UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/319,040	01/18/2019	Guillaume Desjardins	45288-8015US1	2007

151793 7590 03/04/2025
FISH & RICHARDSON P.C.
PO BOX 1022
MINNEAPOLIS, MN 55440-1022

EXAMINER	
TRIEU, EM N	

ART UNIT	PAPER NUMBER
2128	

NOTIFICATION DATE	DELIVERY MODE
03/04/2025	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

PATDOCTC@fr.com

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

Ex parte GUILLAUME DESJARDINS, RAZVAN PASCANU,
RAIA THAIS HADSELL, JAMES KIRKPATRICK,
JOEL WILLIAM VENESS, and NEIL CHARLES RABINOWITZ

Appeal 2024-000567
Application 16/319,040
Technology Center 2100

Before HUNG H. BUI, NABEEL U. KHAN, and JOHN F. HORVATH, *Administrative Patent Judges*.

Opinion for the Board filed by HUNG H. BUI.

Opinion Concurring-in-part filed by JOHN F. HORVATH.

DECISION ON APPEAL

Appellant[1] appeals under 35 U.S.C. § 134(a) from the Examiner's final rejection of claims 1–6, and 8–20. Claim 7 is cancelled. Appeal Br. 16–22 (Claims App.). We have jurisdiction under 35 U.S.C. § 6(b).

We affirm and enter a New Ground of Rejection ("NGR") in accordance with 37 C.F.R. § 41.50(b).[2]

---

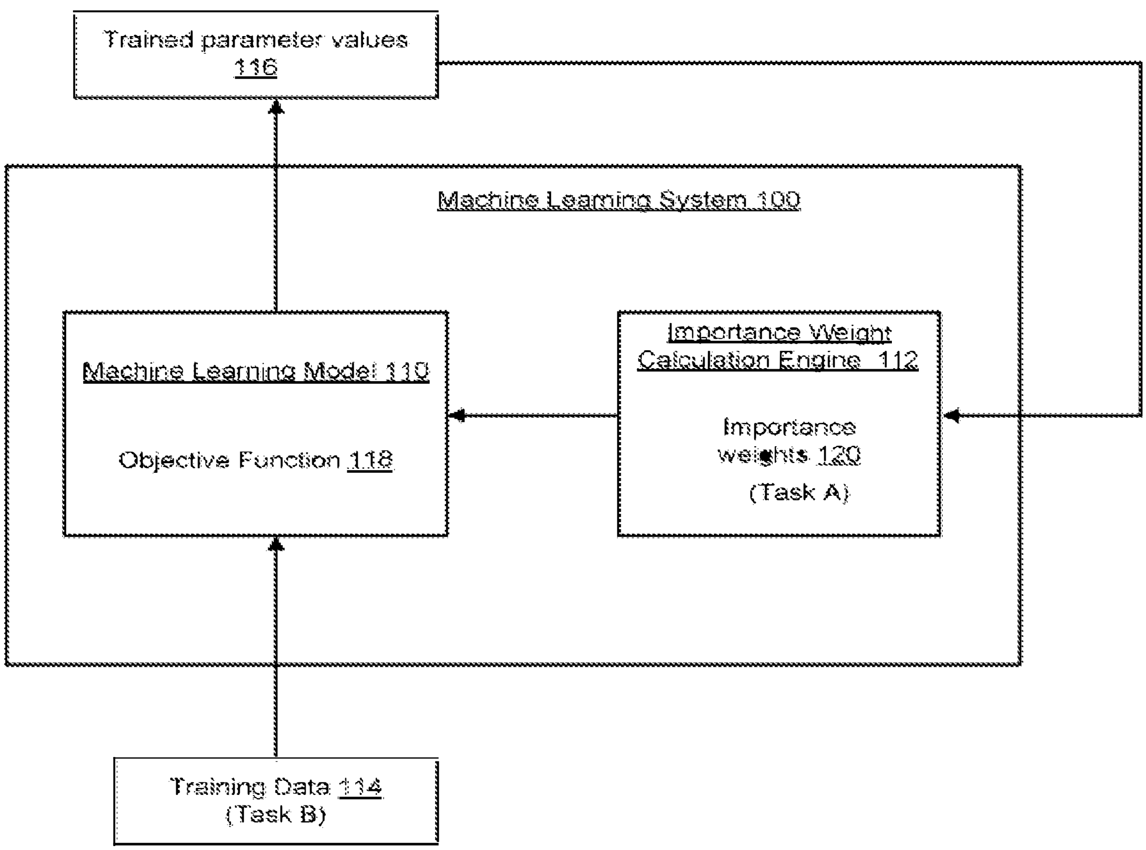
[1] "Appellant" refers to "applicant" as defined in 37 C.F.R. § 1.42(a) (2022). Appellant identifies DeepMind Technologies Limited as the real party in interest. Appeal Br. 1.

[2] We refer to Appellant's Appeal Brief filed May 24, 2023 ("Appeal Br."); Reply Brief filed November 13, 2023 ("Reply Br."); the Examiner's Answer

STATEMENT OF THE CASE

Technology

Appellant's disclosed invention relates to "a system [shown in Figure 1] implemented as computer programs on one or more computers" that can train "a [single] machine learning model on multiple machine learning tasks" such that "once the model has been trained, the model can be used for each of the multiple tasks with an acceptable level of performance" for each task. Spec. ¶¶ 6, 21. Appellant's Figure 1 is reproduced below:



mailed September 19, 2023 ("Ans.") and Final Office Action mailed December 27, 2022 ("Final Act."); and the Specification filed January 18, 2019 ("Spec.").

Appellant's Figure 1, reproduced above, depicts machine learning system 100 configured to (1) train machine learning model 110 on a first machine learning task (Task A) using first training data to determine first values of a plurality of parameters of the machine learning model 110; (2) determine, for each of the plurality of parameters, a respective importance weight 120 that represents a measure of an importance of the parameter to the machine learning model 110 achieving acceptable performance on Task A; (3) train machine learning model 110 on a second machine learning task (Task B) using new training data 114 to adjust the first values of the plurality of parameters to optimize performance of machine learning model 110 on the second machine learning task (Task B) while protecting performance of machine learning model 110 on the first machine learning task (Task A). Spec. ¶¶ 35, 37, 42. According to Appellant, "[t]he system 100 can train the model 110 to learn a sequence of multiple machine learning tasks," i.e., "learn new tasks without forgetting previous tasks . . . to optimize the performance of the model 110 on a new task while protecting the performance in previous tasks by constraining the parameters to stay in a region of acceptable performance (e.g., a region of low error) for previous tasks based on information about the previous tasks. Spec. ¶ 35.

Illustrative Claim

Claims 1–6 and 8–20 are pending. Claims 1, 18, and 19 are independent. Claim 1, reproduced below with disputed limitations emphasized and bracketed letters added for clarity, is illustrative:

1. A computing-implemented method of training a machine learning model, wherein the machine learning model has at least a plurality of parameters and has been trained on a first machine learning task using first training data to determine first values of the plurality of parameters of the

machine learning model, and wherein the method comprises:

> [A] determining, for each of the plurality of parameters, a respective measure of an importance of the parameter to the first machine learning task, comprising:
>
> [A1] computing, based on the first values of the plurality of parameters determined by training the machine learning model on the first machine learning task, an approximation of a posterior distribution over possible values of the plurality of parameters,
>
> [A2] *assigning, using the approximation, a value to each of the plurality of parameters, the value being the respective measure of the importance of the parameter to the first machine learning task and approximating a probability that the first value of the parameter after the training on the first machine learning task is a correct value of the parameter given the first training data used to train the machine learning model on the first machine learning task*;
>
> [B] obtaining second training data for training the machine learning model on a second, different machine learning task; and
>
> [C] training the machine learning model on the second machine learning task by training the machine learning model on the second training data to adjust the first values of the plurality of parameters to optimize performance of the machine learning model on the second machine learning task while protecting performance of the machine learning model on the first machine learning task,
>
> [D] wherein adjusting the first values of the plurality of parameters comprises adjusting the first values of the plurality of parameters [D1] *to optimize an objective function that depends in part on a penalty term that is based on the determined measures of importance of the plurality of parameters to the first machine learning task.*

Appeal Br. 16 (Claims App.). Independent claims 18 and 19 recite a corresponding system and computer-readable medium provided with instructions to perform substantially the same limitations of claim 1. *Id.* at 20–22.

REJECTIONS AND REFERENCES

(1) Claims 1, 2, 4, 9, 15, and 17–20 stand rejected under 35 U.S.C. § 103 as obvious over the combined teachings of Marcheret (US 2013/0254153 A1; published Sep. 26, 2013), Jamaluddin et al. ("Effect of Penalty Function Parameter in Objective Function of System Identification," International Journal of Automotive and Mechanical Engineering (UAME), Vol. 7, pp. 940–954, January 2013; "Jamaluddin"), Rousset et al. ("Neural Networks with a Self-Refreshing Memory: Knowledge Transfer in Sequential Learning Tasks without Catastropic Forgetting," Connection Science, Vol. 12, pp. 1–19, 2000; "Rousset"), Gordon et al. (US 2014/0101090 A1; published Apr. 10, 2014; "Gordon"), and Mehanna et al. (US 2016/0092786 A1; published Mar. 31, 2016; "Mehanna"). Final Act. 4–27.

(2) Claims 3 and 14 stand rejected under 35 U.S.C. § 103 as obvious over the combined teachings of Marcheret, Jamaluddin, Rousset, Gordon, Mehanna, and Aslan et al. (US 2017-0132528 A1; "Aslan"). Final Act. 39–42.

(3) Claims 6 and 10–12 stand rejected under 35 U.S.C. § 103 as obvious over the combined teachings of Marcheret, Jamaluddin, Rousset, Gordon, Mehanna, and Cao et al. ("A Practical Transfer Learning Algorithm for Face Verification," "Cao"). Final Act. 27–39.

(3) Claims 5, 8, 13, and 16 stand rejected under 35 U.S.C. § 103 as obvious over the combined teachings of Marcheret, Jamaluddin, Rousset,

Gordon, Mehanna, and Sinyavskiy et al. (US 9,146,546 B2; issued Sep. 29, 2015; "Sinyavskiy"). Final Act. 42–53.

ANALYSIS

In support of the obviousness rejection, the Examiner finds (1) the combination of Marcheret, Jamaluddin, Rousset, Gordon, and Mehanna teaches or suggests all of the limitations [A]–[D] of Appellant's claim 1, and similarly, independent claims 18 and 19, and articulates (2) reasoning with a rational underpinning to support the combination. Final Act. 4–27. Of particular relevance, the Examiner finds Mehanna teaches or suggests limitation [A2] of claim 1, as reproduced below:

> [A2] ***assigning, using the approximation, a value to each of the plurality of parameters, the value being the respective measure of the importance of the parameter to the first machine learning task and approximating determining an approximation of a probability that the first value of the parameter after the training on the first machine learning task is a correct value of the parameter given the first training data used to train the machine learning model on the first machine learning task.***

Final Act. 15 (citing Mehanna ¶¶ 38, 44).

The Examiner also finds the combination of Marcheret and Jamaluddin teaches or suggests limitation [D]–[D1] of claim 1, as reproduced below:

> [D] wherein adjusting the first values of the plurality of parameters comprises adjusting the first values of the plurality of parameters to optimize an objective function that depends in part on a penalty term that is based on the determined measures of importance of the plurality of parameters to the first machine learning task.

Id. at 7–8 (citing Marcheret ¶ 29 for teaching limitation [D] of claim 1; Jamaluddin pp. 941–942 for teaching limitation [D1] of claim 1).

Appellant disputes (1) the Examiner's findings regarding Mehanna's and Jamaluddin's teachings and (2) the Examiner's proffered reason to incorporate Mehanna's teachings into the proposed combination of Marcheret, Jamaluddin, Rousset, and Gordon. In particular, Appellant presents three principal arguments against the application of Jamaluddin and Mehanna.

First, Appellant contends Mehanna does not teach or suggest limitation [A2] of claim 1, as reproduced below:

> [A2] <u>assigning, using the approximation, a value to each of the plurality of parameters, the value being the respective measure of the importance of the parameter to the first machine learning task and approximating a probability that the first value of the parameter after the training on the first machine learning task is a correct value of the parameter given the first training data used to train the machine learning model on the first machine learning task.</u>

Appeal Br. 8–10 (citing Mehanna ¶¶ 5–6, 38, 44); Reply Br. 2–4. In particular, Appellant acknowledges Mehanna teaches (1) "features [as inputs to a machine learned model] . . . may include <u>***attributes of a user retrieved from a user profile***</u> (e.g., demographic information), <u>***actions associated with the user, and connections between the user and other users***</u>" as well as "<u>***characteristics of content items***</u>"; and (2) "a measure of features impact associated with a feature" as "<u>***a measure of the feature's importance to the machine learned model.***</u>" Appeal Br. 9 (citing Mehanna ¶¶ 5–6, 38). However, Appellant argues (1) Mehanna's "measure of feature impact" cannot be equated to "a value [assigned] to each of the plurality of parameters" recited in claim 1, and (2) even if one of Mehanna's features can be equated to a claimed "parameter,"

> "nowhere does Mehanna teach or suggest that the measure of the feature's importance **"approximat[es] a probability that the first value of the parameter after the training on the first machine learning task is a correct value of the parameter given the first training data used to train the machine learning model on the first machine learning task"**

as recited in Appellant's claim 1. *Id.* at 9–10.

Second, Appellant contends Jamaluddin does not teach or suggest limitations [D]–[D1] of Appellant's claim 1, as reproduced below:

> [D] wherein adjusting the first values of the plurality of parameters comprises adjusting the first values of the plurality of parameters to optimize an objective function that depends in part on a penalty term that is based on the determined measures of importance of the plurality of parameters to the first machine learning task.

Id. at 10–12. Specifically, Appellant acknowledges Jamaluddin teaches the use of an objective function in the context of machine learning to optimize model structure selection. *Id.* at 10 (citing Jamaluddin pp. 941–942). According to Jamaluddin, the objective function depends in part on a penalty—a fixed value termed penalty function parameter in which the penalty function is used to "penalize[s] terms with the absolute values of the estimated parameter less than the *penalty*." Jamaluddin 942. However, Appellant argues (1) Jamaluddin's penalty is different from the claimed "**<u>penalty term that is based on the determined measures of importance of the plurality of parameters to the first machine learning task</u>**" and, as such, (2) Jamaluddin does not teach limitations [D]–[D1] of Appellant's claim 1. Appeal Br. 11–12.

Third, Appellant contends the Examiner's proffered reason to incorporate Mehanna's teachings into the proposed combination of Marcheret, Jamaluddin, Rousset, and Gordon is "insufficient to support a *prima facie* obviousness rejection because (i) the references are not analogous in arts and (ii) the proposed rationale does not even relate to the features being claimed." *Id.* at 12–14. According to Appellant, neither Jamaluddin nor Mehanna teaches or suggests "machine learning models" or "multiple machine learning tasks, let alone training a single machine learning model on multiple machine learning tasks" and, as such, "are not analogous in arts and cannot be combined." *Id.* at 13–14. Similarly, Appellant acknowledges the Examiner's proffered reason to incorporate Mehanna's teachings into the combination is "to minimize an error of a machine learning model" (Final Act. 17 (citing Mehanna ¶ 44)), but argues the Examiner's proffered reason

> does not even relate to the claimed features of 'assigning, using the approximation, a value to each of the plurality of parameters, the value being the respective measure of the importance of the parameter to the first machine learning task and approximating a probability that the first value of the parameter after the training on the first machine learning task is a correct value of the parameter given the first training data used to train the machine learning model on the first machine learning task'

as recited in Appellant's claim 1. *Id.* at 14.

Appellant's arguments are not persuasive of reversible Examiner error. *See In re Jung*, 637 F.3d 1356, 1365 (Fed. Cir. 2011) ("[I]t has long been the Board's practice to require an applicant to identify the alleged error in the Examiner's rejections."). Instead, we find the Examiner's findings regarding Marcheret, Jamaluddin and Mehanna, including the Examiner's responses and explanations to Appellant's arguments, are supported by a preponderance of the evidence on this record. Ans. 54–64. As such, we adopt the Examiner's findings provided therein.[3] *Id.*

At the outset, we note that claim terms (not prior art terms) are given their broadest reasonable interpretation consistent with the specification during examination. *In re Am. Acad. of Sci. Tech Ctr.*, 367 F.3d 1359, 1364 (Fed. Cir. 2004). Neither the term "parameters" nor the term "values of the [] parameters" is expressly defined by Appellant's Specification. Instead, Appellant's Specification describes (1) a machine learning model 110, shown in Figures 1–2, that receives "an input and generate[s] an output, e.g., a predicted output, based on the received input"; (2) "the machine learning model 110 [may be] a parametric model having multiple parameters" and such "machine learning model 110 generates the output based on the received input and on values of the parameters of the model 110"; and (3) such "machine learning system 100 trains the machine learning model 110 on a particular task, i.e., to learn the particular task, by adjusting the values of the parameters of the machine learning model 110 to optimize

[3] We review the appealed rejection for error based upon the issues identified by Appellant, and in light of the contentions and evidence produced thereon. *Ex parte Frye*, 94 USPQ2d 1072, 1075 (BPAI 2010) (precedential). As such, arguments not made are forfeited. *See In re Google Tech. Holdings LLC*, 980 F.3d 858, 863 (Fed. Cir. 2020) ("Because Google failed to present these claim construction arguments to the Board, Google forfeited both arguments."); 37 C.F.R. § 41.37(c)(1)(iv) (2023) ("Except as provided for in §§ 41.41, 41.47 and 41.52, any arguments or authorities not included in the appeal brief will be refused consideration by the Board for purposes of the present appeal.").

performance of the model 110 on the particular task, e.g., by optimizing an objective function 118 of the model 110." Spec. ¶¶ 31–34, Figures 1–2.

Based on Appellant's Specification, the term "parameters" can be broadly, but reasonably, interpreted to encompass Mehanna's features as inputs to the machine learning models, or alternatively, Gordon's posterior distribution. Gordon ¶ 13; Spec. ¶ 50.

Second, we note Appellant cannot show non-obviousness by attacking references individually where the rejection is based on a combination of references. *See In re Keller*, 642 F.2d 413 (CCPA 1981). Here, the combination of Marcheret and Jamaluddin (rather than solely Jamaluddin) is cited for teaching or suggesting limitation [D]–[D1] of claim 1. Final Act. at 10–12. For example, Marcheret (not Jamaluddin) is cited for teaching limitation [D] of Appellant's claim 1: "***wherein adjusting the first values of the plurality of parameters comprises adjusting the first values of the plurality of parameters to optimize an objective function . . . to the first machine learning task***." Final Act. 7 (citing Marcheret ¶ 29). Likewise, Jamaluddin (not Marcheret) is cited for teaching limitation [D1] of Appellant's claim 1: "***to optimize an objective function that depends in part on a penalty term that is based on the determined measures of importance of the plurality of parameters***." *Id.* at 8 (citing Jamaluddin pp. 941–942).

Third, we are not persuaded by Appellant's argument that the cited prior art references, including Marcheret, Jamaluddin, Roussett, Gordon, and Mehanna are not analogous to the claimed invention and, therefore, cannot be properly used to support the proffered combination. Appeal Br. 13–14. "Whether a reference in the prior art is 'analogous' is a fact question." *In re Clay*, 966 F.2d 656, 658 (Fed. Cir. 1992) (citing *Panduit Corp. v. Dennison*

Mfg. Co., 810 F.2d 1561, 1568 n.9 (Fed. Cir. 1987)). Two criteria have evolved for determining whether prior art is analogous:

> (1) whether the art is from the same field of endeavor, regardless of the problem addressed, and
> (2) if the reference is not within the field of the inventor's endeavor, whether the reference still is reasonably pertinent to the particular problem with which the inventor is involved.

Id. at 658-59 (citing *In re Deminski*, 796 F.2d 436, 442 (Fed. Cir. 1986); *In re Wood*, 599 F.2d 1032, 1036 (CCPA 1979)). "A reference is reasonably pertinent if … it is one which, because of the matter with which it deals, logically would have commended itself to an inventor's attention in considering his problem." *Id.* at 659. "If a reference disclosure has the same purpose as the claimed invention, the reference relates to the same problem, and that fact supports use of that reference in an obviousness rejection." *Id.* As correctly recognized by the Examiner, all the cited prior art references, including "Marcheret, Jamaluddin, Roussett, Gordon, and Mehanna are analogous in arts because they have the same [field] of endeavor of training the plurality of machine learning models, wherein, each of the training of the machine learning model is considered as the machine learning task." Ans. 61–62. For example, Jamaluddin's model is described in the context of a machine learning model. Jamaluddin p. 953 (citing Golberg, D.E. 1989 "Genetic algorithms in search, optimization and machine learning." Massachusetts: Addison-Wesley). Likewise, Mehanna's machine learning models (shown in Figure 4) are described as being trained based on input features with additional modifications to output results. Mehanna Abs.

Because these references are in the same field of endeavor, we need not address whether each reference is reasonably pertinent to the particular problem with which the inventor is involved. As such, we agree with the Examiner that Marcheret, Jamaluddin, Rousett, Gordon, and Mehanna are analogous because they are within the same field of endeavor as the claimed invention—namely machine learning models.

We recognize that the Examiner must articulate "reasoning with some rational underpinning to support the legal conclusion of obviousness." *In re Kahn*, 441 F.3d 977, 988 (Fed. Cir. 2006). However, "[u]nder the correct [obviousness] analysis, any need or problem known in the field of endeavor at the time of invention and addressed by the patent can provide a reason for combining the elements in the manner claimed." *KSR Int'l Co. v. Teleflex Corp.*, 550 U.S. 398, 420 (2007).

The Examiner has provided reasoning with sufficient rational underpinning to incorporate Mehanna's teachings into the proposed combination of Marcheret, Jamaluddin, Rousset, and Gordon, i.e., "to minimize an error of a machine learning model." Final Act. 17 (citing Mehanna ¶ 44). We agree with the Examiner's reasoning. In contrast, Appellant has not sufficiently demonstrated the Examiner's proffered reason to combine is reached in error or why a person of ordinary skill in the art *would not* have reached the conclusions reached by the Examiner. Consequently, we are not persuaded that the Examiner's proffered reason to combine these references is incorrect.

In the Reply Brief, Appellant argues that "the parameters recited in claim 1 cannot be equated to the features in Mehanna and a measure of feature impact in Mehanna cannot be equated to a 'value' assigned to each parameter in claim 1." Reply Br. 1–4. According to Appellant, "Mehanna's features are retrieved from external sources such as 'a user profile' or 'content items,'" whereas the claimed "parameters" recited in claim 1 "are **internal** variables of the machine learning model" and, as such, "are not obtained from external sources like the features of Mehanna." Reply Br. 2. Likewise, Mehanna's "measure of feature impact" is "a measure of the feature's importance **to the machine learned model**" and "is not task-specific," whereas "the value assigned to each parameter is a measure of the importance of the parameter **to a specific task** (i.e., the first machine learning task) among a plurality of machine learning tasks." *Id.* at 3.

We disagree. The claimed "parameters" recited in claim 1 are not internal variables of the machine learning model, as Appellant argues. Instead, the claimed "parameters" and the "values of the parameters" are input to the machine learning model (shown in Figure 1) similar to the features shown in Mehanna's Figure 4.

Lastly, Appellant raises a new argument, in the Reply Brief, against the application of Mehanna. In particular, Appellant argues, even if Mehanna's feature can be equated with the claimed "parameters" in claim 1,

> nowhere does Mehanna teach or suggest that the measure of the feature's importance '**approximat[es] a probability that the first value of the parameter after the training on the first machine learning task is a correct value of the parameter given the first training data used to train the machine learning model on the first machine learning task**

as recited in claim 1. Ans. 3–4 (citing Mehanna ¶ 3).

However, this argument is not timely. In the absence of a showing of good cause by Appellant, we decline to consider this new argument raised for the first time in the Reply Brief. *See* 37 C.F.R. § 41.41(b)(2) (2018); *In re Hyatt*, 211 F.3d 1367, 1373 (Fed. Cir. 2000) (noting that an argument not first raised in the brief to the Board is waived on appeal); *Ex parte Nakashima*, 93 USPQ2d 1834, 1837 (BPAI 2010) (informative) (explaining that arguments and evidence not timely presented in the principal Brief, will not be considered when filed in a Reply Brief, absent a showing of good cause explaining why the argument could not have been presented in the Principal Brief); *Ex parte Borden*, 93 USPQ2d 1473, 1477 (BPAI 2010) (informative) ("Properly interpreted, the Rules do not require the Board to take up a belated argument that has not been addressed by the Examiner, absent a showing of good cause.").

Nevertheless, for purposes of completeness, we are persuaded that Mehanna does not teach or suggest relevant part of limitation [A2] of claim 1: "approximating a probability that the first value of the parameter after the training on the first machine learning task is a correct value of the parameter given the first training data used to train the machine learning model on the first machine learning task." However, Gordon teaches a machine learning algorithm configured to compute an approximation of a posterior distribution over possible values of the plurality of parameters. Final Act. 13 (citing Gordon ¶¶ 14, 46). An approximation of a posterior distribution is commonly known by those skilled in the art as representing "a probability that the current value is a current value of the parameter." Spec. ¶¶ 49–50.

Based on the teachings of Gordon and the commonly known "approximation of a posterior distribution" (Spec. ¶¶ 49–50), a skilled artisan would understand that Gordon (rather than Mehanna) teaches or suggests the relevant part of limitation [A2] of claim 1.

For these reasons, we are not persuaded of Examiner error. *See supra* note 3. Accordingly, we sustain the Examiner's obviousness rejection of independent claims 1, 18, and 19, and their respective dependent claims 2, 4, 9, 15, 17, and 20, which Appellant does not argue separately. Appeal Br. 15. However, because we have relied on facts and reasoning not raised by the Examiner, we designate our affirmance as a new ground of rejection pursuant to 37 C.F.R. § 41.50(b) to preserve Appellant's procedural safeguards. *In re Stepan Co.*, 660 F.3d 1341, 1346 (Fed. Cir. 2011) ("Had the Board labeled its rejection as a new ground of rejection, Stepan could have reopened prosecution to address the newly-alleged deficiencies in its Declaration with the examiner."); *In re Leithem*, 661 F.3d 1316, 1319 (Fed. Cir. 2011) ("Mere reliance on the same statutory basis and the same prior art references, alone, is insufficient to avoid making a new ground of rejection when the Board relies on new facts and rationales not previously raised to the applicant by the examiner.").

Because Appellant does not dispute the obviousness rejections of (1) claims 3 and 14 as obvious over the combined teachings of Marcheret, Jamaluddin, Rousset, Gordon, Mehanna, and Aslan; (2) claims 6 and 10–12 as obvious over the combined teachings of Marcheret, Jamaluddin, Rousset, Gordon, Mehanna, and Cao; and (3) claims 5, 8, 13, and 16 as obvious over the combined teachings of Marcheret, Jamaluddin, Rousset, Gordon, Mehanna, and Sinyavskiy, we also sustain these obviousness rejections for the same reasons provided in the record. Final Act. 27–53.

I. NEW GROUNDS OF REJECTION

New 35 U.S.C. § 101 Rejection of Claims 1–6 and 17–20 under 37 C.F.R. § 41.50(b)

Patent eligibility is a question of law that is reviewable *de novo*. *Dealertrack, Inc. v. Huber*, 674 F.3d 1315, 1333 (Fed. Cir. 2012). Patentable subject matter is defined by 35 U.S.C. § 101, as follows:

> [w]hoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

In interpreting this statute, the U.S. Supreme Court has emphasized that patent protection should not preempt "the basic tools of scientific and technological work." *Gottschalk v. Benson*, 409 U.S. 63, 67 (1972) ("*Benson*"); *Mayo Collaborative Servs. v. Prometheus Labs., Inc.*, 566 U.S. 66, 71 (2012) ("*Mayo*"); *Alice Corp. Pty. Ltd. v. CLS Bank Int'l*, 573 U.S. 208, 216 (2014) ("*Alice*"). The rationale is that patents directed to basic building blocks of technology would not "promote the progress of science" under the U.S. Constitution, Article I, Section 8, Clause 8, but instead would impede the same. Accordingly, laws of nature, natural phenomena, and abstract ideas, are not patent-eligible subject matter. *Thales Visionix Inc. v. United States*, 850 F.3d 1343, 1346 (Fed. Cir. 2017) (citing *Alice*, 573 U.S. at 216).

The Supreme Court has set forth a two-part *Alice/Mayo* test for subject matter eligibility in *Alice* (*Alice* 573 U.S. at 217–18). The first part is to determine whether the claim is directed to a patent-ineligible concept. *Id.* (citing *Mayo*, 566 U.S. at 76–77). If so, then the eligibility analysis proceeds to the second part of the *Alice/Mayo* test in which we "examine the elements of the claim to determine whether it contains an 'inventive concept' sufficient to 'transform' the claimed abstract idea into a patent-eligible application." *Alice*, 573 U.S. at 221 (internal quotation marks omitted) (quoting *Mayo*, 566 U.S. at 72, 79). There is no need to proceed to the second part, however, if the first part of the *Alice/Mayo* test yields a determination that the claim is directed to patent-eligible subject matter.

The Federal Circuit has described the *Alice* part-one inquiry as looking at the "focus" of the claims, their "character as a whole," and the *Alice* part-two inquiry as looking more precisely at what the claim elements add—whether they identify an "inventive concept" in the application of the ineligible matter to which the claim is directed. *See Elec. Power Grp., LLC v. Alstom S.A.*, 830 F.3d 1350, 1353 (Fed. Cir. 2016); *Enfish, LLC v. Microsoft Corp.*, 822 F.3d 1327, 1335–36 (Fed. Cir. 2016); *Internet Patents Corp. v. Active Network, Inc.*, 790 F.3d 1343, 1346 (Fed. Cir. 2015).

In an effort to achieve clarity and consistency in how the Office applies the Supreme Court's two-part framework, the Office published revised guidance interpreting governing case law and establishing a framework to govern all patent-eligibility analysis under *Alice* and § 101 effective as of January 7, 2019. *2019 Revised Patent Subject Matter Eligibility Guidance*, 84 Fed. Reg. 50–57 (Jan. 7, 2019) ("2019 Revised Guidance, 84 Fed. Reg."). The Manual of Patent Examining Procedure ("MPEP") has incorporated this revised guidance and subsequent updates at § 2106.

2019 Revised Guidance

Under the Revised Guidance, we first look under *Alice* part 1 or "Step 2A" to whether the claim recites:

(1) Prong One: any judicial exceptions, including certain groupings of abstract ideas (i.e., [i] mathematical concepts, [ii] mental processes, or [iii] certain methods of organizing human activity such as a fundamental economic practice or managing personal behavior or relationships or interactions between people); and

(2) Prong Two: additional elements that integrate the judicial exception into a practical application (*see* Manual of Patent Examining Procedure ("MPEP") §§ 2106.05(a)–(c), (e)–(h)).[4]

See 2019 Revised Guidance, 84 Fed. Reg. 51–52, 55, Revised Step 2A, Prong One (Abstract Idea) and Prong Two (Integration into A Practical Application). Only if a claim: (1) recites a judicial exception, and (2) does not integrate that exception into a practical application, do we then evaluate whether the claim provides an "inventive concept" under *Alice* step 2 or "Step 2B." *See* 2019 Revised Guidance 84 Fed. Reg. 56; *Alice*, 573 U.S. at 217–18. For example, we look to whether the claim:

1) adds a specific limitation beyond the judicial exception that is not "well-understood, routine, conventional" in the field (*see* MPEP § 2106.05(d)); or

2) simply appends well-understood, routine, and conventional activities previously known to the industry, specified at a high level of generality, to the judicial exception.

See 2019 Revised Guidance, 84 Fed. Reg. 56.

---

[4] All references to the MPEP are to the Ninth Edition, Revision 08.2017 (rev. Jan. 2018).

Alice/Mayo—Part 1 (Abstract Idea)
Step 2A–Prongs 1 and 2 identified in the Revised Guidance

Step 2A, Prong One

Turning to the first part of the *Alice* inquiry, we find Appellant's independent claims 1, 18, and 19 ("the claims") are directed to a patent-ineligible abstract concept of training a machine learning model (algorithm) on multiple machine learning tasks sequentially.

For example, Appellant's claim 1 recites a method of training a machine learning model wherein the machine learning model is trained on a first machine learning task in which (1) limitations [A]–[A2] require "determining, for each of the plurality of parameters, a respective measure of an importance of the parameter to the first machine learning task" via [A1] "computing . . . an approximation of a posterior distribution over possible values of the plurality of parameters," and [A2] "assigning, using the approximation, a value to each of the plurality of parameters . . . being the respective measure of the importance of the parameter to the first machine learning task"; (2) limitation [B] requires "obtaining second training data for training the machine learning model on a second, different machine learning task; (3) limitation [C] requires "training the machine learning model on the second machine learning task . . . to optimize performance of the machine learning model on the second machine learning task while protecting performance of the machine learning model on the first machine learning task," and (4) limitations [D]–[D1] requires "adjusting the first values of the plurality of parameters to optimize an objective function that depends in part on a penalty term that is based on the determined measures of importance of the plurality of parameters to the first machine learning task."

All limitations [A]–[D] recited in Appellant's claim 1 and, similarly, claims 18 and 19 can be broadly, but reasonably, interpreted as encompassing a mathematical algorithm computing mathematical calculations and manipulating particular information, i.e., values of certain parameters to train a machine learning model, similarly to the claims in *Gottschalk v. Benson*, 409 U.S. 63, 67 (1972).

Information as such is intangible, and data analysis and algorithms are abstract ideas. *See, e.g.*, *Microsoft Corp. v. AT&T Corp.*, 550 U.S. 437, 451 n.12 (2007); *Alice*, 134 S. Ct. at 2355; *Parker v. Flook*, 437 U.S. 584, 594–95 (1978) ("Reasoning that an algorithm, or mathematical formula, is like a law of nature, *Benson* applied the established rule that a law of nature cannot be the subject of a patent."); *Gottschalk*, 409 U.S. at 71–72. Similarly, information collection and analysis, including when limited to particular content, is within the realm of abstract ideas. *See, e.g.*, *Internet Patents Corp. v. Active Network, Inc.*, 790 F.3d 1343, 1349 (Fed. Cir. 2015); *Digitech Image Techs., LLC v. Elecs. for Imaging, Inc.*, 758 F.3d 1344, 1351 (Fed. Cir. 2014); *CyberSource Corp. v. Retail Decisions, Inc.,* 654 F.3d 1366, 1370 (Fed. Cir. 2011). That is, "[w]ithout additional limitations, a process that employs mathematical algorithms to manipulate existing information to generate additional information is not patent eligible." *Digitech*, 758 F.3d at 1349–51 ("Data in its ethereal, non-physical form is simply information that does not fall under any of the categories of eligible subject matter under section 101.").

Thus, under Step 2A, Prong One, we determine that the claims recite an abstract idea identified in the Revised Guidance.

Step 2A, Prong Two (Integration into a Practical Application)

Under *Step 2A, Prong Two* of the Revised Guidance, we must determine if the claims (specifically, using any additional limitations beyond the judicial exception) integrate the judicial exception into a practical application. However, we discern no additional element (or combination of elements) recited in Appellant's claims 1, 18, and 19 that may have integrated the judicial exception into a practical application. *See* Revised Guidance, 84 Fed. Reg. 54–55. For example, Appellant's claimed additional elements (e.g., "one or more computers" and "one or more storage devices" recited in claims 18 and 19) (1) do not improve the functioning of a computer or other technology; (2) are not applied with any particular machine (except for a generic computer); (3) do not effect a transformation of a particular article to a different state; and (4) are not applied in any meaningful way beyond generally linking the use of the judicial exception to a particular technological environment, such that the claim as a whole is more than a drafting effort designed to monopolize the exception. *See* MPEP § 2106.05(a)–(c), (e)–(h);

Likewise, training a machine learning model (algorithm) on multiple machine learning tasks in sequence does not provide any "technical solution to a technical problem" as contemplated by the Federal Circuit in (1) *Enfish, LLC v. Microsoft Corp.,* 822 F.3d 1327 (Fed. Cir. 2016); (2) *McRO, Inc. v. Bandai Namco Games Am. Inc.*, 837 F.3d 1299 (Fed. Cir. 2016); (3) *Amdocs (Isr.) Ltd. v. Openet Telecom, Inc.*, 841 F.3d 1288 (Fed. Cir. 2016); and (4)

Thales Visionix Inc. v. United States, 850 F.3d 1343 (Fed. Cir. 2017). *See* MPEP § 2106.05(a).

For example, none of the limitations [A]–[D] recited in Appellant's claim 1 requires, and nowhere in Appellant's Specification is there any description or explanation as to how the training of a machine learning model (algorithm) on multiple machine learning tasks in sequence provides: (1) "a specific improvement to the way computers operate," as explained in *Enfish*, 822 F.3d at 1336; (2) "a process specifically designed to achieve an improved technological result in conventional industry practice," as explained in *McRO*, 837 F.3d at 1316; or (3) an "unconventional technological solution (enhancing data in a distributed fashion) to a technological problem (massive record flows which previously required massive databases)" and "improve the performance of the system itself" as explained in *Amdocs*, 841 F.3d at 1300, 1302; or even (4) "a non-conventional manner to reduce errors in measuring the relative position and orientation of a moving object on a moving reference frame," as explained in *Thales,* 850 F.3d at 1349. There is no evidence in the record to establish that claims 1 and 18–19 recite a specific technological improvement to computers. *See Enfish*, 822 F.3d at 1336.

The focus of Appellant's invention is not to improve the performance of computers or any underlying technology; instead, the focus of Appellant's invention is to train a machine learning model (algorithm) on multiple machine learning tasks in sequence, including "a first machine learning task using first training data" and "a second machine learning task using second training data." Spec. ¶¶ 2–3, 21, 27–29, 30–35. Any improvements embodied by the claims are merely improvements to this abstract idea.

For these reasons, we do not find any "additional elements" recited in claims 1, 18, and 19 integrate the abstract idea into a practical application.

Alice/Mayo—Part 2 (Inventive Concept)
Step 2B identified in the Revised Guidance

Under the 2019 Revised Guidance, only if a claim: (1) recites a judicial exception; and (2) does not integrate that exception into a practical application, do we then look to whether the claim adds a specific limitation beyond the judicial exception that is not "well-understood, routine, conventional" in the field (*see* MPEP § 2106.05(d)); or, simply appends well-understood, routine, conventional activities previously known to the industry, specified at a high level of generality, to the judicial exception. *See* 2019 Revised Guidance, 84 Fed. Reg. 56. However, we find no element or combination of elements recited in Appellant's claims 1, 18 and 19 that contain any "inventive concept" beyond the abstract concept or add anything "significantly more" to transform the abstract concept into a patent-eligible application. *Alice*, 573 U.S. 208 at 221.

Utilizing generic computer components (e.g., processors and memory) to train a machine learning model (algorithm) on multiple machine learning tasks in sequence does not alone transform an otherwise abstract idea into patent-eligible subject matter. As our reviewing court has observed, "after *Alice*, there can remain no doubt: recitation of generic computer limitations does not make an otherwise ineligible claim patent-eligible." *DDR*, 773 F.3d at 1256 (citing *Alice*, 573 U.S. at 222); *see* Spec. ¶¶ 2–3, 21, 27–29, 30–35.

Because Appellant's claims 1, 18, and 19 are directed to no more than a patent-ineligible abstract concept and do not recite something "significantly more" under the second prong of the *Alice* analysis, we issue a new ground of rejection of claim 1, 18, and 19 and their respective dependent claims 2–6, 8–17 and 20 under 35 U.S.C. § 101 as being directed to non-statutory subject matter.

CONCLUSION

On the record before us, Appellant does not persuade us of Examiner error in rejecting (1) claims 1, 2, 4, 9, 15, and 17–20 as obvious over the combined teachings of Marcheret, Jamaluddin, Rousset, Gordon, and Mehanna; (2) claims 3 and 14 as obvious over the combined teachings of Marcheret, Jamaluddin, Rousset, Gordon, Mehanna, and Aslan; (3) claims 6 and 10–12 as obvious over the combined teachings of Marcheret, Jamaluddin, Rousset, Gordon, Mehanna, and Cao; and (4) claims 5, 8, 13, and 16 as obvious over the combined teachings of Marcheret, Jamaluddin, Rousset, Gordon, Mehanna, and Sinyavskiy. However, because we have relied on facts and reasoning not raised by the Examiner, we designate our affirmance as a new ground of rejection of claims 1, 2, 4, 9, 15, and 17–20 as obvious over the combined teachings of Marcheret, Jamaluddin, Rousset, Gordon, and Mehanna.

Separately, pursuant to our authority under 37 C.F.R. § 41.50(b), we enter a new ground of rejection for claims 1–6 and 8–20 under 35 U.S.C. § 101 for lack of patent-eligible subject matter.

Rule 37 C.F.R. § 41.50(b) states that "[a] new ground of rejection pursuant to this paragraph shall not be considered final for judicial review." Further, § 41.50(b) also provides that Appellants, WITHIN TWO MONTHS FROM THE DATE OF THE DECISION, must exercise one of the following two options with respect to the new grounds of rejection to avoid termination of the appeal as to the rejected claims:

(1) *Reopen prosecution*. Submit an appropriate amendment of the claims so rejected or new evidence relating to the claims so rejected, or both, and have the matter reconsidered by the examiner, in which event the prosecution will be remanded to the examiner. . . .

(2) *Request rehearing*. Request that the proceeding be reheard under § 41.52 by the Board upon the same record. . . .

No time period for taking any subsequent action in connection with this appeal may be extended under 37 C.F.R. § 1.136(a). *See* 37 C.F.R. § 1.136(a)(1)(iv) (2023).

II. DECISION SUMMARY

In summary:

Claim(s) Rejected	35 U.S.C. §	Reference(s)/Basis	Affirmed	Reversed	New Ground
1, 2, 4, 9, 15, 17–20	103	Marcheret, Jamaluddin, Rousset, Gordon, Mehanna	1, 2, 4, 9, 15, 17–20		
3, 14	103	Marcheret, Jamaluddin, Rousset, Gordon, Mehanna, Aslan	3, 14		
6, 10–12	103	Marcheret, Jamaluddin, Rousset, Gordon, Mehanna, Cao	6, 10–12		
5, 8, 13, 16	103	Marcheret, Jamaluddin, Rousset, Gordon, Mehanna, Sinyavskiy	5, 8, 13, 16		
1–6, 8–20	101	Eligibility			1–6, 8–20
Overall Outcome			1–6, 8–20		1–6, 8–20

AFFIRMED
37 C.F.R. § 41.50(b)

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

Ex parte GUILLAUME DESJARDINS, RAZVAN PASCANU,
RAIA THAIS HADSELL, JAMES KIRKPATRICK,
JOEL WILLIAM VENESS, and NEIL CHARLES RABINOWITZ

Appeal 2024-000567
Application 16/319,040
Technology Center 2100

Before HUNG H. BUI, NABEEL U. KHAN, and JOHN F. HORVATH, *Administrative Patent Judges*.

HORVATH, *Administrative Patent Judge*, Concurring-in-part

I concur with the majority's decision to reject the pending claims under 35 U.S.C. § 101 as directed to unpatentable subject matter and to designate that rejection a new ground of rejection. However, I would reverse the Examiner's rejection under 35 U.S.C. § 103.

As discussed above, Appellant's invention is directed to training a parameterized machine learning model on a first task, then subsequently training the model on a second (different) task while protecting the model's performance on the first task. To do this, Appellant (1) determines a measure of importance for each of the parameters trained in the performance of the first task by calculating an approximate posterior distribution of the parameters, (2) uses the approximation to assign a probability to each parameter indicating the likelihood that the parameter is correctly valued and reflecting its importance to performance of the first task, and (3) adjusts the parameter values to optimize the model's performance of the second task using an objective function having a penalty term that is based on the assigned parameter probabilities, i.e., on how important the parameter values are in the model's performance of the first task.

According to Appellant's Specification, a posterior distribution over the parameters of a model "assigns a value to the current value of [each] parameter in which the [assigned] value represents a probability that the current value is a correct value of the parameters." Spec. ¶ 50. That is, the step of determining a posterior distribution (step (1) above) is identical to the step of using the posterior distribution to assign a probability value to each parameter indicating the likelihood that the parameter is correctly valued for the model's performance of task 1 (i.e., step (2) above). Said differently, the result of the limitation [A1] step of computing a posterior distribution over parameter values of the plurality of (task 1 trained) parameters is the limitation [A2] step of assigning a value to each of the parameters reflecting the importance of the parameter to the model's performance of task 1, where the assigned value is the probability that the parameter's value was correctly determined by training the model to perform task 1.

As the Examiner correctly finds, Gordon teaches limitation [A1]. *See* Final Act 13 (citing Gordon ¶ 14); Gordon ¶ 14 (disclosing for a "[g]iven set of training data d=(x,y), Bayes' rule expressions may be obtained for computing a posterior distribution p(w|d, h)" and that "[t]his Bayesian model represents a wide variety of machine learning tasks"). Thus, Gordan also teaches limitation [A2] because, according to Appellant's own Specification, Gordon's posterior distribution *is* a measure of the importance of a trained parameter that expresses the probability that the trained parameter has a correct value (i.e., that it has been correctly trained).

The Examiner errs finding Mehanna teaches limitation [A2] for the reasons stated by Appellant. *See* Appeal Br. 8–10 (arguing Mehanna's "features" are inputs to its machine learning model, not parameters of the model). The Majority disagrees with Appellant's argument, finding the "inputs to" and the "parameters of" a machine learning model have the same meaning under a broadest reasonable interpretation. Maj. Op. 10–11 (citing Spec. ¶¶ 31–34, Figs. 1–2). I respectfully disagree.

The Specification discloses a machine learning model can "receive an input and generate an output . . . based on the received input." Spec. ¶ 31. It also discloses the machine learning model may be a parameterized model, in which case the model receives an input and "generates the output based on the received input and on values of the parameters of the model." *Id.* ¶ 32. By using two different terms differently, the Specification discloses the terms have different meanings, even under a broadest reasonable interpretation standard. Accordingly, I find the Examiner errs in finding Mehanna teaches limitation [A2], and would reverse the Examiner's § 103 rejection for that reason.