

'AI-volution' in Life Sciences?

Unveiling trends in worldwide patent filings

June 2024



Computing, including AI, applied to life sciences is undeniably a rapidly growing area, driven by significant investment, recent technological advancements and heightened public interest. A look at published patents from 2014 to 2023 reveals trends in patent filings across various life sciences sectors and highlights geographical hotspots.

Patent filings for computing applications in life sciences outpaced growth in general patent filings between 2017 and 2021. The largest number of these filings is in the healthcare field (over 75%), followed by the areas of bioinformatics and computational chemistry.

China is the largest global contributor of ICT in life sciences patent filings, followed by the US. The majority of Chinese applicants with the highest number of filings were universities, in contrast to the US where industry dominated this patenting space.

Surprisingly, the growth in the number of filings attributable to AI (and machine learning) is not as large as perhaps might have been expected in view of the significant investment, development and hype surrounding AI in recent years. Nor does AI appear to be the main driver of growth in either bioinformatics or computational chemistry in the time window studied. It may be that it is too early for the real growth in AI life science technology to be reflected in the available patent data, or that applicants are keeping the AI technology as a trade secret.

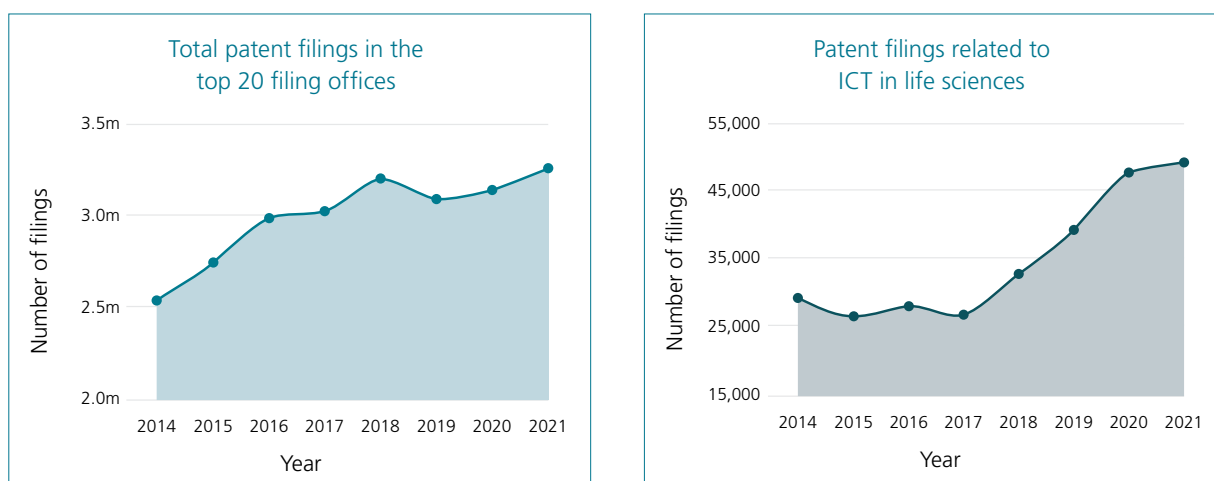


Computing in life sciences:

Surging ahead in growth in patent filings

Compared to patent filings across all technology areas (according to data from the top 20 patent filing offices), life sciences-applied ICT patent filings experienced rapid growth between 2017 and 2021, averaging 20% growth year-on-year from 2017 to 2020, but slowing to 3% in 2021. This growth far outpaced the growth seen in the total number of filings in the same period (see *Exhibit 1*). Overall, this reflects the significant investment in ICT in life sciences as well as the developments in the practical application of this technology in this field.

Exhibit 1: Filings in ICT applied to life sciences compared to the total number of patent filings from top 20 patent offices¹



Patent filings for computing applications in life sciences have seen rapid growth in recent years, outpacing growth seen in general patent filings between 2017 and 2021.

¹ Data for 2022 and 2023 are not included as the data set is incomplete because of the 18-month delay between patent filing and publication.

The pulse of progress:

Healthcare leads filings in computing in life sciences

To take a deeper look at the areas driving the growth in ICT in life sciences, the numbers of worldwide ICT patent filings in the fields of bioinformatics, computational chemistry and healthcare were assessed. Data categorised by IPC codes G16B, G16C and G16H were considered representative of ICT patent filings in these fields (see *Exhibit 2*).

Exhibit 2: Examples of subject-matter falling within IPC classification codes

	Bioinformatics (G16B)	Computational chemistry (G16C)	Healthcare (G16H)
Definition	Bioinformatics, i.e. Information and Communication Technology (ICT) specially adapted for genetic or protein-related data processing in computational molecular biology	Computational chemistry; chemoinformatics; computational materials science	Healthcare informatics, i.e. Information and Communication Technology (ICT) specially adapted for the handling or processing of medical or healthcare data
Examples	Protein structure prediction	Predicting acid dissociation constant	Health management system
	Identifying and validating antigens	Calculating synthetic accessibility of a molecule	Method for selection of substances in treatment
	Classifying and predicting cancer prognosis using patient samples	Predicting activity of a target compound	Stateless decision support for clinical evaluation
	Analysing ligation-based sequencing data	Means for electronically drawing and editing representations of chemical structures	System for efficient creation of customized healthcare application modules

When the ICT filing data are broken down by field, it becomes apparent that healthcare accounted for the vast majority of ICT filings in life sciences (see *Exhibit 3*), followed by bioinformatics and computational chemistry. This is perhaps unsurprising as this field has enormous breadth, ranging from diagnostic and prognostic methods to IT systems for handling patient data, image processing and epidemiological data.

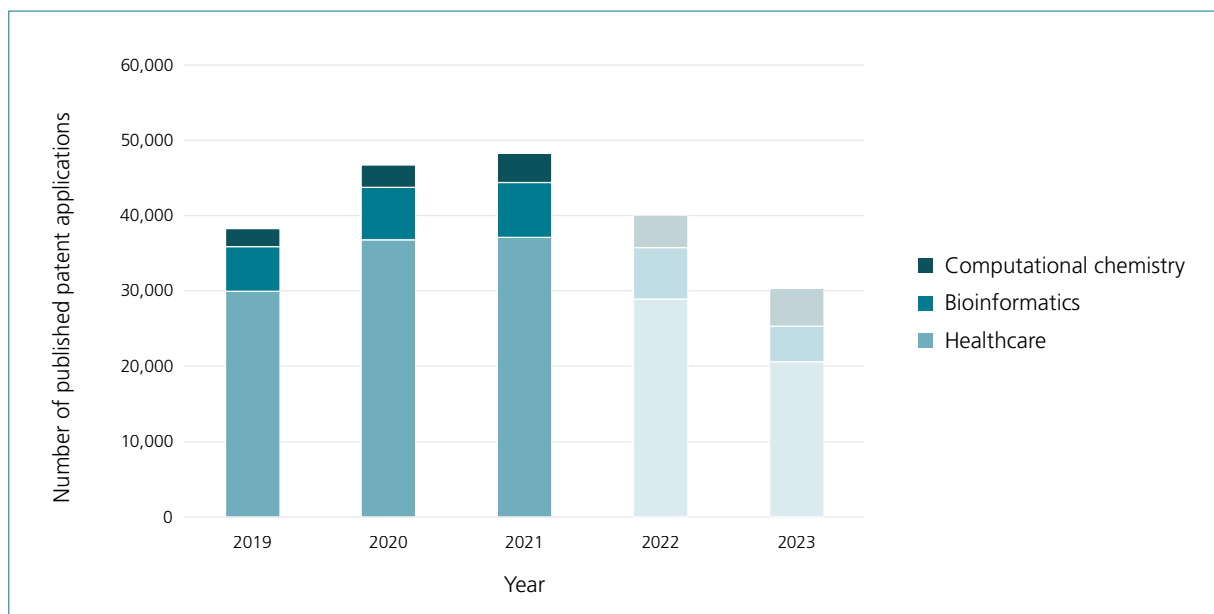
Between 2019 and 2022 bioinformatics was the more active area for ICT filings than computational chemistry. However, the percentage growth in the number of filings in bioinformatics is smaller than the corresponding growth in the number of filings in computational chemistry in that same period, such that the total number of filings in computational chemistry looks set to overtake those in bioinformatics for the first time in 2023 (although it remains to be seen what the final filing data will be as data for 2022 and 2023 are incomplete because of the 18-month delay between patent filing and publication).

Perhaps the differences in the growth of filings in bioinformatics and computational chemistry reflects in part the differences in the complexity of biological systems versus chemical ones. For example, the use of ICT, including AI for small molecule design has been around for much longer than e.g. de novo design of antibodies, first reported in 2024.



The largest number of filings in ICT in life sciences are in the healthcare field, followed by bioinformatics and computational chemistry.

Exhibit 3: The number of published filings seen in applied ICT subcategories²

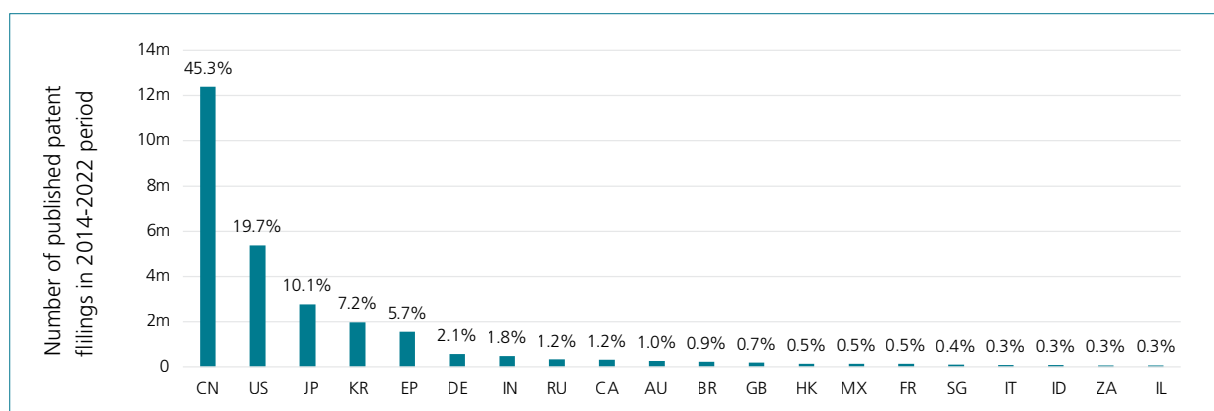


² Healthcare remains the largest field, followed by bioinformatics and computational chemistry. Data for 2022 and 2023 are incomplete because of the 18-month delay between patent filing and publication.

China leads the way

China is the largest global contributor of overall patent filings in the world, contributing 2.3-fold more filings than the USA (see *Exhibit 4*).

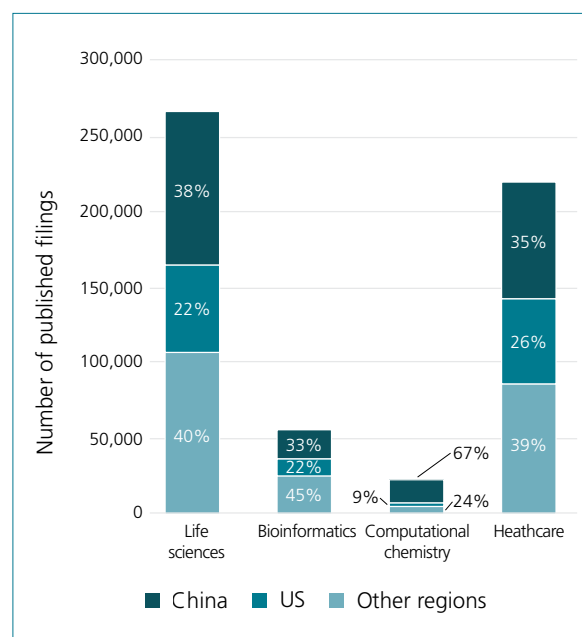
Exhibit 4: Number of published patent applications filed between 2014-2022 for top 20 filing offices



This is also the case in the field of ICT applied to life sciences from 2014-2023, where China leads the way, followed by the US. China also remains the top filing country when the ICT in life sciences filing data are further split into ICT filings in bioinformatics, computational chemistry and healthcare (see *Exhibit 5*).

Interestingly, the US's contribution to the number of filings in the ICT in life sciences field is similar to its contribution in the total number of patent filings (22% vs 19% respectively) and remains similar for ICT filings in the bioinformatics and healthcare fields (22% and 26% respectively). In contrast, the US's contribution in the computational chemistry field is significantly smaller (9%) with China's filings in this field (67%) outnumbering those from the US by over 7-fold. It is worth noting that applications can be categorized under multiple IPC codes therefore, some applications may have been duplicated in the bioinformatics, computational chemistry and healthcare fields. This overlap may account for the increase in the total number of applications classed under bioinformatics, computational chemistry and healthcare, as compared to the total number of ICT in life science applications.

Exhibit 5: Filings split into ICT filings in bioinformatics, computational chemistry and healthcare by country between 2014 and 2023



China's overall dominance in patent filings is perhaps unsurprising, given its status as the world's most populous country up until 2023 and that it provides multiple non-market driven incentives for filings.

However, given that the US is the largest pharmaceutical market in the world, China's dominance in the ICT in life sciences field might be less expected. Nevertheless, it appears that being the largest market does not compensate for China's size and patenting incentives. China also dominates the chemical industry market as compared to the US (43% vs 11% market share in 2021³), which likely plays a role in the vast difference seen in computational chemistry ICT filings.

Looking at the top 50 assignees for patent filings for ICT in life sciences, 19 are American, 15 are Chinese, 9 are Japanese and 2 are German (see *Exhibit 6*). There is also 1 entity per country for France, Hong Kong, Korea, Netherlands and Switzerland. Interestingly, out of the 15 Chinese entities, 10 are universities, whereas for the US only a single university (University of California) appears on the list.

Perhaps China's dominance in the number of filings in this field results from non-market driven incentives provided in China for filing patent applications (such as requirements for research grants), such that this sector is flooded by filings from academic institutions.

In contrast, the US's contribution appears to arise mainly from industry. It would be interesting to see whether the subject-matter of these filings reflects this difference. Either way, the sheer volume of filings shows the significant investment in ICT in life sciences in these countries making them market leaders.

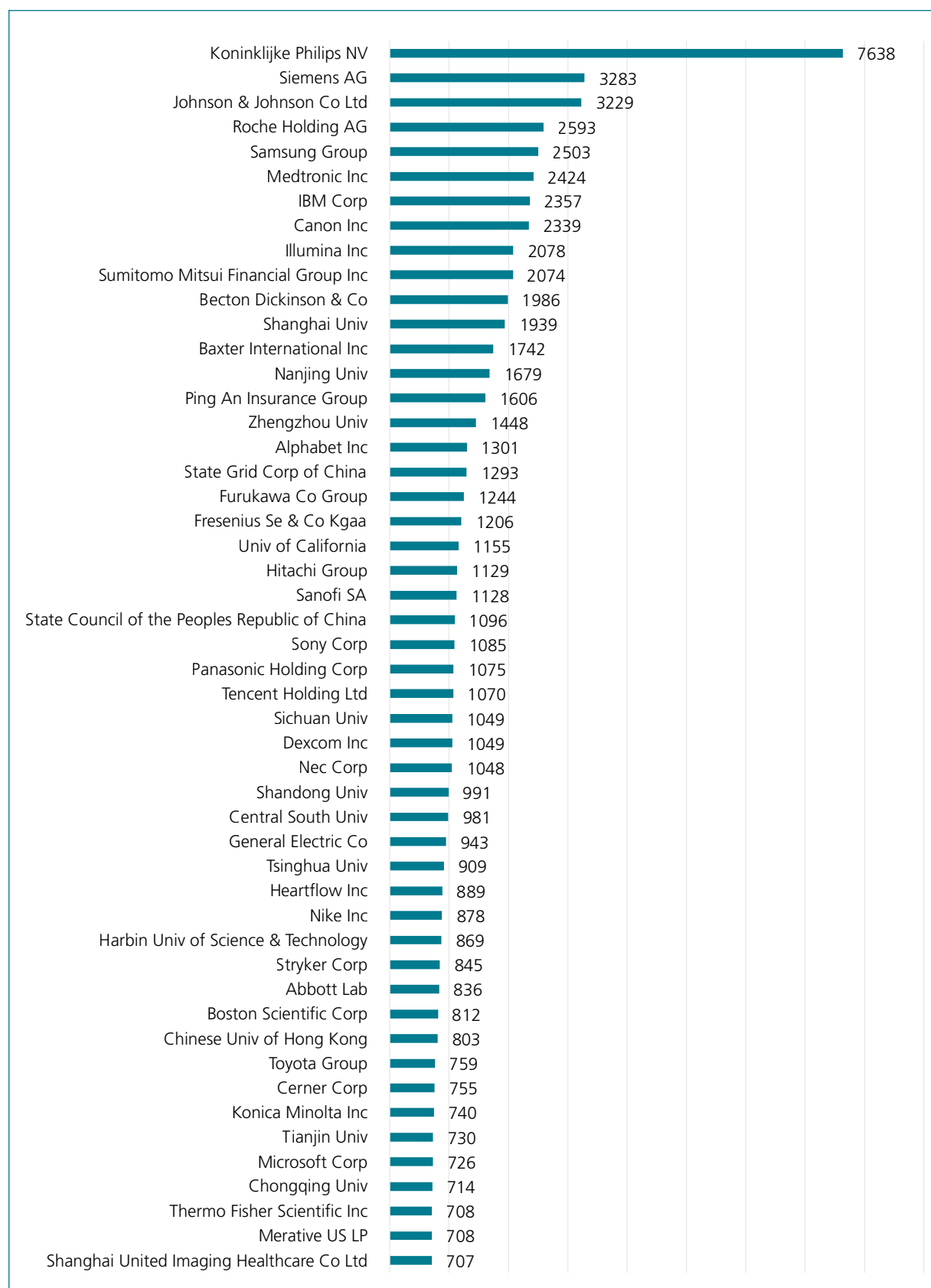
The European Patent Office (EPO) lags behind China and the US in its contribution to the number of filings in the ICT in life sciences (data not shown). The currently proposed EU AI Act could pose a significant hindrance to AI development in the EU since some of the uses of AI in the life sciences sector will be classed as 'high-risk'. Accordingly, enhanced AI regulation requirements will apply to both developers and users of technologies such as AI systems intended to be used as safety components of medical devices or in vitro diagnostic medical devices. This could lead to a downturn in AI life sciences patents originating in the EU. For as long as there is no consistent global approach, we might expect jurisdictional differences in the approach to AI regulation to further influence patent filing trends.



China is the largest global contributor of ICT in life sciences patent filings followed by the US. The majority of Chinese applicants with the highest number of filings were universities, in contrast to the US where industry dominated this patenting space.

³ <https://www.statista.com/statistics/1247223/global-chemical-sales-market-share-by-country/> – accessed April 2024.

Exhibit 6: Top 50 applicants by number of filings for ICT in life sciences between 2014 and 2023



Is the 'AI-volution' in life science patents over-hyped?

To obtain an understanding of the contribution of AI in bioinformatics and computational chemistry, filings relating to machine learning in these categories were analysed. Data categorised by IPC codes G16B 40/20, G16B 40/30 and G16 C20/70 were considered to provide an indication of the impact of AI in bioinformatics and computational chemistry (see *Exhibit 7*). Due to the apparent lack of an equivalent IPC machine learning code for healthcare, the impact of AI in healthcare was not analysed (see *Methods section*).

Exhibit 7: Examples of subject-matter falling within specific IPC codes

	AI in bioinformatics		AI in computational chemistry
	G16B 40/00 ICT specially adapted for biostatistics; ICT specially adapted for bioinformatics-related machine learning or data mining, e.g. knowledge discovery or pattern finding		G16 C20/70
	G16B 40/20	G16B 40/30	
Definition	Supervised data analysis	Unsupervised data analysis	Machine learning, data mining or chemometrics
Examples	Analysing genotyping data	Diagnosing brain injury based on a metabolite profile	Improving accuracy of drug toxicity prediction
	Diagnosing brain injury based on a metabolite profile	Characterising extracellular vesicle population	Suggesting chemical compounds using artificial intelligence
	Generating configurable text strings based on raw genomic data	Taxonomic classification system	Designing polymer materials of desired physical properties
	Identifying and validating antigens	Classifying tumours	Learning model for discovery of preferable combination of water-repellent agents

AI in bioinformatics

The number of AI filings in the bioinformatics field grew between 2019-2022 (see *Exhibit 8*), with the highest year-on-year growth (36%) seen in 2020 (for available data). However, the year-on-year growth of this sub-field appears to have subsequently slowed with 5% growth in 2021, and so far 8% growth in 2022 and 16% fewer filings seen in 2023 (although these figures are subject to increase once the complete data for these years are available).

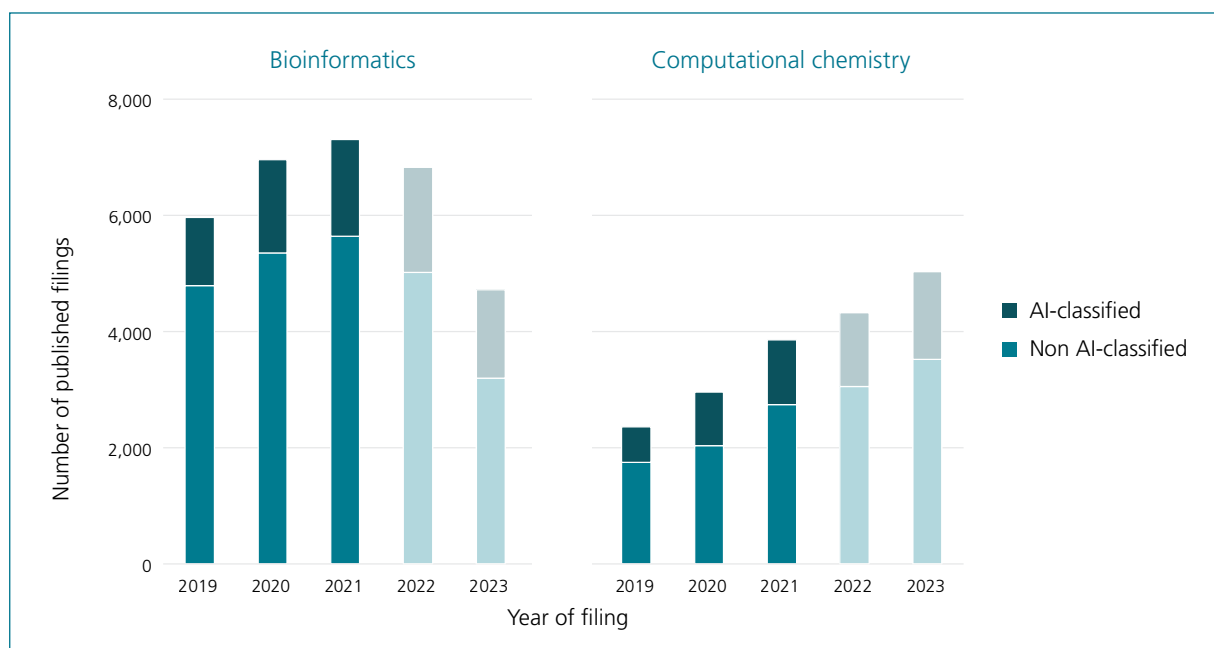
AI as a proportion of the bioinformatics field has also grown, accounting for 20% of all ICT bioinformatics filings in 2019 to (currently) 32% of filings in 2023 (see *Exhibit 8*).

Overall, the AI in bioinformatics field has experienced growth between 2019 and 2022 (1.5-fold increase) (see *Exhibit 8*) and outpaced the growth in the overall bioinformatics field (1.1-fold increase), indicating an increased contribution of AI to the bioinformatics field.



Surprisingly, the growth in the number of filings attributable to AI in bioinformatics was not as large as perhaps would have been expected in view of the significant investment and developments in AI in recent years.

Exhibit 8: Published patent filings in the area of bioinformatics and computational chemistry⁴



⁴ Data for 2022 and 2023 are incomplete in view of the 18-month delay between patent filing and publication.

AI in computational chemistry

While a smaller field than bioinformatics, computational chemistry has experienced larger growth in the number of filings over the last years. This growth was also seen in the number of filings in the field of AI in computational chemistry.

The number of AI filings in the computational chemistry field grew between 2019-2023, with 52% and 21% year-to-year growth with respect to the number of filings in 2020 and 2021. The data for 2022 and 2023 so far show 14% and 19% growth to date, suggesting the filings have continued to grow, albeit at a slower pace, in contrast to the more static growth seen in the AI in bioinformatics subgroup.

Moreover, in line with the general growth in the computational chemistry field (which looks set to overtake bioinformatics for the first time in 2023 (see *Exhibit 3*)), year-to-year growth in AI filings is higher for computational chemistry than bioinformatics. In fact, for the first time, the number of filings in AI in computational chemistry in 2023 are currently almost equal to those in AI in bioinformatics. Again, this suggests that the computational chemistry field, and specifically the subgroup of AI within this field, may be a more active area of development than AI in bioinformatics.

It is interesting to speculate whether patent office requirements in Europe whereby a patent claim has to be limited to a technical purpose has an impact on the data. Finding a technical purpose in computational chemistry may well be easier than in bioinformatics.

However, unlike AI as a proportion of the bioinformatics field, AI as a proportion of the computational chemistry field grew from 26% of all computational chemistry in 2019 to about 30% in 2020 and has remained approximately at that level since⁵. Compared to the average year-to-year growth in the field of computational chemistry, the growth in this field indicative of AI cannot be seen as the sole driver of overall computational chemistry filings.



Despite the buzz about AI in recent years, AI-related technology does not appear to be the main driver of growth in either the field of bioinformatics or computational chemistry.

Machine learning in bioinformatics

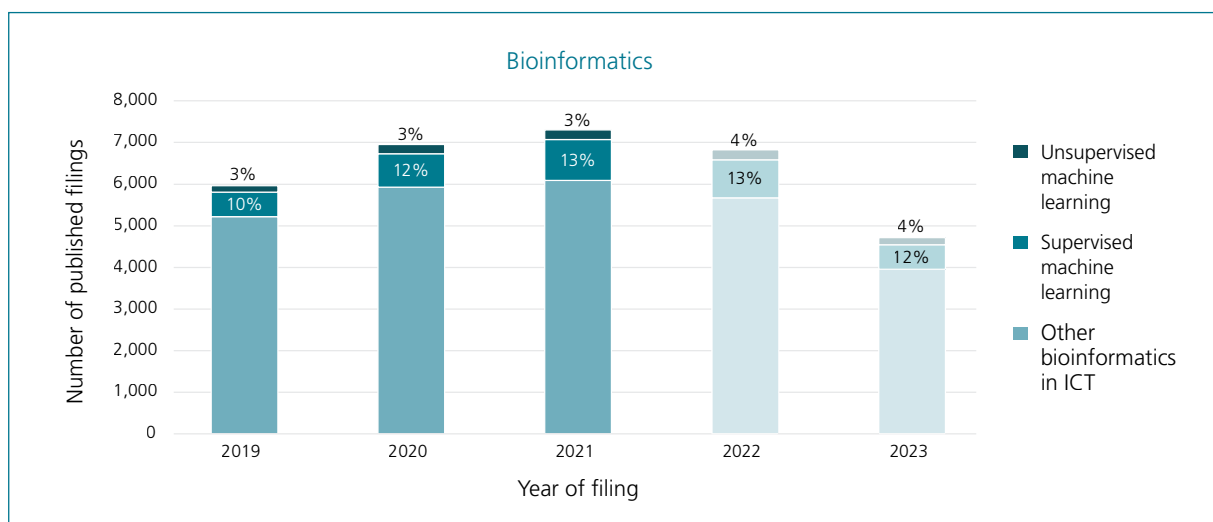
Looking at IPC codes reflecting supervised and unsupervised machine learning as a subgroup of AI in the bioinformatics field, the proportion of supervised machine learning filings fluctuated between 10-13% between 2019-2023 and the proportion of unsupervised machine learning filings fluctuated between 3% and 4% in the same period (see *Exhibit 9*).

This appears to be in line with the general machine learning trends where supervised learning techniques continue to dominate in the field of machine learning. Given the amount of attention given to the machine learning overall, it is interesting that the proportion of filings attributed to supervised or unsupervised machine learning did not appear to experience a large change in the period analysed⁶.

⁵ Bearing in mind incomplete data for 2022 and 2023.

⁶ One has to caveat, that some machine learning subject-matter in bioinformatics may fall into "Other bioinformatics category" in the graph below, due to the IPC classification.

Exhibit 9: Published patent filings in bioinformatics⁷



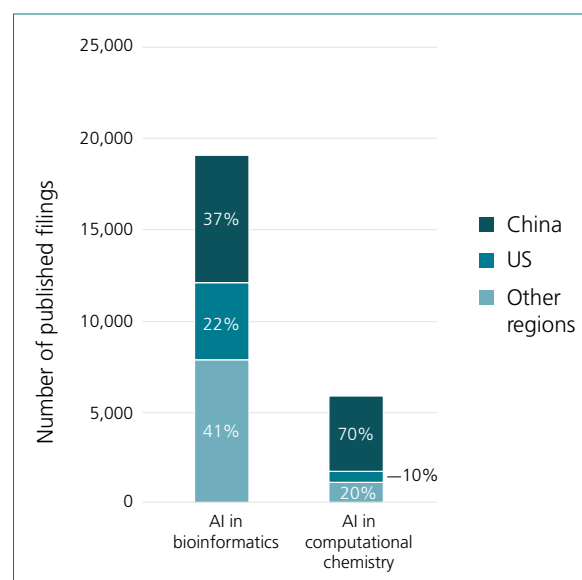
AI in bioinformatics and computational chemistry – distribution of filings

In terms of the distribution of filings reflective of AI in bioinformatics and in computational chemistry across different geographical regions, the earlier-demonstrated patterns for the broader fields of bioinformatics and computational chemistry were reflected (see Exhibit 10), with China leading the way in the number of filings.

Exhibit 10: Filings in AI In bioinformatics and AI in computational chemistry with numbers attributable to China, US and other regions



While AI-related filings are undoubtedly a significant part of computing in life sciences, they do not account for the majority of filings – or at least not yet.



⁷ Supervised machine learning filings continue to outnumber unsupervised machine learning. Comparative data is not available for computational chemistry due to the lack of subcategories for the G16C 20/70 code.



The future facing AI in life sciences

It is surprising that AI-related patent filings in life sciences have not (yet) experienced the levels of growth one might possibly have expected in view of the significant interest and investment in this technology. Despite the buzz about AI in recent years, AI-related technology does not appear to be the main driver of growth in either the field of bioinformatics or computational chemistry.

It might well be the case that real growth in this technology is still too new to be reflected in the current patent data given the 18-month delay between patent filing and publication, and that there will be a much larger number of filings in the near future.

A second factor is that applicants may not prioritise AI applications given the pace of development, and the relative difficulty in identifying infringement as compared to, for example, more straightforward physical products or even manufacturing methods.

Third, applicants may opt to seek patent protection only for the product developed with the use of AI. It is not uncommon for AI algorithms to be kept as trade secrets by clients.

Fourth, it could be that AI has yet to translate to a significant range of patent-worthy or patent eligible uses. It will be interesting to see what changes the next years will bring.

Finally, it is possible that there will be a shift in strategy concerning AI-related patent filings in the life science space with the advent of the USPTO guidance on AI assisted inventions that was released in February this year. Generally speaking, the USPTO guidance means that where an invention is created with the assistance of an AI tool, US patent protection is only available if a human inventor made a significant contribution to the invention. At this point it is not clear whether other jurisdictions will have similar guidance, especially since the criteria for inventorship differ between jurisdictions. At least the USPTO guidance appears to mean that US patent protection is unavailable for outputs from AI tools where there is no significant contribution from a human inventor. In contrast, if the AI tool itself is created by a human it would be potentially available for patent protection assuming it is new, inventive and meets other criteria for patent protection.

Authors



Dr. Lauren Gregory

Senior Associate

T +44 20 7367 3338

E lauren.gregory@cms-cmno.com



Marek Blaszyk

Part-Qualified Patent Attorney

T +44 20 7067 3223

E marek.blaszyk@cms-cmno.com



Dr. Sophie Skidmore

Part-Qualified Patent Attorney

T +44 20 7067 3173

E sophie.skidmore@cms-cmno.com

Special thanks to Dr. Rachel Free for additional remarks

A version of this article was published by In Vivo (see [here](#))



Methodology

Data on the overall number of patent applications filed between 2014 and 2023 were acquired from WIPO IP Statistics Data Centre. All other data were acquired from PatSeer. The search was conducted in February 2024, limiting the filing period to 01/01/2014 to 31/12/2023.

The complete filing data for 2022 and 2023 are not yet available due to the 18-month delay between the filing and publication of the patent application.

The IPC codes were used to filter the search results by providing the following search term: IPC: G06F19* OR G16C* OR G16B* OR G16H* OR G16Z*.

Assumptions

We assume G16B is a code indicative of filings in bioinformatics in general, with G16B 40/00 (or G16B 40*) related to a narrow field of bioinformatics applications including biostatistics, machine learning and data mining, based on the code definition provided by WIPO. Similarly, code G16C is assumed to be indicative of computational chemistry, with G16C 20/70 indicative of the patent filings related to a narrower area of machine learning, data mining and chemometrics. Although the IPC publication does not explicitly provide for categorisation of AI methods in bioinformatics and computational chemistry, we assume that the codes containing machine learning methods might be considered reflective of the developments in the field of AI. We assume that G16B 40/00 and G16C 20/70 are equivalents of each other in bioinformatics and computational chemistry. We also assume that codes G16B 40/20 and G16B 40/30 (supervised and unsupervised data analysis respectively) are reflective of supervised and unsupervised machine learning filings in the field of bioinformatics.

We assume that G16H is indicative of the ICT applications in healthcare. For G16H, the ICT subgroups were not analysed due to a lack of a code which would appear to be equivalent to G16B 40/00 or G16C 20/70 and indicative to some extent of the filings where machine learning was the subject of the application.

To obtain the data on the applied ICT filings in life sciences before 2019, we assume that filings which are now classified under G16B, G16C, G16H or G16Z would have previously been categorized under G06F 19/00, which we base on the IPC publication Revision Concordance List. Thus, the data presented in Exhibit 1 under "ICT in life sciences" label are summarised filings in IPC codes G06F 19/00, G16B, G16C, G16H and G16Z.

In Exhibit 8 the filings labelled as "AI-classified" are filings including either G16B 40/00 or G16C 20/70 codes respectively. The filings labelled as "Non AI-classified" are G16B or G16C filings excluding G16B 40/00 or G16C 20/70.

In Exhibit 9, the filings labelled as "Unsupervised machine learning", "Supervised machine learning" and "Other bioinformatics" are filings classified under G16B 40/30, G16B 40/20 and G16B filings excluding either G16B 40/20 or G16B 40/30. Where machine learning methods other than supervised or unsupervised were used, some machine learning filings might be included in the "Other bioinformatics" group.

IPC code definitions (short)

1. G06F19

Digital computing/data processing adapted for specific applications (including chemistry / life sciences but also others; transferred in 2018 and 2019).

2. G16H

Digital computing/data processing adapted for healthcare data (created in 2018).

3. G16B

Bioinformatics (created in 2019).

- b. G16B 40/00 (or G16B 40*) – Biostatistics/machine-learning/data mining; includes any of 40/10; 40/20; 40/30.
 - c. G16B 40/10 – Signal processing
 - d. G16B 40/20 – **Supervised data analysis**
 - e. G16B 40/30 – **Unsupervised data analysis**

4. G16C

Computational chemistry (created in 2019).

- a. G16C 20/70 – **Machine learning**, data mining or chemometrics

5. G16Z

Misc which don't fit anywhere else (created in 2019).

IPC code definitions (long)

1. G06F19

Digital computing or data processing equipment or methods, specially adapted for specific applications (G06F 17/00 takes precedence; data processing systems or methods specially adapted for administrative, commercial, financial, managerial, supervisory or forecasting purposes G06Q) [2011.01].

2. G16H

(Split from G06F19 in 2018; HEALTHCARE INFORMATICS, i.e. INFORMATION AND COMMUNICATION TECHNOLOGY [ICT] SPECIALLY ADAPTED FOR THE HANDLING OR PROCESSING OF MEDICAL OR HEALTHCARE DATA [2018.01]).

3. G16B

(Split from G06F19 in 2019; BIOINFORMATICS, i.e. INFORMATION AND COMMUNICATION TECHNOLOGY [ICT] SPECIALLY ADAPTED FOR GENETIC OR PROTEIN-RELATED DATA PROCESSING IN COMPUTATIONAL MOLECULAR BIOLOGY [2019.01]).

- b. G16B 40/00 (or G16B 40*) – ICT specially adapted for biostatistics; ICT specially adapted for **bioinformatics-related machine learning or data mining**, e.g. knowledge discovery or pattern finding [2019.01]; any of G16B 40/00; 40/10; 40/20; 40/30
 - c. G16B 40/10 – Signal processing, e.g. from mass spectrometry [MS] or from PCR [2019.01]
 - d. G16B 40/20 – **Supervised data analysis** [2019.01]
 - e. G16B 40/30 – Unsupervised data analysis [2019.01]

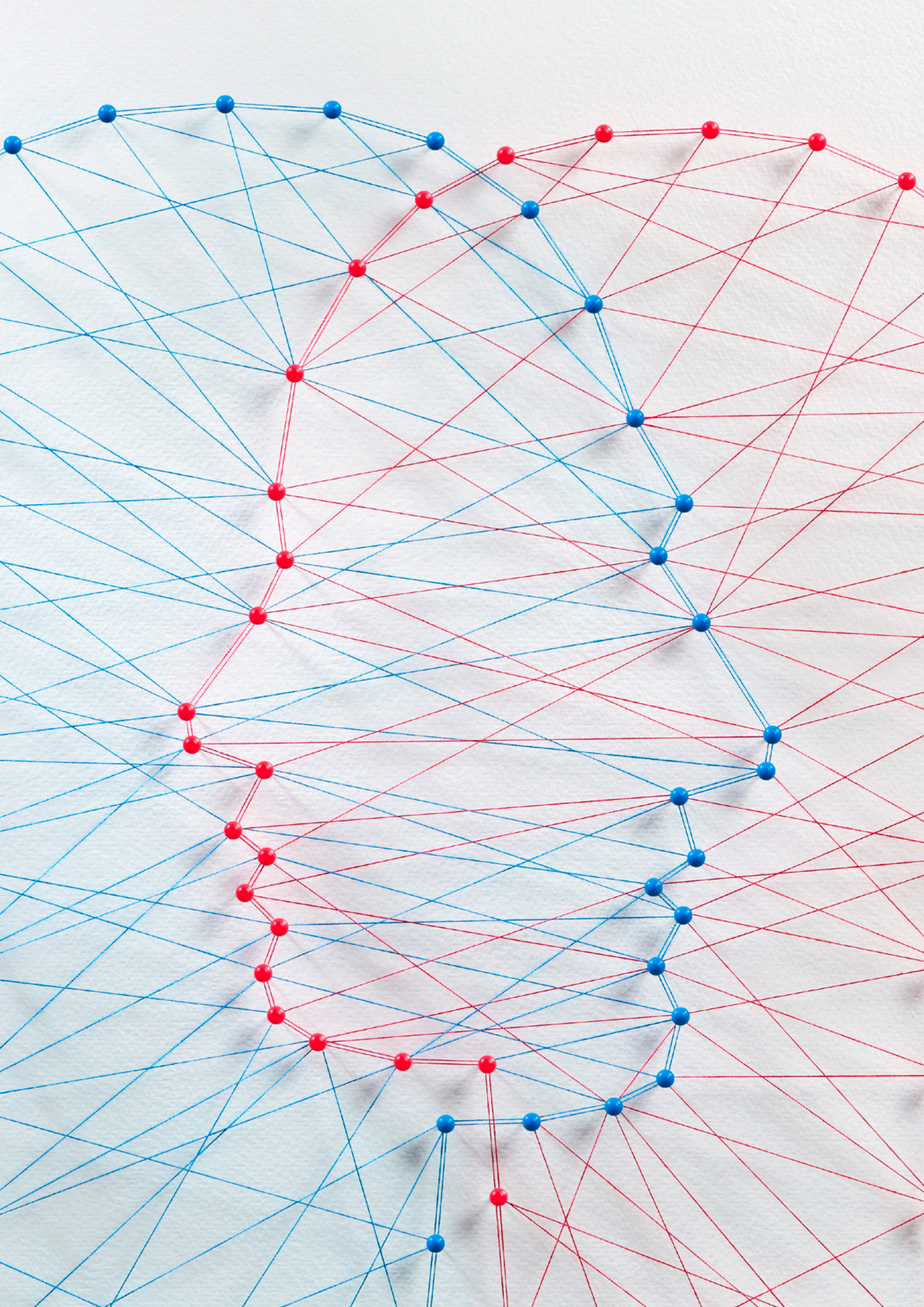
4. G16C

(Split from G06F19 in 2019; COMPUTATIONAL CHEMISTRY; CHEMOINFORMATICS; COMPUTATIONAL MATERIALS SCIENCE [2019.01]).

- a. G16C 20/70 – **Machine learning**, data mining or chemometrics [2019.01]

5. G16Z

(Split from G06F19 in 2019; INFORMATION AND COMMUNICATION TECHNOLOGY [ICT] SPECIALLY ADAPTED FOR SPECIFIC APPLICATION FIELDS, NOT OTHERWISE PROVIDED FOR [2019.01]).





Your free online legal information service.

A subscription service for legal articles on a variety of topics delivered by email.
cms-lawnow.com

The information held in this publication is for general purposes and guidance only and does not purport to constitute legal or professional advice. It was prepared in co-operation with local attorneys.

CMS Legal Services EEIG (CMS EEIG) is a European Economic Interest Grouping that coordinates an organisation of independent law firms. CMS EEIG provides no client services. Such services are solely provided by CMS EEIG's member firms in their respective jurisdictions. CMS EEIG and each of its member firms are separate and legally distinct entities, and no such entity has any authority to bind any other. CMS EEIG and each member firm are liable only for their own acts or omissions and not those of each other. The brand name "CMS" and the term "firm" are used to refer to some or all of the member firms or their offices; details can be found under "legal information" in the footer of cms.law.

CMS locations:

Aberdeen, Abu Dhabi, Amsterdam, Antwerp, Barcelona, Beijing, Belgrade, Bergen, Berlin, Bogotá, Bratislava, Brisbane, Bristol, Brussels, Bucharest, Budapest, Casablanca, Cologne, Cúcuta, Dubai, Dublin, Duesseldorf, Edinburgh, Frankfurt, Funchal, Geneva, Glasgow, Gothenburg, Hamburg, Hong Kong, Istanbul, Johannesburg, Kyiv, Leipzig, Lima, Lisbon, Liverpool, Ljubljana, London, Luanda, Luxembourg, Lyon, Madrid, Manchester, Maputo, Mexico City, Milan, Mombasa, Monaco, Munich, Muscat, Nairobi, Oslo, Paris, Podgorica, Poznan, Prague, Reading, Rio de Janeiro, Riyadh, Rome, Santiago de Chile, São Paulo, Sarajevo, Shanghai, Sheffield, Singapore, Skopje, Sofia, Stavanger, Stockholm, Strasbourg, Stuttgart, Tel Aviv, Tirana, Vienna, Warsaw, Zagreb and Zurich.

cms.law