

UNITED STATES PATENT AND TRADEMARK OFFICE

---

BEFORE THE PATENT TRIAL AND APPEAL BOARD

---

MICROSOFT CORPORATION,  
Petitioner,

v.

REALD SPARK, LLC  
Patent Owner.

---

Case IPR2023-01046  
Patent No. 10,740,985

---

**SUPPLEMENTAL DECLARATION OF DR. ELI SABER**

## TABLE OF CONTENTS

|       |                                                                                                                                           |    |
|-------|-------------------------------------------------------------------------------------------------------------------------------------------|----|
| I.    | INTRODUCTION .....                                                                                                                        | 1  |
| II.   | QUALIFICATIONS .....                                                                                                                      | 1  |
| III.  | MATERIALS REVIEWED .....                                                                                                                  | 6  |
| IV.   | A PERSON OF ORDINARY SKILL IN THE ART .....                                                                                               | 8  |
| V.    | RELEVANT LEGAL STANDARDS .....                                                                                                            | 9  |
| VI.   | BRIEF OVERVIEW OF MACHINE LEARNING .....                                                                                                  | 15 |
| VII.  | OVERVIEW OF THE '985 PATENT .....                                                                                                         | 18 |
| VIII. | OVERVIEW OF THE ASSERTED PRIOR ART .....                                                                                                  | 28 |
|       | A.    Ganin .....                                                                                                                         | 28 |
|       | B.    Chalom.....                                                                                                                         | 29 |
|       | C.    Aslan.....                                                                                                                          | 30 |
|       | D.    Grau .....                                                                                                                          | 31 |
|       | E.    Pace.....                                                                                                                           | 31 |
| IX.   | CLAIM CONSTRUCTION .....                                                                                                                  | 32 |
| X.    | CLAIMS 1-14 AND 16-19 ARE PATENTABLE OVER THE ASSERTED<br>GROUNDS .....                                                                   | 32 |
|       | A.    Ground I: Ganin Does Not Anticipate Claims 1-3, 5, 10, 12-14, and 18....                                                            | 32 |
|       | 1.    Ganin does not disclose the use of a second machine learning<br>algorithm as claimed in element 1[E]/18[E] of the '985 Patent. .... | 32 |
|       | B.    Ground III: Claims 1-3, 5-6, 8-14, and 16-19 are not obvious over Chalom<br>in view of Aslan. ....                                  | 48 |
|       | 1.    Chalom and Aslan Cannot be Combined as Petitioner Suggests. ...                                                                     | 48 |
|       | 2.    The Combination of Chalom and Aslan Does Not Render Claims 1-<br>3, 5-6, 8-14, and 16-19 Obvious. ....                              | 56 |
| XI.   | CONCLUSION .....                                                                                                                          | 64 |

## **PATENT OWNER’S EXHIBIT LIST**

| <b>Exhibit #</b> | <b>Reference Name</b>                                                                                                                                                                                                                                                      |
|------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| EX-1001          | U.S. Patent No. 10,740,985 to Eric Sommerlade and Alexandros Neophytou (“’985 Patent”)                                                                                                                                                                                     |
| EX-1002          | Patent Prosecution History of U.S. Patent No. 10,740,985                                                                                                                                                                                                                   |
| EX-1003          | Declaration of L. Karam Regarding Invalidity of U.S. Patent No. 10,740,985                                                                                                                                                                                                 |
| EX-1004          | Curriculum Vitae of Lina Karam                                                                                                                                                                                                                                             |
| EX-1005          | U.S. Patent No. 10,664,949 to Edmond Chalom and Or Shimshi (“Chalom”)                                                                                                                                                                                                      |
| EX-1006          | Ganin et al. “Deepwarp: Photorealistic image resynthesis for gaze manipulation.” ECCV 2016: 14th European Conference on Computer Vision, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II (pp. 311-326). Springer International Publishing. (“Ganin”) |
| EX-1007          | U.S. Patent Application Publication No. 2017/0132528 A1 to Aslan et al. (“Aslan”)                                                                                                                                                                                          |
| EX-1008          | Webpage printout at <a href="https://link.springer.com/chapter/10.1007/978-3-319-46475-6_20">https://link.springer.com/chapter/10.1007/978-3-319-46475-6_20</a> (DeepWarp: Photorealistic Image Resynthesis for Gaze Manipulation)                                         |
| EX-1009          | U.S. Patent Application Publication No. 2018/0158246 A1 to Oliver Grau and Daniel Pohl (“Grau”)                                                                                                                                                                            |
| EX-1010          | U.S. Patent No. 8,942,283 to Charles P. Pace (“Pace”)                                                                                                                                                                                                                      |
| EX-1011          | U.S. Provisional Application No. 62/542,661 filed August 8, 2017 (“’661 Prov. Appl.”)                                                                                                                                                                                      |
| EX-1012          | U.S. Provisional Application No. 62/543,587 filed August 10, 2017 (“’587 Prov. Appl.”)                                                                                                                                                                                     |
| EX-1013          | Prosecution of European Patent Application No. 18844651                                                                                                                                                                                                                    |
| EX-1014          | [Reserved]                                                                                                                                                                                                                                                                 |
| EX-1015          | [Reserved]                                                                                                                                                                                                                                                                 |
| EX-1016          | Liu et al. “Optical flow and principal component analysis-based motion detection in outdoor videos.” Eurasip Journal on Advances in Signal Processing, volume 2010, article id 680623, 6 pages, 2010. Hindawi Publishing Corporation. (“Liu”)                              |

| <b>Exhibit #</b> | <b>Reference Name</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| EX-1017          | Webpage printout at <a href="https://www.reald.com/privacypolicy">https://www.reald.com/privacypolicy</a> (RealD Privacy Policy)                                                                                                                                                                                                                                                                                                                                                              |
| EX-1018          | Webpage printout at <a href="https://web.archive.org/web/20161231162955/https://www.skoltech.ru/en/2016/11/ckolteh-litsenziroval-tehnologiyu-obespecheniya-zritelnogo-kontakta-v-videokonferentsiyah-v-kompaniyu-reald/">https://web.archive.org/web/20161231162955/https://www.skoltech.ru/en/2016/11/ckolteh-litsenziroval-tehnologiyu-obespecheniya-zritelnogo-kontakta-v-videokonferentsiyah-v-kompaniyu-reald/</a> (Skoltech Licenses Eye Contact Videoconferencing Technology to RealD) |
| EX-1019          | Kononenko, Daniil. “Learnable Warping-Based Approach To Image Re-Synthesis With Application to Gaze Redirection.” (2017): 1-131. (Kononenko Doctoral Thesis)                                                                                                                                                                                                                                                                                                                                  |
| EX-1020          | Sotera Stipulation                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| EX-1021          | Yang et al. “Eye gaze correction with stereovision for video-teleconferencing.” ECCV 2002: 7th European Conference on Computer Vision, Copenhagen, Denmark, May 28–31, 2002, Proceedings, Part II (pp. 479-494). Springer Berlin Heidelberg, 2002. (“Yang”)                                                                                                                                                                                                                                   |
| EX-1022          | Cham et al. “Analogous view transfer for gaze correction in video sequences.” 7th International Conference on Control, Automation, Robotics and Vision, Singapore, December 2-5, 2002, Proceedings (pp. 1415-1420). IEEE Xplore. (“Cham”)                                                                                                                                                                                                                                                     |
| EX-1023          | Webpage printout at: <a href="https://ieeexplore.ieee.org/abstract/document/1234980">https://ieeexplore.ieee.org/abstract/document/1234980</a> (Analogous view transfer for gaze correction in video sequences)                                                                                                                                                                                                                                                                               |
| EX-1024          | Wood et al. “Gazedirector: Fully articulated eye gaze redirection in video.” arXiv: 1704.08763v1 [cs.CV], 27 April 2017.                                                                                                                                                                                                                                                                                                                                                                      |
| EX-1025          | Zitnick et al. “Manipulation of video eye gaze and head orientation for video teleconferencing.” Technical Report MSR-TR-99-46, Microsoft Research, 16 June 1999.                                                                                                                                                                                                                                                                                                                             |
| EX-1026          | Webpage printout at: <a href="https://www.microsoft.com/en-us/research/publication/manipulation-of-video-eye-gaze-and-head-orientation-for-video-teleconferencing/">https://www.microsoft.com/en-us/research/publication/manipulation-of-video-eye-gaze-and-head-orientation-for-video-teleconferencing/</a> (Manipulation of Video Eye Gaze and Head Orientation for Video Teleconferencing – Microsoft Research)                                                                            |

| <b>Exhibit #</b> | <b>Reference Name</b>                                                                                                                                                                                                                        |
|------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| EX-1027          | Gemmell et al. “Gaze awareness for video-conferencing: A software approach.” IEEE Multimedia, volume 7, issue 4, October-December 2000 (pp. 26-35). (“Gemmell”)                                                                              |
| EX-1028          | Kononenko et al. “Learning to look up: Realtime monocular gaze correction using machine learning.” CVPR 2015: IEEE/CVF Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7-12 June 2015 (pp. 4667-4675). (“Kononenko”) |
| EX-1029          | Webpage printout at:<br><a href="https://github.com/BlueWinters/DeepWarp">https://github.com/BlueWinters/DeepWarp</a> (GitHub - BlueWinters_DeepWarp_Tensorflow implement for DeepWarp)                                                      |
| EX-1030          | Aleš Hladnik, “Image Compression and Face Recognition: Two Image Processing Applications of Principal Component Analysis.” International Circular of Graphic Education and Research, No. 6, January 2013 (pp. 56-61).                        |
| EX-1031          | Complaint: RealD Spark, LLC v. Microsoft Corporation, WDWA-2-22-cv-00942, Dkt. 1.                                                                                                                                                            |
| EX-1032          | Hinton et al. Distilling the Knowledge in a Neural Network, arXiv:1503.02531v1 [stat.ML], March 2015. (“Hinton”)                                                                                                                             |
| EX-1033          | C. Bucilua, R. Caruana, and A. Niculescu-Mizil. “Model Compression.” KDD 2006: In Proceedings of the 12th ACM SIGKDD International Conference on Knowledge                                                                                   |
|                  | Discovery and Data Mining, (KDD), Philadelphia, PA, USA, August 20-23, pages 535–541, 2006 (pp. 535–541). (“Bucilua”)                                                                                                                        |
| EX-1034          | [Reserved]                                                                                                                                                                                                                                   |
| EX-1035          | Fleet et al., “Design and Use of Linear Models for Image Motion Analysis.” International Journal of Computer Vision, volume 36, issue 3, 2000 (pp. 171-193). Kluwer Academic Publishers. (“Fleet”)                                           |
| EX-1036          | Wulff et al. “Efficient Sparse-to-Dense Optical Flow Estimation using a Learned Basis and Layers.” CVPR 2015: IEEE/CVF Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7-12 June 2015 (pp. 120-130). (“Wulff”)       |

| <b>Exhibit #</b> | <b>Reference Name</b>                                                                                                                                                                                                                  |
|------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| EX-1037          | Burt et al. “A Multiresolution Spline With Application to Image Mosaics.” ACM Transactions on Graphics, volume 2, issue 4, October 1983 (pp. 217-236). (“Burt”)                                                                        |
| EX-1038          | Giger et al., “Gaze correction with a single webcam.” ICME 2014: IEEE International Conference on Multimedia and Expo, Chengdu, China, July 14-18, 2014. (“Giger”)                                                                     |
| EX-1039          | Ho et al. “Gaze correction using 3D video processing for videoconferencing.” ChinaSIP 2015: China Summit and International Conference on Signal and Information Processing, Chengdu, China, July 12-15, 2015 (pp. 496-499). (“Ho”)     |
| EX-1040          | Jain. Fundamentals of Digital Image Processing. Prentice-Hall, Inc., 1989 (excerpted). (“Jain”)                                                                                                                                        |
| EX-1041          | Vetterli et al. Wavelets and Subband Coding. Prentice-Hall, Inc., 1995 (excerpted). (“Vetterli”)                                                                                                                                       |
| EX-1042          | Dodge and Karam. Introduction to Machine Learning and Deep Learning: A Hands-On Starter’s Guide, 2017; revised 2019. Available online at <a href="https://deeplearningtextbook.org/">https://deeplearningtextbook.org/</a> . (“Dodge”) |
| EX-1043          | Goodfellow et al. Deep Learning, MIT Press, 2016. Available online at <a href="https://www.deeplearningbook.org/">https://www.deeplearningbook.org/</a> (excerpted). (“Goodfellow”)                                                    |
| EX-1044          | Murphy. Machine Learning: A Probabilistic Perspective, MIT Press, 2012. (“Murphy”)                                                                                                                                                     |
| EX-1045          | Sonka et al. Image Processing, Analysis, and Machine Vision, 4th edition, Cengage Learning, 2015.                                                                                                                                      |
| Ex. 1046         | U.S. Patent No. 10,043,088 (“Odry”)                                                                                                                                                                                                    |
| EX-1047          | Microsoft Computer Dictionary, fourth edition, 1999 (excerpted)                                                                                                                                                                        |
| EX-1048          | Internet Archive affidavit w exhibits                                                                                                                                                                                                  |
| EX-1049          | [Reserved]                                                                                                                                                                                                                             |
| EX-1050          | Webpage printout at: <a href="https://link.springer.com/book/10.1007/978-3-319-46475-6">https://link.springer.com/book/10.1007/978-3-319-46475-6</a> (Webpage Computer Vision – ECCV 2016)                                             |
| EX-1051          | Webpage printout at: <a href="https://asp-urasipjournals.springeropen.com/articles/10.1155/2010/680623">https://asp-urasipjournals.springeropen.com/articles/10.1155/2010/680623</a> (SpringerOpen, Liu)                               |

| <b>Exhibit #</b> | <b>Reference Name</b>                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| EX-1052          | Webpage printout at:<br><a href="https://dl.acm.org/doi/abs/10.1155/2010/680623#">https://dl.acm.org/doi/abs/10.1155/2010/680623#</a> (ACM, Liu)                                                                                                                                                                                                                                                                                                        |
| EX-1053          | U.S. Patent No. 9,338,449 to Joel Sole Rojals, Rajan Laxman Joshi, and Marta Karczewicz (“Rojals”)                                                                                                                                                                                                                                                                                                                                                      |
| EX-1054          | U.S. Patent Application Publication No. 2008/0043833 A1 to Jane Perera and Joseph D. Young (“Perera”)                                                                                                                                                                                                                                                                                                                                                   |
| EX-1055          | Joint Claim Construction and Prehearing Statement: RealD Spark, LLC v. Microsoft Corporation, WDWA-2-22-cv-00942, Dkt. 59.                                                                                                                                                                                                                                                                                                                              |
| EX-1056          | George Wolberg, Digital Image Warping, IEEE Computer Society Press, 1990.                                                                                                                                                                                                                                                                                                                                                                               |
| EX-1057          | Webpage printout at:<br><a href="https://ieeexplore.ieee.org/document/7299098">https://ieeexplore.ieee.org/document/7299098</a> (Learning to look up: Realtime monocular gaze correction using machine learning)                                                                                                                                                                                                                                        |
| EX-1058          | Webpage printout at:<br><a href="https://ieeexplore.ieee.org/document/6890306">https://ieeexplore.ieee.org/document/6890306</a> (Gaze correction with a single webcam)                                                                                                                                                                                                                                                                                  |
| EX-2001          | Declaration of Dr. Eli Saber, PhD                                                                                                                                                                                                                                                                                                                                                                                                                       |
| EX-2002          | <i>Curriculum Vitae</i> of Dr. Eli Saber, PhD                                                                                                                                                                                                                                                                                                                                                                                                           |
| EX-2003          | Microsoft's Answering Claim Construction Brief as filed in <i>RealD Spark, LLC v Microsoft Corporation</i> , U.S. District Court, Western District of Washington, Case No. 2:22-cv-00942-TL                                                                                                                                                                                                                                                             |
| EX-2004          | Exhibit 1 to Microsoft's Answering Claim Construction Brief as filed in <i>RealD Spark, LLC v Microsoft Corporation</i> , U.S. District Court, Western District of Washington, Case No. 2:22-cv-00942-TL): Kononenko <i>et al.</i> “Learning to look up: Realtime monocular gaze correction using machine learning.” <i>CVPR 2015: IEEE/CVF Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7-12 June 2015 (pp. 4667-4675).</i> |
| EX-2005          | Exhibit 2 to Microsoft's Answering Claim Construction Brief as filed in <i>RealD Spark, LLC v Microsoft Corporation</i> , U.S. District Court, Western District of Washington, Case                                                                                                                                                                                                                                                                     |

| Exhibit # | Reference Name                                                                                                                                                                                                                                                                    |
|-----------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|           | No. 2:22-cv-00942-TL: File History, Restriction Notice, Aug. 15, 2019; Response to Restriction Notice, Oct. 15, 2019                                                                                                                                                              |
| EX-2006   | Exhibit 3 to Microsoft's Answering Claim Construction Brief as filed in <i>RealD Spark, LLC v Microsoft Corporation</i> , U.S. District Court, Western District of Washington, Case No. 2:22-cv-00942-TL: U.S. Provisional Application 62/542,661                                 |
| EX-2007   | Exhibit 4 to Microsoft's Answering Claim Construction Brief as filed in <i>RealD Spark, LLC v Microsoft Corporation</i> , U.S. District Court, Western District of Washington, Case No. 2:22-cv-00942-TL: Adriana Romero <i>et al.</i> , Fitnets: Hints for Thin Deep Nets (2014) |
| EX-2008   | Supplemental Declaration of Dr. Eli Saber                                                                                                                                                                                                                                         |
| EX-2009   | Deposition Transcript of Dr. Lina J. Karam, Expert for Petitioner                                                                                                                                                                                                                 |
| EX-2010   | A. Murat Tekalp, “Digital Video Processing”, Prentice Hall 1995                                                                                                                                                                                                                   |

## **I. INTRODUCTION**

1. I have been retained on behalf of Patent Owner RealD Spark, LLC (“RealD” or “Patent Owner”) as an expert in this proceeding before the United States Patent and Trademark Office (“USPTO”) involving U.S. Patent No. 10,740,985 (“the ’985 Patent”). I am personally knowledgeable about the matters stated herein and am competent to make this declaration.

2. In my opinion, claims 1-14 and 16-19 are patentable over the asserted prior art raised by Petitioner Microsoft Corporation (“Petitioner”).

3. In summary, my opinions are:

- **Regarding Ground I, Ganin does not disclose the “second machine learning algorithm” elements recited in all of the challenged independent claims.**
- **Regarding Ground III, the combination of Chalom and Aslan fails because (1) there is no motivation to combine Chalom and Aslan, and (2) the combination does not disclose a second machine learning algorithm that uses the editing instructions and/or selection instructions of the first machine algorithm, and which are different from each other as required by all of the challenged independent claims.**

4. My analysis and conclusions regarding each ground are presented in more detail below.

## **II. QUALIFICATIONS**

5. My CV is attached as EX-2002 and I incorporate it here in full.

6. I am a Professor in the Department of Electrical and Microelectronic Engineering (EME) and the Chester F. Carlson Center for Imaging Science at the

Rochester Institute of Technology (RIT) in Rochester, New York. I also serve as the Director of the Image, Video and Computer Vision Laboratory. I joined RIT's full-time faculty in 2004. I have over thirty-four (34) years of industry and academic experience, thirty (30) of which are in the fields of Computer Vision, Digital Image/Video Processing, Pattern Recognition, and Machine Learning.

7. I received a Ph.D. in Electrical Engineering from the University of Rochester in 1996. My concentration was in Signal/Image/Video Processing, Pattern Recognition, and Computer Vision. I received a Master of Science Degree in Electrical Engineering from the University of Rochester in 1992 in the same concentration as listed above, and I received a Bachelor of Science Degree in Electrical and Computer Engineering, Summa Cum Laude, from the State University of New York at Buffalo in 1988.

8. Before becoming a full-time professor, I worked for Xerox Corporation from 1988 to 2004. During my 16 years at Xerox, I was responsible for: a) delivering color management solutions, image processing innovations, architectures, and algorithms for high-end color printing systems, b) xerographic sub-systems for a variety of color products and c) control systems for toner production facilities. Among other roles, as an Advanced Development Scientist and Manager, I established the Advanced Design Laboratory—an Imaging/Xerographics lab—and provided technical and managerial leadership for the Electrical, Imaging and Xerographics Department. As a Product

Development Scientist and Manager, I led the research and development of color management algorithms and image quality metrics for various product platforms. I also led the Image Science, Analysis, and Evaluation area, with twelve to fifteen direct reports and a budget of approximately \$2 million.

9. From 1997 until 2004, I was an adjunct faculty member at the Electrical Engineering Department of the Rochester Institute of Technology and at the Electrical & Computer Engineering Department of the University of Rochester. I was responsible for teaching undergraduate and graduate coursework in signal, image, video processing, pattern recognition, and communications and performing research in multimedia applications, pattern recognition, image understanding and color engineering.

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

11. I proposed, developed and introduced the Digital Video Processing course and reestablished the Pattern Recognition course at the Electrical Engineering Department at RIT. These graduate courses were instrumental in providing MS/PhD students with the proper foundation and knowledge in all aspects of the subject areas in order to perform in- depth research in the field and continue to advance the state of the art.

12. The Digital Video Processing course is designed to provide a thorough and comprehensive coverage of the field. Topics covered include the mathematical representation of digital video including motion and time-varying image formation models, spatio-temporal sampling/reconstruction, sampling structure conversions, motion estimation using optical flow, block-based, pel-recursive, Bayesian and point correspondence type methodologies, segmentation, motion tracking, filtering, restoration, super-resolution and compression.

13. The Pattern Recognition course, on the other hand, covers pertinent concepts in Bayes decision theory, linear and nonlinear classifiers, artificial neural networks, feature selection/generation/extraction and various methodologies of clustering.

14. I am currently to developing a graduate course titled: “Advanced Neural Networks and Deep Learning: Theory, Mathematics and Algorithms.” The course is intended to provide an in-depth comprehensive study of the theory, mathematics and algorithms for advanced neural networks and their usage in deep learning. It is primarily targeted for upper graduate students that are planning to pursue advanced studies/research in the field. The course provides a thorough study of shallow and deep neural networks architectures and algorithms, activation functions, backpropagation, initialization, optimization strategies, etc. It then progresses to convolutional and recurrent style neural networks and, if time permits, will provide an overview of attention

mechanisms/transformers and deep reinforcement learning. The course is being taught for the first time in Spring 2024.

15. My current research focuses on machine learning for object tracking/remote sensing applications; image/video processing for multimedia, military & biomedical applications, and computer vision and three-dimensional scene reconstruction. As a principal investigator (PI) or co-principal investigator, I have acquired almost \$6 million in research funding since joining RIT and have managed multiple government grants from the Department of Defense as well as several corporate grants from Hewlett-Packard, Lenel, and Data Physics. I am currently managing (as PI) a multiyear government grant from the Department of Defense title “Target Detection/Tracking and Activity Recognition from Multimodal Data.” In this project, we propose to develop operational target detection/tracking techniques and activity recognition/understanding methods that can leverage multimodal data in a fusion framework using deep learning algorithms while providing accurate and near real-time performance.

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

17. I am the author or co-author of thirty-eight (38) peer-reviewed journal publications.

18. I have also authored or co-authored one hundred and two (102) conference and workshop publications and a book entitled Advanced Linear Algebra for Engineers

with MATLAB, published by CRC Press in February 2009, and am a named inventor on multiple U.S. and foreign patents.

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

### **III. MATERIALS REVIEWED**

20. In forming my opinions, I have reviewed:

- a) EX-1001: U.S. Patent No. 10,740,985
- b) EX-1002: Patent Prosecution History of U.S. Patent No. 10,740,985
- c) Petition For Inter Partes Review Of U.S. Patent No. 10,740,985 (Claims 1-14 and 16-19) filed by Petitioner Microsoft Corporation (the “Petition”)
- d) Exhibits 1003-1058 to the Petition, including but not limited to:
  - i. Declaration of L. Karam Regarding Invalidity of U.S. Patent No. 10,740,985
  - ii. EX-1005: U.S. Patent No. 10,664,949 to Edmond Chalom and Or Shimshi (“Chalom”)
  - iii. EX-1006: Ganin *et al.* “Deepwarp: Photorealistic image resynthesis for gaze manipulation.” ECCV 2016: *14th European Conference on Computer Vision, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II* (pp. 311-326). Springer International Publishing. (“Ganin”)
  - iv. EX-1007: U.S. Patent Application Publication No. 2017/0132528 A1 to Aslan *et al.* (“Aslan”)

- v. EX-1009: U.S. Patent Application Publication No. 2018/0158246 A1 to Oliver Grau and Daniel Pohl (“Grau”)
- vi. EX-1010: U.S. Patent No. 8,942,283 to Charles P. Pace (“Pace”)
- vii. EX-1016: Liu *et al.* “Optical flow and principal component analysis-based motion detection in outdoor videos.” *Eurasip Journal on Advances in Signal Processing*, volume 2010, article id 680623, 6 pages, 2010. Hindawi Publishing Corporation. (“Liu”)
- e) EX-2003: Microsoft’s Answering Claim Construction Brief as filed in *RealD Spark, LLC v Microsoft Corporation*, U.S. District Court, Western District of Washington, Case No. 2:22-cv- 00942-TL
- f) EX-2004: Exhibit 1 to Microsoft's Answering Claim Construction Brief as filed in *RealD Spark, LLC v Microsoft Corporation*, U.S. District Court, Western District of Washington, Case No. 2:22-cv-00942-TL): Kononenko *et al.* “Learning to look up: Realtime monocular gaze correction using machine learning.” *CVPR 2015: IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, 7- 12 June 2015 (pp. 4667-4675)
- g) EX-2005: Exhibit 2 to Microsoft's Answering Claim Construction Brief as filed in *RealD Spark, LLC v Microsoft Corporation*, U.S. District Court, Western District of Washington, Case No. 2:22-cv-00942-TL: File History, Restriction Notice, Aug. 15, 2019; Response to Restriction Notice, Oct. 15, 2019
- h) EX-2006: Exhibit 3 to Microsoft's Answering Claim Construction Brief as filed in *RealD Spark, LLC v Microsoft Corporation*, U.S. District Court, Western District of Washington, Case No. 2:22-cv-00942-TL: U.S. Provisional Application 62/542,661
- i) EX-2007: Exhibit 4 to Microsoft's Answering Claim Construction Brief as filed in *RealD Spark, LLC v Microsoft Corporation*, U.S. District Court, Western District of Washington, Case No. 2:22-cv-00942-TL: Adriana Romero *et al.*, Fitnets: Hints for Thin Deep Nets (2014)

- j) IPR2023-01046 Institution Decision
- k) EX-2009: Deposition Transcript of Dr. Lina J. Karam, Expert for Petitioner
- l) EX-2010: A. Murat Tekalp, “Digital Video Processing”, Prentice Hall 1995

#### **IV. A PERSON OF ORDINARY SKILL IN THE ART**

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

22. In my opinion, a Person of Ordinary Skill in the Art (“POSITA”) at the time of the invention in August 2017 would have had at least a bachelor or master degree in electrical engineering, computer engineering, electrical and computer engineering, computer science, or similar educational background; or equivalent on the-job training, including approximately two to three years of experience in machine learning and/or pattern recognition. This level of skill is approximate and more experience would compensate for less formal education and vice versa. For example, an individual having no degree in engineering, but several years of experience developing machine learning and/or pattern recognition systems would qualify as a POSITA.

23. My qualifications and experience exceed those of the hypothetical POSITA defined above.

24. I have been asked to address the issues from the perspective of a POSITA in the field of the '985 Patent. I have done so throughout this declaration.

## **V. RELEVANT LEGAL STANDARDS**

25. As a technical expert, I am not offering any legal opinions. I am offering technical assessments and opinions. In rendering my analysis, I have been informed by counsel regarding various legal standards for determining patentability. I have applied those standards in forming my technical opinions expressed in this report.

26. The patent claims describe the invention made by the inventors and describe what the patent owner owns and what the owner may prevent others from using. There are two types of claims: independent claims and dependent claims. An independent claim sets forth all the requirements that must be met in order to be covered by that claim. Thus, it is not necessary to look at any other claim to determine what an independent claim covers. A dependent claim does not itself recite all of the requirements of the claim, but refers to another claim for some of its requirements. In this way, the claim depends on another claim. The law considers a dependent claim to incorporate all of the requirements of the claims to which it refers. The dependent claim then adds its own additional requirements. To determine what a dependent claim covers, it is necessary to look at both the dependent claim and any other claims to which it refers.

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

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

29. I understand that the first step in determining validity is to properly construe the claims. I further understand that, in this proceeding, the Board must determine the scope of the claims by using the same claim construction standard that United States District Courts use. Under this standard, words of a claim are given their “ordinary” and “customary” meaning as understood by a person of skill in the art at the time of the invention. I understand that such a person looks to the claims, specification, and file history in understanding the meaning of the claim terms. I have applied this understanding in this declaration.

30. I understand that a patent cannot be invalidated based on one person’s testimony alone without corroborating evidence, particularly documentary evidence.

Moreover, I understand that corroboration is required of any witness whose testimony alone is asserted to invalidate a patent, regardless of his or her level of interest in relation to that patent. Accordingly, even if the testifying individual has no interest in the litigation, I understand that corroboration is still required.

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

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

33. I understand that for a claim to be anticipated, each claim limitation must be

disclosed in a single prior art reference, and the claimed arrangement or combination of those limitations must also be disclosed in that same prior art reference. I understand that for a reference to be anticipatory, the reference must contain an enabling disclosure. I understand that merely naming or describing the subject matter is insufficient if it cannot be produced without undue experimentation.

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

35. To be obvious in light of a single prior art reference or multiple prior art references, there must be a reason either to modify either the single prior art reference or to combine the two or more prior art references to achieve the claimed invention. I

understand that this reason may come from a teaching, suggestion, or motivation to combine or may come from the reference or references themselves, the knowledge or “common sense” of one skilled in the art, or from the nature of the problem to be solved, and may be explicit or implicit from the prior art as a whole. The combination of familiar elements according to known methods is likely to be obvious when it does no more than yield predictable results.

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

37. I understand that the degree of teaching away will depend on the particular facts; in general, a reference will teach away if it suggests that the line of development flowing from the reference’s disclosure is unlikely to be productive of the result sought by the applicant. I understand that a reference teaches away, for example, if (1) the

combination would produce a seemingly inoperative device, or

(2) the references leave the impression that the product would not have the property sought by the applicant.

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

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

40. I understand that some articulated reasoning with some rational underpinning to support the legal conclusion of obviousness must be made. For example, I understand that there may be no rational underpinning to support the legal conclusion of obviousness where systems are already complete and the party challenging validity seeks to mix and match components, which each exist in the separate systems, from these complete systems.

41. I understand that objective indicia of non-obviousness must be taken into account in a proper obviousness analysis and that these objective indicia of nonobviousness may include the following: (1) failure of others; (2) long-felt but unmet need for the claimed invention; (3) commercial success attributable to the claimed

invention; (4) unexpected results of the claimed invention; (5) initial skepticism; (6) prior art teaching away from the claimed invention; (7) professional approval; (8) licensing; (9) acquiescence to the patent's validity; (10) copying of the claimed invention by others, and (11) progress through the Patent Office. I further understand that a so-called nexus between the merits of the claimed invention and evidence of objective indicia of non-obviousness is required in order for the evidence to be given substantial weight in an obviousness determination. As an example, I understand that nexus can be established if there was commercial success and if the invention disclosed in the patent was that which was commercially successful.

## **VI. BRIEF OVERVIEW OF MACHINE LEARNING**

42. In layman terms, machine learning is a subfield of artificial intelligence and relates to the capability of a machine to imitate intelligent human behavior. It provides computers with the ability to learn complex operations such as classifying patterns, translating languages, or recognizing human faces without having to be explicitly programmed to do so. For instance, to perform facial recognition, computers would traditionally need to be programmed with detailed instructions explaining how to accomplish this task. However, in many cases, writing a precise program for a machine to follow (i.e. such as training a computer to recognize pictures of people) may be a daunting or even impossible task. Hence, machine learning takes the approach of allowing computers to learn to program themselves through experience by utilizing the

appropriate data, which is generally termed as training data. Training data usually includes the proper inputs and their corresponding outputs, which is generally referred to as the ground truth. Put another way, a machine learning algorithm is a computer model that trains itself to perform a specific task by using training data.

43. Typically, machine learning starts with the proper training data (e.g. numbers, photos, text, pictures of faces, etc.). The training data usually includes many sets of input-output combinations that are carefully chosen a priori to be used for training purposes. Generally speaking, larger training data sets typically result in better training and result in higher accuracy.

44. Human designers then select an appropriate machine learning model, supply the training data in a repetitive fashion over many iterations and allow the computer model to train itself to find patterns or make proper predictions/adjustments from the training data at hand. A typical machine learning model usually includes an input layer that is designed to acquire the input data, multiple hidden layers to process the data, and an output layer to provide the appropriate results.

45. The accuracy of the model is typically measured and carefully monitored during the training cycle as the training data is being presented to the model in an iterative fashion. For instance, given a particular input, a machine learning model generates a corresponding output using its current state of weights and parameters. The model-computed output is then compared to the one supplied by the training data (e.g. the ground

truth) and the corresponding error between the model-computed output and the ground truth is generally backpropagated through the model using proper optimization techniques in order to continue to optimize the weights and parameters of the model. This process is then repeated multiple times using the entire training dataset to optimize the weights/parameters to yield higher accuracy results. Over time, human programmers can tweak the model, change/adjust parameters, and/or supply additional training data as needed to yield more accurate results. Once the training of the machine learning model is complete and all of the parameters and weights have been optimized for accuracy, it can then be utilized to perform the task that it was trained to do and can be further updated/upgraded in due time as additional training data becomes available.

46. Applying the foregoing overview to the context of gaze correction, a machine learning algorithm for adjusting images would be provided with training data that is a specific set of data that includes input images (sometimes called “before” images by Petitioner) and output images (sometimes called “after” images by Petitioner). *See, e.g.,* EX-2003 at pp. 6-8. A “before” image may be one where the user is gazing off-camera and the “after” image may be one where the user is gazing directly into the camera. The machine learning algorithm would then train itself with this information to take inputs similar to the “before” image from the training data and make adjustments so that the output is similar to the “after” image from the training data. In other words, the machine learning algorithm will take the training data and learn how to turn input images

into output images. After training is complete, the machine learning algorithm will generate instructions to perform the gaze correction during run-time, for example, by taking a received video stream with a user gazing off-camera, turn it into a corrected video stream where the user is gazing directly into the camera, and output that corrected video stream to the end user.

## **VII. OVERVIEW OF THE '985 PATENT**

47. The '985 Patent is titled “Adjusting a digital representation of a head region.” EX-1001 at Title.

48. The '985 Patent contains six independent claims, and thirteen dependent claims. *Id.* at 26:31-30:59. Of the independent claims, five are method claims, and one is a device claim.

49. The Abstract describes the subject matter and invention as “generating reference data for adjusting a digital representation of a head region, and methods and devices for adjusting the digital representation of a head region.” *Id.* at Abstract. The digital representation may be an image or a three- dimensional geometrical representation. *Id.* at 1:6-9. The “application relates particularly to adjusting target features of the digital representation of the head region, for example to correct a perceived gaze direction of eyes, or to modify the texture and/or shape of features such as the nose, mouth, chin or neck.” *Id.* at 1:9- 13.

50. By way of overview, the '985 Patent employs two machine learning

algorithms that adjust the head region of a user. *See, e.g., id.* at Figure 9, 16:45- 17:18, 26:45-57. This is accomplished by the '985 Patent's instructions, which adjust the digital representation of the head region. Both the first machine learning algorithm and the second machine learning algorithm of the '985 Patent provide these instructions. *See, e.g., id.* at 26:45-49 (reciting that the first machine learning algorithm provides instructions "for adjusting the digital representation of the head region") and 26:50-57 (reciting that the second machine learning algorithm provides instructions "for adjusting the digital representation of the head region."). Both of the '985 Patent's machine learning algorithms are trained using training data and then provide instructions to adjust the head region based on the training data.

51. The '985 Patent's machine learning algorithms adjust a digital representation (such as image) of a head region (such as the eye, facial, and/or neck region) in order to correct and/or improve aspects of a computer-generated display of the head region with minimal demands on computer hardware. *See id.* at 1:6-13, 1:46-55. For example, the algorithms may adjust the eyes or eye regions, nose, etc. of the digital representation of the head region, for example to correct a perceived gaze direction of eyes, or to modify the texture and/or shape of features such as nose, mouth, chin or neck. *Id.*

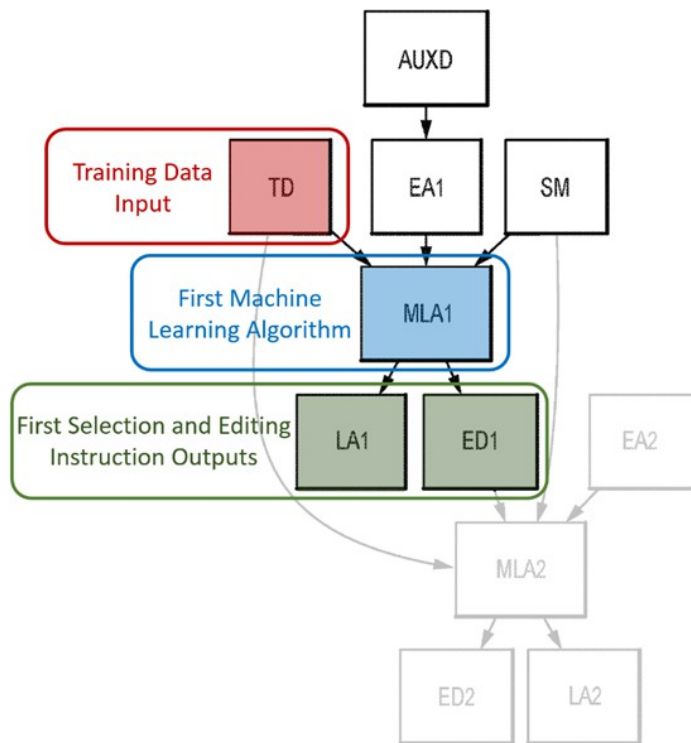
52. As one particular example, the '985 Patent's technology corrects the perceived gaze of a person's eyes during a video conference so that it appears that the

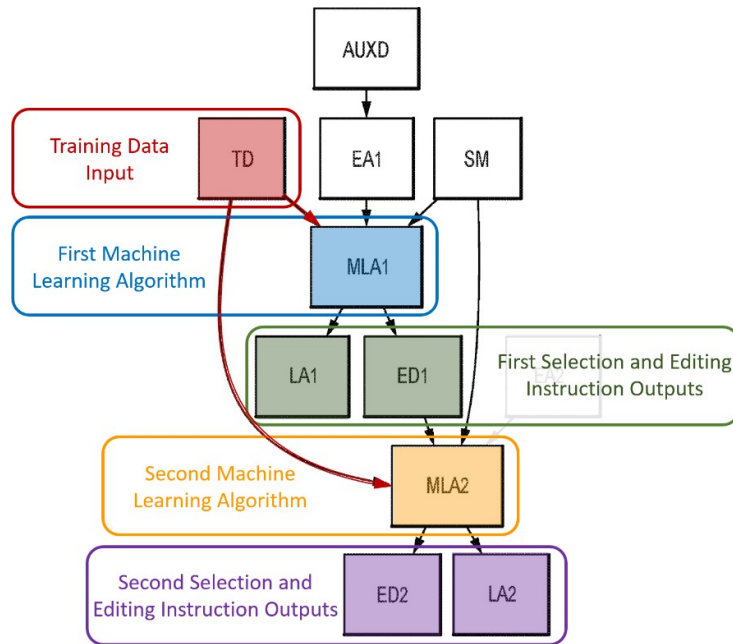
person is looking directly into the camera, and thus making eye contact with other attendees, rather than staring elsewhere at the screen. *Id.* at 1:30-45. The '985 system can also alter the texture and/or shape of facial features as desired, such as “mak[ing] [the] nose look smaller and/or slimmer,” “remov[ing] [a] double chin appearance,” “remov[ing] [neck] wrinkles,” or “adjusting a color of the hair.” *Id.* at 13:62-14:15.

53. To achieve the above, the '985 Patent provides a methodology for “generating reference data for adjusting a digital representation of a head region.” *Id.* at 26:31-27:3. This is accomplished by utilizing “training data [that] compris[es] a set of input patches [where] each input patch compris[es] a target feature of a digital representation of a head region prior to adjustment of the digital representation of the head region, wherein the target feature is the same for each input patch; and a set of output patches in one-to-one correspondence with the input patches, [where] each output patch compris[es] the target feature of the digital representation of a head region after adjustment of the digital representation of the head region”. *Id.* at 26:34-44. The '985 Patent then employs properly designed machine learning algorithms to generate appropriate reference data from the acquired training data. *Id.* at 45-57. The reference data comprises appropriate “editing instructions for adjusting the digital representation of the head region for a range of possible digital representations of the head region.” *Id.* at 26:45-49, 26:50- 57.

54. An example of this methodology is shown in Figure 9 of the '985 Patent,

reproduced with emphasis below. The '985 claimed methods include a **first machine learning algorithm** that receives **training data** and then generates first reference data. The reference data includes one or both of the **first selection instructions** and **first editing instructions** for determining how the head region is to be adjusted. *See, e.g., id.* at 26:45-49 and 26:58-63.





55. The **second machine learning algorithm** is similar. It receives **training data** as input, and outputs **second selection instructions and/or second editing instructions** for determining how the head region is to be adjusted. *See, e.g., id.* at 26:50-57 and 26:64-27:3. The second machine learning algorithm also receives the **selection and/or editing instructions** output from the first machine learning algorithm. *See, e.g., id.* at 26:50-54. Both machine learning algorithms receive the training data as input, and output instructions for adjusting the head region of a user. *See, e.g., id.* at 26:34-27:3.

56. The '985 Patent explains that the use:

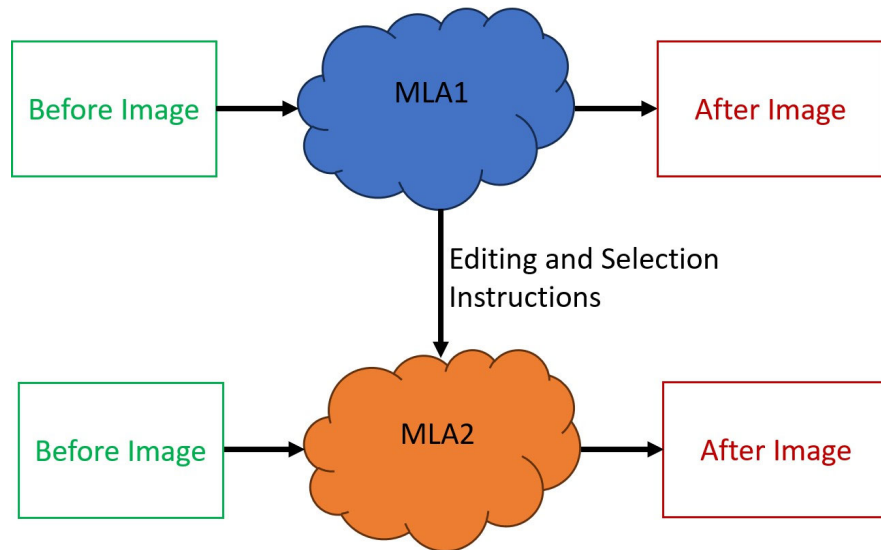
“of two machine learning algorithms [allows] an improved balance to be achieved between accuracy of the adjustment process and speed of execution. The first machine learning algorithm can be configured to provide highly detailed first reference data, which provides high accuracy. Use of this first reference data directly in a processing technique (e.g. image processing technique) to adjust a digital representation

(e.g. image or three-dimensional geometrical representation) of a head region would be relatively expensive in terms of computational resources because of the high level of detail. By arranging instead for the first reference data to be provided to a second machine learning algorithm, which provides, based on the first reference data and the original training data, the reference data (the second reference data) that is to be used for the adjustment of the digital representation of the head region, it is possible to benefit to an extent from the high accuracy of the first machine learning algorithm whilst also providing reference data that is less detailed and thus easier to process efficiently when performing the adjustment of the digital representation of the head region. The quality of the reference data provided by the second machine learning algorithm is found to be significantly improved when the second machine learning algorithm is provided with both the first reference data and the training data in comparison to when the second machine learning algorithm is provided only with the training data.”

*Id.* at 2:17-43.

57. One reason for having two machine learning algorithms that both receive training data and output instructions, as in the '985 Patent, is that the first machine learning algorithm may create its instructions using a computationally intensive process to create highly accurate results, which can be fed to the second machine learning algorithm to create instructions using less complexity while still arriving at highly accurate results. *Id.* at 16:12-44.

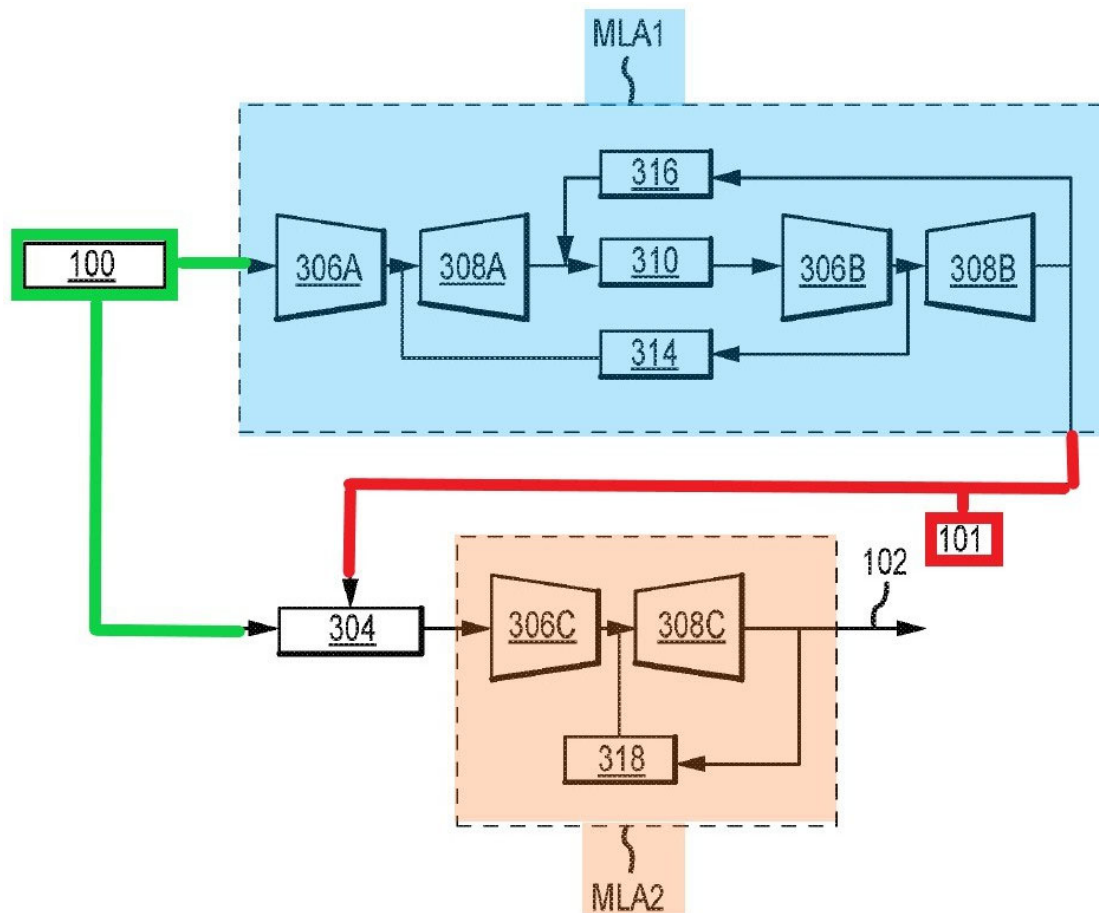
58. The below image is a simplified graphic showing the relationship of the '985 Patent's two machine learning algorithms:



59. The '985 Patent's first machine learning algorithm (in blue) provides its instructions to the second machine learning algorithm (in orange), essentially operating as shortcuts so that the second machine learning algorithm may still provide highly accurate "after" images (in red) even though it may use less complex instructions or fewer parameters than the first machine learning algorithm to do so. *Id.* at 16:8-44; *see also id.* at 17:15-17 (explaining that the "second machine learning algorithm MLA2 thus gets the editing instructions to match or to approximate, and does not have to infer these from the matched input images as MLA1."); 21:40-45 (explaining that the second machine learning algorithm "will be able to provide accurate conversion . . . using a simpler trained machine learning model (which can be stored and operated using fewer computing resources than the first machine learning algorithm MLA1).").

60. One innovative way to use two machine learning algorithms as outlined above is described in detail in an embodiment of the '985 Patent. "Figure 14 depicts a further embodiment of a method of generating reference data for adjusting a digital representation of a head region using a framework of the type depicted in FIG. 8." *Id.* at 19:40-43.

61. Figure 14, with color added for clarification, is as follows:

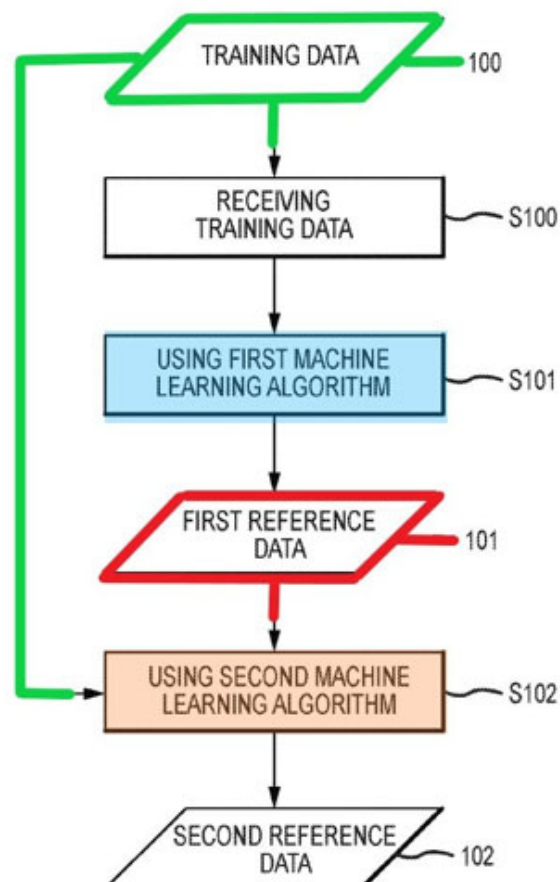


62. Figure 14's first machine learning algorithm is shown in blue. As depicted, the training data, identified as "100" (depicted in green) is input to the first machine learning algorithm, and then a number of steps occur within the first machine learning

algorithm, resulting in the output of the “first reference data” identified as “101” (depicted in red): “Once matching has been achieved or a predetermined number of iterations have been performed, the first machine learning algorithm MLA1 is considered trained and the **resulting output from the second predictor 308B can be used to provide the first reference data 101 . . .**” *Id.* at 21:1-8 (emphasis added).

63. Figure 14 shows the second machine learning algorithm in orange. As depicted, both the training data, identified as “100” (depicted in green), and the first reference data, identified as “101” (depicted in red) are input into the second machine learning algorithm (depicted in orange).

64. Figure 8, with emphasis added for clarification, also depicts this:



65. In one embodiment, the '985 Patent teaches that “the first reference data 101 comprises first editing instructions (e.g. first image editing instructions) for a range of possible configurations of the target feature (e.g. different gaze directions and/or eye morphologies in the case where the target feature comprises an eye region) and first selection instructions for selecting editing instructions (from the first editing instructions) for a particular input patch (e.g. input image patch) based on the configuration of the target feature of the input patch (e.g. the particular gaze direction and/or particular eye morphology of that input patch).” *Id.* at 15:1-11. As one example, the first reference data comprises editing instructions such as “displacement vector fields that represent transformations of [a given] image patch, or other representations of transformations of the image patch . . .” *Id.* at 7:48-54. The displacement vector field, in essence, may represent an appropriate transformation of a given input image patch (e.g. an input eye region) into a gaze corrected output patch (e.g. a gaze corrected eye region). *Id.* at 8:47-9:12. Other examples of editing instructions, including the use of filter fields or a set of typical image components, are also disclosed in the '985 Patent. *Id.* at 9:13-38.

66. The '985 Patent also teaches an embodiment where “the second reference data comprises second editing instructions (e.g. second image editing instructions) for a range of possible configurations of the target feature and second selection instructions for selecting editing instructions (from the second editing instructions) for a particular input patch (e.g. input image patch) based on the configuration of the target feature of the input patch.” *Id.* at 15:12-19. The second editing instructions may also correspond to

“displacement vector fields that represent transformations of [a given] image patch, or other representations of transformations of the image patch.” *Id.* at 7:48-54.

67. The '985 Patent teaches, in one example, that “efficiency is further improved by providing editing instructions (e.g. image editing instructions) in the second reference data in a compressed representation. The use of a compressed representation reduces data storage and bandwidth requirements during use of the editing instructions to perform adjustment of a digital representation of a head region.” *Id.* at 2:44-50.

## **VIII. OVERVIEW OF THE ASSERTED PRIOR ART**

### **A. Ganin**

68. Ganin describes an approach for generating “highly realistic images of a given face with a redirected gaze.” EX-1006 at Abstract. The approach considers “a specific instance of conditional image generation and suggest[s] a new deep architecture that can handle [the task of generating highly realistic images of a given face with a redirected gaze] very well as revealed by numerical comparison with prior art and a user study. . . . [The] deep architecture performs coarse-to-fine warping with an additional intensity correction of individual pixels. All these operations are performed in a feed-forward manner, and the parameters associated with different operations are learned jointly in the end-to-end fashion. After learning, the resulting neural network can synthesize images with manipulated gaze, while the redirection angle can be selected arbitrarily from a certain range and provided as an input to the network.” *Id.*

69. As it relates to the Petition’s arguments, Ganin teaches, at best, a single machine learning algorithm. Ganin’s use of the terms “single,” “feed-forward architecture,” “two stage[]” network, and “end-to-end” (*id.* at p. 4), confirm that Ganin does not disclose two machine learning algorithms, as I will explain in more detail below.

**B. Chalom**

70. Chalom describes apparatus, systems, and methods “related to eye contact correction to provide a virtual user gaze aligned with a camera while the user views a display . . . Such techniques may include determining and reducing histogram of oriented gradient features for an eye region of a source image to provide a feature set, applying a pretrained classifier to the feature set to determine a motion vector field for the eye region, and warping and inserting the eye region into the source image to generate an eye contact corrected image.” EX-1005 at Abstract.

71. As it relates to the Petition’s arguments, Chalom teaches a single machine learning algorithm that has an “offline[] and a real time component . . . [that] provides eye contact correction” in real time. *Id.* at 3:21-24. The offline training takes a “database of faces with a known gaze angle . . . [and] train[s] a machine learning system.” *Id.* at 3:24-27. That system includes a “random forest classifier,” which then determines which random forest to implement based on the camera’s angle. *Id.* at 10:35-39. For example, if the camera is above the display, one random forest model may be implemented by the classifier, while if the camera is to the left of the display, a different random forest model may be implemented by the classifier. *Id.*; *see also id.* at 10:40-64.

**C. Aslan**

72. Aslan is directed to systems and methods enabling multiple machine learning models to be jointly trained in parallel. “An example process for jointly training multiple machine learning models includes providing a set of machine learning models that are to learn a respective task, the set of machine learning models including a first machine learning model and a second machine learning model. The process can initiate training of the first machine learning model to learn a task using training data. During the training of the first machine learning model, information can be passed between the first machine learning model and the second machine learning model. Such passing of information (or ‘transfer of knowledge’) between the machine learning models can be accomplished via the formulation, and optimization, of an objective function that comprises model parameters that are based on the multiple machine learning models in the set.” EX-1007 at Abstract.

73. As it relates to the Petition’s arguments, Aslan discloses a student- teacher model that can jointly train two machine learning models. *Id.* at ¶ 20. The models “pass[] information” to one another through “an objective function.” *Id.* at ¶ 28. The objective function can “measure the agreement/disagreement . . . of the two models” with a penalty term. *Id.* at ¶ 31. To force the two models to agree, the objective function and its penalty term may be optimized. *Id.*

**D. Grau**

74. Grau describes “[a] system, article, and method of providing user facial displays in virtual or augmented reality for face occluding head-mounted displays.” EX-1009 at Abstract. The “present method and system propose to augment the image of one or more external cameras with internal images of the occluded areas captured by one or more cameras mounted inside the HMD. This may be performed with closed virtual or augmented reality HMDs that completely cover a part of a user’s face . . . where an image 100 from an external camera has a face [] of a user covered by a virtual reality HMD []. [The i]mage . . . shows the user’s face [] without the HMD (and without occlusions) and as desired for the face model in the virtual reality as discussed below. However, the methods herein also may apply to augmented reality formed by using see-through glasses that still may block the view of the face in the images of an external camera.” *Id.* at ¶ 30.

**E. Pace**

75. Pace describes systems and methods related to processing video data. “Video data having a series of video frames is received and processed. One or more instances of a candidate feature are detected in the video frames. The previously decoded video frames are processed to identify potential matches of the candidate feature. When a substantial amount of portions of previously decoded video frames include instances of the candidate feature, the instances of the candidate feature are aggregated into a set. The candidate feature set is used to create a feature-based model. The feature-based model

includes a model of deformation variation and a model of appearance variation of instances of the candidate feature. The feature- based model compression efficiency is compared with the conventional video compression efficiency.” EX-1010 at Abstract.

## **IX. CLAIM CONSTRUCTION**

76. As explained in the Institution Decision, “Petitioner contends that it ‘applied plain and ordinary meanings of all terms and no express constructions are required here.’” Institution Decision at p. 16. The Board then concluded that “[a]t this stage in the proceeding, and without further argument from the parties, we determine that we do not need to expressly construe any term for purposes of this Decision.” *Id.* As with my previous declaration, I continue to apply the plain and ordinary meaning of all terms in my analysis.

## **X. CLAIMS 1-14 AND 16-19 ARE PATENTABLE OVER THE ASSERTED GROUNDS**

### **A. Ground I: Ganin Does Not Anticipate Claims 1-3, 5, 10, 12-14, and 18.**

#### **1. Ganin does not disclose the use of a second machine learning algorithm as claimed in element 1[E]/18[E] of the ’985 Patent.**

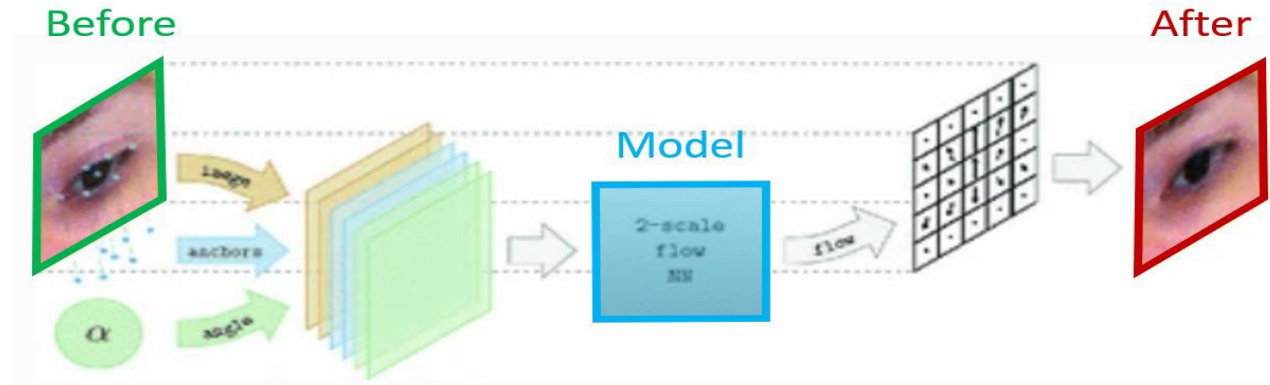
77. Claim elements 1[E] and 18[E] require “using a second machine learning algorithm to generate second reference data using the same training data as the first machine learning algorithm and the first reference data generated by the first machine learning algorithm, the second reference data comprising editing instructions for adjusting the digital representation of the head region for a range of

possible digital representations of the head region,” (EX-1001 at 26:50-57, 29:58-30:5) and thus, as already explained above, the ’985 Patent utilizes two machine learning algorithms.

78. Ganin does not disclose the “second machine learning algorithm” element recited in 1[E] and 18[E], and instead, discloses at best a single alleged “machine learning algorithm.” Ganin’s single “algorithm” uses a two-step process, but this is not the same as having two machine learning algorithms. This is clear from Ganin for a variety of reasons that I will outline in the following paragraphs.

*a) Ganin discloses a single end-to-end architecture with two steps, not two machine learning algorithms.*

79. **Overview of Ganin’s Single MLA:** Ganin teaches, at best, a single machine learning algorithm. Ganin’s single algorithm uses a two-step process composed of a coarse module and a fine module. EX-1006 at p. 4 (stating that “**the** network . . . predict[s] [the gaze redirection] **in two stages** in a coarse-to-fine manner.”). In this two-step model, there is one set of training data that includes before images (in green) that are images prior to entering the coarse module and the after images (in red) that are images after exiting the fine module as shown below:



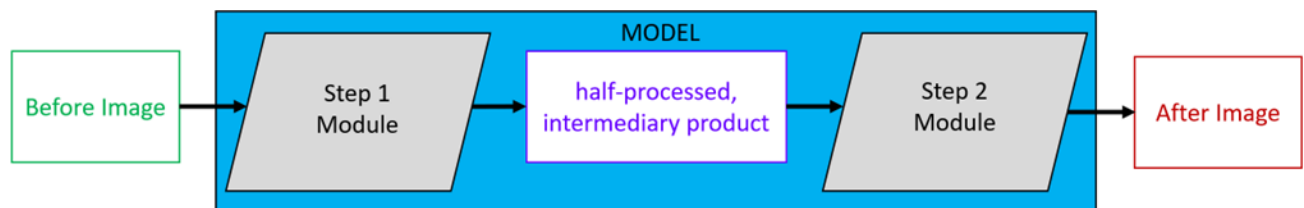
*Id.* at Fig. 2 (annotated).

80. Ganin’s training data provides a single set of before images (in green) and after images (in red). The before images are the inputs with an off-camera gaze. They are input to the single model (shown in blue as the “2-scale flow”). Ganin’s “coarse” step provides a coarse adjustment to the image. Then that semi-adjusted image is passed to the “fine” step to adjust the image on, for example, a pixel-by- pixel basis so that the eyelashes, pupil, and iris are rendered correctly. *Id.* at p. 4 (explaining that “the coarse module gives a rough estimate [while] . . . problematic regions [] can only be fixed by a neural network operating at a finer scale.”). The corrected gaze images are the final, adjusted images (in red). Only at the end of the second step does Ganin provide a usable, adjusted “after” image.

81. Specifically, Figure 2 above illustrates that “[t]he **proposed system** takes an input eye region, feature points (anchors) as well as a correction angle  $\alpha$  and sends them to **the multi-scale neural network** (*see* Sect. 3.2) predicting a flow field. The flow field is then applied to the input image to produce an image of a redirected eye. Finally, the output

is enhanced by processing with the lightness correction neural network (see Sect. 3.4).” *Id.* at Figure 2, p. 4. This does not depict the use of two machine learning algorithms, but instead, a single architecture. Ganin explicitly states that it uses a single architecture with two steps: “the network . . . predict[s] [the gaze redirection] in two stages in a coarse-to-fine manner.” *Id.* at p. 4.

82. A simplified illustration of a single end-to-end architecture with two steps is below:



83. The above illustration depicts a single model that may, for example, adjust the gaze of a headshot image. In step one, the model takes the background of the image and adjusts it to correct, for example, the lighting or focus. The model then passes the semi-completed image to a second step that adjusts the foreground to correct the lighting or focus. Between steps one and two, there is no completed output, no highly-accurate corrected image. There is simply a half-processed, unusable intermediary product as shown above in purple.

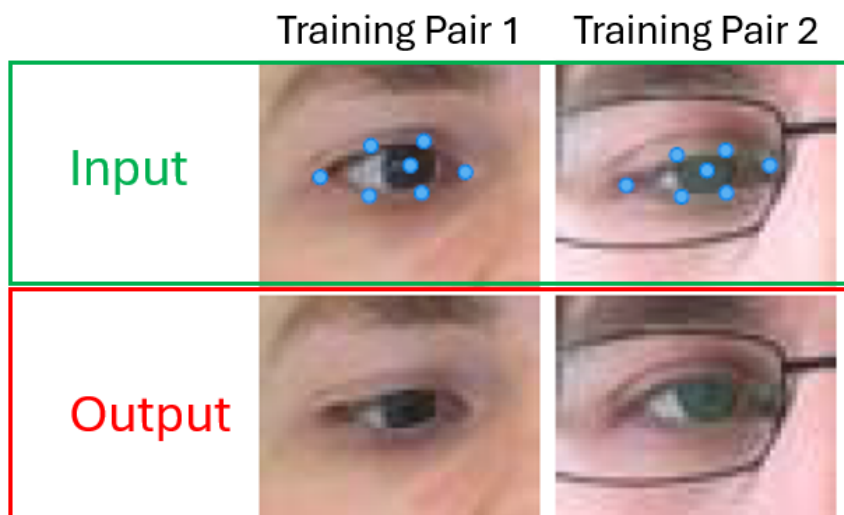
84. The two-step model (shown in the blue box in both Ganin and the above illustration) operates this way in both training and application at run-time. During training, the model receives the before images (in green) and the after images (in red).

This training data allows the model to train itself yielding an optimized set of parameters that can be utilized to produce an “after” image for a given input through the use of multiple steps. When training is complete, the model may be used in run- time. In the run- time scenario, the model receives an input stream with a before image (in green) and uses its optimized set of parameters to create an after image (in red) that will ultimately be displayed to a user. In this two-step model, there is one set of training data, one training to complete one task, one error measurement, and thus one machine learning algorithm.

85. **Ganin’s Training Dataset Confirms One MLA:** One way to confirm that Ganin discloses, at best, a single algorithm is because Ganin does not disclose (1) training data representing the output of the coarse module so that the coarse module can train itself as a first machine learning algorithm or (2) training data representing the input of the fine module so that the fine module can train itself as a second machine learning algorithm. Without these separate and distinct training sets, Ganin’s fine module cannot train itself as a second machine learning algorithm because it does not have the proper training data to take a half-processed, intermediary product from the coarse module and convert it into a final “after” image. Accordingly, Ganin’s two-step model does not utilize two machine learning algorithms as claimed in the ’985 Patent.

86. This conclusion is supported by Ganin’s description of the training data. For example, Ganin describes the training data collection process in Figure 4. That figure shows a training pair for two people (shown below) where Ganin explains that the top

image is an “input image with superimposed feature points on top” and the bottom image is the “output image.” EX-1006 at Fig. 4.



87. Ganin describes the training data as a training “pair”—meaning there is only a single input image (green) and a single output image (red). EX-1006 at Fig. 4; *id.* at p. 10 (“The model is trained on **pairs** of images corresponding to eye appearance **before and after** the redirection.”) (emphasis added); *id.* at p. 11 (“At training time, our dataset allows us to mine **pairs** of images . . . The head pose, the lighting, and all other nuisance parameters are [] the same between the **two images in the pair**. . . . The resulting image **pair** serves as a training example for the learning procedure (Figure 4-Right).”). There is no coarse input, coarse output, fine input, and fine output. Otherwise, (1) we would see them in Figure 4 and the related description so that the training data could be used to train each of the coarse module and the fine module and (2) it would be a “quad” not a “pair.” Moreover, Ganin’s training procedure shows that the training process seeks to optimize the error between the “synthesized output  $O_{\text{output}}$  and the ground-truth  $O_{\text{gt}}$ .” EX-1006 at

p.14. As Ganin only describes one synthesized output and one ground truth, there is no opportunity to assess the error between a coarse output and a coarse ground truth, or to assess the error between a fine output and a fine ground truth. Ganin’s training datasets and procedures therefore confirm that it contains, at best, one machine learning algorithm and not two as Petitioner argues.

88. **Ganin’s Description Confirms One MLA:** Ganin’s descriptions throughout the publication show that it uses a single architecture, network, model, or algorithm, and not two machine learning algorithms. For instance, Ganin repeatedly refers to the architecture in the singular, and as a “feed-forward” or “end-to-end” architecture (emphasis added for all citations listed below):

- a) “To meet these challenges, we develop a new deep **feed-forward** architecture that combines several principles of operation (coarse-to-fine processing, image warping, intensity correction). The architecture is trained **end-to-end** in a supervised way . . .” EX-1006 at p. 2.
- b) “In this section, we discuss the architecture of our deep model for re-synthesis . . .” *Id.* at p. 4.
- c) “The proposed system takes an input eye region, feature points (anchors) as well as a correction angle  $\alpha$  and sends them to the multi-scale neural network (see Sect. 3.2) predicting a flow field.” *Id.* at Fig. 2.

- d) “The task of the network is therefore the prediction of the warping field. This field is predicted in two stages in a coarse-to-fine manner. . . .” *Id.* at p. 4.
- e) “The model was trained **end-to-end** on 128-sized batches using Adam optimizer [14]. We used a regular l2-distance between the synthesized output  $O_{\text{output}}$  and the ground-truth  $O_{\text{gt}}$  as the objective function. We tried to improve over this simple baseline in several ways.” *Id.* at p. 8.
- f) “We treat this problem as a specific instance of conditional image generation and suggest a new deep architecture that can handle this task very well as revealed by numerical comparison with prior art and a user study. . . . All these operations are performed in a **feed-forward** manner, and the parameters associated with different operations are **learned jointly in the end-to-end fashion**. After learning, the resulting neural network can synthesize images with manipulated gaze, while the redirection angle can be selected arbitrarily from a certain range and provided as an input to the network.” *Id.* at Abstract.
- g) “[T]he bulk of gaze redirection is accomplished via warping the input image (Fig. 2). The task of the network is therefore the prediction of the warping field. . . . All operations outlined above are implemented

in **a single feed-forward architecture** and are trained jointly **end-to-end**. We now provide more details on each stage[] of **the** procedure, starting with [a] more detailed description of the data used to train **the** architecture.” *Id.* at p. 4.

89. Hence, it is clear from the disclosures cited above that Ganin is referring to a single machine learning algorithm, not two machine learning algorithms as suggested by Petitioner.

90. Petitioner’s expert suggests that “single does not mean single neural network . . . Ganin never said single neural network.” EX-2009 (Karam Deposition Transcript) at 86:3-7. I disagree. Ganin repeatedly describes his system using singular terms like “a,” “the,” “end-to-end,” and “single” as noted above. Petitioner’s expert later concedes this point, explaining that if Ganin were to disclose a coarse MLA and fine MLA, “[y]ou have to give the ground truths there” for the coarse MLA and you would have to “estimate the ground truth” for the fine MLA. As she admits, “you have to account for both in the training because you would like them to estimate the network that would best match the ground truths.” *Id.* at 106:1-21.

91. The above is also supported by several disclosures by Ganin. To this effect, Ganin states that “After learning, the resulting neural network [*no “s” at the end of the word network*] can . . . as an input to the network [*again no “s” at the end of the word network*].” EX-1006 at Abstract. Ganin refers to its neural network in singular terms (there is no “s” on the end of the work network). This is also supported by Ganin’s

description of Figure 2 where Ganin again describes the neural network in single terms: “The proposed system takes as input ... multiscale neural network” (no “s” at the end of network).

92. The convolutional layers depicted in Figure 10(d) do not form a separate machine learning algorithm but are subcomponent of the same neural network proposed by Ganin and shown in Figure 10(a). EX-1006 at p. 18. The convolutional layers shown in Figure 10(d) represent the subcomponent labeled as “conv layers” shown in Figure 10(b) and 10(c) and therefore are part of the overall architecture. *Id.* They do not represent a second machine learning algorithm for the many reasons discussed above with regards to training data, cost function, etc.

93. In summary, not only do Ganin’s explicit disclosures show that it does not contain two machine learning algorithms, but even Petitioner’s reading of Ganin does not meet the ’985 Patent’s elements 1[E] and 18[E]. This is because Ganin’s fine module does not have the proper training data to train itself as a second machine learning algorithm. It cannot do so because Ganin’s training data is simply the before image and after image without any intermediary product images for training, as I will explain in more detail below. Ganin therefore does not disclose the second machine algorithm element as claimed in independent claims 1 and 18.

***b) Petitioner’s argument that Ganin’s end-to-end training could use the same training data is incorrect.***

94. Petitioner alleges that the 0.5x (coarse) and 1x (fine) scale warping modules

are two machine learning algorithms, but that is incorrect for all the reasons I discuss above. However, even if each module were a machine learning algorithm (again, I disagree), the separate trainings for Ganin’s coarse and fine modules would necessarily require different training data, meaning that Ganin does not teach using the same training data for both the first and second machine learning algorithms as required by the independent claims.

95. As I explained above, to train a machine learning algorithm in a supervised fashion as disclosed by Ganin (EX-1006 at p. 2 “The architecture is trained end-to-end in a supervised way . . .”), it needs a set of training data that includes “before” and “after” images. The algorithm then trains itself in order to take a “before” image and create the “after” image so that at run-time an unknown “before” image can be altered into an accurate “after” image. If Ganin’s coarse module were a separate machine learning algorithm, it would need training data specific to the task of creating a coarse output image, i.e., a training set of “before” images and coarse “after” images. And if Ganin’s fine module were also a separate machine learning algorithm, it would need training data to teach it to take the coarse module’s coarse “after” image and create a final “after” image. But this is not what Ganin discloses as I explain in the paragraphs below.

96. **Training Datasets:** Ganin discloses that its “architecture is trained end-to-end in a supervised way using a specially collected dataset [*referring to the dataset in a singular tense*] . . .” EX-1006 at p. 2. Subsequently, Ganin outlines its training datasets and procedure in EX-1006 at p. 8. In particular, Ganin states that “we record several

images with eyes looking in different fixed directions (about 200 for one video sequence) using a webcam mounted in the middle of the screen.” EX-1006 at p. 8. Ganin then states that “[t]raining pairs are collected, taking two images with different gaze directions from one sequence.” *Id.* Ganin’s Figure 4 illustrates the data collection process. Hence, Ganin’s training images provide an input image (top right) and its corresponding gaze corrected output image as shown in Figure 4 (bottom right). No other intermediate images are provided or disclosed by Ganin.

97. As such, Ganin does not provide any training type imagery specific to any of the modules that are contained upstream within its overall architecture or disclose any training methodologies that focus on specific sub-modules within its end-to-end architecture. In other words, Ganin does not provide a separate coarse type training image dataset, a separate fine type training image dataset, or any other type of imagery that would serve to separately train subsets of its architecture. Thus, although Ganin’s coarse module would require training data that would include the “before” image and the intermediate, half-processed coarse output, there is no such “coarse” training data. Similarly, although the fine module’s training data would require the intermediate, half-processed coarse output and the “after” image output, there is no such “fine” training data.

98. To summarize, even if Ganin’s teachings supported two machine learning algorithms, it would require two training sets, one for the coarse module and one for the fine module. However, claim elements 1[E] and 18[E] require that the “second machine

learning algorithm . . . us[e] **the same training data** as the first machine learning algorithm.” EX-1001 at 26:50-52; 29:58-60 (emphasis added). As I explained above, the coarse module’s training data and the fine module’s training data are not “the same training data.” This means that Ganin’s fine module cannot be the claimed “second machine learning algorithm” because it does not use “the same training data as the first machine learning algorithm” as required by the independent claims.

99. **Training Procedure:** The above arguments are further supported by the training procedure proposed by Ganin. Ganin states that “[t]he architecture is trained end-to-end in a supervised way” and that “[t]he model was trained **end-to-end**,” not in sections. EX-1006 at pp. 2, 8. Similarly, Ganin teaches using “a [single] **regular  $l_2$ -distance**,”<sup>11</sup> not separate ones for the coarse and fine modules. *Id.* In other words, Ganin provides only one “ $l_2$ -distance,” which is a measure of how good the algorithm was at taking the “before” image and creating the “after” image. *Id.* There is no coarse  $l_2$ -distance and no fine  $l_2$ -distance, such that there is no ability to assess the performance of Ganin’s coarse module or Ganin’s fine module. Ganin’s coarse and fine modules therefore cannot be two machine learning algorithms.

100. Moreover, a POSITA would have understood that Ganin’s synthesized output is the final output of the end-to-end architecture rather than that “the synthesized

---

<sup>1</sup> The  **$l_2$ -distance** is one of the most commonly used distances in machine learning. It represents a distance measure between, for instance, two vectors **x** and **y** in an  $N$ - dimensional space. Its use in machine learning is prevalent for quantifying the difference between an actual output computed using a given machine learning algorithm and the corresponding target (the ground truth) that is provided by the training dataset.

output  $O_{\text{output}}$  could refer to both coarse estimate  $O_{\text{coarse}}$  (output of the coarse warping module) and final output  $O$  (output of the fine warping module)” as claimed in the Petition. Petition at p. 16. This is supported by several disclosures in Ganin.

101. First, the training images of Ganin do not provide for any in-between images that would result from the output of the coarse module. Ganin’s training dataset does not disclose an input image, a corresponding output ground truth image at the output of the coarse module, and another separate ground truth output image as the final output. To the contrary, Ganin’s dataset provides an input image and its corresponding final output image as shown in Figure 4 without any images for any of the steps in-between. This is also supported by the images in Figure 3 where the images on the right of the figure provide the corresponding ground truth (see the column labeled “GT”) for each of “three challenging redirection cases.” *Id.* at Fig. 3, p. 7. This is also confirmed in Figure 7 at p. 12. None of the images displayed in the results in Ganin provide, discuss, or reference separate ground truth imagery for the coarse and fine processing modules.

102. Second, in its efforts to incorporate registration, Ganin discloses that the “results can be further perceptually improved [] if **the** objective is slightly modified to take into account misalignment between inputs and ground-truth images. To that end, we enlarge the bounding-box  $B$  that we use to extract the output image of a training pair by  $k = 3$  pixels in all the directions. Given that now  $O_{\text{gt}}$  has the size of  $(H + 2k) \times (W + 2k)$ , the **new objective** is defined as:

$$L(O_{\text{output}}, O_{\text{gt}}) = \min \text{dist} (O_{\text{output}}, O_{\text{gt}}[i : i + H, j : j + W]), \quad (7)$$

where  $\text{dist}(\cdot)$  can be either  $l_2$  or  $l_1$ -distance (the latter giving slightly sharper results), and  $O_{\text{gt}}[i : i+H, j : j+W]$  corresponds to a  $H \times W$  crop of  $O_{\text{gt}}$  with top left corner at the position  $(i, j)$ . Being an alternative to the offline registration of input/ground-truth pairs [] which is computationally prohibitive in large-scale scenarios, this small trick greatly increases robustness of the training procedure against small misalignments in a training set.” *Id.* at p. 13 (emphasis added). Note that Equation (7) provides an objective function calculation that is accomplished between the ground truth image  $O_{\text{gt}}$  that is available from the training dataset and the corresponding output image  $O_{\text{output}}$  that is calculated at the output of the Ganin architecture. There is no disclosure, no mention, and no mechanism for computing an objective function between the ground truth image  $O_{\text{gt}}$  (which only provides for a final output image) and any images that are calculated upstream in the architecture such as  $O_{\text{coarse}}$  as noted in the Petition. Petition at p. 16. Because Ganin does not disclose or provide any mechanism or corresponding ground truth imagery for computing an objective function upstream in its architecture (i.e. Ganin does not disclose separate objective functions for the coarse module and the fine module, and does not provide corresponding training datasets as explained earlier), Ganin supports one, not two, machine learning algorithms. For all the reasons discussed above, Ganin does not disclose a second machine learning algorithm, and therefore cannot anticipate the challenged independent claims, including elements 1[E] and 18[E].

103. Even if Ganin’s teachings supported two training sets (which it does not), one for the coarse module and one for the fine module, they would not be the same set of

training data as required by the independent claims. Claim elements 1[E] and 18[E] require that the “second machine learning algorithm . . . us[e] **the same training data as the first machine learning algorithm.**” EX-1001 at 26:50-52; 29:58-60. However, if Ganin uses a set of training data for the coarse module and a set of training data for the fine module (which it clearly does not) as Petitioner suggests, these training datasets would necessarily be different as follows. The coarse module’s training data would include the “before” image and the intermediate, half-processed coarse output identified by Petitioner’s expert as  $O_{\text{coarse}}$ . *See* EX-1003 at ¶ 116 (noting that the distance calculation would be calculated using  $O_{\text{coarse}}$ ). In contrast, the fine module’s training data would include the intermediate, half-processed coarse output and the “after” image output identified by Petitioner’s expert as  $O$ . *See* EX-1003 at ¶ 116 (noting that the distance calculation for the fine module would be calculated using  $O$ ). As a result, the coarse module’s training data and the fine module’s training data are not “the same training data.” This means that Ganin’s fine module cannot be the claimed “second machine learning algorithm” because it does not use “the same training data as the first machine learning algorithm” as required by the claims.

**B. Ground III: Claims 1-3, 5-6, 8-14, and 16-19 are not obvious over Chalom in view of Aslan.**

**1. Chalom and Aslan Cannot be Combined as Petitioner Suggests.**

***a) A POSITA would not have been motivated to combine Chalom and Aslan.***

104. All of the '985 Patent's independent claims require both a "first machine algorithm" and a "second machine algorithm." EX-1001 at 26:45-50; 28:55-57; 29:20-25; 29:53-58; 30:40-45. Chalom does not disclose a second machine learning algorithm. Petition at p. 56; *see also* EX-1003 at ¶ 209 (admitting that Chalom does not disclose "the training of two MLAs").

105. Petitioner incorrectly argues that a POSITA would allegedly be motivated to combine Chalom and Aslan to arrive at a two-algorithm system because of "accuracy and/or efficiency gains" in Chalom. Petition at p. 42. I disagree that a POSITA would be motivated to combine Chalom and Aslan, or that they could (or would) be combined as suggested by Petitioner.

106. First, nothing in Chalom indicates it is inaccurate or inefficient, so there is no reason to look to Aslan to fix this nonexistent problem. In fact, Chalom claims superior performance that is both accurate and efficient. Chalom states that "current techniques may not be fast enough . . . , or may not be robust enough in terms of quality." EX-1005 at 1:16-21. To overcome this problem, Chalom offers an "offline component" and "online component" to "provide improved corrected eye images, robustness . . . , real time

capability (e.g., faster operation), and flexibility.” *Id.* at 3:21-43. Chalom explains that its techniques “provide for high quality eye contact correction in real time” (*id.* at 9:50-55) and “provide a robust appearance model for an eye region” (*id.* at 13:37-40). As a result, Chalom does not suffer from the inaccuracies or inefficiencies suggested by Petitioner.

107. Second, even if a POSITA were to concede either issue (alleged inefficiency and/or inaccuracy) in Chalom, adding another machine learning algorithm as Petitioner describes in its Petition (Petition at pp. 41-43) does not solve either problem as I explain below.

***(1) A POSITA would not look to Aslan to add a second machine learning algorithm to Chalom’s system to improve its accuracy.***

108. To begin, adding another machine learning algorithm as described in the Petition does not yield higher accuracy. Petition at pp. 41-43. If Chalom’s machine learning algorithm is inaccurate, a POSITA would look to the training data and resolution of the classifier. For instance, if the training data is not of sufficient quantity or diversity, then Chalom’s algorithm will be inadequately trained and would, as a result, yield inaccurate results. Adding a second machine learning algorithm would not improve the training accuracy given the same training dataset. To the contrary, it will increase complexity, add more epochs<sup>2</sup> to the training process, and still provide similar, inaccurate

---

<sup>2</sup> In this context, the term “epoch” refers to one entire passing of training data through the algorithm.

results.

109. Petitioner's expert agrees that the training data must be sufficient before making any other efforts to improve the accuracy of a machine learning algorithm. For instance, Petitioner's expert agrees that "it's [a] no-brainer that you need a good training set to start that's representative of your task because – in addition to all of this stuff in here, so this is a no-brainer. We're assuming we have a good training dataset." EX-2009 at 99:13-19 (emphasis added).

110. Petitioner's expert then goes on to list several other options for improving MLA accuracy, but adding another machine learning algorithm is not among them. She explains that one option would be to "restart the training with different initialization of the weight or some hyperparameters change by tuning the model hyperparameters . . . or even adding more layers to the network." *Id.* at 98:3-7. Another option she mentioned was to change the architecture of the model because "the structure of the network is not really compatible [or] [m]aybe it's too small to learn it." *Id.* at 98:8-21. In summary, Petitioner's expert finds that a machine learning algorithm may function poorly "because you're not using the right objective function, you're not using the right hyperparameters, you're not using the right initialization. Maybe you're not using the right structure." *Id.* at 99:18-24. Therefore, the natural consequence is to address these issues, not add another machine learning algorithm to the problem.

111. Thus, a POSITA looking to improve the accuracy of Chalom's system would consider other simpler and well known technical choices such as: 1) increasing the size

of the training dataset (e.g. add more training examples—an approach known as data augmentation in the machine learning literature), 2) increasing the diversity of the training dataset (e.g. finer resolution of gaze angles in the training dataset—again data augmentation), 3) improving the random forest classifier performance by adding more branches (finer resolution) to its structure, and/or 4) changing the structure, objective function, or parameters of the machine learning algorithm. A POSITA would not naturally be led to add another machine learning algorithm to an already poorly functioning one.

112. To this end, Chalom states that “the pretrained classifier may be a pretrained random forest classifier. In such embodiments, each leaf of the tree of the random forest classifier may represent a likelihood map (e.g., a SAD map) for every pixel in the eye region and, at each branch of the tree, the training process may minimize the entropy between the likelihood maps that have arrived at the branch. The random forest classifier may include any suitable characteristics. In some embodiments, the random forest classifier may have about 4 to 8 trees having about 6 to 7 levels each.” EX-1005 at 4:60-5:2. Hence, to improve the accuracy of Chalom’s classifier, a POSITA would seek to add more levels/leaves to the currently existing random forest classifier to further improve its accuracy rather than introduce a whole new second machine learning random forest classifier algorithm that would need to be properly designed, trained, and structured to interact with the currently existing machine learning classifier. Thus, a POSITA would not be motivated to combine Chalom with Aslan to increase accuracy.

**(2) *A POSITA using Chalom's system would not seek to add a second machine learning algorithm to improve the efficiency of the system.***

113. A POSITA working with Chalom's system and looking to further improve its efficiency would not seek to add a second machine learning algorithm to the training and/or implementation phases of Chalom as proposed by the Petition (Petition at pp. 41-47) since this will increase complexity and consequently reduce efficiency. Chalom teaches that the camera may be above, below, left, or right of the display, and thus it may appear that the user is looking below, above, right, or left of the camera. EX-1005 at 5:37-50. The gaze correction will be different depending on which direction the user is looking, so Chalom explains that the random forest classifier can choose a model that is particular to the appropriate camera orientation. *Id.* For example, if the camera is located above the display, the user will be looking down and thus the gaze correction will be to lift the gaze upward. In contrast, if the camera is to the left of the display, the user will be looking right and thus the gaze correction will be to shift the gaze left. The most efficient way to deploy Chalom's random forest models is to select the most appropriate one based on the camera's location rather than adding more machine learning algorithms such that each model for each camera location is run by a separate machine learning algorithm in parallel. Selecting the most appropriate model for the camera's location, which is exactly what Chalom teaches, would provide Chalom's system with a more efficient operation, prevent a significant increase in complexity that would result from adding a second machine

learning algorithm, maintain the required accuracy and still operate in real time as taught by Chalom.

114. Further, a POSITA looking to compress Chalom's random forest classifier (RFC) would not do so by adding a second machine learning algorithm as alleged by Petitioner. On the contrary, a POSITA would seek to: (1) reduce the size of the RFC (i.e. less trees, less branches, etc.); (2) utilize the most probable trees/branches; or (3) create a smaller RFC for proper gaze adjustment.

115. A properly designed RFC as taught by Chalom would serve POSITA's purposes without the need for adding another machine learning algorithm. For example, Chalom teaches that once the RFC is optimized, it "may be compressed using any suitable technique or techniques" EX-1005 at 15:4-7. A POSITA, using Chalom's methodology, would follow Chalom's advice and thereby create a more compressed version of the already optimized RFC by selecting the most probable trees and/or branches rather than embarking on an entirely new second machine learning algorithm that would require training, optimization, etc. to achieve the above.

***b) Petitioner's expert reasoning to look to Aslan because of its  
"limited resource" environment is incorrect.***

116. Petitioner's expert states that a POSITA would combine Chalom with Aslan "because Chalom is not considering limited resources and low power." EX-2009 at 154:12-155:4; *see also id.* at 156:6-19 (testifying that "Chalom . . . [d]id not mention about lower power; did not mention limited resources . . . But if you want limited

resources and low power, it requires different techniques which Aslan solves . . .”); *id.* at 159:2-12 (testifying that Chalom “can support [real time] but here there is no limit on resources . . . So it can support real-time implementation given the right resources. There’s no limit on resources. But if you want to limit to low power and have limited resources for mobile, then you have to seek to improve on Chalom in order to have that.”).

117. This is wrong. Chalom does consider limited resources, and explicitly discloses high-quality, real-time performance in a mobile device, battery-operated environment. For instance, Chalom discloses that its system “may be implemented as a mobile computing device [] having wireless capabilities. A mobile computing device may refer to any device having a processing system and a mobile power source or supply, such as one or more batteries, for example.” EX-1005 at 22:40-45; *id.* at 19:1-14; 3:4-10 (explaining that “video conferencing applications [may operate] on laptops or other devices such as mobile devices . . .”). Chalom gives examples of mobile computing devices, including a “cellular telephone, combination cellular telephone/PDA, smart device, . . . and so forth.” *Id.* at 22:46-55. Chalom also states that mobile computing devices “include computers that are arranged to be worn by a person, such as [] wrist computers . . . or other wearable computers.” *Id.* at 22:56-61. Chalom therefore discloses a high-quality, real-time machine learning algorithm that performs well in a limited power, mobile device scenario such that Petitioner’s expert assumptions are incorrect. For this reason, there is no motivation to look to Aslan for combination with Chalom.

*c) Chalom and Aslan could not or would not be combined as  
Petitioner suggests.*

118. Petitioner argues that a POSITA would be “motivated to implement [Aslan’s] similar techniques for Chalom’s image-editing context.” Petition at p. 43. The alleged reasoning is that “Chalom’s training process teaches all the possible motion vector fields to its MLAs, and these possible motion vector fields, although they will be used for image editing, are classes/labels themselves.” Petition at p. 43. This is incorrect. Motion vector fields do not represent classes/labels and thus a POSITA could not combine Chalom and Aslan as Petitioner suggests.

119. On the contrary, motion vector fields represent two-dimensional displacement vectors (i.e. displacement in the x-direction  $dx$  and displacement in the y-direction  $dy$ ) for a pixel located at  $(i,j)$  in given frame. See EX-2010 at p. 76. In other words, these motion vectors are intended to provide the appropriate amount of displacement in the x and y directions for the pixel located at  $(i,j)$ . In addition, motion vectors are typically expressed via decimal numbers (i.e.  $dx = 13.5$  pixels and  $dy = 12.6$  pixels) and usually require the use of interpolation techniques (such as bilinear interpolation) to fit a pixel grid. However, classes/labels are usually integers (such as a number or index) and represent one dimension. Petitioner’s expert agrees. EX-2009 at 199:8-200:19. Hence, motion vector fields do not represent classes or labels and cannot form a basis or way of combining Chalom and Aslan.

120. Finally, Petitioner alleges that “[d]uring runtime, finding the optimal motion vector field corresponds to finding the best class/label and therefore involves a classification process.” Petition at p. 43. Given that motion vectors are not classes or labels as outlined above, finding the optimal motion vectors during runtime does not correspond to finding the best class/labels and does not represent a classification process. Petitioner’s argument amounts to nothing more than attempting to fit a square peg in a round hole.

**2. The Combination of Chalom and Aslan Does Not Render Claims 1-3, 5-6, 8-14, and 16-19 Obvious.**

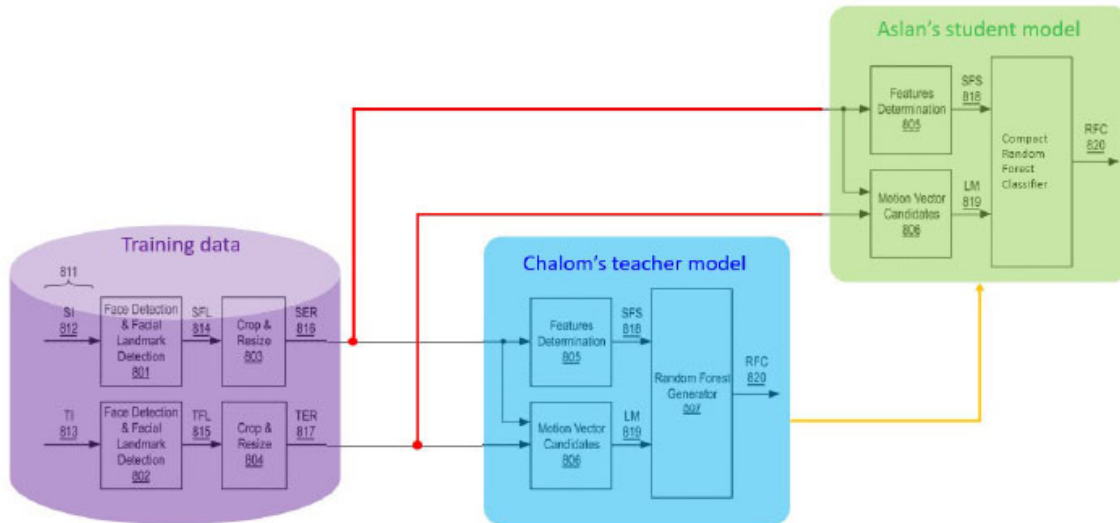
*a) The combination of Chalom and Aslan does not disclose a second machine learning algorithm that uses “first reference data [that] compris[es] [] editing instructions” as required by claim elements 1[E,F], 16[E,F], 17[D], 18[D], and 19[D].*

121. The challenged independent claims all require a second machine learning algorithm that uses the first reference data’s editing instructions from the first machine learning algorithm. EX-1001 at 26:45-54; 28:57-61; 29:20-29; 29:53- 30:2; 30:40-49. The combination of Chalom and Aslan does not disclose this claimed feature, nor has Petitioner described what that combination is with sufficient particularity.

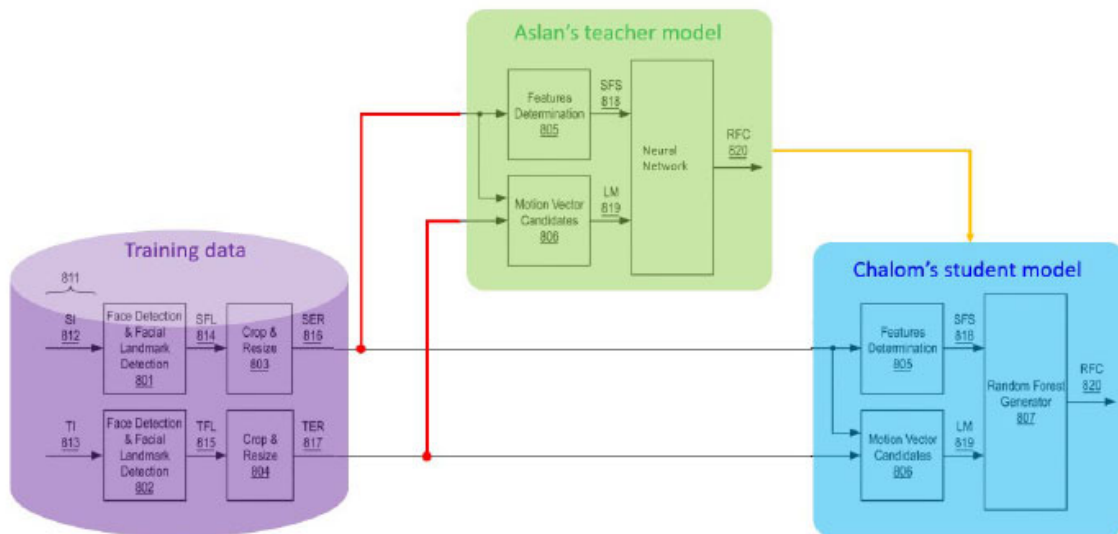
122. To begin with, Petitioner and its expert do not provide a thorough technical explanation to show how the combination of Chalom and Aslan produces first reference data that comprises editing instructions as required by all of the ’985 Patent’s independent

claims. Petitioner summarily suggests that Aslan can be added to supplement Chalom's disclosure in one of two ways below:

### Option 1:



### Option 2:



123. However, Petitioner provides little explanation of how this combination would actually work beyond these color-coded diagrams. This is explained in the following paragraphs with regards to each of the above options.

124. Option 1: For option one, the Petition simply states that a POSITA would “us[e] Chalom’s already trained RFC as a teacher to train a more compact RFC with fewer and/or less complex decision trees.” Petition at p. 44. Thereafter a POSITA “need[] only follow the instructions provided in Aslan . . . [to] feed the same training data (purple) and the output of Chalom’s random forest (orange) to a student model, e.g., a smaller or more compact random forest (green).” *Id.* at p. 45.

125. However, the Petition does not provide any technical description or details with regards to the proposed combination, which results in the Petition being unclear and ambiguous, rather than sufficiently detailing how the output of Chalom’s random forest would interact with a student model. For instance, the Petition does not identify what output is portrayed with the orange arrow shown in the figures above. Is this the output of the random forest classifier RFC 820? If so, the Petition does not so state, nor does the Petition state what that output contains. Nor does the Petition state how or where the undescribed output fits in Aslan’s student model (the green box). Nor does the Petition state how the output interacts with Aslan’s student model, or how it yields “a smaller or more compact random forest (green)” student model.

126. Further, as seen in the figure for Option 1 above, both Chalom’s teacher model and Aslan’s student model have the same construction with regards to the Features

Determination (box 805) and Motion Vector Candidates (box 806). As such, there is no basis to conclude that Aslan's student model would yield "a smaller or more compact random forest (green)" classifier than the teacher model of Chalom. Rather, this could have been done directly by Chalom's teacher model since it has access to the same input as Aslan's student model.

127. In summary, the Petition lacks sufficient technical details necessary to understand Petitioner's proposed combination of Chalom/Aslan.

128. Option 2: For option 2, the Petition suggests a POSITA would have "train[ed] a new accurate teacher model first, which is then used to train Chalom's RFC as a student model." *Id.* at 45. Again, according to the Petition, a POSITA need only follow Aslan's instructions. *Id.* at p. 47. The Petition is silent about what the orange arrow actually includes, except to say that it is the "output of Chalom's random forest" and the claimed "first editing instructions" used to train the student. *Id.* at p. 58.

129. In short, Petitioner does not articulate how the random forest output of Chalom's Figure 8 (RFC 820) integrates into the student model of Aslan (102), what the exact output of the student and teacher models would be, and how these outputs correspond to the first reference data (which comprises first editing instructions and/or first selection instructions) and the second reference data (which comprises second editing instructions and/or second selection instructions) as defined in the claims. To the extent Petitioner suggests this combination discloses the claimed editing instructions, it is a product of hindsight rather than the teachings of the references themselves.

130. Moreover, if a POSITA were to follow Aslan's instructions as Petitioner argues, the combination of Chalom and Aslan would result in a system that provides an "objective function" between the teacher and student models, and Petitioner has neither suggested nor proven that the "objective function" bears any resemblance to the claimed "first editing instructions."

131. Petitioner argues that the combination of Chalom and Aslan passes "information" from the teacher machine learning model to the student machine learning model, and that "information" must be the claimed first reference data. Petition at pp. 62-63. This is incorrect. Chalom makes no teaching about what passes between two machine learning algorithms because it does not have two machine learning algorithms. And, as noted above, Petitioner states only that a POSITA "would have needed only to follow the instructions provided in Aslan" to arrive at the inventions claimed in the '985 Patent. *Id.* at 45; *id.* at 47. This means that a POSITA would have adopted Aslan's teacher-student architecture as-is into the alleged combination. Yet the information passed between models in Aslan is within the realm of the objective function. EX-1007 at ¶ 28 (explaining that "'passing information' between machine learning models [] comprises formulating an objective function."); *id.* at ¶ 39 (explaining that "passing information between the models 100 and 102 [occurs] via the objective function"). Thus, a POSITA would adopt Aslan's "objective function" that is employed between the teacher and student models.

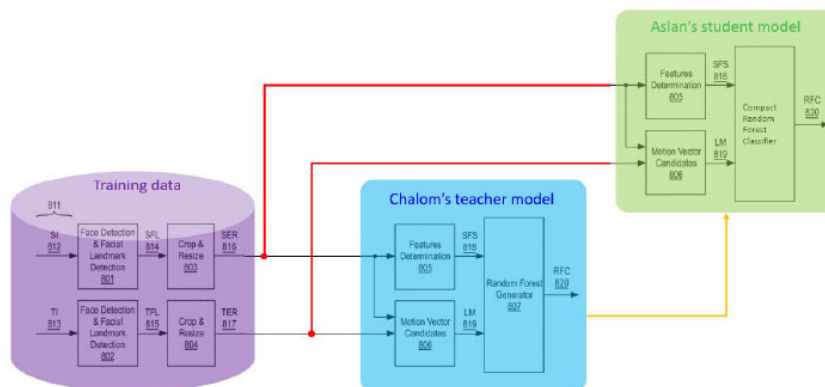
132. Petitioner's expert, Dr. Karam, describes an objective function as a way to

“assess” how well the algorithm is learning its task, i.e., a “performance measure.” EX-1003 at ¶ 23. Dr. Karam describes the objective function as including a “cross- entropy loss function,” or an “ $l_2$ -distance” that determines the error between the training data’s desired output and the actual output generated by the machine learning model. *Id.* at ¶ 28. Aslan also equates its “objective function” with error or “distance”. *See, e.g.*, EX-1007 at ¶ 29 (stating that the model can be biased “via a penalty (or distance) term in the objective function”); *id.* at ¶ 31 (stating that “[t]he objective function . . . can include a penalty term (sometimes called a ‘distance term’) that optimizes the objective function.”).

133. However, Petitioner explains that “editing instructions [are] for adjusting the digital representation of the head region.” Petition at p. 16. Petitioner has failed to allege or explain how a performance measure of how well the models are learning during training could be interpreted as editing instructions that adjust the digital representation of the head region during run-time. Further, neither Chalom nor Aslan discloses editing instructions and selection instructions as an outcome of the training process. The combination of Chalom and Aslan therefore does not render obvious the claimed second machine learning algorithm that “us[es] . . . the first reference data . . . [that] comprise first editing instructions for a range of possible configurations of the target feature.” EX-1001 at 26:45-54; 28:57-61; 29:20-29; 29:53-30:2; 30:40-49.

*b) The combination of Chalom and Aslan does not disclose first reference data and second reference data that are different from each other as required by claim elements 1[F], 16[F], 17[D], 18[D], and 19[D].*

134. As explained above, The Petition lacks sufficient technical details to explain its position that it would be obvious to combine Chalom and Aslan, such that integrating them would produce the desired results as claimed in the '985 Patent. Instead, the Petition provides only high-level conclusory remarks. Dr. Karam's declaration merely repeats the same high-level descriptions, again without any technical details. For instance, as already described above, Petitioner's figure merely places an orange arrow pointing from Chalom's teacher model (blue) to Aslan's Student model (green) without providing detail as to what the orange arrow allegedly represents nor how or where it would integrate into Aslan:



135. Further, both the Chalom (blue) and Aslan (green) models shown in the figure utilize the same Features Determination 805 and Motion Vector Candidates 806 modules; and both modules receive the same inputs from the training data (purple) as shown in the above figure. Based on this, the Random Forest Generator 807 would yield the same output from both models. In short, when using the same inputs, both the teacher model and student model would provide the same outputs.

136. The Petition alleges that the models' outputs are the claimed "reference data." Additionally, Petitioner's expert stated in her deposition that the first and second reference data could be the same. EX-2009 at 120:7-14. However, this position is nonsensical. If the "first reference data" were the same as the "second reference data", then there is no need to distinguish between "first" and "second"—the '985 Patent would merely disclose "reference data". Instead, the '985 claims require that the "reference data" from each of the first machine learning algorithm and second machine learning algorithm be different (hence the use of "first" and "second"). Further, the '985 Patent extensively discloses the use, advantages, and benefits of two machine learning algorithms that utilize two different reference data (i.e., the first reference data and second reference data) that comprise two different selection instructions and two different corresponding editing instructions. See, e.g., EX-1001 at 2:17-50. The differences and advantages of using MLA1 versus MLA2 are also clearly articulated in the '985 Patent (*see id.* at 22:14-33) which describes MLA1 as providing "more highly detailed reference data at the expense of higher computational demands, relative to the second machine

learning algorithm MLA2” (*id.* at 22:14-17) and that “only a subset of the input data ID ... are used to update the second machine learning algorithm MLA2...” *Id.* at 22:25-29. Thus, the first reference data (comprising first selection instructions and editing instructions) are designed to be distinctively different from the second reference data (comprising second selection instructions and editing instructions), for the purpose of striking the proper balance between accuracy and efficiency.

137. The combination of Chalom and Aslan would not result in first reference data and second reference data that are different from each other. Instead, as explained above, the “outputs” of both models would be the same, rather than different. As such, the combination of Chalom and Aslan would not yield a first algorithm generating “first reference data” and a second algorithm generating “second reference data” as required by the ’985 Patent’s independent claims.

## **XI. CONCLUSION**

138. As I explained above, in my opinion, all of the challenged claims (1-14 and 16-19) are patentable over the asserted Grounds raised by Petitioner.

I hereby declare that under the laws of the United States of America, and under penalty of perjury, that all statements made of my own knowledge are true, that all statements made on information and belief are believed to be true, and that the foregoing is true and correct. I further declare that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment,

or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of this proceeding.

Respectfully submitted,

Date: March 13, 2024

By: 

Place: Webster, NY