

Patterns in Practice



Cultures of data mining in pharmaceutical drug discovery

A work in progress report from Patterns in Practice:
cultures of data mining in science, education and the arts

Itzelle Medina Perea, Jo Bates, Monika Fratzczak,
Helen Kennedy, Erinma Ochu

December 2023

Background

Algorithms that identify patterns in, and learn and predict patterns from, datasets play a growing role in practice across sectors. While many data scientists believe these 'AI' (Artificial Intelligence) systems and technologies are likely to deliver new insights and efficiencies, some practitioners view them as overhyped and with the potential for negative material consequences. These perceptions are shaped by practitioners' beliefs, values and emotions. Understanding these factors is crucial to unravel the adoption and application of such algorithmic technologies in different contexts, including how practitioners engage with them. Ultimately, these beliefs, values and emotions shape practitioners' ethical considerations and preferences in using - and not using - such technologies. This work in progress report shares findings from research we conducted exploring these issues at the pharmaceutical company GSK.

Who we are and what we did

Patterns in Practice is a research project funded by the Arts and Humanities Research Council (AHRC). It explores how practitioners' beliefs, values and emotions interact to shape how they engage with and in data mining and machine learning, techniques sometimes labelled as forms of AI.

We examined these cultures of practice across three contrasting contexts: pharmaceutical drug discovery, learning analytics in higher education, and arts practice. Here, we report early findings from the pharmaceutical drug discovery case study. For this case, we worked in partnership with GlaxoSmithKline (GSK) - a multinational pharmaceutical company and our data collection was entirely within this organisation.

Through our research we aim to develop a foundation for engaging people who work with data mining and machine learning - or its results - in critical and reflective dialogue. Our working assumption is that if we want to contribute to the development of more responsible cultures of AI practice, we first need to understand these cultures of practice.

The project will help GSK to become better informed about how practitioners' beliefs, values and emotions shape how they engage with predictive machine learning techniques and outputs at different stages of the drug discovery pipeline. These insights will also be transferable to other organisations within the pharmaceutical sector.

Through exploring the perspectives of practitioners in different roles, we aimed to build a rich picture about cultures of data mining and machine learning practice in the pharmaceutical industry. To achieve this, we carried out interviews and focus groups with practitioners working across three separate projects within one organisation, GSK - a multinational pharmaceutical company. We interviewed 18 participants in different roles on three projects: computational chemists and biologists, medicinal and physical chemists, molecular biologists, scientific managers, and one computational scientist. We also conducted two focus groups, each with five people in different roles.

Data were analysed using a combination of thematic analysis (Braun & Clarke, 2006), and close critical reading around key findings. Here we report three themes forming one of the thematic narratives identified through our analysis.



1. Industry hype

In recent years, the pharmaceutical sector has witnessed a significant increase in the adoption of AI, driven by the exponential growth of available data, enhanced computing power, and refined algorithms. An indication of this trend is the growing number of pharmaceutical companies adopting AI methods to enhance their drug discovery processes (Ferreira & Andricopulo, 2019) and forming partnerships with technology firms.

We found that practitioners' beliefs, values, and emotions regarding the use of predictive machine learning in the sector—and consequently, their own work—were strongly influenced by their perceptions of what was happening externally to the organisation, particularly the current hype about AI.

Scientists in different roles repeatedly stated that AI had been overhyped over recent years, leading to large commercial investments in start-ups and research projects that promise to produce significant pharmaceutical advances from the use of such techniques. Some participants positively evaluated projects such as Google Deepmind's AlphaFold (Deepmind, n.d.), however 'Big Tech' companies more generally were believed by some to be a key driver of this hype.

This hype has generated feelings of scepticism about recent developments in the sector among some practitioners.

“

We are in an extreme hyper-phase at the moment. I think that doesn't sit well with me, there's so – too many start-ups and whatever else who sell this as the best thing since sliced bread and it will solve all our problems and blah, blah, blah and it doesn't. And the other element is that because of that hype there is not enough appreciation of the fact that the output of a model is only as good as the input of the data.”

— *computational chemist, manager*

Some believed that despite these concerns, these external factors were generating competitive pressure on managers in the pharmaceutical sector to invest in AI adoption throughout the drug discovery pipeline. This was the case even when the techniques did not always appear to be delivering immediately on their promises, which could foster feelings of frustration and worry.

Some practitioners were concerned that the hype has led to the overuse of these algorithms even when it may not be the most suitable approach. There was a perception from some scientists and computational chemists that it was seen within the field as too big of a commercial opportunity to miss, and that a fear of being left behind the competition - rather than scientific rigour - was driving some important decision making.

“

And I think there's been a lot of hype and I think a lot of people at the top have heard that and they have a fear that they'll miss out, or they have a fear that they'll get left behind if they don't embrace it.”

— *computational chemist*



This left some feeling worried about the impact that this hype could have in the field. For example, there were concerns that the hype could build up expectations that would ultimately be disappointed resulting in disinvestment, rather than patiently playing the long game of developing a sustainable AI culture in the sector.



I'm worried that there's too much hype. I'm worried that it's fine for now and in a couple of years it's going to not be popular, and things will not have delivered as much as they should have done, and then we'll have problems. ...it can just put the field back quite a long way when people stop thinking it can work because they feel like they've tried it and it didn't work."

— *computational chemist*

Beyond the implications for investment decisions, some computational practitioners also reported a feeling of pressure to live up to the promise of AI and the hype around it. While finding it exciting to be working with predictive machine learning, they also felt a strong sense of responsibility to meet high expectations, which could lead to feelings of overwhelm and fear of falling short or failing to deliver expected results.



I think it's very exciting but I feel quite a lot of pressure around it as well as an expert in this because I feel like we can't really afford to get it wrong. I feel like it's coming but if we do the wrong thing and everyone else – all our business partners lose belief in it now then that could be really quite a problem for us because getting people back on side in the future when things have improved could be quite challenging."

— *computational chemist, manager*

Such cycles of hope and hype around emergent technologies are common in innovation pathways, often motivated by a desire to generate investment - although research suggests the productive value of these hype cycles is questionable and can result in worries about the implications for people and practices. There are therefore potential lessons to be learned from other innovation hype cycles, for example in bioinformatics.

In a nutshell

- Practitioners held diverse views on AI techniques in pharmaceutical advances. Some perceived them as overhyped, generating feelings of scepticism and frustration, while others believed the hype was justified to some extent due to the opportunities presented by data.
- Some were worried that the hype could generate disappointment leading to disinvestment, rather than long term sustainable investment in AI in the sector.
- Some computational practitioners were enthusiastic about predictive machine learning but felt pressured to fulfil AI's promises, leading to a sense of responsibility to meet high expectations and potential feelings of overwhelm and fear of falling short.
- Overall, participants agreed that the pharmaceutical field needed to strike a balance between excitement and healthy scepticism when employing predictive machine learning.

2. Human v machine – creativity and complexity

The public discourse surrounding AI and its potential impact on jobs has been a topic of concern and debate across various sectors, including healthcare, financial services and education. The fear of AI "coming for people's jobs" has been a significant aspect of the broader conversation about the future of work. Within the pharmaceutical sector, there have been past reports that some practitioners have been concerned about potential job losses if there was widespread adoption of AI in the field (Allen & Bannigan, 2019). Yet others report that the feelings of fear and anxiety have now largely disappeared as the claim that AI does not represent a threat to scientists' jobs continues to be reinforced (Kirchmair, 2019; Schneider, 2019).

In our fieldwork, some people felt that there was occasional discomfort from medicinal chemists about using predictive models. At times this was assumed to derive from concern about the threat to jobs.



I think a lot of it is about people feeling disengaged for one like the chemists, that they – the chemists – some chemists feel like the computer's taking over their job. I don't think they're right to think that, but some do feel that, and that is a big challenge."

— *computational chemist, manager*

However, differing perspectives also emerged regarding the reason for the occasional discomfort from medicinal chemists. For example, some computational practitioners attributed it to the inaccurate predictions the models can produce, while others linked it to the impact on their work practice.



You will get the occasional bad result and sometimes that has led to people saying, 'Oh, this is rubbish,' ...and sort of putting them off."

— *computational chemist, manager*



Chemistry... there's a lot of creativity involved in that. They're [chemists] very strong mechanistic thinking but laid on top a lot of creativity...so that's a big challenge of getting people to buy into machine learning, because it's around computation and it's around...particularly for people with this creative background, we do see that what they consider to be a creative process and we're turning it into an engineering process."

— *computational chemist, manager*

Medicinal chemists generally felt confident that their jobs were not at immediate risk due to AI adoption. This confidence seemed to stem from the belief that AI could complement their work rather than replace it entirely. They believed they had something to offer beyond what a predictive model could.

Instead of worries about their jobs, we found a deep critique of the claim that AI was capable of replacing human insight in fields such as drug discovery. For the medicinal chemists we spoke to, they had a strong belief that an experienced chemist could outperform a computer, and that a computer could only contribute to particular tasks - not do the full job.



My job isn't being taken over by these tools...at the moment it absolutely feels like my job and what I do is worthwhile and a machine cannot do the full job that I do and that's good."

— *medicinal chemist*

One aspect of their critique was the belief that many existing applications of AI techniques were conservative in nature, meaning that they tend to stick too closely to what is already known. For example, one medicinal chemist observed that in general they thought humans were more likely to have breakthrough creative moments than AI and see the bigger picture. Another, when asked about their thoughts on the direction of the field as a whole, was generally positive, observing that AI is getting good at predicting things already known much more efficiently than humans. However, they also had a desire for AI to give them "an idea [they'd] never think of."



It's getting quite good at predicting things which we could predict before but on a more slow and laborious process. What I would really want from an AI/ML [artificial intelligence/ machine learning] is it to give me an idea I'd never think of...What I think – it's often the case with medicinal chemistry, and chemical science, and medicine, that the best things are things that are surprising that you wouldn't expect. And a lot of breakthroughs have come from unexpected discoveries and if you narrow down too much towards let's only make – predict, model things we understand on that track, I think you lose the opportunity for that serendipity."

— *medicinal chemist*

One of the three projects we explored involved working with a system for automated molecular design. The aim of this project was to run experiments as instructed by the algorithms in order to increase the variety of data - including negative results - in the database. Despite finding predictive machine learning techniques useful for particular tasks on other projects, the medicinal chemists on this project expressed concerns that their non-computational insights were devalued on this project. They felt that there were fewer opportunities for them to use their creativity and knowledge, and this change in their work practice led them to sometimes experience a sense of boredom and lack of motivation. While not worried about losing their jobs, they at times felt less satisfied with their work.



For me I think it just kind of crushed any sort of creativity that you can have in your job, because you were reliant on compounds coming through Bradshaw... and for me I found it soul destroying."

— *medicinal chemist*

Their experience on this project contrasted with the values many practitioners held in relation to scientific work. They reflected on how it was important for them to feel that their work was enjoyable, to some extent autonomous, and enabled them to use their creativity. They also valued feeling they were involved in decision making and their contributions were valued by colleagues and the organisation, and contributed to society.



In a nutshell

- Medicinal chemists expressed confidence in AI adoption not threatening their jobs, viewing AI as a useful tool to complement human expertise. They believed that computational techniques were conservative compared to human creativity and the ability to see the bigger picture. When working on the project which required interacting with a system for automated molecular design, some chemists felt their creativity and knowledge were underutilised, leading to feelings of boredom and lack of motivation.
- Enjoying their work, having autonomy, and using their creativity were all highly valued by these practitioners. They also emphasised the importance of being involved in decision-making and having their contributions recognised and valued.
- On the computational side, there were contrasting opinions regarding medicinal chemists' discomfort with predictive models. Some believed it was due to occasional inaccuracies, while others attributed it to the impact on their work process.

3. Promoting realistic expectations

While practitioners in all roles believed that predictive machine learning could sometimes be useful and they were pleased to report that they thought the models were getting better, many of them valued a culture that fostered a sense of realism about what it was possible for the techniques to deliver in the sector.

One participant, for example, felt amused regarding the hype surrounding deep learning, and tended to challenge the assumption that the complex algorithms were always the answer to the problems being grappled with.



People are very naive about machine learning and, they seem to think that's the only technique out there, right, and so, ..., you do see it a lot. It's like, "Oh, we can do deep learning on this." And actually that dataset's probably not suited for deep learning, but this algorithm here is – traditional informatics...I laugh every time someone says deep learning I bring out my magic wand..."Oh, yes, we'll shake my magic wand of deep learning and everything will be fixed."

— *computational chemist, manager*

Some participants also reflected on the uncertainty they felt around how long they might have to wait to see a more significant impact in practice. They suggested that a certain amount of confidence in the techniques was required to be an advocate even in the face of sometimes disappointing results.



I think long term it's really hard to predict. I think almost anything is possible... So I really couldn't predict what can happen in the future. I think all kinds of incredible things could happen. That doesn't mean they will."

— *computational chemist, manager*



Some practitioners described how they have gone on a journey with their expectations of machine learning while working in this context. For example, one recalls how they initially had grand expectations when they began working with this technology, however, over time, they adopted a more realistic view regarding its capabilities.



I think if you go in thinking that it's going to solve all your issues – and maybe I think I did think that when I worked on the first kind of AI/ML programme that I worked on here. I thought, oh well, these guys have all the answers. So I think that was my very first thought and then I quickly realised that actually, you need to put everything into a perspective of what that machine learning is doing...I mean, like I said earlier, it's never going to be able to predict from a billion compounds in a big soup, "That is going to be the candidate for this biological target." It might be able to draw every single structure alive, but it's never going to be able to do that."

— *medicinal chemist*

In a nutshell

- Practitioners across different roles, including computational and medicinal chemists, believed that predictive machine learning was useful and that models have improved over time. However, they value maintaining a sense of realism about the technology's limitations in the sector.
- Some practitioners in different roles shared their experiences of initially having high expectations but gradually adopting a more realistic perspective on machine learning capabilities over time.
- Some reflected on the uncertainty they felt around how long they might have to wait to see a more significant impact in practice.



What's next?

Data analysis and findings sharing: Over the coming months, we will continue analysing the data we have collected and an end of project report for each case explored will be published in mid 2024. We will also be working on a number of papers that will be submitted to academic journals for publication in 2024.

Practitioner and public dialogue events: To facilitate the sharing of knowledge, in late 2023/early 2024 we are running a series of dialogue events and releasing a podcast to engage practitioners and the public to reflect on the findings in the three contexts explored.

Artist residency: We have partnered with the Watershed's Pervasive Media Studio to host an artist in residence, composer and improviser, Craig Scott, to respond to emerging themes across all case studies. The arts residency aims to engage public audiences with our research, and there will be a musical performance and sharing of his human-machine learning response in a hybrid format in February 2024. Find out more: <https://www.watershed.co.uk/studio/events/2024/02/09/lunchtime-talk-craig-against-machine>

Recent outputs

Storytelling performance: Data/ opium

In 2022, we collaborated with Otis Mensah, musician/writer and the first Poet Laureate of Sheffield, and ENON Films to produce a short storytelling performance in response to early findings of our pharma case study.

The film is available on our website:
<https://lifeofdata.org/site/patterns-in-practice/data-opium/>



Conference paper

Beliefs, Values and Emotions in Pharmaceutical Practitioners' Engagements with Narrow AI Adoption. MYC23 Proceedings. 2023 ASIS&T Mid-Year Conference. pp. 1-4.
<https://doi.org/10.5281/zenodo.7733344>

In this paper, we report early findings from this case study.



Acknowledgments

We would like to thank our participants and our partner GSK for their time and commitment to the project.

For more information about the project and our upcoming events, please visit:
<https://lifeofdata.org/site/patterns-in-practice/>

The support of the Arts and Humanities Research Council is gratefully acknowledged. Grant number AH/T013362/1.