

UNITED STATES PATENT AND TRADEMARK OFFICE

---

BEFORE THE PATENT TRIAL AND APPEAL BOARD

---

TESLA, INC.,

Petitioner,

v.

PERCEPTIVE AUTOMATA LLC,

Patent Owner.

---

Case No. IPR2025-01576

Patent No. 11,467,579

---

**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 .....                           | 12 |
| A.    | Overview .....                                  | 12 |
| B.    | Claims.....                                     | 13 |
| C.    | Burden of Proof for Invalidity .....            | 14 |
| D.    | Claim Construction.....                         | 15 |
| E.    | Person of Ordinary Skill in the Art .....       | 15 |
| F.    | Prior Art.....                                  | 15 |
| G.    | Anticipation.....                               | 16 |
| H.    | Obviousness .....                               | 17 |
| V.    | PERSON OF ORDINARY SKILL IN THE ART .....       | 20 |
| VI.   | THE '579 PATENT AND THE CHALLENGED CLAIMS ..... | 21 |
| A.    | Overview of the '579 Patent.....                | 21 |
| B.    | Claim Overview .....                            | 24 |
| VII.  | CLAIM CONSTRUCTION.....                         | 27 |
| VIII. | GROUND 1, 3, AND 5 .....                        | 27 |
| A.    | Djuric.....                                     | 28 |
| B.    | AA.....                                         | 29 |

|     |                                                                                                                                                                                                                                                                                                                             |    |
|-----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| C.  | Djuric and AA Do Not Teach Elements 1[a-iii], 13[a-iii], and 20[a-iii] (“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”) .....       | 31 |
| 1.  | Djuric Does Not Teach “hidden context” .....                                                                                                                                                                                                                                                                                | 31 |
| 2.  | Djuric Does Not Teach “the output comprising a plurality of values” .....                                                                                                                                                                                                                                                   | 35 |
| 3.  | Djuric Does Not Teach “each value representing a likelihood of receiving a particular user response” .....                                                                                                                                                                                                                  | 37 |
| D.  | Djuric Does Not Teach Elements 1[c], 13[c], and 20[c] (“executing the probabilistic neural network to generate output representing hidden context for at least a traffic entity displayed in the new image”) .....                                                                                                          | 39 |
| E.  | Djuric Does Not Teach “determining a measure of uncertainty for each of the plurality of values” (Limitations 1[d], 13[d], and 20[d])                                                                                                                                                                                       | 39 |
| F.  | A POSITA Would Not Have Been Motivated to Combine Djuric and AA in the Manner Tesla and Dr. Janét Allege .....                                                                                                                                                                                                              | 41 |
| IX. | FOUNDATIONS 2, 4, AND 6 .....                                                                                                                                                                                                                                                                                               | 45 |
| A.  | Cox .....                                                                                                                                                                                                                                                                                                                   | 46 |
| B.  | Djuric, AA, and Cox Do Not Teach Elements 1[a-iii], 13[a-iii], and 20[a-iii] (“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”) ..... | 51 |
| 1.  | Djuric, AA, and Cox Do Not Teach “hidden context” .....                                                                                                                                                                                                                                                                     | 51 |
| 2.  | Djuric and Cox Do Not Teach “the output comprising a plurality of values, each value representing a likelihood of receiving a particular user response” .....                                                                                                                                                               | 53 |

|     |                                                                                                                                                                                                                    |    |
|-----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| C.  | Djuric Does Not Teach Elements 1[c], 13[c], and 20[c] (“executing the probabilistic neural network to generate output representing hidden context for at least a traffic entity displayed in the new image”) ..... | 56 |
| D.  | Djuric Does Not Teach “determining a measure of uncertainty for each of the plurality of values” (Limitations 1[d], 13[d], and 20[d]).....                                                                         | 56 |
| E.  | A POSITA Would Not Have Been Motivated to Combine Djuric, AA, and Cox in the Manner Tesla and Dr. Janét Allege.....                                                                                                | 57 |
| X.  | BOTH THE PETITION AND DR. JANÉT’S DECLARATION FAIL TO ANALYZE THE FACTORS REQUIRED FOR A PROPER OBVIOUSNESS ANALYSIS.....                                                                                          | 62 |
| XI. | CONCLUSION.....                                                                                                                                                                                                    | 63 |

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. 11,467,579 (the “’579 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 1-20 of the ’579 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 1-6 of the Petition, render obvious claims 1-20 of the ’579 patent. In my opinion, the alleged prior art references, as presented in Grounds 1-6 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 '579 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 '579 patent.

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

| <u><b>Document</b></u>           | <u><b>Exhibit/Paper #</b></u> |
|----------------------------------|-------------------------------|
| '579 patent                      | 1001                          |
| File history for the '579 patent | 1002                          |
| Declaration of Jason Janét       | 1003                          |

|                                                                                                                                                                                                                                                                                                                                                                                    |      |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| U.S. Patent Application Publication No.<br>2019/0049970 (“Djuric”)                                                                                                                                                                                                                                                                                                                 | 1004 |
| U.S. Patent Application Publication No.<br>2020/0160535 (“AA”)                                                                                                                                                                                                                                                                                                                     | 1005 |
| International Publication WO 2014/210334<br>(“Cox”)                                                                                                                                                                                                                                                                                                                                | 1008 |
| U.S. Patent Application Publication No.<br>2018/0247200 (“Rolfe”)                                                                                                                                                                                                                                                                                                                  | 1009 |
| U.S. Patent No. 10,496,091 (“Ross”)                                                                                                                                                                                                                                                                                                                                                | 1015 |
| Petition                                                                                                                                                                                                                                                                                                                                                                           | 1    |
| Patent Owner’s Preliminary Response                                                                                                                                                                                                                                                                                                                                                | 9    |
| Behshad Mohebbali, Amirhessam Tahmassebi,<br>Anke Meyer-Baese, Amir H. Gandomi (2020),<br>Chapter 14 - Probabilistic neural networks: a brief<br>overview of theory, implementation, and<br>application, <i>Handbook of Probabilistic Models</i> ,<br>347-367, <a href="https://doi.org/10.1016/B978-0-12-816514-0.00014-X">https://doi.org/10.1016/B978-0-12-816514-0.00014-X</a> | 2050 |

|                                                                                                                                                                                                                                                                                                 |      |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| Specht, D. F. (1990), Probabilistic Neural Networks, <i>IEEE Transactions on Neural Networks</i> , 1(1), 109-118                                                                                                                                                                                | 2051 |
| Duda, R. O., Hart, P. E., & Stork, D. G. (2001), <i>Pattern Classification</i> (2nd ed.), Wiley                                                                                                                                                                                                 | 2052 |
| Haykin, S. (2009), <i>Neural Networks and Learning Machines</i> (3rd ed.), Pearson                                                                                                                                                                                                              | 2053 |
| Guo, Z.; Huang, Y.; Hu, X.; Wei, H.; Zhao, B. A Survey on Deep Learning Based Approaches for Scene Understanding in Autonomous Driving. <i>Electronics</i> <b>2021</b> , <i>10</i> , 471. <a href="https://doi.org/10.3390/electronics10040471">https://doi.org/10.3390/electronics10040471</a> | 2054 |
| Puja Shashi, Suchithra R. Review Study on Digital Image Processing and Segmentation. <i>American Journal of Computer Science and Technology</i> . Vol. 2, No. 4, 2019, pp. 68-72. doi: 10.11648/j.ajcst.20190204.14                                                                             | 2055 |

|                                                                                                                                                                                                  |      |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| M. Wang, W. Deng, Deep Face Recognition: A Survey, <i>Neurocomputing</i> (2020), doi: <a href="https://doi.org/10.1016/j.neucom.2020.10.081">https://doi.org/10.1016/j.neucom.2020.10.081</a>    | 2056 |
| 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 | 2057 |
| 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      | 2058 |
| C. Aggarwal, “Neural Networks and Deep Learning”, Springer ISBN 978-3-319-94462-3                                                                                                                | 2059 |

27. I have been asked to consider the time of the inventions of the challenged claims to be around February 6, 2019, which I understand is the filing date of the earliest provisional patent application to which the '579 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 '579 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 '579 patent claims priority to U.S. Provisional Patent Application No. 62/822,269, which was filed on March 22, 2019. I understand that the '579 patent also claims priority to U.S. Provisional Patent Application No. 62/802,151, which was filed on February 6, 2019. My opinions herein do not differ whether the time of the inventions of the challenged claims is around February 6, 2019, March 22, 2019, or February 6, 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 '579 patent.

29. As I discuss in detail below, it is my opinion that claims 1-20 of the '579 patent are not rendered obvious by the alleged prior art references, as presented in Grounds 1-6 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 '579 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 ’579 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 '579 Patent as of February 6, 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, ¶ 66.

53. For purposes of providing my opinions in connection with this declaration, I do not dispute Dr. Janét's definition, 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 '579 patent as of February 6, 2019 under either prong of Dr. Janét's definition. *See* Section II above and my CV (Ex. 2049).

## **VI. THE '579 PATENT AND THE CHALLENGED CLAIMS**

### **A. Overview of the '579 Patent**

54. The inventions disclosed and claimed in the '579 patent generally relate to systems, computer-implemented methods, and computer readable media associated with navigating autonomous vehicles by “predicting hidden context of traffic entities” “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 1:20-23, 4:8-21. The '579 patent discloses an implementation for achieving this result by training a probabilistic neural network to generate output representing the hidden context for the traffic entity. *Id.* at 1:66-2:9. A probabilistic neural network (PNN) differs fundamentally from a conventional fully connected neural network (FCNN) and convolutional neural networks (CNNs) in both its learning method and decision process. *See* Ex. 2050 at pp. 347-367; *see also* Ex. 2051 at pp. 109-118, Ex. 2052 at pp. 15-30, Ex. 2053. A PNN is a kernel-based classifier derived from Bayesian decision theory that estimates the probability density function of each class using Parzen window or similar style estimations and classifies new observations according to the maximum posterior probability without iterative weight

optimization. *See id.* Hence, rather than learning weights through iterative backpropagation, a PNN utilizes representative training samples resulting in rapid training and convergence to the Bayes-optimal classifier as the amount of training data increases. *See id.* In contrast, an FCNN consists of layers of interconnected neurons whose weights are optimized iteratively through gradient-based backpropagation to minimize a loss function. *See id.* Likewise, a CNN is a deep learning architecture that automatically learns hierarchical spatial features from structured data such as images through convolutional filters also optimized via gradient-based backpropagation. *See id.* Hence, FCNNs and CNNs typically require substantially longer training times, careful hyperparameter tuning, and often converge to local minima, while PNNs emphasize fast training and probabilistic classification based on similarity to stored training examples. *See id.* A PNN is “a multi-task neural network,” and in the context of the ’579 patent, it is “configured to predict a plurality of output values representing different hidden context attributes” (Ex. 1001 at 6:4-6) that “may generate different outputs for the same input if the neural network is executed repeatedly [and] the outputs generated have a particular statistical distribution, for example, mean and standard deviation” (*id.* at 7:22-24). The PNN disclosed and claimed in the ’579 patent is capable of performing several operations including “receiving as input, an image of traffic, the image displaying a traffic entity belonging to the traffic” and “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.” *Id.*, claim 1. Hidden context in an image 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,” “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 autonomous vehicle by a user” (*id.* at 2:64-3:15), “awareness of a bicyclist that a vehicle is driving close to the bicyclist,” or the “intent of a pedestrian” (*id.* at 5:42-46). Additionally, the system of the ’579 patent also relies on user responses that are “request[ed] from users observing images of traffic entities,” where “each user response is one of a plurality of values [that] corresponds to a rating that the user provides to a hidden context of the traffic entity . . . indicat[ing] what the user believes is the value of the hidden context attribute, for example, on a scale from 1-5, a number indicating how likely the user believes, a pedestrian is likely to cross the street or how likely a bicyclist is aware of a vehicle.” *Id.* at 7:4-15.

## B. Claim Overview

55. Independent claims 1, 13, and 20 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 for navigating autonomous vehicles, comprising:                                                                                                                                                                              |
| 1[a]     | training a probabilistic neural network, the probabilistic neural network configured to perform steps comprising:                                                                                                                     |
| 1[a-i]   | receiving as input, an image of traffic, the image displaying a traffic entity belonging to the traffic,                                                                                                                              |
| 1[a-ii]  | generating a feature vector for a plurality of features, the feature vector comprising values describing statistical distribution for each feature, and                                                                               |
| 1[a-iii] | 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; |
| 1[b]     | receiving a new image captured by a camera mounted on an autonomous vehicle navigating through traffic;                                                                                                                               |
| 1[c]     | executing the probabilistic neural network to generate output representing hidden context for at least a traffic entity displayed in the new image;                                                                                   |
| 1[d]     | determining a measure of uncertainty for each of the plurality of values; and                                                                                                                                                         |
| 1[e]     | navigating the autonomous vehicle to avoid the traffic entity displayed in the new image, the navigation based on at least the                                                                                                        |

|                 |                                                                                                                                                                                                                                                                                                                  |
|-----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                 | measure of uncertainty generated by the probabilistic neural network.                                                                                                                                                                                                                                            |
| <b>Claim 13</b> | <b>Limitations</b>                                                                                                                                                                                                                                                                                               |
| 13[pre]         | A non-transitory computer readable storage medium storing instructions, that when executed by a processor, cause the processor to perform steps comprising:                                                                                                                                                      |
| 13[a]           | training a probabilistic neural network, the probabilistic neural network configured to perform steps comprising:                                                                                                                                                                                                |
| 13[a-i]         | [the probabilistic neural network configured to perform steps comprising:] receiving as input, an image of traffic, the image displaying a traffic entity belonging to the traffic,                                                                                                                              |
| 13[a-ii]        | [the probabilistic neural network configured to perform steps comprising:] generating a feature vector for a plurality of features, the feature vector comprising values describing statistical distribution for each feature, and                                                                               |
| 13[a-iii]       | [the probabilistic neural network configured to perform steps comprising:] 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; |
| 13[b]           | receiving a new image captured by a camera mounted on an autonomous vehicle navigating through traffic;                                                                                                                                                                                                          |
| 13[c]           | executing the probabilistic neural network to generate output representing hidden context for at least a traffic entity displayed in the new image;                                                                                                                                                              |
| 13[d]           | determining a measure of uncertainty for each of the plurality of values; and                                                                                                                                                                                                                                    |
| 13[e]           | navigating the autonomous vehicle to avoid the traffic entity displayed in the new image, the navigation based on at least the measure of uncertainty generated by the probabilistic neural network.                                                                                                             |

| <b>Claim 20</b> | <b>Limitations</b>                                                                                                                                                                                                                                                                                               |
|-----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 20[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:                                                                                                      |
| 20[a]           | training a probabilistic neural network, the probabilistic neural network configured to perform steps comprising:                                                                                                                                                                                                |
| 20[a-i]         | [the probabilistic neural network configured to perform steps comprising:] receiving as input, an image of traffic, the image displaying a traffic entity belonging to the traffic,                                                                                                                              |
| 20[a-ii]        | [the probabilistic neural network configured to perform steps comprising:] generating a feature vector for a plurality of features, the feature vector comprising values describing statistical distribution for each feature, and                                                                               |
| 20[a-iii]       | [the probabilistic neural network configured to perform steps comprising:] 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; |
| 20[b]           | receiving a new image captured by a camera mounted on an autonomous vehicle navigating through traffic;                                                                                                                                                                                                          |
| 20[c]           | executing the probabilistic neural network to generate output representing hidden context for at least a traffic entity displayed in the new image;                                                                                                                                                              |
| 20[d]           | determining a measure of uncertainty for each of the plurality of values; and                                                                                                                                                                                                                                    |
| 20[e]           | navigating the autonomous vehicle to avoid the traffic entity displayed in the new image, the navigation based on at least the measure of uncertainty generated by the probabilistic neural network.                                                                                                             |

## **VII. CLAIM CONSTRUCTION**

56. I understand that, to date, neither Perceptive nor Tesla has proposed any constructions for any terms in the challenged claims of the '579 patent. Based on my review of the '579 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.

57. 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 '579 patent specification, I understand that the plain and ordinary meaning of the term “hidden context,” as used in the '579 patent, “represent[s] intentions and future plans of a traffic entity” (Ex. 1001 at 14:53-55) and “includes factors that affect the behavior of[] traffic entities, for example, a state of mind of a pedestrian, a degree of awareness of the existence of the autonomous vehicle in the vicinity (for example, whether a bicyclist is aware of the existence of the autonomous vehicle in the proximity of the bicyclist), and so on” (*id.* at 4:12-18).

## **VIII. GROUNDS 1, 3, AND 5**

58. I understand that Tesla alleges (and Dr. Janét opines) that claims 1-20 of the '579 patent are obvious in view of Djuric and AA. *See* Pet. at 16-55, 65-74; Ex. 1003, ¶¶ 68-152; 170-188. For the reasons I discuss below, I disagree. In my

opinion, none of claims 1-20 are obvious in view of Djuric and AA, either alone or in combination.

**A. Djuric**

59. In arriving at my opinions concerning Tesla and Dr. Janét's assertions regarding claims 1-20 of the '579 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.

60. 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].

61. 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.

62. 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 ’579 patent. Ex. 2054 at p. 1.

**B. AA**

63. In arriving at my opinions concerning Tesla and Dr. Janét’s assertions regarding claims 1-20 of the ’579 patent and AA, I have considered AA’s disclosure from the perspective of a POSITA. Below I discuss some aspects of AA that are relevant to my opinions.

64. As AA indicates, it is directed to a method for predicting subject motion” based on “training data comprising a set of subject pose trees . . . wherein each subject pose tree . . . indicates a location along the axis of the image at which each of a plurality of joints of a subject is located” in order “to generate a probability distribution . . . indicating a likelihood that the joint will be located at each location along the axis.” Ex. 1005 at [0005], claim 1.

65. AA uses a convolutional neural network to predict motion of a subject based on the subject’s body poses. *See id.* at [0001]-[0005]. AA explains that its convolutional neural network is “designed to recognize visual features from an image” in a “hierarchy.” *Id.* at [0029], [0034]. The first layer learns to “recognize relatively simple features, such as edges.” *Id.* at [0029]. The second layer “taking the output of the first layer as input, may learn to recognize combinations of features, such as simple shapes for visual data.” *Id.* “[H]igher layers may learn to represent complex shapes in visual data or . . . may learn to recognize common visual objects.” *Id.* I note that these extracted image features are used to identify and track the locations of a subject’s joints and body poses. *See, e.g., id.* at [0062].

66. The training data of AA consists of pose trees identifying the locations of joints within an image, and the output of AA is a probability distribution predicting the location of each joint (*see id.* [0005], claim 1), and thus, both the inputs to and outputs of AA concern a subject’s physical pose and motion. AA does

not teach inferring hidden context from an image, predicting information that is not tied to the observed pose of a subject, or generating representations of latent contextual information underlying image content.

**C. Djuric and AA Do Not Teach Elements 1[a-iii], 13[a-iii], and 20[a-iii] (“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”)**

67. Djuric and AA do not teach “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” as required by element [a-iii] of each of claims 1, 13, and 20. Below I discuss why different limitations of this claim element are not met by Djuric or AA.

**1. Djuric Does Not Teach “hidden context”**

68. 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 teaches the claimed “hidden context,” and they instead rely exclusively on Djuric for this limitation. I also understand that Tesla and Dr. Janét 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.

69. As discussed above, in my opinion, the term "hidden context" does not require any construction and should be given its plain and ordinary meaning. Based on my review of the '579 patent specification, in my opinion, the plain and ordinary meaning of the term "hidden context," as used in the '579 patent, "represent[s] intentions and future plans of a traffic entity" (Ex. 1001 at 14:53-55) and "includes factors that affect the behavior of[] traffic entities, for example, a state of mind of a pedestrian, a degree of awareness of the existence of the autonomous vehicle in the vicinity (for example, whether a bicyclist is aware of the existence of the autonomous vehicle in the proximity of the bicyclist), and so on" (*id.* at 4:12-18). As such, in my opinion, the plain and ordinary meaning of "hidden context" encompasses information that is not directly observable from an image.

70. The '579 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," "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 autonomous vehicle by a user" (Ex. 1001 at 2:64-3:15), "awareness of a bicyclist that a vehicle is driving close to the

bicyclist,” or the “intent of a pedestrian” (*id.* at 5:42-46). In my opinion, neither Djuric does not teach these or similar concepts.

71. Djuric’s “nominal pathways correspond to ‘common patterns of object travel.’” Pet. at 29-30; Ex. 1003 at ¶ 93. 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.” Ex. 1004 at [0081]. Therefore, the “goal” in Djuric is for the object to travel along a pathway in a commonly recognized pattern. *See id.*

72. The “goal-based prediction” of Djuric is fundamentally different from the “hidden context” disclosed in the ’579 patent. As previously discussed, the ’579 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 4:19-22. The ’579 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.

73. 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.

74. I have considered the paragraphs 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 the claimed “generating output representing hidden context.” *See* Pet. at 29-31; Ex. 1003, ¶¶ 92-96.

75. At least certain of the portions of the Petition and Dr. Janét’s declaration discussed above relate to claim 1 of the ’579 patent and Ground 1. Both the Petition and Dr. Janét’s declaration merely refer back to these discussions when discussing elements 13[a-iii] and 20[a-iii] for Ground 1, as well as for its discussions of this (and all) elements of the challenged independent claims for Grounds 3 and 5. *See*

Pet. at 53, 55, 67, 70; Ex. 1003, ¶¶ 132, 148, 175, 182. In my opinion, the combination of Djuric and AA fails to teach or suggest this aspect of elements 13[a-iii] and 20[a-iii] for Ground 1, as well as elements 1[a-iii], 13[a-iii], and 20[a-iii] for Grounds 3 and 5 for the same reasons as discussed above.

76. Based on my analysis of the Petition and Dr. Janét’s declaration, Tesla and Dr. Janét do not identify where Djuric supposedly teaches “generating output representing” hidden context. As Djuric does not teach the claimed “hidden context,” then neither can generate an output representing such hidden context. Therefore, Djuric does not teach “generating output representing hidden context.”

**2. Djuric Does Not Teach “the output comprising a plurality of values”**

77. I understand that Tesla does not contend that AA teaches the claimed “the output comprising a plurality of values,” and that Tesla only contends that Djuric teaches this limitation. *See* Pet. at 29-31, 53, 55.

78. I understand that Tesla’s argument as to this limitation is that “[t]he model output in Djuric also compris[es] a plurality of values because the output includes ‘a trajectory confidence level for each predicted trajectory.’” Pet. at 30; Ex. 1003 at ¶ 94. Based on my analysis of the Petition and Dr. Janét’s declaration, I understand that this is the extent of Tesla’s argument and Dr. Janét does not offer any additional explanation. *See id.*

79. Tesla and Dr. Janét’s arguments fail to consider the proper relationship between the limitations of the claim element. The claims of the ’579 patent require the “output representing hidden context” to comprise “a plurality of values.” Thus, the claimed values are tied to, and describe, the hidden context represented by the output – the claimed values are not simply values of any kind. Therefore, Tesla and Dr. Janét’s argument that Djuric “compris[es] a plurality of values because the output includes ‘a trajectory confidence level’” (Pet. at 30; Ex. 1003 at ¶ 94) is technically incorrect because it does not satisfy the claim requirement that the “values” be tied to and describe the hidden context represented by the output.

80. Further, as previously discussed, Djuric does not teach “generating output representing hidden context.” The output of Djuric is predicted trajectories, and confidence levels are associated with those trajectories. *See* Pet. at 30; Ex. 1003 at ¶ 94. Djuric’s output does not represent hidden context, and therefore its confidence levels do not describe hidden context.

81. I have considered the paragraphs 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 the claimed “the output comprising a plurality of values.” *See* Pet. at 29-31; Ex. 1003, ¶¶ 92-96.

82. At least certain of the portions of the Petition and Dr. Janét’s declaration discussed above relate to claim 1 of the ’579 patent and Ground 1. Both the Petition

and Dr. Janét’s declaration merely refer back to these discussions when discussing elements 13[a-iii] and 20[a-iii] for Ground 1, as well as for its discussions of this (and all) elements of the challenged independent claims for Grounds 3 and 5. *See* Pet. at 53, 55, 67, 70; Ex. 1003, ¶¶ 132, 148, 175, 182. In my opinion, the combination of Djuric and AA fails to teach or suggest this aspect of elements 13[a-iii] and 20[a-iii] for Ground 1, as well as elements 1[a-iii], 13[a-iii], and 20[a-iii] for Grounds 3 and 5 for the same reasons as discussed above.

83. In my opinion, Djuric does not teach “the output comprising a plurality of values.”

**3. Djuric Does Not Teach “each value representing a likelihood of receiving a particular user response”**

84. Based on my review of the Petition and Dr. Janét’s declaration, I do not understand Tesla and Dr. Janét to allege that AA teaches the claimed “each value representing a likelihood of receiving a particular user response,” and that Tesla and Dr. Janét instead only contend that Djuric teaches this limitation. *See* Pet. at 29-31, 53, 55.

85. Tesla and Dr. Janét allege that Djuric teaches using “supervised training techniques,” including using “labeled” data and that “[s]uch labeled data used in supervised training is data labeled by a human.” Pet. at 30-31; Ex. 1003 at ¶ 95. I disagree. Tesla does not identify any teaching in Djuric that states that the labels are provided by human users as Tesla alleges, much less any teaching of a value

representing a likelihood of receiving a particular user response, as required by the challenged claims. In the Petition, Tesla only cites to Dr. Janét's declaration, which states "[i]t is my opinion that such labeled data used in supervised training is data labeled by a human." *Id.* But Dr. Janét's declaration does not provide any supporting analysis or evidence.

86. I have considered the paragraphs 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 limitation *See* Pet. at 29-31; Ex. 1003, ¶¶ 92-96.

87. At least certain of the portions of the Petition and Dr. Janét's declaration discussed above relate to claim 1 of the '579 patent and Ground 1. Both the Petition and Dr. Janét's declaration merely refer back to these discussions when discussing the corresponding limitations for claims 13 and 20 for Ground 1, as well as for its discussions of this (and all) elements of the challenged independent claims for Grounds 3 and 5. *See* Pet. at 53, 55, 67, 70; Ex. 1003, ¶¶ 132, 148, 174, 182. In my opinion, the combination of Djuric and AA fails to teach or suggest the corresponding limitations for claims 13 and 20 for Ground 1, as well as these limitations in all three independent claims for Grounds 3 and 5 for the same reasons as discussed above.

88. Therefore, Djuric does not teach “each value representing a likelihood of receiving a particular user response.”

**D. Djuric Does Not Teach Elements 1[c], 13[c], and 20[c] (“executing the probabilistic neural network to generate output representing hidden context for at least a traffic entity displayed in the new image”)**

89. As discussed in Section VIII.C above, in my opinion, Djuric does not teach hidden context. Tesla and Dr. Janét again allege for these elements that Djuric’s goal-based prediction teaches hidden context. *See* Pet. at 33, 53, 55; Ex. 1003 at ¶¶ 101, 134, 150. In my opinion, Djuric fails to teach the hidden context recited in these elements for the same reasons discussed in Section VIII.C above.

**E. Djuric Does Not Teach “determining a measure of uncertainty for each of the plurality of values” (Limitations 1[d], 13[d], and 20[d])**

90. In my opinion, given how Tesla and Dr. Janét have attempted to map Djuric’s teachings to the language of the challenged claims, Djuric does not teach “determining a measure of uncertainty for each of the plurality of values,” as required by the challenged claims.

91. As discussed above with respect to elements 1[a-iii], 13[a-iii], and 20[a-iii], Tesla and Dr. Janét allege that Djuric’s trajectory confidence levels are the claimed “plurality of values.” *See* Section VIII.C.2. Here, however, Tesla and Dr. Janét allege that the claimed “measure of uncertainty” is Djuric’s “way-point confidence level.” Pet. at 34, 53, 55; Ex. 1003, ¶¶ 101, 134, 150. However, Djuric

makes clear that these way-point confidence levels relate to way-points within a predicted trajectory. *See* Ex. 1004 at [0085]. Nothing in Djuric describes determining these way-point confidence levels for any trajectory confidence level, i.e., Tesla and Dr. Janét’s alleged “plurality of values.”

92. I have considered the paragraphs 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 limitation *See* Pet. at 33, 53, 55; Ex. 1003 at ¶¶ 101, 134, 150.

93. At least certain of the portions of the Petition and Dr. Janét’s declaration discussed above relate to claim 1 of the ’579 patent and Ground 1. Both the Petition and Dr. Janét’s declaration merely refer back to these discussions when discussing the corresponding limitations for claims 13 and 20 for Ground 1, as well as for its discussions of this (and all) elements of the challenged independent claims for Grounds 3 and 5. *See* Pet. at 53, 55, 67, 70; Ex. 1003, ¶¶ 135, 151, 174, 182. In my opinion, the combination of Djuric and AA fails to teach or suggest the corresponding limitations for claims 13 and 20 for Ground 1, as well as these limitations in all three independent claims for Grounds 3 and 5 for the same reasons as discussed above.

**F. A POSITA Would Not Have Been Motivated to Combine Djuric and AA in the Manner Tesla and Dr. Janét Allege**

94. 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 AA.

95. Tesla and Dr. Janét’s first proposed modification to Djuric in view of AA is that “[t]he model in the Djuric-AA combination teaches a probabilistic convolutional neural network.” Pet. at 20; Ex. 1003 at ¶ 78. However, neither Tesla nor Dr. Janét explain *what* problem in Djuric would have prompted a POSITA to turn to AA for this modification. In addition, Tesla and Dr. Janét claim that both Djuric and AA already teach a PNN. Pet. at 23-24; Ex. 1003 at ¶¶ 86, 87. If that is the case, Tesla and Dr. Janét never explain why a POSITA considering Djuric would modify Djuric in view of AA to implement a PNN.

96. Petitioner’s second proposed modification to Djuric in view of AA is that “[t]he model in the Djuric-AA combination generates a feature vector as described in AA.” Pet at 20. Petitioner does not identify any shortcoming of Djuric that would be addressed through this modification.

97. Tesla and Dr. Janét also make many generic assertions like that the combination would have been “a natural, predictable refinement,” that implementation “would also be straightforward,” that the combination would use

“known techniques,” and that it would be a “predictable combination of well-known prior art elements.” Pet. 21-22, 59-60; Ex. 1003 at ¶¶ 80-82. Tesla and Dr. Janét only assert that a POSITA *could* combine the references, not that a POSITA *would have been motivated* to do so.

98. Djuric and AA themselves teach away from combination, for several reasons. Below I detail why.

99. 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].

100. Alternatively, AA relies on sophisticated image-processing techniques, including pose estimation, image segmentation, registration, and modeling detailed relationships among body joints. *See* Ex. 1005 at [0035]-[0038], [0063], [0070]; *see also* Ex. 2055 at p. 1 (“Some common technique of image segmentation such as edge detection, thresh holding, region growing and clustering . . . . Image segmentation is one of the most important steps in digital image processing. It also involves great area of research in computer vision. The main and foremost objective of image segmentation is to filter and extract region of interest for analysis of image. Hence it may also be defined as division and sub division of an image into multiple structure and sub structure . . . . Segmentation also depends on various characteristics such as colour or texture.”). For example:

Deep learning methods, such as convolutional neural networks, use a cascade of multiple layers of processing units for feature extraction and transformation. They learn multiple levels of representations that correspond to different levels of abstraction. The levels form a hierarchy of concepts, showing strong invariance to the face pose, lighting, and expression changes, as shown in Fig. 2. It can be seen from the figure that the first layer of the deep neural network is somewhat similar to the Gabor feature found by human scientists with years of experience. The second layer learns more complex texture features. The features of the third layer are more complex, and some simple structures have begun to appear, such as high-bridged nose and big eyes. In the fourth, the network output is enough to explain a certain facial attribute,

which can make a special response to some clear abstract concepts such as smile, roar, and even blue eye.

Ex. 2056 at p. 1. Incorporating AA into Djuric would substantially increase computational complexity and resource consumption, which would directly undermine what Djuric identifies as important advantages of its system. Thus, Djuric teaches away from combination with AA.

101. Similarly, AA teaches away from Djuric's approach. AA explains that models that "use predicted [] locations at one point in time to predict [] locations at a future point in time" are not meant to handle "the complexity of subject body motion, occlusion of joints in captured images, false detections of joints, and the like." Ex. 1005 at [0051]. Djuric is such a system. Djuric "use[s] predicted [] locations at one point in time to predict [] locations at a future point in time," as it is concerned with predicting trajectories of moving objects. *See* Ex. 1004 at [0002]. AA instead focuses on detailed modeling of body poses and joint relationships. AA's detailed pose-tree analysis therefore serves a fundamentally different purpose than Djuric's trajectory-prediction framework. A POSITA seeking Djuric's streamlined trajectory prediction would have little reason to adopt AA's significantly more complex architecture.

102. In my opinion, Tesla and Dr. Janét's theories were constructed using hindsight. The Petition alleges that Djuric teaches a PNN because its trained model may include a Bayesian network, while AA allegedly teaches a PNN because it

generates probability distributions. *See* Pet. at 23-24; Ex. 1003 at ¶¶ 86, 87. Tesla and Dr. Janét then state:

**The '579 Patent[] explains that its neural network can be a “probabilistic convolutional neural network.” EX1001, 5:56-60. In view of this teaching and as described above, it would have been obvious to a POSITA to implement Djuric using a PNN as taught by AA (a probabilistic neural network).**

*Id.* (emphasis added). Tesla and Dr. Janét say that, in view of the teachings of the '579 patent, it would have been obvious to a POSITA to implement Djuric using the PNN of AA. But this expressly uses the teachings of the '579 patent as the reason to combine the prior-art references. Thus, Tesla and Dr. Janét began their analysis with the patented solution, and worked backwards to construct their combination.

103. In my opinion, Tesla and Dr. Janét have not demonstrated that a POSITA would have been motivated to combine Djuric and AA.

## **IX. GROUNDS 2, 4, AND 6**

104. I understand that Tesla alleges (and Dr. Janét opines) that claims 1-20 of the '579 patent are obvious in view of Djuric, AA, and Cox. *See* Pet. at 55-74; Ex. 1003, ¶¶ 153-188. For the reasons I discuss below, I disagree. In my opinion, none of claims 1-20 are obvious in view of Djuric, AA, and Cox, either alone or in combination.

## A. Cox

105. In arriving at my opinions concerning Tesla and Dr. Janét’s assertions regarding claims 1-20 of the ’579 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.

106. As Cox indicates, its “invention relates to machine learning, in particular to machine learning augmented by measurements acquired from human populations.” Ex. 1008, [0001].

107. 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

'579 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. 1008, [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.

108. 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. 1008, [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.*

109. 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. 1008, [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. 1008, [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. 1008, [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. 1008, [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. 1008, [0009].

110. 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. 1008, [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. 1008, [0035].

111. 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  $w_t(x)$  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. 1008, [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 taught in Cox (*see* [0047]-[0049] and [0052]-[0055]) and illustrated in figures 4A/4B, 5, 6A/6B, 7A/7B, 8A/8B, 9-12.

112. 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. 1008, [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, AA, and Cox Do Not Teach Elements 1[a-iii], 13[a-iii], and 20[a-iii] (“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”)**

113. Djuric, AA, and Cox do not teach “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” as required by element [a-iii] of each of claims 1, 13, and 20. Above I have discussed why different limitations of this claim element are not met by Djuric and AA, and those deficiencies apply equally here. Below I discuss the different limitations of this claim element are also not met by Cox.

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

114. I understand that Tesla does not contend that Cox (or AA) teaches the claimed “hidden context,” and that Tesla only contends that Djuric teaches this limitation. *See* Pet. at 29-31, 53, 55-64. As I discuss in Section VIII.C above, Djuric does not teach hidden context, and those same reasons apply equally to these grounds. Even though Tesla and Dr. Janét do not allege that Cox teaches hidden context, I have nevertheless considered Cox’s disclosure, and in my opinion, Cox does not teach “hidden context.”

115. 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. 1008 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].

116. 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.* at [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.

117. The claimed “hidden context” of the ’579 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,” “a near term goal,” “a degree of awareness of the autonomous vehicle by a user,” “awareness of a bicyclist that a vehicle is driving close to the bicyclist,” or the “intent of a pedestrian” (Ex. 1001, 2:64-3:15, 5:42-46)). The claimed “hidden context” requires inferences regarding mental state, awareness, intentions, or likely behavior. Cox does not makes 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 teach “hidden context.”

118. Based on my analysis of the Petition and Dr. Janét’s declaration, Tesla and Dr. Janét do not identify where Cox supposedly teaches “generating output representing” hidden context. As none of Cox, Djuric, or AA teach the claimed “hidden context,” then correspondingly, none can generate an output representing such hidden context. Therefore, none of Cox, Djuric, nor AA teach “generating output representing hidden context.”

**2. Djuric and Cox Do Not Teach “the output comprising a plurality of values, each value representing a likelihood of receiving a particular user response”**

119. Like Djuric and AA, Cox does not teach “the output comprising a plurality of values, each value representing a likelihood of receiving a particular user response.”

120. The Petition and Dr. Janét’s Declaration attempt to rewrite the claim language such that “the model’s output *represent[] [s] a likelihood of receiving a particular user response from a user presented with the image.*” Pet. at 61; Ex. 1003 at ¶ 163 (emphasis in original). But this is not what the claim requires. The claim language plainly states “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.” Ex. 1001, claim 1. The grammatical structure of the claim is straightforward: 1) the output represents hidden context; 2) the output comprises a plurality of values; and 3) each value represents a likelihood of receiving a particular user response. These are three separate requirements that must each be satisfied.

121. Instead, Tesla and Dr. Janét have rewritten the claim to treat the output itself as representing a likelihood of receiving a user response. That treatment eliminates the distinction inherent in the claim language between the output and the values contained within the output. In the claim language, the output represents hidden context, while the values within that output represent likelihoods of receiving particular user responses. This distinction is significant because Tesla and Dr. Janét do not contend that Cox teaches an output comprising “a plurality of values” – instead, they rely on Djuric for that portion of the limitation. *See, e.g.*, Pet. at 61, 64. In my opinion, this is an attempt to characterize the output itself as a likelihood of

receiving a user response in order to avoid the actual claim structure in an attempt to make the combination of Cox, Djuric, and AA appear more closely aligned with the claims.

122. I have considered the paragraphs 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 limitation. *See* Pet. at 60-64; Ex. 1003, ¶¶ 162-167.

123. At least certain of the portions of the Petition and Dr. Janét's declaration discussed above relate to claim 1 of the '579 patent and Ground 2. Both the Petition and Dr. Janét's declaration merely refer back to these discussions when discussing the corresponding limitations for claims 13 and 20 for Ground 2, as well as for its discussions of this (and all) elements of the challenged independent claims for Grounds 4 and 6. *See* Pet. at 65, 67, 70; Ex. 1003, ¶¶ 168, 174, 182. In my opinion, the combination of Djuric, AA, and Cox fails to teach or suggest the corresponding limitations for claims 13 and 20 for Ground 2, as well as these limitations in all three independent claims for Grounds 4 and 6 for the same reasons as discussed above.

124. None of Djuric, AA, or Cox teach "the output comprising a plurality of values, each value representing a likelihood of receiving a particular user response."

**C. Djuric Does Not Teach Elements 1[c], 13[c], and 20[c] (“executing the probabilistic neural network to generate output representing hidden context for at least a traffic entity displayed in the new image”)**

125. As discussed in Section VIII.C above, in my opinion, Djuric does not teach hidden context. Tesla and Dr. Janét again allege for these elements that Djuric’s goal-based prediction teaches hidden context. *See* Pet. at 60, 65, 67, 70; Ex. 1003 at ¶¶ 162, 168, 174, 182. In my opinion, Djuric fails to teach the hidden context recited in these elements for the same reasons discussed in Section VIII.C above.

**D. Djuric Does Not Teach “determining a measure of uncertainty for each of the plurality of values” (Limitations 1[d], 13[d], and 20[d])**

126. As discussed in Section VIII.E above, in my opinion, given how Tesla and Dr. Janét have attempted to map Djuric’s teachings to the language of the challenged claims, Djuric does not teach “determining a measure of uncertainty for each of the plurality of values,” as required by the challenged claims. As to these grounds too, Tesla and Dr. Janét allege that Djuric’s trajectory confidence levels are the claimed “plurality of values” and the claimed “measure of uncertainty” is Djuric’s “way-point confidence level.” *See* Pet. at 60, 65, 67, 70; Ex. 1003 at ¶¶ 162, 168, 174, 182. In my opinion, this is an incorrect reading of Djuric onto the claim language for the same reasons I discuss in Section VIII.E above.

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

127. In my opinion, numerous claim elements are absent from the asserted prior art, and therefore the challenged claims are not obvious. As previously explained, Tesla and Dr. Janét do not establish that a POSITA would have been motivated to combine Djuric and AA. Tesla and Dr. Janét also do not establish that a POSITA would have been motivated to further combine Djuric and AA with Cox.

128. 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 taught in Cox aspire to “human-like” performance. *See* Ex. 2050 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 taught 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.

129. 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 at [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.

130. 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.

131. 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. Ex. 1004 at [0021]. 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.

132. 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. 2057 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. 2058 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.

133. In addition, Cox states that it does not deal in 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).

134. As previously explained, AA requires complex processes such as the segmentation of frames and registration of features across imagery. AA already

performs much more complex image recognition and classification processes than does Cox. To this effect, AA is focused on sophisticated image-processing techniques, including pose estimation, image segmentation, registration, and modeling detailed relationships among body joints. *See* Ex. 1005 at [0035]-[0038], [0063], [0070]; *see also* Ex. 2055 at p. 1. A POSITA would have no reason to look to Cox, because AA's systems already utilize convolutional neural networks and probabilistic models to perform its multi-class classifications. In other words, the probabilistic models of AA would be more than suitable to perform appropriate classifications because they would not only provide the proper class for a given element that is being classified, but also its probability of belonging to that class. Furthermore, the probabilistic models of AA are suited for multi-class classification given their ability to provide both an appropriate class label for a given element as well as its probability of belonging to a given class. This is more desired than just a simple binary style classification as described in Cox.

135. As they similarly failed to do with respect to Djuric and AA alone, Tesla and Dr. Janét do not explain *what* problem or technical deficiency found in Djuric and AA would have prompted/motivated a POSITA to turn to Cox. Djuric discloses a fully functioning system; AA discloses a distinct, fully functioning system; and Cox also discloses a distinct, fully functioning system. The working system of Cox has as its primary focus the classification of attributes. *See, e.g.*, Ex.

1008 at [0049]. This addresses a different problem than both the working system of Djuric, which is focused on predicting the trajectory of a moving object (*see, e.g.*, Ex. 1004 at [0031]), and the working system of AA, which is focused on predicting the motion of a subject based on the subject's body poses (*see, e.g.*, Ex. 1005 at [0001]-[0005]). Further, Tesla and Dr. Janét do not explain how, for example, the human-weighted loss function of Cox for binary classification of attributes would be integrated into the trajectory-predicting system of Djuric. As Djuric, AA, and Cox all independently perform their intended functions, a POSITA would seemingly have no reason to combine them.

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

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

137. 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 1-6 does Dr. Janét discuss any differences between Djuric, AA, or Cox, and the challenged claims of the '579 patent.

138. As I discussed above, 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. AA's disclosure is focused on using a convolutional neural network to predict motion of a subject based on the subject's body poses. 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 '579 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, AA, 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, AA, 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, AA, and Cox do not teach or suggest.

## **XI. CONCLUSION**

139. 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.

140. 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.