

Petition for *Inter Partes* Review of U.S. Patent No. 11,126,889

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

TESLA, INC.

Petitioner

v.

PERCEPTIVE AUTOMATA LLC

Patent Owner

IPR2025-01574

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

**PETITION FOR *INTER PARTES* REVIEW OF CLAIMS 1-20 OF U.S.
PATENT NO. 11,126,889**

Mail Stop PATENT BOARD
Patent Trial and Appeal Board
United States Patent and Trademark Office
Submitted Electronically

TABLE OF CONTENTS

PETITIONER’S EXHIBIT LIST.....	VI
I. MANDATORY NOTICES UNDER 37 C.F.R. 42.8(A)(1)	1
A. Real Party-in-Interest under 37 C.F.R. 42.8(b)(1)	1
B. Related Matters under 37 C.F.R. 42.8(b)(2)	1
C. Lead and Back-Up Counsel under 37 C.F.R. 42.8(b)(3)	1
D. Service Information under 37 C.F.R. 42.8(b)(4).....	2
II. CLAIM LISTING.....	3
III. INTRODUCTION	10
IV. PAYMENT OF FEES UNDER 37 C.F.R. § 42.103.....	10
V. CERTIFICATION OF GROUNDS FOR STANDING	10
VI. OVERVIEW OF CHALLENGE AND RELIEF REQUESTED	10
A. Prior Art Printed Publications	10
B. Identification of Challenge and Statement of Precise Relief Requested	12
VII. SUMMARY OF THE ’889 PATENT	12
A. Prosecution History	12
B. Technology Background	13
C. The Alleged Invention.....	14
VIII. LEVEL OF ORDINARY SKILL IN THE ART	16
IX. CLAIM CONSTRUCTION.....	16
X. THE CHALLENGED CLAIMS ARE UNPATENTABLE.....	17
A. Ground 1: Claims 1-20 are obvious under 35 U.S.C. § 103 over Huval and Ross.....	17

Petition for *Inter Partes* Review of U.S. Patent No. 11,126,889

1.	Summary of Prior Art	18
	a. Huval–EX1004.....	18
	b. Ross–EX1005	19
2.	Motivation to Combine Huval and Ross.....	20
3.	Claim 2.....	26
	a. 2[Pre].....	26
	b. 2[a]	26
	c. 2[b]	29
	d. 2[c]	33
	e. 2[d]	37
	f. 2[e].....	40
	g. 2[f].....	40
	h. 2[g].....	41
4.	Claim 3	42
5.	Claim 4.....	43
6.	Claim 5.....	44
7.	Claim 6.....	45
8.	Claim 7.....	46
9.	Claim 8.....	48
10.	Claim 9.....	49
11.	Claim 10.....	50
12.	Claim 11.....	51
13.	Claim 12.....	51
	a. 12[Pre].....	51
	b. 12[a]	52

Petition for *Inter Partes* Review of U.S. Patent No. 11,126,889

c.	12[b]	52
d.	12[c]	52
e.	12[d]	52
f.	12[e]	52
g.	12[f]	52
h.	12[g]	52
14.	Claim 13	53
15.	Claim 14	53
16.	Claim 15	53
17.	Claim 16	53
18.	Claim 17	53
19.	Claim 18	53
20.	Claim 19	53
21.	Claim 20	53
22.	Claim 1	53
a.	1[Pre]	53
b.	1[a]	54
c.	1[b]	54
d.	1[c]	54
e.	1[d]	54
f.	1[e]	55
g.	1[f]	55
h.	1[g]	55
B.	Ground 2: Claims 1-20 are rendered obvious under 35 U.S.C. § 103 over Huval, Ross, and Cox.	55

Petition for *Inter Partes* Review of U.S. Patent No. 11,126,889

1.	Summary of Prior Art	56
a.	Cox–EX1006.....	56
2.	Motivation to Combine Huval, Ross, and Cox.....	57
3.	Claim 2	62
a.	2[Pre].....	63
b.	2[a]	63
c.	2[b]	63
d.	2[c]	63
e.	2[d]	65
f.	2[e]:.....	67
g.	2[f].....	67
h.	2[g]:.....	67
4.	Claim 10	67
5.	Claims 1, 3-9, and 11-19	68
6.	Claim 20.....	68
C.	Ground 3: Claims 1-20 are rendered obvious under 35 U.S.C. § 103 over Huval, Ross, and Munro.....	68
1.	Summary of Prior Art	69
a.	Munro–EX1009	69
2.	Motivation to Combine Huval, Ross, and Munro.....	69
3.	Claim 2	72
a.	2[Pre]-2[b].....	72
b.	2[c]	72
c.	2[d]	73
d.	2[e]	74

Petition for *Inter Partes* Review of U.S. Patent No. 11,126,889

e.	2[f].....	74
f.	2[g].....	74
4.	Claim 12.....	74
5.	Claim 1	74
6.	Claims 3-11 and 13-20.....	74
D.	Ground 4: Claims 1-20 are rendered obvious under 35 U.S.C. § 103 over Huval, Ross, Cox, and Munro.....	75
1.	Motivation to Combine Huval, Ross, Cox, and Munro.....	75
2.	Claim 2	76
a.	2[Pre]-2[b].....	76
b.	2[c]	76
c.	2[d]	76
d.	2[e]	77
e.	2[f].....	77
f.	2[g].....	77
3.	Claims 10 and 20.....	77
4.	Claim 12.....	78
5.	Claim 1	78
6.	Claims 3-9, 11, and 13-19.....	78
E.	Printed Matter.....	78
XI.	CONCLUSION	81

PETITIONER’S EXHIBIT LIST

Ex[No.]	Description of Documents
1001	U.S. Patent No. 11,126,889 (“’889 Patent”)
1002	File History of U.S. Patent No. 11,126,889
1003	Declaration of Dr. Jason Janét
1004	U.S. Patent Application Publication No. 2018/0373980 (“Huval”)
1005	U.S. Patent No. 10,496,091 (“Ross”)
1006	International Publication No. WO2014/210334 (“Cox”)
1007	File History of U.S. Patent No. 10,614,344
1008	File History of U.S. Patent No. 10,402,687
1009	U.S. Patent Application Publication No. 2016/0162456 (“Munro”)
1010	Complaint in <i>Perceptive Automata LLC v. Tesla, Inc.</i> , Case No. 2:25-cv-00742 (E.D. Tex. filed July 23, 2025)
1011	CV of Dr. Jason Janét

I. MANDATORY NOTICES UNDER 37 C.F.R. 42.8(a)(1)

A. Real Party-in-Interest under 37 C.F.R. 42.8(b)(1)

The real party-in-interest is Tesla, Inc. No other parties exercised or could have exercised control over this Petition; no other parties funded or directed this Petition.

B. Related Matters under 37 C.F.R. 42.8(b)(2)

As of the filing date of this Petition, and to the best knowledge of Petitioner, U.S. Patent No. 11,126,889 (the “’889 Patent”) is involved in *Perceptive Automata LLC v. Tesla, Inc.*, Case No. 2:25-cv-00742 (E.D. Tex. filed July 23, 2025) (“EDTX Litigation”). EX1010. To the best knowledge of Petitioner, the ’889 Patent is not involved in any other post-grant proceedings.

C. Lead and Back-Up Counsel under 37 C.F.R. 42.8(b)(3)

Petitioner provides the following designation of counsel:

Lead Counsel	Backup Counsel
Roger Fulghum, Reg. No. 39,678 Baker Botts L.L.P. One Shell Plaza 910 Louisiana Street Houston, TX 77002 Phone: (713) 229-1234 roger.fulghum@bakerbotts.com	Mark Speegle, Reg. No. 77,512 Baker Botts L.L.P. 401 S. 1 st St., Suite 1300 Austin, TX 78704-1296 Phone: (512) 322-2536 mark.speegle@bakerbotts.com Ellyar Y. Barazesh, Reg. No. 74,096 Baker Botts L.L.P. 401 S. 1 st St., Suite 1300 Austin, TX 78704-1296 Phone: (512) 322-2507 ellyar.barazesh@bakerbotts.com

Petition for *Inter Partes* Review of U.S. Patent No. 11,126,889

	William Gaines, Reg. No. 83,138 Baker Botts L.L.P. 401 S. 1st St., Suite 1300 Austin, TX 78704-1296 Phone: (512) 322-2550 william.gaines@bakerbotts.com Lance J. Goodman, Reg. No. 77,989 Baker Botts L.L.P. 401 S. 1st St., Suite 1300 Austin, TX 78704-1296 Phone: (512) 322-2629 lance.goodman@bakerbotts.com Ashraf Fawzy, Reg. No. 67,914 Tesla, Inc. 800 Connecticut Ave. NW Washington, DC 20006 Phone: (202) 905-9221 afawzy@tesla.com
--	---

D. Service Information under 37 C.F.R. 42.8(b)(4)

A copy of this entire Petition, including all Exhibits and a power of attorney, is being served by FEDERAL EXPRESS, costs prepaid, to the correspondence address of record for the '889 Patent at the USPTO:

758 - FENWICK & WEST LLP
c/o Rajendra B. Panwar (rpanwar@fenwick.com)
Silicon Valley Center
801 California Steet
Mountain View, CA 94041

and by email to the attorney or agent of record for Patent Owner in the EDTX Litigation:

Patrick J. Conroy (pat@nelbum.com)
Andrea L. Fair (andrea@millerfairhenry.com)

Please address all correspondence to lead and back-up counsel. Petitioner consents to service at lead counsel's address provided above. Petitioner consents to electronic service, provided it is made to all of the following e-mail address:

Roger Fulghum (roger.fulghum@bakerbotts.com)
Mark Speegle (mark.speegle@bakerbotts.com)
Ellyar Y. Barazesh (ellyar.barazesh@bakerbotts.com)
William Gaines (william.gaines@bakerbotts.com)
Lance J. Goodman (lance.goodman@bakerbotts.com)
Ashraf Fawzy (afawzy@tesla.com)
DLTeslavPerceptiveAutomataIPRs@bakerbotts.com

A Power of Attorney is filed concurrently herewith under 37 C.F.R. §42.10(b).

II. CLAIM LISTING

Claim 1

1[Pre]	A computer system comprising: a computer processor; and a non-transitory computer readable storage medium storing instructions that when executed by the computer processor cause the computer processor to perform steps of a computer-implemented method for controlling an autonomous vehicle based on a predicted state of mind of road users in a scene captured by a camera of the autonomous vehicle, the steps comprising:
1[a]	receiving a plurality of images displaying road scenes captured by one or more vehicles;
1[b]	receiving a plurality of user responses, each user response describing a state of mind of a road user displayed in one or more images;
1[c]	generating a training dataset comprising summary statistics of user responses describing the state of minds of road users displayed in the plurality of images;
1[d]	training, using the training dataset, a supervised learning based model configured to predict summary statistics describing a state of mind of a road user displayed in an input image;
1[e]	receiving, by an autonomous vehicle, a new image captured by a camera of the autonomous vehicle, the new image of a scene including a road user;
1[f]	predicting, by the autonomous vehicle, using the supervised learning

	based model, summary statistics describing a state of mind of the road user in the new image; and
1[g]	controlling the autonomous vehicle based on the prediction of the supervised learning based model.

Claim 2

2[pre]	A computer-implemented method for controlling an autonomous vehicle based on a predicted state of mind of road users in a scene captured by a camera of the autonomous vehicle, the method comprising:
2[a]	receiving a plurality of images displaying road scenes captured by one or more vehicles;
2[b]	receiving a plurality of user responses, each user response describing a state of mind of a road user displayed in one or more images;
2[c]	generating a training dataset comprising summary statistics of user responses describing the state of minds of road users displayed in the plurality of images;
2[d]	training, using the training dataset, a supervised learning based model configured to predict summary statistics describing a state of mind of a road user displayed in an input image;
2[e]	receiving, by an autonomous vehicle, a new image captured by a camera of the autonomous vehicle, the new image of a scene including a road user;
2[f]	predicting, by the autonomous vehicle, using the supervised learning based model, summary statistics describing a state of mind of the road user in the new image; and
2[g]	controlling the autonomous vehicle based on the prediction of the supervised learning based model.

Claim 3

3	The computer-implemented method of claim 2, wherein the road user in the input image is a pedestrian and the state of mind predicted by the supervised learning based model indicates whether the pedestrian is likely to perform an action, the action including one of: staying in place, or crossing a street.
----------	---

Claim 4

4	The computer-implemented method of claim 2, wherein the road user in the input image is a cyclist and the state of mind predicted by the supervised learning based model indicates whether the cyclist is likely to cross a lane.
---	---

Claim 5

5	The computer-implemented method of claim 2, wherein the summary statistics comprises one or more of: a central tendency, a variance, a skew, a kurtosis, a scale, or a histogram.
---	---

Claim 6

6	The computer-implemented method of claim 2, wherein receiving a user response comprises: sending one or more images for display to a viewer via a user interface, the one or more images including a road user, wherein the user interface is configured to prompt the viewer to describe how the road user will act.
---	---

Claim 7

7	The computer-implemented method of claim 2, wherein receiving a user response comprises: generating a stimulus comprising a set of images; sending the generated stimulus for display via a user interface; and receiving, via the user interface, a user response describing the stimulus.
---	---

Claim 8

8	The computer-implemented method of claim 7, further comprising: modifying an image to generate a stimulus, the modifying comprising extracting a subset of the image contained in a bounding box around the road user displayed in the image.
---	---

Claim 9

9	The computer-implemented method of claim 8, wherein the supervised learning based model uses one or more features comprising:
---	---

	dimensions of the bounding box; location of the bounding box; shape of the bounding box; or a change in size or position of the bounding box across two images representing video frames.
--	--

Claim 10

10	The computer-implemented method of claim 2, wherein the user response is determined based on: an explicit response to a question asked via a user interface; and an implicit data representing one or more of: time taken to respond; whether user deleted keystrokes; or whether user moved a mouse anywhere other than a location corresponding to a selected response.
-----------	--

Claim 11

11	The computer-implemented method of claim 2, wherein the supervised learning based model is one of: a random forest regressor, a support vector regressor, a simple neural network, a deep convolutional neural network, a recurrent neural network, or a long short-term memory (LSTM) neural network.
-----------	--

Claim 12

12[Pre]	A non-transitory computer readable storage medium storing instructions that when executed by a computer processor cause the computer processor to perform steps of a computer-implemented method for controlling an autonomous vehicle based on a predicted state of mind of road users in a scene captured by a camera of the autonomous vehicle, the steps comprising:
12[a]	receiving a plurality of images displaying road scenes captured by one or more vehicles;
12[b]	receiving a plurality of user responses, each user response describing a state of mind of a road user displayed in one or more images;
12[c]	generating a training dataset comprising summary statistics of user responses describing the state of minds of road users displayed in the

	plurality of images;
12[d]	training, using the training dataset, a supervised learning based model configured to predict summary statistics describing a state of mind of a road user displayed in an input image;
12[e]	receiving, by an autonomous vehicle, a new image captured by a camera of the autonomous vehicle, the new image of a scene including a road user;
12[f]	predicting, by the autonomous vehicle, using the supervised learning based model, summary statistics describing a state of mind of the road user in the new image; and
12[g]	controlling the autonomous vehicle based on the prediction of the supervised learning based model.

Claim 13

13	The non-transitory computer readable storage medium of claim 12, wherein the road user in the input image is a pedestrian and the state of mind predicted by the supervised learning based model indicates whether the pedestrian is likely to perform an action, the action including one of: staying in place, or crossing a street.
-----------	--

Claim 14

14	The non-transitory computer readable storage medium of claim 12, wherein the road user in the input image is a cyclist and the state of mind predicted by the supervised learning based model indicates whether the cyclist is likely to cross a lane.
-----------	--

Claim 15

15	The non-transitory computer readable storage medium of claim 12, wherein the summary statistics comprises one or more of: a central tendency, a variance, a skew, a kurtosis, a scale, or a histogram.
-----------	--

Claim 16

16	The non-transitory computer readable storage medium of claim 12, wherein receiving a user response comprises:
-----------	---

	sending one or more images for display to a viewer via a user interface, the one or more images including a road user, wherein the user interface is configured to prompt the viewer to describe how the road user will act.
--	--

Claim 17

17	The non-transitory computer readable storage medium of claim 12, wherein receiving a user response comprises: generating a stimulus comprising a set of images; sending the generated stimulus for display via a user interface; and receiving, via the user interface, a user response describing the stimulus.
-----------	---

Claim 18

18	The non-transitory computer readable storage medium of claim 17, further comprising: modifying an image to generate a stimulus, the modifying comprising extracting a subset of the image contained in a bounding box around the road user displayed in the image.
-----------	--

Claim 19

19	The non-transitory computer readable storage medium of claim 18, wherein the supervised learning based model uses one or more features comprising: dimensions of the bounding box; location of the bounding box; shape of the bounding box; or a change in size or position of the bounding box across two images representing video frames.
-----------	---

Claim 20

20	The non-transitory computer readable storage medium of claim 12, wherein the user response is determined based on: an explicit response to a question asked via a user interface; and an implicit data representing one or more of: time taken to respond; whether user deleted keystrokes; or whether user moved a mouse anywhere other than a location
-----------	---

Petition for *Inter Partes* Review of U.S. Patent No. 11,126,889

	corresponding to a selected response.
--	---------------------------------------

III. INTRODUCTION

Tesla, Inc. (“Petitioner”) requests *inter partes* review (“IPR”) of claims 1-20 (the “Challenged Claims”) of U.S. Patent No. 11,126,889 (the “’889 Patent”) (EX1001). This Petition demonstrates that the Challenged Claims are unpatentable.

IV. PAYMENT OF FEES UNDER 37 C.F.R. § 42.103

The undersigned authorizes the Office to charge the fee set forth in 37 C.F.R. §42.15(a)(1) for this Petition to Deposit Account No. 02-0384. The undersigned further authorizes payment for any additional fees that may be due in connection with this Petition.

V. CERTIFICATION OF GROUNDS FOR STANDING

Petitioner certifies under 37 C.F.R. § 42.104 that the ’889 Patent is available for IPR and that Petitioner is not barred or estopped from requesting IPR of the Challenged Claims based on the grounds identified in this Petition.

VI. OVERVIEW OF CHALLENGE AND RELIEF REQUESTED

A. Prior Art Printed Publications

The ’889 Patent was filed on March 24, 2020 and claims priority via several continuation applications to a provisional patent application filed on July 5, 2017

(the “771 Provisional”).¹ EX1001, Cover Page. Petitioner does not believe that the ’889 Patent is entitled to the July 5, 2017 priority date (*see* IPR2025-01573). However, this issue is moot because Huval, Ross, Cox, and Munro are prior art to the ’889 Patent before July 5, 2017.

- U.S. Patent Application Publication No. 2018/0373980 (“Huval”) (EX1004); filed June 27, 2017; prior art under 35 U.S.C. § 102(a)(2).
- U.S. Patent No. 10,496,091 (“Ross”) (EX1005); filed on August 17, 2016; prior art under 35 U.S.C. § 102(a)(2).
- International Publication No. WO2014/210334 (“Cox”) (EX1006); published December 31, 2014 and filed June 26, 2014; prior art under 35 U.S.C. § 102(a)(1) and § 102(a)(2).
- U.S. Patent Application Publication No. 2016/0162456 (“Munro”) (EX1009); published June 9, 2016 and filed December 9, 2015; prior art under 35 U.S.C. § 102(a)(1) and § 102(a)(2).

None of these references were cited or otherwise discussed during prosecution.

¹ Because the application leading to the ’889 Patent was filed after March 16, 2013, patentability is governed by the America Invents Act (“AIA”) version of 35 U.S.C. §§102 and 103. *See* Pub L. No. 112-29, 125 Stat. 284 (2011).

B. Identification of Challenge and Statement of Precise Relief Requested

Petitioner requests cancellation of the Challenged Claims base on the specific grounds of this Petition, which are set forth below and are supported by the declaration of Dr. Janét (EX1003).

Ground	Challenged Claims	Basis	Reference(s)
1	1-20	§ 103	Huval and Ross
2	1-20	§ 103	Huval, Ross, and Cox
3	1-20	§ 103	Huval, Ross, and Munro
4	1-20	§ 103	Huval, Ross, Cox, and Munro

VII. SUMMARY OF THE '889 PATENT

A. Prosecution History

The Office issued a first non-final office action that included only a non-statutory double patenting rejection over U.S. Patent 10,614,344, the parent of the '889 Patent. EX1002, 100-106. In response, the applicant filed a terminal disclaimer and made minor amendments to the claims. *Id.*, 120, 127-137. The Office

subsequently issued a notice of allowance without identifying any particular claim limitations that rendered the claims allowable. *Id.*, 140-146.²

B. Technology Background

The '889 Patent generally relates to supervised machine learning models, which are computer programs that analyze input data and generate an output based on the data. EX1003, ¶57. Such models are “supervised” because they are trained using labeled training data that is known to represent a correct output. *Id.* Human annotators (also referred to as “labelers” or “reviewers” or “annotators”) can be used to label such training data. *Id.* After training, the model can be applied to new, non-training data, to provide an output for the new data that reflects the human annotators’ decision making when labeling training data. *Id.*

For example, to develop a vehicle detection supervised machine learning model, human annotators may label all instances of training data images that display a vehicle. EX1003, ¶58. The model is then trained using the labeled training data, and, after training, the model can detect vehicles in new, non-training data such as camera images. *Id.* The model’s detection of vehicles reflects how the human

² Prosecution of the related '344 Patent and '687 Patent also did not identify allowable claim limitations. EX1007, 91-98; EX1008, 102-103.

annotators would have labeled the new images because the model was trained based on the human annotators' decision making when labeling training data. *Id.*

C. The Alleged Invention

The alleged invention of the '889 Patent relates to (1) training a machine learning model using human labeled training data and (2) applying the trained model to new input data to generate an output that replicates how humans would have labeled the new input data. EX1001, Abstract, 2:4-23, 3:51-64, 6:44-7:44, 11:50-12:25, 12:12-23. But this alleged invention was well known, as demonstrated by the grounds herein. Section X; EX1003, ¶59. Indeed, the *inventor's own prior art* shows that the '889 Patent should not have issued. The Cox reference, which also lists Samuel Anthony as an inventor, discloses machine learning model predictions are "more consistent with the decisions of the human annotators." EX1006, ¶0032. Neither the patentee nor Mr. Anthony brought the Cox reference to the patent office's attention during prosecution of the '889 Patent. *See* EX1001; EX1002.

The substance of the '889 Patent further reveals that it is not new. The patent explains how model training data can be obtained and labeled by human annotators. EX1001, 4:30-55, 5:16-34, 5:46-6:4, Figs. 1, 2B; EX1003, ¶60. This concept was well known, as demonstrated by (at least) Cox. EX1006, ¶¶0008-0009, 0029-0032 (human annotators annotate training images); EX1003, ¶60. The patent then explains that such human labels can be used to generate "summary statistics" that "may

characterize the aggregate responses of multiple human observers” to a particular training image. EX1001, 6:8-18. But these complex-sounding terms refer to simple operations, such as averaging or otherwise combining the various annotator responses, or generating a tally of annotator responses—e.g., “how many” annotators labeled a training image in a certain way. *Id.* This was well known. EX1003, ¶60; EX1004, ¶0085 (averaging or combining human responses); EX1009, ¶0140 (tallying human annotator labels). And in another example, the patent explains that “summary statistics” relate to how long a human annotator took to label training data or to the position of the annotator’s eyes when labeling training data. EX1001, 6:11-15. Yet again, this was well known and at least the Cox reference demonstrates this. *See, e.g.*, EX1006, ¶¶0006-0008, 0018, 0063; EX1003, ¶60.

Thereafter, the patent notes that a machine learning model is trained using the responses and used to make a prediction based on “live” input data. EX1001, 7:12-46. But this was also well known. EX1003, ¶61. For example, the Huval reference explains that a machine learning model trained by human labeled training data is used to make predictions based on image data. EX1004, ¶0011. And the ’889 Patent’s application of these known concepts to autonomous driving does not save it. EX1001, 1:57-60, 3:52-53; EX1003, ¶61. Both the Huval and Ross references teach using predictions of a machine learning model to aid control of an autonomous

vehicle. EX1004, ¶¶0008-0009, 0016-0021, 0036-0039; EX1005, 1:30-2:53, 3:64-4:5, 4:33-5:11, 6:33-59, 7:57-8:31, 9:13-10:30, 15:1-16:52, Fig. 1; EX1003, ¶61.

Thus, the '889 Patent describes nothing more than well-known machine learning concepts. EX1003, ¶62. The grounds of this Petition demonstrate that the Challenged Claims are unpatentable, should not have issued, and should now be canceled. *Id.*

VIII. LEVEL OF ORDINARY SKILL IN THE ART

A person having ordinary skill in the art (“POSITA”) relevant to the '889 Patent as of July 5, 2017 would have had at least: (1) a bachelor’s degree in electrical engineering, computer engineering, computer science, or equivalent course work, with two to three years of work experience in computer vision, image/video processing, and/or machine learning; or (2) a master’s degree in electrical engineering, computer engineering, computer science, or equivalent course work, with a focus in computer vision, autonomous vehicles, and/or machine learning. EX1003, ¶63.

IX. CLAIM CONSTRUCTION

Claim terms in an IPR are construed according to their “ordinary and customary meaning” to those of skill in the art. *Phillips v. AWH Corp.*, 415 F.3d 91303 (Fed. Cir. 2005) (en banc); 37 C.F.R. §42.100(b). Because there is no need for

claim construction to determine that the cited prior art renders all Challenged Claims obvious, Petitioner does not present any constructions.³ EX1003, ¶64.

X. THE CHALLENGED CLAIMS ARE UNPATENTABLE

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

Huval alone teaches most limitations of the Challenged Claims and is further combined with Ross regarding the following limitations: (1) *receiving a plurality of user responses, each user response describing a state of mind of a road user displayed in one or more images*; (2) *training, using the training dataset, a supervised learning based model configured to predict summary statistics describing a state of mind of a road user displayed in an input image*; and (3) *predicting, by the autonomous vehicle, using the supervised learning based model, summary statistics describing a state of mind of the road user in the new image*. As discussed below, these limitations are explicitly taught by Ross, and it would have been obvious to combine Ross's relevant teachings with Huval. EX1003, ¶65.

³ Petitioners do not concede that the terms of the challenged claims are definite under 35 U.S.C. §112.

1. Summary of Prior Art

a. Huval–EX1004

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

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

Huval is analogous art to the ’889 Patent, from the same field of endeavor as the patent (e.g., predictive model techniques), and is reasonably pertinent to the particular problem that the patent was trying to solve (e.g., navigating an autonomous vehicle in an environment including moving objects). EX1004, ¶¶0008-

0015, 0021-0024, 0036-0039, 0043-0050; EX1001, 1:19-21, 1:57-30, 5:46-53, 6:8-49, Fig. 2A; EX1003, ¶68.

b. Ross–EX1005

Ross discloses techniques for “maneuvering” an autonomous vehicle. EX1005, 1:30-31, 3:39-42, 3:64-5:5. Ross’s system generates human influenced training data and uses the data to train an “intent model” for “predicting [the] intent of an object,” such as a vehicle, pedestrian, or bicyclist, in the environment of an autonomous vehicle. *Id.*, 1:30-2:53, 4:33-5:11, 9:13-10:10, 15:1-16:52. The human influenced training data is formed by “human operators” who manually generate a “list of predetermined actions” based on their “observations of the actions of other road users” in sensor data captured by autonomous vehicles. *Id.*, 9:41-10:10.

Ross teaches that an autonomous vehicle uses the trained intent model for navigation EX1005, 7:57-8:10, 9:13-10:10, 15:1-16:52. Cameras of the vehicle capture “raw” camera data of the environment around the vehicle displaying objects in the environment. *Id.* The trained model is then executed on this data to predict “a set of possible intents for a given road user by looking at all the possible actions that the road user could execute using [the trained] intent models.” *Id.*, 3:64-4:46. Specifically, the model generates a “set of possible intents or hypotheses identifying possible next actions (and predicted trajectories), a given point in time when the action is likely to occur, and associated likelihood values” with respect to each object.

Id., 7:57-8:10, 9:13-10:10, 15:1-16:52 (parenthetical in original). The set of likelihood values, with each value associated with a different prediction, “indicate[s] which of the predictions are more likely to occur (relative to one another).” *Id.*, 9:27-40. The autonomous vehicle of Ross is then controlled accordingly. *Id.*, 1:30-2:53, 3:64-4:5, 4:33-5:11, 6:33-59, 7:57-8:31, 9:13-10:30, 15:1-16:52, Fig. 1.

Ross is analogous art to the ’889 Patent, from the same field of endeavor as the patent (e.g., predictive model techniques), and is reasonably pertinent to the particular problem that the patent was trying to solve (e.g., navigating an autonomous vehicle in an environment including moving objects). EX1005, Abstract, 3:64-5:1; 9:27-47; EX1001, 1:19-21, 1:57-60; EX1003, ¶71.

2. Motivation to Combine Huval and Ross

A POSITA would have been motivated to combine Huval and Ross such that:

- Huval’s annotation labels provided by human annotators label observed road user actions in image data captured by vehicles, where such labeled actions correspond to road user intent;
- Huval’s machine learning model is (1) trained with such human annotation labels and (2) predicts a “set of possible intents or hypotheses identifying possible next actions (and predicted trajectories), a given point in time when the action is likely to occur, and associated likelihood values” for road users displayed in input image data; and

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

EX1004, ¶¶0008-0010, 0016-0024, 0029-0039, 0043-0050, 0096, 0100, Fig. 1; EX1005, 1:30-2:53, 4:33-5:11, 9:13-10:10, 15:1-16:52; EX1003, ¶72.

A POSITA would have been motivated to combine Huval and Ross for several reasons. EX1003, ¶73.

A POSITA would have been motivated to combine Huval and Ross because Ross simply teaches applying Huval’s system to the well-known behavior and intent context. EX1003, ¶74. The combination provides that Huval’s autonomous vehicle system is trained to not only identify objects in the environment of an autonomous vehicle, but also to predict the intent or state of mind of such objects and implement vehicle control accordingly. *Id.*

Specifically, Huval’s machine learning model is trained based on human annotator labels and implemented by an autonomous vehicle to automatically detect other vehicles, pedestrians, and cyclists in new images. EX1004, ¶¶0008-0010, 0016-0024, 0029-0039, 0043-0050, 0096, 0100, Fig. 1. Ross expands on Huval’s disclosure, teaching a predictive model for autonomous vehicles that predicts the intent of other vehicles, pedestrians, and cyclists. EX1005, Abstract. In Ross, human influenced training data is used to train an intent model, where correctly predicted actions performed by a road user are “marked” as being correct. *Id.*, 1:30-2:53, 4:33-5:11, 9:13-10:10, 15:1-16:52. The predicted actions include

various vehicle/pedestrian actions (e.g., following a lane, changing lanes, crossing a roadway at a crosswalk). *Id.*, 9:13-10:10, 15:1-16:22. And Ross explains that such actions correspond to particular intents; for example, a vehicle “moving towards another lane” corresponds to an intent of the vehicle (and its associated driver) to “change lanes when [the] desired lane is free.” *Id.* A “pedestrian waiting to cross the road” corresponds to an intent of the pedestrian “to cross the road when clear.” *Id.*; EX1003, ¶75.

Ross further teaches that Huval’s trained model is used to predict a “set of possible intents or hypotheses identifying possible next actions (and predicted trajectories), a given point in time when the action is likely to occur, and associated likelihood values” for road users of the autonomous vehicle’s environment displayed in input image data. EX1005, 1:30-2:13, 3:64-4:5, 4:33-5:11, 7:57-8:31, 9:13-10:10, 15:1-16:41 (parenthetical in original), Claims 1 and 7. And thereafter, the autonomous vehicle is controlled based on the predictions. *Id.*, 1:55-2:13, 15:1-16:41, 6:33-59, 9:13-10:30, 4:66-5:11, 15:1-16:41, Fig. 1; EX1003, ¶76.

As such, the above description demonstrates that Ross merely provides implementation details as to how Huval’s system is applied to the behavior and intent context. EX1003, ¶77. A POSITA would thus have been motivated to combine the references and have had a reasonable expectation of success. *Id.*

A POSITA would have also been motivated to combine Huval and Ross because each reference provides explicit teachings, suggestions, and motivations to make the combination. EX1003, ¶78. For instance, Huval explains that its machine learning model is implemented so that an autonomous vehicle can “perceive an environment around the autonomous vehicle[] and navigate the autonomous vehicle to a selected destination.” EX1004, ¶0015. At least this disclosure teaches, suggests, and motivates a POSITA to combine Huval’s teachings with the teachings of Ross. EX1003, ¶78. This is because Ross furthers Huval’s teachings by similarly acknowledging that an autonomous vehicle “detect[s] and identif[ies] objects in the vehicle’s environment,” but adds that its system advantageously predicts the intent of road users to determine how they might act in the environment. EX1005, 3:39-4:15, 4:33-5:11, 13:29-14:9, 15:1-16:41. A POSITA would have thus looked to Ross and combined its teachings with Huval to provide this benefit. EX1003, ¶78.

A POSITA would have additionally been motivated to combine Huval and Ross because Ross provides additional benefits to Huval’s system. EX1003, ¶79. As noted above, Huval uses its trained model to “perceive” the surroundings of an autonomous vehicle—including the perception of “dynamic” or moveable objects in the surroundings—and effectuate appropriate vehicle control. EX1004, ¶¶0010-0011, 0015, 0018-0019, 0032-0035, 0043-0050, 0098. Huval’s model is trained

using human annotations, including of behavior or actions in training images (albeit in the context of actions performed by the autonomous vehicle itself rather than by objects in the vehicle's environment). *Id.*, ¶¶0008-0012, 0021-0022, 0036-0050, 0069, 0071-0092, 0099-0120, Figs. 1, 2A, 2B, 3-5. Ross benefits Huval by teaching additional training and model predictions that allow for perceiving the surroundings of an autonomous vehicle. EX1005, 5:6-11, 9:13-10:10, 15:1-16:52; EX1003, ¶80. Ross teaches model training and implementation that predicts the intents of road users within the environment of an autonomous vehicle, which benefits Huval's system by making it safer and more robust. *Id.* As such, Ross provides Huval's system with the ability to determine what moving road users in the autonomous vehicle's environment might do, which would have improved the safety of Huval's vehicle control. *Id.* Huval's vehicle would have better anticipated how road users might act and implemented control to avoid possible collisions. *Id.* At bottom, Huval's autonomous vehicle would have better interacted with objects in its environment by "[r]easoning about the intents of objects, such as other road users" and thus functioned "in a more 'human-like' or 'polite' way." *Id.*; EX1005, 5:6-11, 15:1-16:52. A POSITA would have thus recognized the benefits of combining Huval and Ross, which motivates the combination. EX1003, ¶80.

All the reasons explained above establish that a POSITA would have been motivated to combine Huval and Ross as described. EX1003, ¶81. This

combination would have been the predictable combination of well-known prior art elements (e.g., Huval's human annotation and prediction model; Ross's training and intent model teachings) according to known methods to yield predictable results (Huval's annotation labels provided by human annotators label observed road user actions in image data captured by vehicles, where such labeled actions correspond to road user intent; Huval's machine learning model is (1) trained with such human annotation labels and (2) predicts a "set of possible intents or hypotheses identifying possible next actions (and predicted trajectories), a given point in time when the action is likely to occur, and associated likelihood values" for road users displayed in input image data; Huval's autonomous vehicle is controlled based on the predictions). *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398, 415-421 (2007); EX1005, 15:1-21; EX1003, ¶81. The combination would have further used known techniques (known machine learning model training and intent prediction techniques as taught by Ross) to improve similar devices (Huval's prediction model) in the same way (using Ross's aforementioned techniques). *Id.* Lastly, in addition to the above, because Ross already teaches training a predictive model to generate predictions about vehicles, pedestrians, and bicyclists in an environment of an autonomous vehicle, combining Ross with Huval's teachings about training a model that generates predictions regarding the same objects would have required minimal changes. *Id.* Making this combination is thus well within

the skill of a POSITA, and a POSITA would have had a reasonable expectation of success. *Id.*

3. Claim 2⁴

Huval and Ross render obvious claim 2. EX1003, ¶82.

a. 2[Pre]

To the extent it is limiting, Huval and Ross teach the preamble. EX1003, ¶83. The Huval-Ross combination teaches computer-implemented techniques for controlling an autonomous vehicle based on the predicted intents of road users (e.g., vehicles, pedestrians, and bicyclists) displayed in camera images of the environment of the vehicle, shown further below (*[a] computer-implemented method for controlling an autonomous vehicle based on a predicted state of mind of road users in a scene captured by a camera of the autonomous vehicle*). EX1004, ¶¶0008-0009, 0016; EX1005, 3:64-4:15, 7:57-8:53, 9:66-10:30; *See* Sections X.A.2, X.A.3.b (element 2[a]), X.A.3.e-g (elements 2[d], 2[e], and 2[f]); EX1003, ¶83.

b. 2[a]

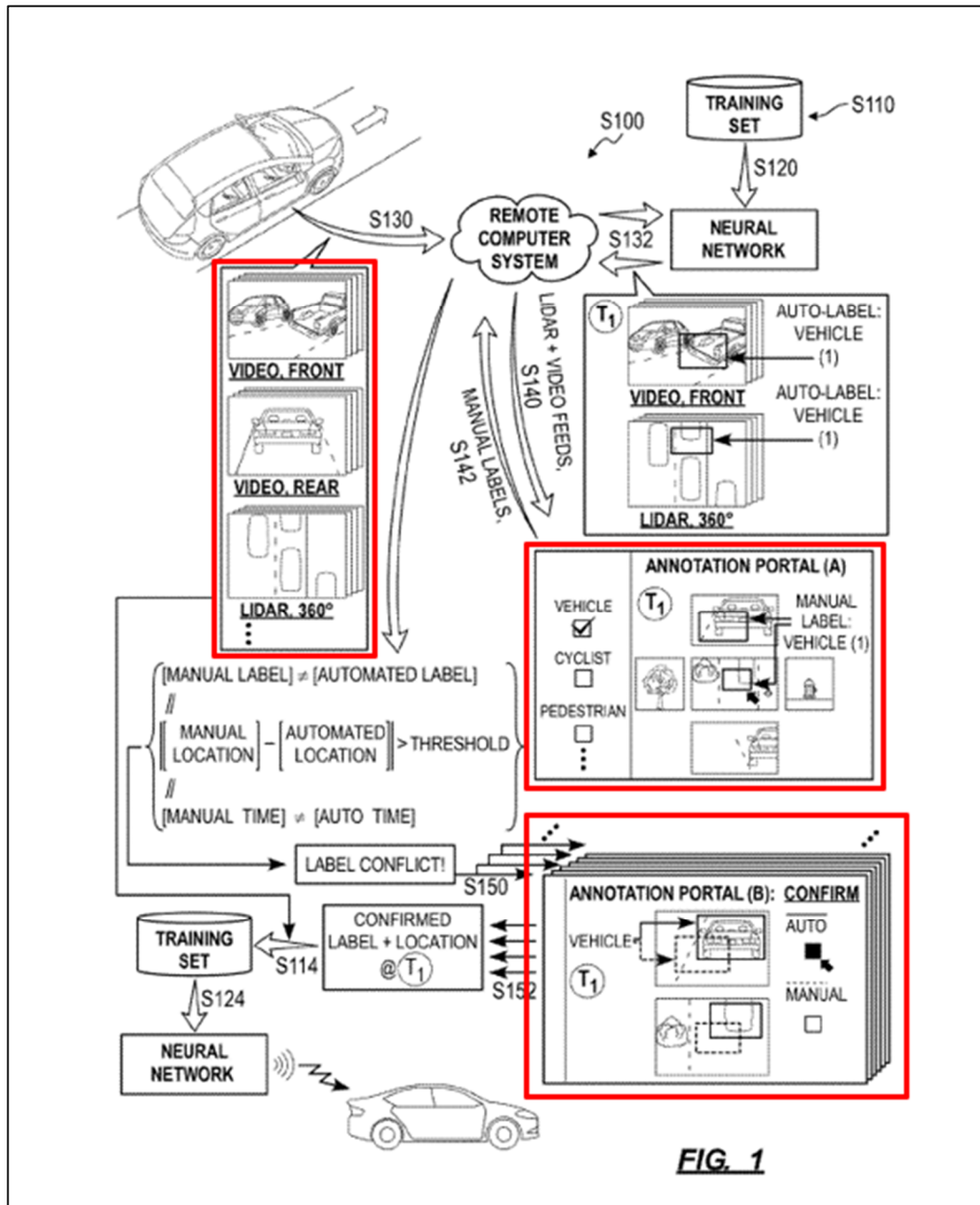
Huval and Ross teach element 2[a]. EX1003, ¶84. Huval describes that its system receives images from a training set of images, and also receives images

⁴ Petitioner's analysis begins with independent claim 2 as claims 3-11 depend from this claim.

from cameras of an autonomous vehicle that are recorded “during operation” (*receiving a plurality of images displaying road scenes captured by one or more vehicles*). EX1004, ¶¶0008-0009, 0017, 0021-0022, 0025, 0029-0030, 0037-0038; EX1003, ¶84. The autonomous vehicle includes cameras that “collect information about the autonomous vehicle’s environment” and specifically “output a video feed containing a sequence of digital photographic images (or ‘frames’).” EX1004, ¶¶0016-0017. The images include objects in the autonomous vehicle’s surroundings such as “vehicles, pedestrians, and cyclists,” which means the image displays a *road scene*. *Id.*, ¶¶0008-0009, 0011, 0016-0018, 0025, 0029-0030, 0071-0092, Figs. 1, 3, 4; EX1003, ¶84. Thus, by capturing images of vehicles, pedestrians, and cyclists surrounding the autonomous vehicle and providing these images to Huval’s system, Huval teaches *receiving a plurality of images displaying road scenes captured by one or more vehicles*. *Id.*

Such images are transmitted by Huval’s remote computer system to a human annotator for manual labeling, which further teaches *receiving a plurality of images displaying road scenes captured by one or more vehicles*. EX1004, ¶0025; EX1003, ¶85. For example, a vehicle can upload images to the remote computer system and the computer system can “store these optical data in a remote database and queue these optical data for transmission to human annotators for manual labeling.” EX1004, ¶0025. The human annotators receive the images in an “annotator portal”

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



EX1004, Fig. 1 (annotated).

Huval also describes *receiving a plurality of images displaying road scenes captured by one or more vehicles* in an additional way. With respect to receiving images from a training set of images, Huval describes a label conflict resolution process to refine training data where human annotators resolve conflicts between automatically generated labels and manual labels. EX1004, ¶¶0008-0009, 0071-0092. For example, Huval's system can determine whether conflicts exist between human (manual) and automated labels of a training image. *Id.*, ¶¶0074-0085. If a conflict exists, the image having conflicting labels is transmitted to multiple human annotators, rendered on their respective annotation portals, to confirm the correct label. *Id.* As such, Huval describes that its human annotators receive images in this manner as well (*receiving a plurality of images displaying road scenes captured by one or more vehicles*). *Id.*; EX1003, ¶86.

c. 2[b]

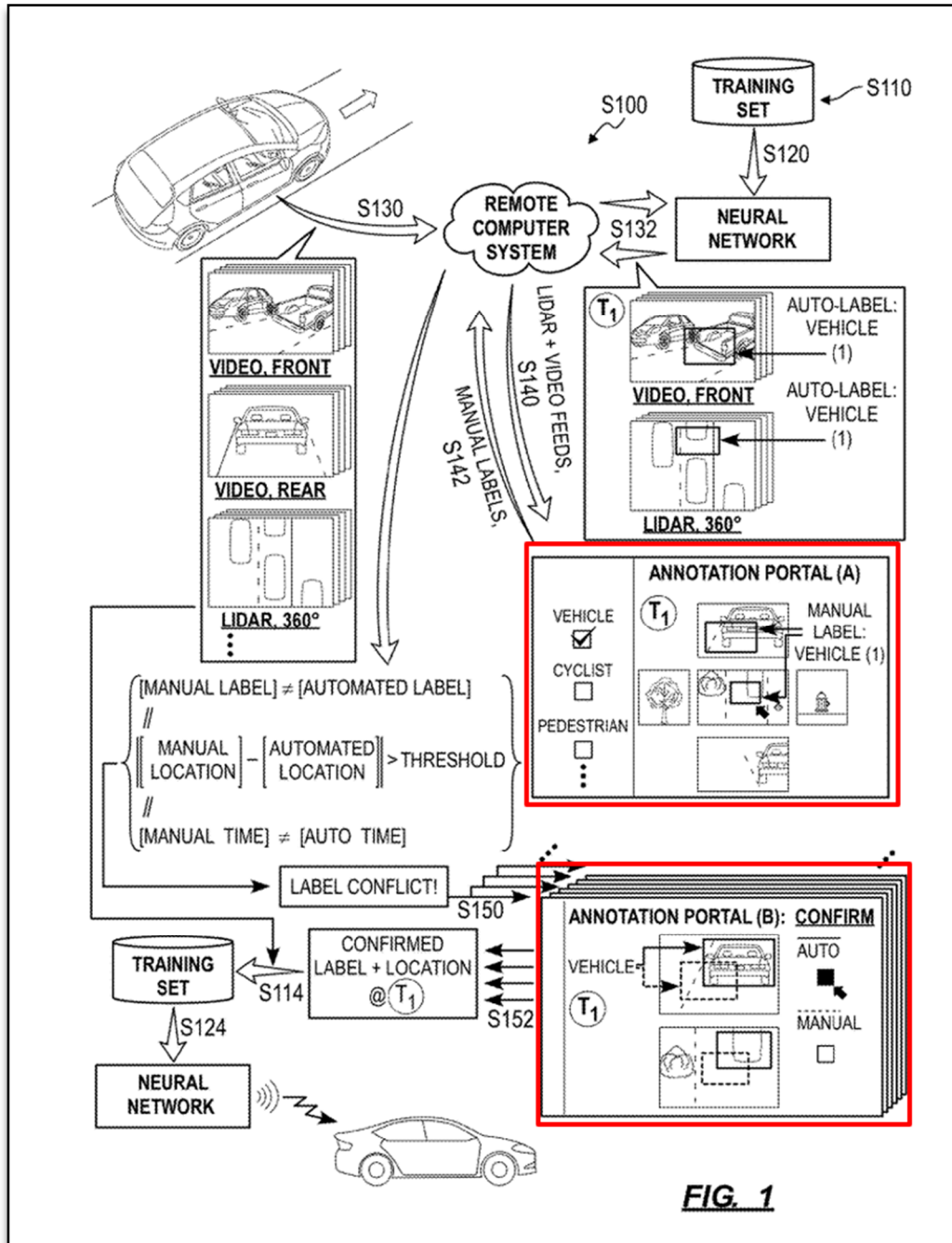
Huval and Ross teach element 2[b]. EX1003, ¶87. Huval describes that images captured by autonomous vehicles are sent to human annotators, who manually label the images and send their responsive labels to Huval's computer system (*receiving a plurality of user responses*). EX1004, ¶¶0008-0012, 0021-0022, 0036-0050, 0069, 0085-0086, 0090, Figs. 1, 2A, 2B, 3; *see also* Section X.A.3.b (element 2[a]). The annotation portal enables the human annotators to label the

images by, for example, “placement of labeled boundary boxes” or selecting a label from “a set of object types related to perception, such as road vehicles (e.g., passenger vehicles, commercial trucks, motorcycles), pedestrian, cyclist . . . and other dynamic [] objects.” EX1004, ¶¶0043-0044, Figs. 1, 2A, 2B, 3, 5.

The manual labels received by Huval’s remote computer system (e.g., blocks 114, 142, 152 of Figures 1, 3, and 4) describe road users present in the images, such as a road vehicle, pedestrian, and cyclist (*each user response describing...a road user displayed in one or more images*). EX1004, ¶¶0008-0012, 0036-0050, Figs. 1, 3, and 4; EX1003, ¶88.

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

a plurality of user responses, each user response describing...a road user displayed in one or more images),*Id.*; EX1003, ¶89.



EX1004, Fig. 1 (annotated)

The combination of Ross and Huval further teaches that each of the annotation responses describes a *state of mind of a road user displayed in one or more images*. EX1003, ¶90. In the Huval-Ross combination, the human annotation process would have included the labeling of observed road user (e.g., vehicle, pedestrian, bicyclist) actions in image data captured by vehicles, where such labeled actions correspond to the intents of road users and thus describe the state of minds of the road users (*each user response describing a state of mind of a road user displayed in one or more images*). EX1004, ¶¶0008-0012, 0021-0022, 0036-0050, 0069, 0071-0092, Figs. 1, 2A, 2B, 3, 5; EX1005, 1:55-2:13, 4:33-46, 9:13-10:10, 15:1-16:41; Section X.A.2; EX1003, ¶90. For example, the actions include, for a vehicle in the image data, following a current lane, changing lanes, or turning at an intersection. EX1005, 9:13-10:10, 15:1-16:41; EX1003, ¶90. A vehicle “moving towards another lane” corresponds to an intent of the vehicle (and its associated driver) to “change lanes when [the] desired lane is free,” indicating a state of mind of the vehicle (and driver) to change lanes when the desired lane is free. *Id.* Actions also include, for a pedestrian in image data, crossing a roadway, crossing a roadway in a crosswalk, and jaywalking, and such actions describe an intent of the vehicle or pedestrian. *Id.* For example, a “pedestrian waiting to cross the road” corresponds to an intent of the pedestrian “to cross the road when clear,”

indicating a state of mind of the pedestrian that they wish to cross the road when no traffic is present. *Id.*; EX1003, ¶90.

In view of the above, Huval and Ross therefore teach element 2[b]. EX1003, ¶90.

d. 2[c]

Huval and Ross teach element 2[c]. EX1003. ¶91. Huval describes that it obtains multiple human annotator manual labels for road vehicles, pedestrians, and cyclists displayed in sample images. EX1004, Abstract, ¶¶0008-0018, 0021-0025, 0029-0030, 0036-0050, 0085-0092, 0095-0098, 0118, Figs. 1, 3, 4; Section X.A.3.c (element 2[b]); EX1003, ¶91. For each image, Huval “combine[s]” the labels by determining an “average” to generate a final label for each image, teaching *summary statistics of user responses describing...road users displayed in the plurality of images. Id.* Because the average label is generated for each image, a collection of average values summarizing the labels of a collection of training images is generated by Huval, which teaches *summary statistics* as claimed. EX1003, ¶91. Because these average values are a collection of numerical measurements summarizing the labels, a POSITA would have recognized that these values form statistics. *Id.*

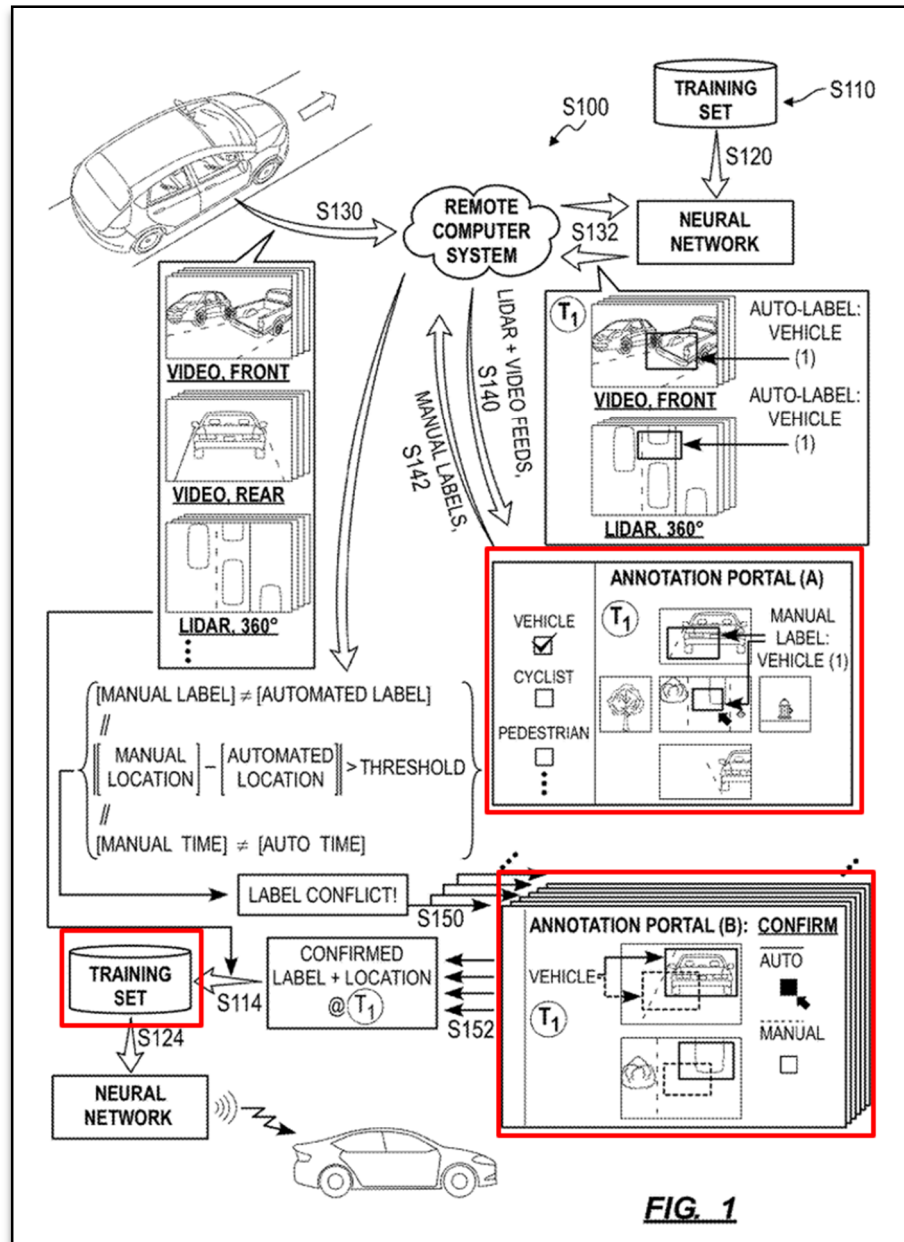
Additionally, Huval’s disclosure aligns with the ’889 patent’s description of *summary statistics*. EX1001, 10:60-67, 6:8-24; EX1003, ¶92. The patent explains

that “summary statistics” associated with user responses include “measurements of the central tendency of the distribution of scores like the mean, median, or mode” and are associated with “a content of a response.” *Id.* The average value generated for each training image, as disclosed by Huval, provides such a central tendency measurement of mean, and further reflect the content of human labels. *Id.*; EX1004, ¶0085.

The training images and associated combined labels (*summary statistics*) are appended to Huval’s “training set,” which means they are part of a *training dataset* (*generating a training dataset comprising summary statistics of user responses describing...road users displayed in the plurality of images*). EX1004, ¶¶0085-0090; EX1003, ¶93. As taught by Huval, the “manual labels” that form the combined labels provided by human annotators are used to “train and refine the neural network over time” and are thus part of the *training dataset* used to train Huval’s model. EX1004, Abstract, ¶¶0008-0018, 0021-0025, 0029-0030, 0036-0050, 0085, 0090, 0095-0098, 0118, Figs. 1, 3, and 4; EX1003, ¶93.

Huval further teaches *summary statistics of [the] user responses* in an additional way. EX1003, ¶94. The training set of Huval comprises *summary statistics of [the] user responses* at least because it includes the manual labels that are “packaged” into the training set and reflect the summarized statistics of the content of all accurate manual human labels that have been obtained and are used

for training. EX1004, Abstract, ¶¶0008-0018, 0021-0025, 0029-0030, 0036-0050, 0085, 0090, 0095-0098, 0118, Figs. 1, 3, and 4; *see id.*, ¶¶0022-0023. For example, Huval’s remote computer system “aggregates labeled optical data, such as stored in a remote database, into a training set on which a localization, perception, motion planning, and/or other neural network can then be trained.” *Id.*, ¶¶0021-0024; *see also id.*, ¶¶0069-0070. Because these aggregated labels are a collection of measurements summarizing how data has been labeled, a POSITA would have recognized that these aggregated labels form statistics. *Id.* EX1003, ¶94. This disclosure further teaches *summary statistics of [the] user responses*. EX1003, ¶94.



EX1004, Fig 1 (annotated)

In combination with Ross, the Huval's response annotation labels would also correspond to the intents and actions of road users in training images and thus describe the state of minds of such road users (*user responses describing the state*

of minds of road users displayed in the plurality of images). EX1003, ¶95; *supra* Section X.A.2 (motivation to combine) and Section X.A.3.c (element 2[b]).

e. 2[d]

Huval and Ross teach element 2[d]. EX1003, ¶96. Huval describes that its training set is used to train its machine learning model (which is a neural network), where the model is supervised and corrected by human label information (*training, using the training dataset, a supervised learning based model*). EX1004, Abstract, ¶¶0008-0020, 0021-0025, 0029-0030, 0036-0050, 0079-0090, 0095-0098, 0118, Figs. 1, 3, and 4; *see* Section X.A.3.d (element 2[c]); EX1003, ¶96.

Huval's model is a *supervised learning based model* because it is “train[ed] and refine[d]” by human labeled training images, including by comparing model predictions and labels provided manually by human annotators, and “retraining” the model to “resolve conflicts” between the model predictions and human labels. EX1004, ¶¶0008-0012, 0085-0096; EX1003, ¶96. Moreover, Huval describes that its model is “train[ed]” using the training set, and also “developed” and “refined” using the training set, so that it detects objects and navigational actions in the scene around its autonomous vehicle (*training, using the training dataset, a supervised learning based model*). EX1004, ¶¶0008-0015, 0021-0024, 0079-0098; EX1003, ¶96.

Additionally, as discussed in Section X.A.2, the Huval-Ross combination teaches that Huval's model is configured to predict a "set of possible intents or hypotheses identifying possible next actions (and predicted trajectories), a given point in time when the action is likely to occur, and associated likelihood values" for road users (e.g., pedestrians, vehicles, bicyclists) displayed in input image data (*supervised learning based model configured to predict summary statistics describing a state of mind of a road user displayed in an input image*). EX1004, Abstract, ¶¶0008-0020, 0021-0025, 0029-0030, 0036-0050, 0085-0090, 0095-0098, 0118, Figs. 1, 3, and 4; EX1005, 1:30-2:53, 3:64-4:5, 4:33-5:11, 7:57-8:31, 9:13-10:10, 15:1-16:52, Claims 1 and 7 (parenthetical in original); EX1003, ¶97; Section X.A.2. This predicted "set of possible intents..." teaches *summary statistics describing a state of mind of a road user displayed in an input image* because it is a summary of the various predicted intents and likelihood values for a particular object in input image data. *Id.*; Section X.A.2, Sections X.A.3.c-d (elements 2[b], 2[c]); *infra* Section X.A.3.g (element 2[f]). The likelihood values summarize how likely each of the predicted intents and trajectories of the object is to occur, where the greatest likelihood value corresponds to the prediction most likely to occur. *Id.*; EX1005, 9:27-40, 15:50-16:52; EX1003, ¶97. Because these predicted intents/trajectories and associated likelihood values are a collection of measurements summarizing the various potential actions of an object and how

likely each is to occur, a POSITA would have recognized that these items form statistics. *Id.*; EX1003, ¶98. Indeed, the predicted intents and associated likelihood values reflect statistics summarizing how an object within the environment of Ross’s autonomous vehicle may behave. *Id.* Additionally these predicted intents describe a *state of mind of a road user* because they describe what the road user in the environment of Huval’s autonomous vehicle might want to do. *Id.*; EX1003, ¶98.

Such predictions are generated both during a training phase of Huval’s system and an operational phase. EX1004, Abstract, ¶¶0008-0020, 0021-0025, 0029-0030, 0036-0050, 0078, 0085-0092, 0095-0098, 0118, Figs. 1, 3, and 4; EX1005, 1:30-2:53, 3:64-4:5, 4:33-5:11, 7:57-8:31, 9:13-10:10, 15:1-16:52, Claims 1 and 7 (parenthetical in original); *supra* Sections X.A.2, X.A.3.c, X.A.3.d (elements 2[b], 2[c]), X.A.3.g *infra*. For example, by correcting and removing inaccurate labels, Huval teaches machine learning training techniques that minimize inaccuracies between the human labeled images and the automatically labeled images, “thereby ‘healing’ the neural network.” EX1004, ¶¶0012, 0073-0094. Moreover, Huval demonstrates that its model is trained because it explains that training involving more a high volume of correctly labeled training data results in a more effective and accurate model. *Id.*, ¶0011 (emphasis added); EX1003, ¶99; *see also* Sections X.A.3.c, X.A.3.d (elements 2[b], 2[c]), X.A.2.

f. 2[e]

Huval and Ross teach element 2[e]. EX1003, ¶100. Huval describes that during an operational phase in which the trained model of Huval has been deployed to an autonomous vehicle, cameras of the autonomous vehicle capture image data of the autonomous vehicle's surroundings that include other vehicles, pedestrians, and cyclists (*receiving, by an autonomous vehicle, a new image captured by a camera of the autonomous vehicle, the new image of a scene including a road user*). EX1004, ¶¶0008-0009, 0011, 0016-0019, 0025, 0029-0030, 0044, 0095-0098, Figs. 1, 3, 4; *see also* Section X.A.3.b (element 2[a]); EX1003, ¶100.

g. 2[f]

Huval and Ross teach element 2[f]. EX1003, ¶101. The Huval-Ross combination teaches that Huval's deployed model of its autonomous vehicle predicts a "set of possible intents or hypotheses identifying possible next actions (and predicted trajectories), a given point in time when the action is likely to occur, and associated likelihood values" for road users (e.g., pedestrians, vehicles, bicyclists) displayed in new input image data as the vehicle operates in an environment (*predicting, by the autonomous vehicle, using the supervised learning based model, summary statistics describing a state of mind of the road user in the new image*). EX1004, Abstract, ¶¶0008-0020, 0021-0025, 0029-0030, 0036-0050, 0085-0090, 0095-0098, 0118, Figs. 1, 3, and 4; *see also* Sections X.A.2, X.A.3.a-

X.A.3.e (elements 2[pre]-2[d]); EX1005, 1:30-2:53, 3:64-4:5, 4:33-5:11, 7:57-8:31, 9:13-10:10, 15:1-16:52, Claims 1 and 7; EX1003, ¶101. This predicted “set of possible intents...” teaches *summary statistics describing a state of mind of the road user in the new image* because it is a summary of the various predicted intents and likelihood values for a particular object in the input image data. *Id.* The likelihood values summarize how likely each of the predicted intents and trajectories of the object is to occur, where the greatest likelihood value corresponds to the prediction most likely to occur. *Id.*; EX1005, 9:27-40, 15:50-16:52, Fig. 10. Because these predicted intents/trajectories and associated likelihood values are a collection of measurements summarizing the various potential actions of an object and how likely each is to occur, a POSITA would have recognized that these items form statistics. *Id.*; EX1003, ¶101. The predicted intents and associated likelihood values thus reflect statistics summarizing how an object within the environment of Ross’s autonomous vehicle may behave. *Id.*; EX1003, ¶102. Additionally these predicted intents describe a *state of mind of the road user* because they describe the predictions of what the road user in the environment of Huval’s autonomous vehicle might want to do. *Id.*

h. 2[g]

Huval and Ross teach element 2[g]. EX1003, ¶103. The Huval-Ross combination describes that Huval’s autonomous vehicle is controlled based on the

predicted intents of vehicles, pedestrians, and cyclists in image data (*controlling the autonomous vehicle based on the prediction of the supervised learning based model*). EX1004, ¶¶0008-0024, 0036-0050, 0085-0086; EX1005, 1:55-2:13, 15:1-16:41, 6:33-59, 9:13-10:30, 4:66-5:11, 15:1-16:41, Fig. 1; EX1003, ¶103; Section X.A.2. Specifically, as taught by Ross, the highest likelihood intent or intents for a road user are determined and controls of the autonomous vehicle is based on these predictions. EX1005, 4:66-5:11, 10:11-30, 15:1-16:41; EX1003, ¶103; Section X.A.2.

4. Claim 3

Huval and Ross render obvious claim 3. EX1003, ¶104. Huval's image data captured by its autonomous vehicle cameras display pedestrians, other vehicles, and cyclists, and the images are a "sequence of digital photographic images (or 'frames') of a video feed" (*wherein the road user in the input image is a pedestrian*). EX1004, ¶¶0008-0009, 0011, 0016-0018 (parenthetical in original), 0025, 0029-0030, 0044-0050, Figs. 1, 3, 4; *see also* Section X.A.3.b, X.A.3.f (elements 2[a], 2[e]); EX1003, ¶104.

The Huval-Ross combination's predicted "set of possible intents..." for a pedestrian in the image data reflects the state of mind of the pedestrian and "includes a corresponding next action to be initiated by the road user," and each predicted intent has an associated likelihood of occurring; this demonstrates that

the predicted intents indicate how likely the pedestrian is to perform a particular action (*and the state of mind predicted by the supervised learning based model indicates whether the pedestrian is likely to perform an action*). EX1005, 1:55-2:13, 4:33-46, 9:13-40, 15:1-21; Section X.A.2, Section X.A.3.g (element 2[f]); EX1003, ¶105. As taught by Ross, the predicted intent can be that the pedestrian “would like to cross the roadway” (*the action including one of: staying in place, or crossing a street*). EX1005, 16:23-32; *see also id.*, 4:47-65, 9:41-10:10; EX1003, ¶105.

5. Claim 4

Huval and Ross render obvious claim 4. Huval teaches *wherein the road user in the input image is a cyclist* for the same reasons discussed in claim 3. Section X.A.4; EX1003, ¶106.

In addition, the Huval-Ross combination’s predicted “set of possible intents...” for a cyclist in the image data reflects the state of mind of the cyclist and the associated likelihoods that a cyclist is likely to perform an action (*and the state of mind predicted by the supervised learning based model indicates whether the cyclist is likely to perform an action*). Section X.A.2, Section X.A.3.g (element 2[f]); EX1005, 3:39-47, 5:1-11, 9:13-10:10, 15:1-16:52, Claims 1, 2, 3, and 7, Fig. 8; EX1003, ¶107. “[E]ach possible intent of the set includes a corresponding next action to be initiated by the road user.” EX1005, 1:55-2:13, 4:33-46, 9:13-40, 15:1-21. As taught by Ross, the predicted intent can be that a bicyclist would like to

“change lanes when [a] desired lane is free of traffic” or that a bicyclist would like to “move through [an] intersection when clear,” and related actions to these intents include the bicyclist is “moving towards another lane” or “waiting at an intersection” (*indicates whether the cyclist is likely to cross a lane*) EX1005, 15:64-16:22; EX1003, ¶107; *see also* EX1005, 15:64-16:22. This disclosure thus provides that the predictions indicate whether the bicyclist is likely to change or cross lanes. *Id.*

6. Claim 5

Huval and Ross render obvious claim 5. EX1003, ¶108. Huval describes that its training data includes human manual labels that are “packaged” into the training set and reflect the summarized statistics of the content of all accurate human manual labels that have been obtained and are used for training. EX1004, Abstract, ¶¶0008-0018, 0021-0025, 0029-0030, 0036-0050, 0085, 0090, 0095-0098, 0118, Figs. 1, 3, and 4; *see id.*, ¶¶0022-0023; *see also* Section X.A.3.d (element 2[c]). Moreover, human annotator label responses for each training image can be combined by averaging them to generate a final annotation label for each image. EX1004, ¶0085; Sections X.A.3.c-X.A.3.e (elements 2[b]-2[d]). By averaging the labels of multiple human annotators for each image, Huval generates a collection of average values summarizing the labels of a collection of training images (*summary statistics*). *Id.* And this collection of values thus includes a central tendency (e.g., an average) of

the human label responses for each image, which teaches *the summary statistics comprises one or more of: a central tendency . . .* EX1003, ¶108.

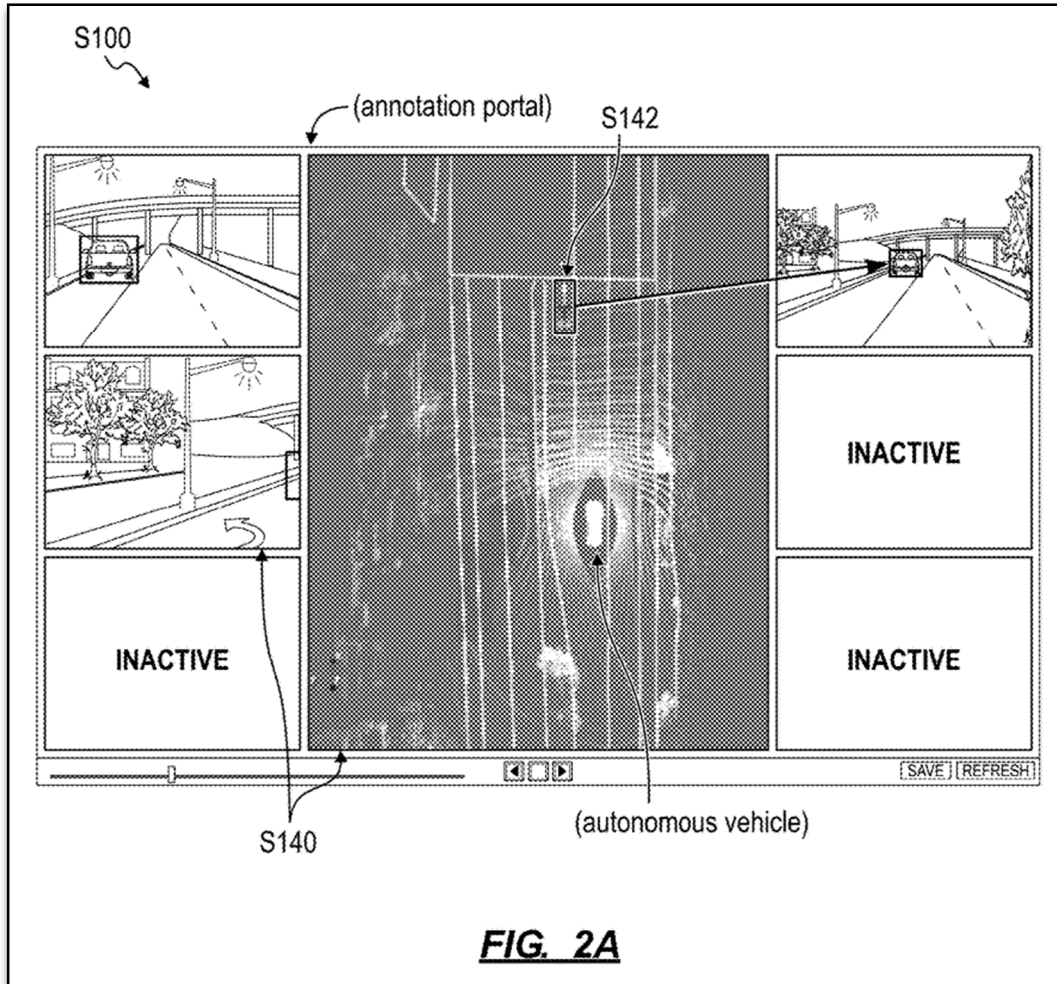
7. Claim 6

Huval and Ross render obvious claim 6. EX1003, ¶109. Huval teaches that human annotators are sent images in an “annotation portal” (*sending one or more images for display to a viewer via a user interface*). EX1004, ¶¶0008-0009, 0037, 0071-0092; *see also* Section X.A.3.b (element 2[a]). The Huval system “render[s] a prompt” to the human annotator using the annotation portal “to select one of the manual and automated labels to confirm a type of an object represented” in the image such as a vehicle, pedestrian, or cyclist (*the one or more images including a road user, wherein the user interface is configured to prompt the viewer*). *Id.*, ¶¶0043-0044, 0052-0055, 0080. The annotator can also manually enter an alternate label for another object type. *Id.*; EX1003, ¶109.

Ross teaches the labeling of road user actions in image data, where such labeled actions correspond to the intents of road users and thus describe the state of minds of the road users. Section X.A.2. Thus, as combined with Huval, Ross teaches that Huval’s human annotators would have been prompted to describe road user actions, including how they are appearing to act in the image (*to describe how the road user will act*). *Id.*; EX1004, ¶¶0008-0009, 0052-0055, 0071-0092; EX1005, 9:41-10:10, 15:50-16:22; EX1003, ¶110.

8. Claim 7

Huval and Ross render obvious claim 7. EX1003, ¶111. Huval describes that, as part of its process for receiving human labels of images from human annotators, a set of video feeds recorded by “the forward-facing, rear-facing, left-facing, and right facing cameras” of a vehicle are served to the annotators’ “local computer system for manual labeling,” where the video images to be labeled are displayed on an “annotation portal” of the computer system and include a road user such as a pedestrian, vehicle, or cyclist (*wherein receiving a user response comprises: generating a stimulus comprising a set of images; sending the generated stimulus for display via a user interface*). EX1004, ¶¶0008-0009, 0011, 0016-0018, 0025, 0029-0030, 0036-0042, 0050, Figs. 1, 2A, 2B, 3, 4; EX1003, ¶111. Moreover, “one or more video feeds exclusively” can be displayed by the portal. EX1004, ¶¶0041-0042, 0050.



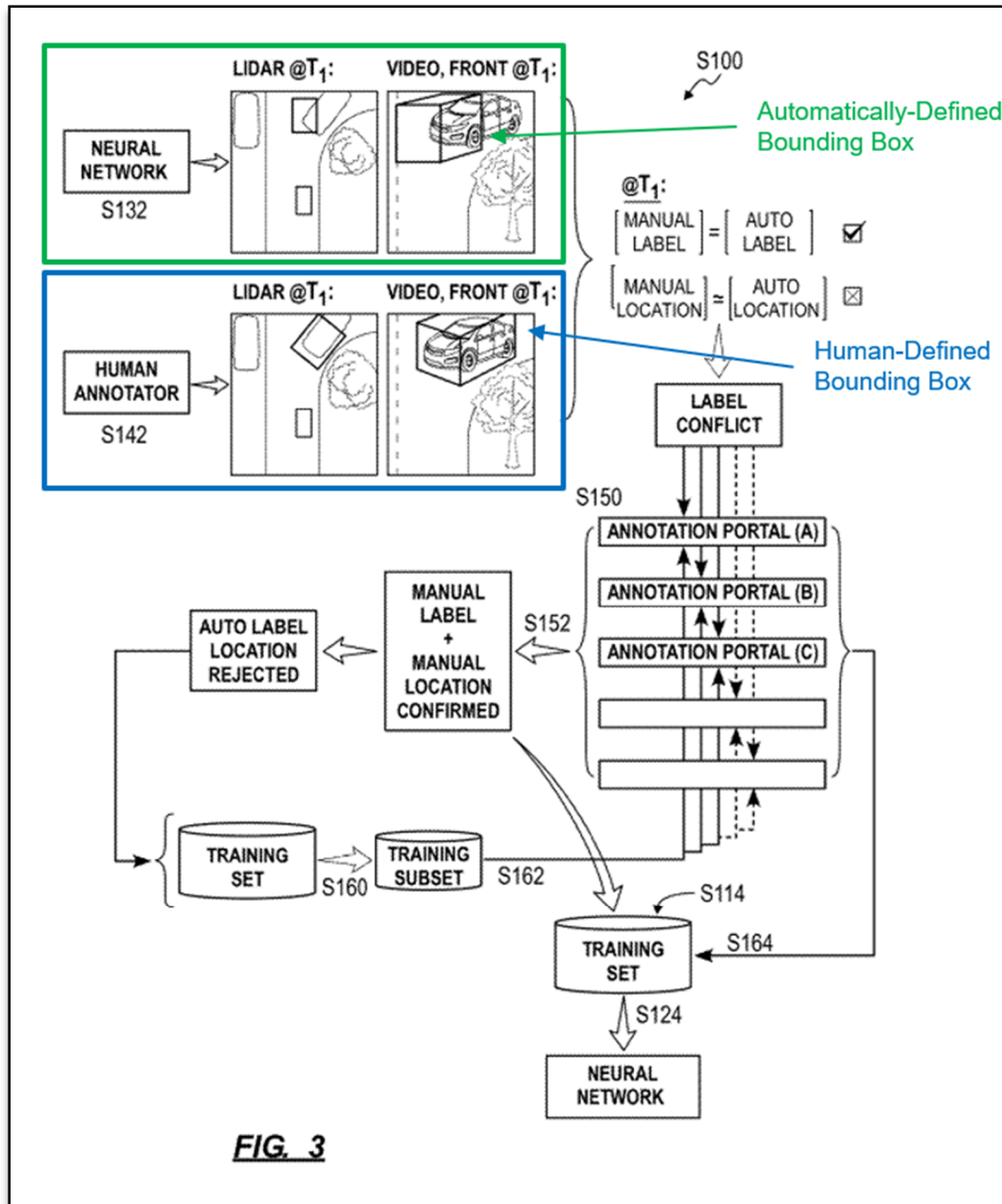
EX1004, Fig. 2A

The human annotators then add, via their respective annotation portals, manual labels to the images in response by, for example, making a selection from “a dropdown menu” and drawing a bounding box “linked to a cursor” around the object (*and receiving, via the user interface, a user response describing the stimulus*). *Id.*; see also *id.*, ¶¶0008-0012, 0037-0050; EX1003, ¶112. Moreover, as taught by Ross, the responses received from Huval’s annotators would have described the training image samples and the observations of what actions a road

user in the sample is doing. *See* Sections X.A.2, X.A.3.c (element 2[b]); EX1003, ¶112.

9. Claim 8

Huval and Ross render obvious claim 8. EX1003, ¶113. As discussed in Sections X.A.3.b (element 2[a]) and X.A.8 (claim 7), Huval describes a label conflict resolution process where human annotators compare conflicting automated and manual labels to confirm the correct label and refine the training dataset. EX1004, ¶¶0008-0009, 0071-0092. As part of this process, “the annotation portal can present” to a human annotator a modified version of a captured image that includes “a pre-generated automated label”, such as a “bounding box,” for an “object detected at an automatically-defined location by the neural network.” EX1004, ¶¶0031-0035, 0043-0050, 0052-0054, 0071-0092, Claim 3, Figs. 1-3. For example, as shown by the annotated green portion in Figure 3 below, a road user vehicle in the image is outlined by an automated label bounding box, which modifies the image (*modifying an image to generate a stimulus*) that the human annotator reviews when comparing label conflicts. *Id.*; EX1003, ¶113. The bounding box extracts a subset of the image around a road user vehicle in the image (*the modifying comprising extracting a subset of the image contained in a bounding box around the road user displayed in the image*). *Id.*



EX1004, Fig. 3 (annotated)

10. Claim 9

Huval and Ross render obvious claim 9. EX1003, ¶114. Huval describes that a training image reviewed for label conflicts includes an automated label bounding box around a road user. EX1004, ¶¶0052-0054, Fig. 3; *see* Section X.A.9 (claim 8).

In addition to confirming the label, the human annotator may “adjust” or move the bounding box on the image. EX1004, ¶0052. Huval’s system can then “recalculate the confidence score for the pre-generated automated label based on feedback received from the human annotator.” *Id.* Huval teaches that its model uses the confidence score associated with bounding box label, which is calculated based on whether the human annotator confirms the location of the automatically-defined location, to “selectively distribute automated labels to human annotators” as part of the training process. *Id.* ¶¶0034, 0052-0054, Fig. 3; EX1003, ¶114. Huval’s model therefore uses the bounding box, and thus its location on an image, shape, and dimensions, to make model training decisions (*wherein the supervised learning based model uses one or more features comprising: dimensions of the bounding box; location of the bounding box; shape of the bounding box; or a change in size or position of the bounding box across two images representing video frames*). *Id.*

11. Claim 10

Huval and Ross render obvious claim 10. EX1003, ¶115. As addressed above for claims 6 and 7, Huval discloses that the process for collecting human annotation responses may include sending prompts to the annotators that elicit explicit responses (*the user response is determined based on: an explicit response to a question asked via a user interface*). *Supra*, Sections X.A.7, X.A.8. A POSITA would have understood that such prompting is a *question*. *Id.*; EX1003, ¶115. This

is because when an annotator is prompted to select a label, the annotator is being asked which label should be selected for the training image that the annotator is reviewing. *Id.*

Huval further discloses that the relative time taken for human annotators to apply labels is also used by Huval's system to determine how annotator labels were applied to training images (*the user response is determined based on... an implicit data representing one or more of ... time taken to respond*). EX1004, ¶0076; EX1003, ¶115.

12. Claim 11.

Huval and Ross render obvious claim 11. EX1003, ¶116. The Huval-Ross model is, for example, a neural network (*wherein the supervised learning based model is one of: a random forest regressor, a support vector regressor, a simple neural network, a deep convolutional neural network, a recurrent neural network, or a long short-term memory (LSTM) neural network*). EX1004, ¶¶0008-0009, 0016; Section X.A.3.e. (element 2[d]); EX1003, ¶116.

13. Claim 12

Huval and Ross render obvious claim 12. EX1003, ¶117.

a. 12[Pre]

To the extent limiting, the Huval-Ross combination teaches element 12[pre] for the same reasons discussed in Section X.A.3.a (element 2[Pre]). EX1003, ¶118. Moreover, Huval describes its systems and methods “can be embodied and/or

implemented at least in part as a machine configured to receive a computer-readable medium storing computer-readable instructions” that “can be executed by computer-executable components” such as “a processor” or “any suitable dedicated hardware device” (*[a] non-transitory computer readable storage medium storing instructions that when executed by a computer processor cause the computer processor to perform steps of a computer-implemented method*). EX1004, ¶0121; EX1003, ¶118.

b. 12[a]

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

c. 12[b]

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

d. 12[c]

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

e. 12[d]

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

f. 12[e]

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

g. 12[f]

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

h. 12[g]

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

14. Claim 13

See Section X.A.4 (claim 3).

15. Claim 14

See Section X.A.5 (claim 4).

16. Claim 15

See Section X.A.6 (claim 5).

17. Claim 16

See Section X.A.7 (claim 6).

18. Claim 17

See Section X.A.8 (claim 7).

19. Claim 18

See Section X.A.9 (claim 8).

20. Claim 19

See Section X.A.10 (claim 9).

21. Claim 20

See Section X.A.11 (claim 10).

22. Claim 1

Huval and Ross render obvious claim 1. EX1003, ¶134.

a. 1[Pre]

To the extent it is limiting, the Huval-Ross combination teach it for the same reasons discussed in Section X.A.3.a (element 2[Pre]) and X.A.13.a (element

12[Pre]). EX1003, ¶135. Moreover, Huval teaches that its disclosed method “can be executed by a computer network” that includes a “remote computer system,” where its described systems and methods “can be embodied and/or implemented at least in part as a machine configured to receive a computer-readable medium storing computer-readable instructions” that “can be executed by computer-executable components” such as “a processor” or “any suitable dedicated hardware device” (*[a] computer system comprising: a computer processor; and a non-transitory computer readable storage medium storing instructions that when executed by the computer processor cause the computer processor to perform steps of a computer-implemented method*). EX1004, ¶¶0008-0015, 0021-0025, 0029-0030, 0031-0035, 0121; *See* Sections X.A.3.a (element 2[Pre]), X.A.13.a (element 12[Pre]); EX1003, ¶135.

b. 1[a]

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

c. 1[b]

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

d. 1[c]

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

e. 1[d]

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

f. 1[e]

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

g. 1[f]

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

h. 1[g]

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

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

The Huval-Ross combination renders obvious the Challenged Claims. However, Huval and Ross are further combined with Cox regarding the following limitations in Ground 2: (1) *generating a training dataset comprising summary statistics of user responses describing the state of minds of road users displayed in the plurality of images*; (2) *training, using the training dataset, a supervised learning based model configured to predict summary statistics describing a state of mind of a road user displayed in an input image*; and (3) *predicting, by the autonomous vehicle, using the supervised learning based model, summary statistics describing a state of mind of the road user in the new image*. As discussed below, these limitations are explicitly taught by Cox, and it would have been obvious to combine Cox's relevant teachings with Huval and Ross. EX1003, ¶143.

1. Summary of Prior Art

a. Cox–EX1006

Cox is a PCT patent application by Harvard College, which lists Samuel Anthony as an inventor—the same Samuel Anthony listed as an inventor of the ’889 Patent. EX1001, Cover; EX1006, Cover.

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

Cox is analogous art to the ’889 Patent, from the same field of endeavor as the patent (e.g., predictive model techniques), and is reasonably pertinent to the

particular problem that the patent was trying to solve (e.g., navigating an autonomous vehicle in an environment including moving objects). EX1006, ¶¶0029-0032, Figs. 1A-B, 3; EX1001, 1:19-21, 1:57-60, 5:46-53; 6:8-49; Fig. 2A; EX1003, ¶146.

2. Motivation to Combine Huval, Ross, and Cox

A POSITA would have been motivated to combine Huval and Ross as discussed in Section X.A.2. A POSITA would have been further motivated to combine Huval and Ross with Cox such that:

- Cox's psychometric data is generated from the Huval-Ross combination's human annotator response labels and used for model training (EX1006, ¶¶0002-0009, 0010, 0014, 0017-0020, 0025-0032, 0035, 0049, 0063, Claim 11, Fig. 1B, 3; EX1003, ¶147; *See* Section X.A.3.d (element 2[c]));
- Cox's distribution of data and labels (e.g., statistics) and other statistical information are generated from the Huval-Ross combination's human annotator response labels and used for model training (EX1006, ¶¶0002-0006, 0014, 0035, 0048; EX1003, ¶147; *See* Section X.A.3.d (element 2[c])); and
- The training of the Huval-Ross model would have included comparing the predicted intents to the human labeled training data to identify prediction errors and optimize the Huval-Ross model such that the prediction errors are minimized. (EX1006, ¶¶0008, 0019, 0029-0038, 0060; Fig. 4B; *See* Section X.A.3.e (element

2[d])). EX1003, ¶147.

A POSITA would have been motivated to combine Huval, Ross, and Cox in this manner for several reasons. EX1003, ¶148.

A POSITA would have been motivated to combine Huval, Ross, and Cox because Cox expands on the disclosure of Huval and Ross. EX1003, ¶149. The Huval-Ross combination teaches the collection of human annotator responses and training of a predictive model that predicts road user intents. Section X.A.2. Cox merely teaches implementation details regarding the training and optimization of the Huval-Ross combination's model, which would have been routine and readily implemented by a POSITA. EX1003, ¶149. In particular, Cox provides details as to generating additional information (psychometric data, a distribution of data and labels, end of test statistics) from human annotator responses—like those of Huval and Ross—and teaches how this information is added to training data and used to train Huval-Ross combination's predictive model. *Id.*; Section X.A.2.; EX1006, ¶¶0002-0008, 0010, 0014, 0017-0019, 0025-0037, 0048-0049, 0063, Figs. 2B, 3, 4B. The psychometric data is formed by response times for annotations, the accuracy of annotations, and/or the presentation time—the amount of time a training image is presented to each annotator—that is obtained during human annotations. *Id.*

Moreover, Cox further teaches model training techniques to the Huval-Ross combination, including by generating predictions of training data using a model and

then optimizing the model such that the differences between the predictions and the training data is minimized. *Id.* As such, the above description demonstrates that Cox merely provides implementation details regarding how the training of the Huval-Ross combination is effectuated. EX1003, ¶150. A POSITA would thus have been motivated to combine the references and have had a reasonable expectation of success. *Id.*

A POSITA would have further been motivated to combine Huval, Ross, and Cox due to the significant overlap in their teachings. EX1003, ¶151. All three references relate to machine learning models within the context of autonomous vehicles. EX1004, ¶¶0016-0020; EX1005, 1:30-31, 3:39-42, 3:64-5:5; EX1006, ¶¶0056, 0060. And more generally, Huval, Ross, and Cox are all directed to machine learning technologies and involve gathering training data, training a model with the training data, and executing the model, providing that a POSITA seeking to implement the Huval-Ross combination would have looked to Cox and been motivated to combine their teachings for this additional reason. EX1003, ¶151; *see* Section X.A.2.

Indeed, a POSITA would have also been motivated to combine the Huval-Ross combination with Cox because the references provide explicit teachings, suggestions, and motivations to make the combination. EX1003, ¶152. For instance, Huval explains that its model is trained using human annotations and that its trained

model “may exhibit effectiveness and accuracy . . . proportional to a volume and quality of labeled data” provided by human annotators. EX1004, ¶¶0008-0012, 0021-0022, 0036-0050, 0069, 0071-0092, 0099-0120, Figs. 1, 2A, 2B, 3-5; EX1003, ¶152. Ross furthers these teachings, explaining that its model is also trained based on human observations such that its autonomous vehicle “function[s] in a more ‘human-like’ or ‘polite’ way.” *Id.*; EX1005, 5:7-12, 9:13-10:10, 15:1-16:41. And like these two references, Cox notes that its system obtains human annotations and generates statistics associated with the annotations to aid in model training. EX1006, ¶¶0004, 0029-0032.

Cox also explains that its training provides for model predictions that “mimic human performance” and are “more consistent with the decisions of the human annotators.” EX1006, ¶¶0004, 0029-0032. And Cox further expressly discloses that its teachings may be applied to “driverless” (i.e., autonomous) vehicle applications, just like those of Huval and Ross. *Id.*, ¶¶0056, 0027; EX1003, ¶153. Thus, due to the disclosures Huval, Ross, and Cox, a POSITA would have looked to Cox’s annotation and model training techniques when considering Huval and Ross, and would have combined the references with a reasonable expectation of success. EX1003, ¶153.

A POSITA would have additionally been motivated to combine Huval and Ross with Cox because Cox provides benefits to the Huval-Ross combination. EX1003, ¶154. Cox’s teachings provide further training techniques and training data

(i.e., psychometric data, generated statistics from annotation responses) applicable to the training of the Huval-Ross combination's predictive model, as well as model optimization techniques. EX1004, ¶¶0008-0010, 0016-0024, 0029-0039, 0043-0050, 0096, 0100, Fig. 1; EX1005, 1:30-2:53, 4:33-5:12, 9:13-10:10, 15:1-16:52; EX1006, ¶¶0019, 0029-0038, 0060; Fig. 4B; EX1003, ¶154. Cox's teachings thus would have improved the Huval-Ross combination's model by training it to provide more accurate outputs. *Id.* Indeed, the resulting trained model would have been more accurate and human-like as taught by Cox, which would have improved Huval's system by allowing it to better identify objects and their behavior in the environment of its autonomous vehicle. *Id.* This would have improved the safety of the Huval-Ross system by reducing collisions between the autonomous vehicle and objects in the environment. *Id.* The trained model as taught by Cox would have further comported with Ross's goal of autonomous vehicle control that "function[s] in a more 'human-like' or 'polite' way." *Id.*

All the reasons explained above establish that a POSITA would have been motivated to combine Huval and Ross as described. This combination would have been the predictable combination of well-known prior art elements (e.g., Cox's annotation and training details, and Huval-Ross's prediction model and annotation details, discussed above) according to known methods to yield predictable results (Cox's psychometric data is generated from the Huval-Ross combination's human

annotator response labels and used for model training; Cox's distribution of data and labels (e.g., statistics) and other statistical information are generated from the Huval-Ross combination's human annotator response labels and used for model training. The training of the Huval-Ross model would have included comparing the predicted intents to the human labeled training data to identify prediction errors and optimize the Huval-Ross model such that the prediction errors are minimized). *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398, 415-421 (2007); EX1003, ¶155. The combination would have further used known techniques (known annotation and machine learning model training techniques as taught by Cox) to improve similar devices (Huval-Ross combination's prediction model) in the same way (using Cox's aforementioned training techniques). *Id.* Moreover, because Ross and Huval each already teach training a predictive model using labeled training data to generate predictions about objects in an environment of an autonomous vehicle (*see* Section X.A.2), combining Huval and Ross with Cox's teachings about training a predictive model using labeled training data would have required minimal changes. *Id.* Making this combination is thus well within the skill of a POSITA, and a POSITA would have had a reasonable expectation of success. *Id.*

3. Claim 2

Huval, Ross, and Cox render obvious claim 2. EX1003, ¶156.

a. 2[Pre]

See Section X.A.3.a (element 2[Pre]).

b. 2[a]

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

c. 2[b]

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

d. 2[c]

Huval, Ross, and Cox teach element 2[c]. EX1003, ¶160. The Huval-Ross-Cox combination provides that psychometric data is generated from the Huval-Ross combination's human annotator response labels regarding road user actions in image data and added to training data for Huval's model (*generating a training dataset comprising summary statistics of user responses describing the state of minds of road users displayed in the plurality of images*). *See* Sections X.B.2., X.A.3.d (element 2[c]); EX1006, ¶¶0002-0009, 0010, 0014, 0017-0020, 0025-0032, 0035, 0049, 0063, Claim 11, Fig. 1B, 3; EX1003, ¶160. This psychometric data summarizes statistics on how human annotators annotated training images because it includes "response times" for annotations, "accuracy" of annotations, and "presentation time" indicating how long an annotator viewed a training image, and thereby teaches *summary statistics of user responses*. *Id.* A POSITA would have recognized that such a collection of data regarding annotations forms statistics. *Id.*; EX1003, ¶160. Moreover, Cox's description of "psychometric data" is consistent

with the '889 Patent's explanation. Just like Cox, the '889 Patent explains that "summary statistics" may "characterize the human observer responses in terms of certain parameters associated with the statistics, such as a content of a response" and "a time associated with entering a response." EX1001, 6:18-22; EX1003, ¶160.

Cox further provides additional examples of generating *summary statistics* based on annotator responses. EX1003, ¶161. Cox discloses "item-response curves across large populations of humans (e.g., how consistent are judgment[s] across a population)," which provide a statistical representation of how human annotators annotated training images. *Id.*; EX1006, ¶¶0002-0008, 0010, 0014, 0018, 0025-0032, 0035, 0049, 0063, Fig. 3. This teaching applies to the Huval-Ross combination's disclosure regarding annotator responses identifying observed road user actions in a sample training image. *Id.*; Sections X.A.3.c (element 2[b]), X.A.3.d (element 2[c]); EX1003, ¶161. Here, Cox teaches that the Huval-Ross combination generates an item response curve describing the human annotated/labeled actions corresponding to the intents of road users. *Id.*; EX1003, ¶161. Such intents describe the state of minds of road users in the images. *Id.* And as taught by Huval, the item response curve would be included as part of training data for the Huval-Ross model. *Id.*; EX1004, Abstract, ¶¶0008-0018, 0021-0025, 0029-0030, 0036-0050, 0085-0090, 0095-0098, 0118, Figs. 1, 3, and 4. This disclosure further teaches *generating a training dataset comprising summary statistics of user responses describing the*

state of minds of road users displayed in the plurality of images. Id.; EX1003, ¶161.

And finally, the Huval-Ross-Cox combination further provides that a distribution of data and labels (e.g., statistics) and “end-of-test statistics” are generated from the Huval-Ross combination’s human annotator response labels regarding road user actions in image data and added to training data for Huval’s model, further teaching *generating a training dataset comprising summary statistics of user responses describing the state of minds of road users displayed in the plurality of images* . EX1006, ¶¶0002-0009, 0010, 0014, 0017-0020, 0025-0035, 0048-0049, 0063, Claim 11; EX1003, ¶162; *See* Sections X.B.2., X.A.3.d (element 2[c]).

e. 2[d]

Huval, Ross, and Cox teach element 2[d]. EX1003, ¶163. As applied to Huval and Ross, Cox teaches that the training of the Huval-Ross model would have included executing the Huval-Ross model on the training data (*see* Section X.B.3.d (element 2[c]) during a training phase (*training, using the training dataset, a supervised learning based model*). EX1004, ¶¶0008-0011, 0016, 0021-0024, 0088-0092; EX1005, 4:47-65, 7:57-8:31, 9:13-10:10, 15:1-16:52, 17:43-18:2, Fig. 10; EX1006, ¶¶0008, 0019, 0029-0038, 0060; Fig. 4B; EX1003, ¶163. In the training, the Huval-Ross model would have predicted a “set of possible intents or hypotheses identifying possible next actions (and predicted trajectories), a given

point in time when the action is likely to occur, and associated likelihood values” for road users in image data (*supervised learning based model configured to predict summary statistics describing a state of mind of a road user displayed in an input image*). *Id.* And as taught by Cox, the training would have included comparing the predicted intents to the human labeled training data to identify prediction errors and optimize the Huval-Ross model. *Id.*

Specifically, Cox teaches the training of a machine learning model using psychometric data from the human annotators, including through the use of a “loss function.” EX1006, ¶¶0031-0032, Fig. 4B; EX1003, ¶164; Section X.B.2. The loss function penalizes misclassification by the model in relation to human annotated training data and thus reduces model error. EX1006, ¶¶0032, 0034, 0060. Using the loss function, Cox teaches to Huval and Ross model training (*training, using the training dataset, a supervised learning based model*) by predicting classifications of training data and optimizing the model such that the difference between the predicted classifications output by the model and the training data is minimized (*supervised learning based model configured to predict summary statistics describing a state of mind of a road user displayed in an input image*). EX1006, ¶¶0008, 0019, 0029-0038, 0060; Fig. 4B; EX1003, ¶164. This teaching is consistent with the ’889 Patent, which states that “trained” means, “in some embodiments,” that “the difference between the summary statistics output by the neural network and the summary

statistics calculated from the responses of the human observers [] is minimized.”

EX1001, 12:20-25.

f. 2[e]:

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

g. 2[f]

Huval, Ross, and Cox teach element 2[f]. EX1003, ¶166. As combined, the Huval-Ross trained model would have operated on new images to predict the actions and possible intents of vehicles, bicyclists, and pedestrians in the images, teaching element 2[f] for the same reasons discussed in Section X.A.3.g (element 2[f]). EX1003, ¶166.

h. 2[g]:

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

4. Claim 10

Huval, Ross, and Cox render obvious claim 10. EX1003, ¶168. Cox teaches to the Huval-Ross combination that “sets of queries are posed to a crowd” of human annotators on a display module and responses are collected via the display module, processed, and analyzed (*wherein the user response is determined based on: an explicit response to a question asked via a user interface*) to train the Huval model. EX1006, ¶¶0016-0018, 0032, 0048; Section X.B.2; EX1003, ¶168.

The Huval-Ross-Cox combination further provides that psychometric data is generated from human annotator response labels regarding road user actions in

image data and added to training data for Huval’s model. *See* Sections X.B.2., X.B.3.d (element 2[c]); EX1006, ¶¶0002-0009, 0010, 0014, 0017-0020, 0025-0032, 0035, 0049, 0063, Claim 11, Fig. 1B, 3; EX1003, ¶169. The psychometric data includes “response times” for annotations, “accuracy” of annotations, and “presentation time” indicating how long an annotator viewed a training images (*an implicit data representing one or more of: time taken to respond . . .*). *Id.*

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

Huval, Ross, and Cox render obvious claims 1, 3-9, and 11-19 for the same reasons discussed in Sections X.A.22, X.A.4-X.A.10, and X.A.12-X.A.20, respectively. EX1003, ¶170.

6. Claim 20.

Huval, Ross, and Cox render obvious claim 20 for the same reasons discussed in Section X.B.4. EX1003, ¶171.

C. Ground 3: Claims 1-20 are rendered obvious under 35 U.S.C. § 103 over Huval, Ross, and Munro.

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

and it would have been obvious to combine Munro’s relevant teachings with Huval and Ross. EX1003, ¶172.

1. Summary of Prior Art

a. Munro–EX1009

Munro is generally directed to techniques for collecting human annotator responses to a training sample, where the responses are used to train a machine learning model. EX1009, Abstract, ¶¶0005, 0042-0048, 0132-0142. For example, the system performs an “aggregation process” that involves determining how many human annotators classified the stimulus as belonging to a first label and how many annotators classified the stimulus as belonging to a second label. *Id.*, ¶¶0140-0142. Munro’s machine learning model is then trained based on the aggregated responses. EX1009, ¶¶0140-0145.

Munro is analogous art to the ’889 Patent, from the same field of endeavor as the patent (e.g., predictive model techniques), and reasonably pertinent to the particular problem that the patent was trying to solve (e.g., navigating an autonomous vehicle in an environment including moving objects). EX1009, Abstract, ¶¶0005, 0042-0048, 0132-0146; EX1001, 1:19-21, 1:57-60, 5:46-53; 6:8-48; EX1003, ¶174.

2. Motivation to Combine Huval, Ross, and Munro

A POSITA would have been motivated to combine Huval, Ross, and Munro such that Huval-Ross’s received human annotator response labels for sample images

would have been analyzed to generate tallies of the number of annotators who classified the behavior or activity as having various labels. EX1009, Abstract, ¶¶0005, 0042-0048, 0132-0142 ; *see* Sections X.A.2; X.A.3.d (element 2[c]); EX1003, ¶175. In this configuration, the tallies would have been part of Huval's model training data. *Id.* A POSITA would have been motivated to combine Huval, Ross, and Munro in this manner for several reasons. EX1003, ¶175.

A POSITA would have been motivated to combine Huval, Ross, and Munro because Munro simply provides implementation details to the Huval-Ross combination. EX1003, ¶176. Specifically, Munro merely expands on the Huval-Ross combination by explaining that the human annotations received by Huval-Ross are tallied to determine the number of annotators that labeled training images in particular ways. EX1004, ¶¶0008-0020, 0016-0024, 0029-0039, 0036-0050, 0085-0090, 0095-0100, 0118, Fig. 1; EX1005, 1:30-2:53, 4:33-5:11, 9:13-10:10, 15:1-16:52; EX1009, Abstract, ¶¶0005, 0042-0048, 0132-0142; EX1003, ¶176. A POSITA would have thus recognized that Munro's teachings provide nothing more than a routine and well-known method of organization to the Huval-Ross annotations, and the combination thus has a reasonable expectation of success. *Id.*

A POSITA would have further combined the references due to the overlap in their teachings. EX1003, ¶177. Huval, Ross, and Munro are all directed to machine learning technologies and involve gathering human annotator responses, providing

that a POSITA seeking to implement the Huval-Ross combination's teachings would have looked to Munro. EX1003, ¶177.

A POSITA would have further been motivated to combine Huval, Ross, and Munro because Munro would provide performance benefits to the Huval and Ross combination. Section X.A.2; EX1003, ¶178. For example, Munro's tallying would provide additional data regarding annotator responses that could be used to better train the Huval-Ross prediction model. *Id.*; EX1004, ¶¶0008-0020, 0016-0024, 0029-0039, 0036-0050, 0085-0090, 0095-0100, 0118, Fig. 1; EX1005, 1:30-2:53, 4:33-5:11, 9:13-10:10, 15:1-16:52; EX1009, ¶¶0140-0142. By training the model with such additional training data, the model would have in turn undergone additional training, which would have improved model accuracy. *Id.*; EX1003, ¶178. And, training the Huval-Ross predictive model using the implementation details taught by Munro would have ensured that the model iteratively improves in accuracy. *Id.*

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

sample images would have been analyzed to generate tallies of the number of annotators who classified the behavior or activity as having various labels, as taught by Munro). *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398, 415-421 (2007); EX1003, ¶179. The combination would have further used known techniques (known annotation organization techniques taught by Munro) to improve similar devices (Huval and Ross combination's prediction system that obtains human annotator labels) in the same way (using Munro's aforementioned techniques). *Id.* A POSITA would have had a reasonable expectation of success in combining Munro with Huval and Ross because Munro's teachings simply provide details about how the Huval-Ross annotator responses would be organized and used. *Id.*

3. Claim 2

Huval, Ross, and Munro render obvious claim 2. EX1003, ¶180.

a. 2[Pre]-2[b]

Huval, Ross, and Munro teach elements 2[Pre] through 2[b] for the same reasons discussed in Sections X.A.3.a through X.A.3.c. EX1003, ¶181.

b. 2[c]

Huval, Ross, and Munro teach element 2[c]. EX1003, ¶182. Huval and Ross teach received human annotator response labels for sample images including observed road user actions. *See* Section X.A.3.d (element 2[c]); EX1003, ¶182. Huval, Ross, and Munro teach that these response labels would have been used to generate tallies of the number of annotators who classified the behavior or activity

in the images as having various labels in their responses, and the tallies would have been part of Huval’s model training data (*generating a training dataset comprising summary statistics of user responses describing the state of minds of road users displayed in the plurality of images*). *Id.*; EX1009, Abstract, ¶¶0005, 0042-0048, 0132-0142; EX1003, ¶182. A POSITA would have recognized that these tallies are *summary statistics* as claimed because they are a collection of numerical measurements summarizing how annotators classified behavior or activity. *Id.* This combined teaching of Huval, Ross, and Munro aligns with the ’889 Patent’s disclosure. EX1003, ¶182. Specifically, the ’889 Patent explains, for example, that the generation of “summary statistics” that “characterize the aggregate responses of multiple human observers to a particular derived stimulus” (i.e., training image) includes “how many human observers” classified the stimulus as displaying various content. EX1001, 6:8-18; EX1003, ¶182.

c. 2[d]

Huval, Ross, and Munro teach element 2[d]. EX1003, ¶183. As combined, the training of the Huval-Ross model would have been performed as discussed in Section X.A.3.e (element 2[d]), except the training data includes the generated summary statistics as taught by Munro, and the model is trained on such training data in the same manner discussed in Section X.A.3.e (element 2[d]). Section X.A.3.e, X.C.3.b; EX1003, ¶183.

d. 2[e]

Huval, Ross, and Munro teach element 2[e] for the same reasons discussed in Sections X.A.3.f (element 2[e]). EX1003, ¶184.

e. 2[f]

Huval, Ross, and Munro teach element 2[f]. EX1003, ¶185. As combined, the Huval-Ross trained model would have operated on new images to determine the actions and possible intents of vehicles, bicyclists, and pedestrians in the images, teaching element 2[f] for the same reasons discussed in Section X.A.3.g (element 2[f]). EX1003, ¶185; Sections X.A.3.g, X.C.3.b-X.C.3.d.

f. 2[g]

Huval, Ross, and Munro teach element 2[g] for the same reasons discussed in Sections X.A.3.h (element 2[g]). EX1003, ¶186.

4. Claim 12

See Section X.C.3 (claim 2), X.A.13 (claim 12), and X.A.3 (claim 2).

5. Claim 1

See Section X.C.3 (claim 2), X.A.22 (claim 1), X.A.13 (claim 12) and X.A.3 (claim 2).

6. Claims 3-11 and 13-20

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

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

The Huval-Ross-Cox combination renders obvious the Challenged Claims. However, Huval, Ross, and Cox are further combined with Munro regarding the following limitation in Ground 4: *generating a training dataset comprising summary statistics of user responses describing the state of minds of road users displayed in the plurality of images*. As discussed below, this limitation is explicitly taught by Munro, and it would have been obvious to combine Munro's relevant teachings with Huval, Ross, and Cox. EX1003, ¶190.

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

A POSITA would have been motivated to combine Huval, Ross, Cox and Munro for the same reasons discussed in Sections X.B.2 and X.C.2. Sections X.B.2, X.C.2; EX1003, ¶191. The same motivations that would apply to add Munro to Huval and Ross apply equally to the combination of Huval, Ross, and Cox. *Id.* Similarly, the same motivations that would apply to add Cox to Huval and Ross apply equally to the combination of Huval, Ross, and Munro. *Id.* There is no inconsistency with adding the Munro's annotator tallying and training teachings, and adding Cox's training data, statistical, and model training/optimization teachings, to the Huval and Ross combination. *Id.* Moreover, combining Munro and Cox together with Huval and Ross would have had no impact on a POSITA's reasonable expectation of success of combining Munro and Cox separately with Huval and Ross.

Id.

2. Claim 2

Huval, Ross, Cox, and Munro render obvious claim 2. EX1003, ¶192.

a. 2[Pre]-2[b]

Huval, Ross, Cox, and Munro teach elements 2[Pre] through 2[b] for the same reasons discussed in Sections X.B.3.a through X.B.3.c. EX1003, ¶193.

b. 2[c]

Huval, Ross, Cox, and Munro teach element 2[c]. EX1003, ¶194. The combination teaches that the generated training dataset of Huval would have included statistical data of annotator responses from human annotators regarding road user actions in images, including tallies of the number of annotators who classified road user actions as having various labels in their responses with respect to various images, as taught by Munro (*generating a training dataset comprising summary statistics of user responses describing the state of minds of road users displayed in the plurality of images*). See X.B.3.d (element 2[c]), X.C.3.b (element 2[c]); EX1003, ¶194.

c. 2[d]

Huval, Ross, Cox, and Munro teach element 2[d]. EX1003, ¶195. As combined, the training of Huval-Ross model would have been performed as discussed in Section X.A.3.e (element 2[d] – Ground 1) and Section X.B.3.e (element 2[d] – Ground 2), except the training data includes the generated summary

statistics as taught by Munro, and the Huval-Ross model is trained on such training data in the same manner discussed in Section X.A.3.e (element 2[d] – Ground 1) and Section X.B.3.e (element 2[d] – Ground 2). EX1003, ¶195; Sections X.A.3.e, X.B.3.e, X.C.3.b, X.C.3.c.

d. 2[e]

Huval, Ross, Cox, and Munro teach element 2[e] for the same reasons discussed in Sections X.B.3.f (element 2[e]). EX1003, ¶196.

e. 2[f]

Huval, Ross, Cox, and Munro teach element 2[f]. EX1003, ¶197. As combined, the Huval-Ross model would have operated on new images to determine the actions and possible intents of vehicles, bicyclists, and pedestrians in the images. *See* Sections X.B.3.g (element 2[f]), X.C.3.e, X.D.2.b-X.D.2.d.

f. 2[g]

Huval, Ross, Cox, and Munro teach element 2[g] for the same reasons discussed in Sections X.B.3.h (element 2[g]). EX1003, ¶198.

3. Claims 10 and 20

Huval, Ross, Cox, and Munro render obvious claims 10 and 20 for the same reasons discussed for Claim 10 and 20 in Ground 2. *Supra* Sections X.B.4, X.B.6; EX1003, ¶199.

4. Claim 12

See Section X.D.2 (claim 2), X.A.13 (claim 12), X.C.4 (claim 12), and X.B.3 (claim 2).

5. Claim 1

See Section X.D.2 (claim 2), X.A.22 (claim 1), X.A.13 (claim 12), X.C.5 (claim 1), and X.B.3 (claim 2).

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

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

E. Printed Matter

Many limitations of the Challenged Claims are not entitled to patentable weight because they merely recite the content of information, i.e. printed matter. *Praxair Distribution, Inc. v. Mallinckrodt Hosp. Prods. IP Ltd.*, 890 F.3d 1024, 1032 (Fed. Cir. 2018) (“Claim limitations directed to the content of information and lacking a requisite functional relationship are not entitled to patentable weight because such information is not patent eligible subject matter under 35 U.S.C. § 101.”). In particular, the following limitations are printed matter:

- 1[Pre], 2[Pre], 12[Pre] – “a predicted state of mind of road users in a scene captured by a camera of the autonomous vehicle”
- 1[a], 2[a], 12[a] – “a plurality of images displaying road scenes”

- 1[b], 2[b], 12[b] – “a plurality of user responses, each user response describing a state of mind of a road user displayed in the one or more images”
- 1[c], 2[c], 12[c] – “summary statistics of user responses describing the state of minds of road users displayed in the plurality of images”
- 1[d], 2[d], 12[d] – “training dataset” and “summary statistics describing a state of mind of a road user displayed in an input image”
- 1[e], 2[e], 12[e] – “a new image of a scene including a road user”
- 1[f], 2[f], 12[f] – “summary statistics describing a state of mind of the road user in the new image”
- 3, 13 – “the road user in the input image [being] a pedestrian” and “the state of mind predicted by the supervised learning based model indicates whether the pedestrian is likely to perform an action, the action including one of: staying in place, or crossing a street”
- 4, 14 – “the road user in the input image [being] a cyclist” and “the state of mind predicted by the supervised learning based model indicates whether the cyclist is likely to cross a lane”
- 5, 15 – “the summary statistics comprises one or more of: a central tendency, a variance, a skew, a kurtosis, a scale, or a histogram”

- 6, 16 – “images including a road user” and “the user interface is configured to prompt the viewer to describe how the road user will act”
- 7, 17 – “stimulus comprising a set of images” and “user response describing the stimulus”
- 8, 18 – “subset of the image contained in a bounding box around the road user displayed in the image”
- 10, 20 – “an explicit response to a question asked via a user interface” and “an implicit data representing ... time taken to respond”

Each of the above limitations recites a data element and then defines that data element based on its content. But a patent cannot cover what “images display[],” or what a “summary statistics [...] describe[e]”—such limitations merely cover “information claimed for its communicative content.” *C R Bard Inc. v. AngioDynamics, Inc.*, 979 F.3d 1372, 1381 (Fed. Cir. 2020). For example, the recitation in limitation 2[b] that the user responses “describe[e] a state of mind of a road user displayed in one or more images” should be ignored because the content of the user response is not patentable. Similarly, the recited limitations about what the predicted “state of mind ... indicates” in claims 3-4 and 13-14 should be ignored—the patentee cannot claim the content of a mental state. *See Praxair*, 890 F.3d at 1034. Because each of the above limitations are defined by the information contained in various data elements, they are patent ineligible printed matter.

Printed matter is only entitled to patentable weight if it is functionally related to the substrate on which the information is present. *See In re Marco Guldenaar Holding B.V.*, 911 F.3d 1157, 1161 (Fed. Cir. 2018) (markings on dice are “not functionally related to the substrate of the dice”). The above printed matter limitations do not have any such functional relationship with the “substrate,” which in this case is computer memory. EX1003, ¶205; *see also* 1[Pre] and 12[Pre]. Rather than impacting how the computer memory operates, the printed matter limitations in the Challenged Claims “merely informs people of the claimed information[.]” *C R Bard Inc.*, 979 F.3d at 1381. As such, none of these limitations can be relied on to distinguish the Challenged Claims from the cited prior art.

XI. CONCLUSION

Petitioner respectfully requests institution of IPR and that the Challenged Claims be canceled as unpatentable pursuant to 35 U.S.C. §318(b).

Respectfully Submitted,

Date: October 2, 2025

/Roger Fulghum/

Roger Fulghum, Reg. No. 39,678
Attorney for Petitioner, Tesla, Inc.

CERTIFICATE OF SERVICE

In accordance with 37 C.F.R. §§42.6(e) and 42.105, the undersigned certifies that on October 2, 2025, a complete and entire copy of the **PETITION FOR *INTER PARTES* REVIEW OF CLAIMS 1-20 OF U.S. PATENT NO. 11,126,889** including exhibits and testimony relied upon and Power of Attorney were served on Patent Owner via FedEx overnight at the correspondence address of record for the '889 Patent and counsel for Patent Owner in the EDTX Litigation, as included below:

758 - FENWICK & WEST LLP
c/o Rajendra B. Panwar (rpanwar@fenwick.com)
Silicon Valley Center
801 California Street
Mountain View, CA 94041

Patrick J. Conroy (pat@nelbum.com)
Andrea L. Fair (andrea@millerfairhenry.com)

Date: October 2, 2025

/Roger Fulghum/

Roger Fulghum, Reg. No. 39,678
Attorney for Petitioner, Tesla, Inc.

CERTIFICATION UNDER 37 C.F.R. §42.24(d)

Pursuant to 37 C.F.R. §42.24(d), the undersigned hereby certifies that the word count under §42.24(a)(1) for the foregoing Petition for *Inter Partes* Review totals 13,775 words, within the 14,000 word limit allowed under §42.24(a)(1)(i).

Date: October 2, 2025

/Roger Fulghum/
Roger Fulghum, Reg. No. 39,678
Attorney for Petitioner, Tesla, Inc.