UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

TESLA, INC.

Petitioner,

v.

PERCEPTIVE AUTOMATA LLC

Patent Owner

IPR2025-01575
U.S. Patent No.: 11,753,046 B2

DECLARATION OF DR. JASON JANÉT

REGARDING U.S. PATENT NO. 11,753,046

TABLE OF CONTENTS

I. INTRODUCTION

A. Engagement

1. My name is Jason Janét, I have been retained by counsel for Tesla, Inc. ("Petitioner" or "Tesla") as an expert witness in the above-captioned proceeding. I have been asked to provide my opinion about the state of the art of the technology described in U.S. Patent No. 11,753,046 ("the '046 Patent") (EX1001) and on the patentability of the claims of this patent. The following is my written declaration on these topics.

B. Background and Qualifications

2. My credentials are set forth in full detail in my Curriculum Vitae which is attached as Exhibit 1013, incorporated here by reference, and summarized as follows: I am currently CTO for Applied Research Associates (ARA). I also hold the rank of Adjunct Associate Professor at Duke University and at North Carolina State University.

3. I received a Bachelor of Science in Mechanical Engineering from the University of Virginia in 1990; a Master of Integrated Manufacturing Systems from the Integrated Manufacturing Systems Engineering Institute at North Carolina State University in 1994; and a Ph.D. in Electrical and Computer Engineering from North Carolina State University in 1998.

4. Since 1991, I have been active in the robotics, sensors, artificial intelligence, pattern analysis, and automation fields. My MS thesis, entitled *Global*

Motion Planning and Self-Referencing for Autonomous Mobile Robots, leveraged graph theory and time-of-flight and/or Doppler sensors to optimize mobile robot movement through, and perception of, environments. My PhD dissertation, entitled *Pattern Analysis, Tracking and Control for Autonomous Vehicles*, leveraged multiple UGVs, time-of-flight and/or Doppler sensors, point clouds, and neural networks to enhance simultaneous localization and mapping (SLAM), as well as the sharing of knowledge between multiple different UGVs. I have authored numerous publications and have co-authored a textbook entitled *Computational Intelligence*.

5. I have designed, built, and marketed toys, robots, automated systems, measuring systems and components thereof, including ground mobile robots, unmanned aerial vehicles (UAV), unattended sensors, automated storage and retrieval systems (ASRS), submersible mobile robots, proof-of-concept extraterrestrial robots, manually-manipulated ("twiddled") submersible foils, wall climbing mobile robots, corner-grip ballgame goals, proximity and ranging systems, point-cloud systems, pattern analysis tools, and augmented reality systems.

6. I have taught the following courses: Introduction to Robotics and Automation (Duke and NCSU); Introduction to Control Theory (Duke and NCSU); Distributed Real-Time Controls (NCSU); and myriad independent studies courses in the areas of robotics, automation, artificial intelligence, autonomy, pattern analysis, sensors, and control systems. I have also served on several MS- and PhD-

level graduate student committees, designed qualifying exam problems, served on the NCSU IMSEI Board, and participated in curriculum development for undergraduate and graduate level programs.

7. I have also started, sponsored, and advised multiple champion student teams for international robot competitions including, but not limited to, the NASA/ASCE Extra-Terrestrial Robotics Competition, the DARPA Grand Challenge, the AUVSI/ONR Autonomous Underwater Vehicle Competition, and the European CLAWAR Wall-Climbing Robot competition.

8. I have initiated multiple unmanned systems projects at both academic and industry levels. Academic unmanned aerial vehicles (UAVs) include, but are not limited to, Quadcopters with Hybrid Remote and Autonomous Control, Marsupial UAVs that Deploy and Recover Unmanned Ground Vehicles, and Wall-Climbing UAVs. Additionally, the AngelFish Cross-Domain Submersible UAV, a DARPA ASW program that I incubated and procured funding for at Teledyne, included a partnership with North Carolina State University. Unmanned underwater vehicles (UUVs) include, but are not limited to, the MicroHunter-class of very small underwater robots, the PilotFish, Gamera, Artemis and other holonomic robots, and submersible crawlers that are used in nuclear boiler water reactors, the LBC hull inspection system, and the HULLBUG hull cleaning system. Unmanned ground

vehicles (UGVs) include, but are not limited to, the miniature PERRY, wallclimbers, all-terrain and subterranean crawlers.

9. I have also supported multiple toy-focused ventures with companies including, but not limited to, Marvel, Hasbro, Twidco, Spin Master, Nottingham-Spirk, IDEO, ARA, and other firms. For Marvel, multiple Spiderman toys were conceived and developed and validated, including the Spiderman Wall-Climbing Car, the Corner-Climbing Spider, and the Inflatable Wall-Climbing Spiderman. For Hasbro, a proprietary corner-gripping technology was employed to place basketball goals in corners. For Twidco, a proprietary flapping foil technology was combined with multiple fish shapes (e.g., shark and clownfish) to enable users to create fishlike motion by "twiddling" a connecting cable from above. For Spin Master and multiple design firms, multiple wall-climbing toys were developed for small wall-climbing robot concepts including simple radio control vehicles and the "Hangtime 540" Tony Hawk aerial skateboard toy. For ARA, augmented reality systems for indoor and outdoor games.

10. In 1999, while I was employed by Nekton Research ("Nekton"), I captured and managed multiple programs sponsored by the Department of Defense (DoD) and private-sector companies that focused primarily on autonomous underwater vehicles (AUV), remotely operated vehicles (ROV), TwiddleFish toys, and indirect-fire projectiles. During my tenure, Nekton entered into a joint venture

agreement with the founders of what eventually became known as Parata Systems – a pharmacy automation solutions company, which I supported launching.

11. After leaving Nekton in March 2002, I joined a then New Jersey-based company called Avionic Instruments, Inc ("Avionic"). While at Avionic, I continued developing and marketing robots and supported engineering related to various design, manufacturing, quality and assembly issues on the core aerospace product lines. Avionic product lines include, but are not limited to, ducted fans, transformer-rectifier units (TRU), regulated TRU (RTRU), auxiliary power unit (APU) control systems, power distribution systems (PDU), frequency converters, corner clamps, and VRAM attractors/thrusters. Customers included DoD, NASA, Boeing, Sikorsky, Augusta-Westland, Dassault, and Lockheed-Martin.

12. At Avionic, I was also tasked with ruggedizing, optimizing, and commercializing two emerging proprietary technologies: the "CFG" and "VRAM". The CFG, short for Corner Friction Grip, was a proprietary technology based on non-marring and non-adhesive materials that provided optimal grip patterns on inside and outside corners. CFG formed the basis of a company called "Shelf Works Technologies," which was spun out of Avionic instruments in 2004, and whose products were sold in Home Depot, Bed Bath and Beyond, etc. CFG also formed the basis of multiple toys including a miniature basketball goal for Hasbro, and a corner climbing Spiderman toy for Marvel.

13. The VRAM, short for Vortex Regenerative Air/Aqueous Movement is a patented method for producing attraction forces between an object and a surface. VRAM is euphemistically referred to as a "tornado in a cup" and generally comprises both cylindrical and toroidal flow patterns that result in vacuum-like regions. The VRAM had several applications, including but not limited to: holding breaching charges and sensors against vertical or inverted surfaces; acting as an attractor for wall and ceiling climbing robots; acting as an attractor for submersible hull-crawling robots; filter-less vacuuming (conceptually similar to the Dyson vacuum); and robotic pick-and-place end-effector (conceptually similar to a gripper) to move articles from one location (e.g., a conveyor) to another (e.g., a collator). The VRAM also formed the basis for multiple toys including: a wall-climbing car for Spin Master, IDEO and Nottingham-Spirk; a Spiderman car for Marvel; an inflatable wall-climbing Spiderman for Marvel; and the Tony Hawk Hangtime 540 aerial skateboard for IDEO.

14. My work with the filter-less vacuum VRAM prompted me to explore a concept related to pharmacy automation, which ultimately led to a Ferris-wheel concept for rapidly dispensing pills. In mid-2003, my team conceived of two prescription pill counting mechanisms that actually benefitted from centrifugal forces and optimally exploited gravity and vacuum: an inner ring with vacuum-based apertures and an inner-bowl with vacuum-based apertures. The inner-bowl with

vacuum-based apertures was deemed most viable, was awarded two US patents, and to this day forms the basis of the RxMedic ADS™ pharmacy robot. The VRAM formed the basis of Vortex HC technologies, many of which have been licensed out to entities including, but not limited to Teledyne SeaBotix, SeaRobotics, and HDT. Both RxMedic and Vortex HC were officially spun out of Avionic in July 2004, at the date Avionic was acquired by Transdigm.

15. In late 2004, soon after Avionic was acquired, I, along with others from Avionic, secured funding to develop an alpha-level multi-dispenser pharmacy robotic system and develop the RxMedic business plan and raise multiple rounds of venture capital. In late 2006, RxMedic (called "APDS" until November 2006) was launched as a stand-alone, sole-focus venture. After the operational launch of RxMedic, I served as General Manager and eventually Chief Technical Officer. Through my roles at RxMedic, I oversaw the development of the RxMedic ADS robot, managed the intellectual property portfolio, coordinated sales and marketing, and provided strategic, fiscal, and operational leadership. In May 2010, J.M Smith Corporation acquired RxMedic, and to assist in the change of ownership, I served as a Director of RxMedic until May 2011.

16. Also in late 2004, after Avionic was acquired, I, along with others from Avionic, secured funding for Vortex HC through DoD contracts and robot sales, to continue developing the VRAM Mobile Robot Platform (VMRP – a wall-climbing

robot), the ARTEMIS AUV (a holonomic submersible robot for counter-mine and counter-obstacle operations), the submersible crawler, the nuclear-grade boiler water reactor (BWR) inspection robot, and other robot products centered around the VRAM. Some DoD programs were/are classified, for which I maintained a SECRET clearance at both the personal and facilities level, and served as the facilities security officer (FSO). In late 2006, corresponding to the full launch of RxMedic, Vortex HC technologies were largely licensed to Teledyne SeaBotix, SeaRobotics, and HDT. However, I have continued to support Vortex HC licensees and customers to-date.

17. After the sale of RxMedic, and after fulfilling my 12-month employment obligation, I joined Teledyne Technologies in Summer 2011. I served as the Senior Manager for the RTP division of Teledyne Scientific, and supported multiple DoD-sponsored robotic-focused programs. Some of these programs were/are classified, for which I maintained a personal SECRET clearance. Among these programs, were cargo unmanned ground vehicles (CUGV); squad-level autonomous unmanned ground vehicles (UGV); autonomous underwater vehicles (AUV); unmanned underwater vehicles (UUV); a cross-domain autonomous vehicle capable of transitioning between air-, surface- and underwater-domains; the EXACTO program for optically tracking and steering bullets to targets; and LIDAR systems for UGVs, as well as the mapping of urban and subterranean environments (along with sister-company Optech). Cross-domain vehicles that were evaluated,

designed and prototyped included, but were not limited to, the AngelFish (later called “EagleRay”) cross-domain submersible UAV for anti-submarine warfare, as well as a ball-shaped robot for beach-, ground- and surf-zone countermine operations. Sensor design, refinement and signal processing was a major component of each program. Sensors employed include, but are not limited to: proximity sensors; ranging sensors; electro-optical imaging; LIDAR. long-, short- and midwave infrared (IR); inertial measurement units (IMU); optical flow; and radiofrequency (RF). Additionally, control systems were designed, refined and integrated into the aforementioned systems. Most control systems were closed-loop, in that they utilized sensor-based feedback; others were open-loop, where states were estimated with little or no feedback.

18. In late 2013, Avionic, a Transdigm business unit at that point, requested that I return to turn around both a supply-chain issue, and the Sikorsky S97 Raider/JMR helicopter programs. The S97 Raider was designed to be the fastest, most maneuverable helicopter, due to its coaxial, counter-rotating variable-pitch wings, and an aft-based push-propeller. Avionic also controlled two business units named Acme Aerospace (Acme) and Aerospace Cooling Solutions (ACS). Avionic, Acme and ACS supported the S97 program, which ultimately met milestones and continues to produce multiple successful demonstrations. In 2013, Transdigm expanded my role to include directorship of the Avionic, Acme and ACS revenue

and engineering teams, and to report operational and financial status at six-week intervals. Transdigm required that I move my family to New Jersey in late 2014, which influenced my decision to resign and assume the CEO role offered to me by Delta Five Systems in Raleigh, NC.

19. As CEO of Delta Five, I led a private-equity backed, hospitality-focused venture, coordinating the company's launch and strategic path including an early-stage pivot. Initially focused on back-end robotics and automation, Delta Five shifted its focus to address the rampant bed bug problem—a top priority for hoteliers—with a novel unattended sensor and trap, called the Telemetered Pest Monitoring System (TPMS), which is based on computer vision and internet of things (IoT), and has proven capable of scaling to other pests. In addition to the TPMS, Delta Five developed a novel means to mass produce a natural, unscented aggregation pheromone that, in concert with placement and heat, lured invertebrates to the TPMS. After leading Delta Five and serving on its Board for four years, I resigned to join ARA but continue serving Delta Five as a share-holder and transition agent.

20. In 2018, I joined ARA as CTO. My primary responsibilities include leading the commercialization and technology transition efforts, which require technical, business and corporate development, as well as the procurement of capital and managing matrixed rosters.

21. I have served as an expert witness on multiple cases including cases involving IPR, ITC, contract disputes, and civil litigation in District and Federal Courts. My resume lists these cases.

C. Compensation and Prior Testimony

22. I am being compensated at a rate of $500.00 per hour for my study and testimony in this matter. My compensation is not contingent on the outcome of this matter or the specifics of my testimony.

23. My CV, attached as Exhibit 1013, lists all other cases in which, during the previous 16 years, I have testified as an expert at trial or by deposition.

D. Information Considered

24. My opinions are based on my years of education, research and experience, as well as my investigation and study of relevant materials. In forming my opinions, I have considered the materials I identify in this report and those listed in the Exhibit List at the end of this declaration.

25. I may rely upon these materials and/or additional materials to respond to arguments raised by the Patent Owner. I may also consider additional documents and information in forming any necessary opinions — including documents that may not yet have been provided to me.

26. My analysis of the materials produced in this matter is ongoing and I will continue to review any new materials as it is provided. This declaration represents only those opinions I have formed to date.

II. LEGAL STANDARDS FOR PATENTABILITY

27. In expressing my opinions and considering the subject matter of the claims of the ’046 Patent, I am relying upon certain basic legal principles that have been explained to me.

28. I understand that for an invention claimed in a patent to be found patentable under 35 U.S.C. §§ 102 and 103, it must be, among other things, new and not obvious from what was known before the invention was made.

29. I understand the information that is used to evaluate whether an invention is new and not obvious is generally referred to as “prior art” and generally includes patents and printed publications (e.g., books, journal publications, articles on websites, product manuals, etc.).

30. I understand that in this proceeding the Petitioner has the burden of proving that the claims of the ’046 Patent are anticipated by or obvious from the prior art by a preponderance of the evidence. I understand that “a preponderance of the evidence” is evidence sufficient to show that a fact is more likely true than it is not.

31. I understand that in this proceeding, the claims should be given their ordinary and accustomed meaning as understood by one of ordinary skill in the art in view of the patent and its file history. The claims after being construed in this manner are then to be compared to the information in the prior art.

32. I understand that in this proceeding, the information that may be evaluated is limited to patents and printed publications. My analysis below compares the claims to patents and printed publications that are prior art to the claims.

33. I understand that there are two ways in which prior art may render a patent claim unpatentable. First, the prior art can be shown to "anticipate" the claim. Second, the prior art can be shown to have made the claim "obvious" to a person of ordinary skill in the art. In this declaration I focus on "obviousness." My understanding of the legal standard for "obviousness" is set forth below.

34. I understand that a claimed invention is not patentable if it would have been obvious to a person of ordinary skill in the field of the invention at the time the invention was made.

35. I understand that the obviousness standard is defined in the patent statute (35 U.S.C. § 103(a)) as follows:

36. A patent may not be obtained, though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the

subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

37. I understand that the following standards govern the determination of whether a claim in a patent is obvious. I have applied these standards in my evaluation of whether the claims of the ’046 Patent would have been considered obvious as of the priority date of the patent.

38. I understand that to find a claim in a patent obvious, one must make certain findings regarding the claimed invention and the prior art. Specifically, I understand that the obviousness question requires consideration of four factors (although not necessarily in the following order):

39. The scope and content of the prior art;

40. The differences between the prior art and the claims at issue;

41. The knowledge of a person of ordinary skill in the pertinent art; and

42. Whatever objective factors indicating obviousness or non-obviousness may be present in any particular case.

43. In addition, I understand that the obviousness inquiry should not be done in hindsight, but must be done using the perspective of a person of ordinary skill in the relevant art as of the effective filing date of the patent claim.

44. I understand the objective factors indicating obviousness or non-obviousness may include: commercial success of products covered by the patent claims; a long-felt need for the invention; failed attempts by others to make the invention; copying of the invention by others in the field; unexpected results achieved by the invention; praise of the invention by those in the field; the taking of licenses under the patent by others; expressions of surprise by experts and those skilled in the art at the making of the invention; and the patentee proceeded contrary to the accepted wisdom of the prior art. I also understand that any of this evidence must be specifically connected to the invention rather than being associated with the prior art or with marketing or other efforts to promote an invention. I am not presently aware of any evidence of "objective factors" suggesting the claimed methods are not obvious, and reserve my right to address any such evidence if it is identified in the future.

45. I understand the combination of familiar elements according to known methods is likely to be obvious when it does no more than yield predictable results. I also understand that an example of a solution in one field of endeavor may make that solution obvious in another related field. I also understand that market demands or design considerations may prompt variations of a prior art system or process, either in the same field or a different one, and that these variations will ordinarily be considered obvious variations of what has been described in the prior art.

46. I also understand that if a person of ordinary skill can implement a predictable variation, that variation would have been considered obvious. I understand that for similar reasons, if a technique has been used to improve one device, and a Skilled Artisan would recognize that it would improve similar devices in the same way, using that technique to improve the other device would have been obvious unless its actual application yields unexpected results or challenges in implementation.

47. I understand that the obviousness analysis need not seek out precise teachings directed to the specific subject matter of the challenged claim, but instead can take account of the "ordinary innovation" and experimentation that does no more than yield predictable results, which are inferences and creative steps that a Skilled Artisan would employ.

48. I understand that sometimes it will be necessary to look to interrelated teachings of multiple patents; the effects of demands known to the design community or present in the marketplace; and the background knowledge possessed by a person having ordinary skill in the art. I understand that all these issues may be considered to determine whether there was an apparent reason to combine the known elements in the fashion claimed by the patent at issue.

49. I understand that the obviousness analysis cannot be confined by a formalistic conception of the words "teaching, suggestion, and motivation." I

understand that in 2007, the Supreme Court issued its decision in *KSR Int'l Co. v. Teleflex, Inc.*, 550 U.S. 398 (2007), where the Court rejected the previous requirement of a "teaching, suggestion, or motivation to combine" known elements of prior art for purposes of an obviousness analysis as a precondition for finding obviousness. It is my understanding that *KSR* confirms that any motivation that would have been known to a person of skill in the art, including common sense, or derived from the nature of the problem to be solved, is sufficient to explain why references would have been combined.

50. I understand that a person of ordinary skill attempting to solve a problem will not be led only to those elements of prior art designed to solve the same problem. I understand that under the *KSR* standard, steps suggested by common sense are important and should be considered. Common sense teaches that familiar items may have obvious uses beyond their primary purposes, and in many cases a person of ordinary skill will be able to fit the teachings of multiple patents together like pieces of a puzzle. As such, the prior art considered can be directed to any need or problem known in the field of endeavor as of the priority date of the patent and can provide a reason for combining the elements of the prior art in the manner claimed. In other words, the prior art does not need to be directed towards solving the same problem that is addressed in the patent. Further, the individual prior art references themselves need not all be directed towards solving the same problem.

51. I understand that obviousness does not require that the features of a secondary reference be bodily incorporated into the structure of the primary reference. Rather, the test is what the combined teachings of those references would have suggested to a Skilled Artisan. The disclosures of the prior art references need not be physically combinable. Combining the teachings of references should be the focus of the analysis.

52. I understand that an invention that might be considered an obvious variation or modification of the prior art may be considered non-obvious if one or more prior art references discourages or lead away from the line of inquiry disclosed in the reference(s). A reference does not "teach away" from an invention simply because the reference suggests that another embodiment of the invention is better or preferred. My understanding of the doctrine of teaching away requires a clear indication that the combination should not be attempted (*e.g.*, because it would not work or explicit statements saying the combination should not be made).

53. I understand that a person of ordinary skill is also a person of ordinary creativity.

54. I further understand that in many fields, it may be that there is little discussion of obvious techniques or combination, and it often may be the case that market demand, rather than scientific literature or knowledge, will drive design trends. When there is such a design need or market pressure to solve a problem and

there are a finite number of identified, predictable solutions, a person of ordinary skill has good reason to pursue the known options within their technical grasp. If this leads to the anticipated success, it is likely the product not of innovation but of ordinary skill and common sense. In that instance the fact that a combination was obvious to try might show that it was obvious. The fact that a particular combination of prior art elements was "obvious to try" may indicate that the combination was obvious even if no one attempted the combination. If the combination was obvious to try (regardless of whether it was actually tried) or leads to anticipated success, then it is likely the result of ordinary skill and common sense rather than innovation.

55. I understand that in *Praxair Distribution, Inc. v. Mallinckrodt Hosp. Prods. IP Ltd.*, the Federal Circuit held that claim limitations directed to the content of information that have no functional relationship are not patent eligible under 35 U.S.C. § 101. 890 F.3d 1024, 1032 (Fed. Cir. 2018) ("Claim limitations directed to the content of information and lacking a requisite functional relationship are not entitled to patentable weight because such information is not patent eligible subject matter under 35 U.S.C. § 101.").

III. BACKGROUND TO THE '046 PATENT

A. Prosecution History

56. The '046 Patent issued from U.S. Patent Application 17/468,516, filed September 7, 2021. EX1001, Cover Page. The Office issued a first non-final office

action that included only a non-statutory double patenting rejection over U.S. Patent 11,126,889, the parent of the ’046 Patent, and noted that the pending claims would be allowable provided this rejection is overcome. EX1002, 93-100. In response, the applicant filed a terminal disclaimer and made minor clarifying amendments to the claims. *Id.*, 115, 122-129. The Office subsequently issued a notice of allowance. *Id.*, 224-230. The Office did not identify any particular claim recitations that rendered the claims allowable. *Id.* But as demonstrated by the grounds of this petition, the Challenged Claims should not have been determined as allowable and should not have issued.

B. Technology Background

57. The ’046 Patent generally relates to supervised machine learning models, which are computer programs that analyze input data and generate an output based on the data. Such models are “supervised” because they are trained using labeled training data that is known to represent a correct output. Human annotators (also referred to as “labelers” or “reviewers” or “annotators”) can be used to label such training data. After training, the model can be applied to new, non-training data, to provide an output for the new data that reflects the human annotators’ decision making when labeling training data.

58. For example, in my opinion, to develop a vehicle detection supervised machine learning model, human annotators may label all instances of training data

images that display a vehicle. The model is then trained using the labeled training data and after training, the model can detect vehicles in new, non-training data such as camera images. The model's detection of vehicles reflects how the human annotators would have labeled the new images because the model was trained based on the human annotators' decision making when labeling training data.

C. The Alleged Invention

59. The alleged invention of the '046 Patent relates to (1) training a machine learning model using human-labeled training data and (2) applying the trained model to new input data to generate an output that replicates how humans would have labeled the new input data. EX1001, Abstract, 6:14-7:50. But this alleged invention was well-known. Section VI; *see also, e.g.*, EX1012, 13:30-14:9. Indeed, the *inventor's own prior art* shows that the '046 Patent should not have issued. The Cox reference, which shares a common inventor with the '046 Patent (i.e., Samuel Anthony), discloses machine learning model predictions that are "more consistent with the decisions of the human annotators." EX1007, ¶0032. Neither the patentee nor Mr. Anthony brought the Cox reference to the patent office's attention during prosecution of the '046 Patent. *See* EX1001; EX1002.

60. In my opinion, the substance of the '046 Patent further reveals that it is not new. The patent explains how model training data can be obtained and labeled by human annotators. EX1001, 4:36-61, 5:22-40, 5:52-6:10, Figs. 1, 2B. This

concept was well known previously, as demonstrated by (at least) Cox. EX1007, ¶¶0008-0009, 0029-0032 (human annotators annotate training images). The patent then explains that such human labels can be used to generate "summary statistics" that "may characterize the aggregate responses of multiple human observers" to a particular training image. EX1001, 6:14-24. But this language simply refers, for example, to generating a tally of the annotator responses—e.g., "how many" annotators labeled a training image in a certain way. *Id.* Again, such tallying was well known. EX1006, ¶0140 (tallying how many annotators labeled a training sample in a certain way). And in another example, the patent explains that "summary statistics" relate to how long a human annotator took to label training data or the position of the annotator's eyes when labeling training data. EX1001, 6:24-28. Yet again, this was well known in the art and at least the Cox reference demonstrates this. *See, e.g.*, EX1007, ¶¶0006-0008, 0018, 0063.

61. Thereafter, the patent notes that a machine learning model is trained using the responses, the model is used to make a prediction based on "live" input data. EX1001, 6:50-7:21. In my opinion, this was also well known, at least because all prediction models trained on human-labeled training data provide predictions for "new" input data. EX1006, ¶¶0140-0142. For example, the Munro reference discloses training a machine learning model based on an "aggregation process" that involves determining how many human annotators classified the stimulus as

belonging to a first label and how many annotators classified the stimulus as belonging to a second label. *Id.*, ¶¶0140-0145. As another example, Cox discloses a model by human annotator responses that provides predictions based on second-order statistical information. EX1007, ¶¶0027, 0032.

62. Thus, the ’046 Patent describes nothing more than well-known machine learning concepts. The grounds of this Petition demonstrate that the Challenged Claims are unpatentable, should not have issued, and should now be canceled.

IV. LEVEL OF ORDINARY SKILL IN THE ART

63. A person having ordinary skill in the art (“POSITA”) relevant to the ’046 Patent as of July 5, 2017 would have had at least: (1) a bachelor’s degree in electrical engineering, computer engineering, computer science, or equivalent course work, with three years of work experience in computer vision, autonomous vehicles, and/or machine learning; or (2) a master’s degree in electrical engineering, computer engineering, computer science, or equivalent course work, with a focus in computer vision, autonomous vehicles, and/or machine learning.

V. CLAIM CONSTRUCTION

64. It is my understanding that claim terms in an IPR are construed according to their “ordinary and customary meaning” to those of skill in the art. *Phillips v. AWH Corp.*, 415 F.3d 9 1303 (Fed. Cir. 2005) (en banc); 37 C.F.R.

§42.100(b). It is my understanding that, for purposes of this proceeding, Petitioner does not present any constructions.

VI. THE ASSERTED GROUNDS OF INVALIDITY

65. My analysis below is based on my understanding of the law and the other information discussed above.

A. Ground 1: Claims 1-19 are obvious over Cox

1. Summary of Prior Art

a. Cox–EX1007

66. Cox is a PCT patent application by Harvard College, which lists Samuel Anthony as an inventor—the same Samuel Anthony listed as an inventor of the ’046 Patent. EX1001, Cover; EX1007, Cover.

67. Cox is directed toward machine learning training and implementation techniques. EX1007, Abstract, ¶¶0001-0004. Crowd-sourcing training data is used to train machine learning algorithms, such as supervised machine learning models. *Id.*, Abstract, ¶¶0001-0004, 0015-0016, 0025-0034, 0044, 0060. Here, human annotators annotate training images and Cox’s model is trained using the annotations. *Id.*, ¶¶0015-19, 0029-0032, Figs. 1A-B, 3. The goal of Cox’s machine learning techniques is to “better mimic human performance.” *Id.*, ¶0004. Cox touts that techniques apply to “driverless” automobile applications. *Id.*, ¶¶ 0056, 0060.

68. Cox is analogous art to the ’046 Patent, from the same field of endeavor as the patent (e.g., predictive model techniques), and is reasonably pertinent to the

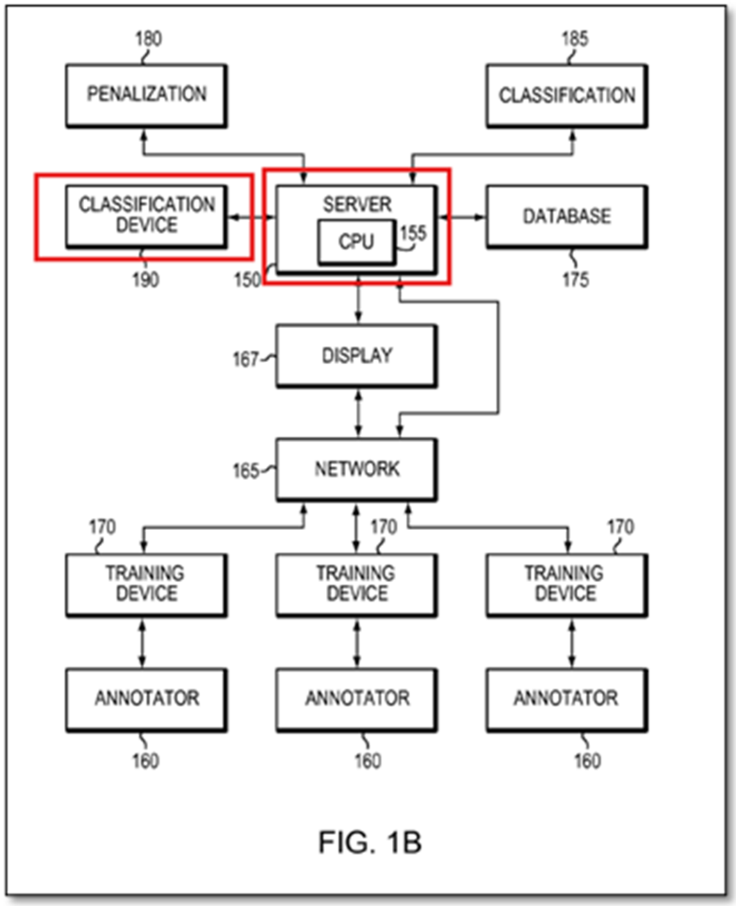
particular problem that the patent was trying to solve (e.g., making more accurate model predictions using human-annotated training data). EX1007, ¶¶0029-0032, Figs. 1A-B, 3; EX1001, 1:27-29, 1:65-2:1, 5:52-59; 6:14-55; Fig. 2A.

2. Claim 1

69. In my opinion, Cox renders obvious claim 1.

a. 1[Pre]

70. In my opinion, to the extent it is limiting, Cox discloses the preamble of claim 1. Cox discloses "classification systems" (or "classifiers") that are implemented on a server 150, using a computer processor 155 and various program modules, where the modules include instructions executed by a computer to implement Cox's techniques (*[a] computer-implemented method*). EX1007, ¶¶0007-0010, 0017-0025, claim 11, Fig. 1B.

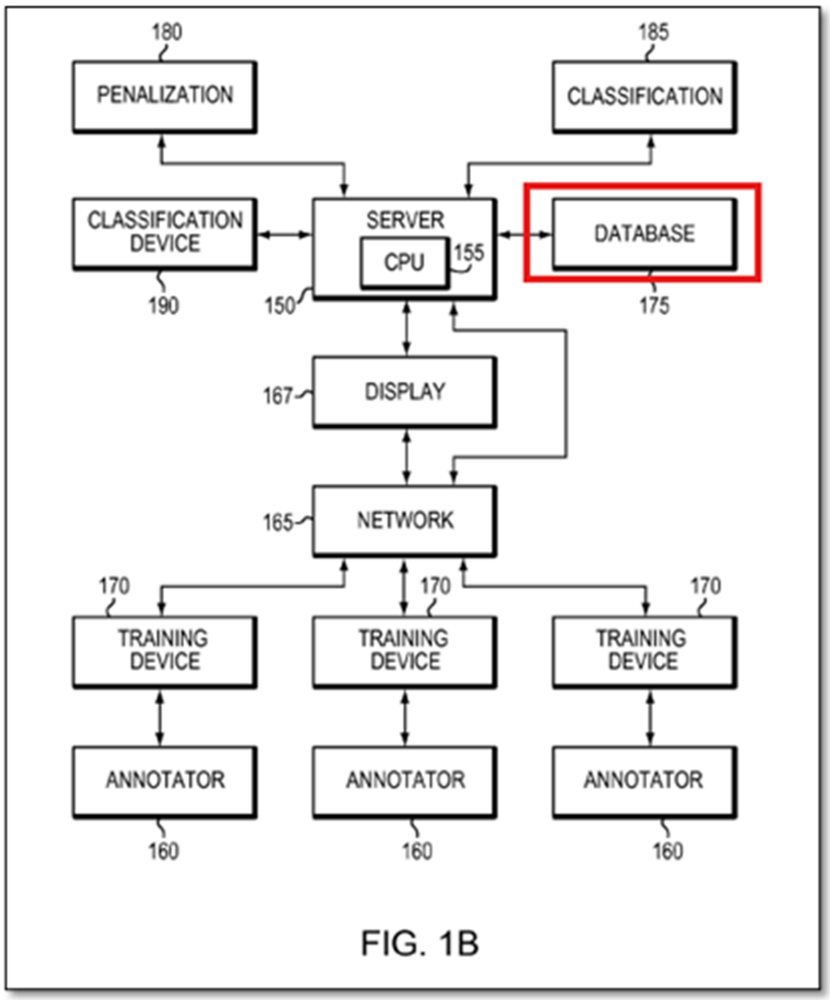


EX1007, Fig. 1B (annotated)

b. 1[a]

71. In my opinion, Cox teaches element 1[a]. Cox describes "a database of training objects" where "[e]ach of the training objects may include or consist essentially of a digital image" including a subject person (*storing a plurality of images, each image displaying one or more users*). EX1007, ¶¶0007-0010, 0017-0018, 0029; *see also* EX1001, 3:60-65 ("human observers[] view sample images of people"). Indeed, the training objects may be images of human faces. EX1007, ¶¶0017, 0029, 0044. A POSITA would have understood or at least found obvious that the people/faces in the images form *one or more users* as claimed. This is at

least because the people/faces reflect people who are partaking in an activity or otherwise using an object, for example as a road user within the context of "driverless" vehicles or a manufacturing facility user within the context of "machine vision for manufacturing," discussed by Cox. EX1007, ¶¶0007-0010, 0017-0018, 0029, 0044, 0056.



EX1007, Fig. 1B (annotated)

c. 1[b]

72. In my opinion, Cox teaches element 1[b]. Cox's purported novelty is "crowd-sourcing" its training data to "dramatically improve the quality, quantity,

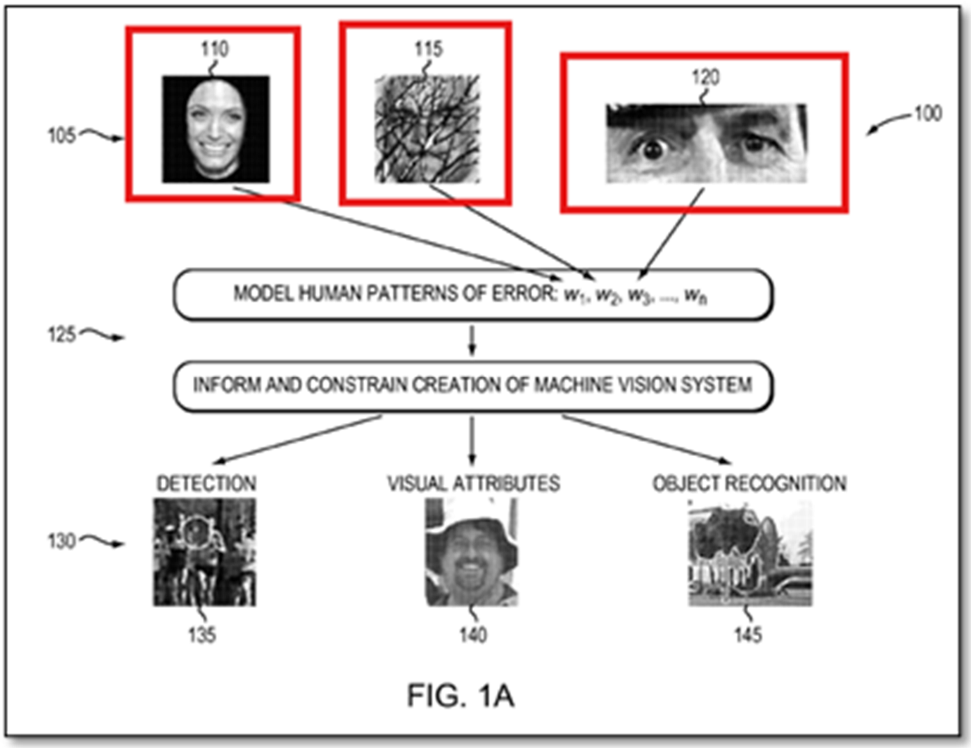
and depth of annotation data available for learning." EX1007, ¶¶0015, 0017, 0025, 0029, 0032, 0034, 0044, 0051. Cox discloses that multiple ("n") training images are presented to ("m") human annotators who in turn annotated the training images. EX1007, ¶0029; *see also* ¶¶0005, 0007-0010, 0017-0018, 0029, 0046 Figs. 3 and 4B. This process in which training images are presented to human annotators for annotation generates training data from the images (*generating training data from the plurality of images, the generating comprising, for each image*). *Id.*

d. 1[b–i]

73. In my opinion, Cox teaches element 1[b-i]. Cox discloses "a crowd-sourced data-acquisition work flow" that involves sending and presenting training images for annotation to multiple human annotators (*[the generating comprising, for each image:] sending the image to a plurality of human observers*). EX1007, ¶¶0017-0018, 0022, 0029. The images are presented to the human annotators on "training devices 170." EX1007, ¶¶0010, 0029.

74. Cox further discloses that when the human annotators are presented with an image, they are asked a question about the image—for example, as to what emotion might correspond to a person shown in the image. EX1007, ¶¶0016 ("**[A] participant may be shown** . . . an image 120 of all or a portion of a person's face **and asked** to select an emotion that corresponds to the image, e.g., jealous, panicked, arrogant, or hateful."), 0018, 0048, Fig. 1A. In my opinion,

when annotators are "asked to select an emotion that corresponds[,]" they are being asked to *answer a question about the state of mind* of the subject. EX1007, ¶¶0016, 0018, 0048, Fig. 1A.



EX1007, Fig. 1A (annotated)

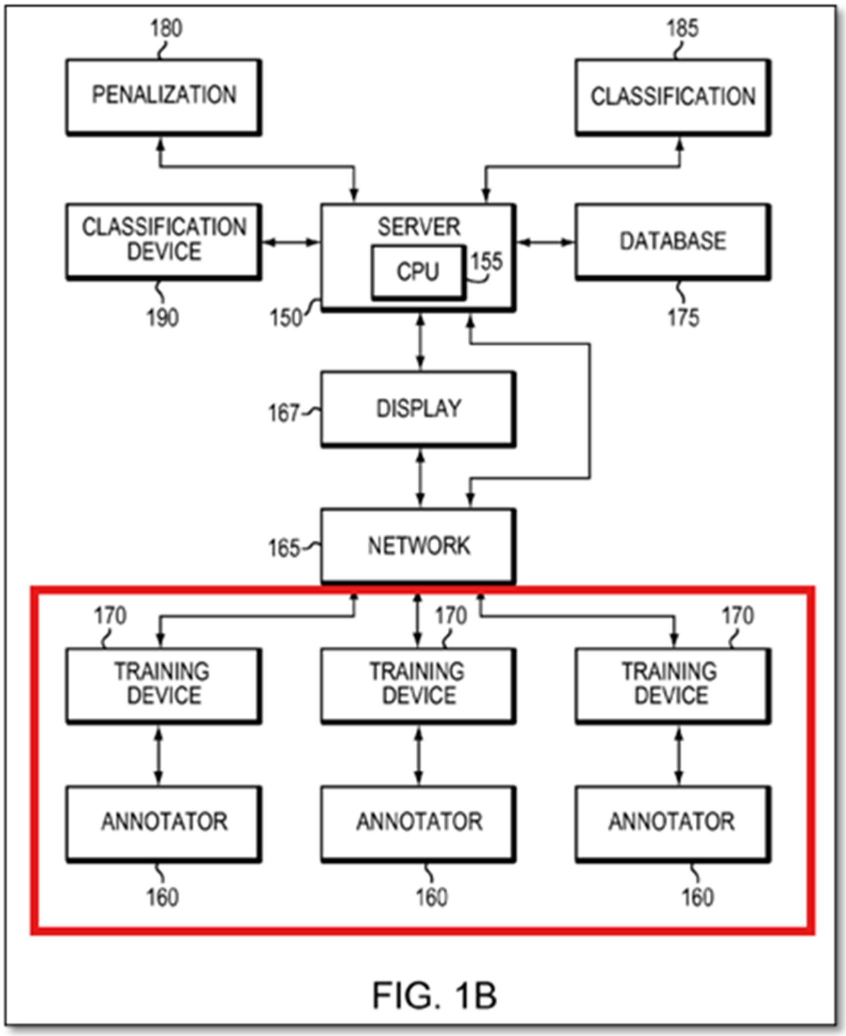
75. Thus, the annotators (i.e., *human observers*) are presented with an image and asked to answer a question about an emotion of the user in the image (*each human observer presented with a request to answer a question about a state of mind of a user in the image*). EX1007, ¶¶0016, 0018, 0048-0049, 0057, Figs. 1A, 1B. In my opinion, this disclosure of emotion in Cox is consistent with the '046 Patent's description of *state of mind*, which is "intention, awareness, **personality, state of consciousness**, level of tiredness, **aggressiveness**, **enthusiasm**,

thoughtfulness or **another characteristic of the internal mental state**" of a user in an image. EX1001, 9:48-54 (emphasis added).

e. 1[b–ii]

76. In my opinion, Cox teaches element 1[b-ii]. As discussed with respect to element 1[b-i], training images are sent to multiple human annotators for annotation. Section VI.A.2.d (element 1[b-i]). The human annotators annotate the images, for example in response to being "asked to select an emotion that corresponds to the image, e.g., jealous, panicked, arrogant, or hateful." EX1007, ¶¶0016-0021, 0049, 0057, Figs. 1A, 1B *see also id.*, ¶¶0007-0010, 0016-0018, 0029. The training devices 170 used by the annotators to input annotations on images "transmit classification data" reflecting the annotations "to the central server 150" and the "classification data [is] received from the human annotators" (*receiving, from each of the plurality of human observers, a response representing a judgment by the human observer of the state of mind of the user in the image*). EX1007, ¶¶0016-0021, 0049, 0057, Figs. 1A, 1B *see also id.*, ¶¶0007-0010, 0016-0018, 0029. In my opinion, the classification data represents *a judgement by the human observer of the state of mind of the user in the image* because the annotations of the classification data are human "judgements," provided by the human annotators, of an emotion of a person in the image. EX1007, ¶¶0002, 0014, 0050, 0061. The

classification data "may be stored in a database 175 of training objects." EX1007, ¶0018, Fig. 1B.

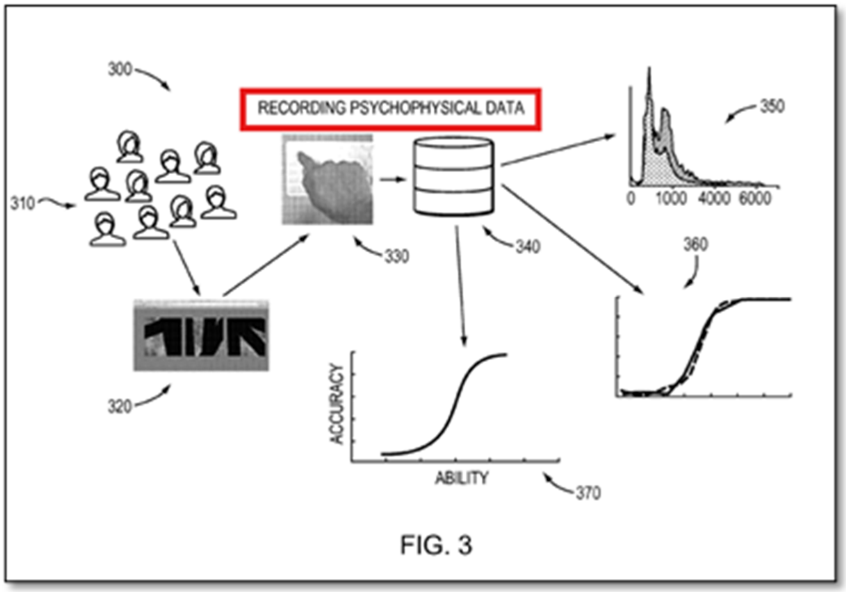


EX1007, Fig. 1B (annotated)

f. 1[b–iii]

77. In my opinion, Cox teaches element 1[b-iii] in multiple ways. As part of Cox's classification techniques, Cox discloses generating "psychometric data" (*generating summary statistics*) describing the emotions of a user in a sample training image (*describing the state of mind of the user in the image*) based on the

annotation responses received from the multiple human annotators (*based on the received responses from the plurality of human observers*). EX1007, ¶¶0002-0008, 0010, 0014, 0018, 0025-0032, 0035, 0049, 0063, Fig. 3; Sections VI.A.2.d-VI.A.2.e (elements 1[b-i], 1[b-ii]). Cox discloses that psychometric data summarizes how human annotators annotate training images because it includes "response times for classifying one or more features, the accuracy of feature classification, and/or the presentation time (i.e., the amount of time presented to each annotator) of one or more training objects." *Id.*, ¶¶0008, 0010, 0018, 0029, Fig. 3. Moreover, in my opinion, because these values are a collection of numerical measurements, a POSITA would have recognized that such values form statistics. *Id.* And finally, Cox's description of "psychometric data" is consistent with the '046 Patent's explanation; just like Cox, the '046 Patent explains that "summary statistics" may "characterize the human observer responses in terms of certain parameters associated with the statistics, such as a content of a response" and "a time associated with entering a response." EX1001, 6:24-28.



EX1007, Fig. 3 (annotated)

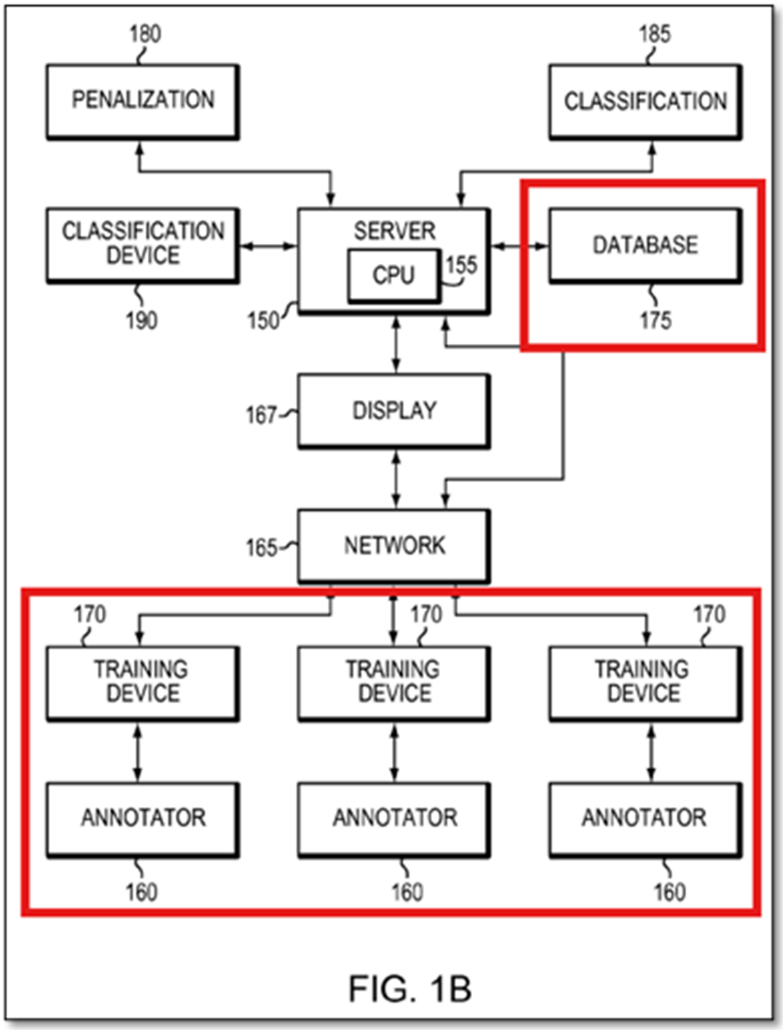
78. Cox further provides additional examples of generating *summary statistics* based on annotator responses. Cox discloses "item-response curves across large populations of humans (e.g., how consistent are judgment[s] across a population)," which Cox describes as a well-known technique associated with human annotator performance. EX1007, ¶0006. Moreover, because these response curves present an analysis of collected numerical measurements, a POSITA would have recognized that the curves form statistics regarding the annotations. *Id.* Cox thus teaches generating an item response curve (*generating summary statistics*) describing the emotions of people in sample training images based on how human annotators labeled the emotions (*describing the state of mind of the user in the image based on the received responses from the plurality of human observers*). *Id.*, ¶¶0002-

0008, 0010, 0014, 0018, 0025-0032, 0035, 0049, 0063, Fig. 3; Sections VI.A.2.d-VI.A.2.e (elements 1[b-i], 1[b-ii]).

79. Cox further discloses a "human-weighted loss function based at least in part on the classification data"—i.e., human annotations—"and the psychometric data." EX1007, ¶0007. Cox also discloses that the responses from human annotators have a distribution of data and labels (e.g., statistics), and that the labels are the responses of "human judgments [that] already provide essential raw material for machine learning, human-generated labels." *Id.*, ¶¶0002-0006, 0014, 0035. Moreover, Cox discloses that "[a]t the end of a series of such queries, end-of-test statistics 820 may be displayed to the annotators 160 via display module 167." *Id.*, ¶0048. In my opinion, a POSITA would have understood from Cox that this data "may be used to inform machine learning in accordance with embodiments of the invention." *Id*. As such, these examples provide further disclosure of how Cox generates summarizing data (*generating summary statistics*) describing the emotions of people in sample training images based on how human annotators labeled the emotions (*describing the state of mind of the user in the image based on the received responses from the plurality of human observers*). EX1007, 0002-0008, 0010, 0014, 0018, 0025-0032, 0035, 0048-0049, 0063, Fig. 3; Sections VI.A.2.d-VI.A.2.e (elements 1[b-i], 1[b-ii]).

g. 1[b–iv]

80. In my opinion, Cox teaches element 1[b-iv]. Cox describes that its system includes a "database of training objects," where the training objects are the images that form training data. EX1007, ¶¶0007-0009, 0017-0020, 0032, Claim 11. Cox also describes that same database as "populated with stored computer records specifying, for each of the plurality of objects, (i) classification data comprising annotations received from a plurality of human annotators" and "(ii) psychometric data characterizing the annotation of the training object by the plurality of human annotators" (*storing the summary statistics in association with the image as part of the training data*). *Id.*; EX1007, ¶0009. Indeed, Cox explains that "[d]uring and/or after the annotation, psychometric data is [] acquired that characterizes the annotation of the training objects by the annotators 160." EX1007, ¶¶0018, 0029, Fig. 3. "The classification and psychometric data may be stored in a database." *Id.*; *see also id.*, ¶0029 ("The psychometric data acquired during step 330, e.g., accuracy of image characterization, response time, presentation time, etc., is recorded in the database 175 in a step 340.").

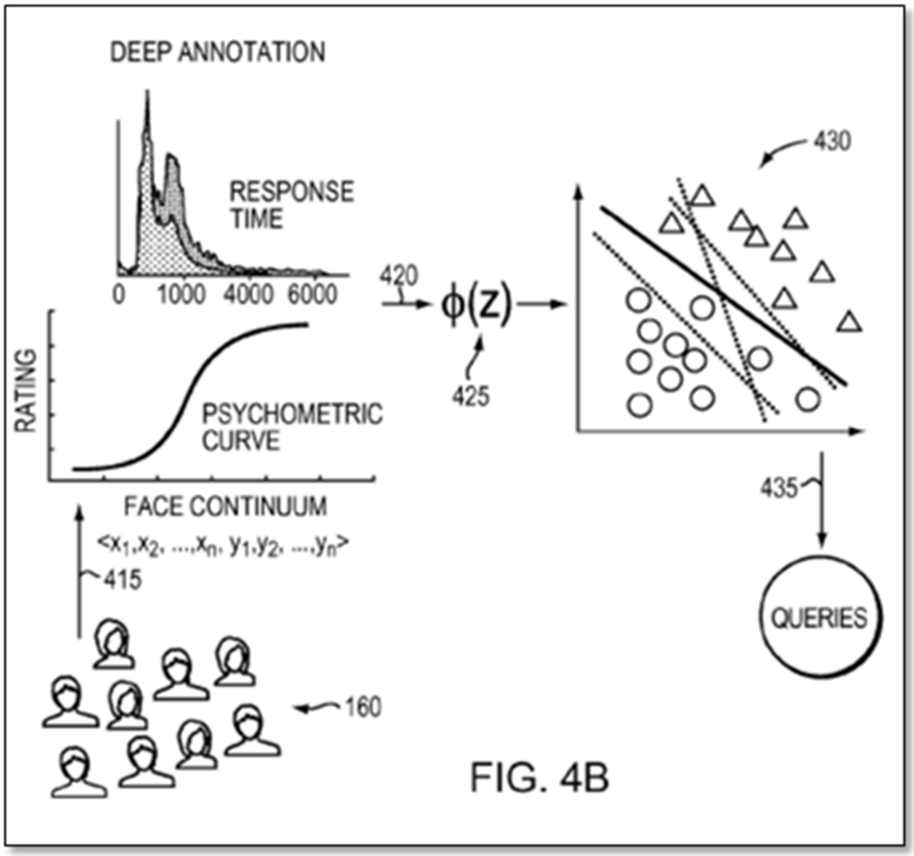


EX1007, Fig. 1B (annotated)

h. 1[c]

81. In my opinion, Cox teaches element 1[c]. Cox discloses that its model is trained by (1) receiving annotated training images showing emotional states, (2) predicting classifications for the training images using the model, and (3) optimizing the model such that differences between the predicted classifications and annotated training images are minimized (*training a model using the training data, the model configured to receive an input image showing a user and predict summary statistics*

describing a state of mind of the user in the input image). EX1007, ¶¶0016-0020, 0029-0037, 0049, 0057, 0060; Fig. 4B (430). Indeed, Cox discloses that "statistical information [] may be collected and applied to the training of computer-vision systems in accordance with embodiments of the invention." EX1007, ¶0027, 0046. The "classification systems for any visual category may be trained with deeply annotated images, by following the learning procedure with human-weighted loss." EX1007, ¶0051; *see also, e.g.,* ¶¶0019, 0029-0032. Cox's disclosure is consistent with the '046 Patent, which states that "trained" means "that the difference between the summary statistics output by the neural network and the summary statistics calculated from the responses of the human observers in step 506 is minimized." EX1001, 12:30-35. This is just like Cox's training approach, as described below and shown in Figure 4B. EX1007, ¶0031, Fig. 4B.



EX1007, Fig. 4B

82. In Cox, annotation responses and psychometric data from the human annotators are utilized to form "a human-weighted loss function" that "includes penalties for misclassification of later presented query objects, as graphically illustrated in graph 430[.]" EX1007, ¶0032. The magnitude of the penalties notably "increases with increasing deviation from the classification data received from the human annotators 160." *Id.*, ¶¶0032, 0034; *see also id.*, ¶¶0008, 0019, 0029-0037, 0060; Fig. 4B (430). These penalties ensure that the trained model is "more consistent with the decisions of the human annotators." *Id.*, ¶0032. Cox further teaches that the "penalties for misclassification may be assigned based at least in

part on the psychometric data." *Id.*, ¶0008. By introducing penalties based on differences from human annotations, Cox optimizes its model to minimize such differences. EX1007, ¶¶0008, 0019, 0029-0037, 0060; Fig. 4B.

83. Accordingly, in my opinion, Cox discloses that its model is trained by predicting classifications of some of the received training objects (images) showing emotional states and optimizing the model such that the difference between the predicted classifications and the training data is minimized (*training a model using the training data, the model configured to receive an input image showing a user and predict summary statistics describing a state of mind of the user in the input image*). EX1007, ¶¶0016-0020, 0029-0037, 0049, 0057, 0060; Fig. 4B (430).

i. 1[d]

84. In my opinion, Cox teaches element 1[d]. Cox discloses training its predictive model based on classifications of training images provided by human annotators and related psychometric data, teaching *the trained model*. *See supra*, Section VI.A.2.h (element 1[c]). Cox describes that the trained model is used to predict the emotional state of a subject in a new, non-training image (*executing the trained model to predict a state of mind of a user in a new image*). EX1007, ¶¶0016, 0020, 0032, Fig. 5.

85. After training, Cox executes its trained model to predict the emotional state of subjects in new images. EX1007, ¶¶0020, 0032. Cox describes that "[o]nce

the human-weighted loss function is determined, one or more 'query objects' may be received by the system for classification." *Id*. The received query objects are "new objects to be classified by the system absent direct human classification." *Id.* As another example, Cox expressly discloses that, in a final step after training the model, "the classification module 185 is utilized to make predictions (based on various query objects) that are more consistent with the decisions of the human annotators 160." *Id*., ¶0032.

86. As explained above with respect to 1[b] and 1[c], Cox explains that its disclosed methods can be applied to any form of supervised learning, including a form of supervised learning that is intended to predict an emotion of a user in an image. EX1007, ¶¶0016, 0032, 0060, Fig. 5. In particular, Cox states a trained "classification module 185 is utilized to make predictions (based on various query objects)," where the query objects can be, for example, "an image 120 of all or a portion of a person's face," and the prediction can be "select[ing] an emotion that corresponds to the image, e.g., jealous, panicked, arrogant, or hateful." EX1007, ¶¶0016, 0032, Fig. 5.

3. Claim 2

87. In my opinion, Cox renders obvious claim 2. Cox describes that training (also referred to as "stimuli") images that are "presented to observers" for labeling can be "degraded by a number of techniques" that involve adjusting pixel values,

including "contrast lowering, brightness lowering, false color, inversion, image scrambling," "blur," and "outline drawings," as well as occlusion such as "thin occluders[] transposed and normalized for a 100% scale" (*the image is manipulated by adjusting values of pixels of the image before presenting to a human observer*). EX1007, ¶¶0046, 0052, 0055, 0064, ¶¶0005-0006, 0014-0016, 0044. In my opinion, Cox's disclosure aligns with the '046 patent that, for example, states "frames can be manipulated by adjusting pixel values" including by blurring, adding occluding bars, bands, and shapes, and removing or changing color information. EX1001, 8:39-54.

4. Claim 3

88. In my opinion, Cox renders obvious claim 3. The emotions in Cox are each a *state of mind*. Section VI.A.2.d (element 1[b-i]). Moreover, a POSITA would have understood that Cox's emotions *indicates whether a user is likely to perform a predetermined action* and a POSITA would have at least understood as much. This is because, for example, a POSITA would have recognized that a "panicked" user would be more likely to flee from perceived danger, and a "hateful" user would be more likely to engage in a fight. *Id*.

5. Claim 4

89. In my opinion, Cox renders obvious claim 4. The emotions in Cox are each a *state of mind*. Section VI.E.2.d (element 1[b-i]). A POSITA would have understood that Cox's emotions *represents a measure of awareness of the*

user regarding an object. For example, a POSITA would have recognized that a user being "panicked" (*state of mind of the user in the image*) would *represent[] a measure of awareness of the user regarding* a source of danger (*an object*), and a user being "jealous" (*state of mind of the user in the image*) would *represent[] a measure of awareness of the user regarding* of a rival (*an object*). *Id*.

6. Claim 5

90. In my opinion, Cox renders obvious claim 5. It would have been obvious to a POSITA that the human annotator response in Cox would include a "*rating on an ordinal scale.*" EX1007, ¶¶0005, 0027-0029. Cox teaches that training considers the relative difficulty of image data. EX1007, ¶¶0005, 0027-0029. A POSITA would have understood from this disclosure that an annotation response would have included information about the relative difficulty associated with an image. *Id.* Because the difficulty information is relative, this information would have been on an ordinal scale (*wherein the response from a human observer comprises a rating on an ordinal scale*). *Id.*

7. Claim 6

91. In my opinion, Cox renders obvious claim 6. Cox's predictive classification model "may be applied to any form of supervised learning, including neural networks, boosting, bagging, random forests, nearest neighbor algorithms, naive bays classifiers, density estimators, and other forms of statistical regression"

and "may also be applied as part of a supervised component of semi-supervised or deep-learning algorithms" (*the model is one of: a random forest regressor, a support vector regressor, a simple neural network, a deep convolutional neural network, a recurrent neural network, or a long short-term memory (LSTM) neural network*). EX1007, ¶0060.

8. Claim 7

92. In my opinion, Cox renders obvious claim 7. Cox discloses psychometric data that forms *summary statistics*. Section VI.A.2.f (element 1[b-iii]). The psychometric data includes annotation response times, annotation accuracy, and the amount of time that a training image was presented to an annotator (*the summary statistics is associated with at least one of a content of a response, a time associated with entering a response*). EX1007, ¶¶0017-0018, 0029. Cox further describes that eye tracking hardware can be used to measure human annotator "saccade-to-target accuracy, saccade-to-target latency, number of saccade hops to target, and total number of saccades," which would have been measured with respect to the display screen displaying the training image that the annotator is viewing *(and a position of an eye of a human observer associated with the response, the position being measured with respect to a display associated with the image*). EX1007, ¶¶0017, 0063. Indeed, the training images of Cox are displayed to annotators on a display of training device 170, which may be computer or cell phone with a display. *Id.*

9. Claim 8

a. 8[pre]

93. *See* Section VI.A.2.a (element 1[Pre]). Cox further explains that its system is implemented on a server that includes a computer processor and "utilizes various program modules" including "computer-executable instructions" that are executed (*[a] non-transitory computer readable storage medium storing instructions that when executed by one or more processors, cause the one or more processors to perform steps*). EX1007, ¶¶0007-0010, 0017-0025, claim 11.

b. 8[a]

94. *See* Section VI.A.2.b (element 1[a]).

c. 8[b]

95. *See* Section VI.A.2.c (element 1[b]).

d. 8[b–i]

96. *See* Section VI.A.2.d (element 1[b–i]).

e. 8[b–ii]

97. *See* Section VI.A.2.e (element 1[b–ii]).

f. 8[b–iii]

98. *See* Section VI.A.2.f (element 1[b–iii]).

g. 8[b–iv]

99. *See* Section VI.A.2.g (element 1[b–iv]).

h. 8[c]

100. *See* Section VI.A.2.h (element 1[c]).

i. 8[d]

101. *See* Section VI.A.2.i (element 1[d]).

10. Claim 9

102. *See* Section VI.A.3 (claim 2).

11. Claim 10

103. *See* Section VI.A.4 (claim 3).

12. Claim 11

104. *See* Section VI.A.5 (claim 4).

13. Claim 12

105. *See* Section VI.A.6 (claim 5).

14. Claim 13

106. *See* Section VI.A.7 (claim 6).

15. Claim 14

107. *See* Section VI.A.8 (claim 7).

16. Claim 15

a. 15[Pre]

108. *See* Sections VI.A.2.a (element 1[Pre]), VI.A.9.a (element 8[Pre]).

b. 15[a]

109. *See* Section VI.A.2.b (element 1[a]).

c. 15[b]

110. *See* Section VI.A.2.c (element 1[b]).

d. 15[b–i]

111. *See* Section VI.A.2.d (element 1[b–i]).

e. 15[b–ii]

112. *See* Section VI.A.2.e (element 1[b–ii]).

f. 15[b–iii]

113. *See* Section VI.A.2.f (element 1[b–iii]).

g. 15[b–iv]

114. *See* Section VI.A.2.g (element 1[b–iv]).

h. 15[c]

115. *See* Section VI.A.2.h (element 1[c]).

i. 15[d]

116. *See* Section VI.A.2.i (element 1[d]).

17. Claim 16

117. *See* Section VI.A.8 (claim 7).

18. Claim 17

118. *See* Section VI.A.4 (claim 3).

19. Claim 18

119. *See* Section VI.A.5 (claim 4).

20. Claim 19

120. *See* Section VI.A.6 (claim 5).

B. Ground 2: Claims 1-19 are obvious over Cox and Ross

121. In my opinion, Cox alone renders obvious the Challenged Claims. However, Cox is further combined with Ross to address the following limitations: (1) *sending the image to a plurality of human observers, each human observer presented with a request to answer a question about a state of mind of a user in the image*; (2) *receiving, from each of the plurality of human observers, a response representing a judgment by the human observer of the state of mind of the user in the image*; (3) *training a model using the training data, the model configured to receive an input image showing a user and predicting summary statistics describing a state of mind of the user in the input image*; and (4) *executing the trained model the predict a state of mind of a user in the new image*. As discussed below, these limitations are addressed by Ross, and it would have been obvious to combine Ross's relevant teachings with Cox.

1. Summary of Prior Art

a. Ross–EX1005

122. Ross discloses techniques for "maneuvering" an autonomous vehicle based on prediction model outputs. EX1005, 1:30-31, 3:39-42, 3:64-5:5. Ross's system generates human influenced training data and uses the data to train an "intent model" for "predicting [the] intent of an object" such as a vehicle, pedestrian, or bicyclist, in the environment of an autonomous vehicle. EX1005, 1:30-2:53, 4:33-5:11, 9:13-10:10, 15:1-16:52. "Human operators" manually generate a "list of

predetermined actions" based on their "observations of the actions of other road users" from sensor data generated by a vehicle 100, and a human observed action of the list that is accurately predicted is "marked as the correct intent" during training. *Id.*, 9:41-10:10.

123. The trained intent model is used by the autonomous vehicle in operation. Cameras of the vehicle capture "raw" camera data of the environment around the vehicle displaying objects in the environment such as other vehicles, pedestrians, and bicyclists. EX1005, 7:57-8:10, 9:13-10:10, 15:1-16:52. The trained model is executed on the camera data to predict a "set of possible intents or hypotheses identifying possible next actions (and predicted trajectories), a given point in time when the action is likely to occur, and associated likelihood values" with respect to each object. *Id.*, 3:64-4:46, 7:57-8:10, 9:13-10:10, 15:1-16:52 (parenthetical in original). The set of likelihood values, with each value associated with a different prediction, "indicate[s] which of the predictions are more likely to occur (relative to one another)." *Id.* The autonomous vehicle of Ross is controlled based on the predictions. *Id.*

124. Ross is analogous art to the '046 Patent, from the same field of endeavor as the patent (e.g., predictive model techniques), and is reasonably pertinent to the particular problem that the patent was trying to solve (e.g., making

more accurate model predictions using human annotated training data). EX1005, Abstract, 3:64-5:1; 9:27-47; EX1001, 1:27-29, 1:65-2:1, 5:52-59; 6:14-55.

2. Motivation to Combine Cox and Ross

125. A POSITA would have been motivated to combine Cox and Ross such that:

- Cox's system transmits a training image to multiple human annotators for annotation with a prompt to answer about an action by a person or road user in the image reflecting a possible intent of the person, as taught by Ross; and
- Cox's system is trained to analyze an action by a person reflecting a possible intent in images, as taught by Ross.

EX1007, 0010, 0016-0018, 0022, 0028-0032, 0048-0049, 0057, Figs. 1A, 1B; EX1005., 1:55-2:13, 4:33-46, 9:13-10:10, 15:1-16:52.

126. In my opinion, a POSITA would have been motivated to combine Cox and Ross in this manner for several reasons.

127. In my opinion, a POSITA would have been motivated to combine Cox and Ross because Ross teaches a known application (driving) for Cox's system. Cox discloses a machine learning prediction system that relies on human annotators to annotate training images regarding emotional states of a person in the images. EX1007, ¶¶0007-0010, 0015-0019, 0025-0035, 0051, 0060. Thus, Cox already provides that human annotators answer prompts with respect to subjects in images

to generate training data, and that a prediction model is trained using the training data. *Id*. Ross simply provides a well-known application for such a system by teaching that the subjects in the images are road users (e.g., vehicles and their associated drivers, pedestrians, bicyclists) performing an action on the road. EX1007, ¶¶0010, 0016-0018, 0022, 0028-0032, 0048-0049, 0057, Figs. 1A, 1B; EX1005., 1:55-2:13, 4:33-46, 9:13-10:10, 15:1-16:52. Ross is only teaching another well-known application of Cox's system—to the behavior and intent context—which would have been well within the capabilities of a POSITA and routine, and the combination would have had a reasonable expectation of success. *Id.* In my opinion, the combination with Ross thus provides an advantageous application for the system in Cox.

128. In my opinion, a POSITA would have further been motivated to combine Cox and Ross due to the similarities and compatibilities of the references because they both represent improvements in the same field of endeavor, e.g., training models to make human-like predictions. For example, Cox identifies its system as relevant to numerous applications, including "driverless" vehicles and "computer-vision systems." EX1007, ¶¶0056, 0027. Cox's model is utilized after training "to make predictions (based on various query objects) that are more consistent with the decisions of the human annotators 160." *Id.*, ¶¶0027-0032 (parenthetical in original), 0051. The goal of Cox's machine learning techniques is

to "better mimic human performance." EX1007, ¶0004. Similar to Cox, the stated goal in Ross is to provide an autonomous vehicle that "function[s] in a more 'human-like' or 'polite' way." EX1005, 5:7-12. Ross similarly has applications to autonomous vehicles, disclosing systems and methods for "behavior and intent estimations of road users for autonomous vehicles." EX1005, Title, Abstract, 1:30-54. To determine the behavior and intent, Ross discloses generating training data and using the training data to train an intent model. EX1005, 3:39-51; 3:64-5:1, 7:57-8:10, 9:13-10:10, 15:1-16:52. The intent model is trained using training data that includes human observations of "road user" actions. *Id.*, 1:55-2:13, 4:33-46, 9:13-10:10, 15:1-16:52. Such actions reflect the "possible intent[s]" associated with an object displayed in the sensor data. *Id*. As demonstrated above, it is my opinion that both Cox and Ross explicitly teach that their disclosures aim to provide more human-like model predictions and performance, which would have led a POSITA to combine their teachings.

129. In my opinion, a POSITA would have been further motivated to combine Cox and Ross at least because the references each provide explicit teachings, suggestions, or motivations for making the combination. Cox itself expressly discloses that its teaching may be "directly applied to several important domains where machine learning is found" and expressly suggests applying the invention to "driverless" vehicle applications. EX1007, ¶0056. And Ross further

explains that its disclosure provides that an autonomous vehicle "function[s] in a more 'human-like' or 'polite' way" (EX1005, 5:7-12), which is just like what Cox aims to achieve as noted above, which is to be "more consistent with the decisions of the human annotators" who annotated the training data used to train the Cox model and to "better mimic human performance." EX1007, ¶¶0004, 0007-0010, 0015-0019, 0025-0035, 0051, 0060. Accordingly, Cox and Ross each provide their own teachings, suggestions, or motivations to make the combination, and a POSITA would have thus combined the references. *Id.*

130. All the reasons explained above establish that a POSITA would have been motivated to combine Cox and Ross. Ross provides driving-specific implementation details to achieve one of the specific applications envisioned in Cox: "driverless […] automobiles." EX1007, ¶0056. This combination would have been the predictable combination of well-known prior art elements (e.g., Cox's training images for annotation and prediction model; Ross's disclosure of observed road user actions demonstrating intents and prediction model predicting road user intents) according to known methods to yield predictable results (Cox's system transmits a training image to multiple human annotators for annotation with a prompt to answer about an action by a person in the image reflecting a possible intent of the person, as taught by Ross; Cox's system trained to analyze an action by a person reflecting a possible intent in images, as taught by Ross). The combination would have further

used known techniques (Ross's known techniques regarding human observations on road user actions) to improve similar devices (Cox's prediction system) in the same way (using Ross's aforementioned techniques). *Id*. Lastly, in addition to the above, because Ross uses human-labeled data to train its model, combining Ross with Cox's teachings about generating and using the same data would have required minimal changes. Making this combination would also be well within the skill of a POSITA, and a POSITA would have had a reasonable expectation of success. *Id*.

3. Claim 1

131. In my opinion, Cox and Ross render obvious claim 1.

a. 1[Pre], 1[a], 1[b]

132. In my opinion, Cox and Ross teach elements 1[Pre], 1[a] and 1[b] for the same reasons discussed in Sections VI.A.2.a (element 1[Pre]), VI.A.2.b (element 1[a]), and VI.A.2.c (element 1[b]). Moreover, with respect to element 1[a], the Cox-Ross combination teaches displaying road users such as vehicles (including their associated drivers), pedestrians, and bicyclists in images, teaching each image displaying one or more users. EX1005, 5:6-11, 9:13-10:10, 15:1-16:52; Section VI.A.2.b (element 1[a]).

b. 1[b–i]

133. Cox and Ross teach element 1[b-i]. Cox discloses transmitting a training image to multiple human annotators for annotation with a prompt to answer about the image (*[the generating comprising, for each image:] sending the image to*

a plurality of human observers, each human observer presented with a request to answer a question). *See* Section VI.A.2.d (element 1[b-i]). Cox's prompt relates to answering a question about emotions of a person in a sample image. *Id.* To the extent the Patent Owner argues, or the Board finds, that such a prompt about the emotion of a person does not relate to a *state of mind* as claimed, the combination of Cox and Ross teaches this concept.

134. Ross discloses that its "intent model" is trained to predict a "set of possible intents or hypotheses identifying possible next actions (and predicted trajectories), a given point in time when the action is likely to occur, and associated likelihood values" using human observation responses to training sensor data (e.g., camera images) indicating possible next actions of an object displayed in the sensor data, such as another road user (e.g., a vehicle, pedestrian, bicyclist). EX1005, 1:55-2:13, 4:33-46, 9:13-10:10, 15:1-16:52. Such actions reflect the "possible intent[s]" associated with the other road users displayed in the sensor data. *Id.* As such, combining Cox with Ross would have provided that Cox's training image would have included an object exhibiting an action reflecting a possible intent, and Cox's human annotator responses would have identified such possible actions associated with a possible intent. *Id.*, EX1007, ¶¶0010, 0016-0018, 0022, 0028-0032, 0048-0049, 0057, Figs. 1A, 1B. As taught by Ross, the possible intents include, for example, for an object vehicle or object bicycle, the intent of a vehicle driver or

bicyclist riding the bicycle to turn the vehicle/bicycle, change lanes, drive through an intersection, or cross and move into the path of a roadway in which the autonomous vehicle is traveling. EX1005, 2:4-13, 9:13-10:10, 15:1-16:52

135. Thus, as applied to Cox, Ross teaches that Cox's system transmits a training image to multiple human annotators for annotation with a prompt to answer about an action by a person or road user in the image reflecting a possible intent of the person (*[the generating comprising, for each image:] sending the image to a plurality of human observers, each human observer presented with a request to answer a question about a state of mind of a user in the image*). EX1007, ¶¶0010, 0016-0018, 0022, 0028-0032, 0048-0049, 0057, Figs. 1A, 1B; EX1005, 1:55-2:13, 4:33-46, 9:13-10:10, 15:1-16:52; Section VI.A.2.d (element 1[b-i]).

c. 1[b–ii] – 1[b–iv]

136. In my opinion, as discussed above, Cox and Ross teach deploying the Ross-Cox model for that collects training data from a plurality of human annotators by presenting them an image with a user and requesting a response to answer a question about a state of mind of the user in the image. *See supra* Sections VI.B.3.b (element 1[b-i]), VI.A.2.d.

137. In my opinion, Cox and Ross teach elements 1[b-ii] – 1[b-iv] for at least the same reasons discussed in Section VI.A.2.e (element 1[b-ii]), VI.A.2.f (element 1[b-iii]), and VI.A.2.g (element 1[b-iv]).

d. 1[c]

138. In my opinion, Cox and Ross teach element 1[c]. As combined, the training of Cox's model would have been performed as discussed in Section VI.A.2.h (element 1[c] – Ground 1) except the training data includes training images and annotations with respect to actions by a person in images reflecting possible intents of the person, and Cox's model is trained on such training data as discussed in Section VI.A.2.h (element 1[c] – Ground 1). *See* Sections VI.B.2, VI.B.3.b, VI.B.3.c, VI.A.2.h.

e. 1[d]

139. In my opinion, Cox and Ross teach element 1[d]. As combined, Cox's model would have operated on new images to determine the actions and possible intents of people in images. *See* Sections VI.B.2, VI.A.2.i (element 1[d] – Ground 1).

4. Claim 3

140. In my opinion, the Cox-Ross combination renders obvious claim 3. Cox and Ross teach an action by a person in the image reflecting a possible intent of the person (*state of mind of the user in the image*). Section VI.B.3.b (element 1[b-i]).

141. Ross describes that an intent prediction model is configured to predict a "set of possible intents or hypotheses identifying possible next actions (and predicted trajectories), a given point in time when the action is likely to occur, and associated likelihood values" from an input camera image of an object, such as a

vehicle or pedestrian. EX1005, 1:55-2:13, 4:33-46, 9:13-40, 15:1-21. The possible next actions reflect actions (e.g., crossing a road, turning on a road) found in a "predetermined list of actions" that are "generated manually, for instance, by human operators based on personal experience or observations of the actions of other road users (by such operators or from sensor data from perception systems of one or more autonomous vehicles such as vehicles 100 or 100A)." *Id.*, 4:47-65, 9:41-10:10 (parenthetical in original). Thus, in my opinion, as applied to Cox, Ross teaches that the action by a person in the image reflecting a possible intent of the person (*state of mind of the user in the image*) would have indicated a likelihood that the person performs a predetermined action (e.g., crossing a road) from a list of predetermined actions (*wherein the state of mind of the user in the image indicates whether the user is likely to perform a predetermined action*), as taught by Ross. *Id.*; *supra* Section VI.B.3.b (element 1[b-i]), VI.B.3.c (element 1[b-ii]). A POSITA would have been motivated to combine Cox and Ross for the reasons discussed in Section VI.B.2.

5. Claim 4

142. In my opinion, the Cox-Ross combination renders obvious claim 4. Cox and Ross teach an action by a person in the image reflecting a possible intent of the person (*state of mind of the user in the image*). Section VI.B.3.b (element 1[b-i]).

143. Ross describes an intent prediction model as discussed in Section VI.B.3 (claim 1). Ross discloses that a predetermined list of actions "may be

generated manually, for instance by human operators based on personal experience or observations of the actions of other road users." EX1005, 9:41-47. For example, predetermined actions regarding a pedestrian in an image may include an action of a "pedestrian waiting to cross the road" which can correspond to an intent of the pedestrian "to cross the road when clear," i.e., when there is no traffic and it is safe for the pedestrian to cross. EX1005, 15:50-16:22. As applied to Cox, Ross further teaches that the predicted action of a person in the image, such a prediction that a pedestrian intends "to cross the road when clear," indicates that the pedestrian is aware of oncoming traffic and will wait to cross until it is safe, i.e., when the traffic has passed. *Id*. Thus, the Cox-Ross combination discloses predicting a measure of awareness of a pedestrian (*state of mind of the user in the image represents a measure of awareness of the user regarding an object*). *Id.* A POSITA would have been motivated to combine Cox and Ross for the reasons discussed in Section VI.B.2.

6. Claim 8

144. *See* Section VI.B.3 (claim 1); *see also* Section VI.A.9.a (element 8[Pre], Ground 1).

7. Claim 10

145. *See* Section VI.B.4 (claim 3).

8. Claim 11

146. *See* Section VI.B.5 (claim 4).

9. Claim 15

147. *See* Section VI.B.3 (claim 1); *see also* Section VI.A.9.a (element 8[Pre]).

10. Claim 17

148. *See* Section VI.B.4 (claim 3).

11. Claim 18

149. *See* Section VI.B.5 (claim 4).

12. Claims 2, 5, 6, 7, 9, 12, 13, 14, 16, and 19

150. In my opinion, Cox and Ross render obvious claims 2, 5, 6, 7, 9, 12, 13, 14, 16, and 19 for the same reasons discussed with respect to Cox in Section VI.A above.

C. Ground 3: Claims 5, 12, and 19 are obvious over Cox and Ellenbogen

151. In my opinion, Cox alone renders obvious all limitations of the Challenged Claims, however, Cox is further combined with Ellenbogen to address the following limitation in Ground 3: *the response from a human observer comprises a rating on an ordinal scale*. As discussed below, this limitation is explicitly taught by Ellenbogen, and it would have been obvious to combine Ellenbogen's relevant teachings with Cox.

1. Summary of Prior Art

a. Ellenbogen–EX1004

152. Ellenbogen is generally directed to "improving machine decision making" and techniques for training and using machine learning models. EX1004, Abstract, ¶¶0053, 0057, 0133-0137, 0168, Fig. 34. Images are transmitted to multiple "human agents" and the agents are queried to "answer[] a question regarding a characteristic of the image." EX1004, ¶¶0005, 0011, 0055-0058, 0062-0074, 0133-0137, 0141, 0149, 0161, 0171, 0180, Fig. 34. The questions can be related to, for example, behavior or activity exhibited by a subject person in a received image. *Id.*, ¶¶0011, 0057, 0141, 0161, claim 1, claim 13; *see id.*, Fig. 9.

153. The agents provide their responses to Ellenbogen's system, and the responses are aggregated into a "composite output." EX1004, ¶¶0135, 0149, 0190, Claim 3, Figs. 9, 34. The composite output, as well as the image that was annotated by the human agents itself, are then used to train Ellenbogen's predictive model. *Id.*, ¶¶0053, 0133-0139, 0162. Ellenbogen's model can then be executed for a "new image," that is not part of training data, in an "operational phase" *Id.*, ¶¶0057, 0133, 0148, 0168.

154. In my opinion, Ellenbogen is analogous art to the '046 Patent, from the same field of endeavor as the patent (e.g., predictive model techniques), and reasonably pertinent to the particular problem that the patent was trying to solve (e.g., making more accurate model predictions using human-annotated training

data). EX1004, Abstract, ¶¶0005, 0042-0048, 0132-0146; EX1001, 1:27-29, 1:65-2:1, 5:52-59; 6:14-55.

2. Motivation to Combine Cox and Ellenbogen

155. In my opinion, a POSITA would have been motivated to combine Cox and Ellenbogen. As discussed above in Section VI.A.2 (claim 1), Cox already discloses presenting training images to human annotators for annotations, prompting the annotators to provide annotations, and receiving annotation responses from these annotators. EX1007, ¶0016-0021, 0029, 0049, 0057, Figs. 1A, 1B; Section VI.A.2 (claim 1). Ellenbogen simply adds to Cox that the annotators are queried to "answer[] a question regarding a characteristic of the image," and responses are received from the agents that include a "confidence measure" of the answer. EX1004, ¶¶0005, 0007-0012, 0055-0058, 0062-0074, 0133-0137, 0141, 0149, 0161, 0171, 0180, Fig. 34. Thus, Ellenbogen teaches implementation details regarding the format of annotation responses of Cox's system, and the combination thus would have been a routine enhancement to Cox's system that would have been easily performed by POSITAs. *Id*. Moreover, a POSITA would have a reasonable expectation of success because the results of the combination would have been predictable: human annotator responses including a "confidence measure" rating. *Id.*

3. Claim 5

156. In my opinion, Cox and Ellenbogen render obvious claim 5. Cox describes receiving human annotators responses in relation to training images. *See* Sections VI.A.2.d-VI.A.2.f. (elements 1[b-i]-1[b-iii]).

157. In my opinion, Ellenbogen also describes receiving human annotators' responses in relation to training images. EX1004, ¶¶0005, 0011, 0055-0058, 0062-0073, 0133-0137, 0141, 0161, 0180, Fig. 34. The response from an annotator includes a "confidence measure" of the result, provided by the human agent, that is a rating on a scale of 0 to 1.0 and can be a value such as "0.9" or "0.5," for example, which are ordinal responses on an ordinal scale. EX1004, ¶¶0007, 0012, 0135, 0149, 0190, 0191, Fig. 34. As applied to Cox, Ellenbogen thus teaches that Cox's responses would have included using an ordinal value (e.g., 0.9 or 0.5) on an ordinal scale (e.g., 0 to 1.0) (*wherein the response from a human observer comprises a rating on an ordinal scale*). *See* Section VI.A.2.d-VI.A.2.f; EX1004, ¶¶0005, 0007, 0011-0012, 0055-0058, 0062-0073, 0133-0137, 0141, 0149, 0161, 0180, 0190-0191, Fig. 34. Ellenbogen's teachings align with the '046 patent, which provides an example ordinal scale as values such as "1" or "4." EX1001, 10:34-53, Fig. 6.

4. Claim 12

158. *See* Section VI.C.3 (claim 5).

5. Claim 19

159. *See* Section VI.C.3 (claim 5).

D. Ground 4: Claims 1-19 are obvious over Cox and Munro

160. In my opinion, Cox alone renders obvious all limitations of the Challenged Claims, however, Cox is further combined with Munro regarding the following limitation in Ground 4: *generating summary statistics describing the state of mind of the user in the image based on the received responses from the plurality of human observers*. As discussed below, these limitations are explicitly taught by Munro, and it would have been obvious to combine Munro's relevant teachings with Cox.

1. Summary of Prior art

a. Munro–EX1006

161. Munro is generally directed to techniques for collecting human annotator responses to a stimulus such as a sample document, where the responses are used to train a machine learning model. EX1006, Abstract, ¶¶0005, 0042-0048, 0132-0142. Munro explains that its system determines summary statistics characterizing annotator responses. *Id.*, ¶¶0140-0142. For example, the system performs an "aggregation process" that involves determining how many human annotators classified the stimulus as belonging to a first label and how many annotators classified the stimulus as belonging to a second label. *Id.* Munro's machine learning model is then trained based on the aggregated responses. EX1006, ¶¶0140-0145.

162. In my opinion, Munro is analogous art to the '046 Patent, from the same field of endeavor as the patent (e.g., predictive model techniques), and reasonably pertinent to the particular problem that the patent was trying to solve (e.g., making more accurate model predictions using human-annotated training data). EX1006, Abstract, ¶¶0005, 0042-0048, 0132-0146; EX1001, 1:27-29, 1:65-2:1, 5:52-59; 6:14-55.

2. Motivation to Combine Cox and Munro

163. In my opinion, a POSITA would have been motivated to combine Cox and Munro such that Cox's generated statistical data from human annotator responses would have included tallies of the number of annotators who classified the images in a certain way. EX1006, Abstract, ¶¶0005, 0042-0048, 0132-0142; EX1007, ¶¶0002-0006, 0014, 0035, 0048. A POSITA would have been motivated to combine Cox and Munro in this manner for several reasons.

164. In my opinion, a POSITA would have been motivated to combine Cox and Munro because Munro simply provides an example implementation of Cox's statistical data generated from annotator responses. Cox discloses a system where model training images are transmitted to multiple human agents, the agents are queried to answer a prompt with respect to an image regarding an emotional state of a subject in the image, and responses from the agents with answers are received by the system. EX1007, ¶¶0002-0006, 0014, 0028-0035, 0048; *see* Section VI.A.2.d-

VI.A.2.f (elements 1[b-i]-1[b-iii]). Munro is similarly directed to techniques for collecting human annotator responses to a training sample, where the responses are used to train a machine learning model. EX1006, Abstract, ¶¶0005, 0042-0048, 0132-0142. Munro's system performs an "aggregation process" for the responses that involves tallying the number of annotators who classified a training sample as having various labels. *Id.*, ¶¶0140-0142.

165. Thus, in my opinion, Munro merely provides an example implementation of Cox's statistical data generated from annotator responses, teaching that it would have included tallies of the number of annotators who classified a behavior or activity in an image as having various labels in their responses. EX1006, Abstract, ¶¶0005, 0042-0048, 0132-0142; EX1007, ¶¶0002-0006, 0014, 0035, 0048; *see* Section VI.A.2.f (element 1[b-iii]). A POSITA would have readily enhanced Cox with these teachings of Munro due to the straightforward and routine nature of the combination teachings. *Id.* In my opinion, including such tallies of human annotator responses as taught by Munro would further improve Cox's ability to optimize a model that "produce[s] more 'human-like' solutions" (EX1007, ¶0005), by recognizing and accounting for trends among the annotator responses.

166. Both Cox and Munro are directed to machine learning technologies and involve gathering human annotator responses, providing that a POSITA seeking to

implement Cox's teachings would have looked to Munro. Munro's teachings simply provide how Cox's annotator responses would have been organized—tallying them by label—which would have been nothing more than a well-known method of organization having a reasonable expectation of success. Moreover, at least due to the straightforward application of Munro's teachings to Cox, the results of the combination would have been predictable. EX1006, Abstract, ¶¶0005, 0042-0048, 0132-0142; EX1007, ¶¶0002-0006, 0014, 0035, 0048. Further, in my opinion, this combination would have at least been the predictable combination of well-known prior art elements (e.g., Cox's annotator responses; Munro's organizing of annotator responses) according to known methods to yield predictable results (Cox's annotator responses organized as taught by Munro).

167. It is my opinion that, due to the straightforward application of Munro's teachings to Cox, the results of the combination would have been predictable. EX1006, Abstract, ¶¶0005, 0042-0048, 0132-0142; EX1007, ¶¶0002-0006, 0014, 0035, 0048. Because Cox already analyzes and uses collections of human annotator responses, combining Cox with Munro's teaching to include tallies of such responses would require minimal modifications to Cox. Making this combination is thus well within the skill of a POSITA, and a POSITA would have had a reasonable expectation of success. *Id.*

3. Claim 1

a. 1[Pre]-1[b–ii]

168. In my opinion, Cox and Munro teach elements 1[Pre] through 1[b–ii] for the same reasons discussed in Sections VI.A.2.a through VI.A.2.e (elements 1[Pre] – 1[b–ii]).

b. 1[b–iii]

169. In my opinion, Cox and Munro teach element 1[b-iii]. The combination of Cox and Munro teaches that Cox's generated statistical data from annotator responses of human annotators regarding the emotional state of a user in a training image would have included tallies to aggregate the number of annotators who similarly classified the emotional state of the user in the training image (*[the generating comprising, for each image:] summary statistics describing the state of mind of the user in the image based on the received responses from the plurality of human observers*). EX1006, Abstract, ¶¶0005, 0042-0048, 0132-0142; EX1007, ¶¶0002-0006, 0014, 0035, 0048; *see* Sections VI.A.2.f (element 1[b-iii]), VI.D.2. A POSITA would have recognized that these tallies are *summary statistics* because they are a collection of numerical measurements as to how annotators classified emotional states. *Id*. This combined teaching of Cox and Munro further aligns with the '046 Patent's disclosure. Specifically, the '046 Patent explains, for example, that the generation of "summary statistics" that "characterize the aggregate responses of multiple human observers to a particular derived stimulus" (i.e., training image)

includes "how many observers" classified the stimulus as displaying various content. EX1001, 6:1-11.

170. In my opinion, a POSITA would have been motivated to combine Cox and Munro for the reasons discussed in Section VI.D.2.

c. 1[b–iv]

171. In my opinion, Cox and Munro teach element 1[b-iv] for at least the same reasons discussed in Section VI.A.2.g (element 1[b-iv]), except the generated summary statistics as taught by Munro would have been stored in Cox's computer record in a database for the statistic's associated image. Sections VI.A.2.g (element 1[b-iv]), VI.D.3.b (element 1[b-iii]).

d. 1[c]

172. In my opinion, Cox and Munro teach element 1[c]. As combined, the training of Cox's model would have been performed as discussed in Section VI.A.2.h (element 1[c]) except the training data includes the generated summary statistics as taught by Munro, and Cox's model is trained on such training data in the same manner discussed in Section VI.A.2.h (element 1[c]). *See* Sections VI.A.2.h (element 1[c]), VI.D.3.b (element 1[b-iii]), VI.D.3.c.

e. 1[d]

173. In my opinion, Cox and Munro teach element 1[d] for the same reasons discussed in Section VI.A.2.i (element 1[d]).

4. Claim 8

174. *See* Section VI.D.3 (claim 1); *see also* Section VI.A.9.a (element 8[Pre], Ground 1).

5. Claim 15

175. *See* Section VI.D.3 (claim 1); *see also* Section VI.A.9.a (element 8[Pre], Ground 1).

6. Claim 2-7, 9-14, and 16-19

176. In my opinion, Cox and Munro render obvious claims 2-7, 9-14, and 16-19 for the same reasons discussed with respect to Cox in Section VI.A above.

E. Ground 5: Claims 1-19 are obvious over Ellenbogen and Munro

177. In my opinion, Ellenbogen teaches most limitations of the Challenged Claims and is further combined with Munro to address the following limitations: (1) *generating summary statistics describing the state of mind of the user in the image based on the received responses from the plurality of human observers*; (2) *storing the summary statistics in association with images as part of the training data*; (3) *training a model using the training data, the model configured to receive an input image showing a user and predict summary statistics describing a state of mind of the user in the input image*. As discussed below, these limitations are explicitly taught by Munro, and it would have been obvious to combine Munro's relevant teachings with Ellenbogen.

1. Motivation to Combine Ellenbogen and Munro

178. In my opinion, a POSITA would have been motivated to combine Ellenbogen, Munro, and Cox such that:

- Ellenbogen's composite output of human agent responses regarding a behavior or activity of a person in a sample image includes tallies of the number of agents who classified the behavior or activity as having various labels in their responses, as taught by Munro (the "Ellenbogen-Munro composite output") (EX1004, ¶¶0135, 0149, 0190; *id.*, ¶¶0005, 0008, 0011, 0054-0058, 0060-0073, 0108, 0133-0137, 0141, 0149, 0161, 0180, Claim 3, Figs. 9, 34; EX1006, ¶¶0140-0142);
- The Ellenbogen-Munro composite output is stored in a computer record as training data, as taught by Munro, and in association with underlying images using the computed score taught by Moreno (EX1004, ¶¶0133-0137, Fig. 34; EX1006, ¶¶0059-0062, 0141-0142); and
- Ellenbogen's predictive model is trained by the Ellenbogen-Munro composite output and image that was annotated, as taught by Munro (EX1006, ¶¶0140-0142; EX1004, ¶¶0057, 0133-0137, 0138-0148, 0166-0168, Fig. 34).

179. In my opinion, a POSITA would have been motivated to combine Ellenbogen and Munro in this manner for several reasons. This combination would have at least been the predictable combination of well-known prior art elements

(e.g., Ellenbogen's annotator responses; Munro's organizing of annotator responses, training, and storage) according to known methods to yield predictable results (Ellenbogen's annotator responses being organized, stored, and used to train a mode, as taught by Munro).

180. In my opinion, a POSITA would have been motivated to combine Ellenbogen and Munro because Munro simply provides implementation details for Ellenbogen's system. Ellenbogen discloses a system where model training images are transmitted to multiple human agents, the agents are queried to "answer[] a question regarding a characteristic of the image," and responses from the agents with answers are received by the system. EX1004, ¶¶0005, 0011, 0055-0058, 0062-0074, 0133-0137, 0141, 0149, 0161, 0171, 0180, Fig. 34. Ellenbogen describes that a "composite output" is generated from the responses. *Id.*, ¶¶0135, 0149, 0190.

181. Like Ellenbogen, Munro is generally directed to techniques for collecting human annotator responses to a training sample, where the responses are used to train a machine learning model. EX1006, Abstract, ¶¶0005, 0042-0048, 0132-0142. As a disclosure for one such technique, Munro's system performs an "aggregation process" for the responses that involve tallying the number of annotators who classified a training sample as having various labels. *Id.*, ¶¶0140-0142. In my opinion, Munro provides an example in which its system determines how many human annotators classified a training sample as having a first label and

how many annotators classified the sample as having a second label. *Id*. And Munro provides routine disclosure in how model training is implemented and training data is stored, referring to storing based on computed stores. EX1006, ¶¶0059-0062, 0140-0142.

182. As such, in my opinion, Munro simply provides implementation details specifying how Ellenbogen's composite output including annotator responses is formed, stored, and used for training, teaching that it would have included tallies of the number of annotators who classified a behavior or activity in an image as having various labels in their responses (the "Ellenbogen-Munro composite output"). EX1006, Abstract, ¶¶0005, 0042-0048, 0132-0142; EX1004, ¶¶0135, 0149, 0190; *id.*, ¶¶0005, 0008, 0011, 0054-0058, 0060-0073, 0108, 0133-0137, 0141, 0149, 0161, 0180, Claim 3, Figs. 9, 34. A POSITA would have readily enhanced Ellenbogen with Munro's teachings due to the routine and straightforward nature of the combined teachings. *Id.* Both Ellenbogen and Munro are directed to machine learning technologies and involve gathering human annotator responses and training a model, providing that a POSITA seeking to implement Ellenbogen's teachings would have looked to Munro.

183. In my opinion, a POSITA would have further been motivated to combine Ellenbogen and Munro because Munro would provide performance benefits to Ellenbogen. For example, Munro's tallying would provide additional data

regarding annotator responses that could be used to better train Ellenbogen's prediction model. EX1004, ¶¶0005, 0008, 0011, 0054-0058, 0060-0073, 0108, 0133-0137, 0141, 0149, 0161, 0180, 0190, Claim 3, Figs. 9, 34; EX1006, ¶¶0140-0142. By training the model with such additional training data, the model would have in turn undergone additional training, which would have improved model accuracy. *Id*. Further, Munro's storage teachings would improved processing data in Ellenbogen because each Ellenbogen-Munro composite output would be stored in association with its underlying image for more efficient access and use. EX1004, ¶¶0057, 0133-0148, 0166-0168, Fig. 34; EX1006, ¶¶0059-0062, 0140-0142. And, training Ellenbogen's predictive model using the implementation details taught by Munro would have ensured that the model iteratively improves in accuracy. *Id*.

184. In my opinion, a POSITA would also have a reasonable expectation of success in combining Munro with Ellenbogen because Munro's teachings simply provide details about how Ellenbogen's annotator responses would be organized and used. Due to the straightforward application of Munro's teachings to Ellenbogen, the results of the combination would have been predictable. EX1004, ¶¶0135, 0149, 0190; *id*., ¶¶0005, 0008, 0011, 0054-0058, 0060-0073, 0108, 0133-0137, 0141, 0149, 0161, 0180, Claim 3, Figs. 9, 34; EX1006, Abstract, ¶¶0005, 0042-0048, 0059-0062, 0132-0142. And implementing these teachings from Munro in Ellenbogen would have been nothing more than a routine and well-known method

of organization, training, and storage, which means that a POSITA would have had a reasonable expectation of success in making the combination. *Id*.

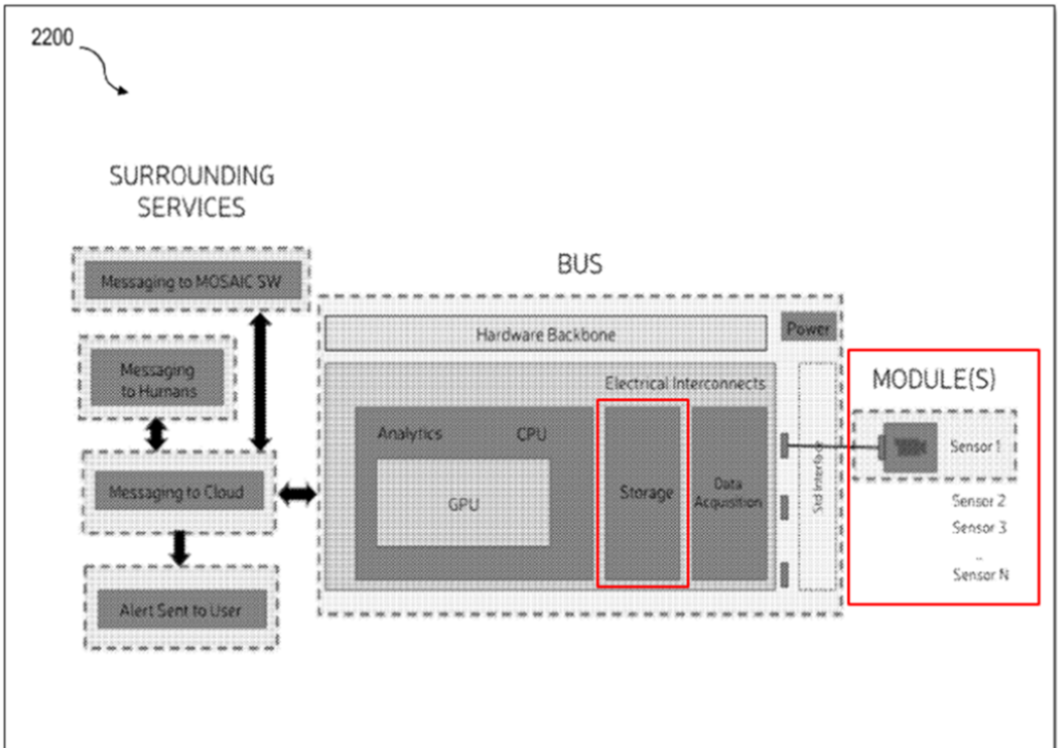
2. Claim 1

a. 1[Pre]

185. In my opinion, to the extent it is limiting, Ellenbogen discloses the preamble. Ellenbogen describes a computerized process for "training a machine computation component" including a machine learning model, and techniques for executing the trained machine learning model (*[a] computer-implemented method*). EX1004, ¶¶0057, 0133-0137, 0168, Fig. 34.

b. 1[a]

186. In my opinion, Ellenbogen teaches element 1[a]. Ellenbogen discloses (1) that "sensor data" such as "a series of images" includes a "person" in the images partaking in an activity, exhibiting "suspicious behavior," or expressing a "sentiment," (2) that the sensor data is produced by an "imaging device, video camera, [or] still camera," and (3) that the sensor data is stored in "storage" such as a database (*storing a plurality of images, each image displaying one or more users*). EX1004, ¶¶0010-0011, 0054, 0057, 0133-0137, 0141, 0161, 0226, Figs. 22, 34.



EX1004, Fig. 22 (annotated)

187. Moreover, in my opinion, a "database" of images to be labeled by human annotators is provided by Ellenbogen's system, and the images include people, which further teaches *storing a plurality of images, each image displaying one or more users*. EX1004, ¶¶0054, 0094, 0137, 0141, 0161-0164.

c. 1[b]

188. In my opinion, Ellenbogen teaches element [1b]. Ellenbogen describes that the images (*see* Section VI.E.2.b (element 1[a])) are labeled by "human agents" to generate training data for a predictive model (*generating training data from the*

plurality of images, the generating comprising, for each image). EX1004, ¶¶0133-0137, Fig. 34.

d. 1[b–i]

189. In my opinion, Ellenbogen teaches element 1[b-i]. An image of the series of images (*see* Section VI.E.2.b (element 1[a])) is transmitted to client devices of multiple human agents (*[the generating comprising, for each image:] sending the image to a plurality of human observers*). EX1004, ¶¶0005, 0011, 0055-0058, 0062-0074, 0133-0137, 0141, 0149, 0161, 0171, 0180, Fig. 34, claim 1, claim 13. Each of the human agents is queried to "answer[] a question regarding a characteristic of the image" such as a behavior of a person in the image or an activity associated with a person in the image (*each human observer presented with a request to answer a question about a state of mind of a user in the image*). *Id*. When Ellenbogen's system is used to determine the activity or behavior of a person in an image, the "question regarding a characteristic of the image" relates to the *state of mind* of the person in the image because it asks the human agent to answer a question about what the person is doing. *Id.*; EX1004, Fig. 9. For example, the determined behavior or activity of the person in the image would have included an intention of the person with respect to what they are doing. *Id.* The determined behavior or activity of the person in the image would have further included behavior, such as "suspicious" behavior, and/or a "vehicle behavior" like "tailgating" or driving in the "wrong

direction." *Id*. Indeed, annotated Figure 9 below provides an example showing that to determine "if people exhibit certain behaviors," then "human questions" are asked regarding "[s]uspicious behavior," "[b]ehavior of people on bus," "[v]ehicle behavior (tailgating, wrong direction)," "[a]ctivity recognition (e.g., person using phone)," and "interactions with retail products/end cap." *Id.*, Fig. 9.

Modality	Description	Requirements	Examples
Intrusion	Here is a physical space, region, or boundary. I'm concerned about intrusion into that space.	• Tripwire specification • Person/vehicle/object description • Person/vehicle/object particulars • Human questions	• Perimeter security • Asset protection • Door monitoring • Animals intruding • People where they shouldn't be
Access	I'm trying to identify people or classes of people that should either be allowed in or prevented from entering. It's a space, region, or boundary where it is normal for people to go in and out, but only certain people should be allowed access.	• Tripwire specification • Person/vehicle/object description • Person/vehicle/object particulars • Watchlist • Human questions	• Watch list • VIP list • Particular delivery vehicle/driver
Loiter/Dwell	People, vehicles, or objects can be there, but if they loiter/dwell that is something I want to know.	• Area of interest • Person/vehicle/object description • Person/vehicle/object particulars • Time thresholds • Human questions	• Person dwell • Vehicle dwell • Object dwell (bag left behind)
Behavior	I want to know if people exhibit certain behavior.	• Behavior of interest • Region of interest • Human questions	• Suspicious behavior • Behavior of people on bus • Vehicle behavior (tailgating, wrong direction) • Activity recognition (e.g., person used phone) • Interaction with retail products/end cap
Tracking	Once I've identified someone or some vehicle, I want to be able to track it from one camera to the next.	• Region of interest • Last location, time, and trajectory • Person/vehicle/object description • Person/vehicle/object particulars • Human questions	• Person tracking • Vehicle tracking

FIG. 9

EX1004, Fig. 9 (annotated)

190. In my opinion, Ellenbogen's disclosure corresponds to the '046 patent's description of a "state of mind;" for example, the patent explains that the "evaluation of the state of mind of a road user depicted" in an image "can be of the intention," "state of consciousness," "aggressiveness," "enthusiasm," or "thoughtfulness," and

at least each of these are represented by a "behavior" described by Ellenbogen. EX1001, 9:48-54, 5:65-6:2. For example, a person partaking in an activity reflects an intention of that person to perform the activity, suspicious behavior at least indicates a "state of consciousness," and a vehicle that is tailgating at least indicates an "aggressiveness" of the vehicle's driver, and each of these examples, as well as those also discussed above in for this element 1[b-i], at least teaches a *state of mind*. EX1004, Fig. 9.

e. 1[b–ii]

191. In my opinion, Ellenbogen teaches element 1[b-ii]. The multiple human agents each respond with a "query result" to the "question regarding a characteristic of the image" that describe a behavior or activity of a person in the image (*[the generating comprising, for each image:] receiving, from each of the plurality of human observers, a response representing a judgment by the human observer of the state of mind of the user in the image*). EX1004, ¶¶0005, 0008, 0011, 0054-0058, 0060-0073, 0108, 0133-0137, 0141, 0149, 0161, 0180, Claim 3, Figs. 9, 34. Ellenbogen describes that these responses by human agents reflect "judgments" that the agents are making with respect to the image, and the judgments thus reflect the behavior or activity. *Id.*, ¶¶0057, 0060, 0180, Fig. 9; *see also* Section VI.E.2.d (element 1[b–i]).

f. 1[b–iii]

192. In my opinion, Ellenbogen and Munro teach element [1b-iii]. Ellenbogen describes that a "composite output" is generated from the query result responses regarding the behavior or activity of a person in the image are received from the multiple human agents. EX1004, ¶¶0135, 0149, 0190; *id.*, ¶¶0005, 0008, 0011, 0054-0058, 0060-0073, 0108, 0133-0137, 0141, 0149, 0161, 0180, Claim 3, Figs. 9, 34; Section VI.E.2.e (element 1[b-ii]).

193. Munro is generally directed to techniques for collecting human annotator responses to a training sample, where the responses are used to train a machine learning model. EX1006, Abstract, ¶¶0005, 0042-0048, 0132-0142. Munro explains that its system performs an "aggregation process" for the responses that involves tallying the number of annotators who classified a training sample as having various labels. *Id.*, ¶¶0140-0142. In my opinion, Munro provides an example in which its system determines how many human annotators classified a training sample as having a first label and how many annotators classified the sample as having a second label. *Id.*

194. Thus, in my opinion, Munro teaches that Ellenbogen's composite output of query result responses regarding the behavior or activity of a person in a sample image would have included an aggregate of the number of annotators who classified the behavior or activity as having various labels in their responses (the "Ellenbogen-Munro composite output") (*[the generating comprising, for each*

image:] summary statistics describing the state of mind of the user in the image based on the received responses from the plurality of human observers). EX1006, Abstract, ¶¶0005, 0042-0048, 0132-0142; EX1004, ¶¶0135, 0149, 0190; *id.*, ¶¶0005, 0008, 0011, 0054-0058, 0060-0073, 0108, 0133-0137, 0141, 0149, 0161, 0180, Claim 3, Figs. 9, 34. A POSITA would have recognized that these tallies are *summary statistics* as claims because they are a collection of numerical measurements summarizing how annotators classified behavior or activity. *Id*. This combined Ellenbogen-Munro teaching further aligns with the '046 Patent's disclosure. The '046 Patent explains, for example, that the generation of "summary statistics" that "characterize the aggregate responses of multiple human observers to a particular derived stimulus" (i.e., training image) includes "how many observers" classified the stimulus as displaying various content. EX1001, 6:1-11.

195. In my opinion, a POSITA would have been motivated to combine Ellenbogen and Munro for the reasons discussed in Section VI.E.1.

g. 1[b–iv]

196. In my opinion, Ellenbogen and Munro teach element 1[b-iv]. Ellenbogen and Munro teach the Ellenbogen-Munro composite output (*summary statistics*). Section VI.E.2.f (element 1[b-iii]). Ellenbogen describes that the "sensor data" reflecting the image having the person in it that was annotated, as well as the "result from the agent computation component"—which is the composite output—

are used to train a "predictive model" and are thus *training data* for the model. EX1004, ¶0136; *id.*, ¶¶0133-0137, Fig. 34. Thus, Ellenbogen and Munro teach that the Ellenbogen-Munro composite output (*summary statistics*) and associated image that was annotated are (*the image*) are training data for Ellenbogen's model (*part of the training data*). EX1004, ¶¶0133-0137, Fig. 34; Section VI.E.2.f (element 1[b-iii]).

197. In my opinion, Munro teaches that to store the Ellenbogen-Munro composite output in association with the image as part of the training data. Specifically, Munro describes that its system includes a "client data store" (i.e., a database of training data) that contains "a repository of data that is used to train and ultimately generate the natural language model." EX1006, ¶0061. This teaching in Munro is also consistent with the disclosure in Ellenbogen that a "database of training images […] can be updated over time with real world images and labels." EX1004, ¶0164. Thus, in the combination with Ellenbogen, a POSITA would have understood that storing the Ellenbogen-Munro composite in the client data store would comprise "*storing the summary statistics […] as part of the training data.*"

198. Further, Munro teaches that a "computed score" can be determined based on the annotation aggregation process, and may be used to determine whether to use a given training reference during model training. EX1006, ¶0141. In the combination with Ellenbogen, this "computed score" thus associates the

Ellenbogen-Munro composite with a training reference (i.e., the *image*), such that the composite is "*stor[ed] […] in association with the image as part of the training data*."

199. To the extent this limitation is interpreted to require that the *summary statistics* and *image* are stored together, in my opinion a POSITA would have found it obvious to store the composite together with the training reference in Munro's client data store. EX1006, ¶¶0061, 0141; EX1004, ¶¶0133-0137, 0164. This is because storing such information together would have improved processing efficiency and speed, and thus been a routine and well-known design choice. *Id.* A POSITA would have had a reasonable expectation of success for this arrangement because storing these together would have been simpler and involved fewer parts than storing the composite and images separately.

h. 1[c]

200. In my opinion, Ellenbogen and Munro teach element 1[c]. Ellenbogen describes that the "sensor data" reflecting the image having the person in it that was annotated, as well as the "result from the agent computation component," that is Ellenbogen's composite output, is used to train Ellenbogen's predictive model. EX1004, ¶0136; *id.*, ¶¶0133-0137, Fig. 34. Ellenbogen describes that the goal of its system is to "progressively train the artificial intelligence to begin to recognize" certain "behavior" in an "environment on its own." *Id.*, ¶¶0057, 0133.

201. Munro teaches that its tally of annotator responses is also used to train a predictive model via a "model training process." EX1006, ¶¶0140-0142. As such, the combination of Ellenbogen and Munro teaches that the Ellenbogen-Munro composite output (*see* Sections VI.E.2.f, VI.E.2.g (elements 1[b-iii], 1[b-iv])) and the associated image that was annotated, as well as other instances of such items, are used as training data to train Ellenbogen's predictive model (*training a model using the training data*). *Id.*; EX1004, ¶¶0057, 0133-0137, Fig. 34; EX1006, ¶¶0140-0142.

202. In my opinion, Ellenbogen and Munro further teach *the model configured to receive an input image showing a user and predict summary statistics describing a state of mind of the user in the input image*. Specifically, Ellenbogen teaches that the artificial intelligence model in Ellenbogen can be trained "to begin to recognize suspicious behavior in that environment on its own." EX1004, ¶0057. Ellenbogen discloses that "[a]s the artificial intelligence component is trained on more real-world data, the artificial intelligence component becomes more accurate and less agent input is required." EX1004, ¶0061. Ellenbogen further teaches to predict a "confidence" and a score related to the confidence, e.g., *summary statistics*. *Id.*, ¶¶0060-61, ¶0166 ("The object classifier can generate a score from 0.0 to 1.0 related to the confidence that the object is not present (0.0) or present (1.0). The classification can be binary or multiclass.").

203. Ellenbogen further teaches that its prediction can be a "multiclass" prediction in which a "score from 0.0 to 1.0 related to the confidence" for a predicted classification is used. EX1004, ¶0166. Multiclass predictions analyze whether a prediction falls within one of multiple classes. *Id.*, ¶¶0166-0168. Thus, in my opinion, a POSITA would have understood or at least found obvious that a multiclass prediction utilizing confidence scores from 0.0 to 1.0 would have provided that Ellenbogen's model generates multiple potential classes of behavior or activity and confidence scores for each of classes, summarizing the various possible classifications and likelihoods for behavior or activity in the image (*summary statistics describing a state of mind of the user in the input image*). EX1004, ¶¶0133-0137, 0138-0148, 0166-0168, Fig. 34. Moreover, in my opinion, the predicted classes of behavior or activity and associated confidence scores are *predicted summary statistics* that correspond to the tallied annotator responses of the Ellenbogen-Munro composite output that are the stored *summary statistics* (*see* Section VI.E.2.g, VI.E.2.h, *supra*) at least because they are the predicted categories (and associated likelihoods) of behavior or activity that would have also been represented by human annotators from their labels during training in the Ellenbogen-Munro composite output.

204. Thus, in combination with Munro, Ellenbogen's model is trained by receiving input images of a person exhibiting a certain behavior or activity (*the*

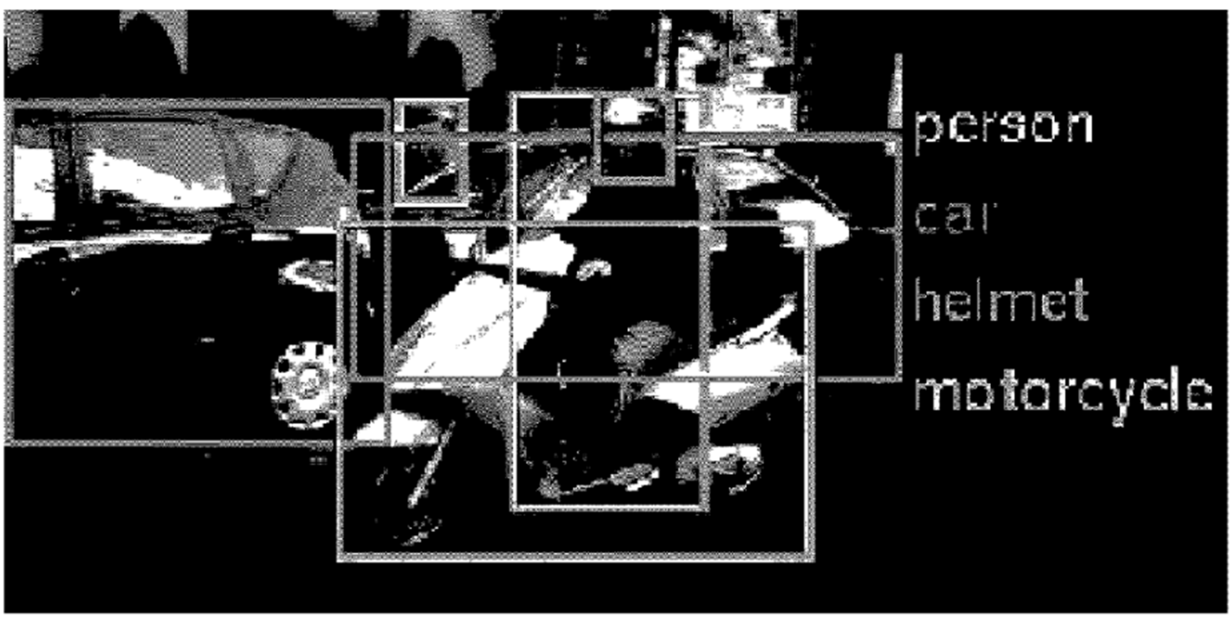
model configured to receive an input image showing a user) and predicting a multiclass behavior or activity prediction of a person in a training image and associated confidence scores (*and predict summary statistics describing a state of mind of the user in the input image*), and then optimizing the model, for example, such that a difference between the multiclass predicted behavior or activity of a person and the Ellenbogen-Munro composite output reflecting tallied annotation responses is minimized. EX1004, ¶¶0133-0137, 0138-0148, 0166-0168, Fig. 34; EX1006, ¶¶0140-0142; Section VI.E.2.f (element 1[b-iii]). The predicted multiclass behavior or activity of a person in a training image and associated confidence scores form *summary statistics describing a state of mind of the user in the input image*.

i. 1[d]

205. In my opinion, the Ellenbogen and Munro combination teaches element [1d]. Ellenbogen's predictive model is trained (*the trained model*). *See* Section VI.E.2.h (element 1[c]). The trained predictive model is then executed in regard to a new image that is not part of training data in an "operational phase" to determine whether a person of the image is exhibiting certain behavior or partaking in an activity, such as whether a driver of a vehicle is driving aggressively by tailgating (*executing the trained model to predict a state of mind of a user in a new image*). EX1004, ¶¶0057, 0133, 0148, 0168.

3. Claim 2

206. In my opinion, Ellenbogen and Munro render obvious claim 2. Ellenbogen teaches that images can be manipulated (e.g., with the addition of bounding boxes) before it is sent to human agents (*wherein the image is manipulated by adjusting values of pixels of the image before presenting to a human observer*). *Id.*; EX1004, ¶0170, *see also id.*, ¶¶0162-063, 166-68. As illustrated in Figure 13 (below), the addition of bounding boxes constitutes alterations to pixels in the digital images. *Id.*



EX1004, Fig. 13

4. Claim 3

207. In my opinion, Ellenbogen and Munro render obvious claim 3. To the extent this element is limiting, it would be obvious to a POSITA that the detection of "suspicious behavior" and "non-suspicious" behavior in Ellenbogen (EX1007, ¶¶0011, 0057, 0161) includes determining a "*state of mind*" that "*indicates whether*

a user is likely to perform a predetermined action." For example, "suspicious behavior" suggests that a user may be likely to commit a crime.

5. Claim 4

208. In my opinion, Ellenbogen and Munro render obvious claim 4. To the extent this element is limiting, it would be obvious to a POSITA that the detection of "suspicious behavior" and "non-suspicious" behavior in Ellenbogen (EX1007, ¶¶0011, 0057, 0161) includes determining a "*state of mind*" that "*represents a measure of awareness of the user regarding an object*." For example, a user appearing lost (identified as "non-suspicious" in Ellenbogen) would represent a lack of awareness of the user's surroundings.

6. Claim 5

209. In my opinion, Ellenbogen and Munro render obvious claim 5. Ellenbogen's human agent response to a query (*the response from a human observer*) includes a query result as well as a "confidence measure" of the result that is a rating from 0 to 1.0 and can be a value such as "0.9" or "0.5," for example, which are ordinal confidence rating values on an ordinal scale (*wherein the response from a human observer comprises a rating on an ordinal scale*). EX1004, ¶¶0007, 0012, 0135 ("confidence measure may be directly supplied by an agent"), 0149, 0190, 0191, Fig. 34. Ellenbogen's teaching aligns with the '046 patent, which provides an example ordinal scale as values such as "1" or "4." EX1001, 10:34-53, Fig. 6.

7. Claim 6

210. In my opinion, Ellenbogen and Munro render obvious claim 6. Ellenbogen's trained predictive model can be a "deep neural network, a convolutional neural network (CNN), a Faster Region-based CNN (R-CNN), and the like," as well as a "deep learning neural network" (*wherein the model is one of: a random forest regressor, a support vector regressor, a simple neural network, a deep convolutional neural network, a recurrent neural network, or a long short-term memory (LSTM) neural network*). EX1004, ¶¶0053, 0136, 0162; EX1007, ¶¶0027-0032, 0060.

8. Claim 7

211. In my opinion, Ellenbogen and Munro render obvious claim 7. The Ellenbogen-Munro composite output is part of *the summary statistics* based on aggregation of feedback from human agents. Section VI.E.2.f (element 1[b-iii]). Ellenbogen further teaches utilizing latency in the response time from human agents in the assessment of agent confidence. EX1004, ¶¶0100-0101; Figure 27. Thus, a POSITA would have understood that the Ellenbogen-Munro composite (*the summary statistics*) are "*associated with at least one of a content of a response, a time associated with entering a response, and a position of an eye of a human observer associated with the response, the position being measured with respect to a display associated with the image*."

9. Claim 8

a. 8[Pre]

212. In my opinion, Ellenbogen and Munro teach element 8[pre] for the reasons explained in Section VI.E.2.a (element 1[Pre]). Moreover, Ellenbogen's system can be implemented by a computer system including non-transitory computer readable storage medium/computer memory storing instructions that are executed by one or more processors to provide Ellenbogen's disclosed functionality (*A non-transitory computer readable storage medium storing instructions that when executed by one or more processors, cause the one or more processors to perform steps*). EX1004, ¶0015.

b. 8[a]

213. *See* Section VI.E.2.b (element 1[a]).

c. 8[b]

214. *See* Section VI.E.2.c (element 1[b]).

d. 8[b-i]

215. *See* Section VI.E.2.d (element 1[b-i]).

e. 8[b-ii]

216. *See* Section VI.E.2.e (element 1[b-ii]).

f. 8[b-iii]

217. *See* Section VI.E.2.f (element 1[b-iii]).

g. 8[b-iv]

218. *See* Section VI.E.2.g (element 1[b-iv]).

h. 8[c]

219. *See* Section VI.E.2.h (element 1[c]).

i. 8[d]

220. *See* Section VI.E.2.i(element 1[d]).

10. Claim 9

221. *See* Section VI.E.3 (claim 2).

11. Claim 10

222. *See* Section VI.E.4 (claim 3).

12. Claim 11

223. *See* Section VI.E.5 (claim 4).

13. Claim 12

224. *See* Section VI.E.6 (claim 5).

14. Claim 13

225. *See* Section VI.E.7 (claim 6).

15. Claim 14

226. *See* Section VI.E.8 (claim 7).

16. Claim 15

227. *See* Sections VI.E.2 (claim 1), VI.E.9.a (element 8[Pre]).

a. 15[Pre]

228. *See* Sections VI.E.2.a (element 1[Pre]), VI.E.9.a (element 8[Pre]).

b. 15[a]

229. *See* Section VI.E.2.b (element 1[a]).

c. 15[b]

230. *See* Section VI.E.2.c (element 1[b]).

d. 15[b-i]

231. *See* Section VI.E.2.d (element 1[b-i]).

e. 15[b-ii]

232. *See* Section VI.E.2.e (element 1[b-ii]).

f. 15[b-iii]

233. *See* Section VI.E.2.f (element 1[b-iii]).

g. 15[b-iv]

234. *See* Section VI.E.2.g (element 1[b-iv]).

h. 15[c]

235. *See* Section VI.E.2.h (element 1[c]).

i. 15[d]

236. *See* Section VI.E.2.i (element 1[d]).

17. Claim 16

237. *See* Section VI.E.8 (claim 7).

18. Claim 17

238. *See* Section VI.E.4 (claim 3).

19. Claim 18

239. *See* Section VI.E.5 (claim 4).

20. Claim 19

240. *See* Section VI.E.6 (claim 5).

F. Printed Matter

241. I understand that Claim limitations directed to the content of information and lacking a requisite functional relationship are not entitled to patentable weight because such information is not patent eligible subject matter

242. Because they recite the content of information without a functional relationship, the following limitations are printed matter:

- 1[a], 8[a], 15[a] – "image displaying one or more users"
- 1[b-i], 8[b-i], 15[b-i] – "request to answer a question about a state of mind of a user in the image"
- 1[b-ii], 8[b-ii], 15[b-ii] – "response representing a judgment by the human observer of the state of mind of the user in the image"
- 1[b-iii], 8[b-iii], 15[b-iii] – "summary statistics describing the state of mind of the user in the image based on the received responses from the plurality of human observers"
- 1[c], 8[c], 15[c] – "input image showing a user" and "summary statistics describing a state of mind of the user in the input image"
- 1[d], 8[d]. 15[d] – "a state of mind of a user in a new image"

- 3, 10, 17 – "the state of mind of the user in the image indicates whether the user is likely to perform a predetermined action"
- 4, 11, 18 – "the state of mind of the user in the image represents a measure of awareness of the user regarding an object"
- 5, 12, 19 – "the response from a human observer comprises a rating on an ordinal scale"

Each of the above limitations recites a data element and then defines that data element based on its content.

243. I understand that printed matter is only entitled to patentable weight if it is functionally related to the substrate on which the information is present. In my opinion, the above printed matter limitations do not have any such functional relationship with the "substrate," which in this case is computer memory. *See* 8[pre] and 15[pre]. Rather than impacting how the computer memory operates, the above limitations merely describe the claimed information.

VII. CONCLUSION

244. It is my opinion that the Challenged Claims be canceled as unpatentable pursuant to 35 U.S.C. § 318(b).

*　　*　　*

I, Jason Janét, do hereby declare and state, that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, under Section 1001 of Title 18 of the United States Code.

Executed on: October 1, 2025

Jason Janét