

Artificial Intelligence in Drug Discovery & the UK Biotech Sector

Insights from the *State of the Discovery Nation 2019*

Introduction

Artificial Intelligence (AI) algorithms have gained prominence in their utility in drug discovery¹, through a synergism of:

- **Technology:** Improvements in computing power
- **Data:** Collection of diverse and large data sets
- **Commercial motivation:** Due to the enormous costs of drug discovery

This interest is fuelling significant investments by large pharma into platforms and AI companies² to tackle diverse questions such as target identification, compound discovery, clinical trial recruitment and post market surveillance^{3,4}.

Enthusiasm needs to be grounded in our deep understanding and appreciation of the AI toolbox, and requires a combination of prior expertise, continuous education, together with appropriate use for onward applications. There undoubtedly exists a great opportunity for AI to speed up and improve decision making, in order to make drug discovery more efficient, whilst enabling the realisation of personalised medicine and tailoring of drug therapy based upon the individual. This rationale will continue to prompt continued interest.

Real strength in sector

The State of the Discovery Nation 2019⁵ analysis of the views of the UK's biotech and pharma sector confirmed that AI approaches are core in all areas of modern drug discovery from target identification to clinical trials. 90% of UK SMEs confirmed that they needed data science, with half specifically needing AI and Machine learning.

The UK was seen positively in the international context, with a strong mix of data science and drug discovery expertise enabling small and growing AI in drug discovery companies to compete internationally, despite lower capital availability than in the US.

The success of BenevolentAI, Exscientia and Genomics plc, amongst others, demonstrates that AI in drug discovery companies can grow and thrive in the UK.

**90% of UK SMEs
need data science, with
half specifically needing
AI and Machine learning⁵**

¹Pannifer and Hodgskiss (2019) https://pharmafield.co.uk/in_depth/ai-the-future-of-medicine/ ²Fleming (2018) Nature, Computer-calculated compounds, 557: 555-7.

³<https://blog.benchsci.com/6-things-we-learned-about-artificial-intelligence-in-drug-discovery-from-330-scientists> ⁴Mak and Pichika (2019) Drug Disc. Today, Artificial intelligence in drug development: present status and future prospects; 24:3, 773-805. ⁵State of the Discovery Nation 2019. Joint report by Medicines Discovery Catapult and the BioIndustry Association

Opportunities for Innovation

Building on a good basis for success, there are many areas where innovation is in high demand:

User experience:

- Users have more confidence in AI models that give outputs that are interpretable by the scientist, as opposed to outputting a single, non-contextual or 'black-box' score.
- There is a need for user interfaces and functionality to be increasingly self-service, that is accessible directly by bench scientists, without needing an informatician to interpret or guide.
- AI which helps the user understand the accuracy (and uncertainty) of predictions is valuable, both to the informatics and bench science community.

Demonstrating impact:

- Any algorithm must be able to show excellent reproducibility across a range of datasets, and at large scale, as the pharmaceutical discovery process is already highly automated and optimised.
- Onward prediction and evaluation with non-training data, rather than retrospective validation, is a key marker for high performing models.
- Automated synthesis and analysis of molecules designed *in-silico* will enable rapid validation of the performance of the assets and rapid feedback to improve the algorithms.
- Although there is increased use of multi-algorithmic approaches with different algorithms being used successfully, there are difficulties in comparing results, the ability for benchmarking is key.

Data availability:

- While there are some AI approaches which make the most of sparse or incomplete data, all data science approaches benefit from increased amounts of relevant, well curated and linked data. Improved data discovery, curation and preparation would further speed progress around three-fold.
- Automation of high-throughput assays to generate relevant datasets and to assess intra-assay variability is needed.
- Key data requests are in the field of ADME and toxicity; synthesis prediction; compound bioactivity; and linked genetic-phenotypic datasets.
- There is high value in the fields of data extraction. Improvement can be made by using natural language processing in curating and enabling better use of public and private scientific and clinical literature.

Key challenges to be addressed

Access to available data:

- The sector needs methods of sharing proprietary data which do not expose precise biological targets or compound structures, while still retaining enough information to inform machine learning.
- As pharma recognise the increased perceived value of their in-house datasets, this creates a barrier to sharing their data; despite a willingness to find a way to share. Although IMI⁶ initiatives have been set up enabling pharma companies to share data, most UK SMEs are not part of these initiatives, and there are no trickle-down economy benefits.

Validation of AI methodologies:

- It remains difficult to assess the value of an AI approach. This is a market failing, preventing customers from independently verifying the benefits of partnering.
- A method to validate through test and standardised datasets, pilot projects, partnerships, expert use, and relevant open competitions, is central to build credibility for AI models.

Talent and skills:

- Companies wanting to access or develop AI approaches feel more hindered by the lack of available talent, than by a lack of investment capital.
- Differing viewpoints exist on how best to commercialise and service the demand for AI approaches; that needs to be coordinated and addressed.

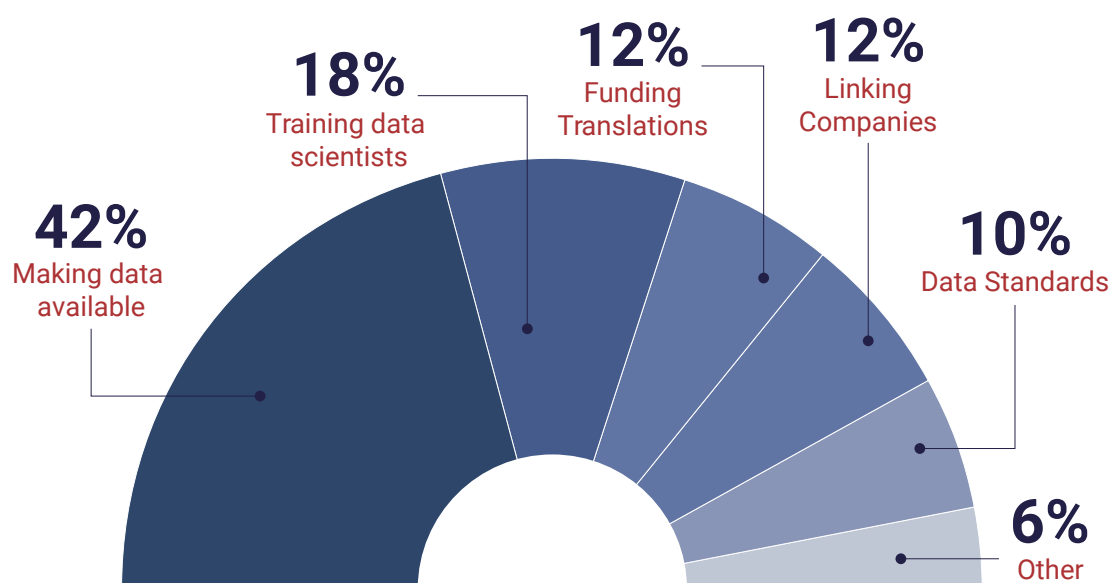
How the Medicines Discovery Catapult can support the sector

Medicines Discovery Catapult (MDC) has chosen AI in drug discovery as one of its strategic priorities. With a deep domain knowledge in drug-discovery, alongside specific expertise in scientific data analysis, information extraction, natural language processing, and machine learning, MDC is set-up to enable the community to access this established expertise.

Using our informatics skills, data extraction, extensive proprietary databases, algorithms, and market understanding, we can help validate and drive adoption of new data-driven approaches. MDC can support the community to address some of the key system-wide challenges faced, and work in partnership to help drive the innovations the sector needs.

⁶Innovative Medicines Initiative, <https://www.imi.europa.eu/>

Recommendations made by the sector on government support



Making data available is the most valued government support for UK data science⁵

MDC can support the sector to overcome challenges and innovate

- Support algorithm developers to access validation sets of data; collecting and making available datasets around ADME, toxicity, and linked genetic and phenotypic data.
- Maximise its position as a well-known and trusted member of the UK AI sector to create links between industry, academia and charities in the field.
- The fast changes in the demand and supply for AI in drug discovery led to requests for MDC to organise events and networking to support the community, improve mutual understanding between producers and customers, and matchmake deals nationally and internationally.
- Offer SME's advice and collaborate on grant applications to support access to additional funding and resources.
- Support the request from industry for 'industrial' training of data scientists, using MDC's knowledge of the sector's requirements for AI to influence future training courses.

Get in touch now

Visit our website for our latest partnership and funding opportunities.

Get in touch for more information, expert guidance, strategic consultancy and access to our technology. Or to discuss a potential project or service.

md.catapult.org.uk



info@md.catapult.org.uk



[@meddisccat](https://twitter.com/meddisccat)



[linkedin.com/company/meddisccat](https://www.linkedin.com/company/meddisccat)



01625 238734



CATAPULT
Medicines Discovery