

Babylon AI Achieves Equivalent Accuracy with Human Doctors

28 June 2018 | News

Transatlantic collaboration between the London-based medical Artificial Intelligence (AI) company and the Royal College of Physicians, Stanford Primary Care and Yale New Haven Health yields significant breakthrough



Babylon Health has announced a world-first during a presentation streamed live from London's Royal College of Physicians: The company's AI, in a series of robust tests (including the relevant sections of the MRCGP exam), has demonstrated its ability to provide health advice which is on-par with practicing clinicians.

The MRCGP exam is the final test for trainee General Practitioners (GPs), set by the Royal College of General Practitioners (RCGP).

Trainee GPs who pass this assessment have demonstrated their competence and clinical skills to a level which is sufficiently high enough for them to undertake independent practice.

A key part of this exam tests a doctor's ability to diagnose.

1. Babylon's technology provides health information, rather than a medical diagnosis, for regulatory reasons. The tests carried out relate to the diagnostic exams taken by doctors as a benchmark for accuracy, however, Babylon's AI service remains an information service, rather than a medical diagnosis.

Babylon took a representative sample-set of questions testing diagnostic skills from publicly available RCGP sources, as well

as independently published examination preparation materials, and mapped these to the current RCGP curriculum in order to ensure the questions resembled actual MRCGP questions as closely as possible.

The average pass mark over the past five years for real-life doctors was 72%. In sitting the exam for the first time, Babylon's AI scored 81%.

As the AI continues to learn and accumulate knowledge, Babylon expects that subsequent testing will produce significant improvements in terms of results.

Important though exams are, doctors are presented with a much wider range of illnesses and conditions in their daily practice.

Therefore, to further test the AI's capabilities, Babylon's team of scientists, clinicians and engineers next collaborated with the Royal College of Physicians, Dr Megan Mahoney (Chief of General Primary Care, Division of Primary Care and Population Health, Stanford University), and Dr Arnold DoRosario (Chief Population Health Officer, Yale New Haven Health) to test Babylon's AI alongside seven highly-experienced primary care doctors using 100 independently-devised symptom sets (or 'vignettes').

Babylon's AI scored 80% for accuracy, while the seven doctors achieved an accuracy range of 64-94%.

The accuracy of the AI was 98% when assessed against conditions seen most frequently in primary care medicine. In comparison, when Babylon's research team assessed experienced clinicians using the same measure, their accuracy ranged from 52-99%.

2. As the RCGP does not publish past papers, Babylon used example questions some published directly by the College, some which were sourced from publicly available resources (which all are referenced) - during its AI exam preparation and testing.
3. Average CSA pass mark by trainee GPs was calculated using publicly available RCGP test result data from the period 2012 - 2017.

Crucially, the safety of the AI was 97%. This compares favourably to the doctors, whose average was 93.1%.