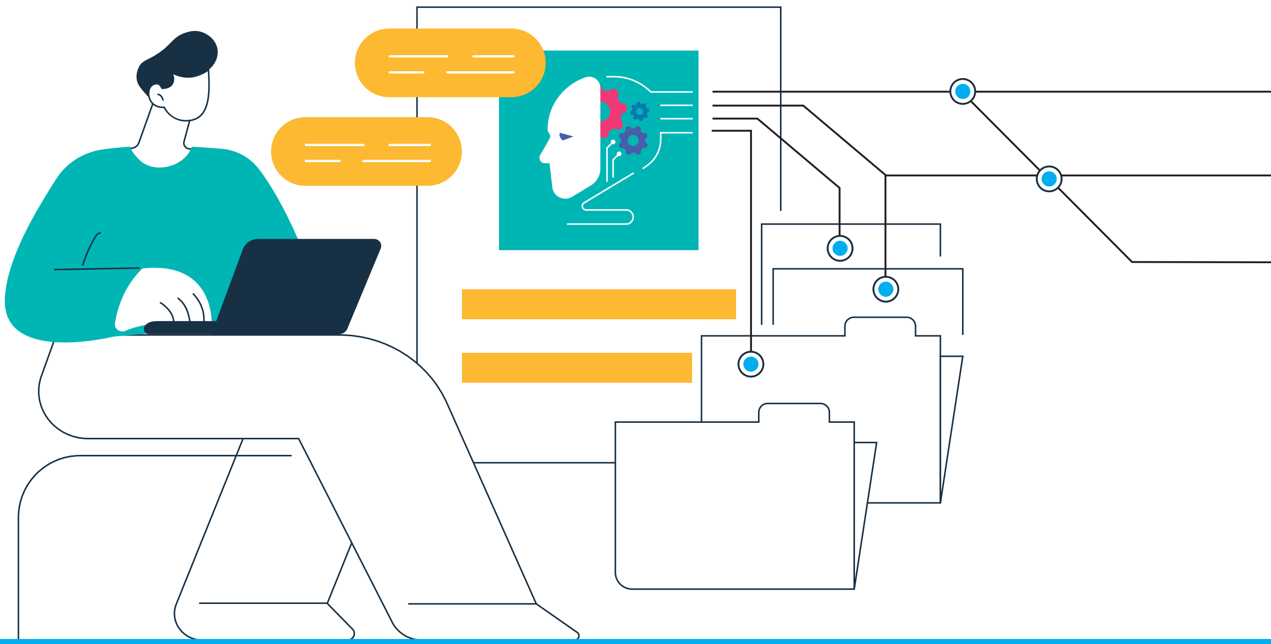
Patenting artificial intelligence

When ChatGPT, Open AI's chatbot, hit the shelves in November 2022, it got people mesmerized by its eerily human conversation skills. But the breathtaking capabilities of Artificial Intelligence (AI), in particular of the generative type such as ChatGPT, stand in baffling contrast to business-as-usual handling of patents drawn to AI tech: they are just another vernacular of computer-implemented inventions (CII). And indeed they are examined as such: that is, if you want your patent application on AI to navigate the EPO's waters successfully, you better let it sail wrapped in a »technical« application. The below will look at facets of this principle in more detail and explore other jurisdictions and the concept of technicity.



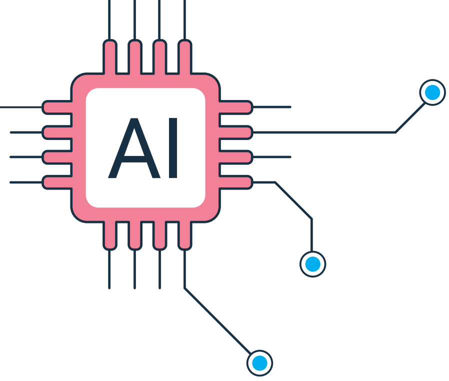
Patenting artificial intelligence

EPO case law review

The recent edition of the Guidelines for Examination in the EPO of April 2025 (in the following »Guidelines«) expressly state[1] that the use of computational models and algorithms **including artificial neural networks** does **not by itself render inventions related to artificial intelligence or machine learning non-patentable** in terms of Art. 52 EPC. Hence, if a claim related to artificial intelligence or machine learning is directed either to a method involving the use of technical means (e.g., a computer) or to a device, its subject-matter has **technical character** as a whole and is thus not excluded from patentability under Art. 52 EPC.

When assessing inventive step (Art. 56 EPC), the established »COMVIK approach« (T 641/00, G 1/19) is applied, meaning that features of computational models and algorithms themselves **contribute to the technical character** of the invention **only if these features contribute to a technical solution of a technical problem**. Specifically, an invention consisting of a mixture of technical and non-technical features and having technical character as a whole is to be assessed with respect to the requirement of inventive step by taking account of all those features which contribute to said technical character, whereas features making no such contribution cannot support the presence of inventive step. For computational models and algorithms to contribute to a technical solution of a technical problem, they may be **applied in a field of technology or be adapted to a specific technical implementation**.[2]

[1] in Part G, Chapter II, 3.3.1
[2] see the Guidelines, Part G, Chapter II, 3.3.1



Another criterion specifically relevant in the context of AI inventions is sufficiency of disclosure (Art. 83 EPC). In this regard, the Guidelines point out[3] that in the technical fields of computers, a clear description of function may be much more appropriate than an over-detailed description of structure.

For AI inventions, decision G 1/19 of the Enlarged Board of Appeal is particularly relevant despite concerning modelling and simulation rather than AI *per se*. The answer to the first question addressed by G 1/19 is that *»a computer-implemented simulation of a technical system or process that is claimed as such can, for the purpose of assessing inventive step,* ***solve a technical problem by producing a technical effect going beyond the simulation's implementation on a computer«***. The criterion *»technical effect going beyond the simulation's implementation«* is understood to mean any *»further technical effect«* going beyond the *»normal«* physical interactions between the program and the computer, namely a *»technical effect going beyond the simulation's straightforward or unspecified implementation on a standard computer system«*, such as *»technical effects on a physical entity in the real world«*, technical effects requiring *»a direct link with physical reality«*, or technical effects within the computer system or network achieved e.g., by adaptations to the computer system (margin nos. 50 and 51 of G 1/19).

In the following, recent EPO case law on inventions related to artificial intelligence or machine learning in various fields of technology is discussed, in particular decisions subsequent to G 1/19.

Selected decisions by the EPO Boards of Appeal

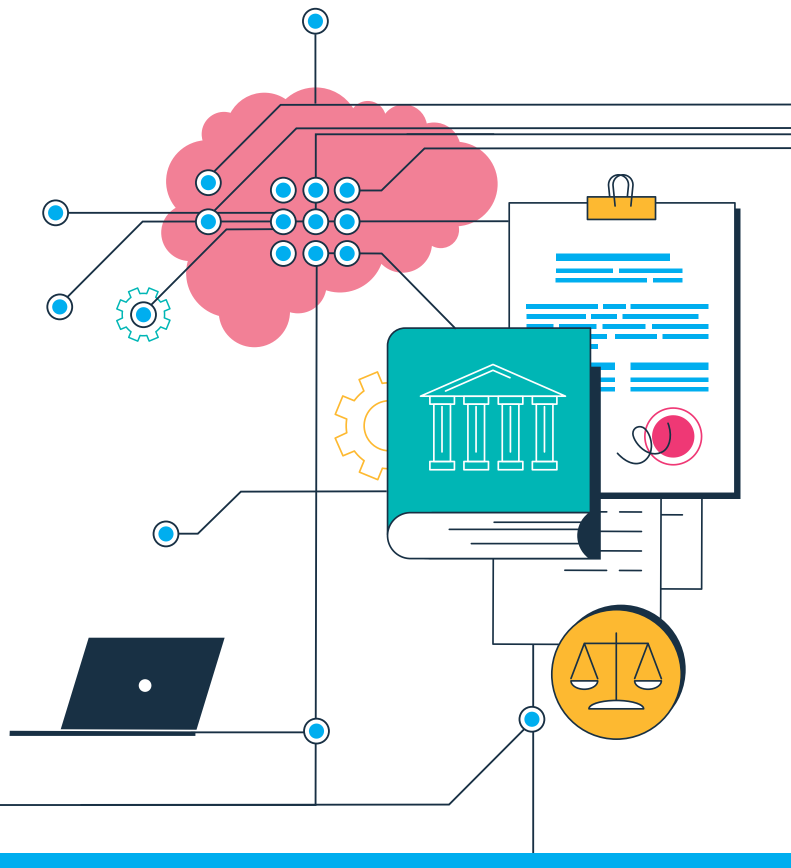
T 161/18

T 161/18 deals with sufficiency of disclosure in connection with neural networks in medical technology and states in its headnote that the invention under review relying on machine-based learning **cannot be put into practice due to lack of disclosure of training of the neural network**. Specifically, the invention underlying T 161/18 concerns a method for determining the heart rate volume, wherein a neural network provides weighting factors determined by learning. The Board found that training of the neural network cannot be carried out by the skilled person, as the application only discloses that input data shall include a broad variety of patients as to age, gender, health condition etc. (Reasons 2.2). According to the Board, **the application fails to disclose input data for training and data suitable for solving the underlying technical problem**. Thus, the Board concluded that the requirements of Art. 83 EPC are not met.

The Board also commented on inventive step: As no improvement relative to the prior art was acknowledged, the objective technical problem was formulated as an alternative to the solution shown in the prior art. In the authors' understanding, this implies that the subject matter of claim 1 was considered to have technical character. However, as the difference relative to the cited prior art was only the neural network, the training of which was not disclosed in detail, no specific technical effect was acknowledged. The alternative solution was therefore not regarded as inventive (Reasons 3.5 and 3.6).

T 702/20

T 702/20 states in its headnote that a neural network defines a class of mathematical functions, which is, as such, excluded from patentability under Art. 52 EPC. As for other »non-technical« matter, it can therefore only be considered for the assessment of inventive step when used to solve a technical problem. **Therefore, the question to be answered is whether the neural network relates only to computer programs and mathematical methods »as such«** or whether it relates to something more, and, in particular, to something that can fulfil the patentability conditions of the EPC (Reasons 10).



[3] in Part F, Chapter III, 1.

In the case at hand, the difference between the invention and D1 resided in that the different layers of the neural network were connected in accordance with an »error code check matrix«. The application referred to the reduction of the number of connections between nodes and »loose coupling« (Reasons 1.1).

Firstly, the Applicant argued that the claimed modification in the neural network structure, in comparison with standard fully-connected networks, reduces the amount of resources, in particular storage, required, and that this should be recognized as a technical effect (Reasons 14). The Board, however, did not acknowledge a technical effect since the modified network is different to a fully-connected network, and does not learn in the same way. Hence, although it may require less storage, it does not do the same thing, e.g., does not learn complex data relationships (Reasons 14.1).

Further, the Applicant argued that neural networks are artificial brains, which solve an automation problem. This argument did not convince the Board (Reasons 15 and 16). Also, the Board saw no **further technical use** (Reasons 18). On this basis, the Board concluded that the claimed subject matter lacks inventive step in terms of Art. 56 EPC (Reasons 19.1).

The Board stressed (in Reasons 20) that there can be no reasonable doubt that **neural networks can provide technical tools useful for automating human tasks or solving technical problems**. In most cases, however, this requires them **to be sufficiently specified, in particular as regards the training data and the technical task addressed**. What specificity is required will typically depend on the problem being considered, as it must be established that the trained neural network solves a technical problem in the claimed generality.

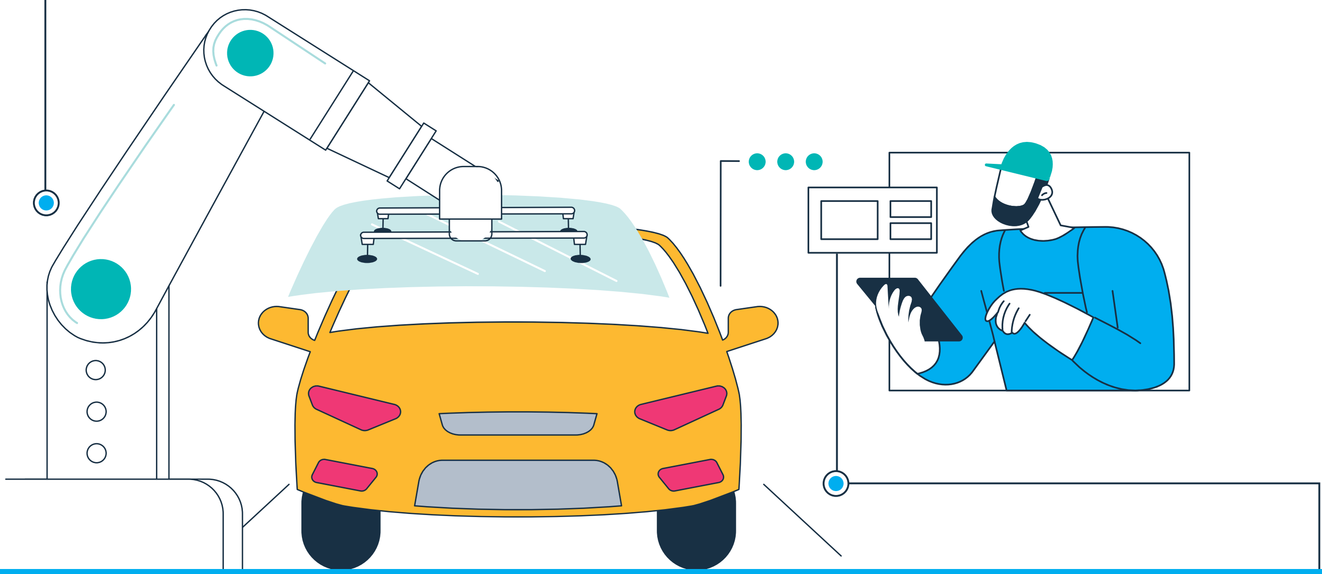
T 2082/22

The claims underlying T 2082/22 defined three distinguishing features, which were handled as partial problems for the analysis of inventive step. One of the distinguishing features defined that classification of detected environment data is performed by means of a self-learning neural network (Reasons 2.2).

The Examining Division concluded in connection with this distinguishing feature that the underlying partial problem may be formulated as the »provision of a mathematical model for the as such non-technical classification« (Reasons 2.9). Apparently, the Examining Division regarded this distinguishing feature as contributing to the technical character and forming the basis for the (partial) problem.

The Board of Appeal was not convinced by the line of argument put forward by the Applicant in the appeal proceedings that the technical effect of this feature was that the processor was specifically adapted for classification. **Rather, the Board found that, for a neural network to be regarded as a technical implementation, the configuration of the neural network needs to be specified. However, the application did not disclose the specific structure of the neural network, nor the required training data**. Here, the Board confirmed that, in line with T 702/20 the required training data needs to be indicated, but the application merely disclosed general and commonly known features of a neural network (Reasons 2.11).

The Board stated that the neural network is merely one out of a plurality of possible pattern recognition methods for classification and concluded that selecting one such known method, namely a method including a neural network, does not require inventive step (Reasons 2.12).



T 814/20

On a positive note, the **Board acknowledged inventive step in T 814/20**. More specifically, the Board acknowledged that claim 1 defining a method for re-identification of objects captured by image cameras represents a **technical purpose**. Specifically, it was regarded as tantamount to an objective measurement in physical reality (Reasons 12).

A key point underlying this case was the question whether the method provides the technical effect over substantially the whole scope of the claims (Reasons 13). Interestingly, the Board stated that the claimed method will not »work« under all imaginable circumstances, but it is probably safe to say that no computer vision method does. For instance, the claimed method may fail to re-identify objects largely changing appearance. However, in the Board's view, the skilled person understands, from the claims and the description, the kind of situations for which the method is designed and their parameters (such as illumination and geometry). The Board held that the method credibly worked over that range of situations. Hence, in the Board's judgement, this is sufficient to **satisfy the requirement that, in the present case, a technical effect was present over substantially the whole scope of the claims** (Reasons 13.4, 13.5).

The Board then concluded that it was not obvious to provide the claimed subject matter (Reasons 15).

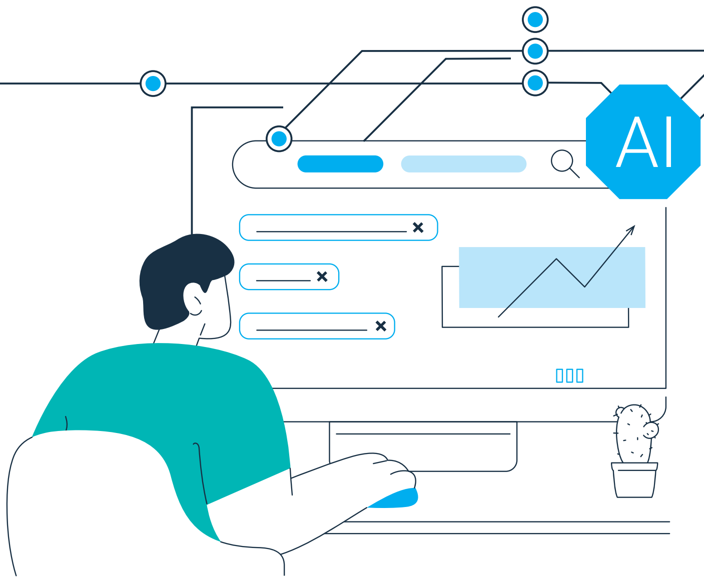
T 748/19

In the case underlying T 748/19, claim 1 was directed to training a neural network-based event identifying operation using stored camera processing data relating to the time periods of the capturing of image frames subjected to the identification of events. The description specified that the neural network may be a recursive neural network (RNN).

The cited prior art did not disclose the use of camera processing data as claimed, and on this basis, the Board acknowledged inventive step(Reasons 3 to 3.3).

Regarding Art. 83 EPC, the Board acknowledged that the skilled person was able to choose an architecture for a neural network and carry out its training if a set of camera parameters and events to be identified were defined. However, this was regarded as **insufficient to establish compliance with Article 83 EPC**, because the claimed subject matter was not characterized only by being a neural network, but also, perhaps primarily, by its purpose of being able to (reliably) identify events. This purpose must be achieved in different scenarios, also corresponding to different sets of events. The Board accepted that there were simple cases for which the claimed purpose can be achieved on the basis of the teaching in the application and common general knowledge. For instance, motion can be detected using the encoding data rate. However, it was not at all clear whether the claimed method can be successful in other, possibly more complex scenarios (Reasons 10 and 11). It was the Board's view that the claims were meant to cover at least the identification of events listed in the application, in their corresponding scenarios. For each considered scenario and corresponding set of events, the skilled person needed to define a set of parameters allowing the identification of said events. A large choice of camera parameters used for event detection were disclosed and claimed. But, according to the Board, it was not trivial to see which (or if any) camera parameters contain the information needed for identification, and the application did not provide any clear example of such selection. The Board also found no guidance provided as to which events may be detected based on which set of parameters (Reasons 14 and 15). The Board concluded that, **to carry out the claimed invention, the skilled person would have to define, without theoretical or practical guidance from the application (e.g. by way of theoretical considerations or concrete examples), in sufficient detail, as a function of the surveillance scenario, the events to be identified, the acquisition setup (e.g. lighting, perspective, resolution, etc.), and test which of the mentioned camera parameters would allow which events to be detected in the given circumstances**. The Board considered this to be an undue burden on the skilled person trying to carry out the invention. In the Board's judgement, this amounted more to conducting a research program than to carrying out the invention in accordance with the teaching provided (Reasons 8). The Board concluded that **the application did not disclose a method of identifying events using a neural network trained with camera processing data** in a manner sufficiently clear and complete for it to be carried out by the skilled person in terms of Art. 83 EPC.

Interestingly, the Board compared, in Reasons 20 and 20.1, the case underlying T 748/19 with the case underlying T 814/20 (as discussed above). In the latter decision, Art. 83 EPC was not an issue and is not discussed in the decision. The Board in T 748/19 stated that the situation in T 814/20 was different in that the algorithm was clearly defined, the theoretical assumptions were sound, and the technical effect was proven for a test scenario, so that it was clear to the Board what the generalization in the claims was meant to cover. In the case underlying T 748/19, however, the Board took the view that without a definition of the camera parameter set, this was not the case.



T 1669/21

While the remaining decisions discussed in this article refer to *ex parte* appeals, T 1669/21 deals with an opposition case in which the invention was not regarded as sufficiently disclosed in terms of Art. 83 EPC. Notably, the burden of proof to establish insufficiency of disclosure generally lies with the opponent (T 182/89, T 63/06).

Claim 1 recited a calculation model which is adapted and with which data and parameters are analyzed. The Board found that an »adapt**ed** calculation model« was not limited to a machine-learning model, as it did not need to be »adapt**ive**« in the sense of self-learning (Reasons 1.2.3). Only according to a dependent claim, was the model a neural network. Hence, the Board also regarded analytical calculation models to be encompassed by claim 1 and concluded that, for such calculation models, the patent included no examples or indications for the modelling. Accordingly, compliance with Art. 83 EPC was denied (Reasons 1.2.4).

When it came to the specific calculation model of machine learning, such as the neural network as defined in the dependent claim, **the patent did not include any description except for a mention of a neural network**. The Board stated that, in this respect, the case differs from the case underlying T 161/18 (as discussed above), in which the neural network was at least specified (Reasons 1.3.2). The common general knowledge submitted by the Proprietor in T 1669/21 may have rendered the skilled person in a position to decide which architectures and types of calculation models for what problems existed in general, but the skilled person still had to select a specifically suitable calculation model (Reasons 1.3.4). The Board further noted that the patent failed to disclose what »adaption« of the calculation model actually meant, although this aspect was not seen as decisive (Reasons 1.5). Because of the merely generalized formulation of data and parameters and the lack of examples, the Board found that the requirements of Art. 83 EPC were not met (Reasons 1.6 to 1.6.6).

The Board specifically commented on training data and came to the conclusion that also the **disclosure regarding the training data, which is vital to the success of the invention, was too general and insufficient**, so that the skilled person was not able to reproduce the invention without undue efforts or over the entire breadth of the invention (Reasons 1.7.6).

T 1425/21

The claims underlying T 1425/21 were directed to the training of a first machine learning model (being a cumbersome machine learning model) and a second machine learning model (being a distilled machine learning model) (Reasons 1, 3).

Differences relative to the cited prior art were identified as (a) the second machine learning model (distilled machine learning model) having fewer parameters than the first machine learning model (cumbersome machine learning model) such that generating output from the second machine learning model required less memory than generating output from the first machine learning model; and (b) the second machine learning model trained based upon a soft score satisfying a particular form as set out in claim 1 (Reasons 3). In the Board's view, features (a) and (b) were mathematical methods which cannot be taken into account for inventive step unless they contribute in a causal manner to a technical effect (Reasons 16), applying the »COMVIK approach«. The Applicant asserted that a technical effect is based on implementation, as the method provides for reduced memory use with »matching« (i.e., the same or equivalent) classification results (Reasons 11). However, the Board took the view that the second model may use fewer resources, but that it cannot be said to produce the same results, since many smaller models will, in fact, be considerably worse (Reasons 20). The Board stated that, in principle, it appeared possible to argue that the smaller model represented a »good« trade-off between resource requirements and accuracy, i.e., that the smaller model may be less accurate but have (predictably) smaller resource requirements. However, as the application lacked any information in that regard, **no technical effect was acknowledged** (Reasons 20.1, 20.2).

The Board also commented on Art. 83 EPC and stated, for the sake of completeness, that the skilled person was not able to provide smaller networks with reduced memory needs and equivalent accuracy with only »few routine tests« for all classification tasks (Reasons 22). Firstly, while the skilled person might have been aware of the various architectures and types of networks available from the common general knowledge, the number of these possibilities was quite large. For each of them, downsizing could have been done in different ways, e.g., by reducing the number of layers, of neurons, of weights etc.. The trial-and-error process would also have had to keep an eye on the desired trade-off between size and accuracy, which was not a simple endeavor (Reasons 23.1, 23.2). Secondly, the Board did not agree that the temperature-based training process as claimed simplified the trial-and-error process (Reasons 24). Against this background, the Board found that the application did also not sufficiently teach how to carry out in practice the invention.

T 1952/21

Claim 1 underlying T 1952/21 was directed to a machine learning system comprising first and second processing paths each comprising a feed-forward neural network. The Board found that the functioning of the computer, or the computer itself, were not adapted, and that no further technical use was implied by the claim. Thus, even if the advantages in reinforcement learning brought forward by the applicant were to be acknowledged, the Board had to conclude, on the basis of G 1/19, that the claimed system did not solve a technical problem (Reasons 24). Inventive step was thus denied.

The Board also referred to T 702/20 and noted that its conclusion in the case at hand was consistent with that in case T 702/20, which was in many ways mirrored the fact pattern of T 1952/21. In T 702/20the same Board (albeit in a different composition) decided, also following G 1/19, that a trained machine learning model, namely a neural network, can »only be considered for the assessment of inventive step when used to solve a technical problem, e.g. when trained with specific data for a specific technical task« (Reasons 24.1).

Discussion and Takeaways

Key hurdles when it comes to patenting AI inventions at the EPO are inventive step (Art. 56 EPC) and sufficiency of disclosure (Art. 83 EPC).

For assessment of inventive step, the EPO's approach to AI inventions implements G 1/19. In this regard, it may be insufficient to only mention a neural network in the application or to disclose general and commonly known features of a neural network. Rather, it is ***advisable to disclose the specific structure of the neural network and the training data to ensure that the neural network is sufficiently specified, in particular as regards the training data and the technical task addressed.***

The same holds true for Art. 83 EPC, Here, it is also advisable to not only mention a neural network, but to disclose training data, so that the skilled person is able to reproduce the invention without undue efforts and over the entire breadth of the invention.

Specifically, ***theoretical or practical guidance by way of theoretical considerations or concrete examples in sufficient detail to achieve the purpose in different scenarios, based on corresponding parameters would be beneficial in the application. It may be helpful to clearly define the algorithm, present the theoretical assumptions and prove the technical effect for a test scenario.***

While EPO case law on AI inventions is developing further, it appears that it is a more or less unsurprising refinement of the EPO's established approach for computer-implemented invention in general.

Dr. Michaela Weigel-Krusemarck
/ Partner
/ German Patent Attorney
/ European Patent Attorney
/ UPC Representative
/ Physicist

Dr. Christian Pioch
/ Partner
/ German Patent Attorney
/ European Patent Attorney
/ UPC Representative
/ Physicist

Patenting artificial intelligence

Beyond the EPO – Emotional about »Programs for Computer«

Is an artificial neural network (ANN) a *»program for a computer«*? That is the question that the UK's highest court, the Supreme Court of the United Kingdom (UKSC), considered in hearings at the end of July 2025. Ever since the UK Patent Office (UKIPO) refused Emotional Perception's (then Mashtraxx) national UK patent application, the case vacillated its way through the courts in the UK's judicial hierarchy: the UKIPO said Emotional Perception's claim to training and using an AI system (which may be referred to herein as a kind of »media file recommender system«) is a mere »program for a computer« and thus allegedly falls foul of the UK Patents Act's 1(2)(c) exclusion. On Appeal, the High Court said no, it is not merely that. And the Court of Appeal[4], in turn, felt that of course it is just that. Now the question is before their UK Supreme Court to consider.[5]

Imagine an AI that hits just the right notes: it apparently understands your emotional setup based on a piece of music (a file) which you happen to like in your current mood, and that magically comes up with another piece of music that ties in quite nicely with how you are feeling today. Such a feat is what Emotional Perception's recommender system may be thought of aiming at.

The media file recommender system at issue is an intricate piece of technology. It is built on ANNs, a mainstay of machine learning that brings to the fore much of AI's prowess, be it the eerily human conversation skills of ChatGPT, or the video and imagery generators of Midjourney and others, or in other, humbler, AI skills such as image recognition in autonomous vehicles or medical diagnostics. An ANN is »as such« a mathematical model inspired by how the brain processes information. It is astonishing in its capabilities: it has been shown by various so called »universal approximation theorems« that ANNs, even in their apparently simplest forms (»perceptron«) are capable of learning any pattern in data, no matter how wayward, given enough data and computing resources.

The media file recommender system at issue deals with music and video. For illustration on what the system does, we focus on music files (tracks) here. The system is based on using ANNs to make textual descriptions (such as reviews, etc.) of music files »converge« with descriptions of the files' spectral or other measurable features, which may be referred to herein as »acoustic data« (timbre, rhythm, etc.), i.e., the sort of thing that can be captured in a spectrogram of the audio file. Music pieces that are close to the way in which they are described in prose, in natural language, ought to have similar spectrogram features. The learning of this pattern, i.e., textual description versus spectrogram description, is enabled by another piece of technology:

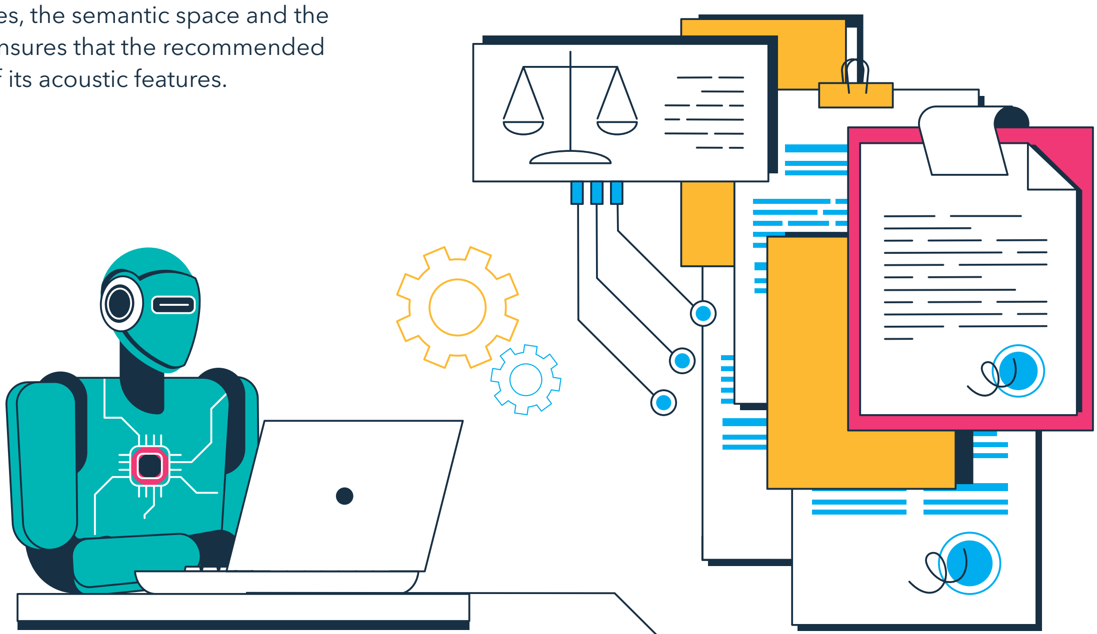
[4] [2024] EWCA Civ 825
[5] *Emotional Perception AI Limited (Appellant) v Comptroller General of Patents (Respondent)*, Case ID UKSC/2024/0131

»embeddings«. Words, sentences, paragraphs, etc., can be mapped into vectors in a suitable vector space, called »embedding space«. And meaning and composition boils down to vector arithmetic. One can add or subtract vectors that stand for texts and get meaningful answers, with vectors for texts being closer than others if their meanings (semantics) are. Embeddings are themselves another machine learning technique.[6]

Embeddings are the way textual data gets number crunched, ultimately also powering much of the LLMs (large language models), such as Open AI's ChatGPT. Using embeddings of text in a »semantic vectors space« and measurable acoustic data in a »physical property space« for given music files, the ANNs can learn their relationship in term of how close their vectors are, using geometric (Euclidean) distance. Then, given a file, it finds in a collection of existing files, the one that is closest in terms of textual description, that is, in terms of semantics. But the way in which the ANN was trained with forced convergence in both spaces, the semantic space and the physical property space, ensures that the recommended file is also close in terms of its acoustic features.

The way CII is approached (»Aerotel test«) by the UKIPO and its courts follows the EPO's former »contribution approach«, whilst the EPO's current approach on CII is defined by »COMVIK«. Here, the UK Court of Appeal, in siding with the Hearing Officer, feels that the recommender system is merely software, in its implementation (ANN) and its purpose (file recommendation). Whether the UK Supreme Court will share this view should be known by autumn 2025.

Anyone in the CII field ought to watch this space for possible knock-on effects this decision may have beyond UK jurisdiction. It may also be said that Emotional Perception's file recommender system may have enjoyed lighter music before the EPO, as the connection with the measured acoustic data might have won an EPO Examiner over. Whilst the particular UK case did not have counterparts, other applications by the same Applicant did succeed at EPO with similar subject matter on media /entertainment recommendation.



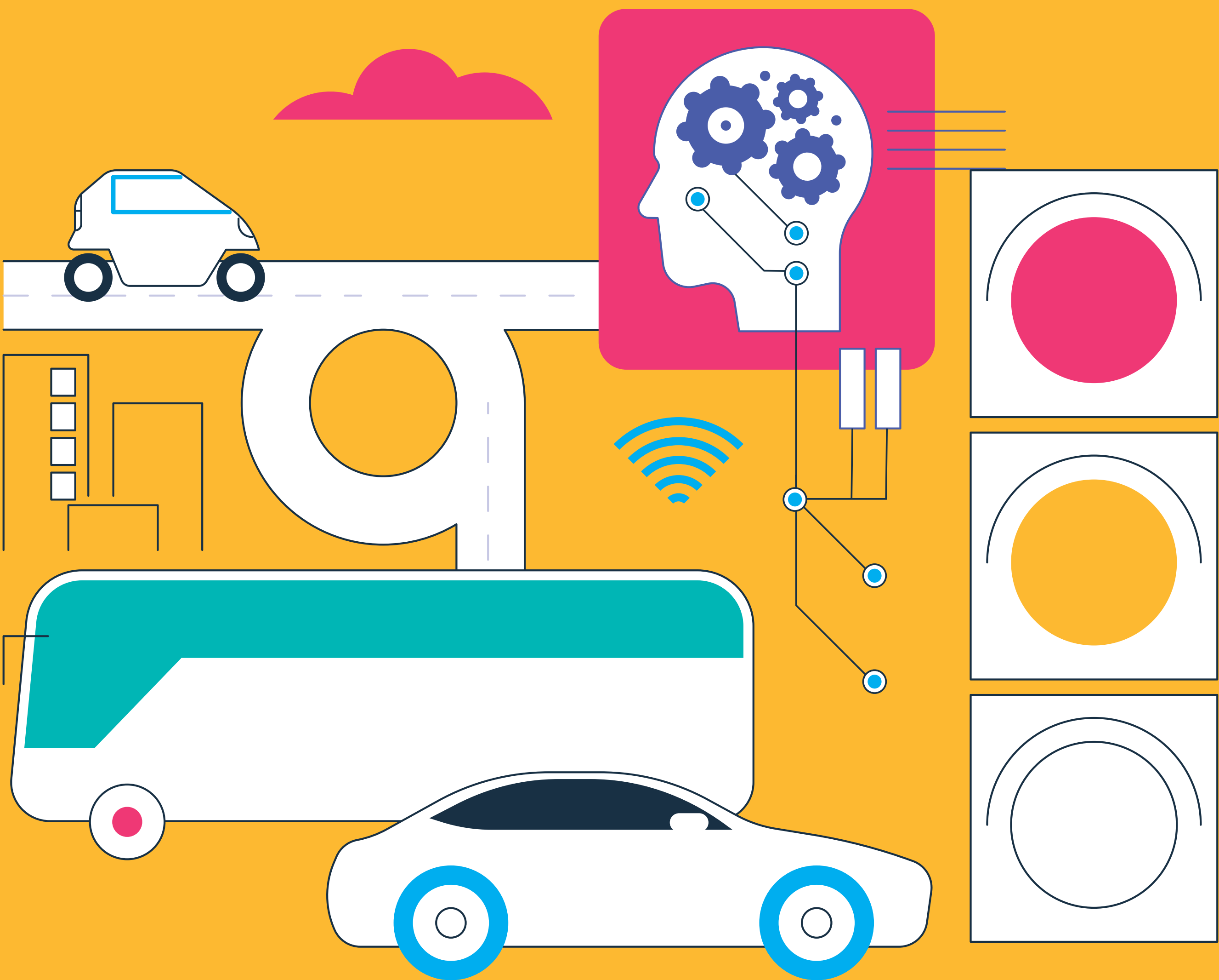
[6] As a side note: The history of science backdrop is fascinating: it was the geometry by the ancient Greeks, of the likes of Euclides, Pythagoras and Apollonius that looked at geometry in whole entities: points, lines, circles, triangles and conics. Then Descartes turned geometry into the language of coordinates, later recognized as vectors. And now it is embeddings that put geometry back the into the language – into the natural language that is!

Attila Kimpan
/ Of Counsel
/ European Patent Attorney
/ UPC Representative
/ Mathematician

Patenting artificial intelligence

Technical perspective: A classification of patentable solutions to problems from the field of automotive engineering

Ready-made artificial intelligence (AI) solutions are not practicable for each and every technical application. Often, **AI engineering** is necessary to adapt the architecture of the respective machine learning model or its utilization to the specific demands of the application. This can lead to patentable inventions.



Examples of patentable solutions to problems encountered in the field of automotive engineering

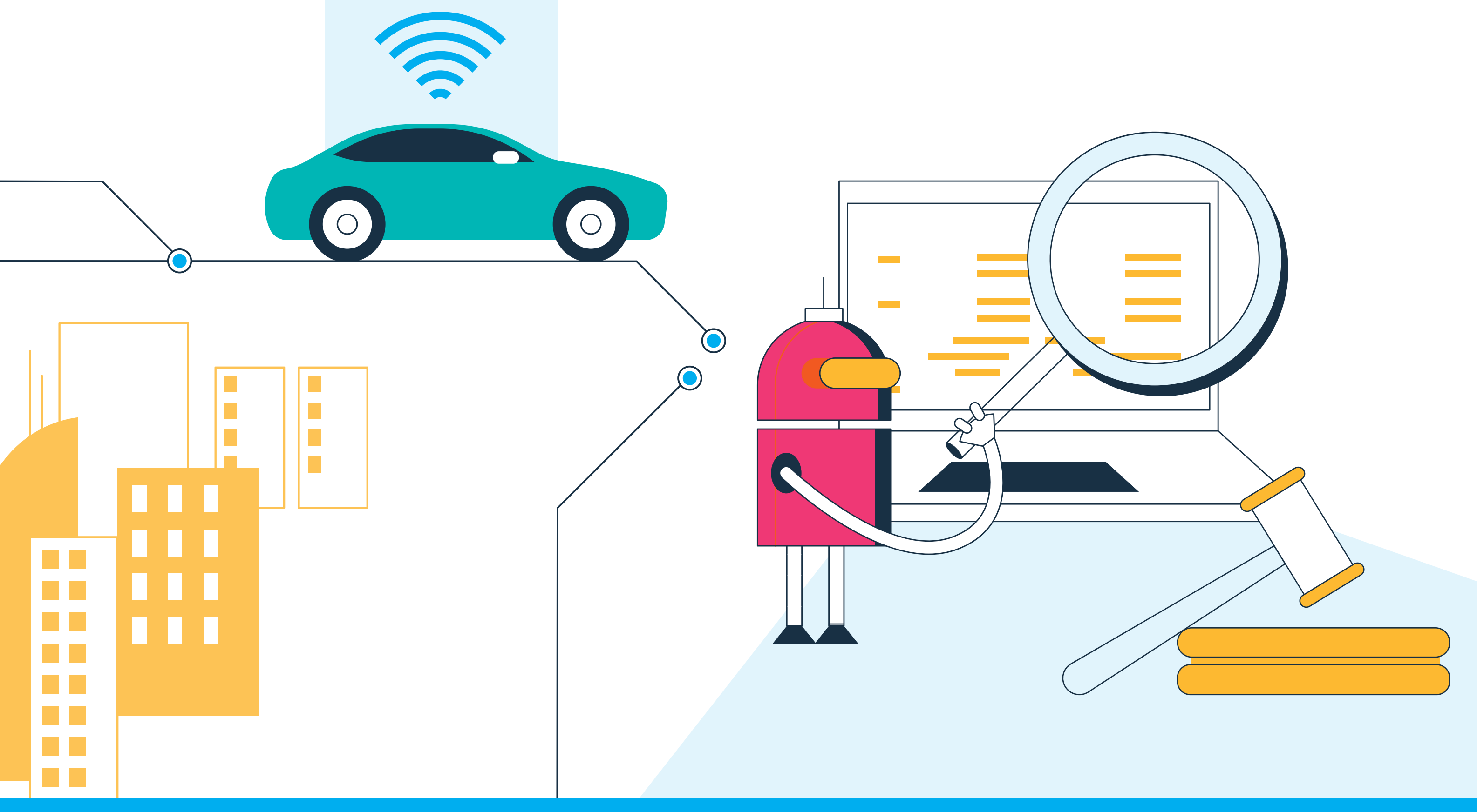
Compared to generative AI such as large language models (LLMs) for generating and editing text and generators for images and videos, AI for use in vehicles or other mobile environments must run on a comparatively spartanly equipped hardware platform. Power consumption, heat generation, overall size, and, last but not least, the cost of the corresponding control units are all relatively constrained. It is therefore desirable to simplify the architecture of a neural network while maintaining performance to the greatest extent possible. For example, weights and other parameters of the network can be quantized to one of only a limited number of possible discrete values. Individual neurons or whole parts of the network that are deemed less important can be pruned. Additionally, certain neurons can be randomly deactivated during runtime using a process known as »dropout«. A network can be composed of several parts in a »mixture of experts«, of which parts, for each input, only the one most competent for processing that particular input becomes active.

The more AI is involved in important decisions, the more important reliability and security become. Many AI models are by default a »black box« whose processing steps from input to end result are difficult to track. »Explainable AI« therefore aims to make it easy to understand exactly what an AI model bases its decision on. For example, a »saliency map« can show which areas of an image were crucial for the decision of an image classifier. In this context, it is also important to know how quickly the decision can »flip« if the input is changed. For example, manipulating a stop sign with a sticker that is barely noticeable to the human eye can cause a traffic sign recognition system to recognize a completely different traffic sign and to disregard the stop sign. The AI model must therefore be made resistant to such »adversarial examples« or at least be able recognize them.

If the data processed by AI is measurement data, these may be affected by noise or by other uncertain signals. In order to address and quantify these uncertain signals, processing can be designed to be entirely or partially probabilistic. That is, instead of specific values for variables, parameters that characterize the distribution functions of these variables are calculated. Concrete values for the variables can then be drawn from these distribution functions. To prevent small uncertain signals from steering the processing in completely different, surprising directions, batch normalization and other regularization methods can also be used.

When processing data with AI, the task of separating important information from unimportant information often arises. Autoencoder architectures can be used in such instances. These architectures use an encoder to encode the input into a representation in a latent space, whose dimensionality is usually drastically reduced compared to the input. A decoder is then used to reconstruct the original input, or another variable of interest, from this latent code. The low dimensionality of the representation forces the input through an »information bottleneck«. The encoder thus learns to extract the most important information from the input. The decoder learns to reconstruct a maximum from the little information it receives.

Language models, in particular LLMs, are increasingly used as a universal tool to find, based on a large amount of existing knowledge, as well-founded an answer to a question as possible. In such instances, the existing knowledge can be incorporated in various ways. For example, a generically pre-trained LLM can be further trained with the existing knowledge (»fine-tuning«). The existing knowledge can be fed into the LLM as additional contextual information, in light of which the answer is to be formulated. Furthermore, the LLM can be given the opportunity to query the existing knowledge itself from a database or other storage (»retrieval augmented generation«).



Another factor that plays a role in the engineering of AI applications is the question of where to obtain the necessary training data. Particularly in supervised training, where the output of the AI model is compared with a target output (»ground truth label«) and this output is evaluated using a cost function (»loss function«), the »labelling« of training examples with the corresponding target outputs is often a manual and therefore time-consuming process. Therefore, data augmentation is used to generate variations of the training examples for which a previously assigned »ground truth label« is still valid, so that no additional effort is required to »label« these new training examples. Using various domain transfer methods, training examples from a specific domain (such as images taken in summer or during the day) can be transferred to another domain (such as images taken in winter or at night) while retaining the respective semantic content (e.g., road users, roads, or other object instances). Generative adversarial networks (GANs), such as CycleGAN, are important tools for such domain transfer. If »ground truth labels« are obtained ad hoc during training, »active learning« can be used to specifically select those training examples whose labels promise the greatest learning success. Comprehensive training on training examples with sufficient variability reduces the likelihood that, when the AI model is later applied, a specific input will suddenly prove to be »out-of-distribution« that cannot be handled correctly.

Training cannot always be provided by a single entity. For one, the resources required are often too large, and for another, the training examples in many applications also contain personal data, such as recognizable faces or license plates. Here, collection by a central entity is subject to legal restrictions, especially when transmitting data across national borders. The »federated learning« approach enables training carried out by many participants with a locally available part of the training examples, such that thereafter only contributions to the training success, but not the training examples themselves, need to be collected centrally.

AI models, such as neural networks, do not always have to be retrained from scratch. They can also be trained to adopt all or part of the knowledge embodied in an existing model, for example through »transfer learning« or a »teacher-student« approach. With such approaches, for example, a »student« network with a smaller architecture that is to be used in a vehicle can adopt just enough knowledge from a much larger »teacher« network to enable it to perform the upcoming driving task.

Conclusion

The abovementioned innovative aspects of AI engineering for exemplary automotive applications require a detailed analysis of the technical aspects in order to render the specific AI models patentable. Drafting patent applications allowing for optimal patent protection often requires a deep understanding of the technical complexity of such innovations in automotive engineering and elsewhere.

Dr. Fabian Kiendl
/ Counsel
/ German Patent Attorney
/ European Patent Attorney
/ UPC Representative
/ Physicist

Dr. Christian Schäflein
/ Partner
/ German Patent Attorney
/ European Patent Attorney
/ UPC Representative
/ Physicist

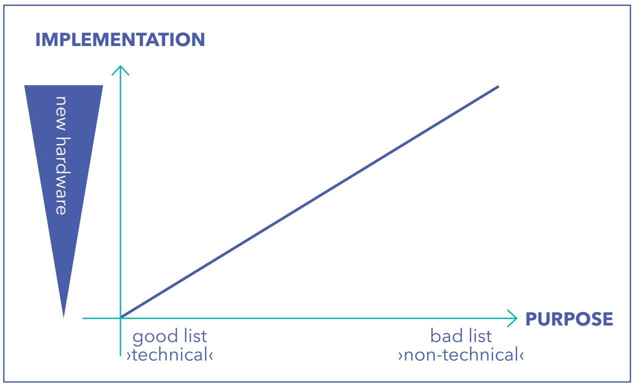
Patenting artificial intelligence

Opinion: What is technical?

Despite the quite negative rap that the AI cases of the EPO's Boards of Appeal decisions received in the case law summary in this issue, it should be borne in mind that a good number of AI related matter makes it through to grant. The EPO is by far the most friendly patent office when it comes to AI or other software or maths heavy matter. AI, despite being based on both maths and computer programs at least implementation-wise, benefits from the EPC's loop-hole language on maths and software only being excluded »as such«. And this language was then developed by the Boards of Appeal under »COMVIK« into yet more receiving terms: non-technical matter such as some maths tricks or other arguably abstract mind games are given weight if they have a tangibly technical effect when used in a technical process.

Neither the Boards of Appeal decisions nor the language of the EPC, even when put under the *»travaux préparatoires«*-looking glass, will tell us just what is it that makes something »technical«. What »technical« really is remains the Cheshire cat of patent law. Instead of a handy definition, one is given a list of examples to muse on what is said to constitute something »technical«: anything image processing, number crunching of all sorts of real world measurements (e.g., sensor data), and, for some reason, cryptography is in the EPO's »good list« of what it feels are technical problems meant to be solved by the CII. Problems that are not deemed technical problems in turn end up on the EPO's »bad list«. And this includes endeavours that sound administrative or »office like«: actuarial, payroll and commercial activities, such as putting ads out, but also, for some reason, language and text processing, documents processing, etc.

But not all is lost for non-technical problems as long as the solution claimed is »deeply« technical. Patent applications for such mattes may may still make it through to grant. It may be helpful to conceptual this in terms of a s sliding scale:



In CII inventions, the more one's claims are drawn to software or maths matters, the more the problem the CII sets itself to solve ought to be technical to be planted solidly in the EPO's »good list« land. And the more the problem is removed from something traditionally seen as technical (such us running payroll more efficiently), the more technical the implementation side should be, possibly down to modifications at the hardware level. Thus, one may get away with a patent on payroll process, but in return the EPO may want to see new machinery with new hardware features being reflected in the patent claims. This is unlikely to be the case of course, as most CIIs are on the software level.

A possible other way forward for matters drawn to solving non-technical problems is when »technical implementation« details are »back-reflected« into the hardware on which the software runs. This is illustrated in an example cited by the

[7] Part G, Chapter II, 3.3

[8] »word size«, also sometimes called »word length«, is the number of bits the computer's processor can handle at any one time.

EPO Guidelines[7]. The example goes back to a case in which a claim to a polynomial reduction algorithm (which, as such, sits deeply in the EPO's »bad list« land) was nevertheless allowed as the algorithm there had a particular adaptation that reflected the »word size« that the computer running the algorithm is using[8].

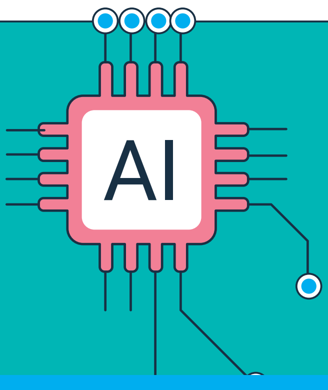
As to dancing along to »COMVIK's« tune, one may ask how there can be non-technical features that have technical effect. This can be illustrated in the case **T 1425/21** (also discussed in the EPO case law review in this issue) on reduced memory use by an ANN. For example, if the ANN there had been cloaked up in a claim for some specific control purpose in the Internet of things (IoT) gadgetry usually light on memory, it is likely the case would have made it through to grant.

The upshot of the above is that there is no escaping the long arm of »technical«: no matter how fast. Efficient, memory preserving the invention may be: but if all this might is directed at non-technical problems, you are shooting with wet gun powder.

According to a recent piece on AI by the Economist (issue September 13, 2025) in about three years' time, AI tech is going to see worldwide investments in excess of $3 trillion. This bonanza is likely to power new marvels of ingenuity: drug discovery, new materials, etc. But such burgeoning hunger for AI is likely to hit a ceiling of limited supplies in data and energy, despite the seemingly infinite cash flow. Training and running machine learning models, such as ANNs that underlie all AI, gobbles up large quantities of energy. Reviving retired atomic power plants or building new one to meet this energy need is unlikely to be the way forward, unless AI is made smart enough to tell us where the nuclear waste is to go.

New inventions around AI will likely be geared to more efficient energy use. A glimpse of this was seen through »DeepSeek« (a Chinese AI start-up) whose large language model (LLM) by the same name can have some of is artificial neurons switched off if not needed in training to save energy. Curiously, it turns out that this particular model is more power-hungry during inference (that is, the use of the trained model after training). This snapshot alone illustrates there is room galore for inventions around AI in a world such as ours, where there is simply not enough energy and data to go around. And yet, the strategies and answers for successful patent protection is the same now as it will be then: a solid foundation of enabling material, a motivating and convincing narrative on why it is thought the AI can deliver on what is claimed, and consideration of some technical application on the EPO's »good list« if the bulk of the invention is in the maths or software. In other words, what will be called for is careful, thoughtful drafting.

Coming back to the Chesire cat of what is »technical«, its elusiveness is not necessarily a bad thing if one does not want to limit the stage too much for yet unknown actors: for fields of endeavours that are at present on no-one's radar, but that may emerge in the future and that one day will be called »technical«, i.e., a »technology«. Thus, ultimately, it is down to users of the patent system, the Applicants, to push the boundary and help shape case law by filling the »negative« mold that EPC's Article 52(2) »as such« exclusions provide. And one way to do this is to file appeals and argue the case with passion.



Attila Kimpan
/ Of Counsel
/ European Patent Attorney
/ UPC Representative
/ Mathematician