

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

TESLA, INC.

Petitioner,

v.

PERCEPTIVE AUTOMATA LLC

Patent Owner

IPR2025-01574

U.S. Patent No.: 11,126,889 B2

**DECLARATION OF DR. JASON JANÉT
REGARDING U.S. PATENT NO. 11,126,889**

TABLE OF CONTENTS

	Page
I. INTRODUCTION	1
A. Engagement	1
B. Background and Qualifications	1
C. Compensation and Prior Testimony	11
D. Information Considered.....	11
II. LEGAL STANDARDS FOR PATENTABILITY	12
III. BACKGROUND TO THE '889 PATENT	20
A. Prosecution History	20
B. Technology Background	20
C. The Alleged Invention.....	21
IV. LEVEL OF ORDINARY SKILL IN THE ART	23
V. CLAIM CONSTRUCTION.....	24
VI. THE CHALLENGED CLAIMS ARE UNPATENTABLE.....	24
A. Ground 1: Claims 1-20 are obvious under 35 U.S.C. § 103 over Huval and Ross.....	24
1. Summary of Prior Art	25
a. Huval–EX1004.....	25
b. Ross–EX1005	26
2. Motivation to Combine Huval and Ross.....	27
3. Claim 2.....	33
a. 2[Pre].....	33

Declaration of Dr. Jason Janét Regarding U.S. Patent No. 11,126,889

b.	2[a]	33
c.	2[b]	37
d.	2[c]	41
e.	2[d]	44
f.	2[e]	47
g.	2[f]	47
h.	2[g]	49
4.	Claim 3	49
5.	Claim 4	50
6.	Claim 5	51
7.	Claim 6	52
8.	Claim 7	53
9.	Claim 8	55
10.	Claim 9	56
11.	Claim 10	57
12.	Claim 11	58
13.	Claim 12	58
a.	12[Pre]	58
b.	12[a]	59
c.	12[b]	59
d.	12[c]	59
e.	12[d]	59
f.	12[e]	59
g.	12[f]	59

Declaration of Dr. Jason Janét Regarding U.S. Patent No. 11,126,889

h. 12[g].....	59
14. Claim 13	59
15. Claim 14.....	60
16. Claim 15	60
17. Claim 16	60
18. Claim 17.....	60
19. Claim 18.....	60
20. Claim 19	60
21. Claim 20	60
22. Claim 1	60
a. 1[Pre].....	60
b. 1[a]	61
c. 1[b]	61
d. 1[c]	61
e. 1[d]	61
f. 1[e].....	61
g. 1[f].....	62
h. 1[g].....	62
B. Ground 2: Claims 1-20 are rendered obvious under 35 U.S.C. § 103 over Huval, Ross and Cox.....	62
1. Summary of Prior Art	63
a. Cox-EX1006.....	63
2. Motivation to Combine Huval, Ross, and Cox.....	64
3. Claim 2	69

Declaration of Dr. Jason Janét Regarding U.S. Patent No. 11,126,889

a.	2[Pre].....	69
b.	2[a]	69
c.	2[b]	70
d.	2[c]	70
e.	2[d]	72
f.	2[e].....	74
g.	2[f].....	74
h.	2[g].....	74
4.	Claim 10	74
5.	Claims 1, 3-9, and 11-19	75
6.	Claim 20	75
C.	Ground 3: Claims 1-20 are rendered obvious over Huval, Ross, and Munro	75
1.	Summary of Prior Art	76
a.	Munro–EX1009	76
2.	Motivation to Combine Huval, Ross, and Munro.....	76
3.	Claim 2	79
a.	2[Pre]-2[b].....	79
b.	2[c]	79
c.	2[d]	80
d.	2[e]	81
e.	2[f].....	81
f.	2[g].....	81
4.	Claim 12	81
5.	Claim 1	81

Declaration of Dr. Jason Janét Regarding U.S. Patent No. 11,126,889

6.	Claims 3-11 and 13-20.....	82
D.	Ground 4: Claims 1-20 and rendered obvious under 35 U.S.C. § 103 over Huval, Ross, Cox, and Munro	82
1.	Motivation to Combine Huval, Ross, Cox, and Munro	82
2.	Claim 2	83
a.	2[Pre]-2[b].....	83
b.	2[c]	83
c.	2[d]	84
d.	2[e]	84
e.	2[f].....	84
f.	2[g].....	84
3.	Claims 10 and 20.....	85
4.	Claim 12	85
5.	Claim 1	85
6.	Claims 3-9, 11, and 13-19	85
E.	Printed Matter.....	85
VII.	CONCLUSION	88

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. 11,126,889 (“the ’889 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 1011, 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.

Declaration of Dr. Jason Janét Regarding U.S. Patent No. 11,126,889

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

Declaration of Dr. Jason Janét Regarding U.S. Patent No. 11,126,889

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.

Declaration of Dr. Jason Janét Regarding U.S. Patent No. 11,126,889

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.

Declaration of Dr. Jason Janét Regarding U.S. Patent No. 11,126,889

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 1011, 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 '889 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 '889 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:

Declaration of Dr. Jason Janét Regarding U.S. Patent No. 11,126,889

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 '889 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.

Declaration of Dr. Jason Janét Regarding U.S. Patent No. 11,126,889

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 is 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 '889 PATENT

A. Prosecution History

56. The '889 Patent issued from U.S. Patent Application 16/828,823, filed March 24, 2020. 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 10,614,344, the parent of the '889 Patent, and noted that the pending claims would be allowable provided this rejection is overcome. EX1002, 100-106. In response, the applicant filed a terminal disclaimer and made minor clarifying amendments to the claims. *Id.*, 120, 127-137. The Office subsequently issued a notice of allowance. *Id.*, 140-146. The Office did not identify any particular claim recitations that rendered the claims allowable.² *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 '889 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 '889 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, 2:4-23, 3:51-64, 6:44-7:44, 11:50-12:25, 12:12-23. But this alleged invention was well known, as demonstrated by the grounds herein. Section VI. Indeed, the *inventor's own prior art* shows that the '889 Patent should not have issued. The Cox reference, which also lists Samuel Anthony as an inventor, discloses machine learning model predictions are "more consistent with the decisions of the human annotators." EX1006, ¶0032.

Declaration of Dr. Jason Janét Regarding U.S. Patent No. 11,126,889

Neither the patentee nor Mr. Anthony brought the Cox reference to the patent office's attention during prosecution of the '889 Patent. *See* EX1001; EX1002.

60. The substance of the '889 Patent further reveals that it is not new. The patent explains how model training data can be obtained and labeled by human annotators. EX1001, 4:30-55, 5:16-34, 5:46-6:4, Figs. 1, 2B. In my opinion, this concept was well known, as demonstrated by (at least) Cox. EX1006, ¶¶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, 6:8-18. But these complex-sounding terms refer to simple operations, such as averaging or otherwise combining the various annotator responses, or generating a tally of annotator responses—e.g., “how many” annotators labeled a training image in a certain way. *Id.* In my opinion, this was well known. EX1004, ¶0085 (averaging or combining human responses); EX1009, ¶0140 (tallying human annotator labels). And in another example, the patent explains that “summary statistics” relate to how long a human annotator took to label training data or to the position of the annotator's eyes when labeling training data. EX1001, 6:11-15. Yet again, in my opinion, this was well known and at least the Cox reference demonstrates this. *See, e.g.,* EX1006, ¶¶0006-0008, 0018, 0063.

61. Thereafter, the patent notes that a machine learning model is trained using the responses and used to make a prediction based on “live” input data. EX1001, 7:12-46. But, in my opinion, this was also well known. For example, the Huval reference explains that a machine learning model trained by human labeled training data is used to make predictions based on image data. EX1004, ¶0011. And the ’889 Patent’s application of these known concepts to autonomous driving does not save it. EX1001, 1:57-60, 3:52-53. In my opinion, both the Huval and Ross references teach using predictions of a machine learning model to aid control of an autonomous vehicle. EX1004, ¶¶0008-0009, 0016-0021, 0036-0039; EX1005, 1:30-2:53, 3:64-4:5, 4:33-5:11, 6:33-59, 7:57-8:31, 9:13-10:30, 15:1-16:52, Fig. 1.

62. Thus, in my opinion, the ’889 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. *Id.*

IV. LEVEL OF ORDINARY SKILL IN THE ART

63. In my opinion, a person having ordinary skill in the art (“POSITA”) relevant to the ’889 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 two to three years of work experience in computer vision, image/video processing, 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. It is my understanding that, for purposes of this proceeding, Petitioner does not present any constructions.

VI. THE CHALLENGED CLAIMS ARE UNPATENTABLE

A. Ground 1: Claims 1-20 are obvious under 35 U.S.C. § 103 over Huval and Ross

65. In my opinion, Huval alone teaches most limitations of the Challenged Claims and is further combined with Ross regarding the following limitations: (1) *receiving a plurality of user responses, each user response describing a state of mind of a road user displayed in one or more images*; (2) *training, using the training dataset, a supervised learning based model configured to predict summary statistics describing a state of mind of a road user displayed in an input image*; and (3) *predicting, by the autonomous vehicle, using the supervised learning based model, summary statistics describing a state of mind of the road user in the new image*. As discussed below, these limitations are explicitly taught by Ross, and it would have been obvious to combine Ross’s relevant teachings with Huval.

1. Summary of Prior Art

a. Huval–EX1004

66. Huval discloses systems and methods for training a machine learning model. EX1004, ¶¶0008-0010, 0016-0024, 0031-0035, 0100, Fig. 1. Specifically, Huval describes an autonomous vehicle that uses a trained model to detect objects in “a scene around the autonomous vehicle.” *Id.*

67. Huval teaches that model training data is generated from human labeled images. EX1004, ¶¶0008, 0029-0030, 0036-0039, 0043-0050. Huval’s model, which may be a neural network, is then trained using the training data. *Id.*, ¶¶0008-0011, 0015-0020, 0021-0025, 0029-0030. Once the model is trained, the Huval system “can then push a copy of the neural network . . . to the road vehicle for implementation during autonomous execution.” *Id.*; *see id.*, ¶¶0095-0096. The trained model is then used to “determine the context of a scene around the autonomous vehicle based on the sensor data” and implement appropriate control (e.g., accelerating, braking, steering) of the vehicle. *Id.* ¶¶0015-0020, 0033, 0095-0098.

68. In my opinion, Huval is analogous art to the ’889 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., navigating an autonomous vehicle in an environment including moving objects). EX1004,

¶¶0008-0015, 0021-0024, 0036-0039, 0043-0050; EX1001, 1:19-21, 1:57-30, 5:46-53, 6:8-49, Fig. 2A.

b. Ross–EX1005

69. Ross discloses techniques for “maneuvering” an autonomous vehicle. 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. The human influenced training data is formed by “human operators” who manually generate a “list of predetermined actions” based on their “observations of the actions of other road users” in sensor data captured by autonomous vehicles. *Id.*, 9:41-10:10.

70. Ross teaches that an autonomous vehicle uses the trained intent model for navigation EX1005, 7:57-8:10, 9:13-10:10, 15:1-16:52. Cameras of the vehicle capture “raw” camera data of the environment around the vehicle displaying objects in the environment. *Id.*, 7:57-8:10, 9:13-10:10, 15:1-16:52. The trained model is then executed on this data to predict “a set of possible intents for a given road user by looking at all the possible actions that the road user could execute using [the trained] intent models.” *Id.*, 3:64-4:46. Specifically, the model generates 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.*, 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.*, 9:27-40. The autonomous vehicle of Ross is then controlled accordingly. *Id.*, 1:30-2:53, 3:64-4:5, 4:33-5:11, 6:33-59, 7:57-8:31, 9:13-10:30, 15:1-16:52, Fig. 1.

71. In my opinion, Ross is analogous art to the ’889 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., navigating an autonomous vehicle in an environment including moving objects). EX1005, Abstract, 3:64-5:1; 9:27-47; EX1001, 1:19-21, 1:57-60

2. Motivation to Combine Huval and Ross

72. In my opinion, a POSITA would have been motivated to combine Huval and Ross such that:

- Huval’s annotation labels provided by human annotators label observed road user actions in image data captured by vehicles, where such labeled actions correspond to road user intent;
- Huval’s machine learning model is (1) trained with such human annotation labels and (2) predicts 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” for road users displayed in input image data; and

- Huval’s autonomous vehicle is controlled based on the predictions.

EX1004, ¶¶0008-0010, 0016-0024, 0029-0039, 0043-0050, 0096, 0100, Fig. 1; EX1005, 1:30-2:53, 4:33-5:11, 9:13-10:10, 15:1-16:52.

73. In my opinion, a POSITA would have been motivated to combine Huval and Ross for several reasons.

74. In my opinion, a POSITA would have been motivated to combine Huval and Ross because Ross simply teaches applying Huval’s system to the well-known behavior and intent context. The combination provides that Huval’s autonomous vehicle system is trained to not only identify objects in the environment of an autonomous vehicle, but also to predict the intent or state of mind of such objects and implement vehicle control accordingly. *Id.*

75. Specifically, Huval’s machine learning model is trained based on human annotator labels and implemented by an autonomous vehicle to automatically detect other vehicles, pedestrians, and cyclists in new images. EX1004, ¶¶0008-0010, 0016-0024, 0029-0039, 0043-0050, 0096, 0100, Fig. 1. Ross expands on Huval’s disclosure, teaching a predictive model for autonomous vehicles that predicts the intent of other vehicles, pedestrians, and cyclists. EX1005, Abstract. In Ross, human influenced training data is used to train an intent model, where correctly

predicted actions performed by a road user are “marked” as being correct. *Id.*, 1:30-2:53, 4:33-5:11, 9:13-10:10, 15:1-16:52. The predicted actions include various vehicle/pedestrian actions (e.g., following a lane, changing lanes, crossing a roadway at a crosswalk). *Id.*, 9:13-10:10, 15:1-16:22. And Ross explains that such actions correspond to particular intents; for example, a vehicle “moving towards another lane” corresponds to an intent of the vehicle (and its associated driver) to “change lanes when [the] desired lane is free.” *Id.* In my opinion, a “pedestrian waiting to cross the road” corresponds to an intent of the pedestrian “to cross the road when clear.” *Id.*

76. In my opinion, Ross further teaches that Huval’s trained model is used 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” for road users of the autonomous vehicle’s environment displayed in input image data. EX1005, 1:30-2:13, 3:64-4:5, 4:33-5:11, 7:57-8:31, 9:13-10:10, 15:1-16:41 (parenthetical in original), Claims 1 and 7. And thereafter, the autonomous vehicle is controlled based on the predictions. *Id.*, 1:55-2:13, 15:1-16:41, 6:33-59, 9:13-10:30, 4:66-5:11, 15:1-16:41, Fig. 1.

77. As such, the above description demonstrates that Ross merely provides implementation details as to how Huval’s system is applied to the behavior and

intent context. In my opinion, a POSITA would thus have been motivated to combine the references and have had a reasonable expectation of success. *Id.*

78. In my opinion, a POSITA would have also been motivated to combine Huval and Ross because each reference provides explicit teachings, suggestions, and motivations to make the combination. For instance, Huval explains that its machine learning model is implemented so that an autonomous vehicle can “perceive an environment around the autonomous vehicle [] and navigate the autonomous vehicle to a selected destination.” EX1004, ¶0015. At least this disclosure teaches, suggests, and motivates a POSITA to combine Huval’s teachings with Ross. This is because Ross furthers Huval’s teachings by similarly acknowledging that an autonomous vehicle “detect[s] and identif[ies] objects in the vehicle’s environment,” but adds that its system advantageously predicts the intent of road users to determine how they might act in the environment. EX1005, 3:39-4:15, 4:33-5:11, 13:29-14:9, 15:1-16:41. In my opinion, a POSITA would have thus looked to Ross and combined its teachings with Huval to provide this benefit.

79. In my opinion, a POSITA would have additionally been motivated to combine Huval and Ross because Ross provides additional benefits to Huval’s system. As noted above, Huval uses its trained model to “perceive” the surroundings of an autonomous vehicle—including the perception of “dynamic” or moveable objects in the surroundings—and effectuate appropriate vehicle control. EX1004,

¶¶0010-0011, 0015, 0018-0019, 0032-0035, 0043-0050, 0098. Huval's model is trained using human annotations, including of behavior or actions in training images (albeit in the context of actions performed by the autonomous vehicle itself rather than by objects in the vehicle's environment). *Id.*, ¶¶0008-0012, 0021-0022, 0036-0050, 0069, 0071-0092, 0099-0120, Figs. 1, 2A, 2B, 3-5.

80. In my opinion, Ross benefits Huval by teaching additional training and model predictions that allow for perceiving the surroundings of an autonomous vehicle. EX1005, 5:6-11, 9:13-10:10, 15:1-16:52. Ross teaches model training and implementation that predicts the intents of road users within the environment of an autonomous vehicle, which benefits Huval's system by making it safer and more robust. *Id.* As such, Ross provides Huval's system with the ability to determine what moving road users in the autonomous vehicle's environment might do, which would have improved the safety of Huval's vehicle control. *Id.* Huval's vehicle would have better anticipated how road users might act and implemented control to avoid possible collisions. *Id.* At bottom, Huval's autonomous vehicle would have better interacted with objects in its environment by "[r]easoning about the intents of objects, such as other road users" and thus functioned "in a more 'human-like' or 'polite' way." *Id.*; EX1005, 5:6-11, 15:1-16:52. In my opinion, a POSITA would have thus recognized the benefits of combining Huval and Ross, which motivates the combination.

81. All the reasons explained above establish that a POSITA would have been motivated to combine Huval and Ross as described. In my opinion, this combination would have been the predictable combination of well-known prior art elements (e.g., Huval's human annotation and prediction model; Ross's training and intent model teachings) according to known methods to yield predictable results (Huval's annotation labels provided by human annotators label observed road user actions in image data captured by vehicles, where such labeled actions correspond to road user intent; Huval's machine learning model is (1) trained with such human annotation labels and (2) predicts 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" for road users displayed in input image data; Huval's autonomous vehicle is controlled based on the predictions). EX1005, 15:1-21. The combination would have further used known techniques (known machine learning model training and intent prediction techniques as taught by Ross) to improve similar devices (Huval's prediction model) in the same way (using Ross's aforementioned techniques). Lastly, in addition to the above, because Ross already teaches training a predictive model to generate predictions about vehicles, pedestrians, and bicyclists in an environment of an autonomous vehicle, combining Ross with Huval's teachings about training a model that generates predictions regarding the same objects would have required minimal

changes. Making this combination is thus well within the skill of a POSITA, and a POSITA would have had a reasonable expectation of success.

3. Claim 2¹

82. In my opinion, Huval and Ross render obvious claim 2.

a. 2[Pre]

83. In my opinion, to the extent the preamble is limiting, Huval and Ross teach it. The Huval-Ross combination teaches computer implement techniques for controlling an autonomous vehicle based on the predicted intents of road users (e.g., vehicles, pedestrians, and bicyclists) displayed in camera images of the environment of the vehicle, shown further below (*[a] computer-implemented method for controlling an autonomous vehicle based on a predicted state of mind of road users in a scene captured by a camera of the autonomous vehicle*). EX1004, ¶¶0008-0009, 0016; EX1005, 3:64-4:15, 7:57-8:53, 9:66-10:30; See Sections VI.A.2, VI.A.3.b (element 2[a]), VI.A.3.e (element 2[d]), VI.A.3.f (element 2[e]), VI.A.3.g (element 2[f]).

b. 2[a]

84. In my opinion, Huval and Ross teach element 2[a]. Huval describes that its system receives images from a training set of images, and also receives images

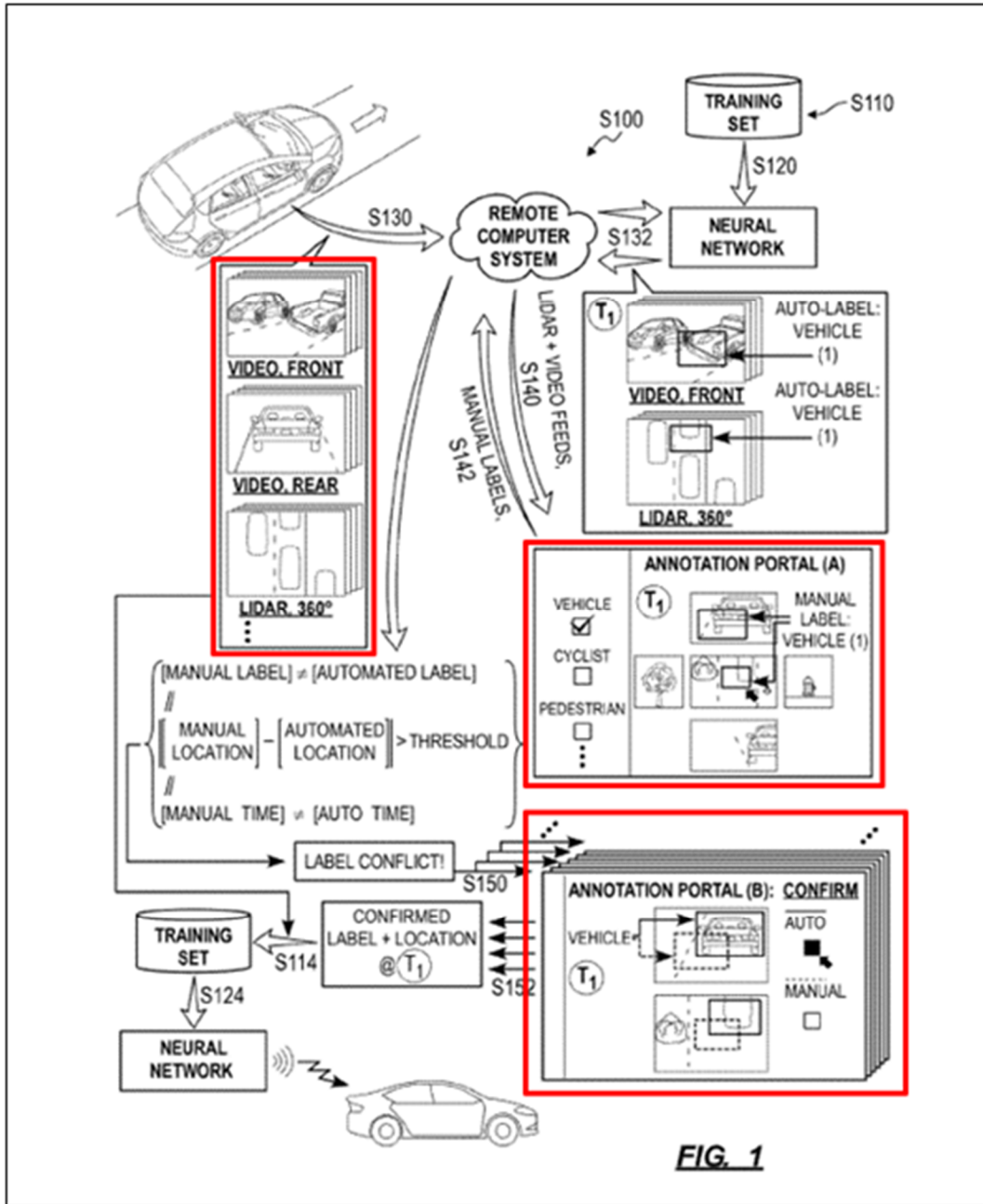
¹ My analysis begins with independent claim 2 as claims 3-11 depend from this claim.

from cameras of an autonomous vehicle that are recorded “during operation” (*receiving a plurality of images displaying road scenes captured by one or more vehicles*). EX1004, ¶¶0008-0009, 0017, 0021-0022, 0025, 0029-0030, 0037-0038. The autonomous vehicle includes cameras that “collect information about the autonomous vehicle’s environment” and specifically “output a video feed containing a sequence of digital photographic images (or ‘frames’).” EX1004, ¶¶0016-0017. The images include objects in the autonomous vehicle’s surroundings such as “vehicles, pedestrians, and cyclists,” which means the image displays a *road scene*. *Id.*, ¶¶0008-0009, 0011, 0016-0018, 0025, 0029-0030, 0071-0092, Figs. 1, 3, 4. Thus, in my opinion, by capturing images of vehicles, pedestrians, and cyclists surrounding the autonomous vehicle and providing these images to Huval’s system, Huval teaches *receiving a plurality of images displaying road scenes captured by one or more vehicles*. *Id.*

85. Such images are transmitted by Huval’s remote computer system to a human annotator for manual labeling, which, in my opinion, further teaches *receiving a plurality of images displaying road scenes captured by one or more vehicles*. EX1004, ¶0025. For example, a vehicle can upload images to the remote computer system and the computer system can “store these optical data in a remote database and queue these optical data for transmission to human annotators for manual labeling.” *Id.* The human annotators receive the images in an “annotator

Declaration of Dr. Jason Janét Regarding U.S. Patent No. 11,126,889

portal” that is “execute[d] on a desktop computer” and render[s] frames of [] video feeds for review by a human annotator.” *Id.*, ¶0037. Multiple instances of queuing and transmitting optical data to annotators occur simultaneously across the Huval system. *Id.*, ¶¶0038 (“Additional instances of the annotation portal can execute on other machines to label other optical images.”), 0085.



EX1004, Fig. 1 (annotated)

86. Huval also describes *receiving a plurality of images displaying road scenes captured by one or more vehicles* in an additional way. With respect to receiving images from a training set of images, Huval describes a label conflict

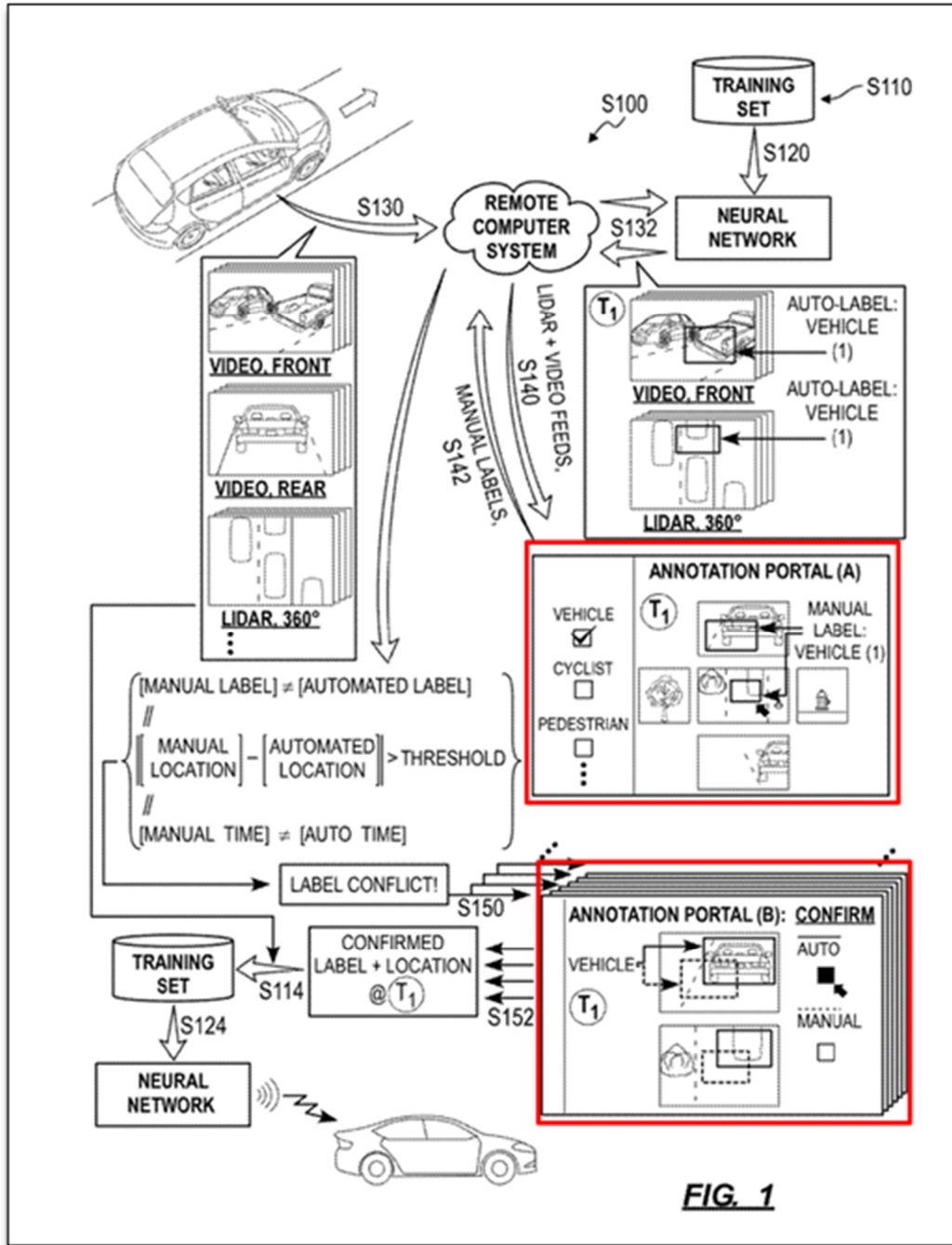
resolution process to refine training data where human annotators resolve conflicts between automatically generated labels and manual labels. EX1004, ¶¶0008-0009, 0071-0092. For example, Huval's system can determine whether conflicts exist between human (manual) and automated labels of a training image. *Id.*, ¶¶0074-0085. If a conflict exists, the image having conflicting labels is transmitted to multiple human annotators, rendered on their respective annotation portals, to confirm the correct label. *Id.* As such, in my opinion, Huval describes that its human annotators receive images in this manner as well (*receiving a plurality of images displaying road scenes captured by one or more vehicles*). *Id.*

c. 2[b]

87. In my opinion, Huval and Ross teach element 2[b]. Huval describes that images captured by autonomous vehicles are sent to human annotators, who manually label the images and send their responsive labels to Huval's computer system (*receiving a plurality of user responses*). EX1004, ¶¶0008-0012, 0021-0022, 0036-0050, 0069, 0085-0086, 0090, Figs. 1, 2A, 2B, 3; *see also* Section VI.A.3.b (element 2[a]). The annotation portal enables the human annotators to label the images by, for example, "placement of labeled boundary boxes" or selecting a label from "a set of object types related to perception, such as road vehicles (e.g., passenger vehicles, commercial trucks, motorcycles), pedestrian, cyclist . . . and other dynamic [] objects." EX1004, ¶¶0043-0044, Figs. 1, 2A, 2B, 3, 5.

88. In my opinion, the manual labels received by Huval’s remote computer system (e.g., blocks 114, 142, 152 of Figures 1, 3, and 4) describe road users present in the images, such as a road vehicle, pedestrian, and cyclist (*each user response describing...a road user displayed in one or more images*). EX1004, ¶¶0008-0012, 0036-0050, Figs. 1, 3, and 4.

89. Additionally, Huval describes a label conflict resolution process where human annotators review conflicting automatically generated labels and manual labels. EX1004, ¶¶0008-0012, 0036-0050, 0071-0092; *see also* Section VI.A.3.b (element 2[a]). When a label conflict is detected, human annotators are sent the training image at issue and additional annotation information is collected from these annotators for the image. *Id.* Here, the system may “expand review of the optical image to multiple (e.g., three, dozens) human annotators” and receive responses from these multiple annotators. *Id.*, ¶¶0085-0086. This disclosure therefore demonstrates another example in which annotation responses from multiple human annotators describing road users in images is received (*receiving a plurality of user responses, each user response describing...a road user displayed in one or more images*). *Id.*



EX1004, Fig. 1 (annotated)

90. In my opinion, the combination of Ross and Huval further teaches that each of the annotation responses describes a *state of mind of a road user displayed in one or more images*. In the Huval-Ross combination, the human annotation

process would have included the labeling of observed road user (e.g., vehicle, pedestrian, bicyclist) actions in image data captured by vehicles, where such labeled actions correspond to the intents of road users and thus describe the state of minds of the road users (*each user response describing a state of mind of a road user displayed in one or more images*). EX1004, ¶¶0008-0012, 0021-0022, 0036-0050, 0069, 0071-0092, Figs. 1, 2A, 2B, 3, 5; EX1005, 1:55-2:13, 4:33-46, 9:13-10:10, 15:1-16:41; Section VI.A.2. For example, the actions include, for a vehicle in the image data, following a current lane, changing lanes, or turning at an intersection. EX1005, 9:13-10:10, 15:1-16:41. A vehicle “moving towards another lane” corresponds to an intent of the vehicle (and its associated driver) to “change lanes when [the] desired lane is free,” indicating a state of mind of the vehicle (and driver) to change lanes when the desired lane is free. *Id.* Actions also include, for a pedestrian in image data, crossing a roadway, crossing a roadway in a crosswalk, and jaywalking, and such actions describe an intent of the vehicle or pedestrian. *Id.* For example, a “pedestrian waiting to cross the road” corresponds to an intent of the pedestrian “to cross the road when clear,” indicating a state of mind of the pedestrian that they wish to cross the road when no traffic is present. *Id.* In view of the above, it is my opinion that Huval and Ross therefore each element 2[b].

d. 2[c]

91. In my opinion, Huval and Ross teach element 2[c]. Huval describes that it obtains multiple human annotator manual labels for road vehicles, pedestrians, and cyclists displayed in sample images. EX1004, Abstract, ¶¶0008-0018, 0021-0025, 0029-0030, 0036-0050, 0085-0092, 0095-0098, 0118, Figs. 1, 3, 4; Section VI.A.3.c (element 2[b]). For each image, Huval “combine[s]” the labels by determining an “average” to generate a final label for each image, teaching *summary statistics of user responses describing...road users displayed in the plurality of images. Id.* Because the average label is generated for each image, a collection of average values summarizing the labels of a collection of training images is generated by Huval, which teaches *summary statistics* as claimed. In my opinion, because these average values are a collection of a numerical measurements summarizing the labels, a POSITA would have recognized that these values form statistics. *Id.*

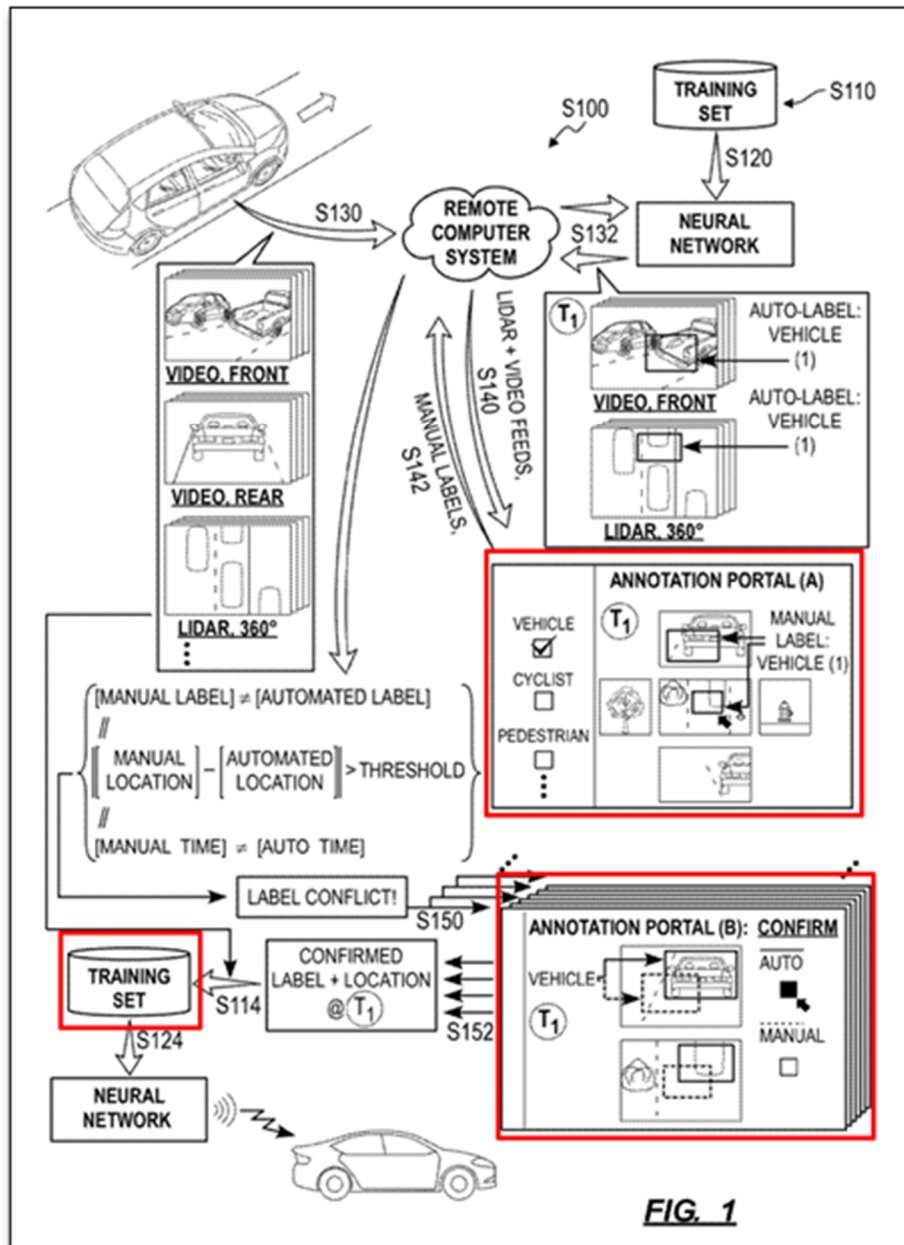
92. Additionally, Huval’s disclosure aligns with the ’889 patent’s description of *summary statistics*. EX1001, 10:60-67, 6:8-24. The patent explains that “summary statistics” associated with user responses include “measurements of the central tendency of the distribution of scores like the mean, median, or mode” and are associated with “a content of a response.” *Id.* In my opinion, the average value generated for each training image, as disclosed by Huval, provides such a

central tendency measurement of mean, and further reflect the content of human labels. *Id.*; EX1004, ¶0085.

93. The training images and associated combined labels (*summary statistics*) are appended to Huval’s “training set,” which means they are part of a *training dataset (generating a training dataset comprising summary statistics of user responses describing...road users displayed in the plurality of images)*. EX1004, ¶¶0085-0090. As taught by Huval, the “manual labels” that form the combined labels provided by human annotators are used to “train and refine the neural network over time” and are thus part of the *training dataset* used to train Huval’s model. EX1004, Abstract, ¶¶0008-0018, 0021-0025, 0029-0030, 0036-0050, 0085, 0090, 0095-0098, 0118, Figs. 1, 3, and 4.

94. In my opinion, Huval further teaches *summary statistics of [the] user responses* in an additional way. The training set of Huval comprises *summary statistics of [the] user responses* at least because it includes the manual labels that are “packaged” into the training set and reflect the summarized statistics of the content of all accurate manual human labels that have been obtained and are used for training. EX1004, Abstract, ¶¶0008-0018, 0021-0025, 0029-0030, 0036-0050, 0085, 0090, 0095-0098, 0118, Figs. 1, 3, and 4; *see id.*, ¶¶0022-0023. For example, Huval’s remote computer system “aggregates labeled optical data, such as stored in a remote database, into a training set on which a localization, perception, motion

planning, and/or other neural network can then be trained.” *Id.*, ¶¶0021-0024; *see also id.*, ¶¶0069-0070. In my opinion, because these aggregated labels are a collection of measurements summarizing how data has been labeled, a POSITA would have recognized that these aggregated labels form statistics. *Id.* This disclosure further teaches *summary statistics of [the] user responses*.



EX1004, Fig. 1 (annotated)

95. In my opinion, in combination with Ross, the response annotation labels would also correspond to the intents and actions of road users and thus describe the state of minds of the road users (*user responses describing the state of minds of road users displayed in the plurality of images*). See *supra* Section VI.A.2 (motivation to combine) and Section VI.A.3.c (element [2b]).

e. 2[d]

96. In my opinion, Huval and Ross teach element 2[d]. Huval describes that its training set is used to train its machine learning model (which is a neural network), where the model is supervised and corrected by human label information (*training, using the training dataset, a supervised learning based model*). EX1004, Abstract, ¶¶0008-0020, 0021-0025, 0029-0030, 0036-0050, 0079-0090, 0095-0098, 0118, Figs. 1, 3, and 4; see Section X.A.3.d (element 2[c]). Huval's model is a *supervised learning based model* because it is “train[ed] and refine[d]” by human labeled training images, including by comparing model predictions and labels provided manually by human annotators, and “retraining” the model to “resolve conflicts” between the model predictions and human labels. EX1004, ¶¶0008-0012, 0085-0096. Moreover, Huval describes that its model is “train[ed]” using the training set, and also “developed” and “refined” using the training set, so that it detects objects and navigational actions in the scene around its autonomous vehicle

(*training, using the training dataset, a supervised learning based model*). *Id.*, ¶¶0008-0015, 0021-0024, 0079-0098.

97. Additionally, in my opinion and as discussed in Section VI.A.2, the Huval-Ross combination teaches that Huval’s 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” for road users (e.g., pedestrians, vehicles, bicyclists) displayed in input image data (*supervised learning based model configured to predict summary statistics describing a state of mind of a road user displayed in an input image*). EX1004, Abstract, ¶¶0008-0020, 0021-0025, 0029-0030, 0036-0050, 0085-0090, 0095-0098, 0118, Figs. 1, 3, and 4; EX1005, 1:30-2:53, 3:64-4:5, 4:33-5:11, 7:57-8:31, 9:13-10:10, 15:1-16:52, Claims 1 and 7 (parenthetical in original); Section VI.A.2. This predicted “set of possible intents...” teaches *summary statistics describing a state of mind of a road user displayed in an input image* because it is a summary of the various predicted intents and likelihood values for a particular object in input image data. *Id.*; Section VI.A.2, Sections VI.A.3.c-VI.A.3.d (elements 2[b], 2[c]); *infra* Section VI.A.3.g (element 2[f]). The likelihood values summarize how likely each of the predicted intents and trajectories of the object is to occur, where the greatest likelihood value corresponds to the prediction most likely to occur. *Id.*; EX1005, 9:27-40, 15:50-16:52.

98. In my opinion, because these predicted intents/trajectories and associated likelihood values are a collection of measurements summarizing the various potential actions of an object and how likely each is to occur, a POSITA would have recognized that these items form statistics. EX1004, Abstract, ¶¶0008-0020, 0021-0025, 0029-0030, 0036-0050, 0085-0090, 0095-0098, 0118, Figs. 1, 3, and 4; EX1005, 1:30-2:53, 3:64-4:5, 4:33-5:11, 7:57-8:31, 9:13-10:10, 15:1-16:52, Claims 1 and 7 (parenthetical in original); Section VI.A.2, Sections VI.A.3.c-VI.A.3.d (elements 2[b], 2[c]); *infra* Section VI.A.3.g (element 2[f]). Indeed, the predicted intents and associated likelihood values reflect statistics summarizing how an object within the environment of Ross's autonomous vehicle may behave. *Id.* Additionally these predicted intents describe a *state of mind of a road user* because they describe what the road user in the environment of Huval's autonomous vehicle might want to do. *Id.*

99. Such predictions are generated both during a training phase of Huval's system and an operational phase. EX1004, Abstract, ¶¶0008-0020, 0021-0025, 0029-0030, 0036-0050, 0078, 0085-0092, 0095-0098, 0118, Figs. 1, 3, and 4; EX1005, 1:30-2:53, 3:64-4:5, 4:33-5:11, 7:57-8:31, 9:13-10:10, 15:1-16:52, Claims 1 and 7 (parenthetical in original); Section VI.A.2, Sections VI.A.3.c-VI.A.3.d (elements 2[b], 2[c]); *infra* Section VI.A.3.g (element 2[f]). For example, by correcting and removing inaccurate labels, Huval teaches machine learning training

techniques that minimize inaccuracies between the human labeled images and the automatically labeled images, “thereby ‘healing’ the neural network.” EX1004, ¶¶0012, 0073-0094. Moreover, Huval demonstrates that its model is trained because it explains that training involving more a high volume of correctly labeled training data results in a more effective and accurate model. *Id.*, ¶0011 (emphasis added); *see also* Sections VI.A.3.c, VI.A.3.d (elements 2[b], 2[c]), VI.A.2.

f. 2[e]

100. In my opinion, Huval and Ross teach element [2e]. Huval describes that during an operational phase in which the trained model of Huval has been deployed to an autonomous vehicle, the cameras of the autonomous vehicle capture image data of the autonomous vehicle’s surroundings that include other vehicles, pedestrians, and cyclists (*receiving, by an autonomous vehicle, a new image captured by a camera of the autonomous vehicle, the new image of a scene including a road user*). EX1004, ¶¶0008-0009, 0011, 0016-0019, 0025, 0029-0030, 0044, 0095-0098, Figs. 1, 3, 4; *see also* Section VI.A.3.b (element 2[a]).

g. 2[f]

101. In my opinion, Huval and Ross teach element 2[f]. The Huval-Ross combination teaches that Huval’s deployed model of its autonomous vehicle predicts 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” for road users (e.g., pedestrians, vehicles, bicyclists) displayed in new input image data as the vehicle operates in an environment (*predicting, by the autonomous vehicle, using the supervised learning based model, summary statistics describing a state of mind of the road user in the new image*). EX1004, Abstract, ¶¶0008-0020, 0021-0025, 0029-0030, 0036-0050, 0085-0090, 0095-0098, 0118, Figs. 1, 3, and 4; *see also* Sections VI.A.2, VI.A.3.a-VI.A.3.e (elements 2[pre]-2[d]); EX1005, 1:30-2:53, 3:64-4:5, 4:33-5:11, 7:57-8:31, 9:13-10:10, 15:1-16:52, Claims 1 and 7. This predicted “set of possible intents...” teaches *summary statistics describing a state of mind of the road user in the new image* because it is a summary of the various predicted intents and likelihood values for a particular object in the input image data. *Id.* The likelihood values summarize how likely each of the intents and trajectories of the object is to occur, where the greatest likelihood value corresponds to the prediction most likely to occur. *Id.*; EX1005, 9:27-40, 15:50-16:52, Fig. 10. Because these predicted intents/trajectories and associated likelihood values are a collection of measurements summarizing the various potential actions of an object and how likely each is to occur, it is my opinion that a POSITA would have recognized that these items form statistics. *Id.*

102. The predicted intents and associated likelihood values thus reflect statistics summarizing how an object within the environment of Ross’s autonomous vehicle may behave. *Id.* Additionally these predicted intents describe a *state of mind*

of the road user because they describe the predictions of what the road user in the environment of Huval's autonomous vehicle might want to do. *Id.*

h. 2[g]

103. In my opinion, Huval and Ross teach element 2[g]. The Huval-Ross combination describes that Huval's autonomous vehicle is controlled based on the predicted intents of vehicles, pedestrians, and cyclists in image data (*controlling the autonomous vehicle based on the prediction of the supervised learning based model*). EX1004, ¶¶0008-0024, 0036-0050, 0085-0086; EX1005, 1:55-2:13, 15:1-16:41, 6:33-59, 9:13-10:30, 4:66-5:11, 15:1-16:41, Fig. 1; Section VI.A.2. Specifically, as taught by Ross, the highest likelihood intent or intents for a road user are determined and used to maneuver the autonomous vehicle. EX1005, 4:66-5:11, 10:11-30, 15:1-16:41; Section VI.A.2.

4. Claim 3

104. In my opinion, Huval and Ross render obvious claim 3. Huval's image data captured by its autonomous vehicle cameras display pedestrians and the images are a "sequence of digital photographic images (or 'frames') of a video feed" (*wherein the road user in the input image is a pedestrian*). EX1004, ¶¶0008-0009, 0011, 0016-0018 (parenthetical in original), 0025, 0029-0030, 0044-0050, Figs. 1, 3, 4; *see also* Section VI.A.3.b, VI.A.3.f (elements 2[a], 2[e]).

105. In my opinion, the Huval-Ross combination's predicted "set of possible intents..." for a pedestrian in the image data reflects the state of mind of the pedestrian and "includes a corresponding next action to be initiated by the road user," and each predicted intent has an associated likelihood of occurring; this demonstrates that the predicted intents indicate how likely the pedestrian is to perform a particular action (*and the state of mind predicted by the supervised learning based model indicates whether the pedestrian is likely to perform an action*). EX1005, 1:55-2:13, 4:33-46, 9:13-40, 15:1-21; Section VI.A.2, Section VI.A.3.g (element 2[f]). As taught by Ross, the predicted intent can be that the pedestrian "would like to cross the roadway" (*the action including one of: staying in place, or crossing a street*). EX1005, 16:23-32; *see also id.*, 4:47-65, 9:41-10:10.

5. Claim 4

106. In my opinion, Huval and Ross render obvious claim 4. Huval teaches *wherein the road user in the input image is a cyclist* for the same reasons discussed in claim 3. *Supra*, Section VI.A.4.

107. In addition, it is my opinion that the Huval-Ross combination's predicted "set of possible intents..." for a cyclist in the image data reflects the state of mind of the cyclist and the associated likelihoods that a cyclist is likely to perform an action (*and the state of mind predicted by the supervised learning based model indicates whether the cyclist is likely to perform an action*). Section VI.A.2, Section

VI.A.3.g (element 2[f]); EX1005, 3:39-47, 5:1-11, 9:13-10:10, 15:1-16:52, Claims 1, 2, 3, and 7, Fig. 8. “[E]ach possible intent of the set includes a corresponding next action to be initiated by the road user.” EX1005, 1:55-2:13, 4:33-46, 9:13-40, 15:1-21. In my opinion, and as taught by Ross, the predicted intent can be that a bicyclist would like to “change lanes when [a] desired lane is free of traffic” and can also be that a bicyclist would like to “move through [an] intersection when clear,” and related actions to these intents include the bicyclist is “moving towards another lane” and “waiting at an intersection” (*indicates whether the cyclist is likely to cross a lane*) EX1005, 15:64-16:22; *see also* EX1005, 15:64-16:22. This disclosure thus provides that the predictions indicate whether the bicyclist is likely to change or cross lanes. *Id.*

6. Claim 5

108. In my opinion, Huval and Ross render obvious claim 5. Huval describes that its training data includes human manual labels that are “packaged” into the training set and reflect the summarized statistics of the content of all accurate human manual labels that have been obtained and are used for training. EX1004, Abstract, ¶¶0008-0018, 0021-0025, 0029-0030, 0036-0050, 0085, 0090, 0095-0098, 0118, Figs. 1, 3, and 4; *see id.*, ¶¶0022-0023; *see also* Section VI.A.3.d (element 2[c]). Moreover, human annotator label responses for each training image can be combined by averaging them to generate a final annotation label for each image. EX1004,

¶0085; Sections VI.A.3.c-VI.A.3.e (elements 2[b]-2[d]). By averaging the labels of multiple human annotators for each image, Huval generates a collection of average values summarizing the labels of a collection of training images (*summary statistics*). *Id.* And, in my opinion, this collection of values thus includes a central tendency (e.g., an average) of the human label responses for each image, which teaches *the summary statistics comprises one or more of: a central tendency*

7. Claim 6

109. In my opinion, Huval and Ross render obvious claim 6. Huval teaches that human annotators are sent images in an “annotation portal” (*sending one or more images for display to a viewer via a user interface*). EX1004, ¶¶0008-0009, 0037, 0071-0092; *see also* Section VI.A.3.b (element 2[a]). The Huval system “render[s] a prompt” to the human annotator using the annotation portal “to select one of the manual and automated labels to confirm a type of an object represented” in the image such as a vehicle, pedestrian, or cyclist (*the one or more images including a road user, wherein the user interface is configured to prompt the viewer*). *Id.*, ¶¶0043-0044, 0052-0055, 0080. Huval also describes a similar process of providing an image with two different automated labels of varying confidence scores to human annotators and “a prompt the human annotator to confirm one of the first and second automated labels or to manually enter an alternate label for

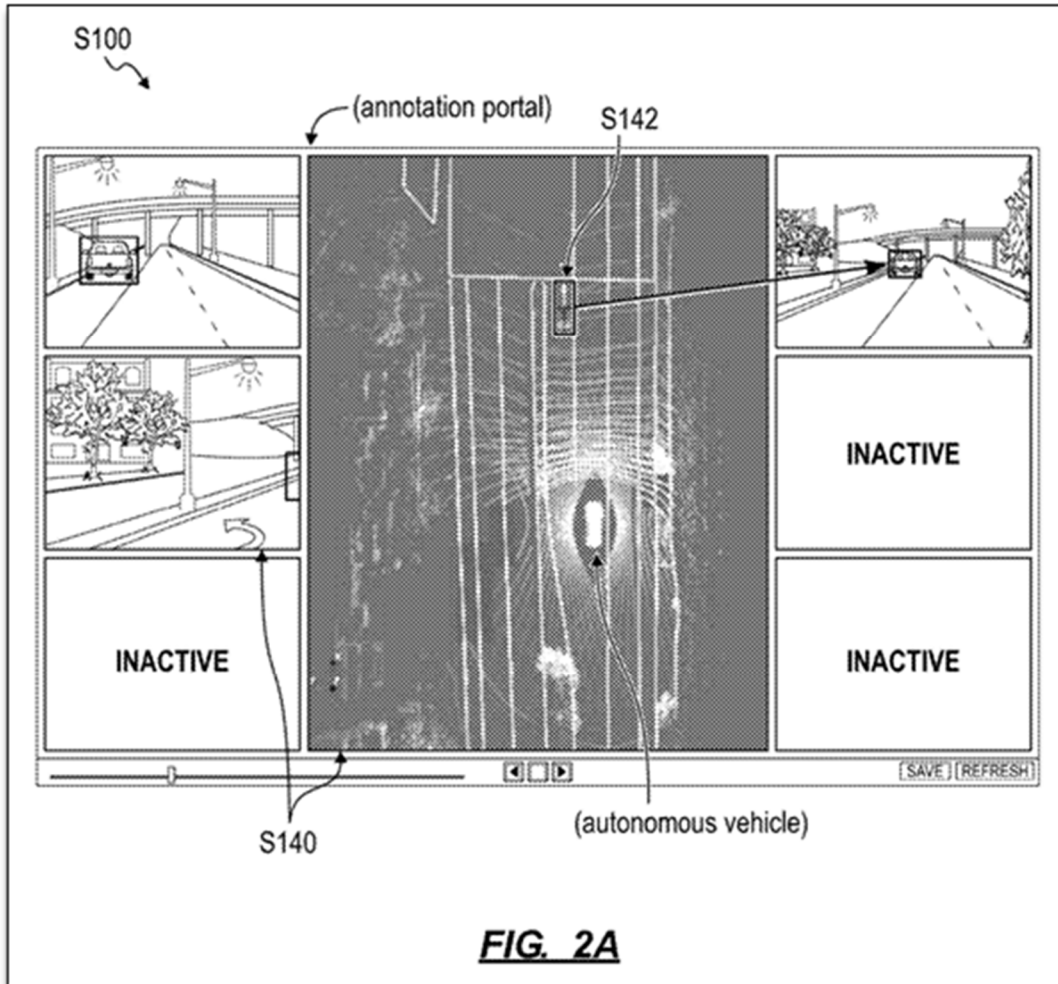
another object type” (*the one or more images including a road user, wherein the user interface is configured to prompt the viewer*). *Id.*, ¶¶0052-0055.

110. In my opinion, Ross teaches the labeling of road user actions in image data, where such labeled actions correspond to the intents of road users and thus describe the state of minds of the road users. Section VI.A.2. Thus, as combined with Huval, Ross teaches that Huval’s human annotators would have been prompted to describe road user actions, including how they are appearing to act in the image (*prompt the viewer to describe how the road user will act*). *Id.*; EX1004, ¶¶0008-0009, 0052-0055, 0071-0092; EX1005, 9:41-10:10, 15:50-16:22.

8. Claim 7

111. In my opinion, Huval and Ross render obvious claim 7. Huval describes that, as part of its process for receiving human labels of images from human annotators, a set of video feeds recorded by “the forward-facing, rear-facing, left-facing, and right facing cameras” of a vehicle are served to the annotators’ “local computer system for manual labeling,” where the video images to be labeled are displayed on an “annotation portal” of the computer system and include a road user such as a pedestrian, vehicle, or cyclist (*wherein receiving a user response comprises: generating a stimulus comprising a set of images; sending the generated stimulus for display via a user interface*). EX1004, ¶¶0008-0009, 0011, 0016-0018,

0025, 0029-0030, 0036-0042, 0050, Figs. 1, 2A, 2B, 3, 4. Moreover, “one or more video feeds exclusively” can be displayed by the portal. *Id.*, ¶¶0041-0042, 0050.



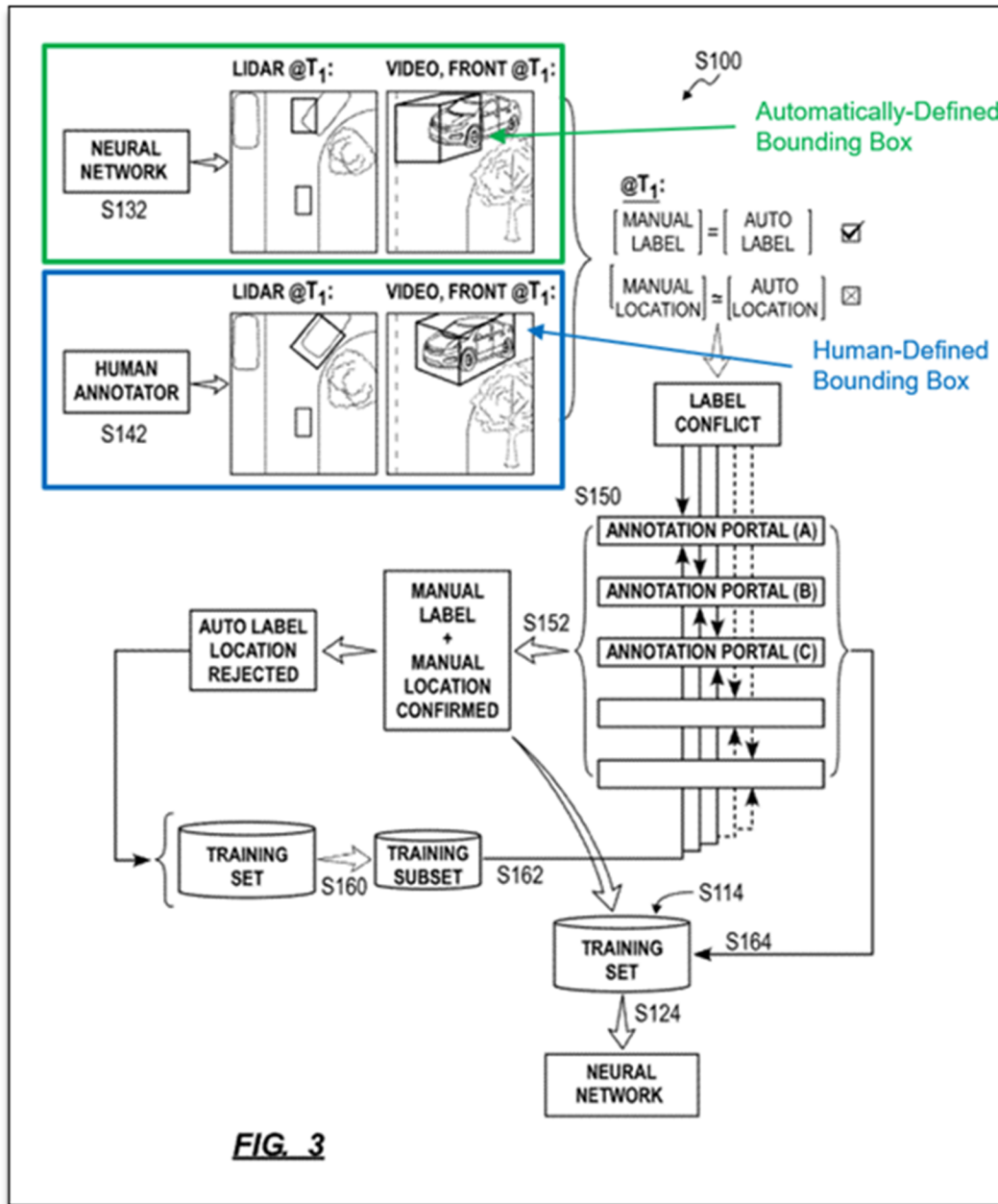
EX1004, Fig. 2A

112. The human annotator then adds, via their respective annotation portals, manual labels to the images in response by, for example, making a selection from “a dropdown menu” and draw a bounding box around the object “linked to a cursor” (*and receiving, via the user interface, a user response describing the stimulus*). *Id.*; *see also id.*, ¶¶0008-0012, 0037-0050. In my opinion, as taught by Ross, the

responses received from Huval's annotators would have described the training image samples and the observations of what actions a road user in the sample is doing. *See* Sections VI.A.2, VI.A.3.c (element 2[b]).

9. Claim 8

113. In my opinion, Huval and Ross render obvious claim 8. As discussed in Sections VI.A.3.b (element 2[a]) and VI.A.8 (claim 7), Huval describes a label conflict resolution process where human annotators compare conflicting automated and manual labels to confirm the correct label and refine the training dataset. EX1004, ¶¶0008-0009, 0071-0092. As part of this process, “the annotation portal can present” to a human annotator a modified version of a captured image that includes “a pre-generated automated label”, such as a “bounding box,” for an “object detected at an automatically-defined location by the neural network.” EX1004, ¶¶0031-0035, 0043-0050, 0052-0054, 0071-0092, Claim 3, Figs. 1-3. For example, as shown by the annotated green portion in Figure 3 below, a road user vehicle in the image is outlined by an automated label bounding box, which modifies the image (*modifying an image to generate a stimulus*) that the human annotator reviews when comparing label conflicts. *Id.* In my opinion, the bounding box extracts a subset of the image around a road user vehicle in the image (*the modifying comprising extracting a subset of the image contained in a bounding box around the road user displayed in the image*). *Id.*



EX1004, Fig. 3 (annotated)

10. Claim 9

114. In my opinion, Huval and Ross render obvious claim 9. Huval describes that a training image reviewed for label conflicts includes an automated label bounding box around a road user. EX1004, ¶¶0052-0054, Fig. 3; *see* Section VI.A.9

(claim 8). In addition to confirming the label, the human annotator may “adjust” or move the bounding box on the image. EX1004, ¶0052. Huval’s system can then “recalculate the confidence score for the pre-generated automated label based on feedback received from the human annotator.” *Id.* Huval teaches that its model uses the confidence score associated with bounding box label, which is calculated based on whether the human annotator confirms the location of the automatically-defined location, to “selectively distribute automated labels to human annotators” as part of the training process. *Id.* ¶¶0034, 0052-0054, Fig. 3. In my opinion, Huval’s model therefore uses the bounding box, and thus its location on an image, shape, and dimensions, to make model training decisions (*wherein the supervised learning based model uses one or more features comprising: dimensions of the bounding box; location of the bounding box; shape of the bounding box; or a change in size or position of the bounding box across two images representing video frames*). *Id.*

11. Claim 10

115. In my opinion, Huval and Ross render obvious claim 10. As addressed above for claims 6 and 7, Huval discloses that the process for collecting human annotation responses may include sending prompts to the annotators that elicit explicit responses (*the user response is determined based on: an explicit response to a question asked via a user interface*). *Supra*, Sections VI.A.7, VI.A.8. In my opinion, a POSITA would have understood that such prompting is a *question*. *Id.*

This is because when an annotator is prompted to select a label, the annotator is being asked which label should be selected for the training image that the annotator is reviewing. *Id.* Huval further discloses that the relative time taken for human annotators to apply labels is also used by Huval's system to determine how annotator labels were applied to training images (*the user response is determined based on... an implicit data representing one or more of ... time taken to respond*). EX1004, ¶0076.

12. Claim 11

116. In my opinion, Huval and Ross render obvious claim 11. The Huval-Ross model is, for example, a neural network (*wherein the supervised learning based 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, ¶¶0008-0009, 0016; Section VI.A.3.e. (element 2[d]).

13. Claim 12

117. In my opinion, Huval and Ross render obvious claim 12.

a. 12[Pre]

118. In my opinion, to the extent limiting, the Huval-Ross combination teaches element 12[pre] for the same reasons discussed in Section VI.A.3.a (element 2[Pre]). Moreover, Huval describes its systems and methods “can be embodied and/or implemented at least in part as a machine configured to receive a computer-

readable medium storing computer-readable instructions” that “can be executed by computer-executable components” such as “a processor” or “any suitable dedicated hardware device” (*[a] non-transitory computer readable storage medium storing instructions that when executed by a computer processor cause the computer processor to perform steps of a computer-implemented method*). EX1004, ¶0121.

b. 12[a]

119. See Section VI.A.3.b (element 2[a]).

c. 12[b]

120. See Section VI.A.3.c (element 2[b]).

d. 12[c]

121. See Section VI.A.3.d (element 2[c]).

e. 12[d]

122. See Section VI.A.3.e (element 2[d]).

f. 12[e]

123. See Section VI.A.3.f (element 2[e]).

g. 12[f]

124. See Section VI.A.3.g (element 2[f]).

h. 12[g]

125. See Section VI.A.3.h (element 2[g]).

14. Claim 13

126. See Section VI.A.4 (claim 3).

15. Claim 14

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

16. Claim 15

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

17. Claim 16

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

18. Claim 17

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

19. Claim 18

131. *See* Section VI.A.9 (claim 8).

20. Claim 19

132. *See* Section VI.A.10 (claim 9).

21. Claim 20

133. *See* Section VI.A.11 (claim 10).

22. Claim 1

134. In my opinion, Huval and Ross render obvious claim 1.

a. 1[Pre]

135. In my opinion, to the extent the preamble is limiting, the Huval-Ross combination teach it for the same reasons discussed in Section VI.A.3.a (element 2[Pre]) and VI.A.13.a (element 12[Pre]). Moreover, Huval teaches that its disclosed method “can be executed by a computer network” that includes a “remote computer

system,” where its described systems and methods “can be embodied and/or implemented at least in part as a machine configured to receive a computer-readable medium storing computer-readable instructions” that “can be executed by computer-executable components” such as “a processor” or “any suitable dedicated hardware device” (*[a] computer system comprising: a computer processor; and a non-transitory computer readable storage medium storing instructions that when executed by the computer processor cause the computer processor to perform steps of a computer-implemented method*). EX1004, ¶¶0008-0015, 0021-0025, 0029-0030, 0031-0035, 0121; *See* Sections VI.A.3.a (element 2[Pre]), VI.A.13.a (element 12[Pre]).

b. 1[a]

136. *See* Section VI.A.3.b (element 2[a]).

c. 1[b]

137. *See* Section VI.A.3.c (element 2[b]).

d. 1[c]

138. *See* Section VI.A.3.d (element 2[c]).

e. 1[d]

139. *See* Section VI.A.3.e (element 2[d]).

f. 1[e]

140. *See* Section VI.A.3.f (element 2[e]).

g. 1[f]

141. See Section VI.A.3.g (element 2[f]).

h. 1[g]

142. See Section VI.A.3.h (element 2[g]).

B. Ground 2: Claims 1-20 are rendered obvious under 35 U.S.C. § 103 over Huval, Ross and Cox

143. In my opinion, the Huval-Ross combination renders obvious the Challenged Claims. However, Huval and Ross are further combined with Cox regarding the following limitations in Ground 2: (1) *generating a training dataset comprising summary statistics of user responses describing the state of minds of road users displayed in the plurality of images*; (2) *training, using the training dataset, a supervised learning based model configured to predict summary statistics describing a state of mind of a road user displayed in an input image*; and (3) *predicting, by the autonomous vehicle, using the supervised learning based model, summary statistics describing a state of mind of the road user in the new image*. As discussed below, these limitations are explicitly taught by Cox, and it would have been obvious to combine Cox's relevant teachings with Huval and Ross.

1. Summary of Prior Art

a. Cox–EX1006

144. 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 '889 Patent. EX1001, Cover; EX1006, Cover.

145. Cox is directed toward “machine learning augmented by measurements acquired from human populations.” EX1006, Abstract, ¶¶0001-0004. Cox teaches crowd-sourcing training data used to train machine learning algorithms, such as supervised machine learning models. *Id.*, Abstract, ¶¶0001-0004, 0015-0016, 0025-0034, 0044, 0060. After transmitting images to a plurality of human annotators/labelers, Cox teaches compiling statistical information about the human responses, including psychometric data associated with the responses, and training a supervised learning model on this information. *Id.*, ¶¶0015-0019, 0029-0032, Figs. 1A-B, 3. The goal of Cox’s machine learning techniques is to “better mimic human performance.” *Id.*, ¶0004. Moreover, Cox touts that its “supervised learning” techniques may be “directly applied to several important domains where machine learning is found” and expressly suggests applying the invention to “driverless” automobile applications. *Id.*, ¶¶0056, 0060.

146. In my opinion, Cox is analogous art to the '889 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., navigating an autonomous vehicle in an environment including moving objects). EX1006, ¶¶0029-0032, Figs. 1A-B, 3; EX1001, 1:19-21, 1:57-60, 5:46-53; 6:8-49; Fig. 2A.

2. Motivation to Combine Huval, Ross, and Cox

147. In my opinion, a POSITA would have been motivated to combine Huval and Ross as discussed in Section VI.A.2. Further, in my opinion, a POSITA would have been further motivated to combine Huval and Ross with Cox such that:

- Cox's psychometric data is generated from the Huval-Ross combination's human annotator response labels and used for model training (EX1006, ¶¶0002-0009, 0010, 0014, 0017-0020, 0025-0032, 0035, 0049, 0063, Claim 11, Fig. 1B, 3; *See* Section VI.A.3.d (element 2[c]));
- Cox's distribution of data and labels (e.g., statistics) and other statistical information are generated from the Huval-Ross combination's human annotator response labels and used for model training (EX1006, ¶¶0002-0006, 0014, 0035, 0048; *See* Section VI.A.3.d (element 2[c])); and
- The training of the Huval-Ross model would have included comparing the predicted intents to the human labeled training data to identify prediction errors and optimize the Huval-Ross model such that the prediction errors are minimized. (EX1006, ¶¶0008, 0019, 0029-0038, 0060; Fig. 4B; *See* Section VI.A.3.e (element 2[d])).

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

149. In my opinion, a POSITA would have been motivated to combine Huval, Ross, and Cox because Cox expands on the disclosure of Huval and Ross. The Huval-Ross combination teaches the collection of human annotator responses and training of a predictive model that predicts road user intents. Section VI.A.2. Cox merely teaches implementation details regarding the training and optimization of the Huval-Ross combination's model, which would have been routine and readily implemented by a POSITA. In particular, Cox provides details as to generating additional information (psychometric data, a distribution of data and labels, end of test statistics) from human annotator responses—like those of Huval and Ross—and teaches how this information is added to training data and used to train Huval-Ross combination's predictive model. *Id.*; Section VI.A.2.; EX1006, ¶¶0002-0008, 0010, 0014, 0017-0019, 0025-0037, 0048-0049, 0063, Figs. 2B, 3, 4B. The psychometric data is formed by response times for annotations, the accuracy of annotations, and/or the presentation time—the amount of time a training image is presented to each annotator—that is obtained during human annotations. *Id.*

150. Moreover, in my opinion, Cox further teaches model training techniques to the Huval-Ross combination, including by generating predictions of training data using a model and then optimizing the model such that the differences

between the predictions and the training data is minimized. *Id.* As such, the above description demonstrates that Cox merely provides implementation details regarding how the training of the Huval-Ross combination is effectuated. In my opinion, a POSITA would thus have been motivated to combine the references and have had a reasonable expectation of success. *Id.*

151. In my opinion, a POSITA would have further been motivated to combine Huval, Ross, and Cox due to the significant overlap in their teachings. All three references relate to machine learning models within the context of autonomous vehicles. EX1004, ¶¶0016-0020; EX1005, 1:30-31, 3:39-42, 3:64-5:5; EX1006, ¶¶0056, 0060. And more generally, Huval, Ross, and Cox are all directed to machine learning technologies and involve gathering training data, training a model with the training data, and executing the model, providing that a POSITA seeking to implement the Huval-Ross combination would have looked to Cox and been motivated to combine their teachings for this additional reason. *See* Section VI.A.2.

152. In my opinion, a POSITA would have also been motivated to combine the Huval-Ross combination with Cox because the references provide explicit teachings, suggestions, and motivations to make the combination. For instance, Huval explains that its model is trained using human annotations and that its trained model “may exhibit effectiveness and accuracy . . . proportional to a volume and quality of labeled data” provided by human annotators. EX1004, ¶¶0008-0012,

0021-0022, 0036-0050, 0069, 0071-0092, 0099-0120, Figs. 1, 2A, 2B, 3-5. Ross furthers these teachings, explaining that its model is also trained based on human observations such that its autonomous vehicle “function[s] in a more ‘human-like’ or ‘polite’ way.” *Id.*; EX1005, 5:7-12, 9:13-10:10, 15:1-16:41. And like these two references, Cox notes that its system obtains human annotations and generates statistics associated with the annotations to aid in model training. EX1006, ¶¶0004, 0029-0032.

153. Cox also explains that its training provides for model predictions that “mimic human performance” and are “more consistent with the decisions of the human annotators.” *Id.* And Cox further expressly discloses that its teachings may be applied to “driverless” (i.e., autonomous) vehicle applications, just like those of Huval and Ross. EX1006, ¶¶0056, 0027. Thus, in my opinion, due to the disclosures Huval, Ross, and Cox, a POSITA would have looked to Cox’s annotation and model training techniques when considering Huval and Ross, and would have combined the references with a reasonable expectation of success.

154. In my opinion, a POSITA would have additionally been motivated to combine Huval and Ross with Cox because Cox provides benefits to the Huval-Ross combination. Cox’s teachings provide further training techniques and training data (i.e., psychometric data, generated statistics from annotation responses) applicable to the training of the Huval-Ross combination’s predictive model, as well as model

optimization techniques. EX1004, ¶¶0008-0010, 0016-0024, 0029-0039, 0043-0050, 0096, 0100, Fig. 1; EX1005, 1:30-2:53, 4:33-5:12, 9:13-10:10, 15:1-16:52; EX1006, ¶¶0019, 0029-0038, 0060; Fig. 4B. In my opinion, Cox's teachings thus would have improved the Huval-Ross combination's model by training it to provide more accurate outputs. *Id.* Indeed, the resulting trained model would have been more accurate and human-like as taught by Cox, which would have improved Huval's system by allowing it to better identify objects and their behavior in the environment of its autonomous vehicle. *Id.* This would have improved the safety of the Huval-Ross system by reducing collisions between the autonomous vehicle and objects in the environment. *Id.* The trained model as taught by Cox would have further comported with Ross's goal of autonomous vehicle control that "function[s] in a more 'human-like' or 'polite' way." *Id.*

155. In my opinion, all the reasons explained above establish that a POSITA would have been motivated to combine Huval and Ross as described. This combination would have been the predictable combination of well-known prior art elements (e.g., Cox's annotation and training details, and Huval-Ross's prediction model and annotation details, discussed above) according to known methods to yield predictable results (Cox's psychometric data is generated from the Huval-Ross combination's human annotator response labels and used for model training; Cox's distribution of data and labels (e.g., statistics) and other statistical information are

generated from the Huval-Ross combination's human annotator response labels and used for model training. The training of the Huval-Ross model would have included comparing the predicted intents to the human labeled training data to identify prediction errors and optimize the Huval-Ross model such that the prediction errors are minimized). The combination would have further used known techniques (known annotation and machine learning model training techniques as taught by Cox) to improve similar devices (Huval-Ross combination's prediction model) in the same way (using Cox's aforementioned training techniques). Moreover, because Ross and Huval each already teach training a predictive model using labeled training data to generate predictions about objects in an environment of an autonomous vehicle (*see* Section VI.A.2), it is my opinion that combining Huval and Ross with Cox's teachings about training a predictive model using labeled training data would have required minimal changes. Making this combination is thus well within the skill of a POSITA, and a POSITA would have had a reasonable expectation of success.

3. Claim 2

156. In my opinion, Huval, Ross, and Cox render obvious claim 2.

a. 2[Pre]

157. *See* Section VI.A.3.a (element 2[Pre]).

b. 2[a]

158. *See* Section VI.A.3.b (element 2[a]).

c. 2[b]

159. See Section VI.A.3.c (element 2[b]).

d. 2[c]

160. In my opinion, Huval, Ross, and Cox teach element 2[c]. The Huval-Ross-Cox combination provides that psychometric data is generated from the Huval-Ross combination's human annotator response labels regarding road user actions in image data and added to training data for Huval's model (*generating a training dataset comprising summary statistics of user responses describing the state of minds of road users displayed in the plurality of images*). See Sections VI.B.2., VI.A.3.d (element 2[c]); EX1006, ¶¶0002-0009, 0010, 0014, 0017-0020, 0025-0032, 0035, 0049, 0063, Claim 11, Fig. 1B, 3. This psychometric data summarizes statistics on how human annotators annotated training images because it includes "response times" for annotations, "accuracy" of annotations, and "presentation time" indicating how long an annotator viewed a training image, and in my opinion thereby teaches *summary statistics of user responses*. *Id.* In my opinion, a POSITA would have recognized that such a collection of data regarding annotations forms statistics. *Id.* Moreover, Cox's description of "psychometric data" is consistent with the '889 Patent's explanation. In my opinion, just like Cox, the '889 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, 6:18-22.

161. In my opinion, 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 provide a statistical representation of how human annotators annotated training images. *Id.*; EX1006, ¶¶0002-0008, 0010, 0014, 0018, 0025-0032, 0035, 0049, 0063, Fig. 3. In my opinion, this teaching applies to the Huval-Ross combination’s disclosure regarding annotator responses identifying observed road user actions in a sample training image. *Id.*; Sections VI.A.3.c (element 2[b]), VI.A.3.d (element 2[c]). Here, Cox teaches that the Huval-Ross combination generates an item response curve describing the human annotated/labeled actions corresponding to the intents of road users. *Id.* In my opinion, such intents describe the state of minds of road users in the images. *Id.* And as taught by Huval, the item response curve would be included as part of training data for the Huval-Ross model. *Id.*; EX1004, Abstract, ¶¶0008-0018, 0021-0025, 0029-0030, 0036-0050, 0085-0090, 0095-0098, 0118, Figs. 1, 3, and 4. Further, in my opinion, this disclosure further teaches *generating a training dataset comprising summary statistics of user responses describing the state of minds of road users displayed in the plurality of images. Id.*

162. And finally, in my opinion, the Huval-Ross-Cox combination further provides that a distribution of data and labels (e.g., statistics) and “end-of-test statistics” are generated from the Huval-Ross combination’s human annotator response labels regarding road user actions in image data and added to training data for Huval’s model, further teaching *generating a training dataset comprising summary statistics of user response describing the state of minds of road users displayed in the plurality of images* EX1006, ¶¶0002-0009, 0010, 0014, 0017-0020, 0025-0035, 0048-0049, 0063, Claim 11; *See* Sections VI.B.2, VI.A.3.d (element 2[c]).

e. 2[d]

163. In my opinion, Huval, Ross, and Cox teach element 2[d]. As applied to Huval and Ross, Cox teaches that the training of the Huval-Ross model would have included executing the Huval-Ross model on the training data (*see* Section VI.B.3.d (element 2[c]) during a training phase (*training, using the training dataset, a supervised learning based model*). EX1004, ¶¶0008-0011, 0016, 0021-0024, 0088-0092; EX1005, 4:47-65, 7:57-8:31, 9:13-10:10, 15:1-16:52, 17:43-18:2, Fig. 10; EX1006, ¶¶0008, 0019, 0029-0038, 0060; Fig. 4B. In the training, the Huval-Ross model would have predicted 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” for road users in image

data (*supervised learning based model configured to predict summary statistics describing a state of mind of a road user displayed in an input image*). *Id.* And as taught by Cox, the training would have included comparing the predicted intents to the human labeled training data to identify prediction errors and optimize the Huval-Ross model. *Id.*

164. Specifically, it is my opinion that Cox teaches to Huval and Ross the training of a machine learning model using psychometric data from the human annotators, including through the use of a “loss function.” EX1006, ¶¶0031-0032, Fig. 4B; Section VI.B.2. The loss function penalizes misclassification by the model in relation to human annotated training data and thus reduces model error. EX1006, ¶¶0032, 0034, 0060. Using the loss function, Cox teaches to Huval and Ross model training (*training, using the training dataset, a supervised learning based model*) by predicting classifications of training data and optimizing the model such that the difference between the predicted classifications output by the model and the training data is minimized (*supervised learning based model configured to predict summary statistics describing a state of mind of a road user displayed in an input image*). EX1006, ¶¶0008, 0019, 0029-0038, 0060; Fig. 4B. It is my opinion that this teaching is consistent with the ’889 Patent, which states that “trained” means, “in some embodiments,” that “the difference between the summary statistics output by

the neural network and the summary statistics calculated from the responses of the human observers [] is minimized.” EX1001, 12:20-25.

f. 2[e]

165. See Section VI.A.3.f (element 2[e]).

g. 2[f]

166. In my opinion, Huval, Ross, and Cox teach element 2[f]. As combined, the Huval-Ross trained model would have operated on new images to predict the actions and possible intents of vehicles, bicyclists, and pedestrians in the images, teaching element 2[f] for the same reasons discussed in Section VI.A.3.g (element 2[f]).

h. 2[g]

167. See Section VI.A.3.h (element 2[g]).

4. Claim 10

168. In my opinion, Huval, Ross, and Cox render obvious claim 10. Cox teaches to the Huval-Ross combination that “sets of queries are posed to a crowd” of human annotators on a display module and responses are collected via the display module, processed, and analyzed (*wherein the user response is determined based on: an explicit response to a question asked via a user interface*) to train the Huval model. EX1006, ¶¶0016-0018, 0032, 0048; Section VI.B.2.

169. In my opinion, The Huval-Ross-Cox combination further provides that psychometric data is generated from human annotator response labels regarding road

user actions in image data and added to training data for Huval's model. *See* Sections VI.B.2, VI.B.3.d (element 2[c]); EX1006, ¶¶0002-0009, 0010, 0014, 0017-0020, 0025-0032, 0035, 0049, 0063, Claim 11, Figs. 1B, 3. The psychometric data includes “response times” for annotations, “accuracy” of annotations, and “presentation time” indicating how long an annotator viewed a training images (*an implicit data representing one or more of: time taken to respond . . .*). *Id.*

5. Claims 1, 3-9, and 11-19

170. In my opinion, Huval, Ross, and Cox render obvious claims 1, 3-9, and 11-19 for the same reasons discussed in Sections VI.A.22, VI.A.4-VI.A.10, and VI.A.12-VI.A.20, respectively.

6. Claim 20

171. In my opinion, Huval, Ross, and Cox render obvious claim 20 for the same reasons as discussed in Section VI.B.4.

C. Ground 3: Claims 1-20 are rendered obvious over Huval, Ross, and Munro

172. In my opinion, the Huval-Ross combination renders obvious the Challenged Claims. However, Huval and Ross are further combined with Munro regarding the following limitation in Ground 3: *generating a training dataset comprising summary statistics of user responses describing the state of minds of road users displayed in the plurality of images*. As discussed below, this limitation

is explicitly taught by Munro, and it would have been obvious to combine Munro's relevant teachings with Huval and Ross.

1. Summary of Prior Art

a. Munro–EX1009

173. 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. EX1009, Abstract, ¶¶0005, 0042-0048, 0132-0142. 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.*, ¶¶0140-0142. Munro's machine learning model is then trained based on the aggregated responses. EX1009, ¶¶0140-0145.

174. In my opinion, Munro is analogous art to the '889 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., navigating an autonomous vehicle in an environment including moving objects). EX1009, Abstract, ¶¶0005, 0042-0048, 0132-0146; EX1001, 1:19-21, 1:57-60, 5:46-53; 6:8-48.

2. Motivation to Combine Huval, Ross, and Munro

175. In my opinion, A POSITA would have been motivated to combine Huval, Ross, and Munro such that Huval-Ross's received human annotator response

labels for sample images would have been analyzed to generate tallies of the number of annotators who classified the behavior or activity as having various labels. EX1009, Abstract, ¶¶0005, 0042-0048, 0132-0142 ; *see* Sections VI.A.2; VI.A.3.d (element 2[c]). In this configuration, the tallies would have been part of Huval's model training data. *Id.* A POSITA would have been motivated to combine Huval, Ross, and Munro in this manner for several reasons.

176. In my opinion, a POSITA would have been motivated to combine Huval, Ross, and Munro because Munro simply provides implementation details to the Huval-Ross combination. Specifically, Munro merely expands on the Huval-Ross combination by explaining that the human annotations received by Huval-Ross are tallied to determine the number of annotators that labeled training images in particular ways. EX1004, ¶¶0008-0020, 0016-0024, 0029-0039, 0036-0050, 0085-0090, 0095-0100, 0118, Fig. 1; EX1005, 1:30-2:53, 4:33-5:11, 9:13-10:10, 15:1-16:52; EX1009, Abstract, ¶¶0005, 0042-0048, 0132-0142. A POSITA would have thus recognized that Munro's teachings provide nothing more than a routine and well-known method of organization to the Huval-Ross annotations, and the combination thus has a reasonable expectation of success. *Id.*

177. In my opinion, a POSITA would have further combined the references due to the overlap in their teachings. Huval, Ross, and Munro are all directed to machine learning technologies and involve gathering human annotator responses,

providing that a POSITA seeking to implement the Huval-Ross combination's teachings would have looked to Munro.

178. In my opinion, a POSITA would have further been motivated to combine Huval, Ross, and Munro because Munro would provide performance benefits to the Huval and Ross combination. Section VI.A.2. For example, Munro's tallying would provide additional data regarding annotator responses that could be used to better train the Huval-Ross prediction model. *Id.*; EX1004, ¶¶0008-0020, 0016-0024, 0029-0039, 0036-0050, 0085-0090, 0095-0100, 0118, Fig. 1; EX1005, 1:30-2:53, 4:33-5:11, 9:13-10:10, 15:1-16:52; EX1009, ¶¶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.* And, training the Huval-Ross predictive model using the implementation details taught by Munro would have ensured that the model iteratively improves in accuracy. *Id.*

179. All the reasons explained above establish that a POSITA would have been motivated to combine Huval, Ross, and Munro as described. This combination would have been the predictable combination of well-known prior art elements (e.g., Huval-Ross's human annotator labels regarding road user behavior and intent; Munro's annotation organization techniques) according to known methods to yield predictable results (Huval-Ross's received human annotator response labels for

sample images would have been analyzed to generate tallies of the number of annotators who classified the behavior or activity as having various labels, as taught by Munro). The combination would have further used known techniques (known annotation organization techniques taught by Munro) to improve similar devices (Huval and Ross combination's prediction system that obtains human annotator labels) in the same way (using Munro's aforementioned techniques). In my opinion, a POSITA would have had a reasonable expectation of success in combining Munro with Huval and Ross because Munro's teachings simply provide details about how the Huval-Ross annotator responses would be organized and used.

3. Claim 2

180. In my opinion, Huval, Ross, and Munro render obvious claim 2.

a. 2[Pre]-2[b]

181. In my opinion, Huval, Ross, and Munro teach elements 2[Pre] through 2[b] for the same reasons discussed in Sections VI.A.3.a through VI.A.3.c.

b. 2[c]

182. In my opinion, Huval, Ross, and Munro teach element 2[c]. Huval and Ross teach received human annotator response labels for sample images including observed road user actions. *See* Section VI.A.3.d (element 2[c]). Huval, Ross, and Munro teach that these response labels would have been used to generate tallies of the number of annotators who classified the behavior or activity in the images as having various labels in their responses, and the tallies would have been part of

Huval's model training data (*generating a training dataset comprising summary statistics of user responses describing the state of minds of road users displayed in the plurality of images*). *Id.*; EX1009, Abstract, ¶¶0005, 0042-0048, 0132-0142. A POSITA would have recognized that these tallies are *summary statistics* as claimed because they are a collection of numerical measurements summarizing how annotators classified behavior or activity. In my opinion, this combined teaching of Huval, Ross, and Munro aligns with the '889 Patent's disclosure. Specifically, the '889 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 human observers" classified the stimulus as displaying various content. EX1001, 6:8-18

c. 2[d]

183. In my opinion, Huval, Ross, and Munro teach element 2[d]. As combined, the training of the Huval-Ross model would have been performed as discussed in Section VI.A.3.e (element 2[d]), except the training data includes the generated summary statistics as taught by Munro, and Huval's model is trained on such training data in the same manner discussed in Section VI.A.3.e (element 2[d]). Section VI.A.3.e, VI.C.3.b.

d. 2[e]

184. In my opinion, Huval, Ross, and Munro teach element 2[e] for the same reasons discussed in Sections VI.A.3.f (element 2[e]).

e. 2[f]

185. In my opinion, Huval, Ross, and Munro teach element 2[f]. As combined, the Huval-Ross trained model would have operated on new images to determine the actions and possible intents of vehicles, bicyclists, and pedestrians in the images, teaching element 2[f] for the same reasons as discussed in Section VI.A.3.g (element 2[f]). Sections VI.A.3.g, VI.C.3.b-VI.C.3.d.

f. 2[g]

186. In my opinion, Huval, Ross, and Munro teach element 2[g] for the same reasons discussed in Sections VI.A.3.h (element 2[g]).

4. Claim 12

187. *See* Sections VI.C.3 (claim 2), VI.A.13 (claim 12), and VI.A.3 (claim 2).

5. Claim 1

188. *See* Sections VI.C.3 (claim 2), VI.A.22 (claim 1), VI.A.13 (claim 12), and VI.A.3 (claim 2).

6. Claims 3-11 and 13-20

189. In my opinion, Huval, Ross, and Munro render obvious claims 3-11, 13-20 for the same reasons discussed with respect to Huval and Ross in Section VI.A above.

D. Ground 4: Claims 1-20 and rendered obvious under 35 U.S.C. § 103 over Huval, Ross, Cox, and Munro

190. In my opinion, the Huval-Ross-Cox combination renders obvious the Challenged Claims. However, Huval, Ross, and Cox are further combined with Munro regarding the following limitation in Ground 4: *generating a training dataset comprising summary statistics of user responses describing the state of minds of road users displayed in the plurality of images*. As discussed below, this limitation is explicitly taught by Munro, and it would have been obvious to combine Munro's relevant teachings with Huval, Ross, and Cox.

1. Motivation to Combine Huval, Ross, Cox, and Munro

191. In my opinion, A POSITA would have been motivated to combine Huval, Ross, Cox and Munro for the same reasons discussed in Sections VI.B.2 and VI.C.2. Sections VI.B.2, VI.C.2. The same motivations that would apply to add Munro to Huval and Ross apply equally to the combination of Huval, Ross, and Cox. *Id.* Similarly, the same motivations that would apply to add Cox to Huval and Ross apply equally to the combination of Huval, Ross, and Munro. *Id.* There is no inconsistency with adding the Munro's annotator tallying and training teachings, and

adding Cox's training data, statistical, and model training/optimization teachings, to the Huval and Ross combination. *Id.* Moreover, combining Munro and Cox together with Huval and Ross would have had no impact on a POSITA's reasonable expectation of success of combining Munro and Cox separately with Huval and Ross. *Id.*

2. Claim 2

192. In my opinion, Huval, Ross, Cox, and Munro render obvious claim 2.

a. 2[Pre]-2[b]

193. In my opinion, Huval, Ross, Cox, and Munro teach elements 2[Pre] through 2[b] for the same reasons discussed in Sections VI.B.3.a through VI.B.3.c.

b. 2[c]

194. In my opinion, Huval, Ross, Cox, and Munro teach element 2[c]. The combination teaches that the generated training dataset of Huval would have included statistical data of annotator responses from human annotators regarding road user actions in images, including tallies of the number of annotators who classified road user actions as having various labels in their responses with respect to various images, as taught by Munro (*generating a training dataset comprising summary statistics of user responses describing the state of minds of road users displayed in the plurality of images*). See VI.B.3.d (element 2[c]), VI.C.3.b. A POSITA would have been motivated to combine Huval, Ross, Cox, and Munro for the reasons discussed in Section VI.D.1.

c. 2[d]

195. In my opinion, Huval, Ross, Cox, and Munro teach element 2[d]. As combined, the training of the Huval-Ross model would have been performed as discussed in Section VI.A.3.e (element 2[d] – Ground 1) and Section VI.B.3.e (element 2[d] – Ground 2), except the training data includes the generated summary statistics as taught by Munro, and Huval’s model is trained on such training data in the same manner discussed in Section VI.A.3.e (element 2[d] – Ground 1) and Section VI.B.3.e (element 2[d] – Ground 2), VI.A.3.e, VI.B.3.e, VI.C.3.b, VI.C.3.c.

d. 2[e]

196. In my opinion, Huval, Ross, Cox, and Munro teach element 2[e] for the same reasons discussed in Sections VI.B.3.f (element 2[e]).

e. 2[f]

197. In my opinion, Huval, Ross, Cox, and Munro teach element 2[f]. As combined, Huval’s model would have operated on new images to determine the actions and possible intents of vehicles, bicyclists, and pedestrians in the images. *See* Sections VI.B.3.g (element 2[f]), VI.C.3.e, VI.D.2.b-VI.D.2.d.

f. 2[g]

198. In my opinion, Huval, Ross, Cox, and Munro teach element 2[g] for the same reasons discussed in Sections VI.B.3.h (element 2[g]).

3. Claims 10 and 20

199. Huval, Ross, Cox, and Munro render obvious claims 10 and 20 for the same reasons discussed for claims 10 and 20 in Ground 2. *Supra* Sections VI.B.4, VI.B.6.

4. Claim 12

200. *See* Section VI.D.2 (claim 2), VI.A.13 (claim 12), and VI.B.3 (claim 2).

5. Claim 1

201. *See* Section VI.D.2 (claim 2), VI.A.13 (claim 12), VI.A.22 (claim 1), and VI.B.3 (claim 2).

6. Claims 3-9, 11, and 13-19

202. In my opinion, Huval, Ross, Cox, and Munro render obvious claims 3-9, 11, and 13-19 for the same reasons discussed with respect to Huval and Ross in Section VI.A above.

E. Printed Matter

203. 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.

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

- 1[Pre], 2[Pre], 12[Pre] – “a predicted state of mind of road users in a scene captured by a camera of the autonomous vehicle”
- 1[a], 2[a], 12[a] – “a plurality of images displaying road scenes”
- 1[b], 2[b], 12[b] – “a plurality of user responses, each user response describing a state of mind of a road user displayed in the one or more images”
- 1[c], 2[c], 12[c] – “summary statistics of user responses describing the state of minds of road users displayed in the plurality of images”
- 1[d], 2[d], 12[d] – “training dataset” and “summary statistics describing a state of mind of a road user displayed in an input image”
- 1[e], 2[e], 12[e] – “a new image of a scene including a road user”
- 1[f], 2[f], 12[f] – “summary statistics describing a state of mind of the road user in the new image”
- 3, 13 – “the road user in the input image [being] a pedestrian” and “the state of mind predicted by the supervised learning based model indicates whether the pedestrian is likely to perform an action, the action including one of: staying in place, or crossing a street”
- 4, 14 – “the road user in the input image [being] a cyclist” and “the state of mind predicted by the supervised learning based model indicates whether the cyclist is likely to cross a lane”

Declaration of Dr. Jason Janét Regarding U.S. Patent No. 11,126,889

- 5, 15 – “the summary statistics comprises one or more of: a central tendency, a variance, a skew, a kurtosis, a scale, or a histogram”
- 6, 16 – “images including a road user” and “the user interface is configured to prompt the viewer to describe how the road user will act”
- 7, 17 – “stimulus comprising a set of images” and “user response describing the stimulus”
- 8, 18 – “subset of the image contained in a bounding box around the road user displayed in the image”
- 10, 20 – “an explicit response to a question asked via a user interface” and “an implicit data representing ... time taken to respond”

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

205. 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* 1[pre] and 12[pre]. Rather than impacting how the computer memory operates, the above limitations merely describe the claimed information.

VII. CONCLUSION

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

Declaration of Dr. Jason Janét Regarding U.S. Patent No. 11,126,889

* * *

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 2, 2025



Jason Janét