

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

TESLA, INC.

Petitioner,

v.

PERCEPTIVE AUTOMATA LLC

Patent Owner

IPR2025-01573

U.S. Patent No.: 10,614,344 B2

**DECLARATION OF DR. JASON JANÉT
REGARDING U.S. PATENT NO. 10,614,344**

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 '344 PATENT	20
A. Priority Date	20
B. Prosecution History	23
C. Technology Background	24
D. The Alleged Invention.....	25
IV. LEVEL OF ORDINARY SKILL IN THE ART	28
V. CLAIM CONSTRUCTION.....	28
VI. THE CHALLENGED CLAIMS ARE UNPATENTABLE.....	29
A. Ground 1: Claims 1-22 are unpatentable under 35 U.S.C. § 103 over Ross and Cox.....	29
1. Summary of Prior Art	29
a. Ross–EX1004.....	29
b. Cox–EX1005.....	32
2. Motivation to Combine Ross and Cox.....	33
3. Claim 1	37
a. 1[Pre].....	37

Declaration of Dr. Jason Janét Regarding U.S. Patent No. 10,614,344

b.	1[a]	38
c.	1[b]	41
d.	1[c]	47
e.	1[d]	50
4.	Claim 2	51
5.	Claim 3	52
6.	Claim 4	52
7.	Claim 5	53
8.	Claim 6	54
9.	Claim 7	55
10.	Claim 8	57
11.	Claim 9	58
12.	Claim 10	58
a.	10[Pre]	58
b.	10[a]	58
c.	10[b]	58
d.	10[c]	59
e.	10[d]	59
13.	Claim 11	59
14.	Claim 12	59
15.	Claim 13	59
16.	Claim 14	59
17.	Claim 15	59
18.	Claim 16	59

Declaration of Dr. Jason Janét Regarding U.S. Patent No. 10,614,344

19.	Claim 17.....	60
20.	Claim 18.....	60
21.	Claim 19.....	60
a.	19[Pre].....	60
b.	19[a].....	60
c.	19[b].....	60
d.	19[c].....	60
e.	19[d].....	60
22.	Claim 20.....	61
23.	Claim 21.....	61
24.	Claim 22.....	61
B.	Ground 2: Claims 1-22 are unpatentable under 35 U.S.C. § 103 over Ross, Cox, and Ogale	61
1.	Summary of Prior Art	61
a.	Ogale–EX1010.....	62
2.	Motivation to Combine Ross, Cox, and Ogale	62
3.	Claim 1	66
a.	1[Pre] and 1[a].....	66
b.	1[b].....	66
c.	1[c]	69
d.	1[d].....	71
4.	Claims 10 and 19.....	71
5.	Claims 2-9, 11-18, and 20-22	71
C.	Ground 3: Claims 1-5, 7-13, and 16-22 are unpatentable under 35 U.S.C. § over Zhang.....	72

Declaration of Dr. Jason Janét Regarding U.S. Patent No. 10,614,344

1.	Summary of the Prior Art	72
a.	Zhang-EX1006	72
2.	Claim 1	73
a.	1[Pre].....	73
b.	1[a]	73
c.	1[b]	76
d.	1[c]	84
e.	1[d]	86
3.	Claim 2.....	87
4.	Claim 3	87
5.	Claim 4.....	87
6.	Claim 7	88
7.	Claim 8	90
8.	Claim 9.....	90
9.	Claim 10	90
a.	10[Pre].....	90
b.	10[a]	91
c.	10[b]	91
d.	10[c]	91
e.	10[d]	91
10.	Claim 11	91
11.	Claim 12.....	91
12.	Claim 13.....	91
13.	Claim 16.....	91

Declaration of Dr. Jason Janét Regarding U.S. Patent No. 10,614,344

14.	Claim 17.....	92
15.	Claim 18.....	92
16.	Claim 19.....	92
	a. 19[Pre].....	92
	b. 19[a].....	92
	c. 19[b].....	92
	d. 19[c].....	92
	e. 19[d].....	92
17.	Claim 20.....	93
18.	Claim 21.....	93
19.	Claim 22.....	93
D.	Printed Matter.....	93
VII.	CONCLUSION	95

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. 10,614,344 (“the ’344 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 1012, incorporated here by reference, and summarized as follows: I am currently CTO for Applied Research Associates (ARA). I also hold the rank of Adjunct Associate Professor at Duke University and at North Carolina State University.

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

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

Declaration of Dr. Jason Janét Regarding U.S. Patent No. 10,614,344

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Declaration of Dr. Jason Janét Regarding U.S. Patent No. 10,614,344

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

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

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

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

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

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

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

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

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

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

Declaration of Dr. Jason Janét Regarding U.S. Patent No. 10,614,344

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 1012, 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 '344 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 '344 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.

Declaration of Dr. Jason Janét Regarding U.S. Patent No. 10,614,344

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

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

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

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

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

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

Declaration of Dr. Jason Janét Regarding U.S. Patent No. 10,614,344

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 negated 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 '344 Patent would have been considered obvious as of the priority date of the patent.

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

- 39. The scope and content of the prior art;
- 40. The differences between the prior art and the claims at issue;
- 41. The knowledge of a person of ordinary skill in the pertinent art; and
- 42. Whatever objective factors indicating obviousness or non-obviousness may be present in any particular case.

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

Declaration of Dr. Jason Janét Regarding U.S. Patent No. 10,614,344

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

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

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

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

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

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

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

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

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

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

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

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

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

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

III. BACKGROUND TO THE '344 PATENT

A. Priority Date

56. I understand the '344 Patent was filed on July 16, 2019, as a continuation of the '687 Patent, which was filed on December 4, 2017 and claims priority to the '771 Provisional filed July 5, 2017. EX1001, Cover Page.¹

57. I understand that no presumption exists that the claims of the '344 Patent are entitled to the provisional filing date. Further, I understand that Patent Owner bears the burden to establish that the Challenged Claims of the '344 Patent are entitled to the benefit of priority to the provisional application by demonstrating that the provisional provides the full support required by 35 U.S.C § 112(a) for each claim. I understand that a claim is only entitled to the benefit of the filing date of an earlier filed provisional application if the claim's subject matter is adequately supported by the earlier filed provisional application as provided by 35 U.S.C. §112. I understand that, to comply with 35 U.S.C. §112(a), the specification "must describe the invention sufficiently to convey to a person of skill in the art that the patentee had possession of the claimed invention at the time of the application, i.e., that the

¹ I understand that, because the application leading to the '344 Patent was filed after March 16, 2013, patentability is governed by the America Invents Act ("AIA") version of 35 U.S.C. §§102 and 103. Pub L. No. 112-29, 125 Stat. 284 (2011).

patentee invented what is claimed.” *Lizardtech, Inc. v. Earth Resource Mapping, Inc.*, 424 F.3d 1336, 1345 (Fed. Cir. 2005).

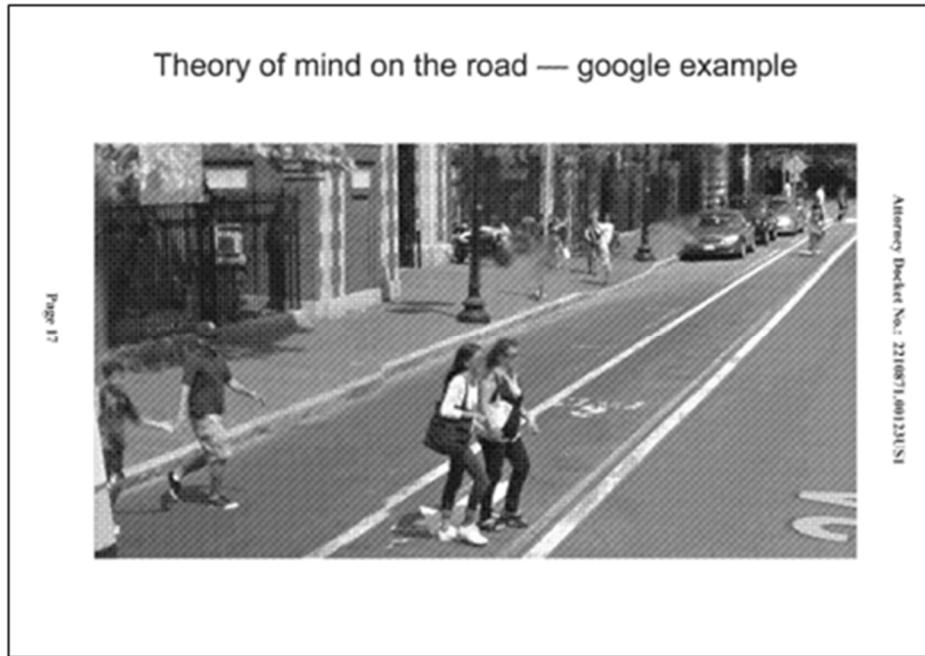
58. In my opinion, no claims of the '344 Patent are entitled to a priority date any earlier than December 4, 2017, because the '771 Provisional does not provide written description support for at least the following claim limitations of claim 1 (and similarly in claims 10 and 19):

- “*the supervised learning based model configured to receive an input sensor data displaying a particular object on a road and predicting an output statistical summary characterizing a distribution of user responses expected to be received responsive to presenting the sensor data to a plurality of users, the user responses associated with the particular object,*”
- “*executing the trained supervised learning based model to generate a statistical summary data characterizing a distribution of user responses expected to be received responsive to presenting the sensor data captured by the sensor installed on the autonomous vehicle to users,*” and
- “*controlling the operation of the autonomous vehicle on the road based on the generated statistical summary data,*”

59. The '771 Provisional includes only PowerPoint slides—the majority of which only show low resolution apparent screenshots of a video without any

Declaration of Dr. Jason Janét Regarding U.S. Patent No. 10,614,344

descriptive content—that are devoid of any reference to at least these claim recitations. *See* EX1007, 9-38. Two exemplary slides showing the lack of detail in the provisional are below.



EX1007, 25



EX1007, 28

60. Notably, the slides in the provisional are all conceptual. None describe the claimed “*supervised learning based model*,” much less the specific claimed features such as “*predicting an output statistical summary characterizing a distribution of user responses*.” Absent any description of these features, conceptual slides do not show that the patentee was in possession of the claimed inventions. Accordingly, in my opinion, the ’771 Provisional does not support any of the Challenged Claims.

B. Prosecution History

61. The ’344 Patent issued from U.S. Patent Application 16/512,560, filed July 16, 2019. EX1001, Cover Page. The Office issued a first non-final office action

that included only a non-statutory double patenting rejection over the application which later issued into the '687 Patent, the parent of the '344 Patent, and noted that the pending claims would be allowable provided this rejection is overcome. EX1002, 91-97. In response, the applicant filed a terminal disclaimer and the Office issued a notice of allowance without identifying any particular claim recitations that rendered the claims allowable. *Id.*, 114, 133-139.² But, the Challenged Claims should not have been allowed and should not have issued.

C. Technology Background

62. In my opinion, the '344 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.

² During prosecution of the parent '687 Patent, which was noted in the double patenting rejection, the Office simply stated that almost all of the recitations of that patent’s claim 1 rendered the claim allowable. EX1008, 102-103.

63. For example, it is my opinion that, 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.

D. The Alleged Invention

64. The alleged invention of the '344 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. *See, e.g.*, EX1001, Abstract, 6:1-7:39. But in my opinion, this invention was well-known, as demonstrated by the unpatentability grounds. Section VI; *see also, e.g.*, EX1010, 13:30-14:9 (explaining how a machine learning model trained by human labeled training data “predicts how a human would have labeled the [new] input sensor data”). Indeed, the Cox reference, cited by the grounds herein—which, also lists Samuel Anthony as an inventor—teaches many aspects of the alleged invention, including to ensure that machine learning model predictions are “more consistent with the decisions of the human annotators” and “better mimic human performance.” EX1005, ¶¶0004, 0032.

Neither the patentee nor Mr. Anthony brought the Cox reference to the patent office's attention during prosecution of the '344 Patent. *See* EX1001; EX1002.

65. The specification of the '344 Patent further confirms that these are not new concepts. The patent explains how model training data can be obtained and labeled by human annotators, but, in my opinion, this concept was well known. EX1001, 4:22-47, 5:8-27, 5:39-64, Figs. 1, 2B. 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, 6:1-11. In my opinion, these complex-sounding terms refer to simple operations, such as the tallying of the various annotator responses. *Id.* Such tallying of annotator responses was well known. EX1009, ¶0140 (tallying how many annotators labeled a training sample in a certain way). And in another example, the patent explains that its "summary statistics" relate to how long a human annotator took to label training data or the position of the annotator's eyes when labeling training data. EX1001, 6:11-15. Yet again, this was well known in the art and something that the inventor's own Cox reference demonstrates. *See, e.g.,* EX1005, ¶¶0006-0008, 0018, 0063.

66. The '344 Patent also says that after a machine learning model is trained using the responses, the model is used to make a prediction based on "live" input data for operation of an autonomous vehicle. EX1001, 7:5-39. For example, the

predictions reflect “predicted summary statistics for the ‘real world’ or ‘live data’ image” where the “predicted summary statistics are the model's best estimation of what the summary statistics would be on the image if the image had human annotations collected.” *Id.*, 12:44-49, Fig. 9. In my opinion, such predictions were also well known. For example, the Ogale reference explains that a machine learning model trained by human labeled training data provided by human annotators “predicts how a human would have labeled the [new] input sensor data.” EX1010, 13:30-14:9.

67. And, in my opinion, the '344 Patent's application of its known concepts to autonomous vehicles does not save it. EX1001, 1:53-56, 3:44-46. Controlling an autonomous vehicle based on the predictions of a machine learning model was well known as demonstrated by at least Ross and Zhang. EX1004, 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; EX1006, ¶¶0054, 0057-0061, Fig. 8. For example, the Ross reference explains how an “intent model” trained using human observation training data is used so that an autonomous vehicle can determine how vehicles, bicyclists, and pedestrians in the environment around the autonomous vehicle will act—for example, whether a vehicle wishes to changes lanes or a pedestrian wishes to cross a road. EX1004, 1:30-2:53, 4:33-5:11, 9:13-10:10, 15:1-16:52. Zhang similarly discloses “predicting the behavior and intent of road users” proximate to an autonomous vehicle using “well-known” neural

networks that are trained using human labeled training data that “associates a label having behavior and direction information with each instance of a vehicle in the training data.” EX1006, ¶0049-0052, Abstract.

68. Thus, in my opinion, the '344 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

69. A person having ordinary skill in the art (“POSITA”) relevant to the '344 Patent as of July 5, 2017 would have had at least: (1) a bachelor’s degree in electrical engineering, computer engineering, computer science, or equivalent course work, with three years of work experience in computer vision, autonomous vehicles, and/or machine learning; or (2) a master’s degree in electrical engineering, computer engineering, computer science, or equivalent course work, with a focus in computer vision, autonomous vehicles, and/or machine learning.

V. CLAIM CONSTRUCTION

70. I understand 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, Petitioners do not present any constructions

because there is no need for claim construction to determine that the cited prior art renders all Challenged Claims obvious.

VI. THE CHALLENGED CLAIMS ARE UNPATENTABLE

A. Ground 1: Claims 1-22 are unpatentable under 35 U.S.C. § 103 over Ross and Cox.

71. In my opinion, Ross alone renders obvious most limitations of the Challenged Claims, and is further combined with Cox to render obvious the remaining limitations, which relate to a *supervised learning model* and predicting outputs that reflect “*user responses expected to be received.*” As discussed below, these limitations are explicitly taught by Cox, and it would have been obvious to combine Cox’s teachings with Ross.

1. Summary of Prior Art

a. Ross—EX1004

72. Ross—a patent issued to Waymo LLC—discloses techniques for “maneuvering” an autonomous vehicle. EX1004, 1:30-31, 3:39-42, 3:64-5:5. The goal of Ross is to provide an autonomous vehicle control system that provides “[r]easoning about the intents of objects, such as other road users” (e.g., other vehicles, pedestrians, bicyclists) and allow the autonomous vehicle to “function in a more ‘human-like’ or ‘polite’ way.” *Id.*, 5:6-11, 15:1-16:52.

73. 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. EX1004, 1:30-2:53, 4:33-5:11, 9:13-10:10, 15:1-16:52. Indeed, “human operators” can manually generate a “list of predetermined actions” based on their “observations of the actions of other road users (by such operators or from sensor data from perception systems of one or more autonomous vehicles such as vehicles 100 or 100A),” and a human observed action of the list that is performed by a road user is “marked as the correct intent” during training. *Id.*, 9:41-10:10.

74. The trained intent model is then used by the autonomous vehicle, where cameras of the vehicle capture “raw” camera data of the environment around the vehicle displaying objects in the environment such as other vehicles, pedestrians, and bicyclists. EX1004, 7:57-8:10, 9:13-10:10, 15:1-16:52. The trained model is executed on the camera data to predict a “set of possible intents or hypotheses identifying possible next actions (and predicted trajectories), a given point in time when the action is likely to occur, and associated likelihood values” with respect to each object. *Id.*, 3:64-4:46, 7:57-8:10, 9:13-10:10, 15:1-16:52 (parenthetical in original). The set of likelihood values, with each value associated with a different prediction, “indicate[s] which of the predictions are more likely to occur (relative to one another),” where “the prediction with the greatest likelihood value may be most likely to occur whereas predictions with lower likelihood values may be less likely to occur.” EX1004, 9:13-10:10. In my opinion, this “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” is thus statistical summary data showing the various road user intents and their associated likelihoods, and as the model was trained using human influenced training data, represents what would have been the human operator’s observations of road user intents to the input raw camera data. *Id.*, 9:13-10:10, 15:1-16:52. The autonomous vehicle of Ross is controlled based on the predictions. EX1004, 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.

75. Ross is thus just like the ’344 Patent, which purports to provide “a set of predictions of the state of mind of road users” and “judgments about the future behavior of road users based on their state of mind.” EX1001, 7:33-39. The patent explains that “state of mind” can refer, for example, to “intention” of a road user, as well as other characteristics of “internal mental state.” *Id.*, 9:33-39.

76. In my opinion, Ross is analogous art to the ’344 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., accounting for moving objects when navigating an autonomous vehicle in an environment). EX1004, Abstract, 3:64-5:1; 9:27-47; EX1001, 1:15-17, 1:53-56.

b. Cox–EX1005

77. 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 ’344 Patent. EX1001, Cover; EX1005, Cover.

78. Cox is directed toward “machine learning augmented by measurements acquired from human populations.” EX1005, Abstract, ¶¶0001-0004. Cox teaches crowd-sourcing training data used to train machine learning algorithms, such as supervised machine learning models. EX1005, Abstract, ¶¶0001-0004, 0015-0016, 0025-0034, 0044, 0060. Just like the ’344 Patent, Cox teaches compiling an n-amount of images (training objects), transmitting the images to a plurality of human annotators (human labelers), compiling statistical information regarding the distribution of human responses (including the actual responses and psychometric data associated with the responses), and training a supervised learning model using gathered information “consistent with the decisions of the human annotators 160.” *Cf* EX1005, ¶¶0029-0032, Figs. 1A-B, 3 *with* EX1001, 5:39-46; 6:1-42; Fig. 2A. The goal of Cox’s machine learning techniques is to “better mimic human performance.” EX1005, ¶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 / semiautonomous automobile[]” applications. EX1005, ¶¶ 0056, 0060.

79. In my opinion, Cox is analogous art to the '344 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., accounting for moving objects when navigating an autonomous vehicle in an environment), describing the gathering training data from a population of number of human labelers, and training a model using the training data, and executing it to output predictions. EX1005, ¶¶0029-0032, Figs. 1A-B, 3; EX1001, 1:15-17, 1:53-56, 5:39-46; 6:1-42; Fig. 2A.

2. Motivation to Combine Ross and Cox

80. In my opinion, a POSITA would have been motivated to combine Ross and Cox such that Ross's intent model is a supervised learning model that is trained with human annotated image data, as taught by Cox (providing the "Ross-Cox model"). EX1004, 4:33-65, 9:27-10:10, 15:1-16:22; EX1005, ¶¶0007-0010, 0015-0019, 0025-0035, 0051, 0060. Based on this training approach, a POSITA would have further configured the Ross-Cox model to generate predictions for objects shown in camera data that are "more consistent with the decisions of the human annotators" who annotated the training data. EX1004, 1:55-2:13, 4:33-46, 9:13-10:10, 15:1-21; EX1005, ¶¶0004, 0007-0010, 0015-0019, 0025-0035, 0051, 0060. A POSITA would have been motivated to combine Ross and Cox in this way for several reasons.

81. First, Cox simply expands on Ross's disclosure of a predictive intent model for an autonomous vehicle that is trained using human observations. EX1004, 9:13-10:10; Section VI.A.1.a. Ross discloses generating training data, using the training data to train an intent model, and executing the model on input raw camera data from cameras of an autonomous vehicle to predict trajectories for vehicles, pedestrians, and bicyclists around the autonomous vehicle. EX1004, 3:39-51; 3:64-5:11, 7:57-8:10, 9:13-10:10, 13:29-33, 15:1-16:52. Cox expands on this by identifying such a model as a supervised learning model, and teaching that the model is trained and optimized using human observations, so that model predictions better reflect human decision making. EX1005, ¶¶0007-0010, 0015-0019, 0025-0038, 0051, 0060, Fig. 4B. The goal of Cox's machine learning techniques is to "better mimic human performance," and Cox explains that its machine learning "may be applied to any form of supervised learning, including neural networks, boosting, bagging, random forests, nearest neighbor algorithms, naive bays classifiers, density estimators, and other forms of statistical regression" and "may also be applied as part of a supervised component of semi-supervised or deep-learning algorithms." EX1005, ¶¶0004, 0060; *see also id.*, ¶¶0008-0010, 0015-0019, 0025-0035, 0051. Cox thus provides a natural expansion and improvement on Ross's base disclosure.

82. In my opinion, a POSITA seeking to implement Ross's system would also have looked to Cox to provide more instructions for training a machine learning

model to predict human behavior. Both references are directed to machine learning technologies and involve (1) gathering training data, (2) training a model with the training data, and (3) executing the model, and as noted above, both are explicitly applicable to autonomous vehicle applications. *Id.* Cox even acknowledges as much, explaining that its disclosure is relevant to many “visual recognition problems” including those related to detecting “objects” such as a “car,” as well as “driverless” vehicle applications. EX1005, ¶¶0016, 0056. Because Cox suggests that its system is relevant for such applications, Cox therefore provides an explicit teaching, suggestion, or motivation to combine with a system for controlling driverless (or autonomous) vehicles, like Ross.

83. The stated goal in Ross to provide an autonomous vehicle that “function[s] in a more ‘human-like’ or ‘polite’ way” would also have motivated a POSITA to combine Ross with Cox. EX1004, 5:7-12. Consistent with Ross’s goal, Cox provides a system for training a model “to better mimic human performance” and to generate predictions that are “more consistent with the decisions of the human annotators.” EX1005, ¶¶0004, 0029-0032. In my opinion, a POSITA would have recognized that using Cox’s teachings would help ensure that Ross’s predictive intent model is more “human-like.” *Id.*; EX1004, 5:7-12.

84. A POSITA would have also been motivated to combine the teachings of Ross and Cox because the combination would have benefitted Ross’s system.

Enhancing Ross's system with Cox's human annotated training data that considers the state of mind of training subjects, as well as Cox's associated model training and optimization processes, would have further refined Ross's prediction model so it better considers the state of mind of such road users and what they might be thinking when present in autonomous vehicle's environment. EX1005, ¶¶0007-0010, 0015-0019, 0025-0035, 0051, 0060. This would have improved the accuracy of Ross's intent predictions, and thus improved safety by helping to reduce collisions between the autonomous vehicle and human drivers or pedestrians.

85. In my opinion, all the reasons explained above establish that a POSITA would have been motivated to combine Ross and Cox as described. This combination would have been the predictable combination of well-known prior art elements (e.g., Ross's prediction model; Cox's model training and optimization techniques) according to known methods to yield predictable results (Ross's intent model being a supervised learning model; the Ross-Cox model generating predictions for objects shown in camera data that are "more consistent with the decisions of the human annotators" who annotated the training data). The combination would have further used known techniques (known machine learning model training and optimization techniques as taught by Cox) to improve similar devices (Ross's model for predicting intents) in the same way (using Cox's aforementioned techniques). Lastly, due to the overlap in Ross's and Cox's

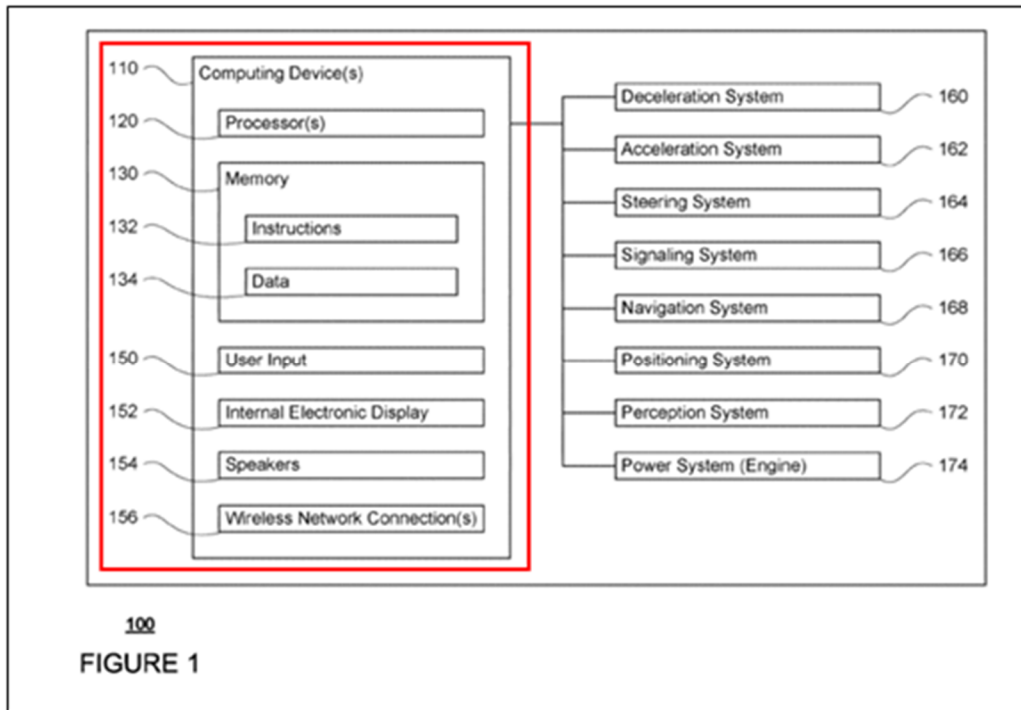
disclosures and their compatibility, making this combination would also be well within the skill of a POSITA, and a POSITA would have had a reasonable expectation of success for at least the same reasons. Indeed, in addition to the above reasons, because Ross uses human-labeled data to train its model, combining Ross with Cox's teachings about generating and using the same type of data in a supervised learning model would have required minimal changes. This would have been well within the skill of a POSITA, and a POSITA would have had a reasonable expectation of success.

3. Claim 1

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

a. 1[Pre]

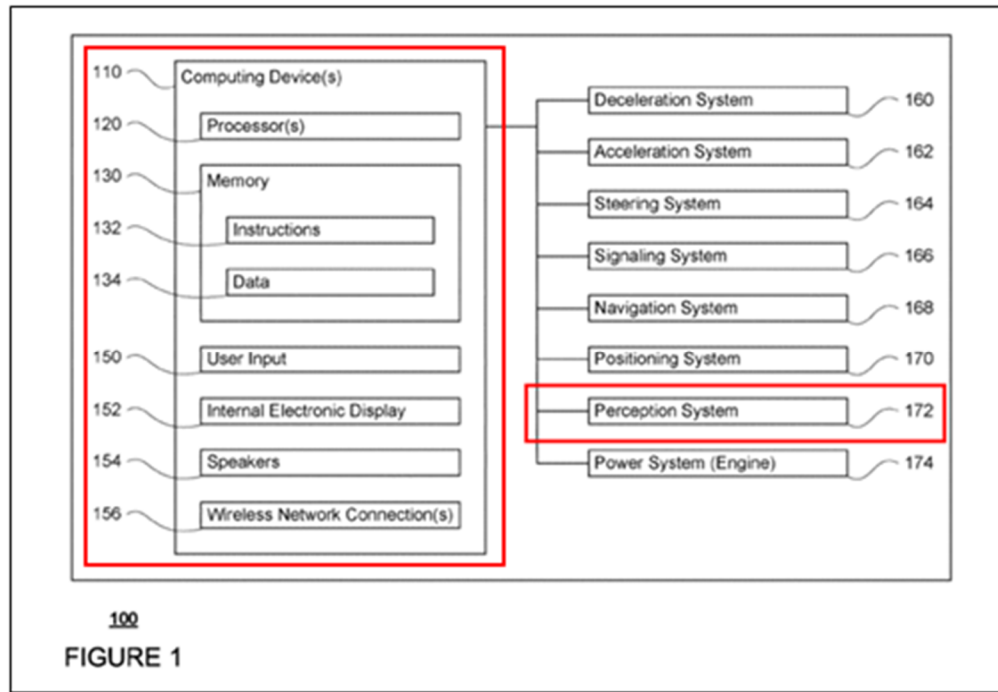
87. In my opinion, to the extent the preamble is limiting, Ross teaches it. Ross describes systems and methods for “maneuvering a vehicle” using a “computing device 110;” computing device 110 is an “autonomous driving computing system” incorporated into a vehicle 100 that controls the vehicle (*[a] computer-implemented method*). EX1004, 1:30-2:53, 3:39-42, 5:13-22, 6:33-59, Fig. 1.



EX1004, Fig. 1 (annotated)

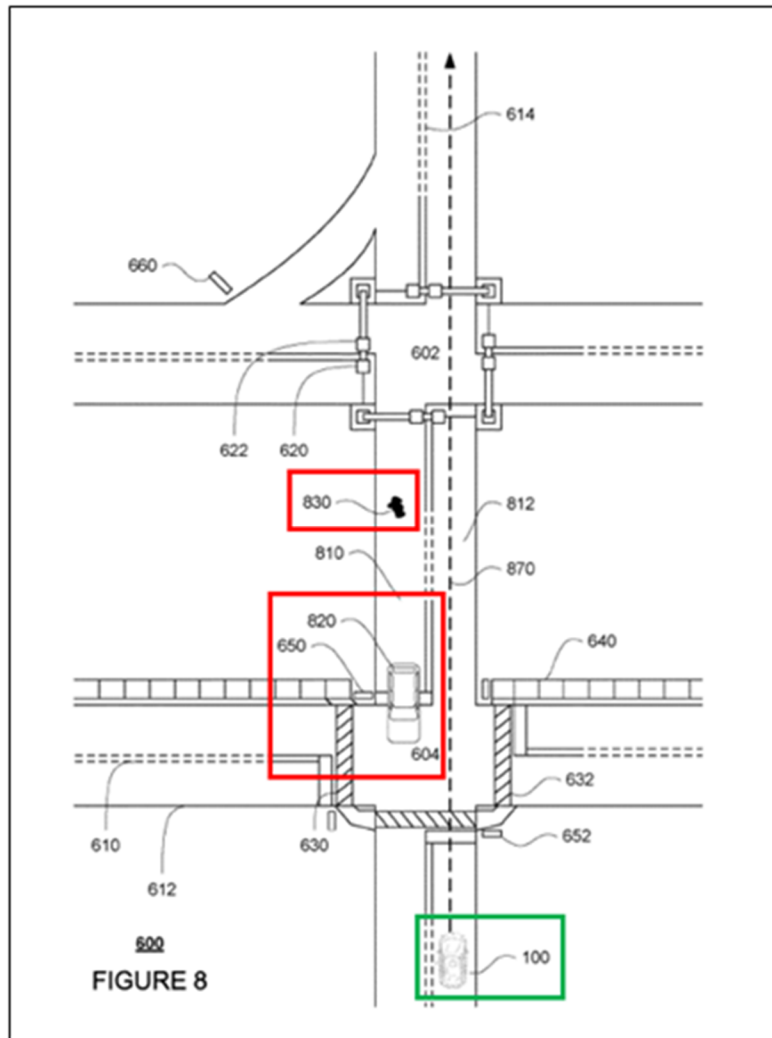
b. 1[a]

88. In my opinion, Ross teaches element 1[a] because vehicle 100 in Ross includes a “perception system 172” that includes “cameras” for “detecting objects external to the vehicle” such as other vehicles, pedestrians, and bicyclists, as well as “obstacles in the roadway, traffic signals, signs, tress, etc.” EX1004, 4:66-5:11, 7:57-8:31, 8:32-52, 9:41-65, 15:50-16:41, Claims 2 and 3, Figs. 1, 3A-3D.



EX1004, Fig. 1 (annotated)

89. As vehicle 100 travels on a road, the cameras of perception system 172 generate “raw data” that displays objects corresponding to other “road user[s]” such as other vehicles, bicycles, or pedestrians—for example, as vehicle 100 travels on a road, the data can display a vehicle 820 and pedestrian 830 in the road as shown by Figure 8. *Id.*, 3:39-47, 4:66-5:11, 7:57-8:31, 8:32-52, 9:41-65, 13:29-14:58, 15:50-16:41, Fig. 8, Claims 1, 2, and 3.



EX1004, Fig. 8 (annotated)

90. The raw camera image data captured by the cameras of perception system 172 are transmitted to computing device 110 of autonomous vehicle 100 operating on a road. EX1004, 7:57-8:31, 9:4-10:10, 13:29-15:21, Claim 1, Fig. 8. Computing device 110 of vehicle 100 operating on a road processes, and thus receives, the “raw data” captured by the cameras of vehicle 100 (*receiving, by a computing device associated with an autonomous vehicle operating on a road,*

sensor data captured by a sensor installed on the autonomous vehicle), where the raw data is camera image data that displays the various objects on the road in the vehicle's environment (e.g., other vehicles, pedestrians, bicyclists) (*the sensor data displaying an object on the road*). *Id.*

c. 1[b]

91. In my opinion, Ross and Cox teach element 1[b] as described below.

- ***supervised learning based model***

92. In my opinion, the combination of Ross and Cox teach a *supervised learning based model* as claimed. Ross's "intent model" is a prediction model trained using human observations to generate likelihoods for various road user intents. EX1004, 4:33-65, 9:27-10:10, 15:1-16:22. Cox similarly teaches a machine learning model that generates predictions and is trained with human labeled image data but specifies that its model is "**any form of supervised learning**, including neural networks" and "random forests" and various other model types. EX1005, ¶¶0027-0032, 0060 (emphasis added). Accordingly, combining the teachings of Cox with Ross provides that Ross's intent model is a supervised learning model, such as a neural network or one of Cox's other explicitly described supervised learning models, that is trained with human labeled image data (the "Ross-Cox model"), thus teaching a *supervised learning model* as claimed. *Id.*; EX1004, 4:33-65, 9:27-10:10, 15:1-16:22; Section VI.A.2.

- *providing the sensor data as input to a supervised learning based model,*

93. In my opinion, Ross describes providing the particular received “raw data” from the cameras of vehicle 100, referred to in Section VI.A.3.b (element 1[a]) (*providing the sensor data*), as the input to the Ross-Cox model (*as input to a supervised learning based model*) to “generate a set of possible intents for a given road user by looking at all the possible actions that the road user could execute.” EX1004, 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; EX1005, ¶¶0026-0032.

- *the supervised learning based model configured to receive an input sensor data displaying a particular object on a road*

94. The '344 Patent recitation *the supervised learning based model configured to receive an input sensor data displaying a particular object on a road* refers to functionality that the '344 Patent's model is configured to perform. In my opinion, this recitation may be interpreted to refer to functionality the model is configured to perform during “live” non-training use (referred to as “operational use” hereafter) as well as during training. EX1001, 11:5-12:20, 12:21-12:60, Figs. 8, 9. In either case, Ross and Cox teach this limitation. *Id.*

95. The Ross-Cox model (*the supervised learning based model*) is configured to receive input “raw data” from cameras of vehicle 100, including the specific raw camera data discussed in Section VI.A.3.b (element [1a]) as well as

other raw camera data captured at other times by the vehicle (*configured to receive an input sensor data*). EX1004, 1:30-2:53, 3:39-47, 3:64-4:5, 4:33-5:11, 7:57-8:31, 8:32-52, 9:13-10:10, 15:1-16:52, Fig. 8, Claims 1, 2, 3, and 7. The camera data pertains to objects corresponding to other road users, such as vehicles, bicycles, and pedestrians on the road (*displaying a particular object on a road*). *Id.* Input raw camera data displaying road objects is provided to the Ross-Cox model in this manner during training of the model as well as during operation. EX1004, 4:33-65, 9:27-10:10, 15:1-16:22.

- ***and predicting an output statistical summary characterizing a distribution of user response expected to be received responsive to presenting the sensor data to a plurality of users, the user responses associated with the particular object.***

96. The '344 Patent limitation *predicting an output statistical summary characterizing a distribution of user responses expected to be received responsive to presenting the sensor data to a plurality of users, the user responses associated with the particular object* again refers to functionality that the '344 Patent's model is configured to perform. In my opinion, this recitation may be interpreted to refer to functionality the model is configured to perform during operational use or during training. EX1001, 11:5-12:29, 12:21-12:60, Figs. 8, 9. In either case, Ross and Cox teach this limitation. *Id.*

Model Functionality During Operational Use

97. After receiving input raw camera data from cameras of autonomous vehicle 100 (including the specific raw camera data discussed in Section VI.A.3.b (element [1a]) as well as other raw camera data captured at other times by the vehicle), the Ross-Cox model predicts for an object (e.g., vehicle, pedestrian, bicyclist) 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” (*predicting an output statistical summary*). EX1004, 1:30-2:53, 3:64-4:5, 4:33-5:11, 7:57-8:31, 9:13-10:10, 15:1-16:52, 17:43-18:2, Fig. 10, Claims 1 and 7 (parenthetical in original). In my opinion, this model output (referred to hereinafter as the “Ross predicted data”) is *an output statistical summary* because it is a summary of the various predicted intents and likelihood values for a particular object in input raw camera image data. EX1004, 4:33-5:11, 9:13-10:10, 15:1-16:52, Fig. 10. The likelihood values summarize how likely each of the intents and trajectories of the object is to occur, based on the training data that was used to train the intent model. *Id.* Because these values are a collection of numerical elements, it is my opinion that a POSITA would have recognized that such values form statistics. *Id.* Moreover, the likelihood values indicate the relative likelihoods of predictions, where the greatest likelihood value corresponds to the prediction most likely to occur. EX1004, 9:27-40, 15:50-16:52. The predicted intents

and associated likelihood values thus reflect statistics summarizing how an object within the environment of Ross's autonomous vehicle may behave.

98. Cox teaches that the Ross predicted data should reflect *user responses expected to be received responsive to presenting the sensor data to a plurality of users*. Cox's purported novelty is using a crowdsourced, "large population" of human annotators to annotate a training image, generating "statistical information" based on the responses received from the human annotators, and training a supervised machine learning model using the statistical information. EX1005, ¶¶0007-0010, 0015-0019, 0025-0035, 0049, 0051, Figs. 1A, 1B. Annotator responses describe a state of mind of subjects displayed in training images. *Id.* The purpose of this training in Cox is to allow the model "to make predictions (based on various query objects) **that are more consistent with the decisions of the human annotators.**" *Id.*, ¶¶0004, 0027-0032 (parenthetical in original, bolding added), 0051. Thus, as applied to Ross in this regard, Cox teaches that the output of the Ross-Cox model should reflect the likely responses from human annotators, including with respect to the state of mind of subjects in analyzed images.

99. In my opinion, the Ross-Cox combination therefore teaches that the Ross predicted data (i.e., the predicted "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 an object

discussed earlier in this section) predicted by the Ross-Cox model (*output statistical summary*), is consistent with the decisions of human annotators during training and thus reflects the distribution of the various human annotator responses received from human annotators (*characterizing a distribution of user responses*). EX1004, 1:55-2:13, 4:33-46, 9:13-10:10, 15:1-21; EX1005, ¶¶0004, 0007-0010, 0015-0019, 0025-0035, 0051; Section VI.A.2. These various human annotator responses thus indicate possible next actions and state of mind of the object (e.g., vehicle, pedestrian, bicyclist) that human annotators would have identified for the object in the raw camera data and responded with when observing the raw camera data input into the Ross-Cox model (*expected to be received responsive to presenting the sensor data to a plurality of users, the user responses associated with the particular object*). *Id.*

Model Functionality During Training

100. In my opinion, Ross and Cox further teach *predicting an output statistical summary characterizing a distribution of user responses expected to be received responsive to presenting the sensor data to a plurality of users, the user responses associated with the particular object* if this recitation is interpreted to refer to functionality the model is configured to perform during training of the model.

101. In Cox, training data in the form of annotation responses and psychometric data from human annotators is utilized for training the model, “resulting in a human-weighted loss function.” EX1005, ¶¶0031-0032, Fig. 4B. The

loss function penalizes misclassification by the model in relation to human annotated training data and thus reduces model error. EX1005, ¶¶0032, 0034. Cox's model is trained 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. EX1005, ¶¶0008, 0019, 0029-0038, 0060; Fig. 4B. Thus, in my opinion, as applied to Ross, Cox teaches model training that would have included comparing Ross's predictions to the human labeled training data to identify prediction errors and optimizing the Ross-Cox model such that the errors are minimized. *Id.*; EX1004, 4:47-65, 7:57-8:31, 9:13-10:10, 15:1-16:52, 17:43-18:2, Fig. 10; Section VI.A.2. This is consistent with the '344 Patent, which discusses training as similarly calculating (and minimizing) differences between training data and model predictions. EX1001, 12:15-20.

102. In view of the above, it is my opinion that the Ross-Cox combination discloses the claimed “*predicting an output statistical summary...*” during training, just like during operational use. In order to optimize the model as taught by Cox, the same type of predictions would be made during training as during operational use. *Id.*

d. 1[c]

103. In my opinion, Ross and Cox teach element 1[c]. The particular raw camera data from the cameras of vehicle 100 (*sensor data captured by the sensor*

installed on the autonomous vehicle) is provided as the input to the trained Ross-Cox model (see Sections VI.A.3.b, VI.A.3.c (elements 1[a], 1[b])), where the model was trained via the human observations, human annotations, and model optimization as discussed for element 1[b]. *Id.* The trained model is then executed on this input data (*executing the trained supervised learning based model*) to generate 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” (*to generate a statistical summary data*), which is the Ross predicted data, for objects in the particular raw camera data. EX1004, 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.3.c (element 1[b]). Specifically, the trained model is executed on the particular raw camera data from the cameras of vehicle 100 input into the model as discussed in element 1[b] in the “Model Functionality During Operational Use” example discussed for element 1[b]. *Id.*

104. In my opinion, the Ross predicted data generated by the trained model forms *a statistical summary data* with respect to the particular raw camera data for the same reasons discussed regarding the *statistical summary* recitation in element 1[b]. See Section VI.A.3.c (element 1[b]). The Ross predicted data output in element 1[c] is data specifically tied to the live operation of the autonomous vehicle (*statistical summary data*), which would have the same general characteristics as the

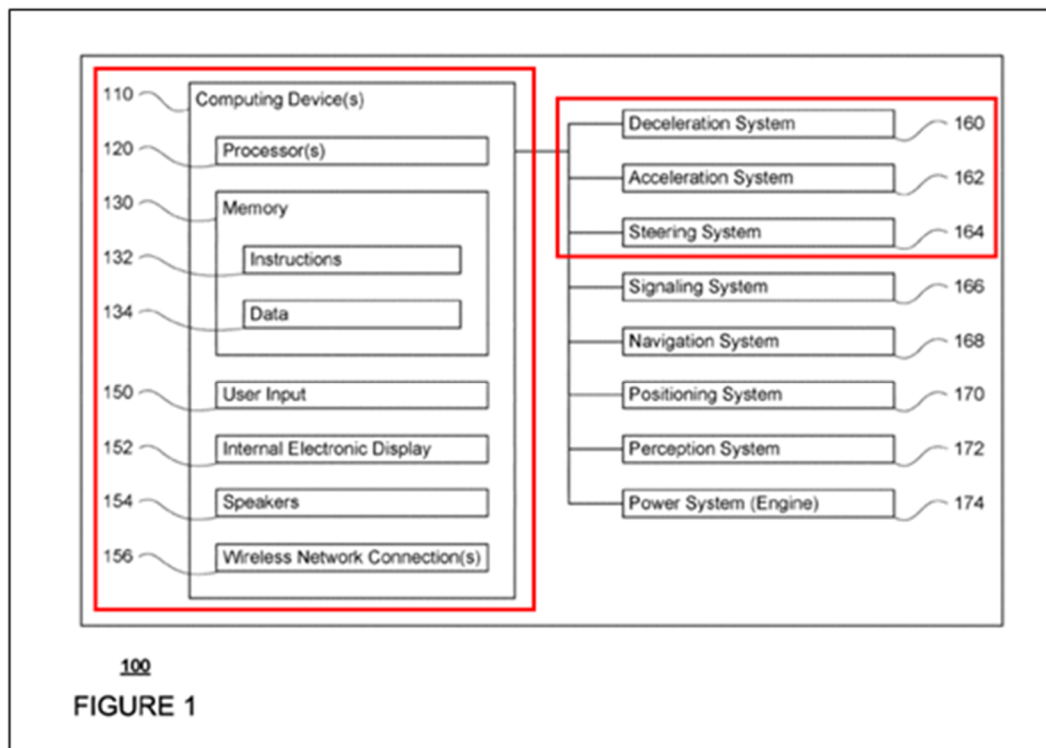
model output described generally in element 1[b] (*statistical summary*). *Id.* Moreover, the Ross predicted data for the particular raw camera data from the cameras of vehicle 100 (*statistical summary data*) is *characterizing a distribution of user responses expected to be received responsive to presenting the sensor data captured by the sensor installed on the autonomous vehicle to users* with respect to the particular raw camera data for the same reasons discussed regarding the *characterizing a distribution* recitation in element 1[b] for both Ross and Cox. *Id.*

105. The Ross predicted data for the particular raw camera data (*output statistical summary*) would have been consistent with the decisions of human annotators during training and thus reflect the distribution of the various human annotator responses received from human annotators (*characterizing a distribution of user responses*). Section VI.A.3.c (element 1[b]). This is because such human annotator responses would have trained the model and the model would have been optimized with respect to the responses, with errors between the model predictions and annotator responses minimized. *Id.* Moreover, the various human annotator responses indicate possible next actions (and thus state of mind) of the object that human annotators would have identified for the object in the particular raw camera data and responded with when observing the particular raw camera data (*expected to be received responsive to presenting the sensor data captured by the sensor installed on the autonomous vehicle to users*). EX1004, 4:47-65, 7:57-8:31, 9:13-

10:10, 15:1-16:52, 17:43-18:2, Fig. 10; EX1005, ¶¶0019, 0029-0038, 0060; Fig. 4B; Section VI.A.2; Section VI.A.3.c (element 1[b])

e. 1[d]

106. In my opinion, Ross and Cox teach element 1[d]. Autonomous vehicle 100's computing device 110 controls the vehicle's deceleration system, acceleration system, and steering system, which controls speed and direction of the vehicle. EX1004, 6:33-59, 10:11-30, Fig. 1.



EX1004, Fig. 1 (annotated)

107. Moreover, in my opinion, executing the Ross-Cox model for the particular input raw camera data 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 an object vehicle, pedestrian, or bicyclist displayed in the camera data (*the generated statistical summary data*). EX1004, 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); *see* Sections VI.A.3.c, VI.A.3.d (elements 1[b], 1[c]). One or more of the predictions with the highest likelihoods are identified and used by the autonomous vehicle’s computing device 110 to “maneuver the vehicle in order to promote the other road user’s intent.” EX1004, 15:50-16:52, 10:11-30, 4:66-5:11. Thus, Ross’s “maneuvering” or control of the vehicle 100 is based on the “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” (*the generated statistical summary data*) generated by the model because the set of intents and trajectories are analyzed to determine those with the “highest likelihoods,” which are then used to control vehicle 100 on the roadway (*controlling the operation of the autonomous vehicle on the road based on the generated statistical summary data*). *Id.*, 9:13-10:30, 15:50-16:52.

4. Claim 2

108. In my opinion, Ross and Cox render obvious claim 2. The Ross-Cox model is, for example, a neural network or one of Cox’s other described supervised learning models, “e.g., any form of supervised learning, including neural networks,

boosting, bagging, random forests, nearest neighbor algorithms, naive bays classifiers, density estimators, and other forms of statistical regression” (*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*). See Section VI.A.3.c (element 1[b]).

5. Claim 3

109. In my opinion, Ross and Cox render obvious claim 3. The camera image data displays objects on the road in the vehicle’s environment including other vehicles, pedestrians, and bicyclists (*wherein the object on the road is one of: a pedestrian, a cyclist, or a motor vehicle*). See Section VI.A.3.b (element 1[a]).

6. Claim 4

110. In my opinion, Ross and Cox render obvious claim 4. Ross describes that the various human observation responses (*user responses*) indicating possible next actions of the object (e.g., a vehicle, pedestrian, bicyclist) that would have been received responsive to humans observing the raw camera data including the object include actions associated with the object (*wherein the user responses describe an action associated with the object*). EX1004, 4:47-65, 9:41-10:10; see Sections VI.A.3.c, VI.A.3.d (elements [1b], [1c]). When the object is a vehicle, the actions include the vehicle “changing lanes” as well as “turning at intersections, make a U-

turn[], entering or exiting driveways, pulling over, [and] pulling out from a parking spot or parked position” (*the action including one of the object staying in place, changing lanes, or crossing a street*). EX1004, 4:47-65, 9:41-10:10. For an object bicyclist, the actions include “all of the actions for vehicles as well as crossing crosswalks and ‘jaywalking’ (for instance, crossing a roadway at a location other than an intersection or crosswalk).” *Id.* For an object pedestrian, the actions include “crossing a roadway” and “crossing a roadway in a crosswalk,” for example. *Id.*

7. Claim 5

111. In my opinion, Ross and Cox render obvious claim 5. Cox teaches that its training involves displaying training images to human annotators for manual annotations/labels. EX1005, ¶0017. Cox describes that “[d]uring and/or after the annotation, psychometric data is also acquired that characterizes the annotation of the training objects by the annotators.” *Id.*, ¶¶0018, 0029, Fig. 3. The psychometric data includes “response times for classifying one or more features, the accuracy of feature classification, and/or the presentation time (i.e., the amount of time presented to each annotator) of one or more training objects.” *Id.*, ¶¶0018-0029 (parenthetical in original). Cox further describes that “[e]ye-tracking hardware may be used to collect measures suitable for deep annotation such as saccade-to-target accuracy, saccade-to-target latency, number of saccade hops to target, and total number of saccades.”

112. Thus, applying Cox's teachings to Ross would have provided that the predicted outputs of the Ross-Cox model during training would have been generated according to training using human annotation/label responses, and would have thus been associated with psychometric data for at least some of the human annotation/label responses (*wherein the statistical data is associated with a parameter of the subset of the plurality of response data*). EX1005, ¶¶0017-0018, 0029; *see* Section VI.A.3.c (element 1[b]). As taught by Cox, the psychometric data would have included response times for annotations, accuracy of the annotation, as well as the amount of time a training image is presented for annotation (*the parameter including at least one of a content of a response, a time associated with entering a response*). *Id.* The psychometric data would have further included information on human annotator eye movements when providing annotation responses, which would have been measured with respect to the display screen displaying the training image that the annotator is viewing (*and a position of an eye of a human observer associated with the response, the position being measured with respect to a display associated with the user interface*). *Id.*; EX1005, ¶0063. A POSITA would have been motivated to combine Ross and Cox for the reasons discussed in Section VI.A.2.

8. Claim 6

113. In my opinion, Ross and Cox render obvious claim 6. The combination of Ross and Cox teach *the parameter* in the form of psychometric data including response times for annotations, accuracy of the annotation, as well as the amount of time a training image is presented for annotation. Section VI.A.7 (claim 5). Cox further teaches that collected “accuracy and reaction time may be measured on a per-image, per-subject, or per-condition level.” EX1005, ¶¶0025-0029. Cox describes that its model may then use statistical information related to these measurements, such as “mean accuracy” or “mean response time” to refine the training data used to train the supervised machine learning model. *Id.* Thus, by determining a “mean accuracy” and “mean response time” of responses from multiple human annotators, Cox teaches to Ross the association of accuracy and response time parameters with central tendencies reflecting average (mean) accuracy and response time (*the parameter is further associated with at least one of a central tendency . . .*). *Id.*; Section VI.A.7 (claim 5).

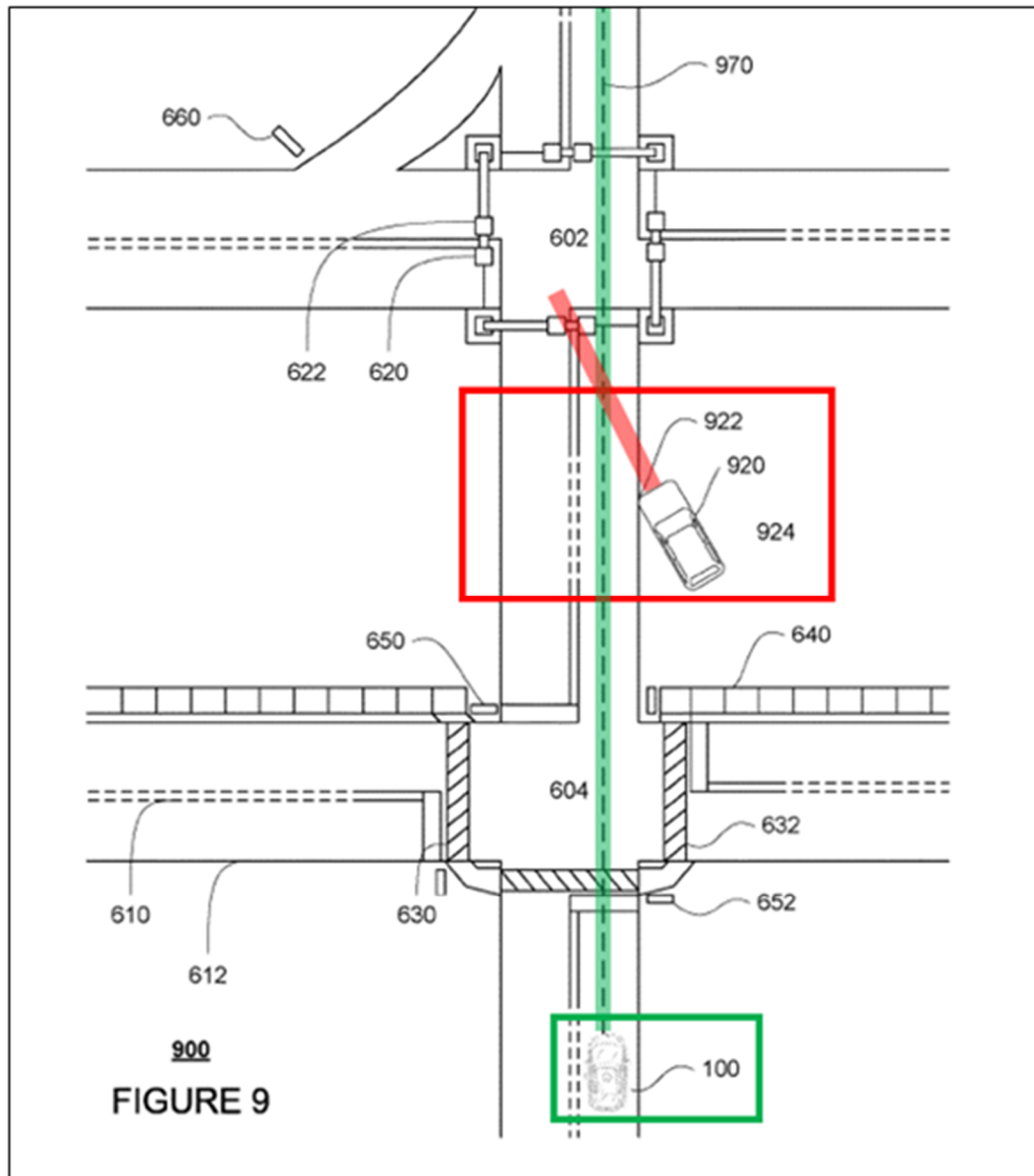
9. Claim 7

114. In my opinion, Ross and Cox render obvious claim 7. The Ross-Cox 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.” EX1004, 1:55-2:13, 4:33-46, 9:13-10:10, 15:1-16:52; EX1005, ¶¶0016, 0056; Sections VI.A.3.c, VI.A.3.d

(elements [1b], [1c]). As discussed with respect to elements 1[b] and 1[c], Cox teaches model training using human annotated training data that provides model predictions are more consistent with human annotation decisions and reflect state of mind of a subject. Sections VI.A.3.c, VI.A.3.d (elements [1b], [1c]). Thus, due to such training, the predictions by the Ross-Cox model reflect predicted intents that human annotators would have responded with when annotating the image data under analysis by the model, teaching *wherein the supervised learning based model is configured to predict user responses. Id.* Moreover, the predicted intents are, for an object vehicle or object bicycle, the intent of a vehicle driver or bicyclist riding the bicycle to turn the vehicle/bicycle, change lanes, drive through an intersection, or cross and move into the path of a roadway in which the autonomous vehicle is traveling. *Id.*; EX1004, 2:4-13, 9:13-10:10, 15:1-16:52. The predicted intents therefore describe *a state of mind of a road user associated with the object* because they describe how a road user would like to operate their vehicle. *Id.*

115. For example, as shown in annotated Figure 9 below, an intent may be that the driver of vehicle 920 intends that vehicle 920 will “exit from parking lot 924 and cross edge 922,” and thereby move (shown in red) into the path of autonomous vehicle 100 (shown in green) when it is “clear.” EX1004, 16:5-41, Fig. 9. In my opinion, the predictions of human responses to viewing the input camera sensor data describe how an annotator would have identified that the driver of vehicle 920 has

an intention to move vehicle 920 into the pathway of vehicle 100 when it is clear, and that the driver of vehicle 920 thus has an awareness of the vehicle 100 driving in the roadway they wish to enter so they can enter the roadway safely. *Id.*



EX1004, Fig. 9 (annotated)

10. Claim 8

116. In my opinion, Ross and Cox render obvious claim 8. *See* the analysis of Section VI.A.9 (Claim 7).

11. Claim 9

117. In my opinion, Ross and Cox render obvious claim 9. *See* the analysis of Section VI.A.9 (Claim 7).

12. Claim 10

118. In my opinion, Ross and Cox render obvious claim 10 for the same reasons discussed with respect to claim 1, as indicated below.

a. 10[Pre]

119. *See* Section VI.A.3.a (element 1[Pre]). Moreover, in my opinion, Ross describes that its computing device 110 contains “one or more processors 120” and “memory 130,” where “the memory 130 stores information accessible by the one or more processors 120, including instructions 132 and data 134 that may be executed or otherwise used by the processor 120” to effectuate the operation of computing device 110 (*[a] non-transitory computer readable storage medium storing instructions that when executed by one or more processors, cause the one or more processors to perform the steps*). EX1004, 5:13-6:5, 6:33-59, Fig. 1, Claim 11.

b. 10[a]

120. *See* Section VI.A.3.b (element 1[a]).

c. 10[b]

121. *See* Section VI.A.3.c (element 1[b]).

d. 10[c]

122. *See* Section VI.A.3.d (element 1[c]).

e. 10[d]

123. *See* Section VI.A.3.e (element 1[d]).

13. Claim 11

124. In my opinion, Ross and Cox render obvious claim 11 for the same reasons discussed above in Section VI.A.4 (claim 2).

14. Claim 12

125. In my opinion, Ross and Cox render obvious claim 12 for the same reasons discussed above in Section VI.A.5 (claim 3).

15. Claim 13

126. In my opinion, Ross and Cox render obvious claim 13 for the same reasons discussed above in Section VI.A.6 (claim 4).

16. Claim 14

127. In my opinion, Ross and Cox render obvious claim 14 for the same reasons discussed above in Section VI.A.7 (claim 5).

17. Claim 15

128. In my opinion, Ross and Cox render obvious claim 15 for the same reasons discussed above in Section VI.A.8 (claim 6).

18. Claim 16

129. In my opinion, Ross and Cox render obvious claim 16 for the same reasons discussed above in Section VI.A.9 (claim 7).

19. Claim 17

130. In my opinion, Ross and Cox render obvious claim 17 for the same reasons discussed above in Section VI.A.10 (claim 8).

20. Claim 18

131. In my opinion, Ross and Cox render obvious claim 18 for the same reasons discussed above in Section VI.A.11 (claim 9).

21. Claim 19

132. In my opinion, Ross and Cox render obvious claim 19 for the same reasons discussed with respect to claims 1 and 10, as indicated below.

a. 19[Pre]

133. *See* Sections VI.A.3.a (element 1[Pre]), VI.A.12.a (element 10[Pre]).

b. 19[a]

134. *See* Sections VI.A.3.b (element 1[a]).

c. 19[b]

135. *See* Sections VI.A.3.c (element 1[b]).

d. 19[c]

136. *See* Sections VI.A.3.d (element 1[c]).

e. 19[d]

137. *See* Sections VI.A.3.e (element 1[d]).

22. Claim 20

138. In my opinion, Ross renders obvious Claim 20 for the same reasons discussed above in Section VI.A.9 (claim 7).

23. Claim 21

139. In my opinion, Ross renders obvious Claim 21 for the same reasons discussed above in Section VI.A.10 (claim 8).

24. Claim 22

140. In my opinion, Ross renders obvious Claim 22 for the same reasons discussed above in Section VI.A.11 (claim 9).

B. Ground 2: Claims 1-22 are unpatentable under 35 U.S.C. § 103 over Ross, Cox, and Ogale

141. Ross and Cox render obvious the Challenged Claims for the reasons discussed in Section VI.A. However, to the extent that the Patent Owner argues, or the Board finds, that the predictions generated by the Ross-Cox model do not correspond to (1) *user responses expected to be received responsive to presenting the sensor data to a plurality of users* or (2) *user responses expected to be received responsive to presenting the sensor data captured by the sensor installed on the autonomous vehicle to users* (as recited by elements 1[b] and 1[c], *infra*), the combination of Ross, Cox, and Ogale teaches these elements.

1. Summary of Prior Art

a. Ogale–EX1010

142. Ogale—another Waymo patent—is generally directed to “autonomous vehicles” and “generating object predictions using a neural network.” EX1010, Abstract, 1:5. Ogale explains that for a neural network trained using human labeled training data prepared by “human participants,” the prediction generated by the neural network based on input sensor data “predicts how a human would have labeled the input sensor data.” *Id.*, 3:29-37, 6:51-7:12, 13:7-11, 13:30-14:9. Ogale’s disclosure aligns with the ’344 Patent’s purported invention, which includes a model predicting information “characterizing a distribution of user responses expected to be received responsive to presenting [] sensor data to a plurality of users.” EX1001, Claim 1. Ogale is analogous art to the ’344 Patent, from the same field of endeavor as the patent (e.g., accounting for moving objects when navigating an autonomous vehicle in an environment), 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). EX1010, 3:29-37, 6:51-7:12, 13:7-11, 13:30-14:9; EX1001, 1:15-17, 1:53-56.

2. Motivation to Combine Ross, Cox, and Ogale

143. In my opinion, a POSITA would have been motivated to combine Ross and Cox as discussed in Section VI.A.2. A POSITA would have been further motivated to combined Ross and Cox with Ogale. The combination would have

made clear that the Ross-Cox model, trained using human annotated training data, would have thus provided predictions that “predict[] how a human would have labeled the input sensor data.” *See* Section VI.A.2; EX1010, 3:29-37, 6:51-7:12, 13:7-11, 13:30-14:9.

144. In my opinion, the Ross-Cox combination already describes model training using “human observations” that are manually marked as correct and human labeled training data that describes state of mind of a training object. *See* Section VI.A.2. The Ross-Cox combination further provides that the Ross-Cox model “function[s] in a more ‘human-like’ or ‘polite’ way” (EX1004, 5:7-12) and “mimic[s] human performance” to ensure that model predictions that are “more consistent with the decisions of the human annotators” (EX1005, ¶¶0004, 0029-0032). *See* Section VI.A.2. Ogale further elaborates that such a model, trained using human participants, would have generated predictions based on input sensor data that “predicts how a human would have labeled the input sensor data.” EX1010, 3:29-37, 6:51-7:12, 13:7-11, 13:30-14:9. In my opinion, Ogale thus merely provides implementation details for the Ross-Cox combination. This explicitly provides that the result of the combination (i.e., model predictions that “predict[] how a human would have labeled the input sensor data.”) would have been predictable to a POSITA, and also had a reasonable expectation of success for at least these same reasons. *Id.*

145. Moreover, in my opinion, a POSITA would have looked to Ogale, Ross, Cox, and Ogale 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 Ross-Cox combination would have looked to Cox. *Id.*; see Sections VI.A.1.a, VI.A.1.b, VI.B.1.a. Ross, Cox, and Ogale overlap in that they are all related to machine learning models within the context of autonomous vehicles and human annotated training data. EX1004, 1:30-31, 3:39-42, 3:64-5:5, 9:30-10:10 (Ross, describing an autonomous vehicle control system); EX1005, ¶¶0027-0032, 0060 (Cox, describing relevance to “visual recognition problems” including those related to detecting “objects” such as a “car”); EX1010, Abstract, 1:5 (Ogale, directed to “autonomous vehicles” and “generating object predictions using a neural network”). And as noted above, each of Ross, Cox, and Ogale are directed to techniques that provide for a predictive model that outputs more “human” predictions, further motivating the combination. Combining Ogale with Ross and Cox would also have improved how “‘human-like’ or ‘polite’” the model behaves—which is a stated goal of Ross—because, as noted above, Ogale teaches model predictions that reflect “how a human would have labeled the input sensor data” and are thus similarly human-like. *Id.*; EX1010, 3:29-37, 6:51-7:12, 13:7-11, 13:30-14:9; EX1004, 5:7-12. This would have improved the Ross-Cox

model's predictions, and in turn, the autonomous vehicle's safe navigation around other humans/vehicles near the autonomous vehicle.

146. Therefore, for the reasons discussed above, it is my opinion that a POSITA would have been motivated to combine Ross, Cox, and Ogale. In addition, as demonstrated above, the combination would have merely involved combining prior art elements (e.g., Ross's model for predicting intents; Cox's model training and optimization techniques; Ogale's prediction model trained using human labels) according to known methods (known software techniques with respect to machine learning model training) to yield the predictable results of model predictions that "predict[] how a human would have labeled the input sensor data." *See* above analysis in this Section VI.B.2; EX1010, 3:29-37, 6:51-7:12, 13:7-11, 13:30-14:9. The combination would have further been use of a known technique (known machine learning model training using human labeling as taught by Ogale) to improve similar devices (Ross-Cox model for predicting intents) in the same way (using Ogale's aforementioned techniques). And a POSITA would have had a reasonable expectation of success at least due to the overlap in Ross's, Cox's, and Ogale's disclosures, compatibility of the references, and motivations provided by the references themselves, as demonstrated above. Indeed, because Ross, Cox, and Ogale all use the same type of human-labeled data to train models to be more human-like (as discussed above), combining Ogale with Ross and Cox would have required

only minimal modifications that were well within the skill of a POSITA, and for which a POSITA would have had a reasonable expectation of success.

3. Claim 1

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

a. 1[Pre] and 1[a]

148. In my opinion, Ross, Cox, and Ogale teach elements 1[Pre] and 1[a] for the same reasons discussed in Section VI.A.3.a and VI.A.3.b (elements 1[Pre] and 1[a]).

b. 1[b]

149. In my opinion, Ross, Cox, and Ogale teach element 1[b]. The Ross-Cox combination teaches the Ross-Cox model (*supervised learning based model*). See Section VI.A.3.c (element 1[b]). Moreover, Ross describes that raw camera data including an object (e.g., another vehicle, bicyclist, pedestrian) on a road is provided as an input to the model (*providing the sensor data as input to a supervised learning based model*), and Ross further explains that the model is configured to receive this specific raw camera data (discussed in Sections VI.A.3.a (elements 1[Pre]) and VI.A.3.b (element 1[a])) as well as other raw camera data captured at other times by the vehicle, for example during training of the model and when the model is analyzing live, non-training data of an environment (*the supervised learning based*

model configured to receive an input sensor data displaying a particular object on a road). See VI.A.3.c (element 1[b]).

150. In my opinion, Ross and Cox teach *predicting an output statistical summary characterizing a distribution of user responses expected to be received responsive to presenting the sensor data to a plurality of users, the user responses associated with the particular object* in the non-training and training contexts for the reasons discussed in Section VI.A.3.c (element 1[b]). But to the extent the Patent Owner argues or the Board finds that Ross and Cox do not teach this recitation, the combination of Ross, Cox, and Ogale does.

151. Like Ross and Cox, Ogale describes that its prediction model for an autonomous vehicle that is trained using human labeled training data and in particular, labeled training data provided by “multiple human participants.” EX1010, 3:29-37, 6:51-7:12, 13:7-11, 13:30-14:9. By training Ogale’s model with such human labeled data, the model outputs predictions based on input sensor data that “predicts how a human would have labeled the input sensor data.” *Id.*

152. Thus, as applied to the Ross-Cox combination, in both the non-training and training contexts discussed in Section VI.A.3.c (element 1[b] of Ground 1), Ogale teaches training the Ross-Cox intent model with human observation data and human annotator responses, expanding on the disclosure taught by Ross individually and the combination of Ross and Cox (*See id.*), such that the Ross-Cox model

predicts how a human would have labeled the input sensor data. EX1010, 3:29-37, 6:51-7:12, 13:7-11, 13:30-14:9 (bracketing added); EX1004, 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); EX1005, ¶¶0004, 0007-0010, 0015-0019, 0025-0035, 0051; Section VI.A.3.c (element 1[b] of Ground 1). The Ross-Cox model’s predicted “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” (*predicting an output statistical summary*) based on input raw camera data characterizes how “multiple human participants” observing training data and labeling the training data would have responded to the input raw camera data and predicted possible next actions (and trajectories) (*id.*); indeed, the predictions would have reflected “how a human would have labeled the input [raw camera] data” as taught by Ogale (*characterizing a distribution of user responses expected to be received responsive to presenting the sensor data to a plurality of users, the user responses associated with the particular object*). *Id.* This is because Ogale makes clear through its disclosure that human annotated training data used to train a predictive model provides that trained model outputs reflect how human annotators would have annotated the input data. EX1010, 3:29-37, 6:51-7:12, 13:7-11, 13:30-14:9.

c. 1[c]

153. In my opinion, Ross, Cox, and Ogale teach element 1[c]. Ross discloses that the particular raw camera data from the cameras of vehicle 100 (*sensor data captured by the sensor installed on the autonomous vehicle*) is provided as the input to the trained Ross-Cox model (*see* Sections VI.A.3.b, VI.A.3.c (elements 1[a], 1[b]) of Ground 1), which was trained via the human observations, human annotations, and model optimization as discussed for element 1[b] in Ground 1 and element 1[b] in Ground 2 with respect to Ross, Cox, and Ogale. *Id.*; *see also* Sections VI.A.3.a-VI.A.3.c (elements 1[Pre], 1[a], and 1[b] of Ground 2), VI.B.3.a-VI.B.3.c. The trained model is then executed on this input data (*executing the trained supervised learning based model*) to generate 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” (*to generate a statistical summary data*), which is the Ross predicted data, for objects in the particular raw camera data. EX1004, 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.3.c (element 1[b] of Ground 1). Specifically, the trained model is executed on the particular raw camera data from the cameras of vehicle 100 input into the model as in the “Model Functionality During Operational Use” example for element 1[b] of Ground 1. *See* Section VI.A.3.c (element 1[b]). As discussed in Section VI.B.3.b(element 1[b] of

Ground 2), the Ross, Cox, and Ogale combination further teaches that the model output characterizes how “multiple human participants” observing training data and labeling the training data would have responded to the input raw camera data and predicted possible next actions (and trajectories), and in turn “how a human would have labeled the input [raw camera] data” as taught by Ogale. Section VI.B.3.b (element 1[b] of Ground 2).

154. In my opinion, the Ross predicted data generated by the trained model forms *a statistical summary data* with respect to the particular raw camera data from the cameras of vehicle 100 for the same reasons discussed regarding the *statistical summary* recitation in element 1[b] of Ground 2. *See* Sections VI.B.3.b (element 1[b]), VI.A.3.c. Moreover, the Ross predicted data for the particular raw camera data from the cameras of vehicle 100 (*statistical summary data*) is *characterizing a distribution of user responses expected to be received responsive to presenting the sensor data captured by the sensor installed on the autonomous vehicle to users* with respect to the particular raw camera data for the same reasons discussed regarding the *characterizing a distribution* recitation in element 1[b] of Ground 2 for Ross, Cox, and Ogale. *Id.* The Ross predicted data generated for the particular raw camera data (*output statistical summary*) characterizes how “multiple human participants” observing training data and labeling the training data would have responded to the particular raw camera data and predicted possible next actions (and trajectories and

thus state of mind) for the object (e.g., a vehicle, pedestrian, bicyclist) in the data (*id.*); the model would have been trained using human annotated data and thus generated predictions that reflect “how a human would have labeled the input [raw camera] data” as taught by Ogale (*characterizing a distribution of user responses expected to be received responsive to presenting the sensor data captured by the sensor installed on the autonomous vehicle to users*). Sections VI.B.3.a-VI.B.3.b (elements 1[Pre]-1[b] of Ground 2), VI.A.3.c (element 1[b] of Ground 1)).

d. 1[d]

155. In my opinion, Ross, Cox, and Ogale teach element 1[d] for the same reasons discussed in Sections VI.A.3.e (element [1d]), VI.B.3.b (element 1[b]), and VI.B.3.c (element 1[c]).

4. Claims 10 and 19

156. In my opinion, Ross, Cox, and Ogale render obvious Claims 10 and 19. Ross, Cox, and Ogale teach elements 10[b], 10[c], and 10[d], and 19[b], 19[c], and 19[d] for the same reasons discussed above in Sections VI.B.3.b(element 1[b]), VI.B.3.c (element 1[c]), and VI.B.3.d (element 1[d]).

157. In my opinion, Ross, Cox, and Ogale teach the remaining elements of claims 10 and 19 for the same reasons discussed in Sections VI.A.12 (claim 10) and VI.A.21 (claim 19).

5. Claims 2-9, 11-18, and 20-22

158. In my opinion, Ross, Cox, and Ogale render obvious claims 2-9, 11-18, and 20-22 for the same reasons discussed in Sections VI.A.4-VI.A.11, VI.A.13-VI.A.20, and VI.A.22-VI.A.24.

C. Ground 3: Claims 1-5, 7-13, and 16-22 are unpatentable under 35 U.S.C. § over Zhang

1. Summary of the Prior Art

a. Zhang–EX1006

159. Zhang is generally directed to a “prediction-based system and method of trajectory planning of autonomous vehicles.” EX1006, Abstract. Zhang describes an autonomous vehicle 105 that implements a trained neural network to analyze image data of the environment around the vehicle, captured by cameras of the autonomous vehicle, and determine a “set of predicted trajectories and confidence levels” of moving vehicles and other “dynamic objects” in the image data. *Id.*, ¶¶0019, 0025, 0029-0030, 0047, 0049-0065, Figs. 1, 3, 4. The neural network is trained using “real world, human behavior data” that represents “human driving behaviors,” and the training data includes images of driving behavior that are labeled by human labelers. EX1006, Abstract, ¶¶0025, 0047, 0049, 0051-0065, claims 1, 4, 5, Figs. 4, 5, 8, 9, 10. Zhang’s model is optimized to correct model predictions during training. *Id.*, ¶0049. In operational use, predicted trajectories are used by autonomous vehicle 105 to control the vehicle. *Id.*, ¶¶0054, 0057-0061, Fig. 8.

160. Zhang is analogous art to the '344 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., accounting for moving objects when navigating an autonomous vehicle in an environment). EX1006, ¶¶0019, 0025, 0029-0030, 0047, 0049-0065, Figs. 1, 3, 4; EX1001, 1:15-17, 1:53-56.

2. Claim 1

161. In my opinion, Zhang renders obvious Claim 1.

a. 1[Pre]

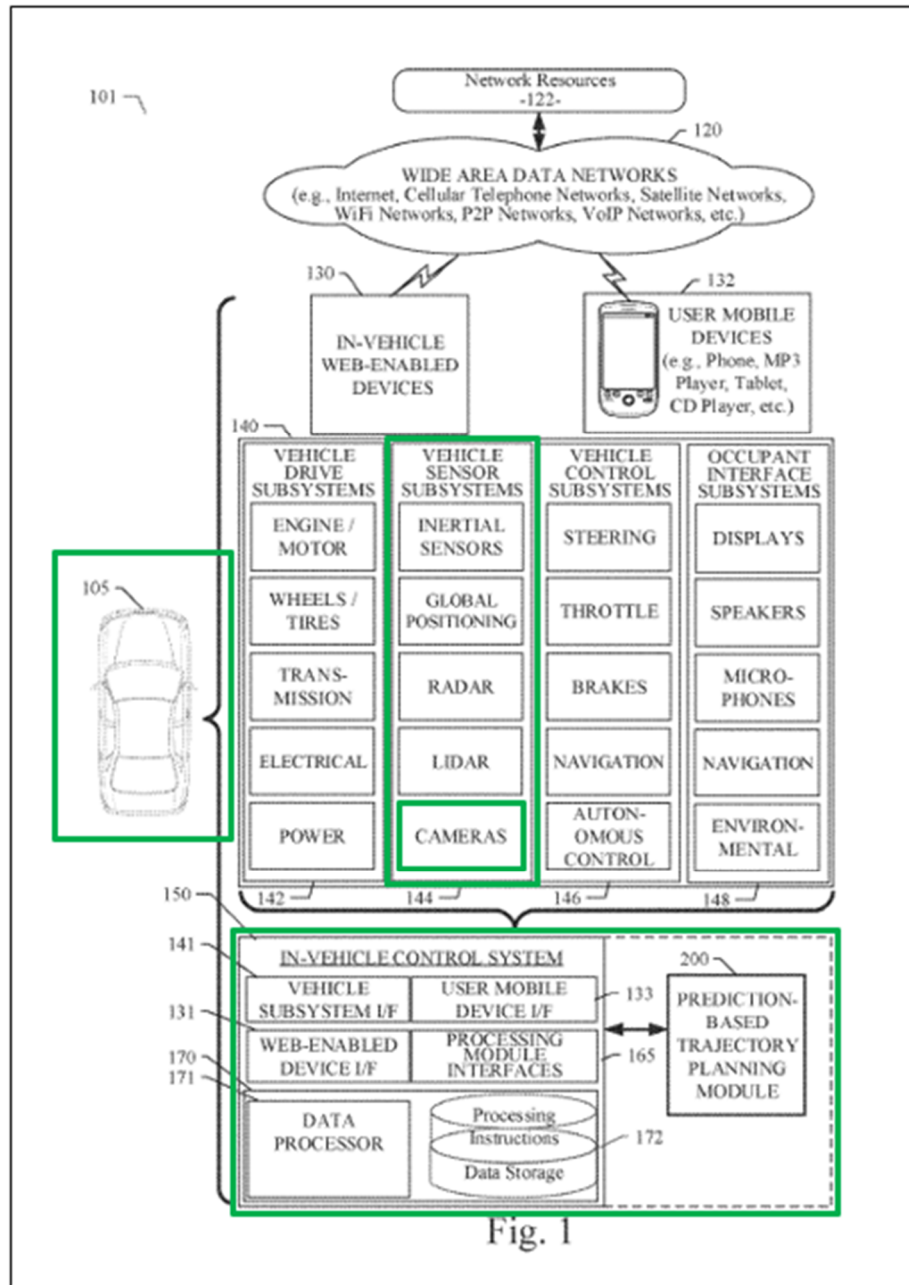
162. In my opinion, to the extent the preamble is limiting, Zhang teaches it. Zhang describes techniques for “trajectory planning, vehicle control systems, and autonomous driving systems,” including “trajectory planning for autonomous vehicles using a prediction-based method” (*[a] computer-implemented method*). EX1006, ¶¶0003,0006, 0018-0032, 0036, 0052-0061, Figs. 1, 5, 6, 9, 11.

b. 1[a]

163. Zhang teaches element 1[a]. Zhang describes a vehicle 105, (a car) that operates in a fully autonomous mode on a “roadway.” EX1006, ¶¶0006, 0018-0026, 0050, 0073, Figs. 1, 3, 11, claim 1, claim 19. The vehicle includes an “in-vehicle control system 150” having a “prediction-based trajectory planning module 200” that is implemented by a “computing system” (*a computing device associated with*

Declaration of Dr. Jason Janét Regarding U.S. Patent No. 10,614,344

an autonomous vehicle operating on a road). *Id.* “[O]ne or more cameras or image capture devices” are attached to vehicle 105 (*a sensor installed on the autonomous vehicle*) and “sense information about an environment” of vehicle 105 to detect “objects” in the environment. EX1006, ¶¶0019, 0029-0030, 0049-0051, 0059, Figs. 1, 3, 4.



EX1006, Fig. 1 (annotated)

164. The in-vehicle control system 150 receives the image data captured by the cameras (*receiving, by a computing device associated with an autonomous vehicle operating on a road, sensor data captured by a sensor installed on the*

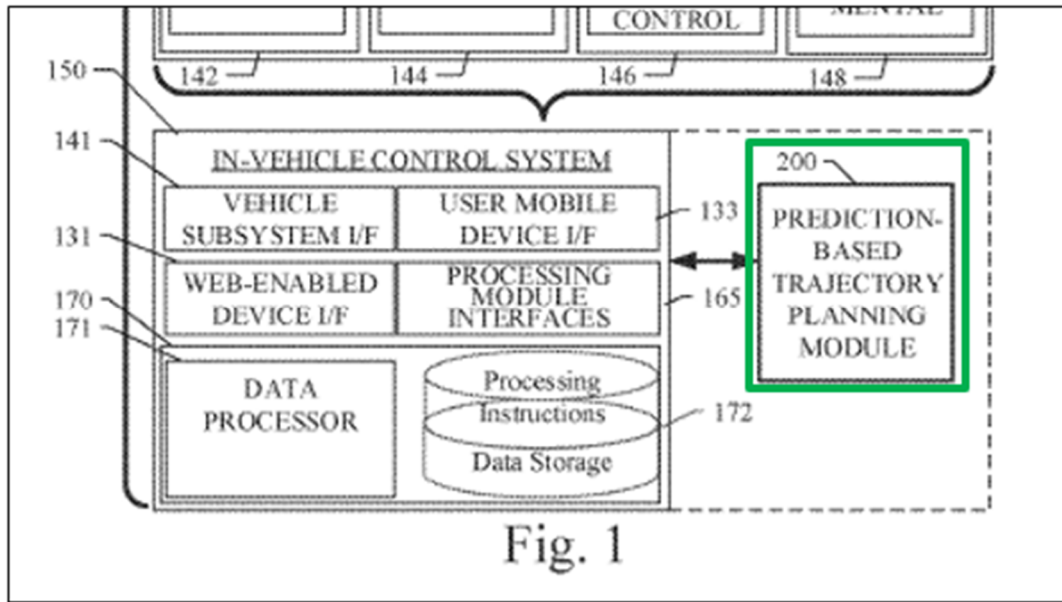
autonomous vehicle). EX1006, ¶¶0019, 0029-0030, 0049-0051, 0058-0059, Figs. 1, 3, 4. The image data displays objects such as “moving vehicles, dynamic objects, or other objects in the proximate vicinity of the vehicle 105” that are on the roadway (*the sensor data displaying an object on the road*). *Id.* In my opinion, this reception of image data occurs during both a “training phase” and an “operational phase” in Zhang’s system. EX1006, ¶¶0019, 0049-0054, 0059-0065, Figs. 1, 2, 3, 7.

c. 1[b]

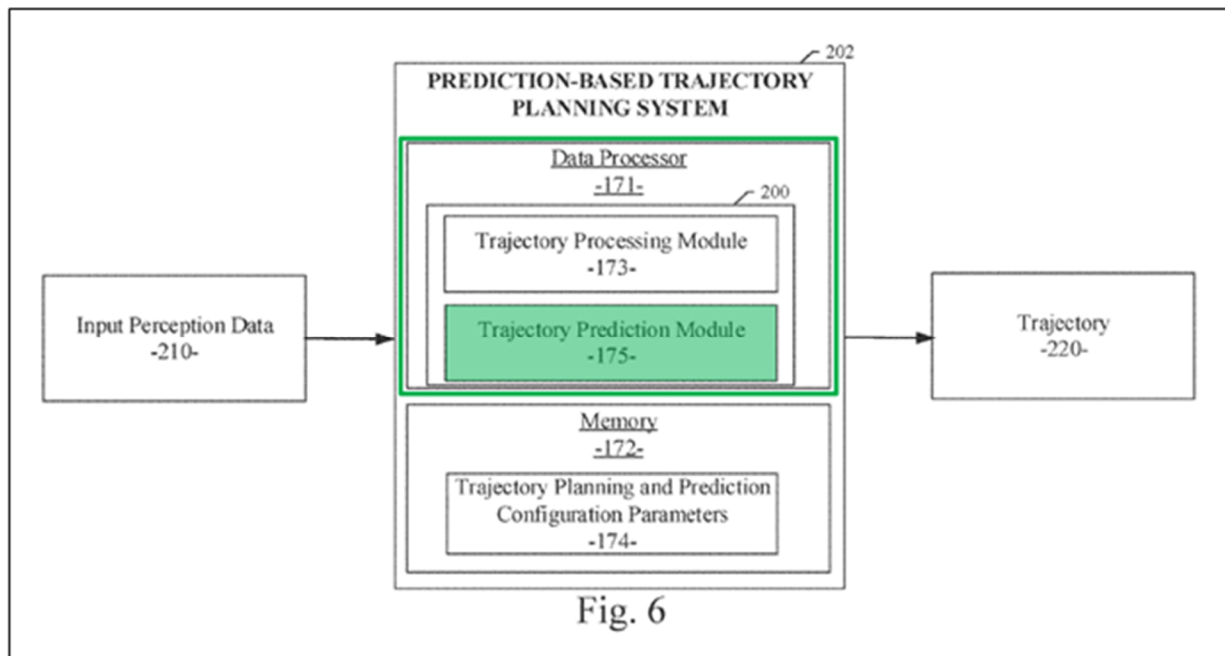
165. In my opinion, Zhang teaches element 1[b].

- ***supervised learning based model***

166. Zhang discloses prediction-based trajectory planning module 200 includes a “trajectory prediction module 175” implemented as a trained “neural network” (*supervised learning based model*). EX1006, ¶¶0045-0059, 0064, Figs. 1, 3, 6, claims 4, 5.



EX1006, Fig. 1 (in part, annotated)



EX1006, Fig. 6 (annotated)

167. Zhang’s neural network model is trained by human labeled training data “to produce [a] configured output based on training data provided to the neural

network or other machine learning system in a training phase.” EX1006, ¶¶0045-0059, 0064, Figs. 1, 3, 6, claims 4, 5. Moreover, a “loss function” is used to “examine and correct the results of the training provided to the trajectory prediction module 175.” *Id.* Loss functions are well known mathematical functions used in model training that measure how well a model’s predicted output matches a known correct output, thereby quantifying model error. *Id.* These model training techniques disclosed by Zhang thus demonstrate that its neural network is supervised to ensure that its outputs are accurate in view of training data. *Id.* Zhang therefore teaches a *supervised learning based model. Id.*

168. In my opinion, Zhang’s neural network forming a *supervised learning based model* also aligns with claim 2 of the ’344 Patent itself, which recites that the *supervised learning based model* is a *simple neural network*. EX1006, ¶¶0045-0059, 0064, Figs. 1, 3, claims 4, 5; EX1001, claim 2.

- ***providing the sensor data as input to a supervised learning based model***

169. Image data from the cameras of Zhang’s autonomous vehicle is part of “perception data 210” that is provided as an input to the trajectory prediction module 175 implemented as a neural network (*providing the sensor data as input to a supervised learning based model*) EX1006, ¶¶0045-0059, 0064, Figs. 1, 3, claims 4, 5; Section VI.C.2.b (element 1[a]). Image data is provided to Zhang’s neural network

during both a “training phase” of the model and an “operational phase” where the trained model is deployed to navigate the autonomous vehicle. *Id.*

- ***the supervised learning based model configured to receive an input sensor data displaying a particular object on a road***

170. In my opinion, the '344 Patent limitation of *the supervised learning based model configured to receive an input sensor data displaying a particular object on a road* may be interpreted to refer to functionality the model is configured to perform during “live” non-training use (referred to as “operational use” hereafter), as well as during a training. In either situation, Zhang teaches the limitation.

171. As noted above, Zhang discloses that its neural network receives image data from the cameras of Zhang’s autonomous vehicle, where the image data displays objects such as “moving vehicles, dynamic objects, or other objects in the proximate vicinity of the vehicle 105” (*the supervised learning based model configured to receive an input sensor data displaying a particular object on a road*). EX1006, ¶¶0045-0059, 0064, Figs. 1, 3, claims 4, 5; Section VI.C.2.b (element 1[a]). The data is received during both training and operational use. *Id.*

- ***predicting an output statistical summary characterizing a distribution of user responses expected to be received responsive to presenting the sensor data to a plurality of users, the user responses associated with the particular object***

172. In my opinion, the '344 Patent limitation *predicting an output statistical summary characterizing a distribution of user responses expected to be received responsive to presenting the sensor data to a plurality of users, the user responses associated with the particular object* may be interpreted to refer to functionality the model is configured to perform during operational use as well as during training. In either situation, Zhang teaches the limitation.

Model Functionality During Operational Use

173. Zhang's neural network of trajectory prediction module 175 is executed to generate a "set of predicted trajectories and confidence levels" for each object (e.g., a moving vehicle) in the image data, where the confidence levels "correspond[] to a likelihood" of the object to "perform the predicted behavior." EX1006, ¶¶0025, 0058. Zhang's "set of predicted trajectories and confidence levels" teaches *an output statistical summary. Id.* The set of predicted trajectories and confidence levels summarize the various trajectories than an object in image data might take and the confidences associated with each prediction. *Id.* This provides a statistical distribution and organization reflecting how an object is predicted to travel within the environment of Zhang's autonomous vehicle, and how likely each prediction is. *Id.* Because these values are a collection of numerical measurements, it is my opinion that a POSITA would have recognized that such values form statistics. *Id.*

174. In my opinion Zhang further teaches, that the “set of predicted trajectories and confidence levels” (*output statistical summary*) is *characterizing a distribution of user responses expected to be received responsive to presenting the sensor data to a plurality of users, the user responses associated with the particular object* as claimed. Zhang describes that its neural network of trajectory prediction module 175 is trained by training data that is “real world, human behavior data” that represents “human driving behaviors.” EX1006, Abstract, ¶¶0025, 0047, 0049, 0051-0065, claims 1, 4, 5, Figs. 4, 5, 9, 10. Zhang explains that “[b]ecause the trajectory prediction module 175 is trained in the training phase using real world, human behavior data, the predicted behavior and trajectories of vehicles or objects produced by the trajectory prediction module 175 are closely correlated to the actual behavior and trajectories of vehicles in real world environments with human drivers.” *Id.* Zhang’s training data includes images that are labeled by human labelers to “associate a label having behavior and direction information with each instance of a vehicle in the training data.” *Id.* The labeled training represents different driving scenarios and includes “human driver behavior data related to how drivers typically react to actions performed by proximate vehicles.” *Id.* A “loss function” is used to “examine and correct the results of the training.” *Id.*, ¶0049. In my opinion, a POSITA would have understood that loss functions are used in machine learning to quantify the error between the correct human labels of training

images and a desired model output, and thus allows for adjustment of the model to reduce the error.

175. Thus, in my opinion, when Zhang's neural network predicts a "set of predicted trajectories and confidence levels" for each object (e.g., moving vehicle) in the image data, the model is predicting the driving behaviors (e.g., trajectories) that human labelers would have identified for the object in the image data and responded with by labeling in the data when presented with the data. EX1006, Abstract, ¶¶0025, 0047, 0049, 0051-0065, claims 1, 4, 5, Figs. 4, 5, 9, 10. The set of confidence values output by the model are the likelihoods of the behaviors/trajectories occurring based on the input camera data of the object. EX1006, Abstract, ¶¶0025, 0047, 0049, 0051-0065, claims 1, 4, 5, Figs. 4, 5, 9, 10. The model predicts these behaviors by being trained on the training data of human labels from humans observing training images of human driving behavior, with the training (and use of a loss function) reducing the error between human labels and the predicted behaviors/trajectories. *Id.* The "set of predicted trajectories and confidence levels" (*output statistical summary*) therefore reflects the distribution of the various human labeler responses (*characterizing a distribution of user responses*) indicating possible trajectories/behaviors of the object (e.g., a moving vehicle) that human labelers would have identified for the object in the image data and responded with by labeling in the data when presented with the data (*expected to be received*

responsive to presenting the sensor data to a plurality of users, the user responses associated with the particular object). EX1006, Abstract, ¶¶0025, 0047, 0049, 0051-0065, claims 1, 4, 5, Figs. 4, 5, 9, 10.

Model Functionality During Training

176. In the training context, Zhang performs the same functionality as discussed above in the operational use example. *See* prior analysis in this Section VI.C.2.c (element 1[b]), *supra*. As noted above, Zhang’s neural network of trajectory prediction module 175 is trained during a training phase by training data, including camera images, that is “real world, human behavior data” that represents “human driving behaviors” and is human labeled, and a “loss function” is used to “examine and correct the results of the training.” *Id.*; EX1006, Abstract, ¶¶0025, 0043-0049, 0051-0065, claims 1, 4, 5, Figs. 2, 4, 5, 9, 10. Moreover, Zhang also teaches that its model is refined and corrected during training. *Id.* Refining the model allows the model predictions to accurately reflect the distribution of the various human labeler responses indicating possible trajectories/behaviors of the object (e.g., a moving vehicle) that human labelers would have identified for the object in the image data and responded with by labeling that data when presented with it. *Id.*

177. In my opinion, the foregoing confirms that Zhang’s “set of predicted trajectories and confidence levels” described above for operational use are also generated during training. EX1006, ¶0058. As noted above, it is my opinion that a

POSITA would have understood that loss functions are used in machine learning to quantify the error between the correct human labels of training images and a desired model output, and thus allows for adjustment of the model to reduce the error. *Id.* Zhang therefore teaches *predicting an output statistical summary characterizing a distribution of user responses expected to be received responsive to presenting the sensor data to a plurality of users, the user responses associated with the particular object* in the training context for the same reasons discussed above in operational use. See prior analysis of this Section VI.C.2.c (element 1[b]), *supra*.

d. 1[c]

178. In my opinion, Zhang teaches element 1[c]. Zhang discloses that the particular camera data from the cameras of vehicle 105 (*sensor data captured by the sensor installed on the autonomous vehicle*) is provided as the input to the Zhang's trained neural network (*see* Sections VI.C.2.b, VI.C.2.c (elements 1[a], 1[b])), which was trained using the human annotated behavior training data and loss function as discussed for element 1[b]. *Id.* The trained model is then executed on this input data (*executing the trained supervised learning based model*) to generate a "set of predicted trajectories and confidence levels" for each object (e.g., moving vehicle) in the image data (*to generate a statistical summary data*). EX1006, ¶¶0025, 0058; Section VI.C.2.c (element 1[b]). For example, the trained model is executed on the particular image data from the cameras of vehicle 105 input into the neural network

as discussed in element 1[b] in the “Model Functionality During Operational Use” example discussed for element 1[b]. *See* Section VI.C.2.c (element 1[b]).

179. In my opinion, Zhang’s “set of predicted trajectories and confidence levels” forms *a statistical summary data* with respect to the particular input image data for the same reasons discussed regarding the *statistical summary* recitation in element 1[b]. *See* Section VI.C.2.c (element 1[b]). Moreover, the “set of predicted trajectories and confidence levels” for the particular input image data from the cameras of vehicle 105 (*statistical summary data*) is *characterizing a distribution of user responses expected to be received responsive to presenting the sensor data captured by the sensor installed on the autonomous vehicle to users* with respect to the particular input image data for the same reasons discussed regarding the *characterizing a distribution* recitation in element 1[b] for Zhang. *Id.* In my opinion, Zhang’s generated “set of predicted trajectories and confidence levels” for the particular input image data (*output statistical summary*) reflects the distribution of the various human labeler responses (*characterizing a distribution of user responses*) indicating possible trajectories/behaviors of the object (e.g., a moving vehicle) that human labelers would have identified for the object in the particular input image data and responded with by labeling in the data when presented with the data (*expected to be received responsive to presenting the sensor data to a plurality*

of users, the user responses associated with the particular object). Section VI.C.2.c (element 1[b]).

e. 1[d]

180. In my opinion, Zhang teaches element 1[d]. Autonomous vehicle 105 generates a “proposed trajectory” for the vehicle 105 itself to follow on the road. EX1006, ¶¶0057-0061, Fig. 8; *id.*, 0019, 0044, 0047, 0050-0051, 0054, Figs. 3, 4. The proposed trajectory of vehicle 105 is analyzed to determine if it will “conflict with any of the predicted trajectories” of the “set of predicted trajectories and confidence levels” (*generated statistical summary data*) for objects in the particular input image data, and if there is a conflict, the proposed trajectory of vehicle 105 is changed to eliminate the conflict. *Id.*, ¶0060, Fig. 8; *see* Section VI.C.2.d (element 1[c]). Thereafter, vehicle 105 is “control[led]” such that the “vehicle 105 is directed along the proposed trajectory, which will not be in conflict with any of the proximate vehicles.” EX1006, ¶0060, Fig. 8. Thus, in my opinion, because the control of vehicle 105 is based on a determination that the proposed trajectory does not conflict with any of the “set of predicted trajectories and confidence levels” (*generated statistical summary data*), the control of vehicle 105 on the road to follow the proposed trajectory is based on the “set of predicted trajectories and confidence levels” (*controlling the operation of the autonomous vehicle on the road based on the generated statistical summary data*). EX1006, ¶¶0057-0061, Fig. 8; *see also id.*,

¶¶0054 (explaining vehicle 105 is controlled based on a predicted trajectory “to more efficiently and safely control the host vehicle 105”), 0019, 0044, 0047, 0050-0051, Figs. 3, 4.

3. Claim 2

181. In my opinion, Zhang renders obvious claim 2. Zhang discloses that its model includes neural networks (*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*). See Section VI.C.2.c (element 1[b]).

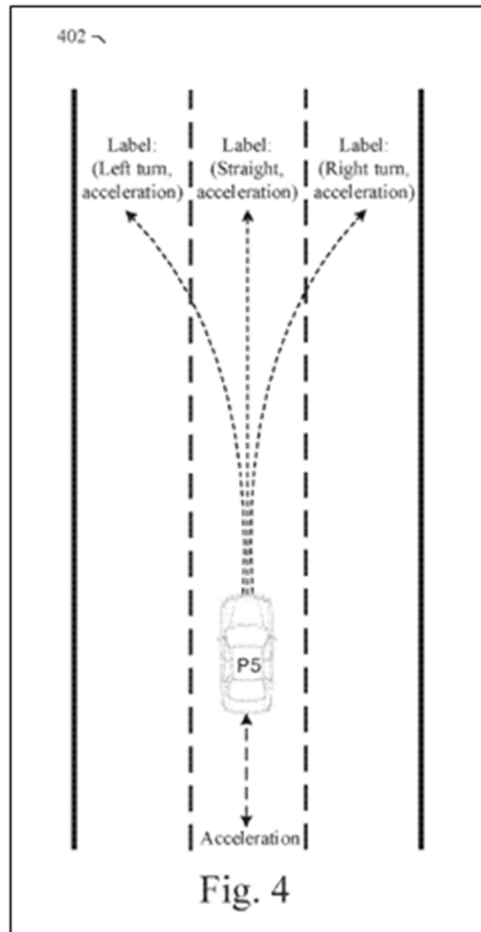
4. Claim 3

182. In my opinion, Zhang renders obvious claim 3. The input perception data 210 from cameras of Zhang’s vehicle displays objects on the road in the vehicle’s environment including other “moving vehicles and dynamic objects” (*wherein the object on the road is one of: a pedestrian, a cyclist, or a motor vehicle*). See Section VI.C.2.b (element 1[a]).

5. Claim 4

183. In my opinion, Zhang renders obvious claim 4. The possible trajectories/behaviors of the object (e.g., a moving vehicle) that human labelers would have identified for the object in the particular input image data when presented with the image include a “left turn” of an object vehicle into a left lane of

a road or a “right turn” of an object vehicle into a right lane of a road, as illustrated by Figure 4 (*wherein the user responses describe an action associated with the object, the action including one of the object...changing lanes....*). EX1006, ¶0051, Fig. 4; Sections VI.C.2.c, VI.C.2.d (elements 1[b], 1[c]).



EX1006, Fig. 4

6. Claim 7

184. In my opinion, Zhang renders obvious claim 7. Zhang’s neural network of trajectory prediction module 175 is executed to generate a “set of predicted

trajectories and confidence levels” for each object (e.g., moving vehicle and its associated driver road user) in the input image data, where the confidence levels “correspond[] to a likelihood” of the object to “perform the predicted behavior” indicated by the trajectories. EX1006, ¶¶0025, 0058. Human labeler responses of “behavior” in training images are used to train the neural network. EX1006, Abstract, ¶¶0025, 0047, 0049, 0051-0065, claims 1, 4, 5, Figs. 4, 5, 9, 10; Section VI.C.2.c (element 1[b]). Therefore, in my opinion, when Zhang’s neural network (*the supervised learning based model*), trained by the human labeled data noted above, predicts the “set of predicted trajectories and confidence levels” for a moving vehicle and its associated road user driver with each trajectory representing behavior of the moving vehicle and driver, the model is predicting the human labeler responses describing this behavior that would have been received when the labeler is presented with the input image data (*predict user responses describing a state of mind of a road user associated with the object*). EX1006, Abstract, ¶¶0025, 0047, 0049, 0051-0065, claims 1, 4, 5, Figs. 4, 5, 9, 10; Section VI.C.2.c (element 1[b]).

185. In my opinion, the behavior (to which predicted human labeler responses relate to) can indicate an intent of a vehicle driver to drive their vehicle into the path of autonomous vehicle 105, such as by making a “left turn” or “right turn” into the vehicle 105’s path, because Zhang’s system identifies predicted trajectories of such vehicles in autonomous vehicle 105’s environment and

determines whether their predicted trajectories conflict with the trajectory of autonomous vehicle 105. EX1006, ¶¶0051, 0057-0061, Figs. 4, 8. Such behavior also indicates an awareness of the driver as to the autonomous vehicle 105, with a predicted trajectory that the driver will turn their vehicle into vehicle 105's path indicating that the driver has a low level of awareness of vehicle 105 or has an awareness of vehicle 105 and believes it can safely make the turn. *Id.*

7. Claim 8

186. In my opinion, Zhang renders obvious claim 8. *See* the analysis of Section VI.C.6 (Claim 7).

8. Claim 9

187. In my opinion, Zhang renders obvious claim 9. *See* the analysis of Section VI.C.6 (Claim 7).

9. Claim 10

188. In my opinion, Zhang renders obvious claim 10 for the same reasons discussed with respect to claim 1, as indicated below.

a. 10[Pre]

189. *See* Section VI.C.2.a (element 1[Pre]). Moreover, Zhang describes that its system is implemented using at least one processor that executes instructions stored in a non-transitory computer readable medium such as a storage device (*[a] non-transitory computer readable storage medium storing instructions that when*

executed by one or more processors, cause the one or more processors to perform the steps). EX1006, ¶¶0003, 0006, 0018-0032, 0036, 0052-0061, 0073, Figs. 1, 5, 6, 9, 11.

b. 10[a]

190. *See* Section VI.C.2.b (element 1[a]).

c. 10[b]

191. *See* Section VI.C.2.c (element 1[b]).

d. 10[c]

192. *See* Section VI.C.2.d (element 1[c]).

e. 10[d]

193. *See* Section VI.C.2.e (element 1[d]).

10. Claim 11

194. In my opinion, Zhang renders obvious claim 11 for the same reasons discussed above in Section VI.C.3 (claim 2).

11. Claim 12

195. In my opinion, Zhang renders obvious claim 12 for the same reasons discussed above in Section VI.C.4 (claim 3).

12. Claim 13

196. In my opinion, Zhang renders obvious claim 13 for the same reasons discussed above in Section VI.C.5 (claim 4).

13. Claim 16

197. In my opinion, Zhang renders obvious claim 16 for the same reasons discussed above in Section VI.C.6 (claim 7).

14. Claim 17

198. In my opinion, Zhang renders obvious claim 17 for the same reasons discussed above in Section VI.C.7 (claim 8).

15. Claim 18

199. In my opinion, Zhang renders obvious claim 18 for the same reasons discussed above in Section VI.C.8 (claim 9).

16. Claim 19

200. In my opinion, Zhang renders obvious claim 19 for the same reasons discussed with respect to claim 1, as indicated below.

a. 19[Pre]

201. *See* Sections VI.C.2.a (element 1[Pre]), VI.C.9.a (element 10[Pre]).

b. 19[a]

202. *See* Section VI.C.2.b (element 1[a]).

c. 19[b]

203. *See* Section VI.C.2.c (element 1[b]).

d. 19[c]

204. *See* Section VI.C.2.d (element 1[c]).

e. 19[d]

205. *See* Section VI.C.2.e (element 1[d]).

17. Claim 20

206. In my opinion, Zhang renders obvious claim 20 for the same reasons discussed above in Section VI.C.6(claim 7).

18. Claim 21

207. In my opinion, Zhang renders obvious claim 20 for the same reasons discussed above in Section VI.C.7 (claim 8).

19. Claim 22

208. In my opinion, Zhang renders obvious claim 20 for the same reasons discussed above in Section VI.C.8 (claim 9).

D. Printed Matter

209. 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.

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

- 1[a], 10[a], 19[a] – “the sensor data displaying an object on the road”
- 1[b], 10[b], 19[b] – “input sensor data displaying a particular object on a road”

Declaration of Dr. Jason Janét Regarding U.S. Patent No. 10,614,344

- 1[b], 10[b], 19[b] – “statistical summary characterizing a distribution of user responses expected to be received responsive to presenting the sensor data to a plurality of users”
- 1[c], 10[c], 19[c] – “statistical summary data characterizing a distribution of user responses expected to be received responsive to presenting the sensor data captured by the sensor installed on the autonomous vehicle to users”
- 4, 13 – “the user responses describe an action associated with the object, the action including one of the object staying in place, changing lanes, or crossing a street”
- 5, 14 – “the parameter including at least one of a content of a response, a time associated with entering a response, and a position of an eye of a human observer associated with the response, the position being measured with respect to a display associated with the user interface”
- 6, 15 – “the parameter is further associated with at least one of a central tendency, a variance, a skew, a kurtosis, a scale, and a histogram”
- 7, 16, 20 – “user responses describing a state of mind of a road user associated with the object”
- 8, 17, 21 – “user responses describing an intention of the road user to move into the path of the autonomous vehicle”

- 9, 18, 22 – “user responses describing awareness of the road user of the autonomous vehicle”

The above limitations each recites a data element and then defines that data element based on its content.

211. 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* 10[pre] and 19[pre]. Rather than impacting how the computer memory operates, the above limitations merely describe the claimed information.

VII. CONCLUSION

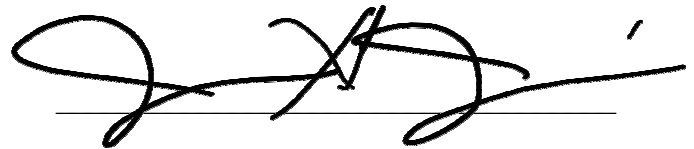
212. It is my opinion that the Challenged Claims be canceled as unpatentable under U.S.C. § 318(b).

Declaration of Dr. Jason Janét Regarding U.S. Patent No. 10,614,344

* * *

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: September 30, 2025

A handwritten signature in black ink, appearing to be 'Jason Janét', written over a horizontal line.

Jason Janét