CENTER FOR THE GOVERNANCE OF CHANGE

CASE STUDY 4: AI IN THE FERTILITY CLINIC

Stevan Cirkovic, Human Fertilisation and Embryology Authority
Dina Halai, Human Fertilisation and Embryology Authority

Innovation, Sustainability, and the Future of Healthcare

Across fertility clinics in Europe, artificial intelligence (AI) has sparked hopes that it could standardize processes and improve outcomes. In this case study we explore how AI has already started to disrupt practice in the fertility lab, review the evidence supporting it, and look at how regulators have responded to challenges. Whatever advantages AI will bring to the fertility sector, the real yardstick for patients and regulators is whether using algorithms increases live birth rates versus standard IVF or not. Yet, often predictive abilities are confused for clinical benefit.

CONVENTIONAL EMBRYO SELECTION

During in-vitro fertilization (IVF), sperm and eggs are mixed in the lab to create embryos. Embryologists are tasked with finding the single most viable embryo to ensure the best chance of success while minimizing the risk of a multiple birth [1], [2], [3]. Traditionally, embryologists remove the embryo from the incubator each day and morphologically assess its development under a microscope. Embryo selection in this way is time-consuming, and there is room for human error and variation between individual practitioners [4], [5]. In recent years, time-lapse imaging (TLI) has been introduced, offering embryologists a continuous view of development without disturbing the embryo by taking hundreds of images [6]. This has not been shown conclusively to make a difference to live birth rates but has created vast amounts of data [7], [8], [9].

AI-ASSISTED EMBRYO SELECTION

The success of AI in radiology, the centrality of imaging to embryo selection, and hopes to increase its consistency with the help of algorithms have created the conditions for AI innovation in the fertility sector.

As early adopters of TLI, private fertility clinic groups across Europe are now investing in AI to train algorithms with TLI imaging data to grade embryos. As an example, the Spanish IVI group, which is also active in Germany, France, Italy and the UK, is partnering with academics to use deep learning for embryo selection. By contrast, UK-based LifeWhisperer claims to only require ordinary microscopic images to be dragged and dropped into an online tool and may thus become accessible more widely [10], [11].

As Table 1 illustrates, AI-powered embryo software is typically a deep learning algorithm using a variety of inputs to produce a single success metric. The algorithms have usually been trained to predict outcomes based on a dataset of embryo images. The Early Embryo Viability Assessment (Eeva) test, for example, was developed using data from 373 women and 3328 embryos and validated on an independent data set for consistency of grading results [12]. Table 1 also shows manufacturers of TLI equipment like Merck Serono and Vitrolife being active in the AI-software space too. There is a variety of software-hardware integrations, ranging from the Eeva test being directly sold with a TLI incubator to LifeWhisperer which does not require TLI at all.

Table 1. The variety of algorithmic embryo selection software in use and development

	Name of AI application	Machine learning method (type of AI)	Input data & parameters	Success metric & output	Developer / distributor	Available with TLI incubator?	CE marked?
In clinical use	Early Embryo Viability Assessment (Eeva) test	Xtend algorithm (deep learning, but this is not confirmed) [12] [13]	74 parametres including cell division timing and egg age [13] [14, p. 282]	Potential of embryo to reach blastocyst stage, displayed in five-tier classification system [13]	Auxogyn (now Progyny) / Merck Serono	Yes, included with Geri+ (Merck's TLI incubator) [15]	Yes [16]
	KIDScore (D3 and D5) and Guided Annotation tool	Morphokinetic scoring algorithm (deep learning); computer vision [17]	Morphokinetic variables based on AI-aided annotation [17]	Chances of embryo implanting successfully, scored from 1 to 9.9 [17]	Vitrolife	Sold separately but requires Embryoscope incubator (Vitrolife)	No
	CAREmaps	Morphokinetic algorithms (unknown type of AI)	Relative time to the start of blastulation and the duration of blastulation (based on annotated data) [10]	Probability of live birth, ranked A-D [10]	CARE Fertility (IVF clinic group), in-house	No, but CARE uses Embryoscope incubator (Vitrolife) [10]	No (not commercially distributed)
In development	LifeWhisperer	Unspecified deep learning method; computer vision [18]	Day 5 blastocysts 2D images taken with phase contrast microscope (TLI not required); morphological parameters [19]	Probability of clinical pregnancy, displayed in percent [11] [19]	Presagen	No, LifeWhisperer is designed not to require TLI equipment	N/A
	STORK	Convolutional neural network based on Google's Inception model, directly using GoogLeNet architecture (deep learning) [20]	Morphological and morphokinetic parameters [20]	Probability of clinical pregnancy (binary good and poor classification) [20]	Weill Cornell University/ IVI group (IVF clinic group) [20] [21]	No, but Embryoscope was used to train model [20]	N/A
	Ivy	Feedforward neural network with back-propagation (deep learning) [22]	Raw time-lapse imaging data [22]	Probability of clinical pregnancy (fetal heart beat), confidence score 0-1 [22]	Virtus Health (IVF clinic group)/ Harrison.AI (acquired by Vitrolife) [23]	Potentially in the future, as Vitrolife could follow Eeva model	N/A

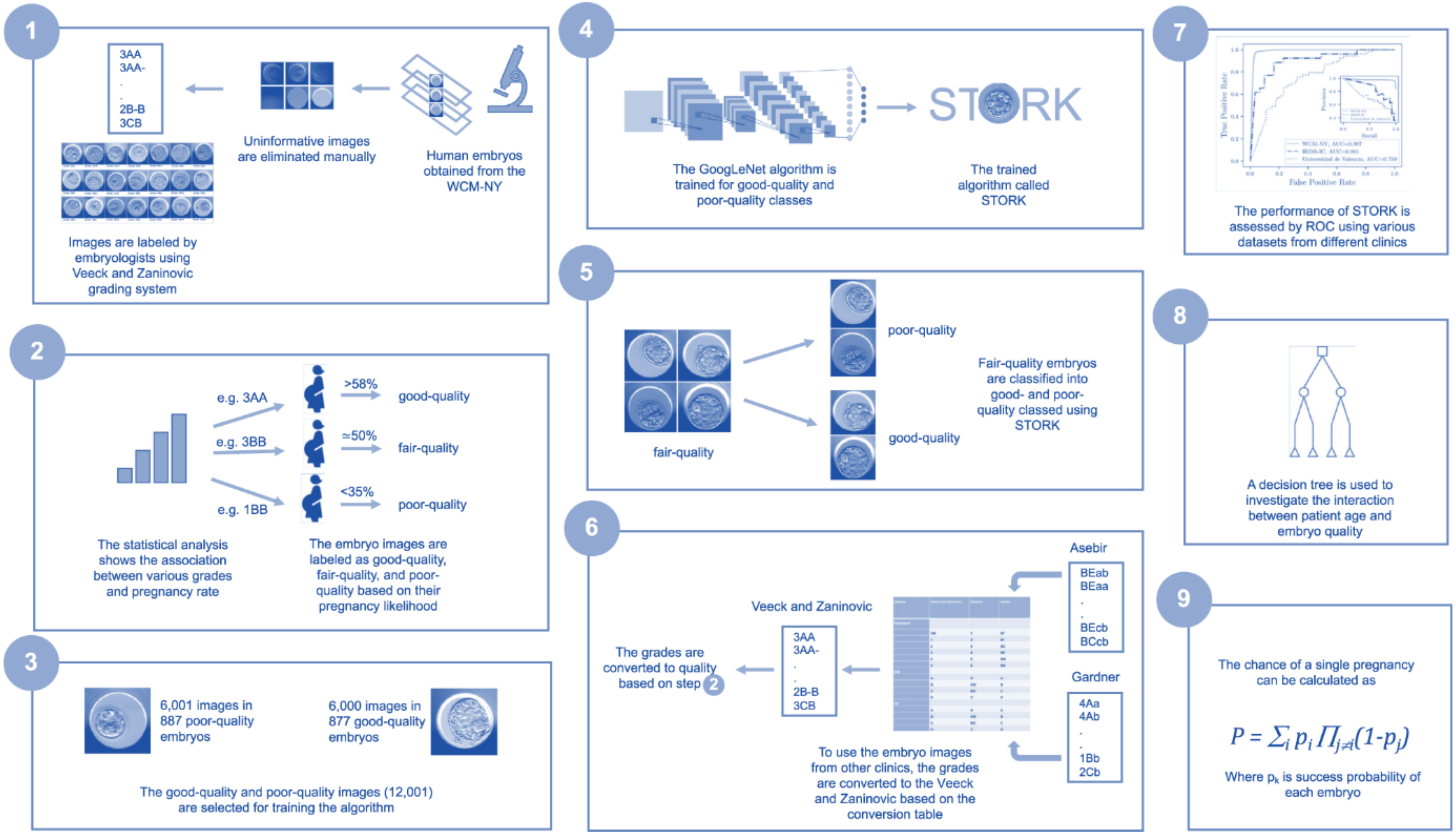

Figure 1. Developing AI-assisted embryo software: Flowchart showing the design and assessment of the STORK algorithm. *Source*: https://www.nature.com/articles/s41746-019-0096-y/figures/1 [20]

REGULATION, EVIDENCE AND LEGAL CHALLENGES

When the Medical Device Regulation (MDR) becomes fully operational across the European Union and the United Kingdom in May 2021, software claiming to make pregnancy more likely will be regulated as medical devices requiring CE marking [24], [25] [26], [27]. Currently, only the Eeva test meets those requirements (see Table 1). Software developed in-house will also have to satisfy core criteria, with the relevant UK guidance currently being finalized by the Medicines and Healthcare products Regulatory Agency (MHRA).

The United Kingdom's fertility regulator, the Human Fertilisation and Embryology Authority (HFEA), has a clear stance on optional technologies claiming to improve patients' chances of having a baby, so-called treatment add-ons: clinics should not offer any unproven add-on at an extra cost or outside of trials [28]. To be recommended for routine practice, add-ons must demonstrate benefit to patient outcomes through high-quality studies. Echoing voices from other government bodies, the same yardstick will be applied to any clinical AI tools [29], [30].

The evidence on patient outcomes is currently lacking, however. While some studies have shown AI to successfully predict live births [10], [31], [32], [33], there is insufficient good quality evidence to show that using embryo selection software improves live birth rates compared to conventional selection by an embryologist [9].

Some of the appeal of TLI has been its potential time- and cost-saving advantages for embryologists [34], [35]. However, this is not the same as improving clinical outcomes and it is unclear why this often leads to patients paying more – despite alleged efficiency gains. Involving algorithms in clinical decision-making also comes with a host of legal and ethical challenges. The most prominent issues are liability for errors and algorithmic explainability [36].

Regulators have thus begun exploring the impact of AI on their statutory duties such as inspections [37], [38]. The HFEA may, for example, probe clinics on how embryologists are trained to challenge AI-generated outputs and whether the AI itself is suitable to support their decisions [38]. This presents a skill challenge for practitioners and regulators alike but, crucially, their duty to ensure patient safety and benefit remains unaffected [39].

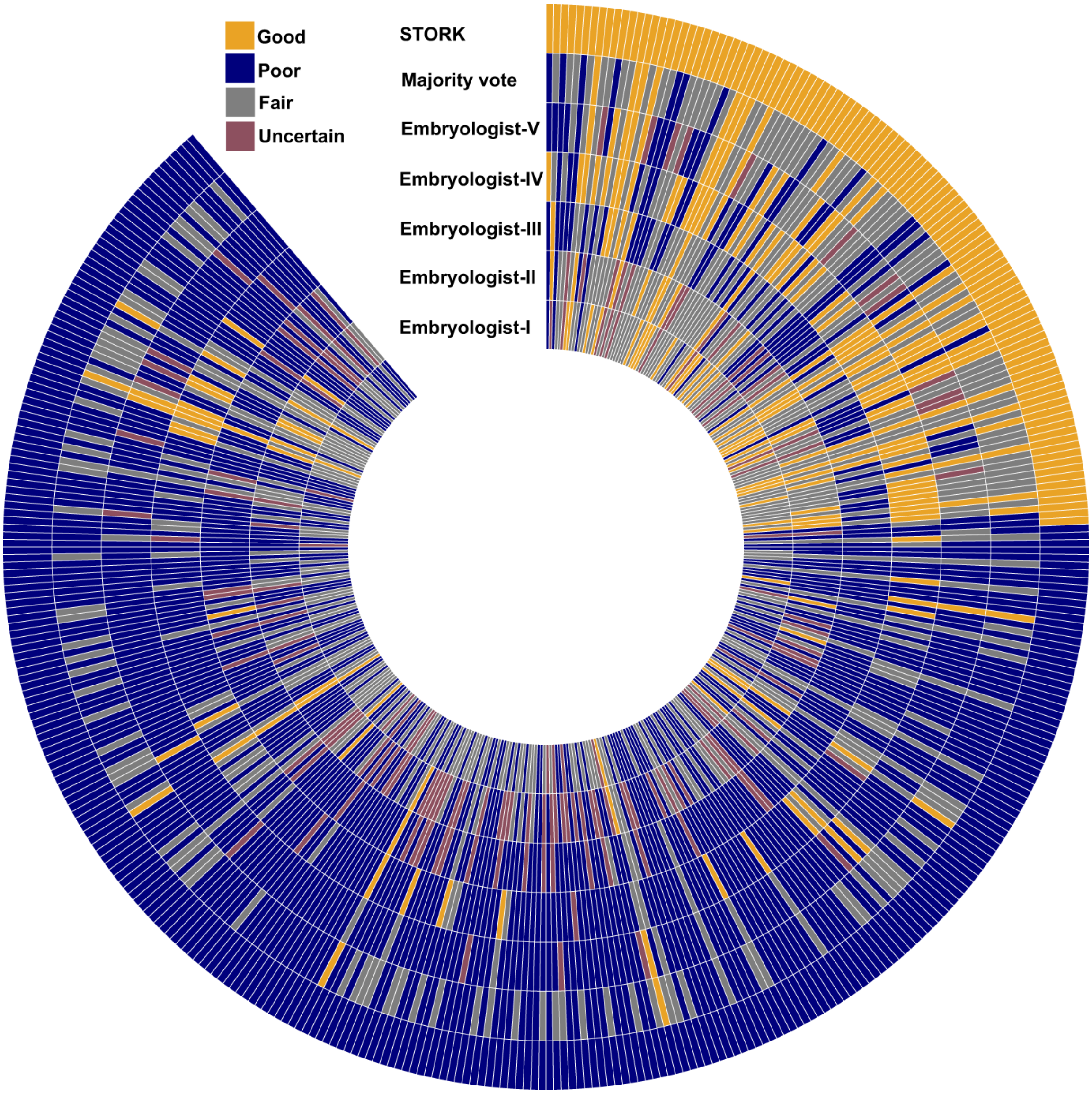

Figure 2. Towards objectivity or false certainty? A circular heatmap showing how the STORK algorithm's binary assessment compares with human embryologists' grading. *Source*: https://www.nature.com/articles/s41746-019-0096-y/figures/5

LOOKING TO THE FUTURE

Embryology is still not an exact science and often yields surprising results. Some studies on TLI, for example, show both an increase in live birth and miscarriages rates [9]. Embryo development may thus not fit into a singular algorithmic function. Yet the AI applications we identified all reduce their outputs to a sole (even if graded) success indicator. Furthermore, they are all deep learning algorithms, which are considered 'black boxes' in terms of explainability [40], [41], [42].

By contrast, Google's Optometrist Algorithm (an AI tool developed to support nuclear fusion research), for example, is not programmed to find objective mathematical functions. Instead

– just like when performing an eye test – this tool presents a human operator with two alternative settings and a variety of predicted parameters. The operator then makes a subjective choice between the alternatives, thereby integrating human intuition and choices into the model, which also makes it explainable [43], [44], [45]. This shows that yet other ways of using AI in embryology are conceivable, and they may be more appropriate given that embryo grading is not an exact science (a low-grade embryo can still result in a live birth).

Whatever advantages AI innovation will bring to the fertility sector, predictive abilities should never be confused with clinical benefit. When patients embark on the often-costly fertility treatment journey, their sole aim is to have a healthy child with the least possible amount of cycles and complications. Above all, algorithms need to robustly show that they deliver improvements on this scale. A more efficient embryo grading technique is only a means to an end after all. Like with IVF treatment add-ons, we recognize a danger that outsized claims about the benefit of AI are not backed up by the (right) evidence. To keep the full range of clinical outcomes in view, including risks such as miscarriages, developers of pioneering algorithms may have to rely on a more established method to demonstrate its benefit to patients and European health systems: the randomized control trial.

REFERENCES

[1] R. Cutting, D. Morroll, S. Roberts, S. Pickering and A. Rutherford, “Elective single embryo transfer: Guidelines for practice British fertility society and association of clinical embryologists,” *Human Fertility,* vol. 11, no. 3, pp. 131-146, 2008.

[2] NICE, “Fertility problems: assessment and treatment [CG156],” 2013.

[3] S. Cirkovic, “The UK Human Fertilisation and Embryology Authority’s "One at a Time" campaign - Centre for Public Impact (CPI),” 2019.

[4] C. L. B. C. L. Curchoe, “What AI Can Do for IVF - Scientific American Blog Network,” 2018. [Online]. Available: https://blogs.scientificamerican.com/observations/what-ai-can-do-for-ivf/.

[5] C. Hickman, “The Future of IVF,” 2018. [Online]. Available: https://www.slideshare.net/chana055/rhg-congress-2018-cristina-hickman.

[6] P. Kovacs, *Embryo selection: The role of time-lapse monitoring,* vol. 12, BioMed Central Ltd., 2014.

[7] T. Nakahara, A. Iwase, M. Goto, T. Harata, M. Suzuki, M. Ienaga, H. Kobayashi, S. Takikawa, S. Manabe, F. Kikkawa and H. Ando, “Evaluation of the safety of time-lapse observations for human embryos,” *Journal of Assisted Reproduction and Genetics,* vol. 27, no. 2-3, pp. 93-96, 2 2010.

[8] K. Kirkegaard, I. E. Agerholm and H. J. Ingerslev, “Time-lapse monitoring as a tool for clinical embryo assessment,” *Human Reproduction,* vol. 27, no. 5, pp. 1277-1285, 1 5 2012.

[9] S. Armstrong, P. Bhide, V. Jordan, A. Pacey, J. Marjoribanks and C. Farquhar, “Time-lapse systems for embryo incubation and assessment in assisted reproduction,” *Cochrane Database of Systematic Reviews,* 29 5 2019.

[10] S. Fishel, A. Campbell, S. Montgomery, R. Smith, L. Nice, S. Duffy, L. Jenner, K. Berrisford, L. Kellam, R. Smith, F. Foad and A. Beccles, “Time-lapse imaging algorithms rank human preimplantation embryos according to the probability of live birth,” *Reproductive BioMedicine Online,* vol. 37, no. 3, pp. 304-313, 1 9 2018.

[11] *How Life Whisperer works.* [Film]. 2019.

[12] Merck Serono, “Merck Serono Introduces New Eeva Test Version Aiming for Optimized Assisted Reproductive Outcomes,” 2015. [Online]. Available: https://merckbiopharmaindia.wordpress.com/2015/07/03/merck-serono-introduces-new-eeva-test-version-aiming-for-optimized-assisted-reproductive-outcomes/.

[13] Merck Serono, “The Eeva™ Test: The potential to change the future of IVF?,” 2015. [Online].

[14] G. Kovacs, A. Rutherford and D. K. Gardner, How to prepare the egg and embryo to maximize IVF success, p. 389.

[15] Merck Group, "Eeva Outcome prediction with undisturbed culture," 2019. [Online]. Available: https://hcp.merckgroup.com/en/fertility/technologies/Eeva/Eeva-Outcome-prediction-with-undisturbed-culture.html.

[16] N. Foderado, "Auxogyn, Inc. Receives CE Mark for Non-Invasive Early Embryo Viability Assessment (Eeva) Test," *BioSpace,* 23 7 2012.

[17] Vitrolife, "Evaluation tools," 2015. [Online]. Available: https://www.vitrolife.com/products/time-lapse-systems/evaluation-tools/.

[18] "Artificial Intelligence Non-Invasively Detects Down Syndrome in Embryos," *PRWeb,* 29 8 2018.

[19] M. VerMilyea, "Artificial Intelligence (AI) technology can predict human embryo viability across multiple laboratories with varying demographics with high accuracy and reproducibility," 2019.

[20] P. Khosravi, E. Kazemi, Q. Zhan, J. E. Malmsten, M. Toschi, P. Zisimopoulos, A. Sigaras, S. Lavery, L. A. D. Cooper, C. Hickman, M. Meseguer, Z. Rosenwaks, O. Elemento, N. Zaninovic and I. Hajirasouliha, "Deep learning enables robust assessment and selection of human blastocysts after in vitro fertilization," *npj Digital Medicine,* vol. 2, no. 1, 12 2019.

[21] M. Meseguer, "Time Lapse Imaging and Artificial Intelligence; the subjectivity is over," 2019. [Online]. Available: https://www.ivi-rmainnovation.com/time-lapse-imaging-and-artificial-intelligence-the-subjectivity-is-over/.

[22] D. Tran, S. Cooke, P. J. Illingworth and D. K. Gardner, "Deep learning as a predictive tool for fetal heart pregnancy following time-lapse incubation and blastocyst transfer," *Human Reproduction,* vol. 34, no. 6, pp. 1011-1018, 4 6 2019.

[23] Vitrolife, "Vitrolife acquires artificial intelligence technology for embryo assessment," 10 4 2019.

[24] C. Holder, M. Iglesias Portella, J. P. Trialle and J.-M. Van Gyseghem, "Legal and regulatory implications of AI: The case of autonomous vehicles, m-health and data mining," 2019.

[25] UK Government, "Regulating medical devices in the event of a no-deal Brexit," 2019. [Online]. Available: https://www.gov.uk/guidance/regulating-medical-devices-in-the-event-of-a-no-deal-scenario.

[26] J. Ordish, H. Murfet and A. Hall, "Algorithms as medical devices," 2019.

[27] MHRA, "Guidance on Medical device stand-alone software including apps (including IVDMDs)".

[28] HFEA; ABA; ACE; BAS; BFS; BICa; ESHRE; FNUK; RCN; RCOG; SING, "The responsible use of treatment add-ons in fertility services: a consensus statement," 2019. [Online]. Available: https://www.hfea.gov.uk/media/2792/treatment-add-ons-consensus-statement-final.pdf.

[29] Department of Health & Social Care, *Code of conduct for data-driven health and care technology,* 2019.

[30] NICE, “Evidence standards framework for digital health technologies,” 2019.

[31] C. Pribenszky, A. M. Nilselid and M. Montag, *Time-lapse culture with morphokinetic embryo selection improves pregnancy and live birth chances and reduces early pregnancy loss: a meta-analysis,* vol. 35, Elsevier Ltd, 2017, pp. 511-520.

[32] N. Zaninovic, P. Khosravi, I. Hajirasouliha, J. Malmsten, E. Kazemi, Q. Zhan, M. Toschi, O. Elemento and Z. Rosenwaks, “Assessing human blastocyst quality using artificial intelligence (AI) convolutional neural network (CNN),” *Fertility and Sterility,* vol. 110, no. 4, p. e89, 9 2018.

[33] Y. Miyagi, T. Habara, R. Hirata and N. Hayashi, “Feasibility of deep learning for predicting live birth from a blastocyst image in patients classified by age,” *Reproductive Medicine and Biology,* vol. 18, no. 2, pp. 190-203, 1 4 2019.

[34] R. J. Paulson, D. E. Reichman, N. Zaninovic, L. R. Goodman and C. Racowsky, *Time-lapse imaging: clearly useful to both laboratory personnel and patient outcomes versus just because we can doesn't mean we should,* vol. 109, Elsevier Inc., 2018, pp. 584-591.

[35] L. van de Wiel, “The datafication of reproduction: time-lapse embryo imaging and the commercialisation of IVF,” *Sociology of Health and Illness,* vol. 41, no. S1, pp. 193-209, 1 10 2019.

[36] Nuffield Council on Bioethics, “Artificial intelligence (AI) in healthcare and research,” 2018.

[37] Care Quality Comission, “Sandbox round 2: Machine Learning (AI) and its use in Radiology, Pathology and other similar diagnostics,” 2019.

[38] HFEA, “Artificial intelligence in fertility treatment,” 2019.

[39] T. Woods, M. Ream, M.-A. Demestihas, S. Hertz, A. Steere and S. Roberts, “Artificial Intelligence: How to get it right - Putting policy into practice for safe data-driven innovation in health and care,” 2019.

[40] House of Lords Select Committee on Artificial Intelligence, “AI in the UK: ready, willing and able?,” 2018.

[41] D. V. Carvalho, E. M. Pereira and J. S. Cardoso, “Machine Learning Interpretability: A Survey on Methods and Metrics,” *Electronics,* vol. 8, no. 8, p. 832, 26 7 2019.

[42] L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter and L. Kagal, “Explaining Explanations: An Overview of Interpretability of Machine Learning,” in *The 5th IEEE International Conference on Data Science and Advanced Analytics (DSAA 2018)*, 2018.

[43] K. Quach, “Amazing new algorithm makes fusion power slightly less incredibly inefficient,” *The Register,* 25: 7, 2017.

[44] J. Bridle, New dark age: technology, knowledge and the end of the future, Verso, 2018, p. 294.

[45] How to Prepare the Egg and Embryo to Maximize IVF Success, Cambridge University Press, 2019.

[46] "Amazing new algorithm makes fusion power slightly less incredibly inefficient," *The Register,* 25: 7. 2017.

This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) License. To view a copy of the license, visit https://creativecommons.org/licenses/by-nc-sa/4.0/"