UCL

AI for People & Planet

Policy Commentary: Discovering New Medicines

Discovering New Medicines

The **Artificial Intelligence (AI) and Discovering New Medicines** discussion was organised as part of the UCL roundtable series 'AI for People and Planet'. The roundtable brought together leading minds from academia and the UK's medicines and pharmaceutical industry to identify how the UK can build on its strengths to meet the challenges and opportunities in integrating AI capabilities across research, drug discovery and clinical development and valuation.

Executive Summary

The key points highlighted during the discussion include:

- AI has the potential to systematically and correctly identify potential new drug candidates. However, a major barrier to the uptake of AI modelling in drug discovery is the lack of publicly available data and integration of health data across the NHS.
- Connecting the specialist expertise in small and medium enterprises (SMEs) and universities with the resourcing and infrastructure of larger companies would support greater data sharing. Having a common set of principles that frames what 'public good' looks like in the context of AI and drug discovery and development may help to accelerate cross-sectoral collaboration.
- There is a skills gap in industry amongst new graduates who lack foundational skills in data science, AI and machine learning. Training programmes that partner with industry would provide students with real world experience. The upskilling of existing employees is also needed.
- Greater diversity (i.e. sex, ethnicity, age and comorbidities) is needed in clinical trials to ensure AI is fit for purpose. Precision and personalised medicine, including new methods such as decentralised clinical trials and smart synthetic data, could support this increase.

These areas are presented in more detail below.

What are the opportunities and barriers to accelerated adoption of AI in drug discovery?

Pre-clinical and clinical applications

AI has the potential to accelerate the development or repurposing of drugs. The predictive ability of AI is beginning to complement wet labs and reduce Design-Make-Test-Analyse (DMTA) times.

The RECOVERY trial

The RECOVERY trial is the world's largest clinical trial assessing treatments for COVID-19 with more than 30,000 participants in the UK. Its success is partially due to its use of existing health datasets. The trial was able to identify that dexamethasone, an inexpensive and readily available drug, reduces death by up to one third in hospitalised patients with severe respiratory complications from COVID-19.

Concurrent to the RECOVERY trial, AI VIVO, a company that uses AI to accelerate drug discovery and development, also identified dexamethasone as a candidate treatment for COVID-19. This demonstrates the potential of AI for systematically and correctly identifying potential drug candidates.

Regulation

The drug discovery and development community need access to healthcare data sets, but the regulations around accessing these data often pose as barriers. While large companies might have the infrastructure and resources required, SMEs may struggle to comply with current governance requirements (for example, not having a qualified clinical epidemiologist on staff). There is an untapped collaborative market pairing smaller companies with larger ones to match specialist expertise from SMEs with the resourcing and infrastructure of larger companies. Health Data Research UK (HDRUK) is an example of an expert service platform that is not only responsible for elements of data curation, but also 'match-makes' companies who need access to data with people with clinical expertise.

While pharmaceutical companies are limited by the scope of current legislation, the COVID-19 pandemic has shown the possibilities when commercial opportunity combines with public utility. What is now needed is a collaborative effort so that all the regulators, authorities and research groups can work in concert. The UK needs to be attractive to pharmaceutical companies, for example, for the purposes of research and development (R&D) or to be a first-to-market launch area. Continuing to develop regulations and ethical protocols that protect people's data - but that are not overly prohibitive - will be important. This is especially the case given the recent proliferation of AI start-ups and SMEs in the UK, many of which have valuable offerings, but this growth could be suffocated if regulations do not account for their size and needs.

Skills training

Data science and AI-related courses are on the rise within academia. As the field becomes more established at both graduate and undergraduate levels, programmes need to have greater consistency across the fundamentals and minimum set of skills that are taught.

Greater emphasis needs to be placed on providing training programmes to equip students with the right skills for the work force. For example, AstraZeneca's newly launched Data Science and AI Graduate programme: R&D provides graduate students with the opportunity to build their technical skills and develop industry knowledge. Additionally, existing employees need to be upskilled so that people with domain expertise in the life sciences can not only understand the results of AI models and applications, but also create new ones.

Data and partnerships

Open data sets within the private and public sector would enable a closer understanding of what has worked and what has not. Current systems for accessing data closed within the NHS can be difficult to navigate and require specialist knowledge that SMEs will not necessarily have. Generally, large corporations see data as proprietary information that should not be freely shared. Work may need to be done to shift perceptions to enable healthcare data to be seen in the same way as information published in academic papers (i.e. available for people to access and use to inform their own hypotheses and decisions).

The potential of 'synthetic data'
Once a team manages to access a dataset, the variation across the country in how the data have been collected and recorded often poses a challenge. One way that universities, including UCL, have been trying to address this is through the development of 'synthetic data' (e.g. datasets that are generated by computer programs) on which to train machine learning models. Hazy is a UCL AI spin out company that focuses on AI-generated smart synthetic data. While anonymised data still pose a risk to re-identification, smart synthetic data are statistically equivalent to raw data, but pose no threat to privacy.

Can AI fulfil the promise of delivering cost effective, fair and equitable access to new medicines?

Ethics

An acknowledged issue with AI is that it may inadvertently exacerbate inequalities that have been unintentionally built into datasets. For a further discussion, refer to the Policy Commentary for the AI for People and Planet roundtable on Equity.

As the field of precision medicine and personalised medicine continues to develop, trials need to be diverse in terms of sex, ethnicity, age and comorbidities. While this might not be difficult in large cities, diversity may pose a challenge to other parts of the UK, particularly in rural areas. More work needs to be done to recruit participants so that cohorts are representative of the target population. Decentralised trials (e.g. where a trial is conducted remotely using telemedicine and mobile/local healthcare providers) have the potential to democratise the clinical trial environment as they are able to recruit people who live far away from research sites.

Social contract

At the core of any work with AI in drug discovery and development is the question of how patients in the NHS can benefit from NHS data. The NHS is faced with a dilemma where they are concerned about the ethics of selling patient data, but equally do not want to give away data for free. If patient data are used to develop a new algorithm, should the NHS then have access to the service it feeds at a reduced rate? Additionally, there is also the question of data ownership; who created the value, and for whom value should be maximised (i.e. the patient, the hospital, the university, the wider NHS etc.)?

The COVID-19 pandemic has seen a drive for open innovation, which has accelerated collaboration and development across universities, regulators and industry. Having a common set of principles that frames what 'public good' looks like in the context of AI and drug discovery and development may help to further accelerate cross-sectoral collaboration in the future.

As part of the principles, one might include placing the value derived from curating and interconnecting data where it would have the greatest impact. For example, this might be a local Trust where enhanced data and their availability would help to improve health outcomes in local populations.

UK ecosystem

Many businesses working with machine learning draw talent from, or are founded by, people with academic backgrounds. Newly developed technologies may originally be housed within a university, which acts as the bedrock for further development. However, a major barrier facing academic and students who wish to found a business based on their research is the time and resources needed to negotiate the university's financial terms for the right to use intellectual property (IP) and data. More permissive public-private technology transfer systems (i.e. systems that facilitate the transfer of technology from universities or government laboratories to companies) would offer a solution to these delays.

The COVID Symptom Study

Throughout the pandemic, the health science company, ZOE, has collaborated with King's College London (KCL) to develop the COVID Symptom Study app. Scientists from KCL are using algorithms and machine learning models to analyse self-reported data from four million people to understand how fast the virus is spreading in different areas, identify high risk areas and identify who is most at risk. The study has shown the successes that can be achieved when a private company, academia and citizen science join forces. For example, the study's data models can rapidly identify hotspots, having identified Leicester as a hotspot ahead of the Government placing the city back in lockdown at the end of June 2020.

Matchmaking larger organisations with SMEs and universities could help. However, the process of matchmaking is not yet optimised for ensuring that technology transfer (e.g. the flow of technical knowledge, data, designs, etc. from one organisation to another) can take place not only between large and small enterprises, but also between public and private organisations.

Conclusion

The Discovering New Medicines roundtable discussed the opportunities and barriers to accelerated adoption of AI in drug discovery and development and explored whether AI can fulfil the promise of delivering cost effective, fair and equitable access to new medicines. While AI has the potential to systematically and correctly identify successful new drugs, more publicly available data and better integration of health data across the NHS is needed. Connecting the specialist expertise in SMEs and universities with the resourcing and infrastructure of larger companies would support greater cross-sectoral data sharing and collaboration. Additionally, having a common set of principles that frame 'public good' in the context of AI and drug discovery and development may help to accelerate cross-sectoral collaboration and democratise the drug discovery and development process. Lastly, addressing the data science skills gap in new graduates, as well as upskilling existing employees in understanding and developing AI models, would create a workforce that is well-equipped to meet the challenges and opportunities in integrating AI capabilities across research, drug discovery and development and clinical development and valuation.

Participants

UCL Participants

Dr Jane Kinghorn (UCL Translational Research Office; *Chair*)
Professor Delmiro Fernandez-Reyes (UCL Department of Computer Science)
Dr Anne Lane (UCLB)
Professor James Thomas (UCL Institute of Education)
Professor Matthew Todd (UCL School of Pharmacy)

External Participants

Dr Junaid Bajwa (UCLH; Microsoft Reach)
Dr Nathan Benaich (Air Street Capital)
Dr Paul Colville-Nash (Medical Research Council)
Dr Sarah Dickson (UKRI)
Dr Aldo Faisal (Imperial College London)
Ms Hilary Newiss (National Voices)
Professor John Overington (Catapult Medicines Discovery)
Dr James Weatherall (AstraZeneca)

This document was prepared under the Chatham House rule by Ms Audrey Tan and Dr Jane Kinghorn. Please get in touch if you would like to know more or contribute to this discussion on Discovering New Medicines.

The AI for People & Planet Discovering New Medicines Roundtable is supported by:

UCL PUBLIC POLICY

GRAND CHALLENGES

UCL INNOVATION & ENTERPRISE

UCL TRANSLATIONAL RESEARCH OFFICE