

Perspective Article

AI agents as the potential next operational layer in reproductive medicine

Zeev Shoham¹¹ The Hebrew University of Jerusalem, Hadassah Medical School, Jerusalem 9112102, Israel

Keywords: artificial intelligence, AI agents, agentic AI, multi-agent systems, reproductive medicine, in vitro fertilization, clinical workflow, governance

<https://doi.org/10.46989/001c.163012>

Journal of IVF-Worldwide

Vol. 4, Issue 3, 2026

Background

Artificial intelligence in reproductive medicine has so far been the domain of targeted, task-specific systems – tools for embryo assessment, sperm analysis, ovarian stimulation support, and outcome prediction. A separate and fast-growing healthcare literature now distinguishes these conventional tools from agentic systems, which retain context, plan, chain actions together, draw on external tools, and coordinate multi-step work under human supervision.

Objective

This research aims to determine whether agentic AI constitutes the next operational layer in reproductive medicine, distinct from earlier predictive applications, and to outline its plausible near-term role, its main risks, and the research questions it raises.

Approach

This article presents a critical perspective and narrative synthesis drawn from the published reproductive medicine and healthcare literature. It addresses how the agentic AI concept should be defined, how the evidence should be read, how plausible translation actually is, and how such systems should be governed – without pooled quantitative estimates.

Key Observations

Reproductive medicine is an unusually informative medical test case. IVF care is longitudinal and data-rich; it draws on several disciplines; it depends heavily on communication; and it is operationally fragmented. The most plausible near-term role for AI agents is supervised orchestration across four domains: evidence generation and scientific publishing, clinical workflow coordination, patient-facing communication, and bedside decision support – possibly organized as multimodal, multi-agent systems. So far, though, almost no evidence specific to reproductive medicine shows that agentic systems improve live birth, cumulative live birth, safety, or patient-centered outcomes. The wider healthcare evidence is still mostly conceptual work, simulations, and early pilots.

Conclusions

For now, AI agents are best treated as supervised collaborators rather than autonomous clinical actors. Adopting them responsibly will require clear definitions, prospective workflow studies using clinically meaningful endpoints, close attention to human factors and operational realities, transparent governance, and a shared, layered framework for validation.

^a Corresponding Author. Zeev Shoham, MD, The Hebrew University of Jerusalem, Hadassah Medical School, Jerusalem 9112102, Israel, +972 52-253-0370, email: zeev1@mail.huji.ac.il

INTRODUCTION

Artificial intelligence (AI) has become a central theme in reproductive medicine, and especially in in vitro fertilization (IVF) and related assisted reproductive technologies (ART). This is largely because the field produces so much data, both structured and semi-structured: embryo images, time-lapse recordings, hormonal trajectories, laboratory variables, and longitudinal treatment outcomes. Reviews of the literature tend to reach a similar verdict: AI may improve objectivity, standardization, and selected workflow steps, but the current evidence base has clear limits – retrospective designs, limited generalizability, inconsistent endpoints, and too little validation against live birth.^{1–3} The performance and validation limitations of these task-specific predictive tools have been examined in detail elsewhere and are not re-evaluated here.⁴

Until recently, most discussion of AI in reproductive medicine concerned task-specific tools: embryo-scoring algorithms, sperm-classification systems, ovarian-stimulation prediction models, and outcome calculators. Each performs a single task or a targeted cluster of tasks, and its role is advisory rather than operational. A separate healthcare literature now makes a different argument – that medical AI is moving beyond these static predictors toward agentic systems that plan, use external tools, retain context, communicate, and coordinate multi-step work under human supervision.^{5–7} This shift is happening as clinical adoption itself rises. Surveys of international fertility specialists report growing use of AI, alongside persistent concerns about cost, training gaps, and overreliance on the technology.⁸

This article is not a formal systematic review. It is a critical perspective and narrative synthesis, and the question it asks is whether emerging agentic AI systems represent the next operational layer in reproductive medicine. The focus is on defining the concept, reading the evidence honestly, judging the plausibility of the translation, and identifying research priorities – not on pooled quantitative estimates.

WHAT DISTINGUISHES AN AI AGENT FROM CONVENTIONAL AI?

The term “AI agent” is still used loosely, and that looseness is itself part of the scientific problem. Recent healthcare reviews tend to describe agentic systems in terms of a connected set of capabilities. Such a system holds context in memory, plans toward a goal, chains several actions together, reaches for external tools or retrieval systems, corrects or reflects on its own output, communicates across steps, and exercises partial autonomy within bounded, human-defined tasks. Conventional medical AI does something more targeted: it classifies, predicts, or ranks within a fixed input–output frame.^{5–7}

The difference is easier to articulate with an example. An embryo-selection model that predicts implantation potential is a tool. A system that pulls prior cycle data, summarizes ovarian response, drafts a monitoring note, flags missing laboratory values, prepares protocol considerations for the clinician to review, and produces a draft patient

message is something else – closer to an agentic workflow assistant. The distinction matters because a task-specific predictor and a system embedded in multi-step clinical or research workflows should not be held to the same evidentiary standard (Figure 1).^{5,6}

WHY REPRODUCTIVE MEDICINE IS A HIGH-RELEVANCE TEST CASE

Reproductive medicine is a particularly informative setting for testing agentic AI, because the field is not only data-rich but also intensely process-dependent. Fertility care runs in repeated cycles: monitoring, interpretation, counseling, procedural coordination, administrative scheduling, communication between the laboratory and the clinic, and follow-up. A single diagnostic event it is not. IVF care unfolds over time and across several professional roles, which makes it a realistic proving ground for workflow-support systems that must maintain context and coordinate tasks.^{1,2}

Yet this same field is a high-risk place for premature deployment. Its decisions concern embryos and gametes, the timing of treatment, costly protocols, and emotionally charged choices made by patients. A single error in summarization, communication, or workflow execution can propagate through several later steps, even when the original recommendation looked minor. Reproductive medicine may therefore be well suited to studying agents and badly suited to running them unsupervised.^{9,10}

There is one more reason the field matters here. Almost all of the reproductive AI evidence available today concerns task-specific tools rather than validated agents. It is worth being blunt about this gap. The problem is not that agentic systems have produced mixed results; rather, agentic systems developed for reproductive medicine have barely been evaluated prospectively.^{2,7}

MULTIMODAL AND MULTI-AGENT SYSTEMS

Two architectural features of current agentic AI deserve particular attention in reproductive medicine: multimodality and multi-agent organization. Much of the healthcare literature is moving away from single-input models toward systems that pull heterogeneous data – images, video, structured measurements, free text, and conversational input – into a single contextual representation.^{5,6} Reproductive medicine is multimodal by nature. One treatment cycle alone generates embryo images, time-lapse recordings, hormonal trajectories, ultrasound findings, laboratory values, clinician notes, and patient messages. A genuinely useful agent would not read any of these streams in isolation; it would assemble them into a coherent, longitudinal picture of the cycle. This is also where the weaknesses are sharpest. An error in fusing modalities – aligning an ultrasound finding with the wrong cycle day, say – can be far harder to catch than an error in a single-modality predictor.³

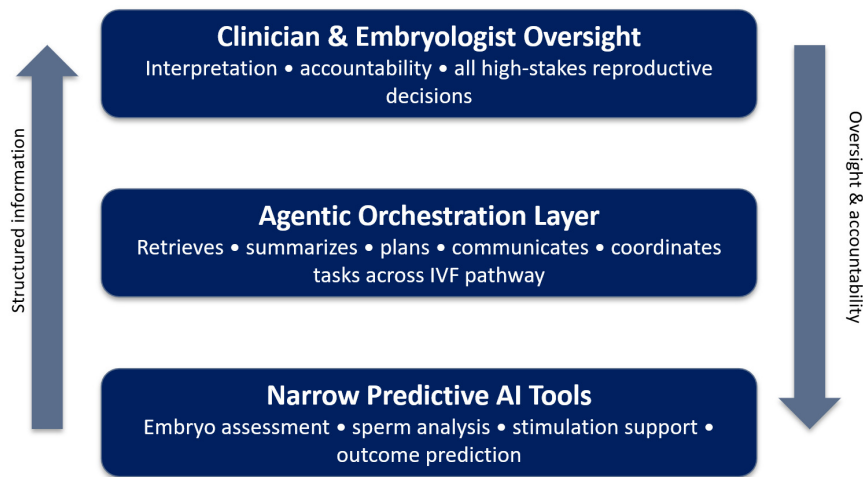


Figure 1. The three-layer architecture of AI in reproductive medicine – from tools to agents to oversight.

The base layer comprises established specific predictive tools, such as those used for embryo assessment, sperm analysis, ovarian stimulation support, and outcome prediction. The middle layer is an emerging agentic orchestration layer that retrieves, summarizes, plans, communicates, and coordinates tasks across the IVF pathway. The top layer is clinician and embryologist oversight, which keeps responsibility for interpretation and for every high-stakes decision. Structured information flows upward; oversight and accountability flow downward. Agentic systems function alongside existing tools and stay subordinate to professional judgment rather than replacing it.

The second feature is multi-agent organization. Instead of a single system trying to do everything, recent work describes architectures in which several specialized agents collaborate under coordination and human supervision.^{5,7} It is not hard to picture this in a fertility unit: a laboratory agent that tracks embryo development and flags anomalies, a scheduling agent that manages cycle timing, a patient-communication agent that handles instructions and questions, a literature agent that retrieves relevant evidence, and a protocol-checking agent that tests a proposed plan against guidelines. The four agentic AI application domains discussed below map onto exactly these kinds of specialized roles. Because each agent has a defined scope, multi-agent designs can improve modularity and auditability. They also open new failure modes – agents that miscommunicate, an early error that propagates down the chain, and responsibility that diffuses when no single component is clearly accountable.^{6,7} For reproductive medicine, the lesson is that multi-agent systems must be evaluated as a whole – not merely as a set of individually competent parts.

DOMAIN 1: EVIDENCE GENERATION AND SCIENTIFIC PUBLISHING

The most immediate domain may be evidence generation itself. AI systems already run through much of the literature pipeline, helping with searches, screening abstracts, organizing studies, supporting drafting, summarizing, and revising manuscripts. In the broader health-research literature, this work is increasingly described as a structured or semi-agentic workflow rather than a set of isolated prompts.^{5,6}

This domain matters especially in reproductive medicine, where the literature is growing fast, varies widely in method, and is prone to overreading surrogate endpoints. An AI agent that speeds up evidence handling can help, but

only if it improves organization without eroding factual accuracy. This is not a hypothetical worry. A review or perspective that discusses fabricated citations discredits itself the moment its own references prove unreliable. AI-assisted evidence generation in reproductive medicine must therefore rest on explicit human verification: of every reference, every factual claim, and every interpretive conclusion.^{2,5}

This has a concrete implication for publishing. Journals in reproductive medicine should make AI-use transparency statements a routine requirement, particularly for reviews and perspectives; they should draw a clear line between AI-assisted language support, AI-assisted evidence handling, and AI-generated scientific claims. This is more than an editorial nicety. It is part of how trust in the reproductive-medicine literature will hold up as agentic drafting tools become ordinary.^{6,7}

Where Domain 1 operates primarily within the research and publishing environment, the remaining three domains move progressively closer to direct patient care.

DOMAIN 2: CLINICAL WORKFLOW COORDINATION

The strongest near-term case for AI agents in reproductive medicine is probably workflow coordination, not autonomous diagnosis or treatment selection. Fertility care is full of tasks that are fragmented yet repetitive: summarizing charts, assembling cycle timelines, reconciling ultrasound and laboratory data, drafting notes, preparing protocols for review, supporting scheduling, and passing structured messages between clinic and laboratory teams. These are exactly the tasks agentic systems are built to support.^{6,7}

A few concrete examples help ground this. An agent might track ovarian stimulation throughout a cycle, prepare cycle summaries before morning rounds, support em-

bryo-transfer scheduling, triage semen-analysis workflows, help document recurrent-implantation-failure follow-up, or pull together prior-cycle responses ahead of clinician review. None of this yet amounts to evidence of validated agentic deployment. It does, however, show where the field is likely to head.^{2,3}

One principle follows from all this. The more a system coordinates steps across time and across roles, the more it should be judged as a workflow intervention rather than as a predictive model. Future studies should not stop at confirming technical accuracy. They should also look at clinician workload, cycle efficiency, task completion, communication failures, system escalation frequency, the depth of users' trust in it, safety events, and, where appropriate, downstream reproductive outcomes.^{6,11}

DOMAIN 3: PATIENT-FACING COMMUNICATION

Fertility care relies heavily on communication, which makes patient-facing AI one of the more plausible homes for agentic systems. Patients have to navigate uncertainty, repeated monitoring, complicated scheduling, lifestyle instructions, changes in medication, and decision points that are emotionally hard. An AI agent might reinforce instructions, prepare visit summaries, answer logistical questions, help triage symptoms, or help a patient frame questions for the next consultation.^{6,10}

Patient trust, though, remains the central constraint. Survey data from fertility care show that familiarity with AI does not, by itself, produce trust in AI-informed reproductive care; patients still prefer that treatment recommendations come from a physician. That is a strong hint that patient-facing AI should be presented as a layer for continuity and communication, not as a stand-in for clinical judgment.¹⁰

This domain also raises ethical questions of its own. What does informed consent mean when communication is AI-mediated? Where does information end and advice begin? How is ambiguity handled, and how do you keep patients from granting fluent output more authority than it has earned? In a field where timing, emotion, and expectation are tightly bound together, the accuracy of communication is not a soft endpoint – it is part of clinical safety.^{9,10}

A further concern, and one that has been studied far less, is psychological. Infertility care is long, emotionally intense, and marked by persistent uncertainty, and patients in that situation often look for reassurance. A conversational agent that is always available and unfailingly responsive could, for some patients, become something they lean on emotionally; its fluent, confident tone may invite anthropomorphism or an inflated sense of its authority.^{9,10} The very features that make such an agent supportive – its availability, its patience, its personalization – are also the ones that can breed dependency or quiet algorithmic persuasion. The evidence here remains thin, so these issues should be read as hypotheses rather than established effects. Even so, given how vulnerable this group of patients can be, the psychological effects of sustained interaction

with conversational agents warrant direct study rather than being assumed away.

DOMAIN 4: BEDSIDE DECISION SUPPORT

The most sensitive and contested application is bedside, or point-of-care, decision support. In principle, an AI agent could integrate prior cycles, patient age, ovarian reserve, laboratory results, embryo-development data, earlier stimulation response, and patient preferences into a structured clinical briefing and could help draft notes, complete checklists, and lay out options for the clinician to review. This is also, however, where AI hallucination, poor calibration, context drift, and probabilistic inconsistency carry the heaviest consequences.^{5,7}

It is worth being explicit here about what large language models cannot reliably do. Current generative systems can produce fluent but incorrect output, invent references, mishandle edge cases, lose fidelity across long context windows, and represent uncertainty poorly. They are probabilistic pattern-generators, not causal reasoners. None of these issues rules out bedside support, but they do mean that any bedside deployment in reproductive medicine must remain tightly bound, transparent, auditable, and under human supervision.^{5,6}

Put plainly, an agent should not change gonadotropin dosing on its own, settle embryo-transfer strategy, or take other high-stakes clinical actions without explicit clinician authorization. Its nearer-term role is preparatory and assistive – organizing the relevant information so that clinical judgment is better supported, not displaced.^{2,3}

HUMAN FACTORS: OVER-TRUST, DESKILLING, AND COMPLACENCY

The risks of agentic AI do not lie only in the technology. They also lie in how clinicians and embryologists interact with it. Automation bias – the tendency to over-rely on automated output and to scrutinize it too little – is well documented in clinical decision-support systems, and it is shaped by trust, workload, time pressure, and task complexity.¹² Agentic systems may increase this risk: fluent, well-formatted output can suggest a reliability the system has not actually earned. The practical result is “rubber-stamping,” where an AI-generated summary or recommendation is signed off with little independent review – especially when staff are pressed for time.

Reproductive medicine is particularly exposed to these effects. Embryologists who come to depend on AI triage of gametes and embryos may, over time, rely less and less on their own morphological judgment – a slow form of deskilling. Clinicians who receive agent-prepared cycle summaries may anchor on them and fail to question what has been left out or framed misleadingly. Alert fatigue and cognitive offloading are further worries: they flag too much, and the system is simply ignored, or they are trusted too readily and quietly take the place of expert reasoning. None of this is an argument against agentic support. It is an argument for measuring the human-factor outcomes directly

– override frequency, appropriate reliance, and the preservation of professional skill – rather than assuming them.⁷

ETHICS, REGULATION, AND ACCOUNTABILITY

The ethical implications of AI in reproductive medicine have been widely acknowledged, but they grow more complex once a system moves from prediction toward agency. The familiar concerns – bias, privacy, explainability, data governance, and inequity – become more pronounced. Agentic systems add others: delegated actions, errors that propagate across multiple steps, internal planning chains that may be hard to audit, responsibility that is ambiguous, and legal liability that is unclear when outputs shape care.^{6, 7,9}

The regulatory implications follow directly. A task-specific embryo-ranking model and a workflow agent that drafts clinical communication or prepares treatment plans do not necessarily fall under the same regulatory framework. As agentic systems grow more dynamic, audit trails, post-deployment surveillance, version control, cybersecurity protection, explicit escalation rules, and human override all matter much more.^{7,11}

One regulatory difficulty stands out: the gap between fixed and adaptive (agentic) systems. Conventional Software as a Medical Device (SaMD) frameworks were primarily designed for locked algorithms whose behavior remains stable after approval. Many agentic systems are different. They rely on large language models that may be continuously updated, fine-tuned, or augmented with new retrieval sources, which makes version control, post-market surveillance, and revalidation genuinely difficult. The Food and Drug Administration (FDA) and European Medicines Agency (EMA) are actively drafting guidance for adaptive AI, but specific pathways for multi-agent reproductive systems remain underdeveloped, and the regulation of adaptive algorithms in reproduction remains unsettled.^{7,13} Until clearer standards are defined, agentic systems used in reproductive care should be version-locked during clinical use and formally revalidated after any material update.

Reproductive medicine carries an added ethical sensitivity, because bias can intersect with age, socioeconomic status, ethnicity, access to treatment, and the commercial interests attached to fertility technology. A system trained on limited datasets from well-resourced clinics may behave quite differently elsewhere. Equity, then, is not an optional consideration; it is a core translational criterion.^{2,9}

OPERATIONAL REALITIES AND EQUITY

Beyond ethics and regulation, the uptake of agentic AI will depend on practical matters that conceptual discussion tends to skip. An orchestration layer is only as good as its connections. It has to interoperate with electronic health records, laboratory and imaging systems and existing time-lapse or andrology platforms – and fragmented or proprietary data environments can make that integration difficult. Running such a system also costs money on an ongoing basis: infrastructure, security, staff training, and

maintenance. Leaning on a single commercial platform carries its own hazards, including vendor lock-in. These burdens fall unevenly. Large, well-resourced IVF centers are far better placed to adopt and sustain agentic systems than smaller units, so the technology could just as easily widen existing disparities in care as narrow them. Broader cost-effectiveness questions for AI in assisted reproduction have been examined elsewhere⁴; the point here is a more specific one – that workflow agents add an integration and orchestration burden, and that burden should be weighed honestly before anyone judges how realistic adoption really is.¹¹ Early health-economics work on AI in assisted reproduction suggests that integration and maintenance costs can be substantial, especially for smaller clinics still running legacy systems. These practical barriers add weight to the same worry: agentic technologies could widen, rather than reduce, the gap in access to high-quality fertility care. The framework that follows is intended to organize both the opportunities and the constraints across the four domains in which agentic systems are most likely to emerge.

A TRANSLATIONAL FRAMEWORK FOR AGENTIC AI IN REPRODUCTIVE MEDICINE

As shown in [Figure 1](#), the agentic AI layer sits in the middle of the three-layer AI architecture. One practical way to organize the agentic layer is by application domain rather than by technology type. As described above, four domains for agentic systems have been identified in the broader healthcare literature.^{6,7} Each represents a distinct horizontal application area within the agentic layer, ranging from evidence generation and scientific publishing (which operates in the research environment rather than the clinical pathway) through clinical workflow coordination and patient-facing communication, to bedside decision support at the most sensitive end of the spectrum. [Table 1](#) summarizes these four domains, their potential applications, current evidence maturity, major risks, and recommended deployment status.

This framework also points to a research agenda. Prospective studies should compare agent-supported workflows against standard ones using outcomes that genuinely matter – clinician time burden, the efficiency of message responses, reduction in protocol error, the quality of communication, patient comprehension, treatment burden, workflow interruptions, and, where appropriate, selected reproductive endpoints. Convenience or speed, on their own, are not evidence of clinical value.^{7,11}

A PROPOSED VALIDATION FRAMEWORK FOR AGENTIC AI IN REPRODUCTIVE MEDICINE

If agentic systems are to be evaluated seriously, the field first needs a shared vocabulary for what “working” even means. Technical accuracy on its own is not enough: an agent can be accurate in isolation and still be unhelpful, unsafe, or badly used in practice. [Table 2](#) defines a layered validation framework meant to make evaluation explicit

Table 1. Domains of agentic AI in reproductive medicine: plausible applications, current evidence maturity, principal risks, and recommended level of deployment.

Domain	Plausible near-term applications	Current evidence maturity	Principal risks	Recommended deployment level
Evidence generation and scientific publishing	Literature search, abstract screening, evidence organization, drafting support, summarization, reference management	Emerging: in active use but with limited formal evaluation	Reference fabrication, factual drift, undisclosed AI authorship, and overinterpretation of surrogate endpoints	Supervised use with mandatory human verification and explicit AI-use disclosure
Clinical workflow coordination	Cycle-summary preparation, chart and timeline assembly, laboratory-clinic reconciliation, note drafting, scheduling support	Early: largely conceptual or piloted, with little IVF-specific validation	Error propagation across steps, automation bias, and workflow disruption	Prospective workflow trials before routine adoption; clinician sign-off
Patient-facing communication	Visit summaries, reinforcement of instructions, logistical questions, question preparation, symptom triage	Early: patient-trust data available, but agentic deployment largely unstudied	Misplaced authority of fluent output, consent ambiguity, and emotional dependency	Cautious piloting with explicit boundaries and escalation rules
Bedside decision support	Structured clinical briefings, synthesis of cycle history, checklist and option preparation for clinician review	Minimal: essentially research-only in reproductive medicine	AI hallucination, poor calibration, context drift, unclear accountability	Research-only; no autonomous high-stakes action

Table 2. A proposed layered framework for validating agentic AI in reproductive medicine. The layers are cumulative – satisfactory performance at a lower layer does not guarantee benefit at a higher one.

Evaluation layer	Question addressed	Representative metrics
1. Technical performance	Is the agent's output accurate, reliable, and reproducible?	Task accuracy; hallucination or fabrication rate; calibration of stated confidence; consistency of output across repeated runs
2. Workflow integration	Does the agent improve how work is done without creating new failure modes?	Time saved; task-completion rates; escalation frequency; workflow interruptions; and communication errors
3. Human factors	Do clinicians and embryologists rely on the agent appropriately?	Clinician trust; override and acceptance rates; indicators of automation bias; evidence of deskilling or over-reliance
4. Patient-centered	Do patients understand and benefit from agent-mediated communication?	Comprehension; satisfaction; decisional confidence; trust; perceived support and emotional impact
5. Clinical safety	Does the agent reduce, rather than introduce, clinical error?	Protocol-error reduction; adverse and near-miss events; incidents of error propagation across steps
6. Reproductive outcomes	Does the agent ultimately improve outcomes that matter to patients?	Live birth rate; cumulative live birth; time to pregnancy; and treatment burden

and comparable across studies. It has six layers, running from low to high – from technical performance to reproductive outcomes. Each layer asks a different question and calls for different metrics. The layers are cumulative, not alternatives. A system can perform well technically and still fail at the human-factors or safety layer – and that failure is exactly what isolated accuracy reporting tends to miss. The framework is deliberately general; specific metrics and thresholds will need to be defined for each application and agreed upon within the field.^{6,7,12}

No single study could cover all six layers, and early work will sensibly concentrate on the lower-level ones. What the framework adds is clarity: it makes explicit which layer a given study actually tests, and it pushes back against the common leap from strong technical performance to assumed clinical or reproductive benefit. Seen this way, vali-

ation becomes a staged program rather than a single endpoint.

DISCUSSION

The argument here is not that reproductive medicine has already entered an era of clinically proven AI agents. It plainly has not. The more defensible claim is more specific, yet forward-thinking: reproductive medicine is an unusually relevant test case for the next phase of AI in medicine, because it brings together data-rich prediction problems, longitudinal workflow complexity, and a heavy communication burden. The current evidence supports cautious conceptual interest – not operational certainty.^{2,6,7}

Two opposing errors are worth avoiding. One is hype – assuming that because an agent can sound coherent, it is ready for autonomous reproductive care. The other is

conceptual inertia – treating every AI application as if it were still a simple classifier. The real opportunity lies in between: recognizing that workflow-support systems may need their own study designs, regulatory pathways, and evaluation metrics. The layered framework proposed above is one attempt to make those metrics explicit.^{5,11} Prospective workflow trials deserve particular care in their design; they should incorporate the layered validation framework from [Table 2](#) to ensure that technical performance is never mistaken for clinical or patient-centered benefit.

A last implication concerns the literature itself. If reproductive medicine is going to discuss AI agents seriously, it also has to model the behavior it expects – disciplined citation, openness about where AI assisted, and rigorous checking of evidence. This matters most for perspective articles, where fluent prose can paper over unsupported claims unless the references behind it are held to the same standard as the claims they support.^{5,6}

As of mid-2026, prospective evaluations of agentic systems built specifically for reproductive medicine remain very scarce. Task-specific predictive tools have built up a modest evidence base. Still, no large multicenter trial has yet tested a multi-agent system against clinically meaningful endpoints – live birth rate, cumulative live birth, treatment burden, or safety events – in a real-world IVF pathway. What the literature offers instead is conceptual frameworks, simulation studies, and early single-center pilots. The need for rigorous translational research before any widespread adoption is hard to overstate.

CONCLUSION

AI agents may yet become the next operational layer in reproductive medicine. For now, though, the literature supports this only as a cautious translational hypothesis – not as an established clinical reality. Their most credible near-term role is as supervised collaborators: handling evidence, coordinating workflow, supporting patient communication, and offering bounded bedside support around the predictive tools already in place, perhaps organized as multi-modal, multi-agent systems.^{2,6}

For now, there is simply not enough reproductive-medicine-specific evidence to justify autonomous deployment in high-stakes care. Progress will require several factors at once: clearer definitions, prospective workflow studies, better worked examples from reproductive medicine, formal governance frameworks, real attention to human factors and operational realities, stronger auditability, and a shared, layered approach to validation – one anchored in the outcomes that genuinely matter to patients, among them safety, reduced treatment burden, trust, and, where relevant, live birth.^{7,9,10}

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

Artificial intelligence tools were used in preparing this manuscript, assisting with language refinement, structural editing, and reference formatting. The author reviewed and verified every scientific claim, interpretation, and reference against the original sources and takes full responsibility for the content of the manuscript.

FUNDING STATEMENT

No specific funding was received for this work.

CONFLICT OF INTEREST STATEMENT

Zeev Shoham is the Editor-in-Chief of the Journal of IVF-Worldwide. No other competing interests were disclosed.

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Conceptualization: Zeev Shoham. Writing – original draft: Zeev Shoham. Writing – review and editing: Zeev Shoham. Supervision: Zeev Shoham.

ATTESTATION STATEMENTS

- The subjects in this trial have not concomitantly been involved in other randomized trials: Not applicable.
- Data related to any of the subjects in the study has not been published previously: Not applicable.
- All study and manuscript data will be made available to the journal editors upon request before and/or after manuscript publication for review or query.
- The authors followed the appropriate checklist for this study design: Not applicable (narrative synthesis and critical perspective).
- Ethics approval was not required for this study as it is a perspective article based on previously published literature.

DATA SHARING STATEMENT

No original datasets were generated or analyzed in this narrative synthesis. All referenced data are available in the cited published literature.

Submitted: May 28, 2026 CDT. Revised: June 05, 2026 CDT. Accepted: June 07, 2026 CDT. Published: July 07, 2026 CDT.



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

REFERENCES

1. Wang R, Pan W, Jin L, Li Y, Geng Y, Gao C, et al. Artificial intelligence in reproductive medicine. *Reproduction*. 2019;158(4):R139-R154. doi:[10.1530/REP-18-0523](https://doi.org/10.1530/REP-18-0523)
2. Kakkar P, Gupta S, Paschopoulou KI, Paschopoulos I, Paschopoulos I, Siafaka V, et al. The integration of artificial intelligence in assisted reproduction: a comprehensive review. *Front Reprod Health*. 2025;7:1520919. doi:[10.3389/frph.2025.1520919](https://doi.org/10.3389/frph.2025.1520919)
3. Hanassab S, Abbara A, Yeung AC, Voliotis M, Tsaneva-Atanasova K, Kelsey TW, et al. The prospect of artificial intelligence to personalize assisted reproductive technology. *NPJ Digit Med*. 2024;7(1):55. doi:[10.1038/s41746-024-01006-x](https://doi.org/10.1038/s41746-024-01006-x)
4. Shoham Z. Artificial intelligence in reproductive medicine: transforming assisted reproductive technologies. *Journal of IVF-Worldwide*. 2025;3(2):1-8. doi:[10.46989/001c.137620](https://doi.org/10.46989/001c.137620)
5. Liu F, Niu Y, Zhang Q, Wang K, Dong Z, Wong IN, et al. A foundational architecture for AI agents in healthcare. *Cell Rep Med*. 2025;6(10):102374. doi:[10.1016/j.xcrm.2025.102374](https://doi.org/10.1016/j.xcrm.2025.102374)
6. Njei B, Al-Ajlouni YA, Kanmounye US, Boateng S, Nguefang GL, Njei N, et al. Artificial intelligence agents in healthcare research: a scoping review. *PLoS One*. 2026;21(2):e0342182. doi:[10.1371/journal.pone.0342182](https://doi.org/10.1371/journal.pone.0342182)
7. Collaco BG, Haider SA, Prabha S, Gomez-Cabello CA, Genovese A, Wood NG, et al. The role of agentic artificial intelligence in healthcare: a scoping review. *NPJ Digit Med*. 2026;9(1):345. doi:[10.1038/s41746-026-02517-5](https://doi.org/10.1038/s41746-026-02517-5)
8. Shoham G, Alexandroni H, Weissman A, Mizrahi Y. Global trends in the use of artificial intelligence (AI) in reproductive medicine: insights from surveys of international fertility specialists. *Journal of IVF-Worldwide*. 2025;3(3):33-44. doi:[10.46989/001c.140673](https://doi.org/10.46989/001c.140673)
9. Koplin JJ, Johnston M, Webb ANS, Whittaker A, Mills C. Ethics of artificial intelligence in embryo assessment: mapping the terrain. *Hum Reprod*. 2025;40(2):179-185. doi:[10.1093/humrep/deae264](https://doi.org/10.1093/humrep/deae264)
10. Cromack SC, Lew AM, Bazzetta SE, Xu S, Walter JR. The perception of artificial intelligence and infertility care among patients undergoing fertility treatment. *J Assist Reprod Genet*. 2025;42(3):855-863. doi:[10.1007/s10815-024-03382-5](https://doi.org/10.1007/s10815-024-03382-5)
11. Sharma M, Savage C, Nair M, Larsson I, Svedberg P, Nygren JM. Artificial intelligence applications in health care practice: scoping review. *J Med Internet Res*. 2022;24(10):e40238. doi:[10.2196/40238](https://doi.org/10.2196/40238)
12. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. *J Am Med Inform Assoc*. 2012;19(1):121-127. doi:[10.1136/amiajnl-2011-000089](https://doi.org/10.1136/amiajnl-2011-000089)
13. Curchoe CL. Unlock the algorithms: regulation of adaptive algorithms in reproduction. *Fertil Steril*. 2023;120(1):38-43. doi:[10.1016/j.fertnstert.2023.05.152](https://doi.org/10.1016/j.fertnstert.2023.05.152)