

Ethical Implementation of Artificial Intelligence to Select Embryos in *In Vitro Fertilization* (IVF)

Michael Anis Mihdi Afnan¹, Cynthia Rudin², Vincent Conitzer^{3,4}, Julian Savulescu^{5,6,7}, Abhishek Mishra⁵, Yanhe Liu^{8,9,10}, Masoud Afnan¹¹

1. Department of Medicine, Imperial College London, London, UK 2. Departments of Computer Science, Electrical Engineering and Statistical Science, Duke University, Durham, USA 3. Departments of Computer Science, Economics and Philosophy, Duke University, Durham, USA 4. Institute for Ethics in AI and Departments of Computer Science and Philosophy, University of Oxford, Oxford, UK 5. Uehiro Centre for Practical Ethics, University of Oxford, Oxford, UK 6. Wellcome Centre for Ethics and Humanities, University of Oxford, Oxford, UK 7. Murdoch Children's Research Institute, Royal Children's Hospital, Parkville, Australia 8. Monash IVF Group, Southport, Australia 9. School of Human Sciences, University of Western Australia, Crawley, Australia 10. School of Medical and Health Sciences, Edith Cowan University, Joondalup, Australia 11. Qingdao United Family Hospital, Qingdao, China. masoudafnan@me.com

What is IVF and how can AI help?

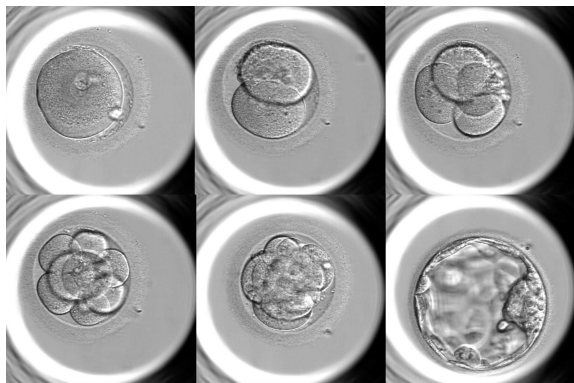


Figure 1: Embryo development from a few hours after fertilisation with 2 pronuclei (top-left) to the "blastocyst" with trophectoderm and inner cell mass (bottom-right)

- IVF is a widely used infertility treatment – over 9 million babies have been born through IVF
- Involves fertilizing eggs with sperm in a lab
- Embryologist selects the best-looking embryo under a microscope for transfer to the uterus to begin a pregnancy. Only ~25% go on to reach live birth
- To maximise the chance of live birth, much attention has been given to new techniques for selecting the embryo most likely to implant
- AI models could leverage static image or time-lapse data to identify important patterns that humans miss and therefore be more accurate
- Algorithms could be more objective than humans

Concerns of using black-box AI to select embryos

Ethical Concerns:

- There are no published Randomised Controlled Trials (RCTs) demonstrating such a tool's safety and effectiveness
- Black-box models would override clinical decision-making (machine paternalism)
- A model may select for certain characteristics (such as male or female sex) unbeknownst to the patient. This selection may not align with patient values.
- Use of a biased algorithm (e.g., biased towards a certain race or sex) could have big societal implications
- It is unclear who would be responsible if a black-box algorithm makes poor selection choices

Epistemic Concerns:

- Black-box models create information asymmetries and are hard to trust
- Confounders are easy to miss in evaluation and reduce an algorithm's generalizability
- Unfavorable economics – a model's ability may only be guaranteed when treatment conditions perfectly match experimental conditions. This could force IVF clinics to purchase specific equipment, perhaps from one supplier, to guarantee success
- Real-time error checking and troubleshooting is difficult if you don't understand a model's reasoning process

Why interpretable AI would be better

- More easily and quickly identify biases
- An aid to clinical decision making (instead of a replacement)
- Responsibility for the decision remains with the clinician
- Combined human + machine performance could be evaluated in a combined arm of a Randomised Controlled Trial
- Confounders and/or an erroneous reasoning process would become apparent
- Clinicians could modulate their interpretation of an algorithm's recommendation under different conditions leaving IVF clinics less susceptible to economic exploitation

Our recommendations

- Replicable, interpretable machine learning tools and data
- Well designed and conducted RCTs
- Post implementation surveillance
- Regulatory oversight requiring interpretable AI when possible
- Funding for public institutions to transparently develop and evaluate AI models with open access to code
- Procedures for maintaining security of patient/embryo data whilst permitting ethical data sharing
- Fully informed consent to use AI
- Include patient values into AI programs where possible
- Participation from the broader AI community