- [29] Parasuraman R, Manzey D H. Complacency and bias in human use of automation: An attentional integration[J]. *Human Factors*, 2010, 52 (3): 381-410.
- [30] Kramer M F, Schaich Borg J, Conitzer V, et al. When do people want AI to make decisions?[C]//Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society. New Orleans LA USA. ACM, 2018: 204-209.
- [31] Nicholson Price W 2nd, Gerke S, Glenn Cohen I. Potential liability for physicians using artificial intelligence[J]. *JAMA*, 2019, 322(18): 1765-1766.
- [32] Schönberger D. Artificial intelligence in healthcare: A critical analysis of the legal and ethical implications[J]. *International Journal of Law and Information Technology*, 2019, 27(2): 171-203.
- [33] Buchanan A B D W. From Chance to Choice[M]. Cambridge, UK: Cambridge University Press, 2000.
- [34] Hofmann B. 'you are inferior!' revisiting the expressivist argument[J]. *Bioethics*, 2017, 31(7): 505-514.
- [35] Voter A F, Meram E, Garrett J W, et al. Diagnostic accuracy and failure mode analysis of a deep learning algorithm for the detection of intracranial hemorrhage[J]. *Journal of the American College of Radiology*, 2021, 18(8): 1143-1152.
- [36] O'Connor M. Algorithm's "unexpected" weakness raises larger concerns about AI's potential in broader populations[J]. Retrieved April, 2021, 29: 2021.
- [37] 中国政府网. 关于印发国家健康医疗大数据标准、安全和服务管理办法(试行)的通知[EB/OL]. (2018-09-14)[2025-02-02]. <https://www.nhc.gov.cn/mohwsbwstjxxzx/s8553/201809/742308fad8ed47fb85b9675ee4b6c6eb.shtml>.
www.gov.cn. Notice on the issuance of the national health medical big data standards, security, and service management measures (Trial) [EB/OL]. (2018-09-14)[2025-02-02]. <https://www.nhc.gov.cn/mohwsbwstjxxzx/s8553/201809/742308fad8ed47fb85b9675ee4b6c6eb.shtml>.
- [38] 国家市场监督管理总局 国家标准化管理委员会. 信息安全技术 个人信息安全规范: GB/T 35273—2020[S]. 北京: 中国标准出版社, 2020.
Standardization Administration of the People's Republic of China. Information security technology—Personal information security specification: GB/T 35273—2020[S]. Beijing: Standards Press of China, 2020.
- [39] 全国人民代表大会常务委员会. 中华人民共和国数据安全法[S/OL]. (2021-06-10) [2025-10-30]. http://www.npc.gov.cn/npc/c2/c30834/202106/t20210610_311888.html.
The Standing Committee of the National People's Congress. Data security law of the People's Republic of China[S/OL]. (2021-06-10) [2025-10-30]. http://www.npc.gov.cn/npc/c2/c30834/202106/t20210610_311888.html.
- [40] 王炎秋, 相俊, 王铂源, 等. 人工智能在辅助生殖中的应用及伦理规范[J]. *科技视界*, 2024(12): 14-16.
Wang Y Q, Xiang J, Wang B Y, et al. Application of artificial intelligence in assisted reproduction and its ethical norms[J]. *Science & Technology Vision*, 2024(12): 14-16.
- [41] 中国政府网. 中共中央办公厅 国务院办公厅 关于加强科技伦理治理的意见[EB/OL]. (2022-03-20)[2025-03-30]. <https://www.gov.cn/>

zhengce/202212/content_6688212.htm.

www.gov.cn. The General Office of the Central Committee of the Communist Party of China and the General Office of the State Council issued the "Opinions on strengthening the ethical governance of science and technology"[EB/OL]. (2022-03-20)[2025-03-30]. https://www.gov.cn/zhengce/202212/content_6688212.htm.

[42] Savulescu J. Procreative beneficence: Why we should select the best children[J]. *Bioethics*, 2001, 15(5-6): 413-426.

[43] 黎燚, 彭飞荣. 论医疗人工智能伦理标准体系的构建[J]. *标准科学*, 2024(7): 70-75.

Li Y, Peng F R. On the construction of the ethical standards system for medical Artificial Intelligence[J]. *Standard Science*, 2024(7): 70-75.

[44] Buda M, Saha A, Walsh R, et al. A data set and deep learning algorithm for the detection of masses and architectural distortions in digital breast tomosynthesis images[J]. *JAMA Network Open*, 2021, 4(8): e2119100.

[45] Chen C, Li O, Tao D, et al. This looks like that: Deep learning for interpretable image recognition[J]. *Advances in neural information processing systems*, 2019, 32: 8930-8941.

Ethical Challenges and Governance of Artificial Intelligence in Embryo Selection for Human Assisted Reproduction

TANG Jian^{1*}, LINGHU Fuyi¹, MANG Lie²

1. School of Medical Humanities, Tianjin Medical University, Tianjin 300070, China;

2. Assisted Reproduction Center, Tianjin United Family Hospital, Tianjin 300074, China

Highlights

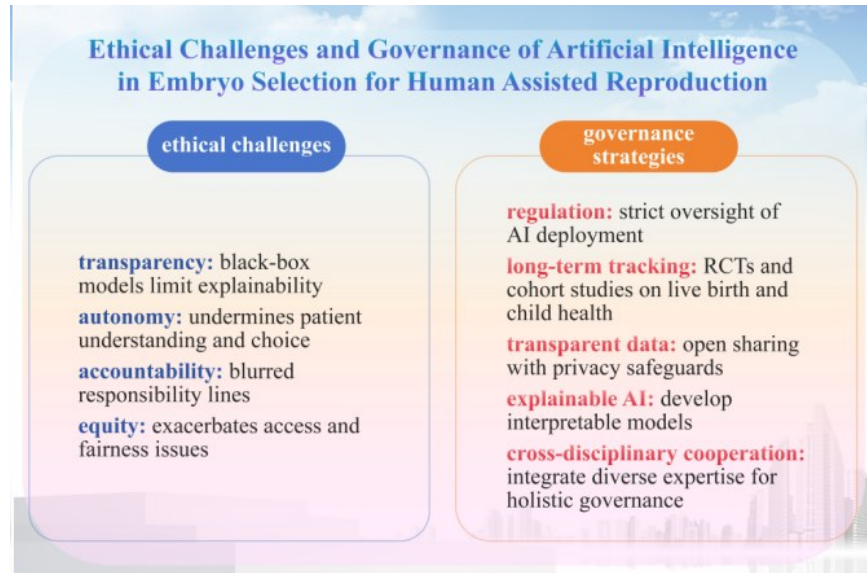
This paper establishes a framework of four core ethical challenges in AI-assisted embryo selection—lack of transparency, compromised autonomy, ambiguous accountability, and equity concerns—providing a structured analytical model.

It innovatively advocates prioritizing explainable AI to enhance transparency, ensure clinical oversight, and strengthen patient trust.

The study emphasizes rigorous validation through randomized trials and long-term cohort studies focused on clinically meaningful outcomes like live birth and child health.

It proposes a dynamic governance framework integrating regulation, interdisciplinary collaboration, secure data sharing, and policy alignment to ensure responsible AI deployment.

Graphical Abstract



Abstract: With the global infertility rate continuing to rise, assisted reproductive technology (ART) has become a crucial solution for overcoming fertility barriers. *In vitro* fertilization and embryo transfer (IVF) is a core component of ART, and embryo selection within this process is a key determinant of pregnancy success. Traditionally, embryologists rely primarily on static morphological assessment criteria (such as the Gardner scoring system) to grade and select embryos. While this method offers advantages such as procedural transparency and ease of explanation to patients, it is inherently limited by its high dependence on the subjective experience of practitioners, resulting in significant inter-observer variability, poor reproducibility,

and limited predictive power. In recent years, the widespread adoption of time-lapse videography(TLV) technology has enabled continuous, non-invasive monitoring of embryonic development, generating vast amounts of image and parameter data. However, human experts struggle to extract all meaningful information from such complex data streams. Against this backdrop, artificial intelligence (AI), particularly deep learning (DL) and computer vision (CV) technologies, due to their exceptional capabilities in processing high-dimensional unstructured data and identifying subtle patterns, have been rapidly introduced into the field of human assisted reproduction to assist or even lead embryo selection decisions. The application of AI is seen as a critical breakthrough for improving IVF success rates and achieving precision reproductive medicine. Yet, every technological innovation brings profound ethical and social transformations. The use of AI in embryo selection not only alters traditional clinical decision-making models but also touches upon fundamental ethical issues concerning the origin of life, individual autonomy, technological accountability, and social equity. Therefore, systematically exploring the ethical challenges posed by AI in embryo selection and constructing corresponding governance frameworks hold significant theoretical value and practical urgency.

This paper offers an in-depth analysis of the ethical challenges posed by artificial intelligence (AI) in human assisted reproductive embryo selection and proposes actionable governance pathways. Using an integrated methodology of literature review and bioethical analysis, it first examines the technical foundations, current advances, and predictive benefits of AI-assisted embryo selection. Building on this, the study critically explores four core ethical dimensions: decision transparency, patient autonomy, responsibility attribution, and social equity. Methodologically, the research begins by classifying AI systems—glass box, matte box, and black box—and analyzing how the unexplainability of deep neural networks in "black box" models undermines interpretability. It then applies core bioethical principles—autonomy, beneficence, non-maleficence, and justice—to assess AI's impact on informed consent, clinician-patient dynamics, accountability mechanisms, and equitable access to care. Finally, drawing on China's regulatory landscape—including the Data Security Law, the Standards, Security, and Service Management Measures for National Health Medical Big Data (Trial), and the Opinions on Strengthening Ethical Governance of Science and Technology—the paper formulates concrete recommendations for a multi-layered, adaptive ethical governance framework.

The study concludes that while AI offers significant technical advantages in embryo selection—particularly through deep learning models that automatically extract subtle, dynamic developmental features from time-lapse videography data, enabling objective, standardized assessments of implantation potential and reducing human bias—it also introduces profound ethical risks. First, the prevailing use of "black box" AI systems severely compromises decision transparency: clinicians and patients cannot understand or explain why a particular embryo was selected or discarded, undermining the traceability, credibility, and regulatory oversight of clinical decisions. Second, this opacity weakens patient autonomy by rendering informed consent largely formalistic; when patients cannot grasp the algorithmic basis of recommendations, and clinicians overly defer to AI outputs, reproductive decision-making may shift toward "algorithmic paternalism", contradicting the principle of shared decision-making. Third, responsibility attribution becomes ambiguous when AI participates in or dominates selection—should implantation fail, complications arise, or offspring health issues emerge, it is unclear how liability should be distributed among clinicians, institutions, and developers, leaving existing accountability frameworks ill-equipped to address harms caused by algorithmic flaws or biased data. Fourth, social equity is at risk: high costs may concentrate AI-enhanced ART among affluent groups, widening access disparities, while algorithms trained on unrepresentative data could perpetuate systemic biases affecting traits like gender or genetics, sending harmful societal signals. Moreover, reducing embryos to algorithmic scores may diminish their perceived intrinsic value as potential lives, raising serious concerns about the dignity of human life.

In conclusion, the application of AI in embryo selection is a double-edged sword, carrying both technological potential and ethical risks. Given the current immaturity of the technology, unknown long-term safety, and lack of uniform evaluation standards, the principle of "cautious advancement and ethics first" must be upheld. A future-oriented ethical governance system should be constructed, centered on strengthening regulation, ensuring transparency, clarifying responsibilities, and promoting fairness. In the short term, priority should be given to developing highly interpretable "transparent box" models, establishing a scientific evaluation system based on randomized controlled trials and long-term birth cohort studies, and promoting limited public access to data and code. In the long run, interdisciplinary collaboration must be deepened, laws and regulations improved, and mechanisms established to ensure that AI technology truly serves human well-being, enabling sustainable and responsible development in the field of assisted reproduction.

Keywords: human assisted reproductive technology; artificial intelligence; human embryo selection; bioethics; ethical governance of technology