Few topics attract excitement and caution like AI. Despite huge investments, pharma is an industry wary of hype, and is doubly cautious about handing over their highly-regimented R&D processes to hard-to-understand algorithms. Scepticism is healthy, but AI is coming. Exactly when will depend on how enthusiastically it is embraced, but once one company gets it right, and proves it, the rest will quickly follow. But whether a full industry transformation happens in two years or 10, organisations need to understand AI, and how it can benefit their business, so they can make informed decisions in the short and long term.

**How will AI transform pharma?**
At the heart of Pharma R&D is the development of new drug molecules, which are effective against a particular biological target involved in disease. This involves huge numbers of experiments, predictive models and expertise, applied across several rounds of optimisation, each with modifications to the best set of potential molecules.

Long term, AI offers the hope of a streamlined and automated approach across these various stages. A future AI may hold within its databases the sum of all knowledge about biology, genes and chemical interactions. It will be able to identify new targets and find candidate molecules for a particular target *in silico* from vast libraries, and develop and refine molecules to home in on the best ones. It will be able to specify how to synthesise the candidate molecules, gather test data and refine further.

That dream is some way off, but AI is already automating many parts of the drug discovery process. Analytics and statistical models have long been used to reduce trial and error in drug discovery. AI has the potential to remove much more, and home in on better answers much quicker than is currently possible. Even short term, AI could conceivably shave a year off the development of many drugs, which would be worth billions.

To benefit from AI short term, companies are looking at how it can deliver across different parts of the discovery process: some in a piecemeal way, some with a view to building towards complete AI driven digitalisation as the technology develops.

**Near-term AI applications**
Target identification and validation can be achieved using AI to identify biological entities to target. For example, AI and natural language processing (NLP) can be used to scan vast tomes of medical literature and genetic datasets to look for clues about gene-disease associations, to identify new targets. Previous literature may also contain clues as to how to tackle this, eg when research made interesting discoveries but did not progress them. A company progressing an early idea can...
scan past literature to see if similar ideas have already been tried, helping focus their direction of research.

With the target identified, AI can be used for molecular generation from scratch, identifying proposed new molecules and performing virtual test cycles. While AI is some way from creating drugs without human guidance, this in invaluable in bringing focus to the idea generation process.

Efficacy is only one part of the puzzle; molecules also need a whole host of other properties prediction – such as absorption, distribution, metabolism, elimination and toxicity (ADMET). For identified candidates, AI can perform in silico property prediction, allowing poor candidates to be eliminated early, and increasing throughput of good quality leads. While in silico modelling is nothing new, it is getting better with more data and better algorithms.

Neural networks can also be used for predicting retrosynthesis routes, assessing the synthesisability of a candidate, and so helping understand how easy the drug is to make. This helps improve planning, and eliminates avenues that will not scale viably.

Beyond drug discovery, it is worth noting that AI is being used across the drug lifecycle, from optimising production processes, to gathering clinical data, to assessing variations in populations that could affect response rates. Life sciences companies are increasingly breaking down silos and viewing the drug development process more holistically. Data and insights from across the value chain could be used to inform the discovery process (and discovery data will be useful further down the line). A holistic approach to technology, data and people will ensure benefits are felt as widely as possible.

**How is AI different to current data-led drug discovery approaches?**

Using data to improve predictions and automation is not new; what AI brings is an ability to deal with new levels of scale and complexity in data. Statistical methods work within the constraints of fixed assumptions based on scientific understanding, but AI can be given a broad brief (eg find a molecule that meets this criteria).

AI represents the coming together of a number of trends. Various machine learning techniques, a subset of AI, have been used in R&D for a while, but advances in algorithms and new, more sophisticated tools, such as such as generative adversarial networks (GANs), are adding the ability to tackle more complex and ambitious tasks. These advances have also been made possible thanks to the explosion in data available to process, and the exponential increases in compute power.

Some of this benefit is simply about bigger and better, or more with less. Some is about replicating
human activity in more efficient ways and automating dull repetitive work. But AI also has some genuinely new applications that could transform how drug discovery is done, thanks to the way it approaches data.

For example, machine learning may discover patterns that humans would not, and come up with solutions no sane human would, but which nevertheless work better than anything humans have discovered.

We only need to look at other industries to see the potential. NASA used generative algorithms to design an antenna which met a set of criteria – the result would never have occurred to a human, but it was better aligned to their needs than anything a human came up with. Airbus is using similar approaches to design improved aircraft parts (non-safety-critical ones for the moment). Electronic circuit design algorithms outperform human designers.

We may soon see this kind of creativity applied to drug design routinely – setting requirements and letting AI loose on data to try to find solutions that meet them. Doing so may take a change of mindset, and may involve going through a few crazy ideas, and refining the approach many times, before finding useful ones.

This is a particular advantage of machine learning. Because the algorithm itself does not have human biases it will search for the best way to meet criteria, and will search in places humans would not. Some of these will be absurd and humans will need to refine the models to redirect thinking. But – if other industries are anything to go by – some will hit on genuinely new and innovative solutions. While humans are put off by human biases it will search for the best way to meet them. Doing so may take a change of mindset, and may involve going through a few crazy ideas, and refining the approach many times, before finding useful ones. A common drug design approach is to use a higher-level description of what we want a molecule to look like. One such description is the reduced graph, which involves specifying in chemistry terminology what structure the molecule should have: for example “an aromatic ring connected to a linker, which in turn is connected to an aliphatic ring acceptor, which in turn will potentially be connected to several other molecular sub-structures with different characterisations”.

This high-level description is useful because it limits the search for molecules to those which meet specified criteria, ie having a similar structure to a known active compound. Creating a reduced graph for a known molecule is easy; the bigger challenge is the opposite process – finding good potential molecules which match the desired reduced graph.

It is a bit like buying a house: if your criterion is ‘any house’, you will never find what you are looking for. But if you specify location, number of bedrooms and price, you have a better chance. Specifying the reduced graph of a molecule is like providing a detailed layout for your ideal home. However, while there are a million or so property ads online, the number of molecules available for drug design is around 1,060, with the overwhelming majority never having been synthesised in a laboratory.

This challenge of generating a set of candidate molecules from a reduced graph description is something AI can help with. Remarkably, we found that this problem can be related to a completely separate AI challenge: translating languages.

Language translation takes advantage of two cutting-edge developments in neural networks: ‘sequence-to-sequence learning’ and ‘attention mechanisms’.

Sequence-to-sequence learning takes a sequence of words – a sentence in English – and outputs another sequence of words – a translation in French. Languages have very different structures, which is why successful machine learning approaches consider sentences in their entirety and generate a new sentence which captures the whole meaning.

It is, of course, also useful to know that particular words in each language relate to each other, and this is where the attention mechanism comes in. Attention mechanisms allow the model to focus on particular words in the input sentence when generating particular words in the output.

Together, this allows translations in which the right words are selected, but also capture the correct overall meaning.

A molecule can be represented as a text sequence using a SMILES string. The same is true of the high-level reduced graph capturing the outline of what the molecule should look like. We created an approach which applied the same basic principles of language translation to ‘translate’ the outline of a molecule into a specified novel molecule matching the outline. In other words, to predict a molecule to match our requirements.
All we required was a dataset with hundreds of thousands of molecules and their equivalent reduced graph outline to train the AI system. Fortunately, there are huge datasets of molecules readily available and generating high-level descriptions of a complete molecule is relatively easy. For any given reduced graph AI can propose new molecules which match the specification and which chemists can use to guide their search for the next drug candidate.

**Becoming an AI-driven organisation**

For every success like the one above, there are many stories of AI failure. So how do we get it right?

At a high level, AI success needs a mindset change. It needs a willingness to take risks and step into new areas. Previous analytics were predictable and easy to understand. AI learns to recognise connections in data, but it is not always easy to see how it works. This creates fear of losing oversight and transparency. Researchers need to become comfortable working with this new approach. Deploying it in the correct way, as we will come to shortly, can go a long way to easing concerns.

AI also needs innovative thinking and a willingness to learn from others in identifying how it can be deployed within an organisation, as we saw from the generative design algorithms used by NASA, and the use of language translation AI in drug design. Establishing a connection between apparently unrelated problems, such as drug discovery and language translation, may seem like a chance occurrence. But many successful applications of AI come from examining related problems in other domains, and understanding how to extend them to new challenges.

That said, humans will remain critical to building, training and overseeing AI. Drug chemistry is hard to predict and human experience will count for a long time. Good AI needs to learn its understanding from human experts and the data they have created. Good models will still need chemists who can use their experience to assess potential problems before spending money to progress AI recommendations. Good AI deployment will require companies to build capability which mixes human expertise with understanding of data and technology. It will require training programmes for users to secure buy-in and ensure these new complex tools are used correctly.
The practicalities of deploying AI

Beneath this big picture change, there are a number of practicalities of deploying AI which also need to be understood.

Each AI deployment will need pragmatic understanding of the intended real-world usage, and must be driven by people who understand the data and the underlying issue being solved, not detached technology professionals. While some niche tools with narrow applications can be simply plugged into your organisation, most AI solutions will need to be custom built to solve very specific problems and trained on your own data. The AI should be robust by design, selecting or creating the right tools and algorithms for the task.

Explainability requirements must also be considered in design. If the user does not understand how AI works, they will struggle to trust the results. Fair, safe systems that can be trusted may require some level of explainable AI (XAI), the principle of designing systems with provision for interpretation and understanding of decisions. This is an actively developing research area, but through a considered approach to data, users and the system, creation of trustworthy automated systems is not an insurmountable challenge.

Even so, in some cases it will just be too complicated for humans to fully understand how an AI reached its decision; trust can only be built through careful validation and testing. For the language translation project, we set aside molecules and reduced graphs from the training data, which were used to provide reduced graphs of drug candidates from published literature which the system had never seen before. If the AI system could take these high-level descriptions and generate a known active compound, this would be a great indication of its value in future discovery programmes. We performed this test with several different known active molecules, which had not been seen by the AI system. In most cases, a known active compound was generated.

Finally, models must be subject to ongoing monitoring through a skilled operations team, with adequate long-term support, retraining processes and controls for model drift over time.
This is all part of the process of AI maturity. The initial push into AI in pharma, as in most industries, was ad hoc and indiscriminate. That is to be expected for any new technology finding its feet. But the industry is now mature enough to get things right.

**Conclusion**

AI is a globally transformative technology, akin to the internet in its potential. But whether it will succeed in every case is down to organisations and their politics.

Pharma has always been good at generating and using data to make better decisions. The growing power of AI will increasingly allow it to make new discoveries. This comes with risks and opportunities that organisations will have to navigate.

The pace of progress remains to be seen. The small stuff – process automation and optimisation – will likely advance rapidly. But the big stuff – using AI to find new drugs – may be slower. There is lots of theoretical work on improving drug design using AI, but developing use cases which lead to big successes is always challenging. It is a bit ‘chicken and egg’ – companies need to see examples of where AI has made a difference before they are willing to make the transformations needed to unleash AI’s full potential, but the big business breakthroughs will take good examples of molecules reaching early trial phases before they are conclusively proven, which could be years.

Some pharma companies have bought into the much talked of digital transformation, breaking down silos and collating and standardising their data and processes, ready to be harnessed for R&D. These will be well placed to make strategic deployments of AI across the board, building entire new platforms from the ground up which will collect all their data and enable AI to be applied across all stages in a closed loop approach.

Others are looking across the existing drug development cycle and identifying specific points where AI can be dropped in to improve processes. This is easier and safer, but less likely to bring truly transformational results.

There is little doubt that AI will transform the industry. Everyone is investing heavily, but views differ on when real change will be seen. It will depend on how grand each company’s vision is.

A real breakthrough, which could see a tipping point in the industry, would be to ID a novel target and then orchestrate finding a drug to hit that target, which was subsequently shown to be right in a trial. It will take a few years to get to this point and even then it would still be years of trials before we can truly say we have a fully AI-developed drug on the market.

In 2018 the flu vaccine ‘turbocharger’, developed by scientists from Flinders University in Australia, went into clinical trials, with the team’s press release hailing it as the first drug designed by Artificial Intelligence. The team used an AI program called SAM (Search Algorithm for Ligands), which was fed information on chemical compounds known to activate the human immune system, as well as compounds known to have no effect on it.

They then developed a computer program that could generate trillions of chemical compounds and let SAM decide which were promising candidates. The team then synthesised some of SAM’s top candidates, one of which proved incredibly effective in animals and has now moved to clinical trial.

It would be a stretch to describe this as an AI-designed drug, since there are so many steps involved in drug development. But the media coverage hints at the growing excitement that a truly AI-developed drug might be possible in a few years. When it does, the industry could see dramatic transformation. In the meantime, incremental improvements in AI are becoming ever more prevalent in optimising processes and helping direct drug discovery.

---

**Dr Mark Roberts** is AI Consultant and AI Lead at Tessella. He is an expert in machine learning, image analysis, scientific computing and large-scale data analysis and holds a PhD in Artificial Intelligence and Computer Science. After leaving academia, Mark became a consultant at Tessella where he worked for 13 years, helping many of the world’s top R&D companies solve their most complex technical and business challenges. He has extensive experience in the pharmaceutical sector where he worked as a scientific computing consultant, business analyst and project manager.

**Dr Sam Genuay** is Principal AI Solutions Engineer at Tessella. He helps organisations exploit innovations in artificial intelligence and develop novel capabilities. Sam has a PhD in Theoretical Physics from Imperial College London, and worked as a Research Fellow at The University of Nottingham before joining Tessella in 2014. He works across drug discovery, clinical development and pharmaceutical manufacturing, to identify transformative opportunities for data-driven decision-making, automation and development of disruptive approaches using technologies in artificial intelligence.