

Petition for *Inter Partes* Review of U.S. Patent No. 11,753,046

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

TESLA, INC.

Petitioner,

v.

PERCEPTIVE AUTOMATA LLC

Patent Owner

IPR2025-01575

U.S. Patent No.: 11,753,046 B2

**PETITION FOR *INTER PARTES* REVIEW OF CLAIMS 1-19 OF U.S.
PATENT NO. 11,753,046**

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

TABLE OF CONTENTS

	Page
PETITIONER’S EXHIBIT LIST	VIII
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	8
IV. PAYMENT OF FEES UNDER 37 C.F.R. § 42.103	8
V. CERTIFICATION OF GROUNDS FOR STANDING	8
VI. OVERVIEW OF CHALLENGE AND RELIEF REQUESTED	8
A. Prior Art Printed Publications	8
B. Identification of Challenge and Statement of Precise Relief Requested	9
VII. SUMMARY OF THE ’046 PATENT	10
A. Prosecution History	10
B. Technology Background	10
C. The Alleged Invention.....	11
VIII. LEVEL OF ORDINARY SKILL IN THE ART	12
IX. CLAIM CONSTRUCTION	13
X. THE ASSERTED GROUNDS OF INVALIDITY	13
A. Ground 1: Claims 1-19 are obvious over Cox	13

Petition for *Inter Partes* Review of U.S. Patent No. 11,753,046

1.	Summary of Prior Art	13
	a. Cox-EX1007.....	13
2.	Claim 1	14
	a. 1[Pre].....	14
	b. 1[a].....	15
	c. 1[b].....	16
	d. 1[b-i].....	17
	e. 1[b-ii].....	19
	f. 1[b-iii]	20
	g. 1[b-iv].....	23
	h. 1[c].....	24
	i. 1[d].....	27
3.	Claim 2.....	28
4.	Claim 3.....	29
5.	Claim 4.....	29
6.	Claim 5.....	30
7.	Claim 6.....	30
8.	Claim 7.....	31
9.	Claim 8.....	31
	a. 8[Pre].....	31
	b. 8[a].....	32
	c. 8[b].....	32
	d. 8[b-i]	32
	e. 8[b-ii].....	32
	f. 8[b-iii].....	32

Petition for *Inter Partes* Review of U.S. Patent No. 11,753,046

g.	8[b-iv]	32
h.	8[c]	32
i.	8[d]	32
10.	Claim 9	32
11.	Claim 10	33
12.	Claim 11	33
13.	Claim 12	33
14.	Claim 13	33
15.	Claim 14	33
16.	Claim 15	33
a.	15[Pre]	33
b.	15[a]	33
c.	15[b]	33
d.	15[b-i]	33
e.	15[b-ii]	33
f.	15[b-iii]	33
g.	15[b-iv]	33
h.	15[c]	34
i.	15[d]	34
17.	Claim 16	34
18.	Claim 17	34
19.	Claim 18	34
20.	Claim 19	34
B.	Ground 2: Claims 1-19 are obvious over Cox and Ross	34
1.	Summary of Prior Art	35

Petition for *Inter Partes* Review of U.S. Patent No. 11,753,046

a.	Ross–EX1005.....	35
2.	Motivation to Combine Cox and Ross.....	36
3.	Claim 1	40
a.	1[Pre], 1[a], 1[b].....	40
b.	1[b–i].....	40
c.	1[b–ii] – 1[b–iv]	42
d.	1[c].....	42
e.	1[d].....	43
4.	Claim 3	43
5.	Claim 4.....	44
6.	Claim 8.....	45
7.	Claim 10.....	45
8.	Claim 11.....	45
9.	Claim 15.....	45
10.	Claim 17.....	45
11.	Claim 18.....	45
12.	Claims 2, 5, 6, 7, 9, 12, 13, 14, 16, and 19	45
C.	Ground 3: Claims 5, 12, and 19 are obvious over Cox and Ellenbogen.....	46
1.	Summary of Prior Art	46
a.	Ellenbogen–EX1004	46
2.	Motivation to Combine Cox and Ellenbogen	47
3.	Claim 5.....	48
4.	Claim 12.....	49
5.	Claim 19.....	49

Petition for *Inter Partes* Review of U.S. Patent No. 11,753,046

D.	Ground 4: Claims 1-19 are obvious over Cox and Munro.....	49
1.	Summary of Prior Art	49
a.	Munro–EX1006	49
2.	Motivation to Combine Cox and Munro.....	50
3.	Claim 1	52
a.	1[Pre]-1[b-ii]	52
b.	1[b–iii].....	52
c.	1[b-iv].....	53
d.	1[c]	54
e.	1[d].....	54
4.	Claim 8.....	54
5.	Claim 15	54
6.	Claims 2-7, 9-14, and 16-19	54
E.	Ground 5: Claims 1-19 are obvious over Ellenbogen and Munro	55
1.	Motivation to Combine Ellenbogen and Munro	55
2.	Claim 1	59
a.	1[Pre].....	59
b.	1[a]	59
c.	1[b].....	60
d.	1[b–i].....	61
e.	1[b–ii].....	63
f.	1[b–iii]	64
g.	1[b–iv].....	65
h.	1[c]	67
i.	1[d].....	70

Petition for *Inter Partes* Review of U.S. Patent No. 11,753,046

3.	Claim 2.....	70
4.	Claim 3.....	71
5.	Claim 4.....	71
6.	Claim 5.....	72
7.	Claim 6.....	72
8.	Claim 7.....	73
9.	Claim 8.....	73
	a. 8[Pre].....	73
	b. 8[a].....	74
	c. 8[b].....	74
	d. 8[b-i].....	74
	e. 8[b-ii].....	74
	f. 8[b-iii].....	74
	g. 8[b-iv].....	74
	h. 8[c].....	74
	i. 8[d].....	74
10.	Claim 9.....	74
11.	Claim 10.....	74
12.	Claim 11.....	75
13.	Claim 12.....	75
14.	Claim 13.....	75
15.	Claim 14.....	75
16.	Claim 15.....	75
	a. 15[Pre].....	75
	b. 15[a].....	75

Petition for *Inter Partes* Review of U.S. Patent No. 11,753,046

c.	15[b]	75
d.	15[b-i]	75
e.	15[b-ii]	75
f.	15[b-iii]	75
g.	15[b-iv]	75
h.	15[c]	75
i.	15[d]	76
17.	Claim 16	76
18.	Claim 17	76
19.	Claim 18	76
20.	Claim 19	76
F.	Printed Matter	76
XI.	CONCLUSION	78

PETITIONER’S EXHIBIT LIST

Ex[No.]	Description of Documents
1001	U.S. Patent No. 11,753,046 (“’046 Patent”)
1002	File History of U.S. Patent No. 11,753,046
1003	Declaration of Dr. Jason Janét
1004	U.S. Patent Application Publication No. 2017/0099200 (“Ellenbogen”)
1005	U.S. Patent No. 10,496,091 (“Ross”)
1006	U.S. Patent Application Publication No. 2016/0162456 (“Munro”)
1007	International Publication WO 2014/210334 (“Cox”)
1008	File History of U.S. Patent No. 10,402,687
1009	File History of U.S. Patent No.10,614,344
1010	File History of U.S. Patent No. 11,126,889
1011	Complaint in <i>Perceptive Automata LLC v. Tesla, Inc.</i> , Case No. 2:25-cv-00742 (E.D. Tex. filed July 23, 2025)
1012	U.S. Patent No. 10,733,506 (“Ogale”)
1013	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. 11,753,046 (the “’046 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”). EX1011. To the best knowledge of Petitioner, the ’046 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. 11,753,046

	<p>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</p> <p>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</p> <p>Ashraf Fawzy, Reg. No. 67,914 Tesla, Inc. 800 Connecticut Ave. NW Washington, DC 20006 Phone: (202) 905-9221 afawzy@tesla.com</p>
--	---

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 '046 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:

Petition for *Inter Partes* Review of U.S. Patent No. 11,753,046

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-implemented method comprising:
1[a]	storing a plurality of images, each image displaying one or more users;
1[b]	generating training data from the plurality of images, the generating comprising, for each image:
1[b-i]	sending the image to a plurality of human observers, each human observer presented with a request to answer a question about a state of mind of a user in the image,
1[b-ii]	receiving, from each of the plurality of human observers, a response representing a judgment by the human observer of the state of mind of the user in the image,
1[b-iii]	generating summary statistics describing the state of mind of the user in the image based on the received responses from the plurality of human observers, and
1[b-iv]	storing the summary statistics in association with the image as part of the training data;
1[c]	training a model using the training data, the model configured to receive an input image showing a user and predict summary statistics describing a state of mind of the user in the input image; and

1[d]	executing the trained model to predict a state of mind of a user in a new image.
------	--

Claim 2

2	The computer-implemented method of claim 1, wherein the image is manipulated by adjusting values of pixels of the image before presenting to a human observer.
---	--

Claim 3

3	The computer-implemented method of claim 1, wherein the state of mind of the user in the image indicates whether the user is likely to perform a predetermined action.
---	--

Claim 4

4	The computer-implemented method of claim 1, wherein the state of mind of the user in the image represents a measure of awareness of the user regarding an object.
---	---

Claim 5

5	The computer-implemented method of claim 1, wherein the response from a human observer comprises a rating on an ordinal scale.
---	--

Claim 6

6	The computer-implemented method of claim 1, wherein the 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 7

7	The computer-implemented method of claim 1, wherein the summary statistics is associated with at least one of a content of a response, a time associated with entering a response, and a position of an eye of a human observer associated with the response, the position being measured with respect to a display associated with the image.
---	--

Claim 8

8[Pre]	A non-transitory computer readable storage medium storing instructions that when executed by one or more processors, cause the one or more processors to perform steps comprising:
8[a]	storing a plurality of images, each image displaying one or more users;
8[b]	generating training data from the plurality of images, the generating comprising, for each image:
8[b-i]	sending the image to a plurality of human observers, each human observer presented with a request to answer a question about a state of mind of a user in the image,
8[b-ii]	receiving, from each of the plurality of human observers, a response representing a judgment by the human observer of the state of mind of the user in the image,
8[b-iii]	generating summary statistics describing the state of mind of the user in the image based on the received responses from the plurality of human observers, and
8[b-iv]	storing the summary statistics in association with the image as part of the training data;
8[c]	training a model using the training data, the model configured to receive an input image showing a user and predict summary statistics describing a state of mind of the user in the input image; and
8[d]	executing the trained model to predict a state of mind of a user in a new image.

Claim 9

9	The non-transitory computer readable storage medium of claim 8, wherein the image is manipulated by adjusting values of pixels of the image before presenting to a human observer.
----------	--

Claim 10

10	The non-transitory computer readable storage medium of claim 8, wherein the state of mind of the user in the image indicates whether the user is likely to perform a predetermined action.
-----------	--

Claim 11

11	The non-transitory computer readable storage medium of claim 8, wherein the state of mind of the user in the image represents a measure of awareness of the user regarding an object.
-----------	---

Claim 12

12	The non-transitory computer readable storage medium of claim 8, wherein the response from a human observer comprises a rating on an ordinal scale.
-----------	--

Claim 13

13	The non-transitory computer readable storage medium of claim 8, wherein the 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 14

14	The non-transitory computer readable storage medium of claim 8, wherein the summary statistics is associated with at least one of a content of a response, a time associated with entering a response, and a position of an eye of a human observer associated with the response, the position being measured with respect to a display associated with the image.
-----------	--

Claim 15

15[Pre]	A computing system comprising: one or more processors; and a non-transitory computer readable storage medium, storing instructions that when executed by the one or more processors, cause the one or more processors to perform steps comprising:
15[a]	storing a plurality of images, each image displaying one or more users;
15[b]	generating training data from the plurality of images, the generating comprising, for each image:
15[b-i]	sending the image to a plurality of human observers, each human observer presented with a request to answer a question about a state of mind of a user in the image,
15[b-ii]	receiving, from each of the plurality of human observers, a response representing a judgment by the human observer of the state of mind of the user in the image,
15[b-iii]	generating summary statistics describing the state of mind of the user in the image based on the received responses from the plurality of human observers, and
15[b-iv]	storing the summary statistics in association with the image as part of the training data;

Petition for *Inter Partes* Review of U.S. Patent No. 11,753,046

15[c]	training a model using the training data, the model configured to receive an input image showing a user and predict summary statistics describing a state of mind of the user in the input image; and
15[d]	executing the trained model to predict a state of mind of a user in a new image.

Claim 16

16	The computing system of claim 15, wherein the summary statistics is associated with at least one of a content of a response, a time associated with entering a response, and a position of an eye of a human observer associated with the response, the position being measured with respect to a display associated with the image.
-----------	--

Claim 17

17	The computing system of claim 15, wherein the state of mind of the user in the image indicates whether the user is likely to perform a predetermined action.
-----------	--

Claim 18

18	The computing system of claim 15, wherein the state of mind of the user in the image represents a measure of awareness of the user regarding an object.
-----------	---

Claim 19

19	The computing system of claim 15, wherein the response from a human observer comprises a rating on an ordinal scale.
-----------	--

III. INTRODUCTION

Tesla, Inc. (“Petitioner”) requests *inter partes* review (“IPR”) of claims 1-19 (the “Challenged Claims”) of U.S. Patent No. 11,753,046 (the “’046 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 ’046 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 ’046 Patent was filed on July 16, 2019 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

¹ The ’046 Patent was filed after March 16, 2013. America Invents Act (“AIA”) 35 U.S.C. §§102 and 103 thus applies.

Petition for *Inter Partes* Review of U.S. Patent No. 11,753,046

the '046 Patent is entitled to the July 5, 2017 priority date (*see* IPR2025-01573). However, this issue is moot because Ellenbogen, Ross, Munro, and Cox are prior art to the '046 Patent before July 5, 2017.

- International Publication No. WO2014/210334 (“Cox”) (EX1007); 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 No. 10,496,091 (“Ross”) (EX1005); filed on August 17, 2016; prior art under 35 U.S.C. § 102(a)(2).
- U.S. Patent Application Publication No. 2017/0099200 (“Ellenbogen”) (EX1004); filed October 6, 2016, published April 6, 2017; prior art under § 102(a)(1) and § 102(a)(2).
- U.S. Patent Application Publication No. 2016/0162456 (“Munro”) (EX1006); 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.

B. Identification of Challenge and Statement of Precise Relief Requested

Petitioner requests cancellation of the Challenged Claims based on the specific grounds set forth below and supported by the declaration of Dr. Janét (EX1003).

Ground	Challenged Claims	Basis	Reference(s)
1	1-19	§ 103	Cox
2	1-19	§ 103	Cox and Ross
3	5, 12, and 19	§ 103	Cox and Ellenbogen
4	1-19	§ 103	Cox and Munro
5	1-19	§ 103	Ellenbogen and Munro

VII. SUMMARY OF THE '046 PATENT

A. Prosecution History

The Office issued a first non-final office action including only a non-statutory double patenting rejection over U.S. Patent 11,126,889, parent of the '046 Patent. EX1002, 93-100. The applicant filed a terminal disclaimer and made minor claim amendments. *Id.*, 115, 122-129. The Office issued a notice of allowance without identifying allowable subject matter. *Id.*, 224-230.

B. Technology Background

The '046 Patent 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”) can 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 reflecting 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 to detect vehicles in new, non-training data. *Id.* The model's detection of vehicles reflects how human annotators would have labeled the new images because the model was trained based on the human annotations. *Id.*

C. The Alleged Invention

The alleged invention of the '046 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, 6:14-7:50. But this alleged invention was well-known. Section X; EX1003, ¶59; *see, e.g.*, EX1012, 13:30-14:9. The *inventor's own prior art* shows that the '046 Patent should not have issued. The Cox reference, which shares a common inventor with the '046 Patent (i.e., Samuel Anthony), discloses machine learning model predictions that are "more consistent with the decisions of the human annotators." EX1007, ¶0032. Neither the patentee nor Mr. Anthony brought the Cox reference to the patent office's attention during prosecution of the '046 Patent. *See* EX1001; EX1002.

The '046 patent explains model training data can be obtained and labeled by human annotators. EX1001, 4:36-61, 5:22-40, 5:52-6:10, Figs. 1, 2B; EX1003, ¶60. This concept was well known previously, as demonstrated by (at least) Cox. EX1007, ¶¶0008-0009, 0029-0032; EX1003, ¶60. The patent 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, 6:14-24. But this simply refers to generating a tally of the annotator responses—e.g., “how many” annotators labeled a training image in a certain way. *Id.* Such tallying was well known. EX1003, ¶60; EX1006, ¶0140. In another example, the patent explains “summary statistics” relate to how long a human annotator took to label training data or the position of the annotator’s eyes when labeling training data. EX1001, 6:24-28. This too was well known and at least Cox demonstrates this. *See, e.g.*, EX1007, ¶¶0006-0008, 0018, 0063; EX1003, ¶60.

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, 6:50-7:21. But Cox discloses the same thing. EX1007, ¶¶0027-0032; EX1003, ¶61. Thus, the '046 Patent describes nothing more than well-known machine learning concepts. EX1003, ¶62.

VIII. LEVEL OF ORDINARY SKILL IN THE ART

A person having ordinary skill in the art (“POSITA”) relevant to the '046 Patent as of July 5, 2017 would have had at least: (1) a bachelor’s degree in electrical

engineering, computer engineering, computer science, or equivalent course work, with three years of work experience in computer vision, autonomous vehicles, and/or machine learning; or (2) a master’s degree in electrical engineering, computer engineering, computer science, or equivalent course work, with a focus in computer vision, autonomous vehicles, and/or machine learning. 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. EX1003, ¶64.

X. THE ASSERTED GROUNDS OF INVALIDITY

A. Ground 1: Claims 1-19 are obvious over Cox

1. Summary of Prior Art

a. Cox–EX1007

Cox, which lists Samuel Anthony as a common inventor with the ’046 Patent, is directed toward machine learning training and implementation techniques. EX1001, Cover; EX1007, Cover, Abstract, ¶¶0001-0004. Crowd-sourcing training data is used to train machine learning algorithms, such as supervised machine learning models. EX1007, Abstract, ¶¶0001-0004, 0015-0016, 0025-0034, 0044, 0060. Here, human annotators annotate training images and Cox’s model is trained

using the annotations. *Id.*, ¶¶0015-19, 0029-0032, Figs. 1A-B, 3. The goal of Cox’s machine learning techniques is to “better mimic human performance.” *Id.*, ¶0004.

Cox is analogous art to the ’046 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., making more accurate model predictions using human-annotated training data). EX1007, ¶¶0029-0032, Figs. 1A-B, 3; EX1001, 1:27-29, 1:65-2:1, 5:52-59; 6:14-55; Fig. 2A; EX1003, ¶68.

2. Claim 1

Cox renders obvious claim 1. EX1003, ¶69.

a. 1[Pre]

To the extent it is limiting, Cox discloses the preamble. EX1003, ¶70. Cox discloses “classification systems” (or “classifiers”) implemented on a server 150, using a computer processor 155 and various program modules located in computer-storage media, where the modules include instructions executed by a computer to implement Cox’s techniques (*[a] computer-implemented method*). EX1007, ¶¶0007-0010, 0017-0025, claim 11, Fig. 1B; EX1003, ¶70.

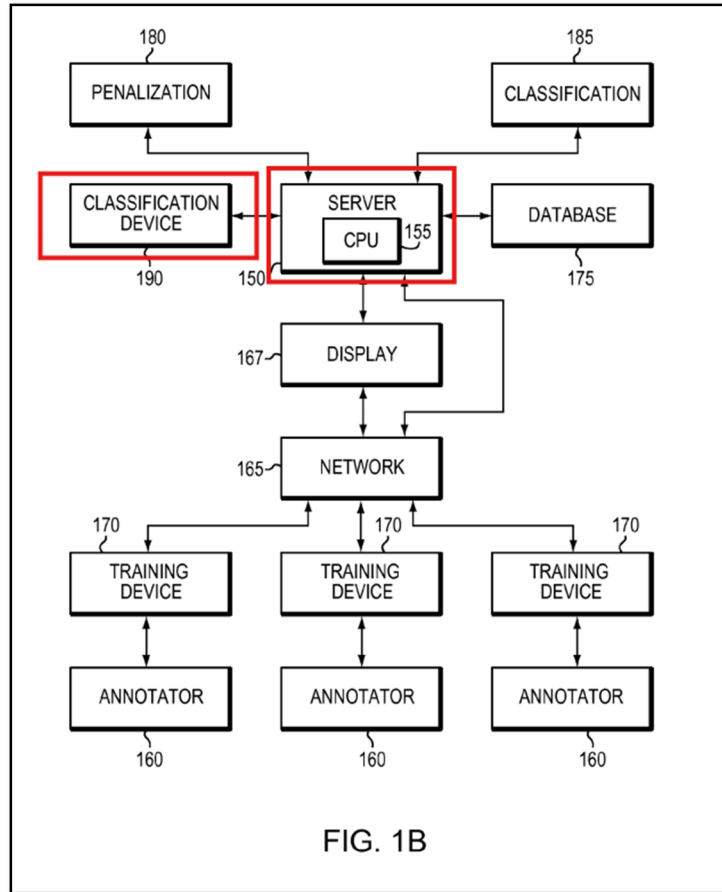


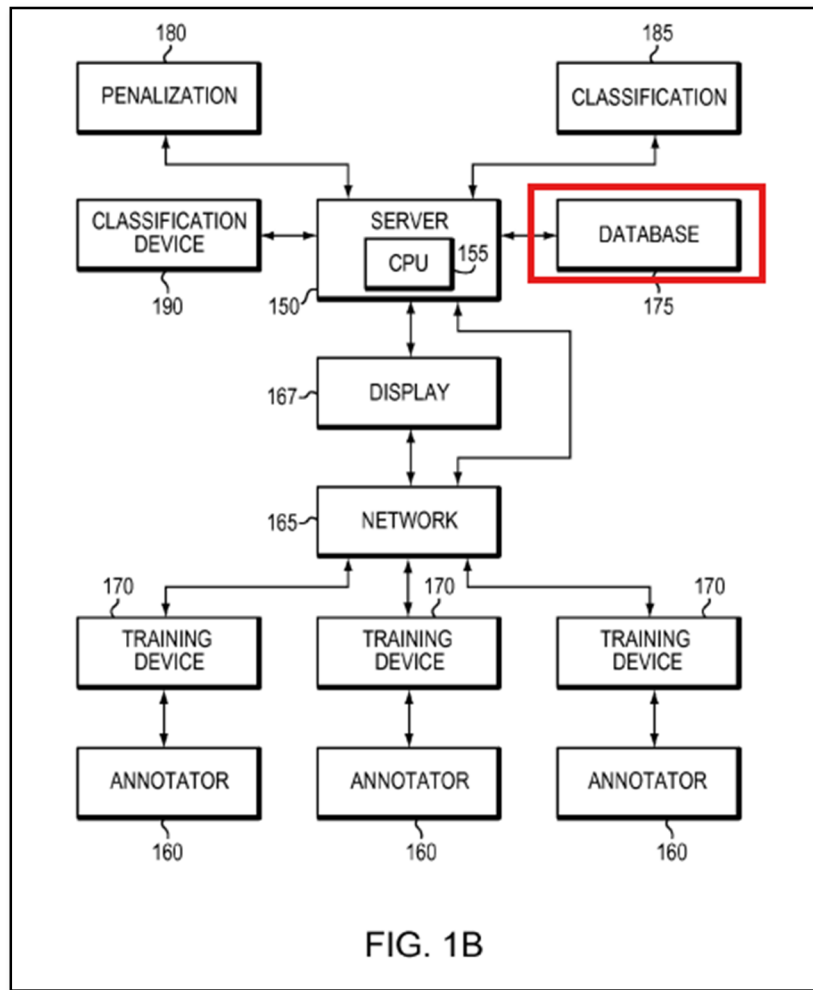
FIG. 1B

EX1007, Fig. 1B (Annotated)

b. 1[a]

Cox teaches element 1[a]. EX1003, ¶71. Cox describes “a database of training objects” where “[e]ach of the training objects may include or consist essentially of a digital image” including a subject person (*storing a plurality of images, each image displaying one or more users*). EX1007, ¶¶ 0007-0010, 0017-0018, 0029; EX1003, ¶71; *see also* EX1001, 3:60-65 (“human observers[] view sample images of people”). The training objects may be images of human faces. EX1007, ¶¶0017, 0029, 0044; EX1003, ¶71. A POSITA would have understood or at least found obvious that the people/faces in the images form *one or more users* as claimed. EX1003, ¶71. This

is at least because the people/faces reflect people who are partaking in an activity or otherwise using an object, for example as a road user within the context of “driverless” vehicles or a manufacturing facility user within the context of “machine vision for manufacturing,” discussed by Cox. EX1007, ¶¶0007-0010, 0017-0018, 0029, 0044, 0056; EX1003, ¶71.



EX1007, Fig. 1B (annotated)

c. 1[b]

Cox teaches element 1[b]. EX1003, ¶72. Cox’s purported novelty is “crowd-sourcing” its training data to “dramatically improve the quality, quantity, and depth

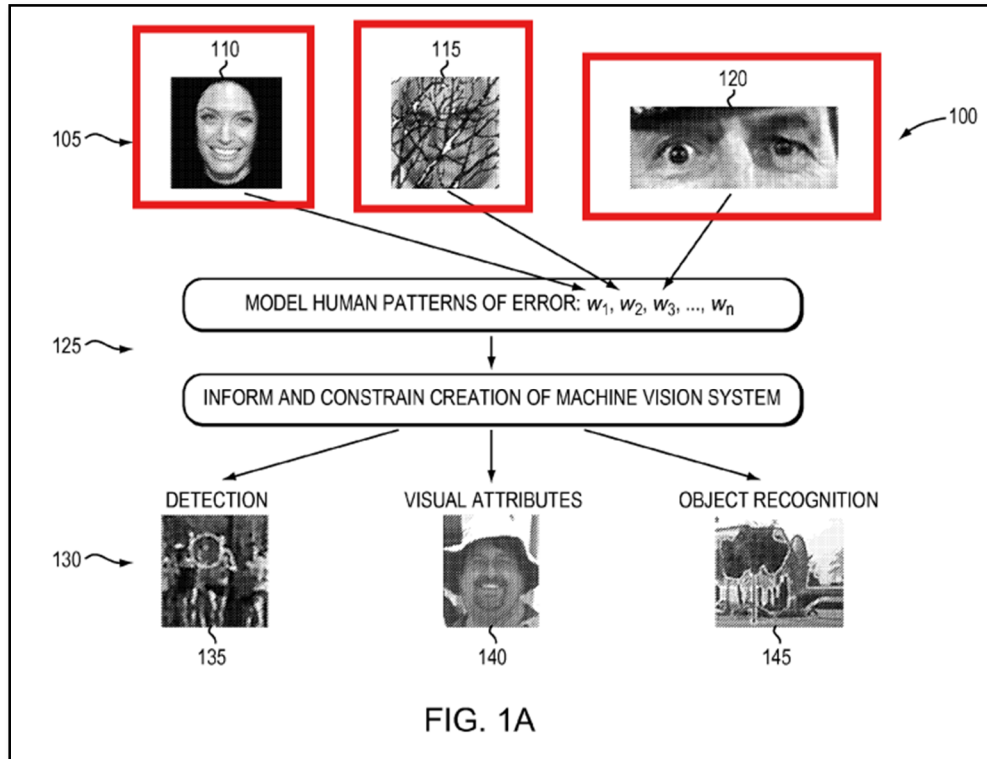
of annotation data available for learning.” EX1007, ¶¶0015, 0017, 0025, 0029, 0032, 0034, 0044, 0051. Cox discloses that multiple (“n”) training images are presented to human annotators, who in turn annotate the training images. EX1007, ¶0029; *see also* ¶¶0005, 0007-0010, 0017-0018, 0029, 0046, Figs. 3 and 4B. This process in which training images are presented to human annotators for annotation generates training data from the images (*generating training data from the plurality of images, the generating comprising, for each image*). *Id.*; EX1003, ¶72.

d. 1[b-i]

Cox teaches element 1[b-i]. EX1003, ¶73. Cox discloses “a crowd-sourced data-acquisition work flow” that involves sending and presenting training images for annotation to multiple human annotators (*[the generating comprising, for each image:] sending the image to a plurality of human observers*). EX1007, ¶¶0017-0018, 0022, 0029; EX1003, ¶73. The images are presented to the human annotators on “training devices 170.” EX1007, ¶¶0010, 0029.

Cox further discloses that when the human annotators are presented with an image, they are asked a question about the image—for example, as to what emotion might correspond to a person shown in the image. EX1007, ¶¶0016 (“**[A] participant may be shown . . . an image 120 of all or a portion of a person’s face and asked** to select an emotion that corresponds to the image, e.g., jealous, panicked, arrogant, or hateful.”), 0018, 0048, Fig. 1A; EX1003, ¶74. When annotators are “asked to select an emotion that corresponds[,]” they are being asked to *answer a*

question about the state of mind of the subject. EX1007, ¶¶0016, 0018, 0048, Fig. 1A; EX1003, ¶74.



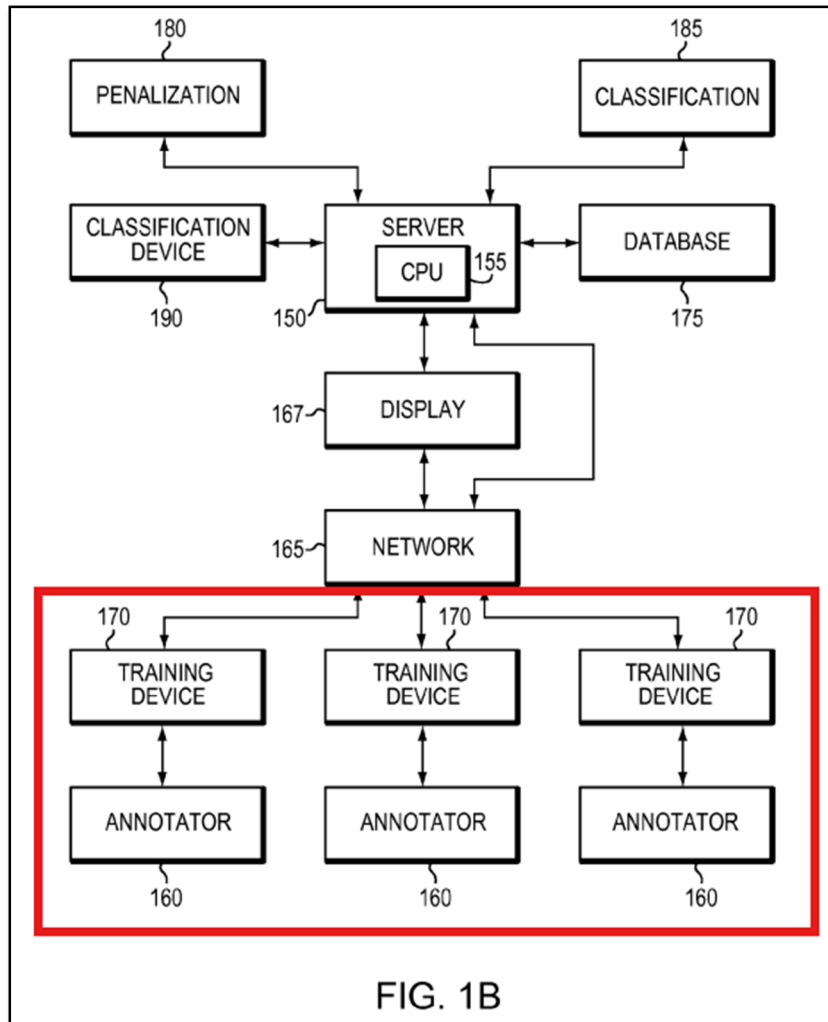
EX1007, Fig. 1A (annotated)

Thus, the annotators (i.e., *human observers*) are presented with an image and asked to answer a question about an emotion of the user in the image (*each human observer presented with a request to answer a question about a state of mind of a user in the image*). EX1007, ¶¶0016, 0018, 0048-0049, 0057, Figs. 1A, 1B; EX1003, ¶75. This disclosure of emotion in Cox is consistent with the '046 Patent's description of *state of mind*, which is “intention, awareness, **personality**, **state of consciousness**, level of tiredness, **aggressiveness**, **enthusiasm**, thoughtfulness or **another characteristic of the internal mental state**” of a user in an image. EX1001, 9:48-54 (emphasis

added); EX1003, ¶75.

e. 1[b-ii]

Cox teaches element 1[b-ii]. EX1003, ¶76. As discussed with respect to element 1[b-i], training images are sent to multiple human annotators for annotation. Section X.A.2.d (element 1[b-i]). The human annotators annotate the images, for example in response to being “asked to select an emotion that corresponds to the image, e.g., jealous, panicked, arrogant, or hateful.” EX1007, ¶¶0016-0021, 0049, 0057, Figs. 1A, 1B *see also id.*, ¶¶0007-0010, 0016-0018, 0029; EX1003, ¶76. The training devices 170 used by the annotators to input annotations on images “transmit classification data” reflecting the annotations “to the central server 150” and the “classification data [is] received from the human annotators” (*receiving, from each of the plurality of human observers, a response representing a judgment by the human observer of the state of mind of the user in the image*). EX1007, ¶¶0016-0021, 0049, 0057, Figs. 1A, 1B *see also id.*, ¶¶0007-0010, 0016-0018, 0029; EX1003, ¶76. The classification data represents *a judgement by the human observer of the state of mind of the user in the image* because the annotations of the classification data are human “judgements,” provided by the human annotators, of an emotion of a person in the image. EX1007, ¶¶0002, 0014, 0050, 0061; EX1003, ¶76. The classification data “may be stored in a database 175 of training objects.” EX1007, ¶0018, Fig. 1B.



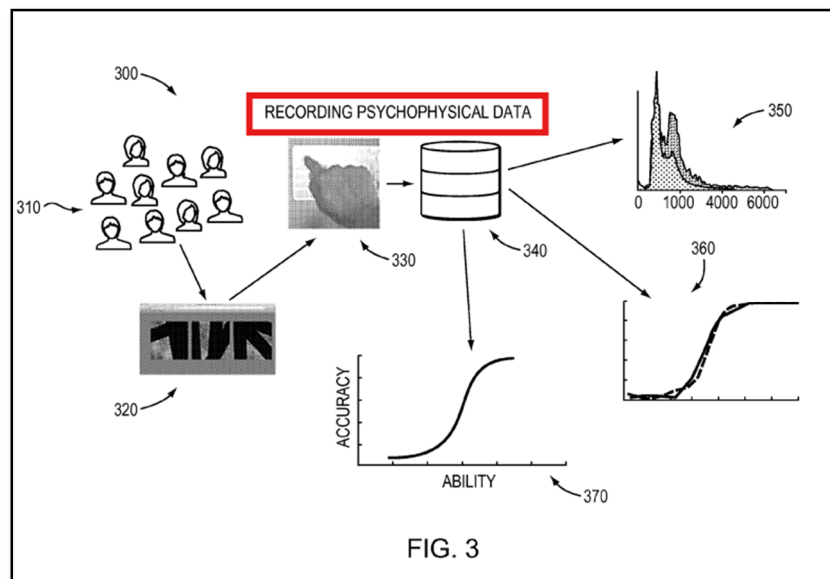
EX1007, Fig. 1B (annotated)

f. 1[b-iii]

Cox teaches element 1[b-iii] in multiple ways. EX1003, ¶77.

As part of Cox’s classification techniques, Cox discloses generating “psychometric data” (*generating summary statistics*) describing the emotions of a user in a sample training image (*describing the state of mind of the user in the image*) based on the annotation responses received from the multiple human annotators (*based on the received responses from the plurality of human observers*). EX1007,

¶¶0002-0008, 0010, 0014, 0018, 0025-0032, 0035, 0049, 0063, Fig. 3; Sections X.A.2.d, X.A.2.e (elements 1[b-i], 1[b-ii]); EX1003, ¶77. Cox discloses that psychometric data summarizes how human annotators annotate training images because it includes “response times for classifying one or more features, the accuracy of feature classification, and/or the presentation time (i.e., the amount of time presented to each annotator) of one or more training objects.” *Id.*, ¶¶0008, 0010, 0018, 0029, Fig. 3. Because these values are a collection of numerical measurements, a POSITA would have recognized that such values form statistics. *Id.*; EX1003, ¶77. And finally, Cox’s description of “psychometric data” is consistent with the ’046 Patent’s explanation; just like Cox, the ’046 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, 6:24-28; EX1003, ¶77.



EX1007, Fig. 3 (annotated)

Cox further provides additional examples of generating *summary statistics* based on annotator responses. EX1003, ¶78. Cox discloses “item-response curves across large populations of humans (e.g., how consistent are judgment[s] across a population),” which Cox describes as a well-known technique associated with human annotator performance. EX1007, ¶0006. Because these response curves present an analysis of collected numerical measurements, a POSITA would have recognized that the curves form statistics regarding the annotations. *Id.*; EX1003, ¶78. Cox thus teaches generating an item response curve (*generating summary statistics*) describing the emotions of people in sample training images based on how human annotators labeled the emotions (*describing the state of mind of the user in the image based on the received responses from the plurality of human observers*). EX1007, ¶¶0002-0008, 0010, 0014, 0018, 0025-0032, 0035, 0049, 0063, Fig. 3; Sections X.A.2.d, X.A.2.e (elements 1[b-i], 1[b-ii]); EX1003, ¶78.

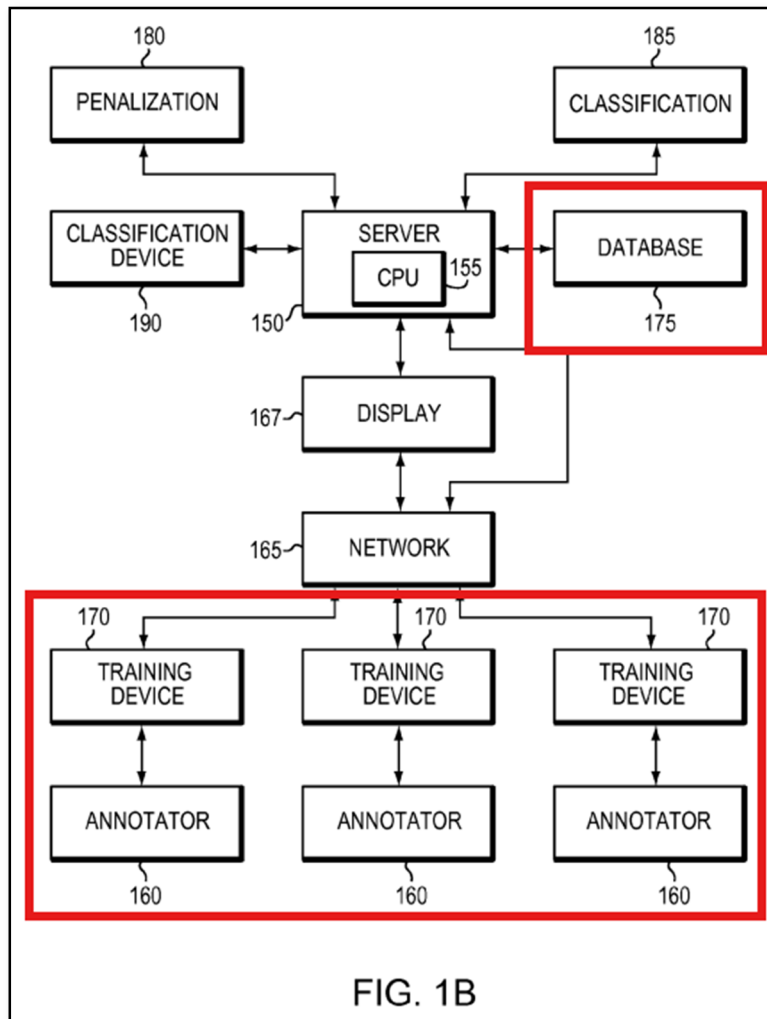
Cox further discloses a “human-weighted loss function based at least in part on the classification data”—i.e., human annotations—“and the psychometric data.” EX1007, ¶0007. Cox also discloses that the responses from human annotators have a distribution of data and labels (e.g., statistics), and that the labels are the responses of “human judgments [that] already provide essential raw material for machine learning, human-generated labels.” *Id.*, ¶¶0002-0006, 0014, 0035. Moreover, Cox discloses that “[a]t the end of a series of such queries, end-of-test statistics 820 may

be displayed to the annotators 160 via display module 167.” *Id.*, ¶0048. A POSITA would have understood from Cox that this data “may be used to inform machine learning in accordance with embodiments of the invention.” *Id.*; EX1003, ¶79. As such, these examples provide further disclosure of how Cox generates summarizing data (*generating summary statistics*) describing the emotions of people in sample training images based on how human annotators labeled the emotions (*describing the state of mind of the user in the image based on the received responses from the plurality of human observers*). EX1007, 0002-0008, 0010, 0014, 0018, 0025-0032, 0035, 0048-0049, 0063, Fig. 3; Sections X.A.2.d, X.A.2.e (elements 1[b-i], 1[b-ii]); EX1003, ¶79.

g. 1[b-iv]

Cox teaches element 1[b-iv]. EX1003, ¶80. Cox’s system includes a “database of training objects,” where the training objects are the images that form training data. EX1007, ¶¶0007-0009, 0017-0020, 0032, Claim 11; EX1003, ¶80. Cox also describes that same database as “populated with stored computer records specifying, for each of the plurality of objects, (i) classification data comprising annotations received from a plurality of human annotators” and “(ii) psychometric data characterizing the annotation of the training object by the plurality of human annotators” (*storing the summary statistics in association with the image as part of the training data*). *Id.*; EX1007, ¶0009. Cox explains that “[d]uring and/or after the annotation, psychometric data is [] acquired that characterizes the annotation of the

training objects by the annotators 160.” EX1007, ¶¶0018, 0029, Fig. 3. “The classification and psychometric data may be stored in a database.” *Id.*; *see also id.*, ¶0029 (“The psychometric data acquired during step 330, e.g., accuracy of image characterization, response time, presentation time, etc., is recorded in the database 175 in a step 340.”).

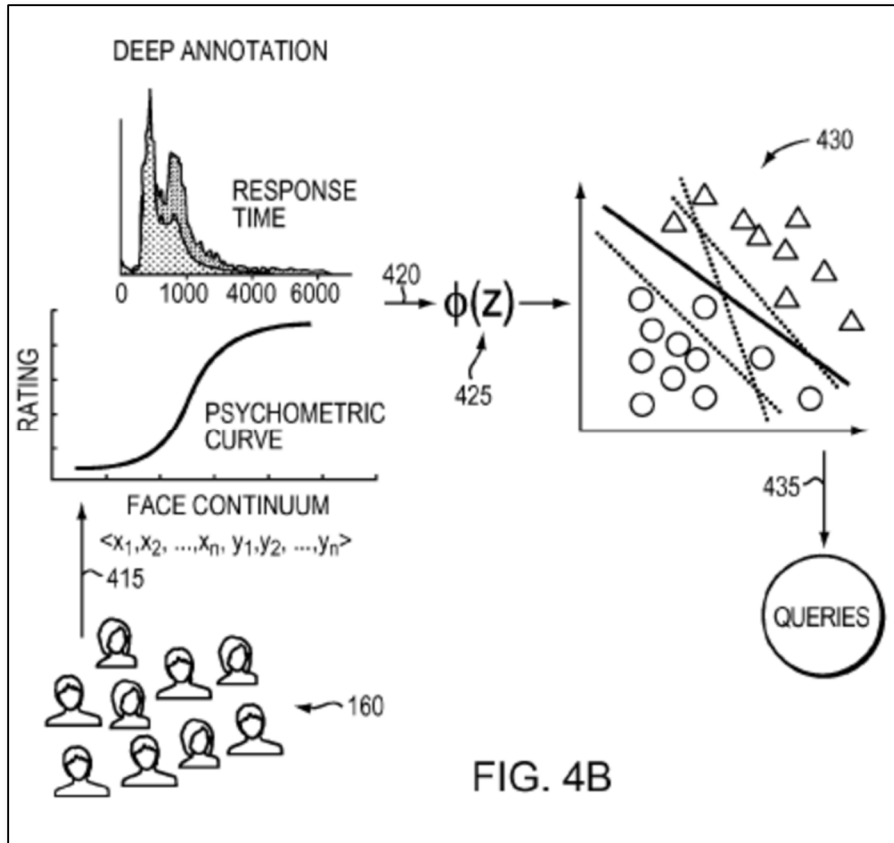


EX1007, Fig. 1B (annotated)

h. 1[c]

Cox teaches element 1[c]. EX1003, ¶81. Cox discloses that its model is trained

by (1) receiving annotated training images showing emotional states, (2) predicting classifications for the training images using the model, and (3) optimizing the model such that differences between the predicted classifications and annotated training images are minimized (*training a model using the training data, the model configured to receive an input image showing a user and predict summary statistics describing a state of mind of the user in the input image*). *Id.*; EX1007, ¶¶0016-0020, 0029-0037, 0049, 0057, 0060; Fig. 4B (430). Cox discloses that “statistical information [] may be collected and applied to the training of computer-vision systems in accordance with embodiments of the invention.” EX1007, ¶¶0027, 0046. The “classification systems for any visual category may be trained with deeply annotated images, by following the learning procedure with human-weighted loss.” EX1007, ¶0051; *see also, e.g.*, ¶¶0019, 0029-0032. Cox’s disclosure is consistent with the ’046 Patent, which states that “trained” means “that the difference between the summary statistics output by the neural network and the summary statistics calculated from the responses of the human observers in step 506 is minimized.” EX1001, 12:30-35. This is just like Cox’s training approach, as described below and shown in Figure 4B. EX1007, ¶0031, Fig. 4B; EX1003, ¶81



EX1007, Fig. 4B

In Cox, annotation responses and psychometric data from the human annotators are utilized to form “a human-weighted loss function” that “includes penalties for misclassification of later presented query objects, as graphically illustrated in graph 430[.]” EX1007, ¶0032. The magnitude of the penalties notably “increases with increasing deviation from the classification data received from the human annotators 160.” *Id.*, ¶¶0032, 0034; *see also id.*, ¶¶0008, 0019, 0029-0037, 0060; Fig. 4B (430). These penalties ensure that the trained model is “more consistent with the decisions of the human annotators.” *Id.*, ¶0032. Cox further teaches that the “penalties for misclassification may be assigned based at least in

part on the psychometric data.” *Id.*, ¶0008. By introducing penalties based on differences from human annotations, Cox optimizes its model to minimize such differences. EX1007, ¶¶0008, 0019, 0029-0037, 0060; Fig. 4B; EX1003, ¶82.

Accordingly, Cox discloses that its model is trained by predicting classifications of some of the received training objects (images) showing emotional states and optimizing the model such that the difference between the predicted classifications and the training data is minimized (*training a model using the training data, the model configured to receive an input image showing a user and predict summary statistics describing a state of mind of the user in the input image*). EX1007, ¶¶0016-0020, 0029-0037, 0049, 0057, 0060; Fig. 4B (430). EX1003, ¶83.

i. 1[d]

Cox teaches element 1[d]. EX1003, ¶84. Cox discloses training its predictive model based on classifications of training images provided by human annotators and related psychometric data, teaching *the trained model*. See *supra*, Section X.A.2.h (element 1[c]). Cox’s trained model is used to predict the emotional state of a subject in a new, non-training image (*executing the trained model to predict a state of mind of a user in a new image*). EX1007, ¶¶0016, 0020, 0032, Fig. 5; EX1003, ¶84.

After training, Cox executes its trained model to predict the emotional state of subjects in new images. EX1007, ¶¶0020, 0032; EX1003, ¶85. Cox describes that “[o]nce the human-weighted loss function is determined, one or more ‘query objects’

may be received by the system for classification.” EX1007, ¶0020. The received query objects are “new objects to be classified by the system absent direct human classification.” *Id.* Cox also expressly discloses that, in a final step after training the model, “the classification module 185 is utilized to make predictions (based on various query objects) that are more consistent with the decisions of the human annotators 160.” *Id.*, ¶0032.

As explained above with respect to 1[b] and 1[c], Cox’s methods can be applied to any form of supervised learning, including a form of supervised learning that is intended to predict an emotion of a user in an image. EX1007, ¶¶0016, 0032, 0060, Fig. 5. Cox states a trained “classification module 185 is utilized to make predictions (based on various query objects),” where the query objects can be, for example, “an image 120 of all or a portion of a person’s face,” and the prediction can be “select[ing] an emotion that corresponds to the image, e.g., jealous, panicked, arrogant, or hateful.” EX1007, ¶¶0016, 0032, Fig. 5; EX1003, ¶86.

3. Claim 2

Cox renders obvious claim 2. EX1003, ¶87. Cox describes that training (also referred to as “stimuli”) images that are “presented to observers” for labeling can be “degraded by a number of techniques” that involve adjusting pixel values, including “contrast lowering, brightness lowering, false color, inversion, image scrambling,” “blur,” and “outline drawings,” as well as occlusion such as “thin occluders[] transposed and normalized for a 100% scale” (*the image is manipulated by adjusting*

values of pixels of the image before presenting to a human observer). EX1007, ¶¶0046, 0052, 0055, 0064, ¶¶0005-0006, 0014-0016, 0044; EX1003, ¶87. Cox’s disclosure aligns with the ’046 patent’s statement that “frames can be manipulated by adjusting pixel values” including by blurring, adding occluding bars, bands, and shapes, and removing or changing color information. EX1001, 8:39-54; EX1003, ¶87.

4. Claim 3

Cox renders obvious claim 3. EX1003, ¶88. The emotions in Cox are each a *state of mind* EX1003, ¶88; Section X.A.2.d (element 1[b-i]). Moreover, a POSITA would have understood that Cox’s emotions *indicates whether a user is likely to perform a predetermined action*. EX1003, ¶88. This is because, for example, a POSITA would have recognized that a “panicked” user would be more likely to flee from perceived danger, and a “hateful” user would be more likely to engage in a fight. *Id.*

5. Claim 4

Cox renders obvious claim 4. EX1003, ¶89. The emotions in Cox are each a *state of mind*. EX1003, ¶89; Section X.A.2.d (element 1[b-i]). A POSITA would have understood that Cox’s emotions *represents a measure of awareness of the user regarding an object*. EX1003, ¶89. For example, a POSITA would have recognized that a user being “panicked” (*state of mind of the user in the image*) would *represent[] a measure of awareness of the user regarding a source of danger (an*

object), and a user being “jealous” (*state of mind of the user in the image*) would represent[] a measure of awareness of the user regarding of a rival (*an object*). *Id.*

6. Claim 5

Cox renders obvious claim 5. EX1003, ¶90. It would have been obvious to a POSITA that the human annotator response in Cox would include a *rating on an ordinal scale*. EX1007, ¶¶0005, 0027-0029; EX1003, ¶90. Cox teaches that training considers the relative difficulty of image data. EX1007, ¶¶0005, 0027-0029. A POSITA would have understood from this disclosure that an annotation response would have included information about the relative difficulty associated with an image. *Id.*; EX1003, ¶90. Because the difficulty information is relative, this information would have been on an ordinal scale (*wherein the response from a human observer comprises a rating on an ordinal scale*). *Id.*

7. Claim 6

Cox renders obvious claim 6. EX1003, ¶91. Cox’s predictive classification model “may be applied to any form of supervised learning, including neural networks, boosting, bagging, random forests, nearest neighbor algorithms, naive bays classifiers, density estimators, and other forms of statistical regression” and “may also be applied as part of a supervised component of semi-supervised or deep-learning algorithms” (*the 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*).

EX1007, ¶0060; EX1003, ¶91.

8. Claim 7

Cox renders obvious claim 7. EX1003, ¶92. Cox discloses psychometric data that forms *summary statistics*. Section X.A.2.f (element 1[b-iii]). The psychometric data includes annotation response times, annotation accuracy, and the amount of time that a training image was presented to an annotator (*the summary statistics is associated with at least one of a content of a response, a time associated with entering a response*). EX1007, ¶¶0017-0018, 0029; EX1003, ¶92. Cox further describes that eye tracking hardware can be used to measure human annotator “saccade-to-target accuracy, saccade-to-target latency, number of saccade hops to target, and total number of saccades,” which would have been measured with respect to the display screen displaying the training image that the annotator is viewing (*and a position of an eye of a human observer associated with the response, the position being measured with respect to a display associated with the image*). EX1007, ¶¶0017, 0063; EX1003, ¶92. The training images of Cox are displayed to annotators on a display of training device 170, which may be computer or cell phone with a display. *Id.*

9. Claim 8

a. 8[Pre]

See Section X.A.2.a (element 1[Pre]). Cox’s system is implemented on a server that includes a computer processor and “utilizes various program modules”

including “computer-executable instructions” (*[a] non-transitory computer readable storage medium storing instructions that when executed by one or more processors, cause the one or more processors to perform steps*). EX1007, ¶¶0007-0010, 0017-0025, claim 11; EX1003, ¶93.

b. 8[a]

See Section X.A.2.b (element 1[a]).

c. 8[b]

See Section X.A.2.c (element 1[b]).

d. 8[b-i]

See Section X.A.2.d (element 1[b-i]).

e. 8[b-ii]

See Section X.A.2.e (element 1[b-ii]).

f. 8[b-iii]

See Section X.A.2.f (element 1[b-iii]).

g. 8[b-iv]

See Section X.A.2.g (element 1[b-iv]).

h. 8[c]

See Section X.A.2.h (element 1[c]).

i. 8[d]

See Section X.A.2.i (element 1[d]).

10. Claim 9

See Section X.A.3 (claim 2).

11. Claim 10

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

12. Claim 11

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

13. Claim 12

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

14. Claim 13

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

15. Claim 14

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

16. Claim 15

a. 15[Pre]

See Section X.A.2.a (element 1[Pre]); Section X.A.9.a (element 8[Pre]).

b. 15[a]

See Section X.A.2.b (element 1[a]).

c. 15[b]

See Section X.A.2.c (element 1[b]).

d. 15[b-i]

See Section X.A.2.d (element 1[b-i]).

e. 15[b-ii]

See Section X.A.2.e (element 1[b-ii]).

f. 15[b-iii]

See Section X.A.2.f (element 1[b-iii]).

g. 15[b-iv]

See Section X.A.2.g (element 1[b-iv]).

h. 15[c]

See Section X.A.2.h (element 1[c]).

i. 15[d]

See Section X.A.2.i (element 1[d]).

17. Claim 16

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

18. Claim 17

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

19. Claim 18

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

20. Claim 19

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

B. Ground 2: Claims 1-19 are obvious over Cox and Ross

Cox alone renders obvious the Challenged Claims. However, Cox is further combined with Ross regarding the following limitations in Ground 2: (1) *sending the image to a plurality of human observers, each human observer presented with a request to answer a question about a state of mind of a user in the image*; (2) *receiving, from each of the plurality of human observers, a response representing a judgment by the human observer of the state of mind of the user in the image*; (3) *training a model using the training data, the model configured to receive an input image showing a user and predicting summary statistics describing a state of mind*

of the user in the input image; and (4) executing the trained model the predict a state of mind of a 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 Cox. EX1003, ¶121.

1. Summary of Prior Art

a. Ross–EX1005

Ross discloses techniques for “maneuvering” an autonomous vehicle based on prediction model outputs. EX1005, 1:30-31, 3:39-42, 3:64-5:5. Human influenced training data is generated and used 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. EX1005, 1:30-2:53, 4:33-5:11, 9:13-10:10, 15:1-16:52. “Human operators” manually generate a “list of predetermined actions” based on their “observations of the actions of other road users” from sensor data generated by a vehicle 100, and a human observed action of the list that is accurately predicted is “marked as the correct intent” during training. *Id.*, 9:41-10:10; EX1003, ¶122.

The trained intent model is used by the autonomous vehicle in operation. Cameras of the vehicle capture “raw” camera data around the vehicle displaying objects such as other vehicles, pedestrians, and bicyclists. EX1005, 7:57-8:10, 9:13-10:10, 15:1-16:52. The trained model is executed on the camera data to predict a “set of possible intents or hypotheses identifying possible next actions (and predicted

trajectories), a given point in time when the action is likely to occur, and associated likelihood values” with respect to each object. *Id.*, 3:64-4:46, 7:57-8:10, 9:13-10:10, 15:1-16:52 (parenthetical in original). The set of likelihood values, with each value associated with a different prediction, “indicate[s] which of the predictions are more likely to occur (relative to one another).” *Id.* The autonomous vehicle of Ross is controlled based on the predictions. *Id.*

Ross is analogous art to the ’046 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., making more accurate model predictions using human-annotated training data). EX1005, Abstract, 3:64-5:1; 9:27-47; EX1001, 1:27-29, 1:65-2:1, 5:52-59; 6:14-55; EX1003, ¶124.

2. Motivation to Combine Cox and Ross

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

- Cox’s system transmits a training image to multiple human annotators for annotation with a prompt to answer about an action by a person or road user in the image reflecting a possible intent of the person, as taught by Ross; and
- Cox’s system is trained to analyze an action by a person or road user, reflecting a possible intent in images, as taught by Ross.

EX1007, ¶¶0010, 0016-0018, 0022, 0028-0032, 0048-0049, 0057, Figs. 1A, 1B; EX1005., 1:55-2:13, 4:33-46, 9:13-10:10, 15:1-16:52; EX1003, ¶125.

A POSITA would have been motivated to combine Cox and Ross in this

manner for several reasons. EX1003, ¶126.

A POSITA would have been motivated to combine Cox and Ross because Ross teaches a known application (driving) for Cox’s system. EX1003, ¶127. Cox discloses a machine learning prediction system that relies on human annotators to annotate training images regarding emotional states of a person in the images. EX1007, ¶¶0007-0010, 0015-0019, 0025-0035, 0051, 0060.; EX1003, ¶127. Ross provides a well-known application for such a system by teaching that the subjects in the images are road users (e.g., vehicles and their associated drivers, pedestrians, bicyclists) performing an action on the road. EX1007, 0010, 0016-0018, 0022, 0028-0032, 0048-0049, 0057, Figs. 1A, 1B; EX1005., 1:55-2:13, 4:33-46, 9:13-10:10, 15:1-16:52; EX1003, ¶127. The combination with Ross thus provides an advantageous application for the system in Cox. *Id.*

A POSITA would have further been motivated to combine Cox and Ross because they both represent improvements in the same field of endeavor, e.g., training models to make human-like predictions. EX1003, ¶128. Cox’s model is utilized after training “to make predictions (based on various query objects) that are more consistent with the decisions of the human annotators 160.” EX1007, ¶¶0027-0032 (parenthetical in original), 0051. The goal of Cox’s machine learning techniques is to “better mimic human performance.” EX1007, ¶0004. Similar to Cox, the stated goal in Ross is to provide an autonomous vehicle that “function[s] in a more ‘human-like’ or ‘polite’ way.” EX1005, 5:7-12. Ross similarly has applications

to autonomous vehicles, disclosing systems and methods for “behavior and intent estimations of road users for autonomous vehicles.” EX1005, Title, Abstract, 1:30-54. To determine the behavior and intent, Ross discloses generating training data and using the training data to train an intent model. EX1005, 3:39-51; 3:64-5:1, 7:57-8:10, 9:13-10:10, 15:1-16:52. The intent model is trained using training data that includes human observations of “road user” actions. *Id.*, 1:55-2:13, 4:33-46, 9:13-10:10, 15:1-16:52. Such actions reflect the “possible intent[s]” associated with an object displayed in the sensor data. *Id.* As demonstrated above, both Cox and Ross explicitly teach that their disclosures aim to provide more human-like model predictions and performance, which would have led a POSITA to combine their teachings. EX1003, ¶128.

A POSITA would have been further motivated to combine Cox and Ross at least because the references each provide explicit teachings, suggestions, or motivations for making the combination. EX1003, ¶129. Cox itself expressly discloses that its teaching may be “directly applied to several important domains where machine learning is found” and expressly suggests applying the invention to “driverless” vehicle applications. EX1007, ¶0056. And Ross provides that an autonomous vehicle “function[s] in a more ‘human-like’ or ‘polite’ way” (EX1005, 5:7-12), which is just like what Cox aims to achieve as noted above, which is to be “more consistent with the decisions of the human annotators” who annotated the training data used to train the Cox model and to “better mimic human performance.”

EX1007, ¶¶0004, 0007-0010, 0015-0019, 0025-0035, 0051, 0060; EX1003, ¶129. Accordingly, Cox and Ross each provide their own teachings, suggestions, or motivations to make the combination, and a POSITA would have thus combined the references. *Id.*

All the reasons explained above establish that a POSITA would have been motivated to combine Cox and Ross. Ross provides driving-specific implementation details to achieve one of the specific applications envisioned in Cox: “driverless [...] automobiles.” EX1007, ¶0056. This combination would have been the predictable combination of well-known prior art elements (e.g., Cox’s training images for annotation and prediction model; Ross’s disclosure of observed road user actions demonstrating intents and prediction model predicting road user intents) according to known methods to yield predictable results (Cox’s system transmits a training image to multiple human annotators for annotation with a prompt to answer about an action by a person in the image reflecting a possible intent of the person, as taught by Ross; Cox’s system trained to analyze an action by a person reflecting a possible intent in images, as taught by Ross). *KSR Int’l Co. v. Teleflex Inc.*, 550 U.S. 398, 415-421 (2007); EX1003, ¶130. The combination would have further used known techniques (Ross’s known techniques regarding human observations on road user actions) to improve similar devices (Cox’s prediction system) in the same way (using Ross’s aforementioned techniques). *Id.* Lastly, in addition to the above, because Ross uses human-labeled data to train its model, combining Ross with Cox’s

teachings about generating and using the same 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 1

Cox and Ross render obvious claim 1. EX1003, ¶131.

a. 1[Pre], 1[a], 1[b]

Cox and Ross teach elements 1[Pre], 1[a] and 1[b] for the same reasons discussed in Sections X.A.2.a (element 1[Pre]), X.A.2.b (element 1[a]), and X.A.2.c (element 1[b]). EX1003, ¶132.

Moreover, with respect to element 1[a], the Cox-Ross combination teaches displaying road users such as vehicles (including their associated drivers), pedestrians, and bicyclists in images, teaching *each image displaying one or more users*. EX1005, 5:6-11, 9:13-10:10, 15:1-16:52; Sections X.B.2, X.A.2.b (element 1[a]).

b. 1[b-i]

Cox and Ross teach element 1[b-i]. EX1003, ¶133. Cox discloses transmitting a training image to multiple human annotators for annotation with a prompt to answer about the image (*[the generating comprising, for each image:] sending the image to a plurality of human observers, each human observer presented with a request to answer a question*). See Section X.A.2.d (element 1[b-i]). Cox's prompt relates to answering a question about emotion of a person in a sample image. *Id.* To

the extent the Patent Owner argues, or the Board finds, that such a prompt about the emotion of a person does not relate to a *state of mind* as claimed, the combination of Cox and Ross teaches this concept. EX1003, ¶133.

Ross discloses that its “intent model” is trained 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” using human observation responses to training sensor data (e.g., camera images) indicating possible next actions of an object displayed in the sensor data, such as another road user (e.g., a vehicle, pedestrian, bicyclist). EX1005, 1:55-2:13, 4:33-46, 9:13-10:10, 15:1-16:52. Such actions reflect the “possible intent[s]” associated with the other road users displayed in the sensor data. *Id.* As such, combining Cox with Ross would have provided that Cox’s training image would have included an object exhibiting an action reflecting a possible intent, and Cox’s human annotator responses would have identified such possible actions associated with a possible intent. *Id.*, EX1007, ¶¶0010, 0016-0018, 0022, 0028-0032, 0048-0049, 0057, Figs. 1A, 1B; EX1003, ¶134. As taught by Ross, the possible intents include, for example, for an object vehicle or object bicycle, the intent of a vehicle driver or bicyclist riding the bicycle to turn the vehicle/bicycle, change lanes, drive through an intersection, or cross and move into the path of a roadway in which the autonomous vehicle is traveling. EX1005, 2:4-13, 9:13-10:10, 15:1-16:52; EX1003, ¶134.

Thus, Ross teaches that Cox's system transmits a training image to multiple human annotators for annotation with a prompt to answer about an action by a person or road user in the image reflecting a possible intent of the person (*[the generating comprising, for each image:] sending the image to a plurality of human observers, each human observer presented with a request to answer a question about a state of mind of a user in the image*). EX1007, ¶¶0010, 0016-0018, 0022, 0028-0032, 0048-0049, 0057, Figs. 1A, 1B; EX1005, 1:55-2:13, 4:33-46, 9:13-10:10, 15:1-16:52; EX1003, ¶135; Section X.A.2.d (element 1[b-i]).

c. 1[b-ii] – 1[b-iv]

As discussed above, Cox and Ross teach deploying the Cox-Ross model that collects training data from a plurality of human annotators by presenting them an image with a user and requesting a response to answer a question about a state of mind of the user in the image. *See supra* Sections X.B.3.b (element 1[b-i]), X.A.2.d.

Cox and Ross teach elements 1[b-ii] – 1[b-iv] for at least the same reasons discussed in Sections X.A.2.e (element 1[b-ii]), X.A.2.f (element 1[b-iii]), and X.A.2.g (element 1[b-iv]). EX1003, ¶137.

d. 1[c]

Cox and Ross teach element 1[c]. EX1003, ¶138. As combined, the training of Cox's model would have been performed as discussed in Section X.A.2.h (element 1[c] – Ground 1) except the training data includes training images and annotations with respect to actions by a person in images reflecting possible intents

of the person, and Cox’s model is trained on such training data as discussed in Section X.A.2.h (element 1[c] – Ground 1). *See* Sections X.B.2, X.B.3.b, X.B.3.c, X.A.2.h.

e. 1[d]

Cox and Ross teach element 1[d]. EX1003, ¶139. As combined, Cox’s model would have operated on new images to determine the actions and possible intents of people in images. *See* Sections X.B.2, X.A.2.i (element 1[d] – Ground 1).

4. Claim 3

The Cox-Ross combination renders obvious claim 3. EX1003, ¶140. Cox and Ross teach an action by a person in the image reflecting a possible intent of the person (*state of mind of the user in the image*). Section X.B.3.b (element 1[b-i]).

Ross’s intent prediction model 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” from an input camera image of an object, such as a vehicle or pedestrian. EX1005, 1:55-2:13, 4:33-46, 9:13-40, 15:1-21. The possible next actions reflect actions (e.g., crossing a road, turning on a road) found in a “predetermined list of actions” that are “generated manually, for instance, by human operators[.]” *Id.*, 4:47-65, 9:41-10:10. Thus, as applied to Cox, Ross teaches that the action by a person in the image reflecting a possible intent of the person (*state of mind of the user in the image*) would have indicated a likelihood that the person performs a predetermined action (e.g., crossing

a road) from a list of predetermined actions (*wherein the state of mind of the user in the image indicates whether the user is likely to perform a predetermined action*), as taught by Ross. *Id.*; EX1003, ¶141.; *supra* Sections X.B.3.b (element 1[b-i]), X.B.3.c (element 1[b-ii]). A POSITA would have been motivated to combine Cox and Ross for the reasons discussed in Section X.B.2.

5. Claim 4

The Cox-Ross combination renders obvious claim 4. EX1003, ¶142. Cox and Ross teach an action by a person in the image reflecting a possible intent of the person (*state of mind of the user in the image*). Section X.B.3.b (element 1[b-i]).

Ross describes an intent prediction model as discussed in Section X.B.3 (claim 1). Ross discloses a predetermined list of actions “may be generated manually, for instance by human operators based on personal experience or observations of the actions of other road users.” EX1005, 9:41-47. Predetermined actions regarding a pedestrian in an image may include an action of a “pedestrian waiting to cross the road” which can correspond to an intent of the pedestrian “to cross the road when clear,” i.e., when there is no traffic and it is safe for the pedestrian to cross. EX1005, 15:50-16:22. As applied to Cox, Ross further teaches that the predicted action of a person in the image, such a prediction that a pedestrian intends “to cross the road when clear,” indicates that the pedestrian is aware of oncoming traffic and will wait to cross until it is safe, i.e., when the traffic has passed. *Id.* Thus, the Cox-Ross combination discloses predicting a measure of awareness of a pedestrian (*state of*

mind of the user in the image represents a measure of awareness of the user regarding an object). *Id.*; EX1003, ¶143. A POSITA would have been motivated to combine Cox and Ross for the reasons discussed in Section X.B.2.

6. Claim 8

See Section X.B.3 (claim 1); *see also* Section X.A.9.a (element 8[Pre], Ground 1)

7. Claim 10

See Section X.B.4 (claim 3)

8. Claim 11

See Section X.B.5 (claim 4)

9. Claim 15

See Section X.B.3 (claim 1); *see also* Section X.A.9.a (element 8[Pre], Ground 1)

10. Claim 17

See Section X.B.4 (claim 3)

11. Claim 18

See Section X.B.5 (claim 4)

12. Claims 2, 5, 6, 7, 9, 12, 13, 14, 16, and 19

Cox and Ross render obvious claims 2, 5, 6, 7, 9, 12, 13, 14, 16, and 19 for the same reasons discussed with respect to Cox in Section X.A above.

C. Ground 3: Claims 5, 12, and 19 are obvious over Cox and Ellenbogen

Cox alone renders obvious the Challenged Claims. However, Cox is further combined with Ellenbogen regarding the following limitation in Ground 3: *the response from a human observer comprises a rating on an ordinal scale*. As discussed below, this limitation is explicitly taught by Ellenbogen, and it would have been obvious to combine Ellenbogen’s relevant teachings with Cox. EX1003, ¶151.

1. Summary of Prior Art

a. Ellenbogen–EX1004

Ellenbogen is generally directed to “improving machine decision making,” and training and using machine learning models. EX1004, Abstract, ¶¶0053, 0057, 0133-0137, 0168, Fig. 34. Images are transmitted to multiple “human agents” who are queried to “answer[] a question regarding a characteristic of the image.” EX1004, ¶¶0005, 0011, 0055-0058, 0062-0074, 0133-0137, 0141, 0149, 0161, 0171, 0180, Fig. 34. The questions can be related to behavior or activity exhibited by a subject person in a received image. *Id.*, ¶¶0011, 0057, 0141, 0161, claim 1, claim 13; *see id.*, Fig. 9.

The agents provide their responses to Ellenbogen’s system, and the responses are aggregated into a “composite output.” EX1004, ¶¶0135, 0149, 0190, Claim 3, Figs. 9, 34. The composite output, as well as the image that was annotated by the human agents itself, are then used to train Ellenbogen’s predictive model. *Id.*, ¶¶0053, 0133-0139, 0162. Ellenbogen’s model can then be executed for a “new

image,” that is not part of training data, in an “operational phase.” *Id.*, ¶¶0057, 0133, 0148, 0168.

Ellenbogen is analogous art to the '046 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., making more accurate model predictions using human-annotated training data). EX1004, Abstract, ¶¶0005, 0042-0048, 0132-0146; EX1001, 1:27-29, 1:65-2:1, 5:52-59; 6:14-55; EX1003, ¶154.

2. Motivation to Combine Cox and Ellenbogen

A POSITA would have been motivated to combine Cox and Ellenbogen. EX1003, ¶155. As discussed above in Section X.A.2 (claim 1), Cox already discloses presenting training images to human annotators for annotations, prompting the annotators to provide annotations, and receiving annotation responses from these annotators. EX1007, ¶¶0016-0021, 0029, 0049, 0057, Figs. 1A, 1B; Section X.A.2 (claim 1). Ellenbogen simply adds to Cox that the annotators are queried to “answer[] a question regarding a characteristic of the image,” and responses are received from the agents that include a “confidence measure” of the answer. EX1004, ¶¶0005, 0007-0012, 0055-0058, 0062-0074, 0133-0137, 0141, 0149, 0161, 0171, 0180, Fig. 34. Thus, Ellenbogen teaches implementation details regarding the format of annotation responses of Cox’s system, and the combination thus would have been a routine enhancement to Cox’s system that would have been easily

performed by POSITAs. *Id.*; EX1003, ¶155. Moreover, a POSITA would have a reasonable expectation of success because the results of the combination would have been predictable: human annotator responses including a “confidence measure” rating. *Id.*

3. Claim 5

Cox and Ellenbogen render obvious claim 5. Cox describes receiving human annotators responses in relation to training images. *See* Sections X.A.2.d-X.A.2.f (elements 1[b-i]-1[b-iii]). Ellenbogen also describes receiving human annotators’ responses in relation to training images. EX1004, ¶¶0005, 0011, 0055-0058, 0062-0073, 0133-0137, 0141, 0161, 0180, Fig. 34. The response from an annotator includes a “confidence measure” of the result, provided by the human agent, that is a rating on a scale of 0 to 1.0 and can be a value such as “0.9” or “0.5,” for example, which are ordinal responses on an ordinal scale. EX1004, ¶¶0007, 0012, 0135, 0149, 0190, 0191, Fig. 34; EX1003, ¶157. As applied to Cox, Ellenbogen thus teaches that Cox’s responses would have included using an ordinal value (e.g., 0.9 or 0.5) on an ordinal scale (e.g., 0 to 1.0) (*wherein the response from a human observer comprises a rating on an ordinal scale*). *See* Sections X.A.2.d-X.A.2.f; EX1004, ¶¶0005, 0007, 0011-0012, 0055-0058, 0062-0073, 0133-0137, 0141, 0149, 0161, 0180, 0190-0191, Fig. 34. EX1003, ¶157. Ellenbogen’s teachings align with the ’046 patent, which provides an example ordinal scale as values such as “1” or “4.” EX1001, 10:34-53, Fig. 6.

4. Claim 12

See Section X.C.3 (claim 5).

5. Claim 19

See Section X.C.3 (claim 5).

D. Ground 4: Claims 1-19 are obvious over Cox and Munro

Cox alone renders obvious the Challenged Claims. However, Cox is further combined with Munro regarding the following limitation in Ground 4: *generating summary statistics describing the state of mind of the user in the image based on the received responses from the plurality of human observers*. As discussed below, these limitations are explicitly taught by Munro, and it would have been obvious to combine Munro’s relevant teachings with Cox. EX1003, ¶160.

1. Summary of Prior Art

a. Munro–EX1006

Munro is directed to techniques for collecting human annotator responses to a stimulus (e.g., a sample document), where the responses are used to train a machine learning model. EX1006, Abstract, ¶¶0005, 0042-0048, 0132-0142. Munro’s system determines summary statistics characterizing annotator responses. *Id.*, ¶¶0140-0142. 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.* Munro’s machine learning model is trained based on the aggregated responses. EX1006, ¶¶0140-0145.

Munro is analogous art to the '046 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., making more accurate model predictions using human-annotated training data). EX1006, Abstract, ¶¶0005, 0042-0048, 0132-0146; EX1001, 1:27-29, 1:65-2:1, 5:52-59; 6:14-55; EX1003, ¶162.

2. Motivation to Combine Cox and Munro

A POSITA would have been motivated to combine Cox and Munro such that Cox's generated statistical data from human annotator responses would have included tallies of the number of annotators who classified the images in a certain way. EX1006, Abstract, ¶¶0005, 0042-0048, 0132-0142; EX1007, ¶¶0002-0006, 0014, 0035, 0048; EX1003, ¶163. A POSITA would have been motivated to combine Cox and Munro in this manner for several reasons. EX1003, ¶163.

A POSITA would have been motivated to combine Cox and Munro because Munro simply provides an example implementation of Cox's statistical data generated from annotator responses. EX1003, ¶164. Cox discloses a system where model training images are transmitted to multiple human agents, the agents are queried to answer a prompt with respect to an image regarding an emotional state of a subject in the image, and responses from the agents with answers are received by the system. EX1007, ¶¶0002-0006, 0014, 0028-0035, 0048; *see* Sections X.A.2.d-X.A.2.f (elements 1[b-i]-1[b-iii]); EX1003, ¶164. Munro is similarly directed to

techniques for collecting human annotator responses to a training sample, where the responses are used to train a machine learning model. EX1006, Abstract, ¶¶0005, 0042-0048, 0132-0142. Munro’s system performs an “aggregation process” for the responses that involves tallying the number of annotators who classified a training sample as having various labels. *Id.*, ¶¶0140-0142.

Thus, Munro merely provides an example implementation of Cox’s statistical data generated from annotator responses, teaching that it would have included tallies of the number of annotators who classified an image as having a given emotional state. EX1006, Abstract, ¶¶0005, 0042-0048, 0132-0142; EX1007, ¶¶0002-0006, 0014, 0035, 0048; *see* Section X.A.2.f (element 1[b-iii]); EX1003, ¶165. A POSITA would have readily enhanced Cox with these teachings of Munro due to the straightforward and routine nature of the combination teachings. *Id.* Including such tallies of human annotator responses as taught by Munro would further improve Cox’s ability to optimize a model that “produce[s] more ‘human-like’ solutions” (EX1007, ¶0005), by recognizing and accounting for trends among the annotator responses. EX1003, ¶165.

Both Cox and Munro are directed to machine learning technologies and involve gathering human annotator responses, providing that a POSITA seeking to implement Cox’s teachings would have looked to Munro. EX1003, ¶166. Munro’s teachings simply provide how Cox’s annotator responses would have been organized—tallying them by label—which would have been nothing more than a

well-known method of organization having a reasonable expectation of success. *Id.* Further, this combination would have at least been the predictable combination of well-known prior art elements (e.g., Cox’s annotator responses; Munro’s organizing of annotator responses) according to known methods to yield predictable results (Cox’s annotator responses organized as taught by Munro). *KSR Int’l Co. v. Teleflex Inc.*, 550 U.S. 398, 415-421 (2007); EX1003, ¶166.

Due to the straightforward application of Munro’s teachings to Cox, the results of the combination would have been predictable. EX1006, Abstract, ¶¶0005, 0042-0048, 0132-0142; EX1007, ¶¶0002-0006, 0014, 0035, 0048; EX1003, ¶167. Because Cox already analyzes and uses collections of human annotator responses, combining Cox with Munro’s teaching to include tallies of such responses would require minimal modifications to Cox. *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 1

a. 1[Pre]-1[b-ii]

Cox and Munro teach elements 1[Pre] through 1[b-ii] for the same reasons discussed in Sections X.A.2.a-X.A.2.e (elements 1[Pre]-1[b-ii]). EX1003, ¶168.

b. 1[b-iii]

Cox and Munro teach element 1[b-iii]. EX1003, ¶169. The combination of Cox and Munro teaches that Cox’s generated statistical data from responses of

human annotators regarding the emotional state of a user in a training image would have included tallies to aggregate the number of annotators who similarly classified the emotional state of the user in the training image (*[the generating comprising, for each image:] summary statistics describing the state of mind of the user in the image based on the received responses from the plurality of human observers*). EX1006, Abstract, ¶¶0005, 0042-0048, 0132-0142; EX1007, ¶¶0002-0006, 0014, 0035, 0048; see Sections X.A.2.f (element 1[b-iii]), X.D.2; EX1003, ¶169. A POSITA would have recognized that these tallies are *summary statistics* because they are a collection of numerical measurements as to how annotators classified emotional states. *Id.* This combined teaching of Cox and Munro further aligns with the '046 Patent's disclosure. EX1003, ¶169. Specifically, the '046 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 observers" classified the stimulus as displaying various content. EX1001, 6:1-11; EX1003, ¶169.

A POSITA would have been motivated to combine Cox and Munro for the reasons discussed in Section X.D.2. EX1003, ¶170.

c. 1[b-iv]

Cox and Munro teach element 1[b-iv] for at least the same reasons discussed in Section X.A.2.g (element 1[b-iv]), except the generated summary statistics as taught by Munro would have been stored in Cox's computer record in a database for

the statistic's associated image. EX1003 ¶171; Sections X.A.2.g (element 1[b-iv]), X.D.3.b (element 1[b-iii]).

d. 1[c]

Cox and Munro teach element 1[c]. EX1003, ¶172. As combined, the training of Cox's model would have been performed as discussed in Section X.A.2.h (element 1[c]) except the training data includes the generated summary statistics as taught by Munro, and Cox's model is trained on such training data in the same manner discussed in Section X.A.2.h (element 1[c]). *See* Sections X.A.2.h (element 1[c]), X.D.3.b (element 1[b-iii]), X.D.3.c.

e. 1[d]

Cox and Munro teach element 1[d] for the same reasons discussed in Section X.A.2.i (element 1[d]). EX1003, ¶173.

4. Claim 8

See Section X.D.3 (claim 1) ; *see also* Section X.A.9.a (element 8[Pre], Ground 1).

5. Claim 15

See Section X.D.3 (claim 1) ; *see also* Section X.A.9.a (element 8[Pre], Ground 1).

6. Claims 2-7, 9-14, and 16-19

Cox and Munro render obvious claims 2-7, 9-14, and 16-19 for the same reasons discussed with respect to Cox in Section X.A above.

E. Ground 5: Claims 1-19 are obvious over Ellenbogen and Munro

Ellenbogen renders teaches most limitations of the Challenged Claims and is further combined with Munro regarding the following limitations: (1) *generating summary statistics describing the state of mind of the user in the image based on the received responses from the plurality of human observers*; (2) *storing the summary statistics in association with images as part of the training data*; (3) *training a model using the training data, the model configured to receive an input image showing a user and predict summary statistics describing a state of mind of the user in the input image*. As discussed below, these limitations are explicitly taught by Munro, and it would have been obvious to combine Munro’s relevant teachings with Ellenbogen. EX1003, ¶177.

1. Motivation to Combine Ellenbogen and Munro

A POSITA would have been motivated to combine Ellenbogen and Munro such that:

- Ellenbogen’s composite output of human agent responses includes tallies of the number of agents who classified the behavior or activity as having various labels in their responses, as taught by Munro (the “Ellenbogen-Munro composite output”) (EX1004, ¶¶0135, 0149, 0190; *id.*, ¶¶0005, 0008, 0011, 0054-0058, 0060-0073, 0108, 0133-0137, 0141, 0149, 0161, 0180, Claim 3, Figs. 9, 34; EX1006, ¶¶0140-0142; EX1003, ¶178);

- The Ellenbogen-Munro composite output is stored in a computer record as training data, as taught by Munro, and in association with underlying images using the computed score taught by Munro (EX1004, ¶¶0133-0137, Fig. 34; EX1006, ¶¶0059-0062, 0141-0142; EX1003, ¶178); and
- Ellenbogen’s predictive model is trained by the Ellenbogen-Munro composite output and image that was annotated, as taught by Munro (EX1006, ¶¶0140-0142; EX1004, ¶¶0057, 0133-0137, 0138-0148, 0166-0168, Fig. 34; EX1003, ¶178).

A POSITA would have been motivated to combine Ellenbogen and Munro in this manner for several reasons. EX1003, ¶179. This combination would have at least been the predictable combination of well-known prior art elements (e.g., Ellenbogen’s annotator responses; Munro’s organizing of annotator responses, training, and storage) according to known methods to yield predictable results (Ellenbogen’s annotator responses being organized, stored, and used to train a model, as taught by Munro). *KSR Int’l Co. v. Teleflex Inc.*, 550 U.S. 398, 415-421 (2007); EX1003, ¶178.

A POSITA would have been motivated to combine Ellenbogen and Munro because Munro simply provides implementation details for Ellenbogen’s system. EX1003, ¶180. Ellenbogen discloses a system where model training images are transmitted to multiple human agents, the agents are queried to “answer[] a question regarding a characteristic of the image,” and responses from the agents with answers are received by the system. EX1004, ¶¶0005, 0011, 0055-0058, 0062-0074, 0133-

0137, 0141, 0149, 0161, 0171, 0180, Fig. 34. Ellenbogen describes that a “composite output” is generated from the responses. *Id.*, ¶¶0135, 0149, 0190.

Like Ellenbogen, 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. EX1006, Abstract, ¶¶0005, 0042-0048, 0059-0062, 0132-0142. As a disclosure for one such technique, Munro’s system performs an “aggregation process” for the responses that involve tallying the number of annotators who classified a training sample as having various labels. *Id.*, ¶¶0140-0142. Munro provides an example in which its system determines how many human annotators classified a training sample as having a first label and how many annotators classified the sample as having a second label. *Id.*; EX1003, ¶181. And Munro provides routine disclosure in how model training is implemented and training data is stored, referring to storing based on computed stores. EX1006, ¶¶0059-0062, 0140-0142.

As such, Munro simply provides implementation details specifying how Ellenbogen’s composite output including annotator responses is formed, stored, and used for training, teaching that it would have included tallies of the number of annotators who classified a behavior or activity in an image as having various labels in their responses (the “Ellenbogen-Munro composite output”). EX1006, Abstract, ¶¶0005, 0042-0048, 0132-0142; EX1004, ¶¶0135, 0149, 0190; *id.*, ¶¶0005, 0008, 0011, 0054-0058, 0060-0073, 0108, 0133-0137, 0141, 0149, 0161, 0180, Claim 3,

Figs. 9, 34; EX1003, ¶182. A POSITA would have readily enhanced Ellenbogen with Munro's teachings due to the routine and straightforward nature of the combined teachings. *Id.* Both Ellenbogen and Munro are directed to machine learning technologies and involve gathering human annotator responses and training a model, providing that a POSITA seeking to implement Ellenbogen's teachings would have looked to Munro. EX1003, ¶182.

A POSITA would have further been motivated to combine Ellenbogen and Munro because Munro would provide performance benefits to Ellenbogen. EX1003, ¶183. For example, Munro's tallying would provide additional data regarding annotator responses that could be used to better train Ellenbogen's prediction model. *Id.*; EX1004, ¶¶0005, 0008, 0011, 0054-0058, 0060-0073, 0108, 0133-0137, 0141, 0149, 0161, 0180, 0190, Claim 3, Figs. 9, 34; EX1006, ¶¶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.* Further, Munro's storage teachings would improved processing data in Ellenbogen because each Ellenbogen-Munro composite output would be stored in association with its underlying image for more efficient access and use. EX1004, ¶¶0057, 0133-0148, 0166-0168, Fig. 34; EX1006, ¶¶0059-0062, 0140-0142; EX1003, ¶183. And, training Ellenbogen's predictive model using the implementation details taught by Munro would have ensured that the model iteratively improves in accuracy. *Id.*

A POSITA would also have a reasonable expectation of success in combining

Munro with Ellenbogen because Munro’s teachings simply provide details about how Ellenbogen’s annotator responses would be organized and used. EX1003, ¶184. Due to the straightforward application of Munro’s teachings to Ellenbogen, the results of the combination would have been predictable. EX1004, ¶¶0135, 0149, 0190; *id.*, ¶¶0005, 0008, 0011, 0054-0058, 0060-0073, 0108, 0133-0137, 0141, 0149, 0161, 0180, Claim 3, Figs. 9, 34; EX1006, Abstract, ¶¶0005, 0042-0048, 0059-0062, 0132-0142; EX1003, ¶184. Implementing these teachings from Munro in Ellenbogen would have been nothing more than a routine and well-known method of organization, training, and storage, which means that a POSITA would have had a reasonable expectation of success in making the combination. *Id.*

2. Claim 1

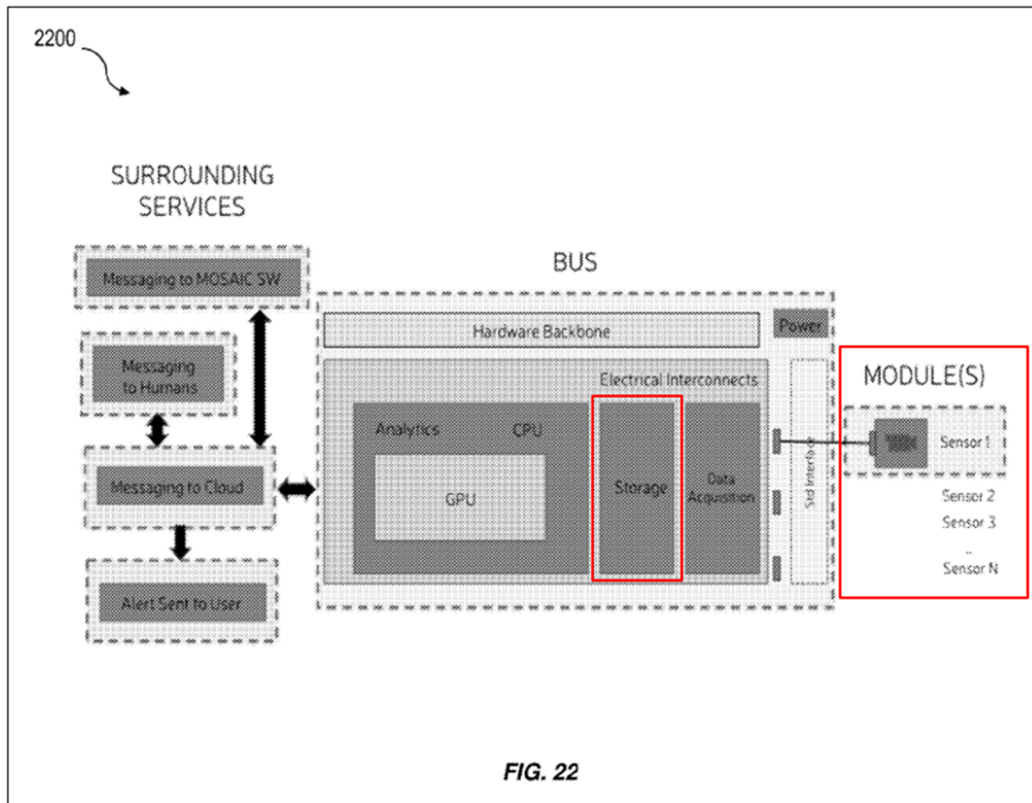
a. 1[Pre]

To the extent it is limiting, Ellenbogen discloses the preamble. EX1003, ¶185. Ellenbogen describes a computerized process for “training a machine computation component” including a machine learning model, and techniques for executing the trained machine learning model (*[a] computer-implemented method*). EX1004, ¶¶0057, 0133-0137, 0168, Fig. 34; EX1003, ¶185.

b. 1[a]

Ellenbogen teaches element 1[a]. Ellenbogen discloses (1) that “sensor data” such as “a series of images” includes a “person” in each of the images partaking in an activity, exhibiting “suspicious behavior,” or expressing a “sentiment,” (2) that

the sensor data is produced by an “imaging device, video camera, [or] still camera,” and (3) that the sensor data is stored in “storage” such as a database (*storing a plurality of images, each image displaying one or more users*). EX1004, ¶¶0010-0011, 0054, 0057, 0133-0137, 0141, 0161, 0226, Figs. 22, 34; EX1003, ¶186.



EX1004, Fig. 22 (annotated)

Moreover, a “database” of images to be labeled by human annotators is provided by Ellenbogen’s system, and the images include people, which further teaches *storing a plurality of images, each image displaying one or more users*. EX1004, ¶¶0054, 0094, 0137, 0141, 0161-0164; EX1003, ¶187.

c. 1[b]

Ellenbogen teaches element 1[b]. EX1003, ¶188. Ellenbogen’s images (*see*

Section X.E.2.b (element 1[a])) are labeled by “human agents” to generate training data for a predictive model (*generating training data from the plurality of images, the generating comprising, for each image*). EX1004, ¶¶0133-0137, Fig. 34; EX1003, ¶188.

d. 1[b-i]

Ellenbogen teaches element 1[b-i]. EX1003, ¶189. An image of the series of images (*see* Section X.E.2.b (element 1[a])) is transmitted to client devices of multiple human agents (*[the generating comprising, for each image:] sending the image to a plurality of human observers*). EX1003, ¶189; EX1004, ¶¶0005, 0011, 0055-0058, 0062-0074, 0133-0137, 0141, 0149, 0161, 0171, 0180, Fig. 34, claim 1, claim 13. Each of the human agents is queried to “answer[] a question regarding a characteristic of the image” such as a behavior of a person in the image or an activity associated with a person in the image (*each human observer presented with a request to answer a question about a state of mind of a user in the image*). *Id.* When Ellenbogen’s system is used to determine the activity or behavior of a person in an image, the “question regarding a characteristic of the image” relates to the *state of mind* of the person in the image because it asks the human agent to answer a question about what the person is doing. *Id.*; EX1004, Fig. 9; EX1003, ¶189. For example, the determined behavior or activity of the person in the image would have included an intention of the person with respect to what they are doing. *Id.* The determined behavior or activity of the person in the image would have further included behavior,

Petition for *Inter Partes* Review of U.S. Patent No. 11,753,046

such as “suspicious” behavior, and/or a “vehicle behavior” like “tailgating” or driving in the “wrong direction.” *Id.* Annotated Figure 9 below provides an example showing that to determine “if people exhibit certain behaviors,” then “human questions” are asked regarding “[s]uspicious behavior,” “[b]ehavior of people on bus,” “[v]ehicle behavior (tailgating, wrong direction),” “[a]ctivity recognition (e.g., person using phone),” and “interactions with retail products/end cap.” *Id.*, Fig. 9; EX1003, ¶189.

Modality	Description	Requirements	Examples
Intrusion	Here is a physical space, region, or boundary. I'm concerned about intrusion into that space.	<ul style="list-style-type: none"> • Tripwire specification • Person/vehicle/object description • Person/vehicle/object particulars • Human questions 	<ul style="list-style-type: none"> • Perimeter security • Asset protection • Door monitoring • Animals intruding • People where they shouldn't be
Access	I'm trying to identify people or classes of people that should either be allowed in or prevented from entering. It's a space, region, or boundary where it is normal for people to go in and out, but only certain people should be allowed access.	<ul style="list-style-type: none"> • Tripwire specification • Person/vehicle/object description • Person/vehicle/object particulars • Watchlist • Human questions 	<ul style="list-style-type: none"> • Watch list • VIP list • Particular delivery vehicle/driver
Loiter/Dwell	People, vehicles, or objects can be there, but if they loiter/dwell that is something I want to know.	<ul style="list-style-type: none"> • Area of interest • Person/vehicle/object description • Person/vehicle/object particulars • Time thresholds • Human questions 	<ul style="list-style-type: none"> • Person dwell • Vehicle dwell • Object dwell (bag left behind)
Behavior	I want to know if people exhibit certain behavior.	<ul style="list-style-type: none"> • Behavior of interest • Region of interest • Human questions 	<ul style="list-style-type: none"> • Suspicious behavior • Behavior of people on bus • Vehicle behavior (tailgating, wrong direction) • Activity recognition (e.g., person used phone) • Interaction with retail products/end cap
Tracking	Once I've identified someone or some vehicle, I want to be able to track it from one camera to the next.	<ul style="list-style-type: none"> • Region of interest • Last location, time, and trajectory • Person/vehicle/object description • Person/vehicle/object particulars • Human questions 	<ul style="list-style-type: none"> • Person tracking • Vehicle tracking

FIG. 9

EX1004, Fig. 9 (annotated)

Ellenbogen’s disclosure corresponds to the ’046 patent’s description of a “state of mind;” for example, the patent explains the “evaluation of the state of mind

of a road user depicted” in an image “can be of the intention,” “state of consciousness,” “aggressiveness,” “enthusiasm,” or “thoughtfulness,” and at least each of these are represented by a “behavior” described by Ellenbogen. EX1001, 9:48-54, 5:65-6:2; EX1003, ¶190. For example, a person partaking in an activity reflects an intention of that person to perform the activity, suspicious behavior at least indicates a “state of consciousness,” and a vehicle that is tailgating at least indicates an “aggressiveness” of the vehicle’s driver, and each of these examples, as well as those also discussed above for this element 1[b-i], at least teaches a *state of mind*. EX1004, Fig. 9; EX1003, ¶190.

e. 1[b-ii]

Ellenbogen teaches element 1[b-ii]. The multiple human agents each respond with a “query result” to the “question regarding a characteristic of the image” that describe a behavior or activity of a person in the image (*[the generating comprising, for each image:] receiving, from each of the plurality of human observers, a response representing a judgment by the human observer of the state of mind of the user in the image*). EX1004, ¶¶0005, 0008, 0011, 0054-0058, 0060-0073, 0108, 0133-0137, 0141, 0149, 0161, 0180, Claim 3, Figs. 9, 34; EX1003, ¶191. Ellenbogen describes that these responses by human agents reflect “judgments” that the agents are making with respect to the image, and the judgments thus reflect the behavior or activity. EX1004, ¶¶0057, 0060, 0180, Fig. 9; *see also* Section X.E.2.d (element 1[b-i]); EX1003, ¶191.

f. 1[b-iii]

Ellenbogen and Munro teach element [1b-iii]. EX1003, ¶192. Ellenbogen describes that a “composite output” is generated from the query result responses regarding the behavior or activity of a person in the image are received from the multiple human agents. EX1004, ¶¶0135, 0149, 0190; *id.*, ¶¶0005, 0008, 0011, 0054-0058, 0060-0073, 0108, 0133-0137, 0141, 0149, 0161, 0180, Claim 3, Figs. 9, 34; Section X.E.2.e (element 1[b-ii]); EX1003, ¶192. 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. EX1006, Abstract, ¶¶0005, 0042-0048, 0132-0142. Munro’s system performs an “aggregation process” for the responses that involves tallying the number of annotators who classified a training sample as having various labels. *Id.*, ¶¶0140-0142. In one example, Munro’s system determines how many human annotators classified a training sample as having a first label and how many annotators classified the sample as having a second label. *Id.*; EX1003, ¶193.

Thus, Munro teaches that Ellenbogen’s composite output of query result responses regarding the behavior or activity of a person in a sample image would have included an aggregate of the number of annotators who classified the behavior or activity as having various labels in their responses (the “Ellenbogen-Munro composite output”) (*the generating comprising, for each image:] summary statistics describing the state of mind of the user in the image based on the received*

responses from the plurality of human observers). EX1006, Abstract, ¶¶0005, 0042-0048, 0132-0142; EX1004, ¶¶0135, 0149, 0190; *id.*, ¶¶0005, 0008, 0011, 0054-0058, 0060-0073, 0108, 0133-0137, 0141, 0149, 0161, 0180, Claim 3, Figs. 9, 34; EX1003, ¶194. 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 Ellenbogen-Munro teaching further aligns with the '046 Patent's disclosure. *Id.* The '046 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 observers” classified the stimulus as displaying various content. EX1001, 6:1-11; EX1003, ¶194.

A POSITA would have been motivated to combine Ellenbogen and Munro for the reasons discussed in Section X.E.1. EX1003, ¶195.

g. 1[b-iv]

Ellenbogen and Munro teach element 1[b-iv]. Ellenbogen and Munro teach the Ellenbogen-Munro composite output (*summary statistics*). Section X.E.2.f (element 1[b-iii]). Ellenbogen describes that the “sensor data” reflecting the image having the person in it that was annotated, as well as the “result from the agent computation component”—which is the composite output—are used to train a “predictive model” and are thus *training data* for the model. EX1004, ¶0136; *id.*, ¶¶0133-0137, Fig. 34; EX1003, ¶196. Thus, Ellenbogen and Munro teach that the

Ellenbogen-Munro composite output (*summary statistics*) and associated image that was annotated are (*the image*) are training data for Ellenbogen's model (*part of the training data*). EX1004, ¶¶0133-0137, Fig. 34; Section X.E.2.f (element 1[b-iii]); EX1003, ¶196.

Munro teaches to store the Ellenbogen-Munro composite output in association with the image as part of the training data. EX1003, ¶197. Specifically, Munro's system includes a "client data store" (i.e., a database of training data) that contains "a repository of data that is used to train and ultimately generate the natural language model." EX1006, ¶0061. This teaching in Munro is also consistent with the disclosure in Ellenbogen that a "database of training images [...] can be updated over time with real world images and labels." EX1004, ¶0164. Thus, in the combination with Ellenbogen, a POSITA would have understood that storing the Ellenbogen-Munro composite in the client data store would comprise "*storing the summary statistics [...] as part of the training data.*" EX1003, ¶197.

Further, Munro teaches that a "computed score" can be determined based on the annotation aggregation process, and may be used to determine whether to use a given training reference during model training. EX1006, ¶0141. In the combination with Ellenbogen, this "computed score" thus associates the Ellenbogen-Munro composite with a training reference (i.e., the *image*), such that the composite is "*stor[ed] [...] in association with the image as part of the training data.*" EX1003, ¶198.

If this limitation is interpreted to require that the *summary statistics* and *image* are stored together, a POSITA would have found it obvious to store the composite together with the training reference in Munro’s client data store. EX1003, ¶199; EX1006, ¶¶0061, 0141; EX1004, ¶¶0133-0137, 0164. This is because storing such information together would have improved processing efficiency and speed, and thus been a routine and well-known design choice. *Id.* EX1003, ¶199. A POSITA would have had a reasonable expectation of success for this arrangement because storing these together would have been simpler and involved fewer parts than storing the composite and images separately. EX1003, ¶199.

h. 1[c]

Ellenbogen and Munro teach element 1[c]. EX1003, ¶200. Ellenbogen describes that the “sensor data” reflecting the image having the person in it that was annotated, as well as the “result from the agent computation component,” that is Ellenbogen’s composite output, is used to train Ellenbogen’s predictive model. EX1004, ¶0136; *id.*, ¶¶0133-0137, Fig. 34. Ellenbogen’s goal is to “progressively train the artificial intelligence to begin to recognize” certain “behavior” in an “environment on its own.” *Id.*, ¶¶0057, 0133.

Munro teaches that its tally of annotator responses is also used to train a predictive model via a “model training process.” EX1006, ¶¶0140-0142. As such, the combination of Ellenbogen and Munro teaches that the Ellenbogen-Munro composite output (*see* Sections X.E.2.f, X.E.2.g (elements 1[b-iii], 1[b-iv])) and the

associated image that was annotated, as well as other instances of such items, are used as training data to train Ellenbogen's predictive model (*training a model using the training data*). *Id.*; EX1004, ¶¶0057, 0133-0137, Fig. 34; EX1006, ¶¶0140-0142; EX1003, ¶201.

Ellenbogen and Munro further teach *the model configured to receive an input image showing a user and predict summary statistics describing a state of mind of the user in the input image*. EX1003, ¶202. Specifically, Ellenbogen teaches that the artificial intelligence model in Ellenbogen can be trained “to begin to recognize suspicious behavior in that environment on its own.” EX1004, ¶0057. Ellenbogen discloses that “[a]s the artificial intelligence component is trained on more real-world data, the artificial intelligence component becomes more accurate and less agent input is required.” EX1004, ¶0061. Ellenbogen further teaches to predict a “confidence” and a score related to the confidence, e.g., *summary statistics*. *Id.*, ¶¶0060-61, ¶0166 (“The object classifier can generate a score from 0.0 to 1.0 related to the confidence that the object is not present (0.0) or present (1.0). The classification can be binary or multiclass.”).

Ellenbogen further teaches that its prediction can be a “multiclass” prediction in which a “score from 0.0 to 1.0 related to the confidence” for a predicted classification is used. EX1004, ¶0166. Multiclass predictions analyze whether a prediction falls within one of multiple classes. *Id.*, ¶¶0166-0168; EX1003, ¶203. Thus, a POSITA would have understood that a multiclass prediction utilizing

confidence scores from 0.0 to 1.0 summarizes the various possible classifications and likelihoods for behavior or activity in the image (*summary statistics describing a state of mind of the user in the input image*). EX1004, ¶¶0133-0137, 0138-0148, 0166-0168, Fig. 34; EX1003, ¶203. Moreover, the predicted classes of behavior or activity and associated confidence scores are *predicted summary statistics* that correspond to the tallied annotator responses of the Ellenbogen-Munro composite output that are the stored *summary statistics* (see Sections X.E.2.g, X.E.2.h, *supra*) at least because they are the predicted categories (and associated likelihoods) of behavior or activity that would have also been represented by human annotators from their labels during training in the Ellenbogen-Munro composite output. EX1003, ¶203.

Thus, in combination with Munro, Ellenbogen's model is trained by receiving input images of a person exhibiting a certain behavior or activity (*the model configured to receive an input image showing a user*) and predicting a multiclass behavior or activity prediction of a person in a training image and associated confidence scores (*and predict summary statistics describing a state of mind of the user in the input image*), and then optimizing the model, for example, such that a difference between the multiclass predicted behavior or activity of a person and the Ellenbogen-Munro composite output reflecting tallied annotation responses is minimized. EX1004, ¶¶0133-0137, 0138-0148, 0166-0168, Fig. 34; EX1006, ¶¶0140-0142; Section X.E.2.f (element 1[b-iii]); EX1003, ¶204. The predicted

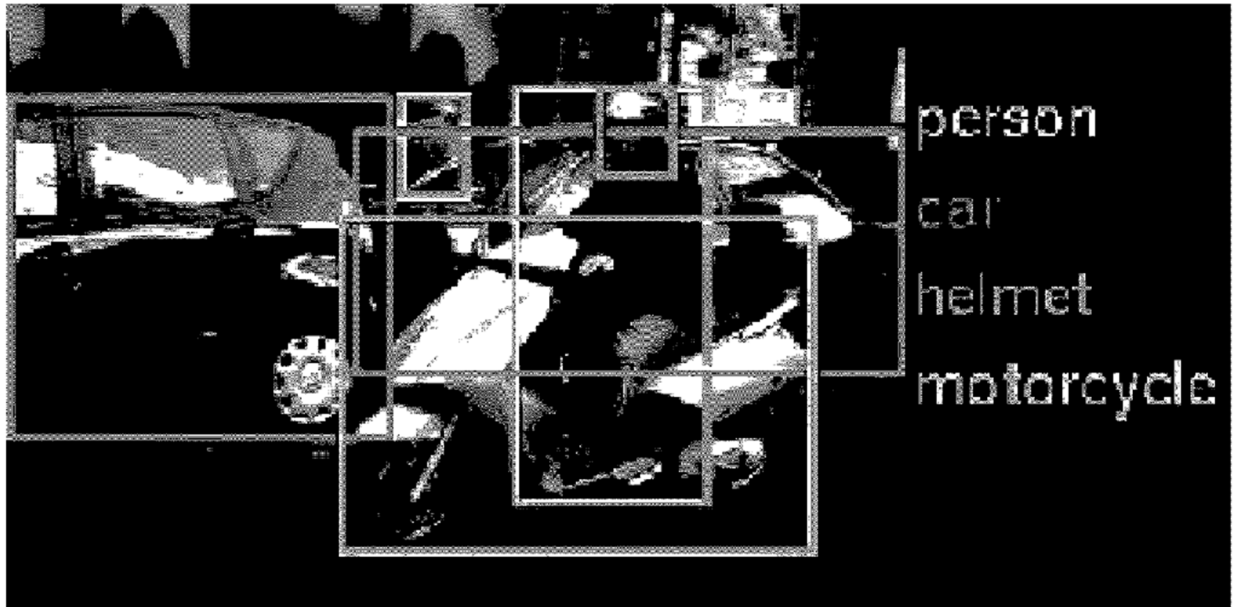
multiclass behavior or activity of a person in a training image and associated confidence scores form *summary statistics describing a state of mind of the user in the input image*. EX1003, ¶204.

i. 1[d]

The Ellenbogen and Munro teaches element [1d]. EX1003, ¶205. Ellenbogen's predictive model is trained (*the trained model*). See Section X.E.2.h (element 1[c]). The trained predictive model is then executed in regard to a new image that is not part of training data in an "operational phase" to determine whether a person of the image is exhibiting certain behavior or partaking in an activity, such as whether a driver of a vehicle is driving aggressively by tailgating (*executing the trained model to predict a state of mind of a user in a new image*). EX1004, ¶¶0057, 0133, 0148, 0168; EX1003, ¶205.

3. Claim 2

Ellenbogen and Munro render obvious claim 2. EX1003, ¶206. Ellenbogen teaches that images can be manipulated (e.g., with the addition of bounding boxes) before it is sent to human agents (*wherein the image is manipulated by adjusting values of pixels of the image before presenting to a human observer*). *Id.*; EX1004, ¶0170, *see also id.*, ¶¶0162-063, 166-68. As illustrated in Figure 13 (below), the addition of bounding boxes constitutes alterations to pixels in the digital images. *Id.*; EX1003, ¶206.



EX1004, Fig. 13

4. Claim 3

Ellenbogen and Munro render obvious claim 3. EX1003, ¶207. It would be obvious to a POSITA that the detection of “suspicious behavior” and “non-suspicious” behavior in Ellenbogen (EX1007, ¶¶0011, 0057, 0161) includes determining a “*state of mind*” that “*indicates whether a user is likely to perform a predetermined action.*” EX1003, ¶207. For example, “suspicious behavior” suggests that a user may be likely to commit a crime. EX1003, ¶207.

5. Claim 4

Ellenbogen and Munro render obvious claim 4. EX1003, ¶208. It would be obvious to a POSITA that the detection of “suspicious behavior” and “non-suspicious” behavior in Ellenbogen (EX1007, ¶¶0011, 0057, 0161) includes determining a “*state of mind*” that “*represents a measure of awareness of the user*

regarding an object.” EX1003, ¶208. For example, a user appearing lost (identified as “non-suspicious” in Ellenbogen) would represent a lack of awareness of the user’s surroundings. *Id.*

6. Claim 5

Ellenbogen and Munro render obvious claim 5. EX1003, ¶209. Ellenbogen’s human agent response to a query (*the response from a human observer*) includes a query result as well as a “confidence measure” of the result that is a rating from 0 to 1.0 and can be a value such as “0.9” or “0.5,” for example, which are ordinal confidence rating values on an ordinal scale (*wherein the response from a human observer comprises a rating on an ordinal scale*). EX1004, ¶¶0007, 0012, 0135 (“confidence measure may be directly supplied by an agent”), 0149, 0190, 0191, Fig. 34; EX1003, ¶209. Ellenbogen’s teaching aligns with the ’046 patent, which provides an example ordinal scale as values such as “1” or “4.” EX1001, 10:34-53, Fig. 6.

7. Claim 6

Ellenbogen and Munro render obvious claim 6. EX1003, ¶210. Ellenbogen’s trained predictive model can be a “deep neural network, a convolutional neural network (CNN), a Faster Region-based CNN (R-CNN), and the like,” as well as a “deep learning neural network” (*wherein the 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, ¶¶0053, 0136, 0162; EX1007, ¶¶0027-0032, 0060; EX1003, ¶210.

8. Claim 7

Ellenbogen and Munro render obvious claim 7. EX1003, ¶211. The Ellenbogen-Munro composite output is part of *the summary statistics* based on aggregation of feedback from human agents. Section X.E.2.f (element 1[b-iii]); EX1003, ¶211. Ellenbogen further teaches utilizing latency in the response time from human agents in the assessment of agent confidence. EX1004, ¶¶0100-0101; Figure 27. Thus, a POSITA would have understood that the Ellenbogen-Munro composite (*the summary statistics*) are “associated with at least one of a content of a response, a time associated with entering a response, and a position of an eye of a human observer associated with the response, the position being measured with respect to a display associated with the image.” EX1003, ¶211.

9. Claim 8

a. 8[Pre]

See Section X.E.2.a (element 1[Pre]). Ellenbogen’s system can be implemented by a computer system including non-transitory computer readable storage medium/computer memory storing instructions that are executed by one or more processors to provide Ellenbogen’s disclosed functionality (*A non-transitory computer readable storage medium storing instructions that when executed by one*

or more processors, cause the one or more processors to perform steps). EX1004, ¶0015; EX1003, ¶212.

b. 8[a]

See Section X.E.2.b (element 1[a]).

c. 8[b]

See Section X.E.2.c (element 1[b]).

d. 8[b-i]

See Section X.E.2.d (element 1[b-i]).

e. 8[b-ii]

See Section X.E.2.e (element 1[b-ii]).

f. 8[b-iii]

See Section X.E.2.f (element 1[b-iii]).

g. 8[b-iv]

See Section X.E.2.g (element 1[b-iv]).

h. 8[c]

See Section X.E.2.h (element 1[c]).

i. 8[d]

See Section X.E.2.i (element 1[d]).

10. Claim 9

See Section X.E.3 (claim 2).

11. Claim 10

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

12. Claim 11

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

13. Claim 12

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

14. Claim 13

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

15. Claim 14

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

16. Claim 15

a. 15[Pre]

See Sections X.E.2.a (element 1[Pre]), X.E.9.a (element 8[Pre]).

b. 15[a]

See Section X.E.2.b (element 1[a]).

c. 15[b]

See Section X.E.2.c (element 1[b]).

d. 15[b-i]

See Section X.E.2.d (element 1[b-i]).

e. 15[b-ii]

See Section X.E.2.e (element 1[b-ii]).

f. 15[b-iii]

See Section X.E.2.f (element 1[b-iii]).

g. 15[b-iv]

See Section X.E.2.g (element 1[b-iv]).

h. 15[c]

See Section X.E.2.h (element 1[c]).

i. 15[d]

See Section X.E.2.i (element 1[d]).

17. Claim 16

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

18. Claim 17

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

19. Claim 18

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

20. Claim 19

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

F. Printed Matter

Many limitations of the Challenged Claims are “directed to the content of information and lacking a requisite functional relationship,” i.e., printed matter, and thus “are not entitled to patentable weight[.]” *Praxair Distribution, Inc. v. Mallinckrodt Hosp. Prods. IP Ltd.*, 890 F.3d 1024, 1032 (Fed. Cir. 2018). In particular, the following limitations are printed matter:

- 1[a], 8[a], 15[a] – “image displaying one or more users”
- 1[b-i], 8[b-i], 15[b-i] – “request to answer a question about a state of mind of a user in the image”
- 1[b-ii], 8[b-ii], 15[b-ii] – “response representing a judgment by the human observer of the state of mind of the user in the image”

Petition for *Inter Partes* Review of U.S. Patent No. 11,753,046

- 1[b-iii], 8[b-iii], 15[b-iii] – “summary statistics describing the state of mind of the user in the image based on the received responses from the plurality of human observers”
- 1[c], 8[c], 15[c] – “input image showing a user” and “summary statistics describing a state of mind of the user in the input image”
- 1[d], 8[d], 15[d] – “a state of mind of a user in a new image”
- 3, 10, 17 – “the state of mind of the user in the image indicates whether the user is likely to perform a predetermined action”
- 4, 11, 18 – “the state of mind of the user in the image represents a measure of awareness of the user regarding an object”
- 5, 12, 19 – “the response from a human observer comprises a rating on an ordinal scale”

The above limitations each recites a data element (sensor data, input sensor data, statistical summary, statistical summary data, parameter, responses, etc.) and then defines that data element based on its content. But a patent cannot cover what “sensor data display[s],” what a “statistical summary characterize[s],” or what “user responses describ[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 1[b-ii] that the user response “represent[] a judgment by the human observer of the state of mind of the

user in the image” should be ignored because the content of the user response is not patentable. Similarly, the recited limitations about what the predicted “state of mind ... represents” in claims 3-5, 10-12, and 17-19 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, ¶243; *see also* 8[Pre] and 15[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).

Petition for *Inter Partes* Review of U.S. Patent No. 11,753,046

Respectfully Submitted,

Date: October 1, 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 1, 2025, a complete and entire copy of the **PETITION FOR *INTER PARTES* REVIEW OF CLAIMS 1-19 OF U.S. PATENT NO. 11,753,046** 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 '046 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 Steet
Mountain View, CA 94041

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

Date: October 1, 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 13,998 words, within the 14,000 word limit allowed under §42.24(a)(1)(i).

Date: October 1, 2025

/Roger Fulghum/

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