This document includes some recent decisions of the EPO in 2024 with regards to software related inventions and shows relevant extracts from the respective decisions.

T 1741/22 (New medical data/ROCHE) 26-07-2024

European Case Law Identifier ECLI:EP:BA:2024:T174122.20240726

System and method for analyzing glucose monitoring data indicative of a glucose level, and a computer program product

Claims - clarity (no)

Inventive step - (no): no credible technical effect produced by deriving additional data from medical measurements

Application number:	16153964.8
IPC class:	G06F 19/00
Distribution:	DISTRIBUTED TO BOARD CHAIRMEN AND MEMBERS (B)
Applicant name:	Roche Diabetes Care GmbH, F. Hoffmann-La Roche AG
Cited decisions:	G 0001/04, G 0001/19, T 2681/16, T 1091/17, T 1910/20, T 0335/21

Board: 3.5.05

Catchword: The mere generation of further data from measurement data already collected from the human body is not a technical effect (T2681/16 and the Guidelines for Examination not followed). See Reasons 2.3 of the decision.

https://www.epo.org/boards-of-appeal/decisions/pdf/t221741eu1.pdf

Claim 1 of the main request reads as follows:

"A system for analyzing glucose monitoring data indicative of a glucose level in a bodily fluid, comprising:

- an input device (3),

- a data processing device (1),

- an output device (4),

- a display device (5), and

- machine readable instructions that are executed by the data processing device, wherein the machine readable instructions cause the data processing device (1) to

- receive continuous glucose monitoring data via the input device (3), the continuous glucose monitoring data

- indicating a glucose level sampled for a person in a bodily fluid at a plurality of sample times over a measurement time period in a continuous glucose level measurement, and

- comprising a plurality of continuous glucose profiles, each of the glucose profiles comprising a plurality of glucose values with a glucose value for each of the plurality of sample times over the measurement period;

- for the plurality of continuous glucose profiles, determine at least one of a plurality of minimum glucose values and a plurality of maximum glucose values for a selected group or each of the plurality of sample times;

- provide first display signals representing at least one of the plurality of minimum glucose values and the plurality of maximum glucose values for the selected group or each of the plurality of sample times;

- output the first display signals via the output device (4) to the display device (5); and

- display a first graphical representation according to the first display signals on the display device (5)."

1. Main request and auxiliary requests 1 to 9

1.1 The board fully agrees with the clarity objections raised in the contested decision. Claim 1 of the main request and auxiliary requests 1 to 4 is practically unintelligible without referring to the description and the drawings. Moreover, the amendments to claim 1 of auxiliary requests 5 to 9 fail to resolve this issue. Most essentially, the temporal scope and relationship of "measurement (time) period(s)" and "sample times" is not clear in claim 1 of any of these claim requests.

The appellants argued that the claims were broad but indeed clear, as the skilled person would not take into consideration "time periods" which would medically not be meaningful. However, the board sees no justification in formulating a claim so abstract that it covers a broad range of meaningless selections of measurement periods, such as one-hour periods on the same day, leaving the reader with an undue burden to speculate on the intended scope of the claims.

1.2 Therefore, claim 1 of the main request and auxiliary requests 1 to 9 does not meet the requirement of Article 84 EPC.

2. Auxiliary request 10 - Inventive step (Article 56 EPC)

2.1 Claim 1 of auxiliary request 10 contains the following limiting features (board's labelling):

(a) A system for analysing glucose monitoring data indicative of a glucose level in a bodily fluid, comprising: an input device, a data processing device, an output device, a display device, and machine-readable instructions that are executed by the data processing device,

(b) [wherein the machine-readable instructions cause the data processing device to] receive continuous glucose monitoring data via the input device, the continuous glucose monitoring data indicating a glucose level sampled for a person in a bodily fluid at a plurality of sample times over a measurement time period in a continuous glucose level measurement, and comprising a plurality of continuous glucose profiles, each of the glucose profiles comprising a plurality of glucose values with a glucose value for each of the plurality of sample times over the measurement period, wherein the plurality of glucose profiles is determined on different days by sampling the glucose level on each day over the measurement period, wherein the measurement period is 24 hours;

(c) [wherein the machine-readable instructions cause the data processing device to] for the plurality of continuous glucose profiles, determine a plurality of minimum glucose values and/or a plurality of maximum glucose values for a selected group or each of the plurality of respective sample times

(d) [wherein the machine-readable instructions cause the data processing device to] provide first display signals representing the plurality of minimum glucose values and/or the plurality of maximum glucose values for the selected group or each of the plurality of respective sample times;

(e) [wherein the machine-readable instructions cause the data processing device to] output the first display signals via the output device to the display device; and display a first graphical representation according to the first display signals on the display device.

2.2 The appellants argued that the distinguishing features of claim 1 of auxiliary request 10 over D1 were features (c) and (d), i.e. determining and displaying minimum/maximum glucose values. They stated that the technical effect of the distinguishing features was to provide an "improved analysis of glucose monitoring data". In particular, "the plurality of minimum/maximum glucose values may correspond to medically relevant outlier values", which "would otherwise be averaged out in the context of known methods employing percentiles as in D1". They emphasised that the technical effect did not lie in a mere "presentation of information" but in that "new data was generated". Accordingly, they formulated the objective technical problem as providing a system for "improved analysis of glucose monitoring data for guidance of a patient or physician".

2.3 However, the board is not convinced that features (c) and (d) contribute to the technical character of the invention.

2.3.1 The appellants persistently emphasised, also at the oral proceedings before the board, that the invention generated "new data" from glucose monitoring data. However, if the mere generation of "new data" were sufficient to contribute to the technical character of the invention, Article 52(2) and (3) EPC would contain meaningless limitations of patentable subject-matter, as e.g. mathematical methods are supposed to constantly generate "new data".

2.3.2 A subset of "new data" that might have been relevant for assessing the contribution to the technical character of the invention in the context of the case at hand could have been a new "collection" of data practised on the human or animal body. Here, the board uses the word "collection" within the same meaning as in G 1/04 referring to "(i) the examination phase [of a diagnostic method] involving the collection of data" (G 1/04, Reasons 5). More recent jurisprudence of the Enlarged Board of Appeal seems to prefer the word "measurement" (G 1/19, Reasons 85 and 99), which involves the calculation of the physical state of an object (i.e. a certain glucose level in a "bodily fluid" in the case at hand). As stated in G 1/19 (Reasons 99), it is generally acknowledged that measurements have technical character since they are based on an interaction with physical reality, such as the human or animal body (see G 1/04, Reasons 6.4.1).

2.3.3 In the case at hand, features (c) and (d) do not involve the actual measurement of the respective glucose level in a bodily fluid. Instead, they process already measured and received continuous glucose monitoring data to generate and display further "new data", namely a plurality of minimum/maximum glucose values, in order to **support a physician in their purely intellectual deductive decision phases of diagnosis and therapy**. Such **subsequent processing of certain measurement data collected from the human or animal body is "predominantly of a non-technical nature"** (ibid.). Thus, it cannot contribute to the technical character of the invention.

2.3.4 This interpretation of the Convention and of the conclusions of the Enlarged Board of Appeal have also been adopted in earlier decisions of this board (see e.g. T 1091/17, Reasons 1.8; T 1910/20, Reasons 1 and 2; T 335/21, Reasons 1.2 and 1.3).

2.3.5 However, at the oral proceedings before the board, the appellants referred to T 2681/16 and to the Guidelines for Examination in the EPO in support of their view. In particular, the appellants considered the case in T 2681/16 to be analogous to the case at hand. The competent board in that case dealt with distinguishing features related to an algorithm to process already acquired, i.e. measured, blood glucose data points. The board acknowledged that these features, when taken in isolation, were non-technical, and could support the presence of an inventive step only if they credibly contributed to producing a technical effect serving a technical purpose (Reasons 3.2.3). However, **the board then accepted the technical effect alleged by the appellant (Reasons 3.2.4), namely "providing an overall measure of the glucose variability (i.e. equally sensitive to both hypo- and hyperglycemic events) and a prediction of glycemic events that were better than, or at least alternative to, those used in [the closest prior art]**". Whereas the board concluded that this technical effect was not achieved over the whole scope of the claim in a higher-ranking request (Reasons 3.2.5 ff.), it was satisfied that this effect was achieved over the whole scope of the claim in a lower-ranking request (Reasons 6.2.1).

2.3.6 **This board is not in agreement with and therefore deviates from the interpretation of the Convention given in T 2681/16.** According to Article 20(1) RPBA, should a board consider it necessary to deviate from an interpretation of the Convention given in an earlier decision of any Board, the grounds for this deviation shall be given, unless such grounds are in accordance with an earlier decision or opinion of the Enlarged Board of Appeal according to Article 112(1) EPC. In particular, **the board disagrees with the finding in T 2681/16 that providing an overall "measure" of the glucose variability and a prediction of glycemic events amounts to a technical effect**. The board is well aware of the tendency of applicants to use the word "measure(ment)" liberally in order to give inventions the veneer of technicality. This is mainly because it is generally acknowledged in the jurisprudence of the Boards of Appeal that "measurements" have technical character. Admittedly, the applicants' use may indeed well correspond to the meaning of the word in common parlance. However, **a prerequisite for a "measurement" with technical character, within the meaning of the jurisprudence of the Boards of Appeal, is an interaction with "physical reality" for the calculation of the physical state of an object, even if the measurement may be carried out indirectly, e.g. by means of measurements of another physical entity (see G 1/19, Reasons 99)**. In the present case and in the case underlying T 2681/16, where the "physical reality" is typically the "patient's blood", the interaction with the physical reality ends once blood glucose measurements are carried out, directly on the relevant physical entity "blood", or indirectly e.g. on another bodily fluid. **The provision of overall glucose variability and a prediction of glycemic events are mathematical steps or intellectual activities which take place in the absence of this interaction with the physical reality and are therefore not "measurements" in this sense.** In other words, the taking of a sample on the patient is an interaction with "physical reality". Generating new data as a consequence of this interaction may result in "measurements" of a technical nature. **But generating (and displaying) further data by an evaluation or interpretation of these measurements (as done according to features (c) and (d) here) amounts to "measurements" generated merely by a cognitive or mathematical exercise that is inherently non-technical.**

2.3.7 As to the Guidelines for Examination in the EPO (in its applicable version of March 2022 and also in its current version of March 2024), section G-II, 3.3, which relates to the technical contribution of mathematical methods, lists

"providing a medical diagnosis by an automated system processing physiological measurements"

among "examples of technical contributions of a mathematical method". **As providing a "medical diagnosis" - whether done by a physician or by an automated system - is devoid of any technical character (see e.g. G 1/04, Reasons 5.3 and 6.3),** this example is clearly erroneous. As there is no further explanation, let alone a reference to any case law, the board sees no reason to speculate on how the Guidelines came up with this example (cf. Article 20(2) RPBA).

2.4 In view of the above, the subject-matter of claim 1 of auxiliary request 10 does not involve an inventive step (Article 56 EPC).

T 1952/21 (Reinforcement learning/BOSCH) 14-06-2024
European Case Law Identifier ECLI:EP:BA:2024:T195221.20240614
MACHINE LEARNING SYSTEM

Claims - clarity (yes)
Reinforcement learning a technical field (no)
Inventive step - all requests (no)

Application number	18174351.9
IPC class	G06N 3/00, G06N 3/04
Applicant name	Robert Bosch GmbH
Cited decisions	G 0001/19, T 1326/06, T 0702/20, T 1294/16
Board	3.5.06

https://www.epo.org/boards-of-appeal/decisions/pdf/t211952eu1.pdf

Claim 1 of the main request defines:

A machine learning system (10), comprising:

an input unit (20);

a processing unit (30);

and an output unit (40);

wherein, the input unit is configured to provide the processing unit with input data;

wherein, the processing unit is configured to process the input data to generate processing path input data; wherein, the processing unit is configured to implement a first processing path comprising a feed-forward neural network to process the processing path input data to generate first intermediate data;

wherein, the processing unit is configured to implement a second processing path comprising a feed-forward neural network to process the processing path input data to generate second intermediate data, wherein said feed-forward neural network comprises stochastic units;

wherein, the processing unit is configured to implement a value output path comprising a feed-forward neural network to process the first intermediate data and the second intermediate data to generate value output data;

wherein, the processing unit is configured to implement a policy output path comprising a feed-forward neural network to process the first intermediate data and the second intermediate data to generate policy output data; and wherein, the output unit is configured to output the value output data and output the policy output data.

The application

Background and prior art

1. The application relates to reinforcement learning. In reinforcement learning, an agent explores the environment according to a policy, determining which action the agent takes (e.g. move right) at every juncture as a function of its current state (e.g. its position in the environment). The agent receives rewards, positive or negative. In this way it can "learn" the value of the various actions and states. The goal of training is to maximize a value function which reflects the expected sum of rewards given a certain action.

2. The application builds upon the method of D1, called A3C (asynchronous advantage actor-critic). That method separately approximates the policy and value models as neural networks. The raw input (describing the environment) is preprocessed in sequence by a feed forward network (CNN - for spatial input description) and a recurrent neural network (LSTM - for time dependencies). The result is fed to the value and the policy networks.

2.1 Developments of that method, termed NoisyNet A3C in the current application, inject randomness into the training by using stochastic weights (e.g. by adding random noise or using stochastic models) in the policy and value networks. This allows for further exploration of the parameter space (see the published application, paragraph 5 and figure 2).

Contribution

3. According to the application, such stochasticity in the dynamics of the controlled system, and also a lack of training data, can lead to imperfect decisions. The application therefore proposes the use of a feed-forward intermediate layer between the LSTM layer of A3C and the policy and value networks, comprising a (standard) deterministic CNN and a CNN with stochastic units, exemplified as neurons with stochastic activation functions, working in parallel. The outputs of the two networks are concatenated and fed to the policy and value networks. Alternatively, the full intermediate layer may be stochastic. This intermediate layer is said to provide for better exploration, faster convergence, and better policies.

Main request: clarity

4. The Examining Division decided (reasons 11) that claim 1 lacked clarity as the term "stochastic unit" did not have a generally accepted meaning in the art and the application did not provide a clear definition either, in particular as to where the stochasticity "originated from".

5. In the Board's view, the skilled person would understand that a "neural network compris[ing] stochastic units" was one comprising neurons the output of which is partly determined by a stochastic element. There is no need to specify the exact origin of stochasticity (e.g. stochastic weight models or added random noise - see points 2 and 3 above)

for the scope of the claim to be clear - it covers any possible "origin". Therefore, the Board does not follow the objection of the Examining Division.

Main request: inventive step

6. The Examining Division acknowledged the differences to D1 as being those related to the intermediate layer containing stochastic units (see the decision, reasons 12.2, but also point 3 above). However, it reasoned (reasons 12.3) that they were "limited in their effect to the way how a mathematical model in the form of a neural network internally processes abstract data", so that they could not contribute to the technical character of the claimed invention. The Examining Division also stated that the claim did not "serve a specific technical purpose".

6.1 In response to the Appellant's arguments, it considered (reasons 12.4) that reinforcement learning was not actually claimed, and anyway it did not "refer to a technical field but to a machine learning approach". Also, the various alleged advantages (see e.g. point 3 above) were not derivable from the claimed matter (decision, reasons 12.6 and 12.8).

6.2 Further, the Examining Division was not convinced that the case law related to simulations or cryptography, esp. RSA, applied to the present case (see the decision, reasons 12.5 and 12.9, respectively 12.7).

The Appellant's arguments

7. The Appellant argued that the distinguishing features contributed to the technical character of the invention for the following reasons.

8. First, the system design was motivated by technical considerations of the internal functioning of the computer. A computer was a deterministic system and therefore limited to deterministic operations, and the claimed stochastic units overcame that limitation by creating the desired stochastic property within the deterministic computer (statement of grounds of appeal, pages 11 and 12; and the Appellant's letter of 14 May 2024, section 1.A.a).

9. Secondly, the claim implicitly defined reinforcement learning: from the required value and policy output paths the skilled person would understand that the claim relates to reinforcement learning (see e.g. the statement of grounds of appeal, pages 8 and 9).

10. The claimed approach brought advantages in this field due to the claimed layer comprising stochastic units. **There should not be a general requirement, neither in the field of reinforcement learning nor, more generally, in the field of artificial intelligence (AI), to provide experiments as evidence for technical effects. The established standard for establishing alleged advantages required only sufficient evidence** (in support, the Appellant referred to the case law book of the Board of Appeals, 10th Edition, Chapter I.D-4.2). The Appellant argued that logical reasoning alone could constitute the required sufficient evidence.

10.1 Accordingly, the Appellant offered theoretical considerations for the present case. The Appellant stated that stochasticity during training allowed for a wider exploration of the field of possible actions. It also allowed the optimization algorithm to escape local optima and find the global optimum. The obtained solution **balanced deterministic and probabilistic signals, and by gaining a probabilistic perspective enabled it to assess ambiguous scenarios more effectively and to generalize better.**

10.2 The corresponding scientific publication of the inventors:

Shang W., van der Wal D., van Hoof H., Welling M. (2020) Stochastic Activation Actor Critic Methods

showed that these advantages could indeed be obtained, using benchmarks typically used in the art.

10.3 Although these results were related to video games, the person skilled in the art was able to obtain these advantages for any technical field; he or she could start with the hyperparameter sets of the prior art and modify them by trial and error, until a working configuration was obtained. This was within the skills of the skilled person, who must be considered to have experience in parametrizing neural networks.

10.4 For example, the invention was grounded in a technical project, namely that of an ABS breaking system. Applying the principles of the application, the inventors were able to reduce the breaking distance significantly.

11. These a**dvantages were to be taken into account as potential technical effects in the sense of G 1/19, as they occurred when the system was used as intended, namely for reinforcement learning**.

12. **The field of machine learning, and in particular reinforcement learning, was technical.**

12.1 The Appellant argued this first by comparison with cryptography (RSA), which according to case law was a technical application. In particular T 1326/06, reasons 6.4, stated the following: "RSA was a breakthrough in the development of cryptography: RSA is regarded as the first practicable, concretely implementable asymmetric cryptosystem and is now a central component in numerous cryptographic security systems. The mathematics underlying RSA thus serves directly to solve a concrete technical problem".

12.2 In the Appellant's view, reinforcement learning was a similar breakthrough for autonomous systems, where it is "the only practicable and concretely implementable solution". It was therefore incorrect to require a limitation to a specific application. In fact **"reinforcement learning has crossed much further the border between technical and non-technical than RSA" - i.e. is more technical than the latter and more remote from mere mathematics - "as it uses many more technical aspects to achieve its purpose**. [...] both an agent and an environment of the agent are required and the agent control is learned as well as improved, while RSA only requires data in terms of public-/private keys and electronic messages". These **parallels should lead to the conclusion that also**

reinforcement learning is a technical field, even if the use of reinforcement learning is not the same as that of RSA.

13. The Board suggested in its preliminary opinion (see also below) that the **appropriate starting point for the assessment of the technical character and potential technical contributions of reinforcement learning was the decision G 1/19**, in which the Enlarged Board of Appeal had made several observations on the examination of computer-related inventions in general, in particular that a specific technical purpose may be needed to establish a technical effect.

14. In response, the Appellant argued the following in its letter of 14 May 2024.

14.1 It might be correct that G 1/19 required a specific technical purpose to establish a technical contribution. But this requirement, strictly applied, would mean that some earlier case law finding certain fields (such as RSA) to be "patent-eligible" would no longer be applicable. The **Appellant questioned whether this development was "aligned with a teleological interpretation of the EPC", in particular when "the understanding of society of the term technical or technology becomes broader over time due to the exponential technologic[al] advancement (also referred to as technological evolution)" and, at the same time, "the case-law of the EPC steadily narrows its understanding" of technology**. The Appellant also asked: "Is this in line with the original intention of the EPC, or is it contrary to the idea of the EPC?"

14.2 In the oral proceedings the Appellant argued that the **EPC was old and written with traditional, for instance mechanical, inventions in mind.** It was understood that such inventions existed to make people's lives easier, for instance by supporting or taking over manual tasks. Nowadays **software implementing artificial intelligence (AI) often has the same purpose, albeit emulating a different class of human capabilities**, and this trend would intensify in the future. **Although AI methods indeed relied heavily on applied mathematics and (big) data processing, they were applicable in many technical fields and thus of independent value.**

14.3 **AI inventions therefore deserved patent protection, which was also desirable in order not to discourage their publication, which was beneficial for the public.**

14.4 Moreover, it was **a question of fairness how narrow a technical application or purpose as required by G 1/19 would have to be**. **Limiting the protection granted to an AI invention to a very specific technical application did not provide fair protection, if it relied on ideas which are broadly applicable.** This was the case here because the invention was an improvement of reinforcement learning which was generally applicable, e.g. to cars and robotics.

14.5 In summary, the Appellant asserted a disconnect between the patent system and the real world. It argued that **"everyone in the real world" would acknowledge AI or machine learning as "technical" and that the case law needed to recognise this evolution of technology**.

15. The distinguishing features therefore had to be accepted as solving a technical problem. They were also not disclosed or rendered obvious by the cited prior art. Hence the claimed invention involved an inventive step.

The Board's opinion

16. The Appellant's allegation that stochastic units overcome the limitations of the "deterministic" computer, goes beyond reinforcement learning and relates to a computer in general.

17. The Board remarks that pseudorandom number generators were known to the person skilled in the art. Their use, in general or in the more specific context of "stochastic units" (which the Appellant acknowledged to be known in the art), does not change in substance the computer, which remains as "deterministic" as any conventional computer. So the Board **cannot see a contribution on this level**.

18. On the more narrow level of reinforcement learning, the non-deterministic behaviour of the claimed system is considered below.

19. The Board agrees with the Appellant that the skilled person would understand the claimed system to be one "for", i.e. meant to be used in, "reinforcement learning". The Appellant submission is, in a nutshell, that this field is technical and that the claimed invention makes improvements in this field.

20. The system for reinforcement learning as claimed is a neural network, comprising various sub-networks, implemented on a computer. The network, as a whole, defines a mathematical function mapping inputs into outputs. Effectively, the claim is to a mathematical method implemented on a computer.

21. Considering this, the Board holds that the Enlarged Board decision **G 1/19, addressing the patentability of computer-implemented mathematical models for simulation, should be the starting point** when assessing the technical character of reinforcement learning. It is commonly accepted that a large part of the findings in G 1/19 apply to any computer implemented inventions.

22. In G 1/19, the Enlarged Board of Appeal stated (reasons 137) that (simulation) models by themselves are not technical but that "they may contribute to technicality if, for example, they are a reason for adapting the computer or its functioning, or if they form the basis for a further technical use of the outcomes of the simulation". However, "such **further use has to be at least implicitly specified in the claim**".

23. The implied use of the system in reinforcement learning requires, as the Appellant argued, an agent acting in an environment (see point 12.1 above). However, the agent and its environment need not exist in the real world, and can be completely virtual, e.g. part of a simulation model (a simulated agent acting within a simulated environment) or even a completely imaginary video game. The Board notes that both the prior art (see D1, section 5.1) and the scientific paper corresponding to the application referred to by the Appellant

present results on video games. **The concept of reinforcement learning in general does not imply a technical context.**

24. The Board has already explained above that the **functioning of the computer, or the computer itself, are not adapted**. A **further technical use is also not implied by the claim.** So, even if the advantages in reinforcement learning brought forward by the Appellant were to be acknowledged (which is not the case, see below from point 32 on), the Board must conclude, **on the basis of G 1/19, that the claimed system does not solve a technical problem**.

24.1 This conclusion is consistent with that in the case **T 702/20**, which is in many ways similar to the present one, where this Board (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"** (T 702/20, Catchword; see also reasons 12 and 17 to 19).

25. The Appellant also argued that reinforcement learning was technical based on an analogy with the case law regarding cryptography, in particular RSA (see 12 above).

25.1 The Board notes that, as the Appellant also acknowledged, notwithstanding certain similarities, RSA and reinforcement learning are different and serve different purposes. In particular, RSA and other cryptographic methods have a specific, and at least implied, **purpose, namely data security**. This is not the case for reinforcement learning. So the findings regarding RSA cannot directly be transferred to reinforcement learning.

25.2 It is therefore immaterial for the present decision whether individual Board of Appeal decisions relating to RSA are still applicable after G 1/19 or whether, as the Appellant seemed to imply, they are now wrong, i.e. "bad law".

26. The Appellant's opinion that decision G 1/19 has narrowed the scope of patentable subject matter and that this is in conflict with the evolution of technology and with a teleological interpretation of the EPC is noted. However, before the Board can deviate from the interpretations or explanations of the EPC given in G 1/19 it has to refer a question to the Enlarged Board (Article 21 RPBA). The Appellant did not propose a question to be referred, nor did it request that a suitable question be referred.

27. The **Board itself sees no reason to deviate from G 1/19 in the present case.**

27.1 The Appellant's argument that it should be possible to patent mostly abstract, mathematical inventions without a limitation to a specific technical application if they are generally applicable and have practical utility for a wide range of new products, may, from a business perspective, be a legitimate one. Although it may be assumed that the Appellant would find substantially less desirable an equally broad patent when held by a competitor.

27.2 But it was the lawmaker's choice to exclude from patentability, albeit only "as such", mathematical methods and programs for computers (see Articles 52(2) and (3) EPC).

27.3 **Mathematical methods have always been generally applicable (e.g. Pythagoras' theorem used to calculate distances) and been applied in many new - and undoubtedly technical - inventions. This did not prevent the legislator to list mathematical methods amongst the things which, as such, are not to be considered inventions. The fundamental nature of mathematical methods and their wide applicability may in fact have been a reason for excluding them from patentability.**

27.4 The Board accepts that **the use of the term technical in the case law of the Boards of Appeal may differ from its use elsewhere in society**, especially from its colloquial use. However, this does not mean that the Boards of Appeal interpret the law incorrectly: it is common place that the legal interpretation of a term may differ from its colloquial meaning. In particular, **the Boards use the term "non-technical" to denote matter excluded under Article 52(2) and (3) EPC. Any alternative interpretation of the terms "technical" and "non-technical" can only be used to justify the patentability of subject-matter to the extent that it does not contradict the law, in particular the exclusion of mathematical methods.**

T 1710/23 (Deep learning GPU/INTEL) 05-08-2024

European Case Law Identifier ECLI:EP:BA:2024:T171023.20240805

COMPUTE OPTIMIZATION MECHANISM FOR DEEP NEURAL NETWORKS

Claims - clarity (no)

Application number 18163807.3
IPC class G06F 9/50 , G06F 8/41, G06T 1/20
Applicant name INTEL Corporation

Board 3.5.06

https://www.epo.org/boards-of-appeal/decisions/pdf/t231710eu1.pdf

Claim 1 of the main request reads as follows:

"An apparatus to facilitate compute optimization, comprising:

a compute mechanism (610);

a central processing unit, CPU, (612) including one or more processors;

a graphics processing unit, GPU, (614) including a plurality of processing units (700(a)-700(n)) each comprising a plurality of execution units, EUs (705, 706, 707), wherein the plurality of EUs (705, 706, 707) comprise a first EU type (705(a)-705(n)) and a second EU type (706(a)-706(n));

wherein the compute mechanism (610) transmits software hints to the graphics processing unit (614);

wherein the hints indicate that the graphics processing unit (614) is to power down, or bypass, higher power EUs (705, 706, 707) if processing of instructions requires less processing intensive EUs (705, 706, 707); wherein

the GPU (614) is implemented to process matrix operations in deep learning applications, wherein the processing units (700(a)-700(n)) being included within memory to eliminate data transfers related to the deep learning matrix operations."

1. The invention

From claim 1 it can be inferred that the application aims to increase the efficiency of graphics processing units (GPUs) for deep learning applications.

To this end, the GPU of the present application includes a plurality of processing units, each comprising a plurality of execution units (EUs), wherein the plurality of EUs comprise a first EU type and a second EU type, the compute mechanism transmits software hints to the graphics processing unit indicating that the GPU is to power down, or bypass, higher power EUs if processing of instructions requires less processing intensive EUs, and the GPU is implemented to process matrix operations in deep learning applications, wherein the processing units are included within memory to eliminate data transfers related to the deep learning matrix operations (claim 1).

2. Clarity; Article 84 EPC

2.1 In the annex to its summons (point 6), the board had raised a number of clarity issues. Only the issues raised in points 6.3 to 6.7 were discussed at length during the oral proceedings. These form the basis for the board's conclusion regarding clarity of the main and auxiliary requests as formulated below. The other issues are not relevant to the board's conclusion and are not dealt with in the present decision.

2.2 It is not clear from claim 1 of the main request what an execution (EU) "type" is meant to refer to, or what is the impact on the claimed apparatus of the existence of two EU types, especially since no reference to those types is made later in the claim.

The appellant pointed out in its response to the summons (points 17 to 20) that the skilled person will read in the remainder of claim 1 that there are "higher power EUs" and "less processing intensive EUs", in claim 5 that the "first type" is selected to process a first type of application workload and the "second type" to process a second type of application workload, and in the description [00154] that the EU types may differ in the number of threads that may be processed, the number of registers per thread, or any other processing characteristic. For instance, 3D applications may require a larger number of threads and smaller thread register space, while media applications may require a smaller number of threads with larger register space.

During the oral proceedings, the appellant additionally observed that according to the description [00155], EU types or configurations may be designed for specific deep learning use models.

According to the board, the above examples merely give a general idea to the skilled person what said EU types could be. They are **not a substitute for a definition of those types in the claim itself**.

In particular, the claim leaves open whether the "EU types" correspond to the "higher power" and "less processing intensive" EUs mentioned later in the claim. The description par. [00155] cited by the appellant would rather tend to indicate that the types are not necessarily linked to power consumption or processing intensity. Indeed, no connection exists between such values and particular deep learning use models.

2.3 It is not clear according to which criteria EUs are considered to be "higher power" or "less processing intensive". This is in particular true if EU types are merely designed for specific deep learning use models (description [00155]; see the above reasoning).

2.4 The feature according to which the compute mechanism transmits software "hints" to the graphics processing unit, wherein the "hints" indicate that the graphics processing unit is to power down, or bypass, higher power EUs if processing of instructions requires less processing intensive EUs, is not clear.

Firstly, the context to be considered when deciding which kind of EUs the processing of instructions requires is not clear. Is the context one of individual instructions, tasks/functions, threads, or something else?

Secondly, it is not clear whether the compute mechanism or the graphics processing unit takes such decision. When taking the claim literally, the hints merely are an indication for the graphics processing unit to power down or bypass higher power instructions if processing of instructions requires less processing intensive EUs, which would seem to leave it to the GPU to make the decision whether less processing intensive EUs are required.

Thirdly, the process which would allow to make such a decision is not clear. Is the decision already made in advance and entered in some table, or is some kind of monitoring performed to allow making the decision, and if so which kind of monitoring?

Fourthly, the nature of the "hints" is not clear. During the oral proceedings, the appellant declared that the term should be understood as "instructions".

Fifthly, the consequence of the transmission of the "hint" is not clear. The claim leaves it for instance open whether or under which conditions the GPU actually powers down or bypasses higher power EUs after having received a corresponding indication in the form of a "hint".

2.5 It is not clear what it means for processing units to be included "within" memory.

The appellant explained during the oral proceedings that the processing units are included in the channels as illustrated in figure 7B of the application. The application however provides no detail about a structure of the channels which would make such inclusion possible. In its simplest form, a channel would actually be nothing more than a conducting wire.

The appellant stressed that the "channels" in case of for instance HBM memory would be a substantially more complex structure than a simple wire. However, even when considering the particular case of HBM memory, which is not part of claim 1 in the main request, it is still not clear what a positioning of the processing units within the channels would mean in practice.

2.6 It is furthermore apparent that data transfers related to deep learning matrix operations will not be "eliminated" because of an inclusion of processing units within memory. Data transfers would still be necessary, even if the path for such transfers were shortened.

2.7 The **main request therefore does not satisfy the requirements of Article 84 EPC (clarity).**

T 0323/21 (Non-decreasing sequence determining device / Nippon Telegraph and Telephone Corp.) 06-08-2024

European Case Law Identifier ECLI:EP:BA:2024:T032321.20240806

NON-DECREASING SEQUENCE DETERMINING DEVICE, NON-DECREASING SEQUENCE DETERMINING METHOD, AND PROGRAM

Patentable invention - technical and non-technical features
Patentable invention - mathematical method

Application number	15849344.5
IPC class	G09C 1/00, G06F 9/44, H04L 9/28
Applicant name	Nippon Telegraph and Telephone Corporation
Cited decisions	T 0154/04

Board 3.4.01

https://www.epo.org/boards-of-appeal/decisions/pdf/t210323eu1.pdf

Background of the invention

1. The invention relates to an applied cipher technique and, in particular, to a method for determining whether a nondecreasing sequence exists, without revealing input data (paragraph [0001] of the published application).

2. There is a method, known from the prior art, called secure computation, for computing encrypted results from encrypted data without decrypting any of the encrypted data. Encryption is performed that distributes pieces of a numerical value among a plurality of secure computers which cooperate to perform a computation in such a manner that the result is distributed among the secure computers without reconstructing the numerical value, that is, with the result and the original value being kept encrypted (paragraph [0002]).

3. According to a known method of pattern matching for character sequences on secure computation, this is accomplished by evaluating a non-deterministic finite pattern sequence, character by character, in an input text.

4. The process for determining whether a text matches a pattern after positions of partial character strings in the pattern have been identified can be abstracted to the problem of determining whether a non-decreasing sequence can be created, by selecting elements one by one, from each set of a sequence of sets. The **invention is thus about determining whether such a non-decreasing sequence can be identified** (paragraph [0006]). It seeks to perform pattern matching in a way that is compatible with encryption.

5. The technique according to the present invention determines, in O(log(m)) rounds, whether or not a nondecreasing sequence exists, by selecting elements one by one from each of m sets (paragraph [0008]). The device and method according to the invention seek efficiently to determine whether such a nondecreasing exists, thus enabling efficient pattern matching for texts.

Inventive step - technicality

6. The device of claim 1 according to the main request consists of a combination of functional units that cooperate to assess whether a nondecreasing sequence can be created from a sequence of sets of numbers. This is achieved by selecting elements, one by one from each set in the sequence. The claim does not contain any reference to any concrete use of the result. The various functional units of the device are defined by their mathematical roles in the determination of whether or not there is a nondecreasing sequence.

7. The claimed invention is a device. It therefore has technical character, as required by Article 52(1) EPC.

8. As recalled in decision T 154/04, Estimating sales activity / DUNS LICENSING ASSOCIATES, OJ EPO 2008, 46, inventive step (and even novelty) can be based only on technical features. "Non-technical features, to the extent that they do not interact with the technical subject matter of the claim for solving a technical problem, i.e. non-technical features 'as such', do not provide a technical contribution to the prior art and are thus ignored

in assessing novelty and inventive step." Moreover, "[f]or the purpose of the problem-and-solution approach, the problem must be a technical problem which the skilled person in the particular technical field might be asked to solve at the relevant priority date" (cf. T 154/04, OJ 2008, 46).

9. This implies that the mathematical functionalities defined in claim 1, to the extent that they do not interact with technical features to produce a technical effect, cannot justify the existence of an inventive step.

10. The technical nature of the claimed subject-matter is limited to the existence of the software code running on the claimed device.

11. In the absence, in claim 1, of reference to any concrete technical use or any physical entity required by the claimed device to carry out said process, **no technical contribution resulting from the various mathematical operations running on the device can be identified**.

12. The **algorithm underlying the claimed device is exclusively of a mathematical nature**. It is without any interaction with the device on which it operates. The technicality of the claimed device lies only in its materiality. Concretely, its technicality is limited to the combination of software code with the associated processing unit. The claimed device with its program code implements a purely mathematical method deprived, in its generality, of any technical purpose.

13. The applicant contested this view and underlined that the recited device and method were used for determining whether a text matched a pattern as in paragraph [0005] of the published application. They further stressed that the disclosed approach had the advantage of allowing a large number of merges to be carried out in parallel, thus reducing the number of rounds to achieve a result.

14. The applicant's arguments did not persuade the Board.

15. Independent claim 1 does not contain any concrete reference to a use of the claimed device and method. Therefore, independently of the fa**ct that text matching is not technical in itself, no technical contribution can be derived from any specific use of the claimed invention if it cannot be derived, explicitly or implicitly,** from the claimed subject-matter.

16. While it is acknowledged that the reduction of the number of rounds to obtain a result can define an advantage over similar approaches, it is noted, **in the absence in the claims of indications regarding the strategy to be followed by appropriately selecting the various indices, i, j, and k, that the limitation regarding the reduced number of merges cannot be derived from the present wording.**

17. More fundamentally, the effect put forward by the applicant is also not sufficient to confer technical character to a method which is mathematical by nature. In the absence of interaction with the device on which the algorithm is to be run, the recited method does not add to its technical character.

18. Hence, the subject-matter of claim 1 of the main request does not define a technical solution to a technical problem. Hence, it is not inventive.

19. Claim 1 of auxiliary request 1 includes additional features regarding the "merging part (50)", which have the effect of limiting the number of merges needed to reach a result. However, the reason developed above regarding the absence of technicality of the claimed method, also apply to this claim. The amendments are not sufficient to define a technical solution to a technical problem. Hence these claims are not inventive.

20. The objections raised above with regard to the main request are not affected by the statement that m is an integer equal to or larger than 3, instead of 2 in the previous requests. The objections apply mutatis mutandis to the claims of the second auxiliary request which are thus not inventive.

T 0279/21 (State-transition-controlled processing of objects/SWISS RE) 30-01-2024

European Case Law Identifier ECLI:EP:BA:2024:T027921.20240130

SYSTEM AND METHOD FOR STATE-TRANSITION-CONTROLLED PROCESSING OF OBJECTS

Inventive step - workflow rules controlling tasks and tags labelling the states of the tasks (no Inventive step - not technical)

Application number	14734190.3
IPC class	G06Q 10/06
Applicant name	Swiss Reinsurance Company Ltd.
Cited decisions	G 0001/19, T 0894/10

Board 3.5.01
DISTRIBUTED TO BOARD CHAIRMEN (C)

Catchword
The appellant considered that ... when [G 1/19] , e.g. at reasons, point 51, states that any technical effect going beyond the implementation of the process on a computer may be considered for inventive step, it means anything beyond a 1:1 mapping between the implementation and a step of the business method being implemented. In other words, any subject-matter that does not "map" to a step in the business method is technical.

The Board agrees that the "implementation" of a business method implies some sort of mapping between non-technical steps of the business method and their technical realisation. Decision G 1/19 has something to say about this mapping, at least in the forward direction, at point 51, when it rephrases the requirement for technical effect as "technical effect going beyond the simulation's straightforward or unspecified implementation on a standard *computer system". Thus, even a 1:1 mapping might be inventive if it is not "straight-forward" (e.g. not standard programming or routine modification of the technical means used), or "unspecified" (e.g. not simply as "means for [carrying out the step]").*

But, looking for a mapping from "implementation" to the step of a business method in the reverse direction does not make sense as the steps of the non-technical activity do not have to be specified explicitly. They would include any steps that the business person would come up with in a non-technical workflow. The way this is handled is by considering the mapping of the implementation to the effect of the step and to examine whether the effect has any technical character, or whether it would be covered by what the business person would consider as part of the non-technical process. This is, in other words, the standard COMVIK approach where one looks at the effect of a feature in order to pose a technical problem, which might simply be the implementation of the feature, for which the above-mentioned mapping in the forward direction meant in G 1/19 applies. (See Reasons 2.18)

https://www.epo.org/boards-of-appeal/decisions/pdf/t210279eu1.pdf

1. Background of the invention

1.1 The invention relates to a central control system for providing automated real-time interaction and state-transition-controlled processing of (data) objects, see page 1, first paragraph.

1.2 Traditional workflow systems comprise at their core a workflow execution engine which controls and monitors the processing of objects. It steers a sequence of activities (work tasks), interactions and signalling with execution devices or means, or in interaction with users or IT resources, as well as rules controlling the progression of processes through the various stages associated with each activity, page 2, second paragraph.

1.3 In practice, workflow execution engines are rarely able to accurately or completely execute all the steps of the process by means of the workflow system alone and human intervention is required, in particular to gather all information needed to decide the next steps of further processing. This is even more complex when the central workflow execution engine is controlling decentralised units, see bridging paragraph, pages 2 and 3.

1.4 Another problem is that it may be necessary to adapt the processing by steps which are not predictable at the beginning of the workflow and which may depend on environmental parameters or parameters of execution devices or other state parameters of a certain work flow state. Prior art systems use interpreters which translate possible process steps into a computer operation code for execution, but this requires resources and time, page 4, 1st paragraph.

1.5 The invention is said, rather generally, to provide a system which is capable of flexibly capturing the external and/or internal factors that may affect the processing of an object within a workflow and that is more capable of being operated by externally or internally occurring boundary conditions or constraints. Furthermore it is able to react dynamically to changing environmental or internal conditions or measuring parameters that are possibly not known or predictable at the beginning of the workflow process, in particular without human interaction, page 4, 2nd paragraph.

1.6 The solution of the invention is a state-transition-controlled processing of objects wherein a selected object 71, 72, 73 is processed from one process state 121, 122 to a subsequent process state 122, 123 by executing the process tasks 131 assigned to the process state of the object, see Figure 1, by the control system. The state transition of the object in the process flow is controlled based upon the operating parameters of an assigned operating tag 132, see page 17, 2nd paragraph. These operating parameters can be changed by authorised assigner units or assignee units. The application explains that a dedicated signalling is done to associated run-time execution modules 50, 51, 52 which serves as means for executing the activated process tasks based on the transmitted control and steering signalling.

2. Article 56 EPC

2.1 The examining division in summary argued, see reasons, point 2.1 of the impugned decision, that the claimed subject-matter related to abstract information modelling concepts at meta-language level in the context of workflows. They pointed out that the design and modelling of workflows for business processes represented activities in the sphere of methods for doing business.

2.2 In appeal, claim 1 of the sole request corresponds to claim 1 on which the decision is based, replacing "control system" with "electronic control system", "object" with "data object", and deleting "at least partially" in combination with the feature of an encapsulated data structure. These changes might give the claim the appearance of a higher level of technicality, but apart from the fact that the Board has doubts that they are clearly defined in the description, in particular in a situation where there is no single technical embodiment explaining how the invention works, the Board judges that they do not add anything inventive and agrees with the examining division's decision on the refused claim.

2.3 The appellant disagrees with this non-technical interpretation. The application does not disclose the modelling of a business process as such. The invention rather relates to the automatic execution of a process with technical means and with an electronic control system. A claim shall be interpreted in good faith and objectively by the person skilled in the art. A missing contribution of particular features, such as "object", "assigner units" and "assignee units", to the technical character shall not be a criterion for an over-broad interpretation of claims. The description and Figures should always be interpreted as a "whole" and an application be regarded in its entirety.

The object of the invention is to provide a technical possibility that allows a workflow to be changed in a controlled manner, that is, the electronic control system enables dynamic reaction to and adjustment of the automated workflow. The solution of the invention involves "operating tags" which represent a protected, dynamically modifiable and tagged area which the electronic control system uses (if possible) to adapt the automation of the workflow.

The appellant further argued that following G 1/19 a method that changes something which is processed, has a technical effect and is therefore technical. The invention introduces the concept of "operating tags" which represent a technical possibility of interaction with a process task. In contrast to D1 the invention does not need a complex folder structure for realising a workflow system.

2.4 Regarding the **question of how to interpret the claimed subject-matter the Board notes** that according to the established jurisprudence of the Boards of Appeal on the general principles for claim interpretation, see Case Law of the Boards of Appeal, Edition 2022, II.A-6, **the wording of the claims should typically be given its broadest technically sensible meaning by a skilled reader.** When an invention is at the boundary between technical and non-technical matter, **also a non-technical interpretation of a claimed feature may have a sensible meaning**, in particular if the description and Figures disclose embodiments which permit such an interpretation. Claims must be read with a mind willing to understand and to make technical sense of them, thereby ruling out illogical or technically meaningless interpretations**. This also means that if a non-technical interpretation of a feature makes sense, then such an interpretation should not be excluded**.

This is exactly how the examining division approached the present case and the Board agrees. As a whole, the application is not limited to "data objects" and to "tasks" which are processed by an electronic system, as mentioned on page 4, lines 27 et seq. of the application. This passage seems to explain that a selected object is processed by executing one or more process tasks by means of a control system from one process state to the subsequent process state. However, at the same time, the application, page 6, lines 14 to 17, and page 11, lines 10 to 12, discloses that a selected object can comprise, e.g., at least one product and/or technical object and/or data and/or claim and/or account and/or job and/or contract and/or request and/or reporting object etc., and according to page 12, lines 20 to 21, tasks may be executed by a dedicated signaling to specific people to perform activities/tasks on the objects, as well as page 17, lines 3 to 9. **This contradicts page 17, second paragraph, where it is said that process tasks are executed by means of runtime execution modules**.

2.5 The appellant's argument that the present invention is not concerned with the modelling of workflows is not convincing, because the application, page 13, lines 7 to 9, **explicitly mentions that the process flow is modelled** and generated by means of the process management engine, including or based upon specific processing rules and technical instructions stored in the database. Moreover, page 13, second paragraph, explains that industrial, scientific, computational or business processes are "automatically operated" by means of the central control apparatus, but the work flow process is said to be composed of, among others, a sequence of process- or work tasks and interactions with human resources. **The term "automatically operated" therefore requires careful interpretation**.

2.6 Regarding the feature "operating tag", it is said on page 10, lines 16 to 19, and page 14, lines 17 to 23, that it is assigned to an object and/or process task and comprises dynamically alterable operating para meters which control the operation of the process task by means of the control system and/or by adding operational constraints to the processing of the pro cess task. An operating para meter may be a label, such as "pending", "cleared", "processing", "in operation" or "done", which describes a task state.

Besides representing a task state, operating tags may vary in their realisation; according to page 15, line 23, to page 16, line 22, an operating tag may be (i) simply a date with the states "due" and "overdue", (ii) reflect an aspect of work e.g. Pricing, Contractual, Reporting with the states "pending" and "done", (iii) refer to the supervision by a user or a group of users e.g. "Supervisor" with the states "watching" or "escalation-pending", a.s.o., (iv) a requirement to retain the audit trail of a task with states "none", "retained" or "expired", (v) a service level agreement with the states "included" or "excluded", (vi) the level of protection with the states "public" or "confidential".

At the same time, "operational tags" seem to serve another purpose. According to page 15, lines 18 et seq., they represent meta data, e.g. comprise non-hierarchical keywords or terms assigned to an object process task, with the purpose of describing the object or process task and allowing it to be found again. They aid in classification, marking ownership, noting boundaries, and indicating identity.

It is said that the purpose of an "operating tag" is to control the operation of an associated process task, which does not seem to change the underlying workflow process, but only the execution of tasks. Operating tags are said to be set by an authorised assigner or assignee of a process task which the application does not exclude to be a human user. In other words, it would seem that a user is given the ability to access and report his/her tasks along the applied dimensions "following a single consistent model", see page 16, first paragraph, in other words following a structured approach. The appellant's argument that the invention is able to dynamically change a workflow process is therefore not convincing.

2.7 The feature "encapsulated", is not technically clear. Page 14, line 30, to page 15, line 3, of the application mentions that operating tags can comprise an encapsulated data structure which is said to provide controlled access, but it does not further explain it. At the same time the encapsulated data structure is said to comprise only partly the dynamically alterable parameters, which seems to mean that other dynamically alterable parameters are stored outside of the encapsulated data structure. **The Board doubts that the technical effect of controlled access, as argued by the appellant, is achieved.** The meaning of "controlled access" in association with encapsulation in the technical field rather seems to be a way to limit direct access to components of an object and to require the use of object methods. This does not seem to imply that access is limited to authorised users. Further security measures would be necessary.

2.8 The **features "stochastic rating module" or "stochastic rating", are not further defined in the application.** The Board agrees with the examining division that **a non-technical, mathematical or business interpretation may be given to these features in terms of a rule for initiating the next process task.**

2.9 D1, see [0023] to [0026], discloses that workflow rules define actions to be taken on a data object at workflow stages based on data elements of an object. The actions are dynamically assigned, see [0055], based on object state, workflow state and user profile. Moreover, a data object in D1 can be used as a representation of a business object such as a contract, an offer, a tender, or an insurance claim, see [0055].

2.10 The examining division considered the subject-matter of claim 1 to be distinguished from D1 by the feature: "and wherein the state-structured process flow is a discrete time stochastic control process, wherein the control system comprises a stochastic rating module and the initiation of the next process task is also based on an additional rating by means of the stochastic rating module".

2.11 The examining division was of the opinion that the claimed rating is nothing more than a non-technical workflow rule according to which a next process task is initiated. Such workflow rules stem from business requirements and their implementation on the D1 system is a straight-forward modular programming based on common general knowledge.

2.12 The appellant in summary argues that the feature "discrete time stochastic control process" must be seen in combination with the system control structure and the operating tags. Such a control is not disclosed in D1 nor is it needed. D1 discloses that different workflow instances may be associated with objects, but these workflow instances are not altered.

2.13 The appellant points out that claim 1 is further distinguished from D1 by the operating tags which "... comprise an encapsulated data structure, wherein controlled access to the operational tag ... for authorized assigner units or assignee units by the encapsulated data structure ... dynamically alterable operating parameters and/or the operational constraint parameters and/or the expanding or the indicating parameters of task states ... dynamically operating the state-structured process flow ... " (listed are the high-lighted portions of claim 1, see page 14 of the grounds). The encapsulated data structure in the operating tag would permit a (dynamic) change of a workflow process during its execution in a way that another subsequent workflow state is reached from a same work flow state in a way which is not defined in advance. A user is able to interact with the process flow without the risk that the system looses control of it. This provides a certain flexibility while maintaining overall control. Furthermore, unauthorised third parties or users outside their granted access rights are prevented from influencing the process flow.

None of the prior art documents D1, D2 or D3 discloses such a data structure. D1 provides only a monitoring of work flow processes which are defined by workflow rules. An "application process" is controllable by the user, see [0020] of D1, but the workflow process must be defined in advance. Two parallel workflows may be generated, [0026] and [0027], but they follow the pre defined workflow rules. Also a dynamic assignment of actions in the workflow, [0055], does not change the workflow as such. Moreover, D1 requires a particular folder-based data structure to represent the workflow.

2.14 **The Board is not convinced by these arguments**. Claim 1 defines a state-structured process flow of how a selected object is processed from a process state to a subsequent process state, but not a (dynamic) change of the workflow process during its execution, in other words, new tasks may be generated and assigned, but the underlying work flow or process flow does not change. This is also reflected by the two-part form of claim 1.

2.15 As mentioned above, "operating tags" serve various purposes, among others, for documentation and for controlling the execution of process tasks, but, again, this does not change the underlying process flow. Moreover, oper**ating tags stem from the business side,**

as they define business conditions determining whether a certain task shall be executed or not, see page 15, line 28, to page 16, line 22. For instance, they denote the aspect of work, such as pricing, contractual, reporting, or that a task is included or excluded in a service level agreement (SLA). **They cannot confer technical character to the claimed subject-matter**.

2.16 The use of an "encapsulated data structure" is a commonly known programming concept, see page 11, second paragraph of the impugned decision, which is known to prevent direct access to internal components of an object. Table 2 of D1 illustrates the data structure of the object used in D1. Paragraphs [0024] and [0025] explain that the values and attributes of the object cannot be accessed directly, but only via actions, e.g. view, modify, data editing or data entry, processing and validation rules. This reflects the properties of an encapsulated data structure, as it is used in the application. It is rather the nature or type of a tag which seems to limit access to an object, see page 16, lines 16 and 18, where it is said that a tag may denote the level of protection with particular states "public" or "confidential".

2.17 As mentioned above, the application does not define how the "stochastic rating module" determines the "stochastic rating". The Board agrees with the examining division that **this merely amounts to a mathematical rule according to which a new process task gets initiated,** see second half on page 7 of the decision, which **cannot contribute to the technical character of claim 1, and thus not to the presence of inventive step** (Article 56 EPC).

2.18 The Board considers that the **appellant did not draw the correct conclusion from the statements in G 1/19.** The appellant considered that when this decision, e.g. at reasons, point 51, states that any technical effect going beyond the implementation of the process on a computer may be considered for inventive step, it means anything beyond a 1:1 mapping between the implementation and a step of the business method being implemented. In other words, any subject-matter that does not "map" to a step in the business method is technical. This was said to apply to all computer-implemented inventions, such as the present case, not just simulations. Accordingly, at least the "operating tags", which did not "map" to a step in the business method, were technical.

The Board agrees that the "implementation" of a business method implies some sort of mapping between non-technical steps of the business method and their technical realisation. **Decision G 1/19 has something to say about this mapping, at least in the forward direction, at point 51, when it rephrases the requirement for technical effect as "technical effect going beyond the simulation's straightforward or unspecified implementation on a standard computer system". Thus, even a 1:1 mapping might be inventive if it is not "straight-forward" (e.g. not standard programming or routine modification of the technical means used), or "unspecified" (e.g. not simply as "means for [carrying out the step]").**

But, looking for a mapping from implementation to a step of the business method in the reverse direction does not make sense as the steps of the non-technical activity do not have to be specified explicitly. They would include any steps that the business person would come up with in a non-technical workflow. The way this is handled is by considering the mapping of the implementation to the effect of the step and to examine whether the effect has any technical character, or whether it would be covered by what the business person would consider as part of the non-technical process. This is, in other words, the standard COMVIK approach where one looks at the effect of a feature in order to pose a technical problem, which might simply be the implementation of the feature, for which the above-mentioned mapping in the forward direction meant in G 1/19 applies.

Thus, **looking at the feature of the "operating tags" in the present case, the effect, as mentioned above at point 2.15, is to define business conditions determining whether a certain task shall be executed or not.** This, of course, **corresponds to a non-technical step of the workflow system**, namely keeping track of the state of a process. Going forward again with the mapping in order to judge inventive step, the implementation is seen to be the use of "operating tags", which even if escaping the "unspecified" classification must surely be "straight-forward".

Furthermore, the **Board cannot recognise that avoiding the folder data structure of D1, as argued by the appellant, represents a technical effect**.

2.19 The present case is rather **comparable to T 894/10**, reasons, points 7 and 8, in which the present Board, in a different composition, held that **all aspects of the idea of modelling and manipulating representations of a workflow are fundamentally non-technical, being essentially aspects of either a business method or an algorithm or both. [...] Technical considerations only come into play when implementing the representation and rules.**

2.20 The Board therefore concludes that the subject-matter of claim 1 lacks an inventive step over D1 within the meaning of Article 56 EPC, because the skilled person would adapt the modules of D1, see [0020][0021], with additional functions to implement new workflow rules or constraints based on common general knowledge.

T 1998/22 (Wide and deep machine learning models/GOOGLE) 20-12-2024

European Case Law Identifier ECLI:EP:BA:2024:T199822.20241220

WIDE AND DEEP MACHINE LEARNING MODELS

Clarity - main request (no)
Inventive step - all requests (no)

Application number	16826643.5
IPC class	G06N 3/04
Applicant name	Google LLC
Cited decisions	G 0001/19, T 0817/16, T 0697/17

Board 3.5.06

https://www.epo.org/boards-of-appeal/decisions/pdf/t221998eu1.pdf

Claim 1 of the main request reads as follows:

"A system comprising one or more computers and one or more storage devices storing instructions that when executed by the one or more computers cause the one or more computers to implement a combined machine learning model (102) for processing a machine learning input comprising a plurality of features (108-122) to generate a predicted output (136) for the machine learning input, the combined machine learning model comprising:

a deep machine learning model (104) configured to process the features to generate a deep model intermediate predicted output;

a wide machine learning model (106) configured to process the features to generate a wide model intermediate predicted output; and

a combining layer (134) configured to process the deep model intermediate predicted output generated by the deep machine learning model and the wide model intermediate predicted output generated by the wide machine learning model to generate the predicted output,

wherein the deep machine learning model and the wide machine learning model have been trained jointly on training data to generate the deep model intermediate predicted output and the wide model intermediate predicted output by backpropagating a gradient determined from an error between a predicted output for a training input and the known output for the training input through the combining layer to the wide machine learning model and the deep machine learning model to jointly adjust the current values of the parameters of the deep machine learning model and the wide machine learning model."

The application

1. The application relates to machine learning. It is proposed to combine a "deep machine learning model" and a "wide machine learning model" with a "combining layer", so as to obtain, for a set of "features" provided as input, a "predicted output" from the outputs of the two models.

2. The "deep machine learning model" may include "a deep neural network 130" and an "embedding layer 150" (paragraph [34]).

The "wide machine learning model" may include "a wide and shallow model, e.g. a generalized linear model 138" and a "cross-product feature transformation 132".

3. According to the description, "in general, a wide machine learning model can memorize feature inter actions through a wide set of cross-product feature transformations and a deep machine learning model can generalize unseen feature combinations by applying embedding functions to the input features", and "by including both deep machine learning model and

wide machine learning model, the wide and deep machine learning model can obtain both benefits of memorization and generalization and thus can perform better on predicting an output from a set of input features" (paragraph [13]).

4. In an embodiment, to which the independent claims of the present requests are limited, the deep and the wide models are (or have been) trained "jointly" based on training data using backpropagation (paragraphs [54]-[62]).

5. In a first application example, the input may be a sequence of words, the features may be tokens representing the words in the sequence and the predicted output may be "a likelihood that a particular word is the next word in the sequence or a prediction for a part of speech or a word sense for a particular word in the sequence" (paragraph [20]).

In a second application example, the input may be "features of a content presentation setting" and the output "a score that represents a likelihood that a particular objective will be satisfied if the content item is presented in the content presentation setting". This may be, for instance, presenting in a web site a product to a user that the user is likely to purchase based on user features, or recommending an app in an online app store (paragraphs [21]-[33]).

Main request - Article 84 EPC

7. "Wide machine learning model"

7.1 The examining division objected under Article 84 EPC that the expression "a wide machine learning model" - used e.g. in claim 1 - was "vague and unclear and [left] the skilled person in doubt as to which technical features it refer[red] to" (decision, point 12.1).

7.2 The appellant argued that the skilled person would be able to attribute a technical meaning to that expression at least in contrast to the expression "deep machine learning model" that is also used in claim 1. Support in the description for the expression "wide machine learning model" was to be found in paragraphs [13], [37], [38] and [52] (statement of grounds of appeal, page 3).

7.3 The board agrees with the examining division that claim 1 is unclear, Article 84 EPC, because of the expression "wide machine learning model", which **does not appear to have an established meaning in the art**.

The appellant's argument is not convincing as the term "wide" is not the opposite of the term "deep" and there is thus no reason to interpret "wide machine learning model" as any machine learning model that is not a "deep machine learning model" (as apparently suggested by the appellant).

Such an understanding of "wide machine learning model" would also not be consistent with the description, in particular the passages cited by the appellant itself. While no general definition is provided for that expression, it is said in paragraph [37] that a wide machine learning model is a "wide and shallow model". Hence, **being "not deep" ("shallow") is not sufficient to be a "wide machine learning model"**. This is also evident when looking at the advantages considered to be associated with the use of a "wide machine learning model" in paragraph [13]: it is not plausible that merely being not "deep" would be sufficient for any machine learning model to have them. The only concrete class of machine learning models disclosed is that of "generalized linear models" (paragraph [37]). It is not clear which other types of models are meant to be encompassed by the expression "wide machine learning model" in claim 1, and which ones are not encompassed by it (e.g. graphical models?). The scope of claim 1 is thus not clear, Article 84 EPC. The same objection applies to the other independent claims.

8. The objection of the examining division against claim 9 concerning the expression "for instance" (decision, point 12.2) has become moot due to the amendment made to that claim in all requests.

Main request - Inventive step

9. The examining division considered that all claims lacked an inventive step, Article 56 EPC, as they failed to solve a technical problem over the disclosure of D1, the distinguishing features of claim 1 over D1 only providing a solution to the non-technical problem of "how to improve/modify the mathematical model used in D1" (decision, point 13).

10. The appellant did not disagree with the differentiating features of claim 1 over D1 identified by the examining division but considered, based on paragraph [13] of the description, that they "allow a reduction of the amount of memory needed and produce an efficient method of memorization" and solve the objective technical problem of how to achieve this technical effect (statement of grounds of appeal, pages 4 and 5).

Following T 697/17, points 5.2.3 and 5.2.4, and T 817/16, point 3.12, determining technicality might be performed by enquiring whether the non-technical features would have been formulated by a technical expert rather than a non-technical expert. In the present case, "it must be a 'technical expert' since neither 'a programmer as such' who is able to implement a non-technical requirements specification on a computer nor a mathematician would have known how to deal with the 'generalization' aspect in the problem and hence none of them would have come up with the distinguishing features of claim 1". Hence, the distinguishing features would have been formulated by a technical expert and must be considered in the assessment of inventive step (statement of grounds of appeal, pages 6-8).

Furthermore, the present invention might be considered to be a "technical implementation" in the meaning of the EPO Guidelines G-II, 3.3, since it affected the input sample storage architecture of the overall architecture, as derivable from paragraph [13] (statement of grounds of appeal, pages 8 and 9).

Additional advantages of the invention were that it was able to deal with mixed kinds of data (paragraphs [35]-[38] and [46]-[48]) and that it could be implemented as a distributed system, e.g. in various cloud services (statement of grounds of appeal, page 9).

11. The **board considers that the system of claim 1 does not make any technical contribution over a conventional general-purpose computer and therefore lacks an inventive step**, Article 52(1) and 56 EPC, already for this reason.

The board notes that this line of argument was already developed in the WOISA (point 2.1) in parallel to the objection starting from D1, the key issue (technicality) being essentially the same.

11.1 The system of claim 1 differs from a conventional general-purpose computer only in the "instructions" stored on it, which amount to a computer program in the meaning of Article 52(2)(c) EPC. The question is whether this computer program **contributes to a technical effect**.

11.2 The **method realised by this computer program takes an abstract input** ("machine learning input comprising a plurality of features") and **produces an abstract output** ("a predicted output for the machine learning input"), **neither of which has an inherent technical character**.

The steps of that method involve the **processing of abstract data** by a "deep" and a "wide" machine learning model followed by a "combining layer". Like a deep neural network and a generalized linear model, **these are abstract computational models of a mathematical nature with no inherent technical character**. **That the two models have been jointly trained based on training data based on backpropagation does not impart any technical character on them**. It is in particular not derivable from the claim that any of these models - nor their combination - has been trained to perform a particular technical function.

It is noted that even the **two application examples** disclosed in the description and mentioned at point 8 above (word prediction and product/app recommendation) **do not appear to be technical applications.**

11.3 The board does **not follow the appellant's argument** which essentially amounts to **consider the invention as an efficient storage method.**

The combination of the deep and wide models applied to an input **does not result in any particular form of storage of that input, but in a prediction** (e.g. a product recommendation). Nor does any of the trained wide models represent a compressed encoding of the corresponding training data. It is **not envisaged** that the **training data can be in any way reconstructed from the trained models**. The trained models are only predictive models derived from the training data.

The board understands the notion of "memorization" used in paragraph [13] in relation to wide models (involving cross-product feature transformations) as an abstract one, relating to learning and to taking into account in the prediction co-occurrences in the historical data ("feature interactions").

11.4 As regards a technical implementation, claim 1 **does not go beyond an implementation of the method by means of corresponding "instructions", i.e. as a computer program**. It does not involve a distributed computing environment or cloud services, and so the corresponding argument of the appellant is not relevant.

11.5 The alleged fact that the invention would be able to deal with mixed kinds of data, e.g. numerical and categorical data, **does not concern a technical property and would thus not imply a technical contribution**, even if it were derivable from claim 1 (it is not).

11.6 As to the argument that only a "technical expert" could have devised the features of the invention, the board notes, as a general word of caution, that this kind of enquiry may be helpful in some cases to separate non-technical features from technical features - in particular to **identify business-related features** - but **does not constitute a definite test as it only concerns which kind of considerations underlie some features of the invention and not which kind of effects are achieved by it**. For instance, a claim to a computer-implemented simulation may involve features which are based on expertise in the technical field of the technical system that is being simulated. This alone would however not be sufficient to conclude that these features contribute to the technical character of the claim (G 1/19, reasons 122, 125, 141 and 142).

In any case, the **board tends to consider that claim 1 does not reflect any considerations beyond computer programming** (which encompasses the design of algorithms) and mathematics.

12. Additionally, the board notes that it does not find fault in the objection raised by the examining division starting from D1. The appellant has not objected to the differentiating features identified in point 13.1.1 of the decision, and they are not considered to solve any technical problem over D1 for the same reasons as those given above.

Auxiliary request 1

13. Claim 1 of auxiliary request 1 differs from claim 1 of the main request only in that it comprises the additional feature "wherein the wide machine learning model (106) is a generalized linear model (132)".

14. Even if the amendment might be accepted as overcoming the objection under Article 84 EPC raised at point 7.3 above, the objection under Article 56 EPC still applies, **a generalized linear model being an abstract model with no inherent technical character**.

Auxiliary request 2

15. Claim 1 of auxiliary request 2 differs from claim 1 of the main request in the following amended features:

- "a machine learning input comprising a plurality of input features";

- "a deep machine learning model (104) configured to process a first set of features included in the machine learning input [deleted: the features] to generate a deep model intermediate predicted output";

- "a wide machine learning model (106) configured to (i) apply a cross-product feature transformation to a subset of a second set of features included in the machine learning input to generate transformed features and (ii) to process, using a generalized linear model, the transformed features and the input features in the second set [deleted: the features] to generate a wide model intermediate predicted output".

16. Even if the amendments might be accepted as overcoming the objection under Article 84 EPC raised at point 7.3 above, the objection under Article 56 EPC still applies, the **amendments only providing further mathematical details of the abstract model with no inherent technical character**.

It is noted that the use of a "cross-product feature transformation" in the "wide machine learning model" may contribute to establishing that the advantages recited in paragraph [13] are actually achieved, as emphasised by the appellant (statement of grounds appeal, pages 10 and 11), but **these advantages remain of a non-technical nature**.

T 0080/23 (Model parameter combination/HUAWEI) 03-12-2024

European Case Law IdentifierECLI:EP:BA:2024:T008023.20241203

MODEL PARAMETER FUSION METHOD AND APPARATUS

Inventive step - (no)

Application number	15908513.3
IPC class	G06N 99/00, G06F 8/10, G06F 15/16, G06F 17/50
Applicant name	Huawei Technologies Co., Ltd.

Board 3.5.06

https://www.epo.org/boards-of-appeal/decisions/pdf/t230080eu1.pdf

Claim 1 reads as follows:

"A model parameter combination method, wherein the method includes: the model parameter training platform perform model parameter combination on data sets carrying data labels, to obtain a total model parameter, and the total model parameter is used to identify a data type of new data, wherein the model parameter training platform includes: the calculation server for iteration calculation; wherein the method is applied to a machine learning system, the machine learning system comprises at least one parameter collection group and at least one parameter delivery group, each parameter collection group is corresponding to at least one parameter delivery group, each parameter collection group comprises at least one node, each parameter delivery group comprises at least one node, a node comprised in the at least one parameter collection group is different from a node comprised in a corresponding parameter delivery group, and the method comprises:

when any parameter collection group meets an intra-group combination condition, combining model parameters of M nodes in the parameter collection group that meets the condition, to obtain a first model parameter of the parameter collection group that meets the condition, wherein a smallest quantity s of combination nodes in the parameter collection group that meets the condition <= M <= a total quantity of nodes comprised in the parameter collection group that meets the condition; and

sending, to N nodes in a parameter delivery group corresponding to the parameter collection group that meets the condition, the first model parameter of the parameter collection group that meets the condition, wherein 1 <= N <= a total quantity of nodes comprised in the parameter delivery group corresponding to the parameter collection group that meets the condition."

The application

1. The application relates to what is commonly referred to as distributed or federated machine learning.

A method is proposed for training a machine learning model (determining "model parameters") using a computer system ("model parameter training platform") comprising a plurality of "calculation servers" - also named "nodes" - which may be "common computers" (paragraphs [0114]-[0116] of the original description; fig. 1).

The training dataset ("original data for model parameter training") comprises labelled data (e.g. animal images carrying an animal label). Each node stores in local storage only a part of the whole training dataset and performs training based on that part (paragraphs [0115]-[0116]; fig. 2: steps 201 and 202).

The proposed method ("model parameter combination method") appears to specifically relate to how to "combine" the model parameters learned separately by each node so as to obtain - after a certain number of iterations of (local) training and combination - "final" or "total" model parameters. The final model parameters are provided for subsequent use of the model (paragraphs [0116]-[0118]).

2. Two prior art methods are acknowledged in the description: a "first method" involving a "parameter server" that collects and combines model parameters, and a "second method" involving a node sending its model parameters to another node where they are combined with the model parameters computed by the other node, this process being repeated from node to node (paragraph [0004]).

The description states that "the first method has a relatively high performance requirement for a parameter server configured to perform model parameter combination, and is prone to cause a shutdown, and the second method requires more data to be stored and a large data transmission volume" (paragraph [0004]).

3. Against this background, the proposed method is said to "resolve a problem that model parameter combination has a high performance requirement for a parameter server, and a large data transmission volume" (paragraph [0005]).

4. In the proposed method, the nodes are assigned to at least one "parameter collection group" (PCG) and at least one "parameter delivery group" (PDG). They are such that:

- each PCG and each PDG comprises at least one node,

- each PCG corresponds to at least one PDG,

- if a PCG corresponds to a PDG, there is at least one node in the former and one in the latter that are different from each other (paragraph [0118]).

An example is given in paragraphs [0202]-[0207] in which 6 nodes numbered 0 to 5 are assigned to 2 PCGs and 3 PDGs as follows:

- PCG 0 = {0, 2, 4},

- PCG 1 = {1, 3, 5},

- PDG 0 = {0, 3},

- PDG 1 = {1, 4},

- PDG 2 = {2, 5}.

5. First, each node receives a part of the training dataset (fig. 1, step 201; paragraph [0119] "data subset from a data set") and trains the model ("performs iterative calculation") based on the respective "data subset" using an "initial model parameter" (step 202; paragraphs [0124]-[0126]).

When any parameter collection group meets an "intra-group condition" - defined as M nodes of the group having completed the current training, where M>=s for a preset minimum value s - the model parameters obtained by these M nodes are "combined" to obtain a "first model parameter" of the parameter collection group (step 203; paragraphs [0127]-[0128]).

It seems that whenever a single "model parameter" ("initial model parameter", "first model parameters") is mentioned it is actually referring to a plurality of parameters or to a vector of parameters (assuming the model involves a plurality of scalar parameters).

6. The description does not explain how (mathematically) the parameters are to be "combined".

It does however indicate that the combination may either be carried out by a "device independent of the parameter collection group" (a "parameter server") or by one or more of nodes of the parameter collection group (a "control node" performing the whole combination or several nodes performing it iteratively, the last node being in that case referred to as a "combination node"; paragraphs [0132]-[0141]).

7. The "first model parameter" obtained by this combination operation is sent to N nodes in a parameter delivery group corresponding to the parameter collection group (step 204; paragraph [0149]-[0150]). The sending may be carried out either in a broadcast or iterative manner (paragraph [0151]).

8. When W parameter collection groups meet an "inter-group combination condition" - which may be defined as a preset number of intra-group combinations having been performed in each of these groups, all model parameters of nodes in each of the W parameter collection groups are "separately combined" so as to obtain a "second model parameter" of each of these W parameter collection groups (step 205; paragraphs [0156]-[0159].

Then, the second model parameters of all parameter collection groups in the W parameter collection groups are "combined" to obtain a "third model parameter (step 206; paragraph [0165]).

It seems that these two combination steps differ from each other in that, in the former, a "separate combination" and, in the latter, an "overall combination" is performed (see e.g. paragraphs [0156], [0165] and [0178]). How these two types of combination actually differ from each other is unclear to the board.

In any case, the combinations in both steps may be performed either by an independent device or by a node. The performing entity is referred to as an "executor" (paragraphs [0160]-[0163], [0167]-[0172]).

9. The third model parameter is then sent to the nodes in the W parameter collection groups and to corresponding parameter delivery groups, in both cases either by broadcast or iteratively (paragraphs [0179]-[0180]).

10. When a "preset condition" is met, which may be that a specific time has passed, "nodes included in a parameter collection group and a parameter delivery group" are "regrouped" (step 207; paragraphs [0183]-[0184]).

After the "regrouping", steps 202 to 207 are repeated "based on a data subset and a current model parameter" until a "final model parameter" is output.

It is not entirely clear to the board what the "regrouping" means. From paragraphs [0186]-[0187], it seems to be an assigning of nodes to PCG and PDGs and of PCGs to PDGs. Perhaps it is a re-assignment different from those underlying the previous executions of steps 202 to 207.

11. The description asserts that thereby "a problem that model parameter combination has a high performance requirement for a parameter server, a large data transmission volume, and a dynamic calculation resource adjustment is resolved" (paragraph [0191]).

Inventive step

12. The examining division considered that the only technical features of claim 1 were a calculation server and a machine learning system, known for instance from D1. The implemented method was by itself a non-technical, abstract mathematical method for "adapting parameters of an abstract model, using abstract parameter groups exchanged between abstract nodes of the model". According to T 49/99, information modelling was not a technical activity. It was not derivable from claim 1 that any technical effect is achieved by this method, in particular not "alleviating a requirement of high performance on a parameter server ... while reducing the amount of data transmission also", as had been argued by the appellant. This **method did thus not solve a technical problem**. The claimed implementation of this non-technical method on a system like that of D1 being obvious, claim 1 lacked an inventive step, Article 56 EPC (decision under appeal, points 1-3).

13. The appellant argued that the invention of claim 1 "does not relate to modelling per se, but rather [to] a method for structuring and operating a model parameter combination apparatus involving the conception and implementation of a complex system that unquestionably includes features of a technical character". The claimed method achieved the "advantageous technical effect of reducing the amount of processing and data transmission required with respect to the cited prior art". Such an effect was technical according to G 1/19, which had "superseded and replaced" T 49/99. Furthermore, the "intra-group [...] conditional combining of model parameters and the sending of the first model parameter to nodes in a parameter delivery group corresponding to the condition-meeting parameter collection group features" conferred technicality upon the claimed method. As D1 disclosed no teaching pointing towards the claimed solution, claim 1 was inventive (statement of grounds of appeal, pages 1-4).

As to the examining division's argument that this effect was not derivable from claim 1, the appellant noted that "the skilled person in the fields of machine learning and model parameter combination would [...] be capable of straightforwardly working the invention based on the claim wording alone so as to solve the [...] formulated objective technical problem of how to provide an improved machine learning system and method, and would experience no doubt as to credibility of [sic] substantially all embodiments encompassed by that claim wording exhibit these performance-improving effects upon which the invention is based" (statement of grounds of appeal, page 3, first paragraph).

At the oral proceedings, the appellant argued that the invention was not about mathematics but concerned a practical computational problem, namely how to improve transmission efficiency and avoid server overload. This was solved by a conditional local combination of model parameters, independently of whether the "nodes" are realised as separate computers or as software on a single computer. The dynamic selection of nodes for the combination of parameters enabled to take into account that some nodes may become faulty and enabled also scalability.

14. The board considers that the subject-matter of claim 1 lacks an inventive step, Articles 52(1) and 56 EPC.

14.1 **Claim 1 is far from reflecting the method proposed in the description**, which has been outlined in points 1 to 11 above.

14.2 Claim 1 is directed to a "model parameter combination method" and refers, on the one hand, to a "machine learning system" and to "nodes", and, on the other hand, to a "model parameter training platform" that includes a (single) "calculation server". **Claim 1 does not establish a clear relationship between these two pairs of entities.**

The board considers that the terms "machine learning system" and the "nodes" used in claim 1 may be interpreted - in the context of that claim - as referring to an abstract system with abstract nodes "of the model", like a neural network or a graphical model with nodes.

While this is not how these terms are used in the description, where "nodes" refer to different "calculation servers" (see point 1 above), the **board considers that claim 1 does not exclude such a broad interpretation**. This is in particular the case as claim 1 refers to a plurality of nodes but only to a single "calculation server". Furthermore, at the oral proceedings, the **appellant did not contest this broad interpretation** of claim 1 and agreed that claim 1 encompassed a realisation of the whole method by a single computer ("the calculation server"), the "nodes" being all realised in software.

Hence, **claim 1 encompasses embodiments in which a single technical entity** - "the calculation server", which may be a conventional computer system - is used to carry out all the claimed method steps.

14.3 The **remainder of claim 1 defines an abstract computation method**, involving abstract "nodes" having "model parameters" (claim 1 does not even specify that the nodes perform computations) and exchanging parameters between them under certain conditions (again similar to how the operation of an abstract neural network or graphical model may be described).

The only operations mentioned in claim 1 as being part of the claimed method are "combining model parameters of M nodes in the parameter collection group" to "obtain a first model parameter" if "an intra-group condition" is met by any parameter collection group (a collection of nodes), and the "sending" of the "first model parameter" to "N nodes of in a parameter delivery group corresponding to the parameter collection group". The "intra-group condition" is defined by reference to a "condition" that is to be met by at least s individual nodes of a "parameter collection group", but that "**condition" remains undefined in claim 1**.

14.4 It is **not apparent to the board that this abstract computation method contributes to solving a technical problem by producing a technical effect** within the context of claim 1. In particular, the appellant's arguments in that respect (see point 13 above) are not convincing.

14.4.1 The **computation of "model parameters" of an abstract "machine learning system", for instance by training based on an abstract data set, does not have any technical character.** The **complexity of the model is irrelevant in this regard, as is the question whether the resulting model is "improved".**

14.4.2 Effects related to "high performance requirements" of a parameter server or "large transmission volume" **cannot be derived from claim 1.**

Claim 1 does not specify how the nodes are technically implemented and so no conclusions can be drawn from claim 1 on how the claimed operations affect "performance requirements" or "transmission volume". This is also the case because claim 1 does not specify under which "condition" a "first model parameter" is obtained by combination of model parameters of certain nodes and is then sent to other nodes, nor what the purpose of the nodes in the "parameter delivery group" is.

As these alleged effects are not derivable from claim 1, it can be left open whether they would qualify as technical effects following G 1/19.

14.4.3 Similarly, effects relating to taking into account the possibility of "faulty" nodes and to the scalability of the method are not derivable from claim 1, for failing to specify a particular technical implementation and to define the "condition" for combination of model parameters and the purpose of the "parameter delivery group".

14.5 The board judges therefore that **claim 1 encompasses a straightforward technical implementation of an abstract computation method on a conventional computer system, where the abstract computation method makes no technical contribution in the context of claim 1 and thus cannot support the presence of an inventive step** within the meaning of Article 56 EPC. It follows that the subject-matter of claim 1 lacks an inventive step.

15. The board notes obiter that **even if the method were interpreted more narrowly in the light of the description, it would not appear to be derivable from it that the technical effects alleged by the appellant are achieved over a technical infrastructure for distributed machine learning like the one disclosed in D1**, for the reasons given in points 10.2 and 10.3 of the communication pursuant to Article 15(1) RPBA dated 14 November 2024.

T 2401/22 (Word salience estimation/SAP) 23-09-2024

European Case Law Identifier ECLI:EP:BA:2024:T240122.20240923

Position-dependent word salience estimation

Inventive step - main request (no)
Inventive step - auxiliary request (no)

Application number	18204942.9
IPC class	G06F 17/27, G06Q 50/00, G06N 3/02
Applicant name	SAP SE
Cited decisions	G 0001/19, T 1351/04, T 2330/13, T 0598/14, T 0755/18, T 0702/20, T 1903/20, T 1952/21

Board 3.5.07

https://www.epo.org/boards-of-appeal/decisions/pdf/t222401eu1.pdf

Claim 1 of the main request reads as follows:

"A computer-implemented method (300) for retrieval of electronic documents being indexed and associated with word salience determined for words in text data, the method being executed by one or more processors (410) and comprising:

receiving (302, 304), by the one or more processors, two or more electronic documents, each electronic document comprising text data, a second electronic document comprising a link to a first electronic document;

processing (306), by the one or more processors, word representations of words of the first electronic document using a first encoder (202) to provide first output and a context vector (210);

processing (308), by the one or more processors, text data of the second electronic document and the context vector using a first decoder (206) to provide second output;

determining (310), by an attention mechanism (204) executed by the one or more processors, a plurality of weights for each word in the text data of the first electronic document based on the first output, and the second output;

providing (312), by the one or more processors, a word salience value for each word, a word salience value comprising a sum of weights of a respective word; and

using the word salience values in a system (100; 400) to perform retrieval of electronic documents being indexed and associated with word salience values of words included in electronic documents."

Invention

1. The invention concerns the determination of the "word salience" of words in electronic documents for the retrieval or summarisation of electronic documents (see original description, page 1, lines 4 to 6; original claim 8).

1.1 **Word salience is "the relative importance of a word within an electronic document"** (page 2, lines 11 to 19). It can be determined based on term frequency-inverse document frequency (TFIDF), which increases proportionally to the number of times a word appears in the document, and is offset by the frequency of the word in the document. According to the description, the traditional TFIDF fails to account for the relative importance of words in different sentences of the document (page 2, lines 11 to 26).

1.2 The invention proposes position-dependent word "salience score" determination which takes into account "secondary data" having an association with the text data in electronic documents (e.g. articles). For example, the secondary data can be social media data or "linking tweets" with uniform resource locator (URL) links to articles. These are expected to reflect reader interest in the articles (page 6, line 31 to page 7, line 29).

1.3 A method according to the invention, as described on page 5, lines 19 to 28, receives electronic documents, each including text data. The electronic documents comprise a first electronic document (e.g. an article) and a second electronic document including a link to the first electronic document (e.g. a linking tweet with URLs to the article). The method uses a "first encoder" (e.g. a recurrent neural network, RNN), a "first decoder" (e.g. a bidirectional grand recurrent unit, GRU) and an attention mechanism. The first electronic document is processed using the "first encoder" to provide "first output" and a context vector. The second electronic document and the context vector are processed using the first decoder to provide "second output". The hidden representation at the last step of the encoder is considered as the context vector (page 9, lines 16 to 31). The attention mechanism determines a plurality of weights for each word in the text data of the first electronic document based on the first output and the second output. These weights are added to provide a word salience value for each word (page 5, lines 15 to 28).

1.4 The word salience values can be used in an information retrieval system in which electronic documents (e.g. articles) can be indexed and associated with word salience values included in the electronic documents. A word can be indexed to multiple electronic documents. In response to a query including a word, the information retrieval system can return ranked results representative of the electronic documents based on the word salience values of the word (page 14, lines 5 to 20).

…

Main request

3. Inventive step

3.1 Claim 1 specifies a computer-implemented method for "retrieval of electronic documents being indexed and associated with word salience determined for words in text data". The method of claim 1 uses an encoder-decoder architecture with an attention mechanism (known in the art, see e.g. document D4, abstract and page 4, third paragraph) to calculate word salience values of words of a first document as explained under point 1.3 above. As a last step of the claimed method, the word salience values are used for retrieving "electronic documents being indexed and associated with word salience values".

3.2 As explained in the background section of the description, on page 1, line 9 to page 2, line 26, at the date of priority of the present application, information retrieval systems were known which used one or more processes to calculate word salience of words and use them in order to identify, retrieve and/or summarise electronic documents based on the words present in the documents and their salience values.

3.3 Such prior-art systems were known to index the electronic documents based on word salience values in order to retrieve electronic documents. In particular, the board notes that **neither the description nor claim 1 describe how the word salience values are used in a system to perform retrieval, or how the documents are indexed in the system**. The application thus assumes that the retrieval and indexing are performed using standard techniques.

Such a prior-art method for retrieval of electronic documents being indexed and associated with word salience values determined for words in the text data (or simply "prior-art method for retrieval of indexed documents") is an appropriate starting point for assessing inventive step.

3.4 The subject-matter of claim 1 differs from the prior-art method for retrieval of indexed documents in that

(a) a second electronic document comprising a link to a first electronic document is also used in calculating the salience values;

(b) the word salience value for each word comprises a sum of weights of a respective word;

and in that the following steps are performed:

(c) processing word representations of words of the first electronic document using a first encoder to provide first output and a context vector;

(d) processing text data of the second electronic document and the context vector using a first decoder to provide second output;

(e) determining, by an attention mechanism, a plurality of weights for each word in the text data of the first electronic document based on the first output, and the second output.

3.5 The appellant argued that the invention of claim 1 achieved high quality word saliency, and thus led to better indexing than in the prior art. Word saliency had a direct impact on how the computer performed a search. It was used to steer the searching towards or away from certain documents. **Searching in large databases was an inherently technical task and data that determined the search outcome was technical data in view of decision T 1351/04.** It was inconsistent to consider that searching and its related infrastructure, such as the indexing which made the searching possible, was technical, yet which documents were produced by the search was irrelevant.

According to the appellant, this dichotomy - searching is technical, but search results are irrelevant - was untenable. Searching existed for a reason. Modern large scale storage systems stored hundreds of thousands of documents, if not millions. Such a storage was technically useless without search. A user who needed a document would have to manually go through those documents one by one until they had found the relevant one. Without an appropriate search function, a storage system beyond a couple of hundred documents would become dysfunctional. They ceased to provide the function that they were created for, namely providing the needed document.

Furthermore, the assumptions that the invention translated linguistic considerations into a mathematical model were incorrect. This did the invention no justice, as in fact, it was rather the opposite. The model did not learn which documents were responsive to which queries from a linguist. The neural models were fully general, and learned from the data itself. Nor did the model learn from a mathematician, as might be the case if a fixed formula were used. Instead, a novel machine learning architecture was defined that allowed the machine to learn the desired mapping itself. The machinery to learn this mapping was not known, or at least not disclosed in the cited documents.

The appellant further argued that the invention provided a technical mechanism to narrow down search results. The invention used word salience values, which were determined not by arbitrary user preferences but through the training of a neural network. This training process ensured that the word salience values reflected the relative importance of terms within the context of the documents themselves, rather than subjective or user-specific criteria. As such, the **word salience values served as a robust, technical solution to the problem of information overload in large-scale document retrieval system**s. Furthermore, the invention's method for calculating and applying these salience values was inherently technical as it involved data processing techniques, including the use of encoder decoder architectures and attention mechanisms, to derive salience values that enhanced the accuracy and relevance of search results.

At the oral proceedings the appellant also argued that the idea of using other documents that refer to the document through links, as in the well-known PageRank approach, was technical. The problem solved by the invention was: how to translate the interest expressed in secondary documents into numerical data that could be used in a normal information retrieval process. The solution used technical data derived from training which was different from linguistic concepts. It was not known how to take secondary documents into account and it was not obvious to devise a way to extract the relevant information from a large set of documents referencing another set of documents.

3.6 **The board does not find the appellant's arguments convincing**. The case law clearly establishes that even though information retrieval in a computer system may include technical tasks, not everything related to searching is technical. **Aspects related to user preferences, linguistics and semantics are in principle not technical. Obtaining search results which better meet the user's interests or more closely match the semantics of the search terms is not a technical effect** (T 598/14, Reasons 2.4). As explained in decision T 598/14, Reasons 2.5, while functional data, such as an index structure, which "controls the computer by directing it to a certain memory location" is considered technical in accordance with decision T 1351/04, not all data used in an information retrieval system is considered functional data contributing to a technical effect. **Index terms which merely correspond to keywords are as such not technical.**

3.7 In the present case, the indexing mechanism itself is not new. The way the electronic documents are indexed and associated with word salience values and the way the word salience values are used to retrieve electronic documents are known from the prior-art method for retrieval of indexed documents (see also points 3.2 to 3.4 above). The link from the second to the first document of feature (a) is well known in the art. In the context of the invention, it is not used for the purpose of accessing data in the computer, but to derive the user interest in a document. **The distinguishing features are thus not functional data, i.e. data making a technical contribution, within the meaning of decision T 1351/04.**

3.8 The result of the distinguishing features is that different salience values are obtained. Word salience values reflect both the relevant information in the electronic documents and the readers' interest in them (see e.g. page 2, lines 11 to 19; page 7, lines 21 to 29; page 10, lines 21 to 26). **A salience value is thus a non-technical mathematical parameter representing users' interests in the content of a document.**

In the claimed method, a second electronic document, e.g. a linking tweet, is used to obtain information about the user interest in the electronic documents (see e.g. page 7, lines 24 to 27). This is the result of a non-technical consideration. Furthermore, the use of e.g. web pages referencing a web page for evaluating the user interest in the web page is well known.

Taking the **above into account, feature (a) merely uses notoriously known technical means for a non-technical purpose and feature (b) does not make a technical contribution.**

3.9 Features (c) to (e) specify features of a computer program for calculating the salience values of words. Computer programs are as such not patentable (Article 52(2)(c) and (4) EPC). The **board does not recognise any technical considerations in the manner a first encoder, a first decoder and an attention mechanism are used to calculate the non-technical salience values of words** (see also T 702/20, Reasons 7 to 21; T 598/14, Reasons 2.3 and 2.4; T 1903/20, Reasons 3.3). Features (c) to (e) **merely produce non-technical salience values, no technical effect being apparent from the claim**.

3.10 The board is not persuaded by the appellant's argument that the claimed method was patentable because it was **based on a machine learning architecture that allowed the machine to learn the desired mapping itself**.

In the board's opinion, the **claim does not specify the features necessary to achieve such a learning effect**. In any case, even if the board recognised that the claimed method caused "the machine to learn the desired mapping itself", **the board would not acknowledge the learning effect to cause a technical effect**. In **decision T 755/18 the board concluded that if neither the output of a machine nor the output's accuracy are technical, an improvement of the machine achieved automatically through supervised learning to generate a more accurate output is not in itself a technical effect** (Reasons 3.2). In the board's opinion, the same holds true for unsupervised machine learning of a machine that produces a non-technical output. In decision T 1952/21, the examined claim specified a machine learning system based on neural networks. The board considered that the claim did not restrict the invention to a technical context. The **board did not recognise reinforcement learning as a further technical use**, even if the advantages in reinforcement learning brought forward by the appellant were to be acknowledged (Reasons 18 to 27.4 and 33.2 to 34).

In line with the cited case law, in the present case, **the use of the results of the method in the machine itself in order to learn how to achieve "enhanced ... accuracy and relevance of search results", as argued by the appellant, is not a further technical use within the meaning of G 1/19.** Even if features (a) to (e) of the claimed method were considered to contribute to a machine capable of learning itself how to produce better salience values, this would **constitute a non-technical algorithmic change of a computer program to produce salience values which better represented the user interests**, which is not a technical effect.

3.11 Therefore, the subject-matter of claim 1 lacks inventive step (Article 56 EPC).
