

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

TESLA, INC.,

Petitioner,

v.

PERCEPTIVE AUTOMATA LLC,

Patent Owner.

Case No. IPR2025-01573

Patent No. 10,614,344

DECLARATION OF DR. ELI SABER, PH.D.

TABLE OF CONTENTS

I.	INTRODUCTION	1
II.	BACKGROUND AND QUALIFICATIONS	2
III.	MATERIALS CONSIDERED AND TIME OF INVENTION	8
IV.	LEGAL STANDARDS	10
A.	Overview	10
B.	Claims.....	11
C.	Burden of Proof for Invalidity.....	12
D.	Claim Construction	12
E.	Person of Ordinary Skill in the Art	13
F.	Prior Art.....	13
G.	Anticipation	14
H.	Obviousness.....	15
V.	PERSON OF ORDINARY SKILL IN THE ART	18
VI.	THE '344 PATENT AND THE CHALLENGED CLAIMS	19
A.	Overview of the '344 Patent.....	19
B.	Claim Overview	22
VII.	CLAIM CONSTRUCTION	24
VIII.	GROUND 1	25
A.	Ross and Cox.....	25
B.	Ross and Cox Do Not Teach Elements 1[b], 10[b], or 19[b] ("providing the sensor data as input to a supervised learning based model, 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”)	31
C.	Ross and Cox Do Not Teach Elements 1[c], 10[c], or 19[c] (“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”)	35
D.	Both the Petition and Dr. Janét’s Declaration Fail to Analyze the Factors Required for a Proper Obviousness Analysis.....	36
E.	A POSITA Would Not Have Been Motivated to Combine Ross and Cox in the Manner Tesla and Dr. Janét Allege	37
IX.	GROUND 2.....	43
A.	Ross, Cox, and Ogale	43
B.	Ross, Cox, and Ogale Do Not Teach Elements 1[b], 10[b], or 19[b] (“providing the sensor data as input to a supervised learning based model, 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”)	46
C.	Ross, Cox, and Ogale Do Not Teach Elements 1[c], 10[c], or 19[c] (“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”)...	49
D.	A POSITA Would Not Have Been Motivated to Combine Ross, Cox, and Ogale in the Manner Tesla and Dr. Janét Allege.....	50
X.	GROUND 3.....	52
A.	Zhang	53

B.	Zhang Does Not Teach Elements 1[b], 10[b], or 19[b] (“providing the sensor data as input to a supervised learning based model, 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”)	55
C.	Zhang Does Not Teach Elements 1[c], 10[c], or 19[c] (“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”)	58
XI.	CONCLUSION	59

I, Eli Saber, Ph.D., hereby declare as follows:

I. INTRODUCTION

1. I have been retained as an independent expert consultant on behalf of Perceptive Automata, LLC (“Patent Owner” or “Perceptive”) in this *Inter Partes* Review proceeding (“IPR”) before the Patent Trial and Appeal Board (“Board” or “PTAB”) of the United States Patent and Trademark Office (“USPTO”) regarding U.S. Patent No. 10,614,344 (the “’344 patent”) (Ex. 1001).

2. I am being compensated for the time I spend on this matter, in addition to expenses. My compensation in this case is in no way contingent on the nature of my findings, the presentation of my findings in testimony, or the outcome of this or any other proceeding. I have no other interest in this proceeding.

3. I understand that Tesla, Inc. (“Petitioner” or “Tesla”) has challenged the patentability of claims 1-22 of the ’344 patent in this proceeding based on certain alleged prior art references identified and discussed in Tesla’s Petition (Paper 1).

4. I have been asked to consider, and provide my opinions regarding, whether the alleged prior art references, as presented in Grounds 1-3 of the Petition, render obvious claims 1-22 of the ’344 patent. In my opinion, the alleged prior art references, as presented in Grounds 1-3 of the Petition, do not render obvious any of the challenged claims.

5. This declaration is based on the information currently available to me. To the extent that additional information becomes available, I reserve the right to continue my investigation and study, which may include review of documents and information that may be produced, as well as testimony from depositions that have not yet taken place. My opinions are set forth below.

II. BACKGROUND AND QUALIFICATIONS

6. I have summarized in this section my educational background, career history, publications, and other relevant qualifications. I understand my full *curriculum vitae* is submitted as Exhibit 2040.

7. I am a Professor in the Department of Electrical and Microelectronic Engineering (EME) at the Chester F. Carlson Center for Imaging Science at the Rochester Institute of Technology (RIT) in Rochester, New York. I also serve as the Director of the Image, Video and Computer Vision Laboratory. I joined RIT's full-time faculty in 2004. I have over thirty-four (34) years of industry and academic experience, thirty-two (32) of which are in the fields of Computer Vision, Digital Image/Video Processing, Pattern Recognition and Machine Learning.

8. I received a Ph.D. in Electrical Engineering from the University of Rochester in 1996. My concentration was in Signal/Image/Video Processing, Pattern Recognition, and Computer Vision. I received a Master of Science Degree in Electrical Engineering from the University of Rochester in 1992 in the same

concentration as listed above, and I received a Bachelor of Science Degree in Electrical and Computer Engineering, Summa Cum Laude, from the State University of New York at Buffalo in 1988.

9. Before becoming a full-time professor, I worked for Xerox Corporation from 1988 to 2004. During my 16 years at Xerox, I was – generally speaking – responsible for: a) delivering color management solutions, image processing innovations, architectures, and algorithms for high-end color printing systems, b) xerographic sub-systems for a variety of color products, and c) control systems for toner production facilities. Among other roles, as an Advanced Development Scientist and Manager, I established the Advanced Design Laboratory – an Imaging/Xerographics lab – and provided technical and managerial leadership for the Electrical, Imaging and Xerographics Department. As a Product Development Scientist and Manager, I led the research and development of color management algorithms and image quality metrics for various product platforms. I also led the Image Science, Analysis, and Evaluation area, with twelve to fifteen direct reports and a budget of approximately \$2 million.

10. From 1997 until 2004, I was an adjunct faculty member at the Electrical Engineering Department of the Rochester Institute of Technology and at the Electrical & Computer Engineering Department at the University of Rochester. I was responsible for teaching undergraduate and graduate coursework in signal,

image, video processing, pattern recognition, and communications and performing research in multimedia applications, pattern recognition, image understanding and color engineering.

11. Since joining RIT full time, I have been responsible for teaching several undergraduate and graduate courses in Digital Signal Processing, Digital Image Processing, Digital Video Processing, Advanced Neural Networks and Deep Learning, Pattern Recognition, Engineering Analysis, Random Signal & Noise, Advanced Engineering Mathematics, Matrix Methods, Pattern Recognition, Communications, Modern Control Theory, and Linear Systems.

12. I proposed, developed and introduced the “Digital Video Processing” and the “Advanced Neural Network and Deep Learning: Part 1” courses; and reestablished the “Pattern Recognition” course at the Electrical Engineering Department at RIT. These graduate courses were instrumental in providing MS/PhD students with the proper foundation and knowledge in all aspects of the subject areas in order to perform in-depth research in the field and continue to advance the state of the art.

13. The Digital Video Processing course is designed to provide a thorough and comprehensive coverage of the field. Topics covered include the mathematical representation of digital video including motion and time-varying image formation models, spatio-temporal sampling/reconstruction, sampling structure conversions,

motion estimation using optical flow, block-based, pel-recursive, Bayesian and point correspondence type methodologies, segmentation, motion tracking, filtering, restoration, super-resolution and compression.

14. The Pattern Recognition course covers pertinent concepts in Bayes decision theory, linear and nonlinear classifiers, artificial neural networks, feature selection/generation/extraction and various types of clustering algorithms and techniques.

15. The Advanced Neural Networks and Deep Learning: Theory, Mathematics and Algorithms course is intended to provide an in-depth comprehensive study of the theory, mathematics and algorithms for advanced neural networks and their usage in deep learning. It is primarily targeted for upper graduate students that are planning to pursue advanced studies/research in the field. The course begins with an in-depth study of the mathematical methodologies for neural networks and optimization (zero-, first- and second-order) techniques. It then progresses to provide a thorough study of shallow and deep neural networks architectures and algorithms, activation functions, regression, two- and multi-class classification, unsupervised learning, feature engineering/selection, backpropagation along with applications to a variety of engineering problems.

16. I am currently working to develop a follow-on course (Part 2) to the Advanced Neural Networks and Deep Learning course described above. This course

is intended to cover convolutional and recurrent style neural networks, principles of feature engineering, kernel methods, tree-based learners, attention mechanisms/transformers, reinforcement learning and generative adversarial networks. This course is targeted to debut in Fall 2027.

17. I am also currently working to revamp the graduate course titled “Image and Video Compression”. The updated version of this course will start by providing a brief history of video coding and video coding standards and then progress to discussing video basics, formats and fundamentals of video compression. It will then progress to cover JPEG, JPEG 2000, H.261, MPEG-1, MPEG-2, H.263, MPEG-4 visual, H.264 and H.265 and if time permits, AV-1 and H.266. The course is intended to debut in Fall 2028.

18. My current research focuses on machine learning for object tracking/remote sensing applications; image/video processing for multimedia, military & biomedical applications; and computer vision and three-dimensional scene reconstruction As a principal investigator (PI) or co-principal investigator, I have acquired almost \$6 million in research funding since joining RIT and have managed multiple government grants from the Department of Defense as well as several corporate grants from Hewlett-Packard, Lenel, and Data Physics. I am currently managing (as PI) a multiyear government grant from the Department of Defense titled “Target Detection/Tracking and Activity Recognition from

Multimodal Data.” In this project, we propose to develop operational target detection/tracking techniques and activity recognition/understanding methods that can leverage multimodal data in a fusion framework using deep learning algorithms while providing accurate and near real-time performance.

19. In 2012, I was awarded the Prestigious Trustees Scholarship, the highest award at RIT with regard to research recognition.

20. I am the author or co-author of forty (40) peer-reviewed journal publications.

21. I have also authored or co-authored 103 conference and workshop publications and a book entitled Advanced Linear Algebra for Engineers with MATLAB, published by CRC Press in February 2009, and I am a named inventor on multiple U.S. and foreign patents.

22. I am a senior member of the Institute of Electrical and Electronic Engineers and a member of the IEEE Signal Processing Society, the Electrical Engineering Honor Society, and Eta Kappa Nu.

23. With respect to the technologies relevant to the '344 Patent, I have extensive experience in developing computer vision and digital image/video processing algorithms and techniques for motion estimation, segmentation, registration, classification, identification, pattern recognition, shape recognition/matching, stereo imaging, facial recognition, biomedical image

processing, compression, surveillance systems, 3d scene reconstruction, machine learning, deep learning for a variety of applications including object recognition and tracking, video analysis, video surveillance, change detection, biomedical, color engineering, grayscale and color document processing.

24. A complete statement of my industrial and academic records including a listing of the above publications and patents is included with my *curriculum vitae*.

III. MATERIALS CONSIDERED AND TIME OF INVENTION

25. The opinions contained in this declaration are based on the documents that I reviewed, my professional judgment, as well as my education, experience, and knowledge regarding technologies related to the '344 patent.

26. In forming my opinions expressed in this declaration, I considered the following materials:

<u>Document</u>	<u>Exhibit/Paper #</u>
'344 patent	1001
File history for the '344 patent	1002
Declaration of Dr. Jason Janét	1003
U.S. Patent No. 10,496,091 ("Ross")	1004
International Publication WO 2014/210334 ("Cox")	1005
U.S. Patent Application Publication No. 2019/0072966 ("Zhang")	1006

U.S. Patent No. 10,733,506 (“Ogale”)	1010
Petition	1
Patent Owner’s Preliminary Response	8
Kooij JFP, Schneider N, Flohr F, Gavrilă DM (2014) Context-based pedestrian path prediction. In: Proceedings of the European Conference on Computer Vision (ECCV). Berlin: Springer, pp. 618–633	2041
A. Zyner, S. Worrall, J. Ward and E. Nebot, "Long short term memory for driver intent prediction," 2017 IEEE Intelligent Vehicles Symposium (IV), Los Angeles, CA, USA, 2017, pp. 1484-1489, doi: 10.1109/IVS.2017.7995919	2042

27. I have been asked to consider the time of the inventions of the challenged claims to be around December 4, 2017, which I understand is the filing date of the earliest non-provisional patent application to which the ’344 patent claims priority. Ex. 1001, Cover. My opinions reflect how a person of ordinary skill in the art (as I define in Section V below) would have understood the claims and the specification of the ’344 patent, the prior art to the patent, and the state of the art, at the time of the alleged invention.

28. I understand that the '344 patent also claims priority to U.S. Provisional Patent Application No. 62/528,771, which was filed on July 5, 2017. My opinions herein do not differ whether the time of the inventions of the challenged claims is around July 5, 2017 or December 4, 2017. In addition, for the sake of clarity, I have been asked to assume that all of the references relied upon by Tesla in support of its IPR petition are prior art to the challenged claims of the '344 patent.

29. As I discuss in detail below, it is my opinion that claims 1-22 of the '344 patent are not rendered obvious by the alleged prior art references, as presented in Grounds 1-3 of the Petition.

IV. LEGAL STANDARDS

A. Overview

30. I am not an attorney or a legal expert. The legal standards I applied in reaching the conclusions expressed throughout this declaration were provided to me by counsel. In preparing this declaration, I have endeavored to faithfully apply these legal standards. I do not intend to offer opinions on the law. Below, I describe the law I applied in considering whether the claims of the '344 patent are patentable.

31. I have been informed that an issued patent is presumed valid and that the challenger of the patent's validity bears the burden of proving invalidity of the patent.

B. Claims

32. I have been informed that patent claims describe the inventions made by the inventors and describe what the patent owner owns and what the owner may prevent others from doing.

33. I have been informed that claims may describe products, such as machines or chemical compounds, or processes for making or using a product. I have been informed that claims are usually divided into parts or steps called limitations.

34. I have been informed that there are two types of claims: independent claims and dependent claims. I have been informed that an independent claim sets forth all the requirements that must be met in order to be covered by that claim. I have been informed that it is not necessary to look at any other claim to determine what an independent claim covers. I have been informed that a dependent claim does not itself recite all of the requirements of the claim, but refers to another claim for some of its requirements. I have been informed that, in this way, the claim depends on another claim. I have been informed that the law considers a dependent claim to incorporate all of the requirements of the claims to which it refers.

35. I have been informed that the dependent claim then adds its own additional requirements. I have been informed that to determine what a dependent claim covers, it is necessary to look at both the dependent claim and any other claims to which it refers.

36. I have been informed that the dependent claims incorporate the limitations of the independent claim upon which they depend and, if a reference fails to disclose any of the limitations of an independent claim, the claims which depend on that claim cannot be made invalid by that reference. I have been informed that, as a result, any of the references at issue in this proceeding that fail to disclose all of the limitations of an independent claim necessarily also fail to disclose all of the limitations of claims which depend on that claim.

C. Burden of Proof for Invalidity

37. I understand that the party challenging the validity of an issued U.S. patent in an IPR bears the burden of proving invalidity through facts supported by a preponderance of the evidence. I understand that a “preponderance of the evidence” means evidence sufficient to prove that a fact is more likely true than it is not.

D. Claim Construction

38. I have been informed that the first step in determining validity is to properly construe the claims. I have been informed that, in this proceeding, the Board must determine the scope of the claims by using the same claim construction standard that United States district courts use. I have been informed that, under this standard, words of a claim are given their ordinary and customary meaning as understood by a person of skill in the art at the time of the invention. I have been informed that such a person looks to the claims, specification, and file history in

understanding the meaning of the claim terms. I have applied this understanding in this declaration.

E. Person of Ordinary Skill in the Art

39. I have been informed that a person having ordinary skill in the art may be defined by factors such as: (a) the levels of education and experience of the inventor and other persons actively involved in the field; (b) the types of problems encountered in the field; (c) prior art solutions to those problems; (d) rapidity with which innovations are made; and (e) the sophistication of the technology.

F. Prior Art

40. I have been informed that a patent cannot be invalidated based on one person's testimony alone without corroborating evidence, particularly documentary evidence. Moreover, I have been informed that corroboration is required of any witness whose testimony alone is asserted to invalidate a patent, regardless of his or her level of interest in relation to that patent. I have been informed that, accordingly, even if the testifying individual has no interest in the litigation, corroboration is still required.

41. In this declaration, I may use the phrase "prior art" when describing specific references. However, my use of this phrase should not be interpreted as agreement that the item(s) described qualify as prior art under the patent laws. I merely use this phrase to address the items labeled as "prior art" by Tesla and its

expert. As I mention above, I have been asked to assume that the references Tesla relies upon in support of this IPR petition are prior art to the challenged claims of the '344 patent.

G. Anticipation

42. I have been informed that a patent claim can be found invalid as anticipated if each and every element of the claim, as properly construed, is found either explicitly or inherently in a single prior art reference. I have been informed that, if the single prior art reference does not disclose any one claim limitation, then that claim is not anticipated. I have been informed that the single prior art reference cannot merely disclose each element; rather, it must disclose all of the elements as arranged in the claim. I have been informed that for a prior art reference to “inherently” disclose something, the reference must necessarily function in accordance with, or include the claimed limitations, to anticipate. I have been informed that the missing descriptive matter cannot merely probably or possibly include the claimed limitations. I have been informed that, with respect to an anticipation analysis, it is improper to combine separate, distinct embodiments of a single prior art reference.

43. I have been informed that for a claim to be anticipated, each claim limitation must be disclosed in a single prior art reference, and the claimed arrangement or combination of those limitations must also be disclosed in that same

prior art reference. I have been informed that for a reference to be anticipatory, the reference must contain an enabling disclosure. I have been informed that merely naming or describing the subject matter is insufficient if it cannot be produced without undue experimentation.

H. Obviousness

44. I have been informed that a patent claim can be found invalid as obvious if the differences between the claim and the prior art references would have been obvious to a person of ordinary skill in the art at the time of the invention, taking into account (1) the scope and content of the prior art, (2) the differences between the prior art and the claims, (3) the level of ordinary skill in the art, and (4) any objective indicia of non-obviousness. I further understand that a claim can be found to be obvious in light of either a single prior art reference or multiple prior art references. I understand that references qualify as prior art for an obviousness determination only when they are analogous to the claimed invention. I understand that two separate tests define the scope of analogous prior art: (1) whether the art is from the same field of endeavor, regardless of the problem addressed; and (2) if the reference is not within the field of the inventor's endeavor, whether the reference still is reasonably pertinent to the particular problem with which the inventor is involved.

45. I understand that to be obvious in light of a single prior art reference or multiple prior art references, there must be a reason either to modify either the single prior art reference or to combine the two or more prior art references to achieve the claimed invention. I understand that this reason may come from a teaching, suggestion, or motivation to combine or may come from the reference or references themselves, the knowledge or “common sense” of one skilled in the art, or from the nature of the problem to be solved, and may be explicit or implicit from the prior art as a whole.

46. The combination of familiar elements according to known methods is likely to be obvious when it does no more than yield predictable results.

47. I further understand that there is no suggestion to combine, however, if a reference teaches away from its combination with another source, and that a reference may be said to teach away when a person of ordinary skill, upon reading the reference, would be discouraged from following the path set out in the reference, or would be led in a direction divergent from the path that was taken by the applicant, or if it suggests that the line of development flowing from the reference’s disclosure is unlikely to be productive of the result sought by the applicant. If when combined, the references would produce a seemingly inoperative device, then they teach away from their combination. I understand that it is improper to combine references where the references teach away from their combination.

48. I understand that the degree of teaching away will depend on the particular facts; in general, a reference will teach away if it suggests that the line of development flowing from the reference's disclosure is unlikely to be productive of the result sought by the applicant. I understand that a reference teaches away, for example, if (1) the combination would produce a seemingly inoperative device, or (2) the references leave the impression that the product would not have the property sought by the applicant.

49. I understand that if the teachings of a prior art reference would lead one skilled in the art to make a modification that would render another prior art device inoperable, then such a modification would generally not be obvious. I understand that if a proposed modification would render the prior art device unsatisfactory for its intended purpose, then the proposed modification would generally not be obvious.

50. I understand that it is improper to use hindsight in making the obviousness determination.

51. I understand that some articulated reasoning with some rational underpinning to support the legal conclusion of obviousness must be made. For example, I understand that there may be no rational underpinning to support the legal conclusion of obviousness where systems are already complete and the party

challenging validity seeks to mix and match components, which each exist in the separate systems, from these complete systems.

V. PERSON OF ORDINARY SKILL IN THE ART

52. I understand that Tesla's expert (Dr. Janét) has opined that:

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.

Ex. 1003, ¶ 69.

53. For purposes of providing my opinions in connection with this declaration, I do not dispute Dr. Janét's definition of a POSITA. However, I note that Dr. Janét offers this definition as of July 5, 2017, even though he opines that the challenged claims of the '344 patent are not entitled to a priority date earlier than December 4, 2017. *See* Ex. 1003, ¶ 58. I have applied Dr. Janét's definition in rendering my opinions herein. I believe that my education and experience qualify me as a person of ordinary skill in the art relevant to the '344 Patent as of either July

5, 2017 or December 4, 2017 under either prong of Dr. Janét’s definition. *See* Section II above and my CV.

VI. THE ’344 PATENT AND THE CHALLENGED CLAIMS

A. Overview of the ’344 Patent

54. The ’344 patent issued on April 7, 2020, from U.S. Patent Application No. 16/512,560, filed July 16, 2019. *See* Ex. 1001. The ’344 patent is a continuation of U.S. Patent Application No. 15/830,549, filed on December 4, 2017, and claims priority to U.S. Provisional Patent Application No. 62/528,771, filed July 5, 2017.

55. The ’344 patent describes and identifies certain problems with autonomously operating a car. As the ’344 patent states, “[t]he ability a driver of a car to look at a person—who is walking, driving another car, or riding a bike on or near a street—and predict what that person wants to do may be the single most important part of urban driving.” ’344 patent, 1:21-24. “For example, when a driver of a car sees people near the car, determining whether one person will cross the street, whether another person will remain standing on a street corner, and whether yet another person will change lanes on his or her bicycle is necessary to safely drive the car and avoid hitting the people.” *Id.* at 1:24-29. “This ability is so fundamental, that operating in cities without it would be nearly impossible.” *Id.* at 1:30-31.

56. The ’344 patent describes certain pre-existing methods that fail to adequately address these shortcomings of autonomous vehicles, including that

“autonomous driving vehicles may rely on methods that make decisions on how to control the vehicles by predicting ‘motion vectors’ of people near the vehicles.” *Id.* at 1:38-40. As the ’344 patent indicates, the conventional methods accomplished this “by collecting data of a person’s current and past movements, determining a motion vector of the person at a current time based on these movements, and extrapolating a future motion vector representing the person’s predicted motion at a future time based on the current motion vector.” *Id.* at 1:41-46. As the ’344 patent recognizes, “[h]owever, the [conventional] methods do not predict a person's actions or movements based on other observations besides his or her current and past movements, which lead to inferior results in predicting the person's future behavior.” *Id.* at 1:46-49.

57. The ’344 patent specifically describes and claims novel methods, non-transitory computer readable storage media, and systems which address the shortcomings of then-existing autonomous vehicle systems. For example, the ’344 patent discloses an embodiment where summary statistics of user responses can be used to train a learning based model so that the learning based model can predict whether an object in the road (e.g., pedestrian) in an image has a given “state of mind—e.g. that a pedestrian is likely to cross the street or that an oncoming vehicle is unlikely to be willing to yield to the vehicle containing the sensor if the vehicle containing the sensor needs to turn.” *Id.* at 10:19-38; *see also id.*, 6:37-7:19, 11:5-

12:20. “The trained model or algorithm makes a prediction of what a pedestrian or other person shown in the ‘live data’ would do based on the summary statistics and/or training labels” *Id.* at 7:20-23; *see also id.*, 12:21-13:48. “The predictions produced by the trained model comprise a set of predictions of the state of mind of road users that can then be used to improve the performance of autonomous vehicles, robots, virtual agents, trucks, bicycles, or other systems that operate on roadways by allowing them to make judgments about the future behavior of road users based on their state of mind.” *Id.* at 7:32-39.

58. For example, claim 1 of the ’344 patent recites “[a] computer-implemented method comprising” the steps of “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, the sensor data displaying an object on the road” and “providing the sensor data as input to a supervised learning based model.”

59. As claim 1 further explains “the supervised learning based model [is] configured to” perform certain functions, *i.e.*, “receive an input sensor data displaying a particular object on a road” and “predict[] 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.” For example, given an image of a pedestrian at an intersection

an autonomous vehicle is approaching, an embodiment of the claimed model is configured to predict an output statistical summary characterizing a distribution of what users expect the pedestrian will do (*e.g.*, enter the crosswalk in front of the vehicle, or wait until the vehicle passes).

60. Claim 1 further provides for “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.”

B. Claim Overview

61. Independent claims 1, 10, and 19 are reproduced below. In this table, I have included the claim element numbering used in Tesla’s IPR petition and Dr. Janét’s declaration for the sake of continuity.

Claim 1	Limitations
1[pre]	A computer-implemented method comprising:
1[a]	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, the sensor data displaying an object on the road;

1[b]	providing the sensor data as input to a supervised learning based model, 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;
1[c]	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
1[d]	controlling the operation of the autonomous vehicle on the road based on the generated statistical summary data.
Claim 10	Limitations
10[pre]	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 comprising:
10[a]	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, the sensor data displaying an object on the road;
10[b]	providing the sensor data as input to a supervised learning based model, 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;
10[c]	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
10[d]	controlling the operation of the autonomous vehicle on the road based on the generated statistical summary data.
Claim 19	Limitations
19[pre]	A computing system comprising: one or more processors; and a non-transitory computer readable storage medium, storing instructions that when executed by the one or more processors, cause the one or more processors to perform steps comprising:
19[a]	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, the sensor data displaying an object on the road;
19[b]	providing the sensor data as input to a supervised learning based model, 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;
19[c]	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
19[d]	controlling the operation of the autonomous vehicle on the road based on the generated statistical summary data.

VII. CLAIM CONSTRUCTION

62. In my opinion, none of the terms of the '344 patent require any specific construction in order to resolve the issues presented in Tesla's petition and in my opinions as expressed below. In rendering the opinions set forth herein, I have applied the plain and ordinary meaning of all terms of the challenged claims to a person of ordinary skill in the art.

VIII. GROUND 1

63. I understand that Tesla alleges (and Dr. Janét opines) that claims 1-22 of the '344 patent are obvious over Ross in view of Cox. *See* Pet. at 18-47; Ex. 1003, ¶¶ 71-140. For the reasons I discuss below, I disagree. In my opinion, none of claims 1-22 are obvious over Ross in view of Cox.

A. Ross and Cox

64. In arriving at my opinions concerning Tesla and Dr. Janét's assertions regarding claims 1-22 of the '344 patent and Ross and Cox, I have considered Ross and Cox's disclosures from the perspective of a POSITA. Below I discuss some aspects of Ross and Cox that are relevant to my opinions.

65. Ross describes "[b]ehavior and intent estimations of road users for autonomous vehicles", Ex. 1005, Title which represents an entirely different application than Cox face/no face detection system. Ross explains that its "disclosure provides a method of maneuvering a vehicle. The method includes receiving, by one or more processors, data from a perception system of the vehicle identifying an

object corresponding to a road user, the information also identifying characteristics of the road user as well as contextual information about an environment in which the vehicle is currently driving; generating, by the one or more processors, a prediction of the intent of the object using the data, wherein the prediction of the intent of the object includes a next action to be initiated by the road user and a point in time for initiation of the next action, wherein the next action is included in a predetermined list of actions; generating, by the one or more processors, a prediction of the behavior of the road user for a predetermined period of time into the future using the data, the prediction of the behavior indicating a likelihood that the road user will initiate the next action during the predetermined period of time, the predetermined period of time including at least a period of time during which the vehicle is expected to have passed the object; and when the prediction of the behavior indicates that the road user is not likely to initiate the next action during the predetermined period of time, maneuvering, by the one or more processors, the vehicle according to the prediction of the intent prior to the vehicle passing the object.” Ex. 1005, 1:30-54. Figure 10 of Ex. 1005 provides an example flow diagram in accordance with aspects of the proposed disclosure.

66. Ross does not disclose or mention any details with regards to summary statistics, probability distributions of users, user responses associated with the object and/or annotations by human observers.

67. As Cox indicates, its “invention relates to machine learning, in particular to machine learning augmented by measurements acquired from human populations.” Ex. 1007, [0001].

68. Cox goes on to explain that “[t]raditional machine-learning algorithms typically incorporate human knowledge (at least implicitly) as an input.” *Id.*, [0002]. Cox gives the specific example of “a machine-learning system trained to decide whether a face is present in an image will be given set of human-generated ‘ground truth’ labels that signal whether a human judged the image to contain a face or not.” *Id.* “In this case, the job of the machine-learning system is to emit analogous labels (for example, ‘face present’ and ‘face not present’) in response to new, previously unseen images.” *Id.* I note that Cox uses this diametric “face present”/“face not present” example throughout its disclosure. *Id.* at [0002], [0008], [0010], [0012], [0013], [0016], [0017], [0026], [0029], [0033], [0039]-[0048], [0050], [0052]-[0056], Figs. 5, 6, 7, 9, 10, 11, 12 (Cox’s disclosures and experiments all describe methodologies related to face detection). In contrast, the challenged claims of the ’344 patent recite determining “a state of mind of a user” in images and are concerned with training and executing a model to predict what road users will do in a given scene (i.e., image or video). While Cox contains a couple passing references to an annotator being asked to identify an emotion that corresponds to the image of

a face (*see* Ex. 1007, [0016], [0049]), nothing in Cox describes or suggests using such emotions to train a model to predict what road users will do in such a scene.

69. As Cox also explains, one limitation of machine learning at the time was “the unavailability of labeled data; it is difficult to provide an algorithm with enough labeled training data to achieve optimal performance.” Ex. 1007, [0003]. As Cox explains, “[w]ithout sufficiently large datasets, machine-learning algorithms tend to ‘overfit’ the data, adapting to spurious structure present in the training set that is not representative of the larger distribution of all examples in the real world.” *Id.* Cox goes on to state that “[m]achine-learning systems typically combat the effects of overfitting by a process called “regularization,” in which penalties are placed on solutions that are thought to be more likely to be the result of overfitting, typically because they are more complex or because they exhibit less stable behavior under injected noise.”, *Id.*

70. To address the above, Cox proposes imposing the use of additional weightings in the objective function of the machine-learning algorithm as a form of regularization to produce more “human-like” solutions via the introduction and use of penalties for misclassification tasks on the part of the machine-learning system for examples that are easily classified by humans. Ex. 1007, [0005]. As a result, Cox proposes the acquisition and use of “[p]sychometric data characterizing the annotation of the training objects by the annotators (i.e. human annotators)”. Ex.

1007, [0007]. To this effect, Cox describes using “psychometric data” as an additional input in the machine learning system, in addition to the human annotations of the image. *See id.* at [0005]-[0010]. As described by Cox, “[t]he psychometric data may include or consist essentially of (i) response time for classifying one or more features, (ii) accuracy of feature classification, and/or (iii) presentation time of one or more training objects.” Ex. 1007, [0008], [0010]. In addition, “[e]ach of the training objects may include or consist essentially of a digital image, and one or more features for classification may include or consist essentially of human faces.” *Id.*, [0008], [0010]. As Cox describes, this information can be used to determine the degree of “penalization” for misclassifying the image (i.e., indicating that the image contains a face when it does not, or vice versa), because “greater penalties may be introduced for misclassification . . . of examples that are easily classified by humans” and “lesser penalties may be introduced for misclassification of examples that are relatively more difficult for humans to classify.” *See id.*, [0005]. These penalties are used in a “human-weighted loss function” to train the model, with the aim of improving its accuracy. *See id.*, [0005]-[0010]. As Cox explains, “[t]he penalization module is executable by the computer processor and is configured to derive a human-weighted loss function based at in part on the classification data in the database and the psychometric data of at least some of the training objects in the database. The loss function includes penalties for misclassification, and the magnitudes of the

penalties increase with increasing deviation from the classification data. The classification module is executable by the computer processor and is configured to classify features of query objects based at least in part on the human-weighted loss function.” Ex. 1007, [0009].

71. Cox goes on to explain that “[o]ne component of a human-regularized maximum-margin kernel machine is the addressing of the issue of the loss function.” Ex. 1007, [0035]. Cox further notes that “[a] prediction during training may be calculated as the output of the classifier for a particular training sample multiplied by its label: $z = yf(x)$.” *Id.* “In accordance with various embodiments of the invention, a loss function that applies a linearly increasing cost for misclassifications (one-sided error) is desirable.” *Id.* “This is embodied by the hinge loss function, which is defined as: $\phi(z) = \max(0, 1 - z)$.” *Id.*

72. Cox further explains that given “the non-linear nature of psychometric curves for visual recognition tasks suggests a model that is much different than linear loss growth when $z < 1$. Therefore, embodiments of the invention assume an alternative model where $wt(x)$ is a weighting function conditioned on a psychometric model for a recognition task t . The weights are determined by the classification data collected from the annotators 160, including the individual and combined statistics of accuracy, response time and presentation time, over a sampling of images across varying degrees of difficulty. Human-weighted loss is thus defined as: $\phi(z) =$

$\max(0, w_t(z))$ ”, Ex. 1007, [0036]. It should be noted that both the hinge loss function and the proposed human-weighted loss function are designed for binary classifications as further disclosed in Cox (see [0047]-[0049] and [0052]-[0055]) and illustrated in figures 4A/4B, 5, 6A/6B, 7A/7B, 8A/8B, 9-12.

B. Ross and Cox Do Not Teach Elements 1[b], 10[b], or 19[b] (“providing the sensor data as input to a supervised learning based model, 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”)

73. In my opinion, Ross and Cox do not disclose elements 1[b], 10[b], or 19[b], which each require “providing the sensor data as input to a supervised learning based model, 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.”

74. At the outset, Tesla and Dr. Janét allege that the “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” limitation “may be interpreted as occurring during operational use or during training.” Pet. 30; Ex. 1003, ¶ 96. I note

that neither Tesla nor Dr. Janét offer any construction for this limitation. Nevertheless, I have considered both of their theories for how this limitation is allegedly met, and each fails.

75. Tesla and Dr. Janét first allege that Ross and Cox teach this element during operational use because “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.’” Pet. 30; Ex. 1003, ¶ 97.

76. Ross does not teach or suggest how the “set of possible intents or hypotheses” or “associated likelihood values” are generated or what form they take. *See* Pet. at 30-31; Ex. 1003, ¶ 97. And, neither of these items constitutes a “distribution of user responses” as required by the challenged claims because neither the set of intents/hypotheses or the likelihood values is a distribution of user responses as claimed.

77. Neither the Petition nor Dr. Janét’s declaration alleges that Ross teaches or suggests that its “set of possible intents or hypotheses” or “associated likelihood values” are a distribution of any kind. *See* Pet. 30-31; Ex. 1003, ¶¶ 97-98. Instead, the Petition and Dr. Janét offer two theories for how this element is supposedly met in view of Cox’s teachings.

78. First, Tesla and Dr. Janét allege that, in view of Cox, “the Ross predicted data . . . 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.” Pet. at 32; Ex. 1003, ¶ 99. Tesla and Dr. Janét allege that this is the case because Cox aims to have its model “make predictions . . . that are more consistent with the decisions of the human annotators.” Pet. 31 (citing Ex. 1005, [0032]).

79. However, merely being “consistent with” such decisions is not the same thing as “predicting an output statistical summary characterizing a distribution of user responses,” as required by the challenged claims. A prediction can be consistent with a training dataset by merely reflecting the same classification as the training dataset, without in any way being “an output statistical summary characterizing a distribution of user responses.” The Petition and Dr. Janét admit that Ross does not disclose this element, but they do not point to anything in Cox that teaches “predicting an output statistical summary characterizing a distribution of user responses,” nor is there any such teaching in Cox based on my review of Cox’s disclosure.

80. Second, “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.” Pet. 33. This identification of prediction errors and optimization (using a loss function) merely amounts to a general description of the training process for the model, and nowhere do Tesla or Dr. Janét identify what in this process involves “predicting an output statistical summary characterizing a distribution of user responses,” as required by the challenged claims.

81. I have considered the portions of Ross and Cox that the Petition and Dr. Janét’s declaration cite in support of their argument, and in my opinion, none of them support the allegation that somehow either Ross or Cox teach “predicting an output statistical summary characterizing a distribution of user responses.” *See* Pet. at 30-31; Ex. 1003, ¶ 97 (both citing Ex. 1007, “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”). Instead, those portions, as discussed above, merely describe a model making predictions more consistent with the decisions of human annotators.

82. The portions of the Petition and Dr. Janét’s declaration discussed above relate to claim 1 of the ’344 patent. Both the Petition and Dr. Janét’s declaration merely refer back to these discussions when discussing element 10[b] and 19[b]. *See* Pet. at 44, 46; Ex. 1003, ¶¶ 121, 135. In my opinion, Cox fails to teach or suggest elements 10[b] and 19[b] for the same reasons as discussed above.

83. Therefore, in my opinion, Cox does not teach or suggest elements 1[b], 10[b], or 19[b] (“providing the sensor data as input to a supervised learning based

model, 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”).

C. Ross and Cox Do Not Teach Elements 1[c], 10[c], or 19[c] (“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”)

84. In my opinion, Cox does not disclose elements 1[c], 10[c], or 19[c], which each require “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.”

85. Tesla and Dr. Janét allege that Ross and Cox teach this element for the same reasons they allege with respect to element 1[b]. *See* Pet. at 34-35; Ex. 1003, ¶¶ 103-104. However, as I discuss above with respect to that element, neither Ross nor Cox teaches or suggests “generat[ing] 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,” as required by both elements 1[b] and 1[c].

86. I have considered the portions of Ross and Cox that Tesla and Dr. Janét identify with respect to this claim element. *See* Pet. at 34-36; Ex. 1003, ¶¶ 103-105. None of them alter my opinions expressed above that neither Ross nor Cox teaches or suggests “generat[ing] a statistical summary data characterizing a distribution of user responses.”

87. The portions of the Petition and Dr. Janét’s declaration discussed above relate to claim 1 of the ’344 patent. Both the Petition and Dr. Janét’s declaration merely refer back to these discussions when discussing element 10[c] and 19[c]. *See* Pet. at 44, 46; Ex. 1003, ¶¶ 122, 136. In my opinion, Cox fails to teach or suggest elements 10[c] and 19[c] for the same reasons as discussed above.

88. Therefore, in my opinion, Cox does not teach or suggest elements 1[c], 10[c], or 19[c] (“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”).

D. Both the Petition and Dr. Janét’s Declaration Fail to Analyze the Factors Required for a Proper Obviousness Analysis

89. As I discuss in Section IV above, one of the factors I understand that is required to be considered in an obviousness analysis is the differences between the prior art and the claims. While I note that Dr. Janét mentions this requirement in his declaration as well (*see* Ex. 1003, ¶ 40), nowhere in his discussion of Ground 1 does

Dr. Janét discuss any differences between Ross and/or Cox and the challenged claims of the '344 patent. And, nowhere do Tesla or Dr. Janét discuss any modification(s) to Ross in view of Cox.

90. As I discussed above, Cox's disclosure is focused on binary face detection processes, and many of its features are specifically geared toward that objective, including its use of a human-weighted loss function as I explain above. In contrast, the challenged claims of the '344 patent relate to training and executing a machine learning model to predict summary statistics that describe the state of mind of a road user in an image. Dr. Janét does not account for these key differences between Cox and the challenged claims. As I discuss above, these differences are not miniscule or insignificant because they are rooted in the specific (but differing) applications that Cox and the challenged claims are each aimed to address, and they are a key factor in the opinions I express above concerning the elements of the challenged claims that Cox does not teach or suggest.

E. A POSITA Would Not Have Been Motivated to Combine Ross and Cox in the Manner Tesla and Dr. Janét Allege

91. In my opinion, a POSITA would not have been motivated to combine Ross and Cox in the manner Tesla and Dr. Janét suggest.

92. Tesla and Dr. Janét allege the following:

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 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.

Pet. at 20-21; Ex. 1003, ¶ 80.

93. Nowhere do Tesla or Dr. Janét describe how a POSITA would have accomplished either of these proposed modifications of Ross in view of Cox. Dr. Janét’s “analysis” of the alleged motivation to combine Cox and Ross spans only a few conclusory paragraphs with no supporting analysis.

94. In my opinion, a POSITA would not have been motivated to combine Ross and Cox in the manner Tesla and Dr. Janét allege. The first alleged motivation is that “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.” Pet. at 20. However, Ross already describes its own training process for its model:

[T]o train a model to generate likelihoods for these intents, rather than using a fixed time horizon and observing what the object does during that fixed time horizon, the training data may look as far into the future as it takes for the road user to perform any action included in the list of predetermined actions. Then, the next action included in the predetermined list of actions that the road user ends up performing is

marked as the correct intent from the object's original position (when the intent would be determined).

Ex. 1004, 4:51-59; *see also id.*, 9:66-10:10 (describing the same process).

95. Ross's disclosure does not make clear if this "mark[ing]" of the "next action" "as the correct intent" is performed by a human annotator or the system. In either scenario, however, a POSITA would not have been motivated to combine Ross with Cox.

96. If this marking is performed by a human annotator, then Ross already discloses using human annotated data in the model training process, and a POSITA would not have been motivated to combine Ross with Cox to add a feature already present in Ross. If, on the other hand, this marking in Ross is performed by the system, there would still not have been any motivation for a POSITA to modify Ross to use its human annotated training data in addition to or in lieu of Ross's training data. As Ross describes, its training data "may look as far into the future as it takes for the road user to perform any action included in the list of predetermined actions." Ex. 1004, 4:54-56. Thus, instead of giving a human annotator an image (e.g., of a pedestrian at an intersection) and asking him or her to make a prediction as to what an object in the image will do (e.g., walk in front of the autonomous vehicle or not), Ross describes using image/video data that looks into the future to see what the object *actually* did in that scenario. A POSITA would not have been motivated to substitute that type of certainty as to the outcome of that scenario with a human

annotator's prediction based on an image of the scene. Moreover, since Ross already describes its own training process,

97. Tesla and Dr. Janét's second alleged motivation to combine Ross and Cox is that "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." Pet. at 20-21. Notably, Tesla and Dr. Janét do not explain what this "further configur[ation]" would entail, as neither Tesla nor Dr. Janét detail any modifications to Ross as part of this alleged motivation to combine. While Ross does not expressly state the purpose of training the model with this data, a POSITA would have known that the purpose of using annotated training data to train Ross's model (or any machine learning model) is to make the model's predictions consistent with the annotations in the training data. Thus, a POSITA would not have been motivated to combine Ross with Cox for this reason, nor did Tesla and Dr. Janét provide any technical details with rational underpinnings as to how this combination would take place.

98. In addition, a POSITA would not have been motivated to combine Ross and Cox because their approaches are fundamentally different. Ross relates to autonomous vehicle intent prediction but uses a fundamentally different approach than Cox. Ross does not use crowd-sourced human annotations to train its models like Cox does, nor does Ross propose the use of any type of loss function for its

training purposes. Instead, Ross's training data (while still annotated) is based on observations of actual road user behavior over time. *See, e.g.*, Ex. 1004, 4:25-32, 4:51-65. Cox, on the other hand, proposes the use of a human-weighted loss function for its binary classifications as described in detail above. In my opinion, this fundamentally different approach means that a POSITA would not have been motivated to combine Cox and Ross in the manner Tesla and Dr. Janét suggest.

99. Supervised machine learning – as described and primarily utilized in Cox for face classification – is a type of machine learning in which a model is trained using specifically labeled data, where each input is paired with a known and appropriately labeled output or target value. The model learns the relationship between the training inputs and outputs by minimizing prediction errors during training via the use of properly designed loss functions, enabling it to generalize and make accurate predictions on new, unseen data for classification purposes as disclosed in Cox. Hence, the pairing of the training data (input-output pair) as well as the design of the network architecture (i.e. the number of inputs and the input layer, the number of hidden layers, the number of nodes per layer, the type of activation function, and the type of output produced by the network, etc.), the regularization methodology and the selection of the loss function are generally unique to the application at hand.

100. Using Cox’s crowd-sourced psychometric annotation framework with Ross’s autonomous vehicle intent prediction requires hindsight reconstruction of the claimed invention. Cox is generally concerned with improving the ability of a machine learning model to predict certain classifications regarding human faces—primarily whether a face is or is not present, but also tangentially with other questions such as the gender of the subject in the image or an emotion conveyed by a face in the image. Ex. 1005, [0016], [0048], [0049]. In contrast, Ross is generally concerned with observing the behavior of road users over time (including past the relevant point in time to make a decision about what the road user will actually do) and using that information to train a machine learning model to predict what such road users will do. *See, e.g.*, Ex. 1004, 4:25-32, 4:51-65. Ross’s system is not concerned or designed for classification purposes as described in Cox but rather for predicting intents and behaviors of road users – a different application altogether that requires a different design of a machine learning system with different training/validation/test datasets. These two technological issues are fundamentally different, and in my opinion, a POSITA looking at Ross’s disclosure would not look to Cox’s disclosure to address either issue Tesla and Dr. Janét identify.

101. Research in the field of intelligent and autonomous vehicle systems also confirms the unique nature of the problems faced by attempting to analyze image, video, or other sensor data in real time driving situations. For example, in 2014 Kooij

et al. noted that “pedestrian path prediction is a challenging problem, due to the highly dynamic nature of pedestrian motion, and systems need to react with limited computation time.” Ex. 2041 at 1. And, in 2017 Zyner et al. noted that “the ability to infer the future intentions of drivers to predict the likelihood of potential collisions” “is a challenging task, particularly in intersections where complex traffic scenarios result in a proportionally high number of accidents.” Ex. 2042 at 1. These examples of other in this field noting the unique challenges of these environments only further reflects that a POSITA would not have been motivated to look to such disclosures to combine them with Cox’s crowdsourced annotation system for improving face detection, which does not present the same unique challenges.

102. Therefore, in my opinion, a POSITA would not have been motivated to combine Ross and Cox in the matter Tesla and Dr. Janét suggest.

IX. GROUND 2

103. I understand that Tesla alleges (and Dr. Janét opines) that claims 1-22 of the ’344 patent are obvious in view of Ross, Cox, and Ogale. *See* Pet. at 47-56; Ex. 1003, ¶¶ 141-158. For the reasons I discuss below, I disagree. In my opinion, none of claims 1-22 are obvious in view of Ross, Cox, and Ogale.

A. Ross, Cox, and Ogale

104. In arriving at my opinions concerning Tesla and Dr. Janét’s assertions regarding claims 1-22 of the ’344 patent and Ross, Cox, and Ogale, I have considered

all of Ross, Cox, and Ogale’s disclosures from the perspective of a POSITA. In my discussion of Ground 1 above, I discussed aspects of Ross and Cox that are relevant to my opinions. Below I discuss some aspects of Ogale that are relevant to my opinions.

105. Ogale indicates that its invention relates to an “[o]bject detection neural network.” Ex. 1010, Title. Ogale’s “specification describes how a computer system can implement a fully-learned neural network to make object predictions from projections of input sensor data. In other words, the system can make accurate object predictions from projections of raw sensor data without using human-programmed logic about how the sensor data should be combined and analyzed.” Ex. 1010, 2:24-30. To this effect, its “specification describes how a vehicle, e.g., an autonomous or semi-autonomous vehicle, can use a fully learned neural network to make object predictions from projections of multiple channels of input sensor data. Each object prediction specifies a region of space that is likely to be occupied by an object. The neural network can be trained to detect any appropriate object class, including vehicles, pedestrians, bicyclists, or lane markers, to name just a few examples.” Ex. 1010, 4:34-42. Its “training neural network subsystem 114 can receive training examples 123 as input. The training examples 123 can include auto-labeled training data 125, human-labeled training data 127, or some combination of the two.” Ex. 1010, 6:51-54.

106. Figure 5 of Ogale “illustrates an example architecture of a neural network 500 for making object predictions from multiple channels of sensor input data.” Ex. 1010, 10:55-57. It depicts a pipeline that receives sensor-derived projection data and progressively transforms it into object detections. The figure shows sensor data entering the network, passing through a space-to-depth transformation that reorganizes spatial information into multiple depth channels. After this preprocessing stage, the data flows through a series of convolutional layers. These layers are intended to extract increasingly complex features from the input, learning patterns that correspond to physical objects in the environment. Each convolutional layer applies learned filters to the data, enabling the network to recognize structures associated with vehicles, pedestrians, and other detectable objects. The processed feature maps are provided as input to a prediction layer. This layer generates an object prediction, identifying regions of space that are likely occupied by objects of interest. In summary, Ogale proposes the use of a convolutional neural network that converts multi-channel sensor projections into spatial object detections by compressing information, extracting features through successive convolutional layers, and producing predictions that identify occupied regions in the environment. *See* Ex. 1010, 10:55-12:2 (discussing Figure 5).

107. In contrast to Ogale, Ross does not propose or describe the use of any type of convolutional neural network (CNN) in its intent prediction methodology.

Cox, on the other hand, discloses the use of support vector machines (SVM) for its binary classification purposes. Ex. 1005, [0038], [0039], [0052]-[0054]. Cox thus describes a completely different architecture than the CNN based approach proposed in Ogale. Generally speaking, support vector machines is a supervised machine-learning algorithm that classifies data by finding the decision boundary (hyperplane) that maximizes the margin between different classes for classification purposes. A convolutional neural network, on the other hand, is a deep-learning model that is generally designed to learn hierarchical features from images by applying convolutional style filters.

B. Ross, Cox, and Ogale Do Not Teach Elements 1[b], 10[b], or 19[b] (“providing the sensor data as input to a supervised learning based model, 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”)

108. In my opinion, the combination of Ross, Cox, and Ogale (as Tesla and Dr. Janét propose) does not teach elements 1[b], 10[b], or 19[b], which each require “providing the sensor data as input to a supervised learning based model, 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.”

109. Tesla and Dr. Janét allege that these elements are met by Ross and Cox for the same reasons as discussed in Ground 1. *See* Pet. at 51-52; Ex. 1003, ¶¶ 149-150. However, Tesla and Dr. Janét then allege that “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.” Pet. at 52; Ex. 1003, ¶ 150. Tesla and Dr. Janét then allege that:

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; EX1003, ¶151. 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.”

Pet. at 52; Ex. 1003, ¶ 150.

110. As I discussed above with respect to Ground 1, in my opinion, Ross and Cox does not teach elements 1[b], 10[b], or 19[b] because, among other reasons, Ross does not teach or suggest how the “set of possible intents or hypotheses” or “associated likelihood values” are generated or what form they take, and neither of these items constitutes a “distribution of user responses” as required by the challenged claims because neither the set of intents/hypotheses or the likelihood

values is a distribution. Tesla and Dr. Janét’s proposed combination of Ross and Cox with Ogale does not alter my opinions in that regard.

111. Tesla and Dr. Janét merely rely on Ogale for its teaching of training its model with “human labeled data” and its disclosure that its trained model “predicts how a human would have labeled the input sensor data.” Pet. at 52; Ex. 1003, ¶ 150. But neither of these statements address the shortcomings of Ross and Cox because neither the fact that Ogale’s training data is labeled by humans or its teaching that its trained model “predicts how a human would have labeled the input sensor data” (*id.*) addresses either the form of Ross’s set of possible intents or hypotheses or associated likelihood values or describe any kind of distribution. Thus, the addition of Ogale to the combination of Ross and Cox still fails to teach these elements of the challenged claims.

112. The portions of the Petition and Dr. Janét’s declaration discussed above relate to claim 1 of the ’344 patent. Both the Petition and Dr. Janét’s declaration merely refer back to these discussions when discussing element 10[b] and 19[b]. *See* Pet. at 56; Ex. 1003, ¶¶ 156-157. In my opinion, Cox fails to teach or suggest elements 10[b] and 19[b] for the same reasons as discussed above.

113. Therefore, in my opinion, Cox and Ross do not teach or suggest elements 1[b], 10[b], or 19[b] (“providing the sensor data as input to a supervised learning based model, 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”).

C. Ross, Cox, and Ogale Do Not Teach Elements 1[c], 10[c], or 19[c] (“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”)

114. In my opinion, the combination of Ross, Cox, and Ogale (as Tesla and Dr. Janét propose) does not teach elements 1[c], 10[c], or 19[c], which each require “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.”

115. Tesla and Dr. Janét merely refer back to their discussion of Cox in Ground 1 and to element 1[b] in Ground 2 for these elements. *See* Pet. at 53-56; Ex. 1003, ¶¶ 153-154. As I discussed above with respect to Ground 1, in my opinion, Ross and Cox does not teach elements 1[c], 10[c], or 19[c] because, among other reasons, Ross does not teach or suggest how the “set of possible intents or hypotheses” or “associated likelihood values” are generated or what form they take, and neither of these items constitutes a “distribution of user responses” as required

by the challenged claims because neither the set of intents/hypotheses or the likelihood values is a distribution. As I discussed above with respect to element 1[b] in Ground 2, Tesla and Dr. Janét’s proposed combination of Ross and Cox with Ogale does not alter my opinions in that regard, as nothing in the portions of Ogale that Tesla and Dr. Janét identify address the shortcomings of Ross and Cox.

116. The portions of the Petition and Dr. Janét’s declaration discussed above relate to claim 1 of the ’344 patent. Both the Petition and Dr. Janét’s declaration merely refer back to these discussions when discussing element 10[c] and 19[c]. *See* Pet. at 56; Ex. 1003, ¶¶ 156-157. In my opinion, Cox fails to teach or suggest elements 10[c] and 19[c] for the same reasons as discussed above.

117. Therefore, in my opinion, Cox and Ross do not teach or suggest elements 1[c], 10[c], or 19[c] (“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”).

D. A POSITA Would Not Have Been Motivated to Combine Ross, Cox, and Ogale in the Manner Tesla and Dr. Janét Allege

118. In my opinion, a POSITA would not have been motivated to combine Ross, Cox, and Ogale in the manner Tesla and Dr. Janét suggest.

119. As I discuss above with respect to Ground 1, a POSITA would not have been motivated to combine Ross and Cox for a number of reasons. Each of those

reasons is equally applicable to the lack of any motivation to combine Ross, Cox, and Ogale as alleged by Tesla and Dr. Janét for Ground 2.

120. In addition, a POSITA would not have been motivated to combine Ogale with Ross and/or Cox. Tesla and Dr. Janét merely allege that “[a] POSITA would have been further motivated to combined Ross and Cox with Ogale” because “[t]he combination would have provided that the Ross-Cox model, trained using human annotated training data, would have provided predictions that ‘predict[] how a human would have labeled the input sensor data.’” Pet. at 48; Ex. 1003, ¶ 143.

121. As I mention above with respect to Ground 1, while Ross does not expressly state the purpose of training its model with annotated training data, a POSITA would have known that the purpose of using annotated training data to train Ross’s model (or any machine learning model) is to make the model’s predictions consistent with the annotations of the training data. Thus, a POSITA would not have been motivated to combine Ross and Cox with Ogale for this reason.

122. Also, as I describe above, Ross does not propose or describe the use of any type of convolutional neural network (CNN) in its intent prediction methodology. Cox, on the other hand, discloses the use of support vector machines (SVM) for its binary classification purposes. Ex. 1005, [0038], [0039], [0052]-[0054]. This represents a completely different architecture than the CNN based approach proposed in Ogale. Generally speaking, support vector machines classify

data by finding the decision boundary (hyperplane) that maximizes the margin between different classes for classification purposes. A convolutional neural network, on the other hand, is a deep-learning model that is generally designed to learn hierarchical features from images by applying convolutional style filters. Tesla and Dr. Janét do not provide any technical descriptions with rational underpinnings as to how the architectures of Ross which is distinctively different from Cox as described earlier would be integrated with the CNN based methodology of Ogale. As an example, I do not understand how the SVM architecture for binary classification of Cox (face/no face) would be integrated with the CNN architecture of Ogale, nor do Tesla or Dr. Janét explain as much.

123. For these reasons (and those I discuss above with respect to Ground 1), in my opinion, a POSITA would not have been motivated to combine Ross, Cox, and Ogale in the matter Tesla and Dr. Janét suggest.

X. GROUND 3

124. I understand that Tesla alleges (and Dr. Janét opines) that claims 1-22 of the '344 patent are obvious in view of Zhang. *See* Pet. 57-76; Ex. 1003, ¶¶ 159-208. For the reasons I discuss below, I disagree. In my opinion, none of claims 1-22 are obvious in view of Zhang.

A. Zhang

125. In arriving at my opinions concerning Tesla and Dr. Janét's assertions regarding claims 1-22 of the '344 patent and Zhang, I have considered Zhang's disclosure from the perspective of a POSITA. Below I discuss some aspects of Zhang that are relevant to my opinions.

126. Zhang indicates that its invention relates to a “[p]rediction-based system and method for trajectory planning of autonomous vehicles”, Ex. 1006, Title. Zhang explains that its system “may include various sensors configured to collect perception data, a computing device, and a trajectory prediction module for predicting a trajectory of other vehicles and/or dynamic objects in the vicinity of (proximate to) a host autonomous vehicle. Initially, the computing device generates a trajectory option, while the trajectory prediction module predicts reactions of vehicles and/or dynamic objects examined using data concerning likely trajectories of each vehicle and/or dynamic objects respectively. Data corresponding to predicted reactions can be sent to the computing device to perfect the trajectory option suggested initially. The computing device may subsequently instruct the trajectory prediction module to further collect data and conduct predictions to complete the trajectory planning process.” Ex. 1006, [0006].

127. Zhang goes on to explain that its system “can be implemented in two phases: an offline training phase and an operational phase. The training phase is used

to train and configure the parameters of the machine learning system or neural network of the trajectory prediction module 175 or any other component of the prediction - based trajectory planning system 202 implemented with a machine learning system or neural network. The operational phase is used after the machine learning components have been trained and are ready to support the generation of predicted vehicle or object trajectories as described in more detail below” in the specification. Ex. 1006, [0045].

128. Zhang explains that its “training data collection system **201** can include a plurality of training data gathering mechanisms ...” and that “[c]orresponding ground truth data can also be gathered by the training data collection system **201**. As a result, the perception information, ground truth data, and other available information can be obtained, processed, and used to build a training dataset for training and configuring the machine learning components of the prediction based trajectory planning system **202**.”, Ex. 1006, [0047]. Zhang goes on to explain that its “training data collection system **201** can thereby collect actual trajectories of vehicles and corresponding ground truth data under different scenarios and different driver actions and intentions in a context”, Ex. 1006, [0048] and that during the training phase, a loss function can be utilized to “examine and correct the results of the training provided to the trajectory prediction module 175 by the training data collection system 201.”, Ex. 1006, [0049]. Zhang states that “images included in the

training data can be labeled using human labelers ...” and that the “labels or labeling data can include context information that defines the directionality and rate behaviors of the vehicles represented in the training data.”, Ex. 1006, [0051].

B. Zhang Does Not Teach Elements 1[b], 10[b], or 19[b] (“providing the sensor data as input to a supervised learning based model, 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”)

129. In my opinion, Zhang does not teach elements 1[b], 10[b], or 19[b], which each require “providing the sensor data as input to a supervised learning based model, 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.”

130. Tesla and Dr. Janét allege that Zhang’s “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.’” Pet. at 63; Ex. 1003, ¶ 173. Nothing about these features teaches that Zhang’s model “predict[s] 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.”

131. Zhang does not teach that its model predicts anything about what users’ responses would be if presented with the sensor data. Instead, Zhang teaches that its training process can use either “real world, human behavior data” (Ex. 1006, [0049]) or training data that is “labeled using human labelers” (*id.*, [0051]). Regardless of which data is used, Zhang does not teach or suggest that its model predicts what users’ responses or labels to the image data would be. Instead, Zhang makes clear that what its model predicts are “vehicle or object trajectories.” *Id.*, [0053]. And, nothing in Zhang suggests that these “vehicle or object trajectories” are “an output statistical summary characterizing a distribution of user responses.”

132. The two paragraphs of Zhang that Tesla and Dr. Janét cite do not support its argument. *See* Pet. at 63; Ex. 1003, ¶ 173. Paragraph [0025] of Zhang merely indicates that “the vehicle may . . . determine a predicted behavior of at least one other vehicle in the context of the environment[] [and] determine a confidence level that may correspond to a likelihood of the at least one other vehicle to perform the predicted behavior.” Ex. 1006, [0025]. And, paragraph [0058] repeatedly discusses predicting trajectories using “trajectory prediction module 175.” *Id.*, [0058]. Nothing in Zhang teaches or suggests that these predicted trajectories (or confidence levels) amount to “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,” as required by the challenged claims.

133. I have considered the portions of Zhang identified by Tesla and Dr. Janét as part of their allegations with respect to this claim element. *See* Pet. at 60-67; Ex. 1003, ¶¶ 165-177. None of those portions of Zhang alter my opinions with respect to these claim elements.

134. The portions of the Petition and Dr. Janét’s declaration discussed above relate to claim 1 of the ’344 patent. Both the Petition and Dr. Janét’s declaration merely refer back to these discussions when discussing element 10[b] and 19[b]. *See* Pet. at 74-75; Ex. 1003, ¶¶ 191, 203. In my opinion, Zhang fails to teach or suggest elements 10[b] and 19[b] for the same reasons as discussed above.

135. Therefore, in my opinion, Zhang does not teach or suggest elements 1[b], 10[b], or 19[b] (“providing the sensor data as input to a supervised learning based model, 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”).

C. Zhang Does Not Teach Elements 1[c], 10[c], or 19[c] (“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”)

136. In my opinion, Zhang does not teach elements 1[c], 10[c], or 19[c], which each require “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.”

137. Tesla and Dr. Janét allege that Zhang teaches this element for the same reasons they allege with respect to element 1[b]. *See* Pet. at 67-69; Ex. 1003, ¶¶ 178-179. However, as I discuss above with respect to that element, Zhang does not teach or suggest “generat[ing] 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,” as required by both elements 1[b] and 1[c].

138. I have considered the portions of Zhang that Tesla and Dr. Janét identify with respect to this claim element, as well as the portions of the Petition and Dr. Janét’s declaration cross-reference. *See* Pet. at 67-69; Ex. 1003, ¶¶ 178-179. None of them alter my opinions expressed above that Zhang does not teach or suggest “generat[ing] 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.”

139. The portions of the Petition and Dr. Janét’s declaration discussed above relate to claim 1 of the ’344 patent. Both the Petition and Dr. Janét’s declaration merely refer back to these discussions when discussing element 10[c] and 19[c]. *See* Pet. at 74-75; Ex. 1003, ¶¶ 192, 204. In my opinion, Zhang fails to teach or suggest elements 10[c] and 19[c] for the same reasons as discussed above.

140. Therefore, in my opinion, Zhang does not teach or suggest elements 1[c], 10[c], or 19[c] (“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”).

XI. CONCLUSION

141. This declaration represents the opinions that I have formed up to date based on the material currently available to me. I reserve the right to revise, supplement, and/or amend my opinions stated in this declaration based on new information that becomes available to me and on my continuing analysis of the materials available to me.

142. I declare that all statements made herein of my knowledge are true, and that all statements made on information and belief are believed to be true, and that

these statements were made with the knowledge that willful false statements and the like so made are punishable by fine, imprisonment, or both, under Section 1001 of Title 18 of the United States Code.

Dated: June 15th, 2026

A handwritten signature in blue ink, appearing to read 'Eli Saber', written over a horizontal line.

Dr. Eli Saber, Ph.D.