Regulatory approval for the use of AI as a medical device

BY ARIEL ONG, JEFFRY HOGG AND ARUN JAMES THIRUNAVUKARASU

A study led by researchers at Moorfields Eye Hospital and UCL Institute of Ophthalmology examined 36 'artificial intelligence as a medical device' tools approved by regulators in Australia, Europe and the US, identifying that 19% had no published peer-reviewed data on accuracy or clinical outcomes [1]. Of the remainder, reporting was often variable and validation was frequently conducted on previously collected datasets which may not necessarily be representative of the populations in which they are going to be used. Can we have confidence that 'approved' AI models are safe and effective to use with patients? Our section editor, Arun Thirunavukarasu, caught up with lead author Ariel Ong at the 2025 Oxford Ophthalmological Congress to discuss what we should take away from her study. The study was also discussed with project supervisor and senior author Jeffry Hogg.

Could you summarise what you and your team did in this project?

We reviewed regulatory databases from the US, EU and Australia to examine AI tools for ophthalmic image analysis that have been approved for use in patients. Our aim was to look at what has been approved and explore the evidence base that supports the use of these devices.

What were your main findings?

We found 36 tools for ophthalmic image analysis that had been granted regulatory approval: 35 in the EU, eight in Australia, but just three in the US. It is unclear why so few devices have approval in America. People have different theories – that this may be related to the costs (regulation is expensive!), market forces, that there are multiple markets within the EU versus just one route to funding (Medicaid) in the US, or other reasons.

Approved tools were exclusively designed for posterior segment disease, with none for anterior segment disease. Most of these were for diabetic retinopathy diagnosis and screening, although there were also some devices for glaucoma and oculomics applications.

We could only scrutinise publicly available evidence, such as information on company websites or in peer-reviewed literature. Some devices had a stronger base of evidence, with many different study designs, external and internal validation, deployment in different settings and in real clinical environments.

However, some had little to no accessible data: up to 20% with no publicly available data at all, others limited to a couple of small studies or conference abstracts with minimal information. It is likely that more data has been submitted to regulators in technical files that has not been shared, but we can't see these documents. This is important because these data are not available to people who want to procure or use the AI tools, or who have the AI tools used on them.

Finally, many validation studies used publicly available datasets which have limited clinical information on sociodemographic data available; this prevents analysis of performance across subgroups, meaning that we don't know whether the AI tool performs well on average and worse in certain populations.

How do these American, Australian, and European findings compare to the UK?

There are several ophthalmic AI tools which have been approved for use in the UK. We are supporting the policy team at the Royal College of Ophthalmologists who have developed an AI directory – you can find more details about real world and research use of approved ophthalmic AI tools in the UK there [2].

The Scottish Diabetic Retinopathy Screening Programme has been using an AI tool to help them to manage their workload for the past decade, but there are no ophthalmic AI tools deployed in routine clinical care in other parts of the NHS, though several are being trialled around the country.

Do you think there is much more evidence kept private, and is this a problem?

We don't know how much more evidence is available; commercial entities may be reluctant to publish any documents that may contain proprietary information. However, manufacturers cannot achieve regulatory approval without submitting evidence, so the presence of products on market without any publicly available evidence shows that a significant body of evidence is left unpublished.

There should be a balance between publishing proprietary information and improving transparency to inform clinicians, patients and people procuring AI tools for use. The situation is also variable between jurisdictions, with no requirement in the EU for the intended use (for which a product receives regulatory clearance) to be made publicly available, unlike in the US where at least a short version (and often high-level performance data) is made available in their regulatory database.

Separate from the issue of evidence availability, the effectiveness of AI is very sensitive to context, so evidence may not always generalise well. Validation studies must be high quality and relevant to local settings and contexts to justify deployment.

Is evidence base for regulatory approval lower for AI than drugs or other medical devices?

It's just different, and it differs by jurisdiction. In general, regulators are still playing catch up with the emergence of these relatively new technologies, and further work is needed to inform what standards are appropriate to grant regulatory approval.

Can clinicians and patients trust that approved AI devices are safe and effective?

It is worth emphasising that regulatory approval does not necessarily make an AI model safe and effective for use. It is only one step in a process, and ongoing post-market surveillance is essential. I recommend reading a commentary by Youssef A and colleagues in *Nature Medicine*, which explains why "recurring local validation" provides more useful information for testing and monitoring AI rather than a one-off external validation study [3].

Local centres, clinicians and manufacturers should all have some role to play in this process, but it is an open question as to who should be responsible for organising these types of studies.

What work needs to be done in light of your findings?

I think researchers should focus on validation beyond simple performance metrics. Interventional studies and silent trials are good ways of measuring real-world performance. There should be more examination of real-world outcomes, patient-centred outcomes, workflow metrics, health economic work, and human-computer interaction studies to understand whether and how AI models should be used.

Do you have any suggestions for how clinicians can help advance and improve AI in ophthalmology?

AI literacy is key. As devices become a possibility for clinical use, it is very important for clinicians to understand that AI is not a monolithic entity, and also not to be afraid of AI or of being replaced. It is essential to understand technologies like generative AI to know not to blindly trust its outputs. In general, I would say it is important to understand what AI is and what it is not, as well as its potential and limitations, in order to use it safely. It's good to see educational offerings being developed for clinicians in this space, like the NHS Fellowship in Clinical AI, CPD, and postgraduate qualifications focusing on the practicalities of bringing AI into practice.

Any other tips?

Our study shows the importance of thinking critically and appraising evidence for yourself. Don't just look at the headlines!

Thanks Ariel, and best of luck to you and your team going forwards!

> "It is essential to understand technologies like generative AI to know not to blindly trust its outputs [...] it is important to understand what AI is and what it is not"

Glossary

- **External validation:** Testing an AI model's performance using data from other populations, institutions, setting, or times than the one used to develop or train the model. This gives us a better idea of whether the model is generalisable or robust.

- **Local validation:** Testing an AI model's performance using data from the specific site or population where deployment is planned, to ensure that it produces results that are relevant to the local clinical context.

- **Post-market surveillance:** Continued monitoring of an AI model after it has been deployed in routine clinical practice. This is helpful for identifying real-world safety issues or changes in performance.

- **Regulatory approval:** Formal authorisation by a regulatory body (e.g. the Medicines and Healthcare products Regulatory Agency (MHRA) and Food & Drug Administration (FDA)) allowing an AI model to be marketed for use in clinical care.

- **Silent trial:** Real-world testing of an AI model in the clinical workflow in real time. Unlike an interventional study, in 'silent' or 'shadow' trials, the AI-generated outputs are not shown to clinicians or used for clinical decision-making. This is helpful for understanding how the model performs and how well it integrates into the clinical workflow without affecting patient care.

References

1. Ong AY, Taribagil P, Sevgi M, et al. A scoping review of artificial intelligence as a medical device for ophthalmic image analysis in Europe, Australia and America. *NPJ Digit Med* 2025;**8(1)**:323.
2. https://www.rcophth.ac.uk/ai-directory [Link last accessed September 2025]
3. Youssef A, Pencina M, Thakur A, et al. External validation of AI models in health should be replaced with recurring local validation. *Nat Med* 2023;**29(11)**:2686−7.

AUTHORS

Ariel Ong,
NIHR Doctoral Fellow, UCL; ST6, Oxford Eye Hospital, UK.

Jeffry Hogg,
Assistant Professor, University of Birmingham, UK.

SECTION EDITORS

Nima Ghadiri,
Medical Ophthalmology Consultant and Honorary Senior Clinical Lecturer, Liverpool, UK.
nima.ghadiri@liverpoolft.nhs.uk

Arun James Thirunavukarasu,
NIHR Academic Clinical Fellow and Resident Ophthalmologist, International Centre for Eye Health, LSHTM, London, UK.
ajt205@cantab.ac.uk

Declaration of competing interests: None declared.