

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

TESLA, INC.,

Petitioner,

v.

PERCEPTIVE AUTOMATA LLC,

Patent Owner.

Case No. IPR2025-01577

Patent No. 11,520,346 B2

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	11
A.	Overview	11
B.	Claims.....	12
C.	Burden of Proof for Invalidity	13
D.	Claim Construction.....	13
E.	Person of Ordinary Skill in the Art	14
F.	Prior Art.....	14
G.	Anticipation.....	15
H.	Obviousness	16
V.	PERSON OF ORDINARY SKILL IN THE ART	19
VI.	THE '346 PATENT AND THE CHALLENGED CLAIMS	20
A.	Overview of the '346 Patent.....	20
B.	Claim Overview	22
VII.	CLAIM CONSTRUCTION.....	25
VIII.	GROUND 1, 3, 5, AND 7.....	26
A.	Djuric.....	26
B.	Djuric Does Not Teach Elements 1[d-ii], 10[d-ii], and 19[d-ii] ("output summary statistics of expected human responses describing a hidden context"); Elements 1[d-iii], 10[d-iii], and 19[d-iii] ("determining the hidden context of the traffic entity"); and Elements	

1[d-v], 10[d-v], and 19[d-v] (“modifying the region based on the hidden context”).....	28
1. Djuric Does Not Teach “hidden context”.....	29
2. Djuric Does Not Teach “output summary statistics of expected human responses describing a hidden context”.....	33
IX. GROUNDS 2, 4, 6, AND 8.....	35
A. Cox.....	35
B. Djuric and Cox Do Not Teach Elements 1[d-ii], 10[d-ii], and 19[d-ii] (“output summary statistics of expected human responses describing a hidden context”); Elements 1[d-iii], 10[d-iii], and 19[d-iii] (“determining the hidden context of the traffic entity”); and Elements 1[d-v], 10[d-v], and 19[d-v] (“modifying the region based on the hidden context”).....	41
1. Djuric and Cox Do Not Teach “hidden context”.....	41
2. Djuric and Cox Do Not Teach “output summary statistics of expected human responses describing a hidden context of the input traffic entity”.....	44
C. A POSITA Would Not Have Been Motivated to Combine Djuric and Cox in the Manner Tesla and Dr. Janét Allege	47
X. BOTH THE PETITION AND DR. JANÉT’S DECLARATION FAIL TO ANALYZE THE FACTORS REQUIRED FOR A PROPER OBVIOUSNESS ANALYSIS.....	56
XI. CONCLUSION.....	57

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. 11,520,346 (the “’346 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 1-23 of the ’346 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 1-8 of the Petition, render obvious claims 1-23 of the ’346 patent. In my opinion, the alleged prior art references, as presented in Grounds 1-8 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 '346 Patent, I have extensive experience in teaching and developing neural network, machine learning and deep learning algorithms for computer vision, digital image/video processing and remote sensing applications. I also have extensive experience in teaching and developing 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. I have also served as an expert on multiple cases involving machine learning algorithms with applications to autonomous devices as well as other entities.

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 '346 patent.

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

<u>Document</u>	<u>Exhibit/Paper #</u>
'346 patent	1001
File history for the '346 patent	1002
Declaration of Jason Janét	1003

U.S. Patent Application Publication No. 2019/0049970 (“Djuric”)	1004
U.S. Patent Application Publication No. 2019/0164018 (“Zhu”)	1005
International Publication WO 2014/210334 (“Cox”)	1006
U.S. Patent No. 10,496,091 (“Ross”)	1015
Petition	1
Patent Owner’s Preliminary Response	9
Guo, Z.; Huang, Y.; Hu, X.; Wei, H.; Zhao, B. A Survey on Deep Learning Based Approaches for Scene Understanding in Autonomous Driving. Electronics 2021, 10, 471.	2041
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	2042

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	2043
C. Aggarwal, “Neural Networks and Deep Learning”, Springer ISBN 978-3-319-94462-3	2044

27. I have been asked to consider the time of the inventions of the challenged claims to be around January 30, 2019, which I understand is the filing date of the earliest provisional patent application to which the '346 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 '346 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 '346 patent claims priority to U.S. Provisional Patent Application No. 62/798,978, which was filed on January 30, 2019. I understand that the '346 patent also claims priority to U.S. Provisional Patent Application No. 62/800,416, which was filed on February 1, 2019. I understand that the '346 patent also claims priority to U.S. Provisional Patent Application No.

62/822,269, which was filed on March 22, 2019. My opinions herein do not differ whether the time of the inventions of the challenged claims is around January 30, 2019, February 1, 2019, March 22, 2019, or January 30, 2020. 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 '346 patent.

29. As I discuss in detail below, it is my opinion that claims 1-23 of the '346 patent are not rendered obvious by the alleged prior art references, as presented in Grounds 1-8 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 '346 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 '346 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 '346 Patent as of January 30, 2019 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, ¶ 68.

53. For purposes of providing my opinions in connection with this declaration, I do not dispute Dr. Janét's declaration, and I have applied his 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 '346 Patent as of January 30, 2019 under either prong of Dr. Janét's definition. *See* Section II above and my CV (Ex. 2049).

VI. THE '346 PATENT AND THE CHALLENGED CLAIMS

A. Overview of the '346 Patent

54. The inventions disclosed and claimed in the '346 patent generally relate to systems, computer-implemented methods, and computer readable media associated with “navigating an autonomous vehicle through traffic” by “predict[ing] hidden context attributes associated with traffic entities that determine behavior of these traffic entities in the traffic.” *See, e.g.*, Ex. 1001 at 1:27-28, 2:1-3. The invention claimed in the '346 patent is directed to navigating an autonomous vehicle so that the autonomous vehicle stays at least a threshold distance away from a specified region where a traffic entity is expected to reach within a threshold time interval. *Id.* at 2:55-3:6, claim 1. The '346 patent discloses an implementation for achieving this result by utilizing a machine learning based model. *Id.* The '346 patent can generate a point cloud representation of the surroundings of the autonomous vehicle and identify traffic entities located in the surroundings of the autonomous vehicle based on sensor data received from sensors mounted on the autonomous vehicle. *Id.* The machine learning model can be trained in part based on a group of users (or human observers) viewing images of traffic entities and indicating their understanding of how they believe the traffic entity will behave. *See id.* at 2:30-34. The machine learning model is configured to receive an input image of a traffic entity, output summary statistics of expected human responses describing a hidden

context of the traffic entity shown in the input image, and determine the hidden context of the traffic entity based on the output. *Id.* at 3:19-35, claim 1. The system of the '346 patent can then determine a region of the point cloud where the traffic entity is expected to reach within a threshold time interval and use the determined hidden context to modify the region. *Id.* For example, the region can be adjusted based on the hidden context such that the size and direction of the region is adjusted along a predicted motion vector of the traffic entity. *See id.* at 3:36-47. The autonomous vehicle is controlled based on the determinations made based on the hidden context. *See id.* at 3:48-52.

55. The '346 patent describes that hidden context can be, for example, “a state of mind of a user,” “a task that a user . . . is planning on accomplishing,” “a goal of a user . . . wherein the user expects to achieve the goal within a threshold time interval,” (*id.* at 2:9-30) “a near term goal . . . for example . . . indicating that the person is likely to pick up an object (e.g., a wallet) dropped on the street but stay on that side of the street,” “a degree of awareness of the person about the autonomous vehicle,” or “whether a bicyclist driving in front of the autonomous vehicle is likely to be aware that the autonomous vehicle is behind the bicycle” (*id.* at 14:67-15:3).

B. Claim Overview

56. 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 method comprising:
1[a]	receiving, by an autonomous vehicle, sensor data from sensors mounted on the autonomous vehicle, the sensor data comprising one or more images;
1[b]	generating a point cloud representation of the surroundings of the autonomous vehicle based on the sensor data;
1[c]	identifying, one or more traffic entities based on the sensor data, the traffic entities representing non-stationary objects in traffic in which the autonomous vehicle is driving;
1[d-i]	for each of the one or more traffic entities: determining one or more motion parameters describing movement of the traffic entity;
1[d-ii]	[for each of the one or more traffic entities:] providing an image of the one or more images, the image showing the traffic entity as input to a machine learning model configured to receive an input image showing an input traffic entity and output summary statistics of expected human responses describing a hidden context of the input traffic entity shown in the input image;
1[d-iii]	[for each of the one or more traffic entities:] determining the hidden context of the traffic entity based on the output of the machine learning based model;
1[d-iv]	[for each of the one or more traffic entities:] determining a region of the point cloud where the traffic entity is expected to reach within a threshold time interval; and

1[d-v]	[for each of the one or more traffic entities:] modifying the region based on the hidden context of the traffic entity; and
1[e]	navigating the autonomous vehicle so that the autonomous vehicle stays at least a threshold distance away from the modified region of each of the one or more traffic entities.
Claim 10	Limitations
10[pre]	A non-transitory computer readable storage medium storing instructions, that when executed by a processor, cause the processor to perform steps comprising:
10[a]	receiving, by an autonomous vehicle, sensor data from sensors mounted on the autonomous vehicle, the sensor data comprising one or more images;
10[b]	generating a point cloud representation of the surroundings of the autonomous vehicle based on the sensor data;
10[c]	identifying, one or more traffic entities based on the sensor data, the traffic entities representing non-stationary objects in traffic in which the autonomous vehicle is driving;
10[d-i]	for each of the one or more traffic entities: determining one or more motion parameters describing movement of the traffic entity;
10[d-ii]	[for each of the one or more traffic entities:] providing an image of the one or more images, the image showing the traffic entity as input to a machine learning model configured to receive an input image showing an input traffic entity and output summary statistics of expected human responses describing a hidden context of the input traffic entity shown in the input image;
10[d-iii]	[for each of the one or more traffic entities:] determining the hidden context of the traffic entity based on the output of the machine learning based model;

10[d-iv]	[for each of the one or more traffic entities:] determining a region of the point cloud where the traffic entity is expected to reach within a threshold time interval; and
10[d-v]	[for each of the one or more traffic entities:] modifying the region based on the hidden context of the traffic entity; and
10[e]	navigating the autonomous vehicle so that the autonomous vehicle stays at least a threshold distance away from the modified region of each of the one or more traffic entities.
Claim 19	Limitations
19[pre]	A computer system comprising: a processor; and a non-transitory computer readable storage medium storing instructions that when executed by the processor, cause the processor to perform steps comprising:
19[a]	receiving, by an autonomous vehicle, sensor data from sensors mounted on the autonomous vehicle, the sensor data comprising one or more images;
19[b]	generating a point cloud representation of the surroundings of the autonomous vehicle based on the sensor data;
19[c]	identifying, one or more traffic entities based on the sensor data, the traffic entities representing non-stationary objects in traffic in which the autonomous vehicle is driving;
19[d-i]	for each of the one or more traffic entities: determining one or more motion parameters describing movement of the traffic entity;
19[d-ii]	[for each of the one or more traffic entities:] providing an image of the one or more images, the image showing the traffic entity as input to a machine learning model configured to receive an input image showing an input traffic entity and output summary statistics of expected human responses describing a hidden context of the input traffic entity shown in the input image;

19[d-iii]	[for each of the one or more traffic entities:] determining the hidden context of the traffic entity based on the output of the machine learning based model;
19[d-iv]	[for each of the one or more traffic entities:] determining a region of the point cloud where the traffic entity is expected to reach within a threshold time interval; and
19[d-v]	[for each of the one or more traffic entities:] modifying the region based on the hidden context of the traffic entity; and
19[e]	navigating the autonomous vehicle so that the autonomous vehicle stays at least a threshold distance away from the modified region of each of the one or more traffic entities.

VII. CLAIM CONSTRUCTION

57. I understand that, to date, neither Perceptive nor Tesla has proposed any constructions for any terms in the challenged claims of the '346 patent. Based on my review of the '346 patent claims, specification, and file history, I agree that, to resolve the issues in this IPR proceeding, no terms in the challenged claims require construction.

58. Given that conclusion, I have applied the plain and ordinary meaning to a POSITA of all terms in the challenged claims, including the term “hidden context.” Based on my review of the '346 patent specification, in my opinion, the plain and ordinary meaning of the term “hidden context,” as used in the '346 patent, “represent[s] intentions and future plans of a traffic entity” (Ex. 1001 at 14:55-57) and “includes factors that affect the behavior of[] traffic entities, for example, a state

of mind of a user represented by a traffic entity such as a pedestrian[,] a degree of awareness of the autonomous vehicle by a user represented by the traffic entity,” and more (*id.* at 2:9-30).

VIII. GROUNDS 1, 3, 5, AND 7

59. I understand that Tesla alleges (and Dr. Janét opines) that certain limitations of claims 1-23 of the ’346 patent are obvious in view of Djuric alone. *See* Pet. at 21-50, 63-67, 70-74; Ex. 1003, ¶¶ 76-153, 179-185, 193-203. Grounds 1, 3, 5, and 7 all rely exclusively on Djuric as allegedly teaching Limitations 1[d-ii], 10[d-ii], 19[d-ii], 1[d-iii], 10[d-iii], 19[d-iii], 1[d-v], 10[d-v], and 19[d-v] of independent claims 1, 10, and 19. *See* Pet. at 21-36, 47-50, 63-67, 70-74. For the reasons I discuss below, I disagree. In my opinion, none of claims 1-23 are obvious in view of Djuric, either alone or in combination with any other reference discussed in Grounds 3, 5, or 7.

A. Djuric

60. In arriving at my opinions concerning Tesla and Dr. Janét’s assertions regarding claims 1-23 of the ’346 patent and Djuric, I have considered Djuric’s disclosure from the perspective of a POSITA. Below I discuss some aspects of Djuric that are relevant to my opinions.

61. As Djuric indicates, it is “directed to predicting the moving trajectory of objects that are within the surrounding environment of an autonomous vehicle.” Ex. 1004 at [0021].

62. Djuric explains that it uses “state data [fused] with data associated with a geographic area to create a combined data set” that is input “into a machine-learned model [that] can output one or more predicted future locations of [an] object (e.g., a predicted object trajectory).” *Id.* The “state data” is “indicative of at least a current or a past state of an object that is within a surrounding environment of an autonomous vehicle.” *Id.* at [0005]. The purpose of fusing the state data with data associated with a geographic area is to “project the data associated with the geographic area onto the state data such that the states of the object can be better evaluated within the context of the object’s surroundings.” *Id.* at [0021]. I note that the geographic data is only used to provide a framework for evaluating the state data: “The combined data set can be specific to a particular object, rather than the entire surrounding environment of the autonomous vehicle This enables the autonomy computing system to focus its processing resources . . . on predicting the future location(s) of an individual object, rather than the future states of the entire surrounding environment.” *Id.* at [0031]. Thus, Djuric only uses its geographic data to improve the prediction of an observed object’s trajectory.

63. I note that Djuric states it predicts a future location of an object based on the object's velocity and acceleration, known current and/or past locations of the object, and elements in the object's environment such as obstructions in the path of the object. *See id.* at [0023], [0025]. Djuric does not disclose analyzing image content to infer hidden context, identifying latent information within an image, or predicting information unrelated to the observed state and motion of a particular object. A POSITA would understand that there is a difference between the "high-level path planning" techniques of Djuric, and the "complex task of outdoor scene understanding" that is accomplished by the '346 patent. Ex. 2050 at pp. 1-2.

B. Djuric Does Not Teach Elements 1[d-ii], 10[d-ii], and 19[d-ii] ("output summary statistics of expected human responses describing a hidden context"); Elements 1[d-iii], 10[d-iii], and 19[d-iii] ("determining the hidden context of the traffic entity"); and Elements 1[d-v], 10[d-v], and 19[d-v] ("modifying the region based on the hidden context")

64. Djuric does not teach "output summary statistics of expected human responses describing a hidden context" (as required by elements 1[d-ii], 10[d-ii], and 19[d-ii]); "determining the hidden context of the traffic entity" (as required by elements 1[d-iii], 10[d-iii], and 19[d-iii]); or "modifying the region based on the hidden context" (as required by elements 1[d-v], 10[d-v], and 19[d-v] of each of claims 1, 10, and 19. Below I discuss why these claim elements are not met by Djuric.

1. Djuric Does Not Teach “hidden context”

65. Based on my review of the Petition and Dr. Janét’s declaration, I understand that Tesla and Dr. Janét do not contend that any combination of references discloses the claimed “hidden context,” and they instead rely exclusively on Djuric for this limitation. *See* Pet. at 25-30, 33-34, 63-64, 70-71; Ex. 1003 at ¶¶ 84-94, 99, 179, 193. Tesla and Dr. Janet allege that Djuric’s teaching that “the model generates predicted trajectories in consideration of ‘the nominal pathway(s) of an object’” meets the claimed “generating output representing hidden context” and that “Djuric’s disclosure of ‘goal-based prediction’ aligns with the ’579 patent’s description of ‘hidden context.’” Pet. at 29-30; Ex. 1003 at ¶ 93. I disagree with Tesla and Dr. Janét.

66. As discussed above, in my opinion, the term “hidden context” does not require any construction. Based on my review of the ’346 patent specification, in my opinion, the plain and ordinary meaning of the term “hidden context,” as used in the ’346 patent, “represent[s] intentions and future plans of a traffic entity” (Ex. 1001 at 14:55-57) and “includes factors that affect the behavior of [] traffic entities, for example, a state of mind of a user represented by a traffic entity such as a pedestrian[,] a degree of awareness of the autonomous vehicle by a user represented by the traffic entity,” and more (*id.* at 2:9-30). As such, in my opinion, “hidden context” encompasses information that is not directly observable from an image.

67. The '346 patent provides examples of hidden context, including: “a state of mind of a user,” “a task that a user . . . is planning on accomplishing,” “a goal of a user . . . wherein the user expects to achieve the goal within a threshold time interval” (*id.* at 2:9-30), “a near term goal . . . for example . . . indicating that the person is likely to pick up an object (e.g., a wallet) dropped on the street but stay on that side of the street,” “a degree of awareness of the person about the autonomous vehicle,” “whether a bicyclist driving in front of the autonomous vehicle is likely to be aware that the autonomous vehicle is behind the bicycle” (*id.* at 14:55-15:3), or what a person “intends” to do (*id.* at 16:46-51, 19:25-31). Djuric does not disclose these or similar concepts.

68. Djuric’s “nominal pathway(s) . . . correspond to common patterns of object travel.” Ex. 1004 at ¶ 81. Djuric explains that the “‘goal’ that an object may aim to accomplish . . . [is] associated with one or more nominal pathways . . . correspond[ing] to common patterns of object travel [such as] vehicles travelling within lanes on a roadway or other travel way, bicycles traveling in a bike lane, pedestrians travelling within the boundaries of a walkway, etc.” *Id.* at [0081]. Therefore, the “goal” in Djuric is for the object to travel along a pathway in a commonly recognized pattern. *See id.*

69. The “goal-based prediction” of Djuric is fundamentally different from the “hidden context” disclosed in the '346 patent. As previously discussed, the '346

patent “uses the hidden context to predict behavior of people near a vehicle in a way that more closely resembles how human drivers would judge the behavior.” Ex. 1001 at 2:23-26. The ’346 patent focuses on the nature of people to behave in ways that are typically unpredictable by conventional systems, whereas Djuric focuses on movement in “common patterns” that was recognized by prior art systems. *See* Ex. 1003 at ¶ 93.

70. The input of Djuric is “state data [fused] with data associated with a geographic area to create a combined data set,” which is input “into a machine-learned model [that] can output one or more predicted future locations of [an] object (e.g., a predicted object trajectory).” Ex. 1004 at [0021]. The purpose of fusing the state data with data associated with a geographic area is to “project the data associated with the geographic area onto the state data such that the states of the object can be better evaluated within the context of the object’s surroundings.” *Id.* As discussed above, I note that the geographic data is only used to provide a framework for evaluating the state data, which “enables the autonomy computing system to focus its processing resources ... on predicting the future location(s) of an individual object, rather than the future states of the entire surrounding environment.” *Id.* at [0031]. Thus, Djuric teaches away from analyzing an entire environment surrounding an object.

71. I have considered the portions of Djuric 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 Djuric teaches "hidden context." *See* Pet. at 25-30, 33-34, 63-64, 70-71; Ex. 1003 at ¶¶ 84-94, 99, 179, 193. Instead, those portions, as discussed above, make clear that Djuric's goal-based prediction is not hidden context as recited in the challenged claims.

72. At least certain of the portions of the Petition and Dr. Janét's declaration discussed above relate to claim 1 of the '346 patent and Ground 1. Both the Petition and Dr. Janét's declaration refer back to these discussions when discussing this claim limitation for the other independent claims in Ground 1 as well as in Grounds 3, 5, and 7. *See* Pet. at 48, 50, 63-64, 70-71; Ex. 1003, ¶¶ 126, 145, 179, 193. In my opinion, Djuric fails to teach or suggest this limitation for the same reasons as discussed above.

73. Based on my analysis of the Petition and Dr. Janét's declaration, Tesla and Dr. Janét do not identify where Djuric supposedly discloses hidden context or determining hidden context. As Djuric does not disclose the claimed "hidden context," then Djuric cannot generate output summary statistics of expected human responses describing the hidden context, nor modify a region based on the hidden context. Therefore, Djuric does not disclose the claimed limitations.

2. Djuric Does Not Teach “output summary statistics of expected human responses describing a hidden context”

74. I understand that Tesla’s argument as to this limitation is that “[b]ecause the confidence levels are a collection of numerical measurements summarizing how likely the trajectories are, a POSITA would have recognized that these values form summary statistics,” and that “[i]t would have been obvious for a POSITA to use” human-labeled training data. Pet. at 26-27; Ex. 1003 at ¶¶ 87-88. Based on my analysis of the Petition and Dr. Janét’s Declaration, this appears to be the extent of Tesla’s argument and Dr. Janét does not offer any additional explanation. *See id.*

75. Tesla and Dr. Janét’s arguments fail to consider the proper relationship between the limitations of the claim element. The claims of the ’346 patent require the “output[ting] summary statistics of expected human responses” that describe “a hidden context of the input traffic entity.” Thus, the claimed summary statistics are tied to, and describe, the hidden context of the traffic entity – the claimed summary statistics are not simply statistics of any kind. Therefore, Tesla and Dr. Janét’s argument that Djuric “[b]ecause the confidence levels are a collection of numerical measurements summarizing how likely the trajectories are, a POSITA would have recognized that these values form summary statistics,” (Pet. at 26-27; Ex. 1003 at ¶ 87-88) is technically incorrect because it does not satisfy the claim requirement that

the “summary statistics” be tied to and describe the hidden context of the traffic entity.

76. Tesla and Dr. Janét allege that Djuric teaches using “supervised training techniques,” including using “labeled” data and that “it is well-known to use human labeled training data for model training.” Pet. at 27; Ex. 1003 at [0088]. But they identify no teaching in Djuric stating that the predicted trajectories and/or associated confidence levels are “of expected human responses describing a hidden context of the input traffic entity,” as required by the challenged claims. Tesla only cites to Dr. Janét’s declaration, which repeats what is stated in the Petition: “it was well-known to use human labeled training data for model training.” Ex. 1003 at [0088].

77. I have considered the portions of Djuric 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 Djuric teaches this claim limitation. *See* Pet. at 26-28; Ex. 1003, ¶¶ 87-91. Instead, those portions, as discussed above, make clear that Djuric does not teach “output summary statistics of expected human responses describing a hidden context.”

78. At least certain of the portions of the Petition and Dr. Janét’s declaration discussed above relate to claim 1 of the ’346 patent and Ground 1. Both the Petition and Dr. Janét’s declaration refer back to these discussions when discussing this claim limitation for the other independent claims in Ground 1 as well as in Grounds 3, 5,

and 7. *See* Pet. at 48, 50, 63-64, 70-71; Ex. 1003, ¶¶ 126, 145, 179, 193. In my opinion, Djuric fails to teach or suggest this limitation for the same reasons as discussed above.

79. In my opinion, nothing in Djuric teaches the claimed “summary statistics of expected human responses describing a hidden context of the input traffic entity.”

IX. GROUNDS 2, 4, 6, AND 8

80. Tesla alleges (and Dr. Janét opines) that certain limitations of claims 1-23 of the '346 patent are obvious in view of Djuric and Cox. *See* Pet. at 51, 56-60, 63-64, 67-68, 70-71; Ex. 1003, ¶¶ 154, 163-169, 171, 179, 185, 187, 193. Grounds 2, 4, 6, and 8 all rely on Djuric and Cox as allegedly teaching Limitations 1[d-ii], 10[d-ii], 19[d-ii], 1[d-iii], 10[d-iii], 19[d-iii], 1[d-v], 10[d-v], and 19[d-v] of independent claims 1, 10, and 19. *See id.* For the reasons I discuss below, I disagree. In my opinion, none of claims 1-23 are obvious in view of Djuric and Cox.

A. Cox

81. In arriving at my opinions concerning Tesla and Dr. Janét's assertions regarding claims 1-23 of the '346 patent and Cox, I have considered Cox's disclosure from the perspective of a POSITA. Below I discuss some aspects of Cox that are relevant to my opinions.

82. As Cox indicates, its “invention relates to machine learning, in particular to machine learning augmented by measurements acquired from human populations.” Ex. 1006, [0001].

83. 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 ’346 patent recite “generating output representing hidden context for the traffic entity, the output comprising a plurality of values, each value representing a likelihood of receiving a particular user response from a user presented with the image.” 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. 1006, [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.

84. 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. 1006, [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.*

85. 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. 1006, [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.

1006, [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. 1006, [0008], [0010] where “[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.” Ex. 1006, [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. 1006, [0009].

86. 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. 1006, [0035] and 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: $z = yf(x)$. In accordance with various embodiments of the invention, a loss function that applies a linearly increasing cost for misclassifications (one-sided error) is desirable. This is embodied by the hinge loss function, which is defined as: $\phi(z) = \max(0, 1 - z)$ ” Ex. 1006, [0035].

87. 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. 1006, [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.

88. Finally, Cox explains that “[e]mbodiments of the present invention incorporate a substantial pool of detection data (e.g., more than ten thousand subjects with one hundred trials per subject across all conditions, tested with approximately 7,500 face images) for occluded faces in very large, noisy, perpendicularly obstructed, and Simoncelli textured configurations. These scenarios represent an important set of difficult circumstances for detection, enabling the location of more faces through a better understanding of difficult examples during classifier training.” Ex. 1006, [0047] and provides results on face classification in figures 9, 10 and 11 with comparisons between the use of the hinge loss function and the proposed alternative Human-weighted loss function. Clearly, from the above disclosures, the focus of Cox (training, testing, use of loss functions, etc.) is primarily on face classification.

B. Djuric and Cox Do Not Teach Elements 1[d-ii], 10[d-ii], and 19[d-ii] (“output summary statistics of expected human responses describing a hidden context”); Elements 1[d-iii], 10[d-iii], and 19[d-iii] (“determining the hidden context of the traffic entity”); and Elements 1[d-v], 10[d-v], and 19[d-v] (“modifying the region based on the hidden context”)

89. Djuric and Cox do not disclose “output summary statistics of expected human responses describing a hidden context” (as required by elements 1[d-ii], 10[d-ii], and 19[d-ii]); “determining the hidden context of the traffic entity” (as required by elements 1[d-iii], 10[d-iii], and 19[d-iii]); and “modifying the region based on the hidden context” (as required by elements 1[d-v], 10[d-v], and 19[d-v] of each of claims 1, 10, and 19. Above I have discussed why different limitations of these claim elements are not met by Djuric, and those deficiencies apply equally here. Below I discuss the different limitations of these claim elements that are also not met by Cox, to the extent the Petition and Dr. Janét allege they are.

1. Djuric and Cox Do Not Teach “hidden context”

90. Based on my review of the Petition and Dr. Janét’s declaration, I do not understand Tesla and Dr. Janét to contend that Cox discloses the claimed “hidden context,” and that Tesla and Dr. Janét only contends that Djuric teaches this limitation. *See* Pet. at 56-59; Ex. 1003, ¶¶ 163-168. As discussed above with respect to Grounds 1, 3, 5, and 7, Djuric does not teach hidden context.” *See* Section VIII.B.

91. While Tesla and Dr. Janét do not allege that Cox teaches hidden context, in my opinion, Cox does not teach hidden context.

92. Cox is directed to using “large sets of human-derived data containing different types of information” to provide “ground-truth labels for regularization of machine-learning systems.” Ex. 1006 at [0005]. Cox incorporates human-generated labels into training data to improve machine learning performance to produce more “human-like” results. *See id.*; *see also id.* at [0016].

93. Cox trains its machine learning systems to identify and classify features that are present in an image. Cox describes presenting participants with images and asking them to identify a person’s gender, determine whether a face is present in a partially obscured image, or select an emotion corresponding to a facial expression. *Id.* [0016]. The participants then assign labels to the images, and those resulting human-generated labels are used to train a machine learning system to detect faces despite image noise or occlusion and assign visual attributes to images. *Id.* As examples, Cox provides that the trained system may determine that an image contains a man who is smiling, has a beard, and is wearing a hat, or may identify objects such as a tree, car, bus, building, or street. *Id.* Thus, Cox is directed to recognizing and labeling observable features and objects in an image and assigning corresponding labels.

94. The claimed “hidden context” of the ’346 patent represents things that are not directly visible in an image (for example, “a state of mind of a user,” “a task that a user . . . is planning on accomplishing,” “a goal of a user . . . wherein the user

expects to achieve the goal within a threshold time interval,” (Ex. 1001 at 2:9-30) “a near term goal,” “a degree of awareness of the person about the autonomous vehicle,” or “whether a bicyclist driving in front of the autonomous vehicle is likely to be aware that the autonomous vehicle is behind the bicycle” (*id.* at 14:67-15:3)). The claimed “hidden context” requires inferences regarding mental state, awareness, intentions, or likely behavior. Cox does not make inferences such as inferring latent information that is not directly observable from the image. Cox also does not predict underlying mental states, intentions, awareness, goals, or other hidden contextual information in an image. Thus, Cox does not disclose “hidden context.”

95. Based on my analysis of the Petition and Dr. Janét’s Declaration, Tesla and Dr. Janét do not identify where Cox supposedly disclose “generating output representing” hidden context.

96. I have considered the portions of Djuric 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 Djuric teaches this claim limitation. *See* Pet. at 56-59; Ex. 1003, ¶¶ 163-168. Instead, those portions, as discussed above, make clear that Djuric and Cox do not teach “hidden context.”

97. At least certain of the portions of the Petition and Dr. Janét’s declaration discussed above relate to claim 1 of the ’346 patent and Ground 2. Both the Petition and Dr. Janét’s declaration refer back to these discussions when discussing this claim

limitation for the other independent claims in Ground 2 as well as in Grounds 4, 6, and 8. *See* Pet. at 56-59, 63-67, 70-71; Ex. 1003, ¶¶ 163-168, 179-184, 193. In my opinion, Djuric and Cox fail to teach or suggest this limitation for the same reasons as discussed above.

98. As neither of Cox or Djuric discloses the claimed “hidden context,” then correspondingly, neither can generate an output representing such hidden context. Therefore, neither Cox nor Djuric discloses “generating output representing hidden context.” And, the grafting of Cox onto Djuric does not remedy the deficiencies of either Djuric or Cox with respect to the claimed “hidden context.”

2. Djuric and Cox Do Not Teach “output summary statistics of expected human responses describing a hidden context of the input traffic entity”

99. Like Djuric, Cox does not disclose “output summary statistics of expected human responses describing a hidden context of the input traffic entity.”

100. The Petition and Dr. Janét’s Declaration attempt to rewrite the claim language such that “the Djuric-Cox model generates predicted trajectories (*hidden context*) and trajectory confidence levels (*summary statistics*) reflective of human-labeled image data that is used to train the model (*expected human responses*).” Pet. at 57; Ex. 1003 at ¶ 164 (emphasis in original). In other words, they state that the Djuric-Cox model generates hidden context and summary statistics reflective of expected human responses. But this is not what the claim requires. The claim

language plainly states “output summary statistics of expected human responses describing a hidden context of the input traffic entity.” Ex. 1001, claim 1. The grammatical structure of the claim is straightforward: 1) the output is summary statistics; 2) the summary statistics are of expected human responses; and 3) the expected human responses describe hidden context of a traffic entity. These are three separate requirements that must each be satisfied.

101. Instead of analyzing the claim language, Tesla and Dr. Janét appear to rewrite the claim language, by substituting (as Tesla and Dr. Janét allege) (1) predicted trajectories for hidden context, (2) trajectory confidence levels for summary statistics, and (3) “reflective of human-labeled image data that is used to train the model” for expected human responses. Below I have shown how this change rewrites the claim language:

<u>Actual Claim Language</u>	<u>Tesla and Dr. Janét’s Interpretation</u>
a machine learning model configured to . . . output summary statistics of expected human responses describing a hidden context of the input traffic entity	a machine learning model configured to . . . generate hidden context and summary statistics reflective of human- expected human responses

Ex. 1001, claim 1; Pet. at 57. This “interpretation” does not reflect the actual language of the challenged claims.

102. Even if Tesla and Dr. Janét’s version of the claims was accepted, the output of their Djuric-Cox model would still have to relate to hidden context of a traffic entity, which it does not. Djuric and Cox do not teach hidden context in an image. Djuric does not teach hidden context of a traffic entity for the same reasons I discuss above. *See* Section VIII.B. Nowhere do Tesla and Dr. Janét allege that Cox teaches this feature. Instead, Tesla and Dr. Janét merely argue that Cox (when combined with Djuric) teaches that Djuric’s “predicted trajectories and trajectory confidence levels” “are generated in accordance with the human-annotated/manually-labeled training data [of Cox].” Pet. at 58; Ex. 1003, ¶ 168.

103. I have considered the portions of Djuric 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 Djuric teaches this claim limitation. *See* Pet. at 57-59; Ex. 1003, ¶¶ 164-168. Instead, those portions, as discussed above, make clear that Djuric and Cox do not teach “output summary statistics of expected human responses describing a hidden context of the input traffic entity.”

104. At least certain of the portions of the Petition and Dr. Janét’s declaration discussed above relate to claim 1 of the ’346 patent and Ground 2. Both the Petition and Dr. Janét’s declaration refer back to these discussions when discussing this claim limitation for the other independent claims in Ground 2 as well as in Grounds 4, 6, and 8. *See* Pet. at 59, 63-64, 70-71; Ex. 1003, ¶¶ 169, 179, 193. In my opinion, Djuric

and Cox fail to teach or suggest this limitation for the same reasons as discussed above.

105. In my opinion, Djuric and Cox do not disclose “output summary statistics of expected human responses describing a hidden context of the input traffic entity.”

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

106. In my opinion, numerous claim elements are absent from the asserted prior art, and therefore the challenged claims are not obvious. Further, Tesla and Dr. Janét do not establish that a POSITA would have been motivated to combine Djuric and Cox.

107. First, Tesla and Dr. Janét’s alleged motivation to combine is essentially that both Cox and Djuric relate to “human-like” predictions, but this is too broad. Nearly every machine learning system (and especially those that are intended for use in autonomous vehicles) including supervised machine learning systems as disclosed in Cox aspire to “human-like” performance. *See* Ex. 2053 at p. VII (“Neural networks were developed to simulate the human nervous system for machine learning tasks by treating the computational units in a learning model in a manner similar to human neurons. The grand vision of neural networks is to create artificial intelligence by building machines whose architecture simulates the computations in the human nervous system.”). In essence, 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.

108. Hence, the specific combination Tesla and Dr. Janét propose, i.e., using Cox’s crowd-sourced psychometric annotation framework with Djuric’s trajectory 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. 1006, [0016], [0048], [0049]. In contrast, Djuric is generally concerned with observing the

environment surrounding an autonomous vehicle and predicting the trajectories of objects. *See, e.g.*, Ex. 1004, [0021]. Hence, Djuric's system is not concerned or designed for classification purposes as described in Cox but rather for predicting trajectories of objects in a road scene—a different application altogether that requires a different design of a machine learning system with different training/validation/test datasets.

109. These two technological issues are fundamentally different, and in my opinion, a POSITA looking at Djuric's disclosure would not look to Cox's disclosure to address either issue Tesla and Dr. Janét identify.

110. In my opinion, Cox and Djuric are also directed to different endeavors within the field of machine learning. Cox is generally concerned with improving machine learning models for analyzing the presence/non-presence (and characteristics of) a potential face in an image. Ex. 1006, [0002], [0013], [0014], [0016], [0017], [0033], [0039]-[0050]. In contrast, Djuric is generally concerned with improving machine learning models for predicting trajectories of objects in a road scene. *See, e.g.*, Ex. 1004 at [0005], claim 1. In my opinion, given the particular challenges of these problems within the field of machine learning, a POSITA would not have been motivated to look to the disclosures of Djuric to combine them with those of Cox.

111. 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. 2051 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. 2052 at 1. These examples of other researchers in the field of intelligent and autonomous vehicle systems noting the unique challenges of autonomous vehicle operating environments only further reflects that a POSITA would not have been motivated to look to disclosures in that field to combine them with Cox’s crowdsourced annotation system for improving face detection, which does not present the same unique challenges.

112. Tesla and Dr. Janét do not identify a specific technical deficiency in Djuric that would have motivated a POSITA to modify Djuric with Cox. Tesla and Dr. Janét assert that “Cox thus provides a natural expansion and improvement on Djuric’s base disclosure,” but they do not explain how such a “natural expansion” would be undertaken. Pet. at 54; Ex. 1003 at ¶¶ 159, 160. As an example, the loss

functions of Djuric and Cox solve fundamentally different mathematical problems. Cox's loss is a classification loss that assigns the cost of misclassifying a discrete, categorical label (face/no-face, attribute present/absent). *See, e.g.*, Ex. 1006 at [0008]. Djuric's confidence model is a probabilistic regression problem. Djuric predicts continuous spatial waypoints drawn from a 2D Gaussian distribution with a learned covariance and is trained by minimizing negative log-likelihood. *See, e.g.*, Ex. 1004 at [0086]. While Djuric does use a cross-entropy classification term, it is only to estimate a per-trajectory probability among multiple predicted modes, not to label a perceptual category. *See* Ex. 1004 at [0090]. A POSITA would recognize that Cox's misclassification-penalty loss formulation has no natural mapping onto Djuric's Gaussian-Mixture/Mixture-Density-Network loss formulation, where the quantities of interest are spatial displacement error and covariance (*see* Ex. 1004 at [0086]-[0087], [0092], [0094]), not a discrete label deviation (*see, e.g.*, Ex. 1006 at [0007]). Attempting to add Cox's weighting onto Djuric's loss therefore would not predictably improve Djuric's already-calibrated outputs. In fact, it would likely distort the output of Djuric. Thus, Cox's human-weighted loss function and image-classification framework could not improve Djuric's trajectory-confidence model. *See id.*

113. Tesla and Dr. Janét simply state that “[a] POSITA seeking to implement Djuric’s model would also have looked to Cox to provide more instructions for

training a machine learning model to predict human behavior.” Pet. at 54; Ex. 1003 at ¶¶ 159, 160. There are several problems with this statement. They never identify a specific technical deficiency in Djuric that would have motivated a POSITA to modify Djuric with Cox, they only state that a POSITA would have “looked to Cox to provide more instructions for training a machine learning model to predict human behavior” (Pet. at 54; Ex. 1003 at ¶ 160), without explaining why a POSITA would need more instructions for training the machine learning model, or what those instructions are.

114. Tesla and Dr. Janét also do not explain *how* Cox would help a POSITA using Djuric “to predict human behavior” (Pet. at 54; Ex. 1003 at ¶ 160), when Cox itself does not predict human behavior. Cox trains machine learning systems to identify and classify features that are present in an image, such as a man who is smiling, has a beard, and is wearing a hat, or may identify objects such as a tree, car, bus, building, or street. Ex. 1006 at [0016]. Thus, Cox is directed to recognizing and labeling observable features and objects contained in an image – Cox does not predict human behavior.

115. Additionally, Djuric and Cox themselves teach away from Tesla and Dr. Janét’s proposed combination, for several reasons. Below I detail why.

116. Djuric emphasizes reducing computational burden and increasing efficiency. For example, Djuric explains:

The systems and methods enable the vehicle technology to more accurately predict future locations of objects, **while reducing the computational resources needed to do so Such a focused processing approach can increase computational speeds** as well as provide more accurately predicted object trajectories. Additionally, use of a machine-learned model as described herein can allow for parallelization of the process, rather than being run in sequence. Moreover, the combined data set **can allow the model's network to be smaller and, thus, use less computational resources (e.g., processing, storage, etc.) and run more efficiently.** This improvement is particularly advantageous for a vehicle computing system because the vehicle computing system has limited computational resources onboard the vehicle and is running in real-time, as the vehicle is travelling.

Ex. 1004 at [0037] (emphases added); *see also id.* at [0021], [0031].

117. In my opinion, incorporating the details of Cox in making binary classifications to label observable features would unnecessarily complicate Djuric. This would increase computational complexity and resource consumption, which would directly undermine the objectives that Djuric identifies as important advantages of its system. Thus, Djuric teaches away from the proposed modification.

118. Cox also teaches away from its combination with Djuric. Cox focuses on binary classifications such as whether a given element is present in an image, or not. *See, e.g.,* Ex. 1006 at [0016]; *see also id.* at [0048] (“annotators may be asked to press a specific key if they observe a face in an image 810 and another key if they

do not observe a face in the image”). The stated goal of Cox is to classify attributes in an image, for example, a person’s hair color or facial expression denoting happiness or sadness. *See id.*; *see also* Ex. 1006 at [0049]. Cox states that it does not deal with higher levels of complexity:

Attributes may be considered individual “building blocks” for higher-level recognition tasks, with their own complexities and challenges regarding features and training. Thus, rather than initially tackling the more complex problem of face identity, embodiments of the invention consider attributes as a distinct problem. Existing algorithms for attributes have not been informed by human behavior by any significant measure, making embodiments of the invention a novel alternative.

Id. (emphasis added). Cox is only concerned with the attribute or feature level of an image, not with processing data to make predictions about the future trajectory of an object, like Djuric. Given this, in my opinion, a POSITA would not combine Cox with Djuric.

119. Cox’s human-weighted loss is a regularizer aimed at the problem of too little labeled data, where machine-learning algorithms “tend to ‘overfit’ the data.” Ex. 1006 at [0003]. Djuric trains on large sets of real-world trajectory data drawn from prior driving events and tailored to the operating region. *See, e.g.*, Ex. 1004 at [0021]-[0023], [0031] (discussing the large amount of data that can be collected and limiting the data set to focus processing resources). A POSITA would understand

that a regularizer designed for the small amount of data and human-perception regime of Cox addresses a problem Djuric does not face. This removes the technical justification for grafting Cox's loss onto Djuric.

120. Additionally, neither Tesla nor Dr. Janét explain *what* the problem is in Djuric that would have prompted a POSITA to turn to Cox. Djuric discloses a complete, functioning system directed to trajectory prediction, and Cox discloses a complete, functioning system directed to the classification of attributes. Djuric and Cox address different technical problems and employ different approaches. As Djuric and Cox independently perform their intended functions, a POSITA would seemingly have no reason to combine them.

121. Tesla and Dr. Janét also do not explain how their proposed combination would actually be implemented, in a technical sense. For example, Tesla and Dr. Janét do not explain how Cox's human-weighted loss function would be integrated into Djuric's trajectory-predicting system. In my opinion, this fact further demonstrates that a POSITA would not have been motivated to combine Djuric and Cox.

122. In my opinion, Tesla and Dr. Janét do not establish that a POSITA would have been motivated to combine Djuric and Cox.

X. BOTH THE PETITION AND DR. JANÉT'S DECLARATION FAIL TO ANALYZE THE FACTORS REQUIRED FOR A PROPER OBVIOUSNESS ANALYSIS

123. 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 Grounds 1-8 does Dr. Janét discuss any differences between Djuric, Cox, Zhu, or Ross and the challenged claims of the '346 patent.

124. As I discussed above, specifically with respect to Djuric and Cox, Djuric's disclosure is focused on predicting a future location of an object based on the object's velocity and acceleration, known current and/or past locations of the object, and elements in the object's environment such as obstructions in the path of the object. 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. In contrast, the challenged claims of the '346 patent relate to predicting hidden context of traffic entities to better control autonomous vehicles in a way that more closely resembles how human drivers would judge the behavior of the traffic entities. Dr. Janét does not account for these key differences between Djuric and 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 Djuric and 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 Djuric and Cox do not teach or suggest.

XI. CONCLUSION

125. 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.

126. 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: July 8, 2026



Dr. Eli Saber, Ph.D.