Curriculum Vitae

Geoffrey E. Hinton

January 6, 2025

Citizenship: Canadian (also British)

Address: Department of Computer Science
University of Toronto
10 Kings College Road
Toronto, Ontario, M5S 3G4
Canada
Email: hinton@cs.toronto.edu

Higher Education and Qualifications

1967 - 1970   Cambridge University, B.A. Hons (Experimental Psychology)

1972 - 1975   Edinburgh University, PhD. in Artificial Intelligence (awarded 1978)

Professional Experience

Jan 76 - Sept 78	Research Fellow
	Cognitive Studies Program, Sussex University, England
Oct 78 - Sept 80	Visiting Scholar
	Program in Cognitive Science, University of California, San Diego
Oct 80 - Sept 82	Scientific Officer
	MRC Applied Psychology Unit, Cambridge, England
Jan 82 - June 82	Visiting Assistant Professor
	Psychology Department, University of California, San Diego
Oct 82 - June 87	Assistant Professor then Associate Professor
	Computer Science Department, Carnegie-Mellon University, Pittsburgh, USA
Jul 87 - June 98	Professor
	Computer Science Department, University of Toronto, Canada
Jul 98 - Sep 01	Founding Director of the Gatsby Computational Neuroscience Unit
	University College London, England
Oct 2001 - Dec 2013	Professor
	Computer Science Department, University of Toronto, Canada
Jan 2014 -	Emeritus Professor
	Computer Science Department, University of Toronto, Canada
Mar 2013 - Sep 16	Distinguished Researcher, Google (half-time).
Oct 2016 - 2023	VP and Engineering Fellow, Google (half-time).
Jan 2017 -	Chief Scientific Adviser, Vector Institute (*pro bono*)

Professional Recognition

Fellowships

2023 Honorary Foreign Member of the US National Academy of Sciences
2016 Honorary Foreign Member of the US National Academy of Engineering
2015 Honorary Foreign Member of the Spanish Real Academia de Ingenieria
2014 Distinguished Fellow, Canadian Institute for Advanced Research
2003 Honorary Foreign Member of the American Academy of Arts and Sciences
2003 Fellow of the Cognitive Science Society
1998 Fellow of the Royal Society
1996 Fellow of the Royal Society of Canada
1991 Fellow, Association for the Advancement of Artificial Intelligence
1987 Fellow, Canadian Institute for Advanced Research (1987-1998; 2004-2014)

Awards

2025 King Charles III Coronation Medal
2024 The Nobel Prize in Physics (jointly with John Hopfield)
2024 The Vinfuture Grand Prize (jointly with Jensen Huang, Fei-Fei Li, Yoshua Bengio and Yann LeCun)
2024 The Ulysses Medal, University College Dublin
2023 The Royal Medal of the Royal Society
2022 The Princess of Asturias award (jointly with Yoshua Bengio, Demis Hassabis and Yann LeCun)
2019 The ACM A. M. Turing Award (jointly with Yoshua Bengio and Yann LeCun)
2019 The Honda Prize
2019 Toronto Region Builder Award
2018 Companion of the Order of Canada (Canada's highest honour)
2017 BBVA Foundation Frontiers of Knowledge Award
2016 The NEC C&C Award
2016 IEEE/RSE James Clerk Maxwell Gold Medal
2014 IEEE Frank Rosenblatt Medal
2012 Killam Prize in Engineering
2010 Gerhard Herzberg Canada Gold Medal
2005 IJCAI Research Excellence Award
2001 The David E. Rumelhart Prize
1998 IEEE Neural Networks Pioneer Award
1992 ITAC/NSERC award for academic excellence.
1990 IEEE Signal Processing Society Senior Award

Honorary Degrees

2022 Honorary Degree of Doctor of Science, University of Toronto
2013 Doctorat honorifique, University of Sherbrooke
2011 Honorary Degree of Doctor of Science, University of Sussex
2001 Honorary Degree of Doctor of Science, University of Edinburgh

Top N lists

2023 Toronto's most influential person, *Toronto Life Magazine*
2019 Toronto's 50 most influential people, *Toronto Life Magazine*
2018 Toronto's 50 most influential people, *Toronto Life Magazine*
2017 The Bloomberg 50, *Bloomberg Business week*
2017 The 50 most powerful people in Canadian business, *The Globe and Mail Report on Business*
2017 Toronto's 50 most influential people, *Toronto Life Magazine*
2016 The WIRED 100 - 2016's most influential people, *Wired Magazine*

Named Lectures

2024: The Romanes Lecture
2019: Honda Prize Lecture
2019: Royal Society of Edinburgh President's Lecture
2019: ACM Turing Lecture
2014: Dertousos Lecture, MIT
2012: Killam Prize Lecture, McGill
2011: The Foundation Lecture, Royal Canadian Institute
2010: The Hans-Lukas Teuber Lecture, MIT
2010: The "Big Thinkers" lecture, Yahoo, San Jose
2010: The Rockwood Memorial Lecture, UC San Diego
2009: The Ed Posner Lecture, NIPS-09, Vancouver
2009: The Ian Howard Lecture, York University
2006: The Graham Lecture, University of Toronto
2003: The Pinkel Lecture, University of Pennsylvania
2001: The David E. Rumelhart Prize Lecture, Edinburgh
1998: The Rockwood Memorial Lecture, UC San Diego
1995: The Rockwood Memorial Lecture, UC San Diego
1993: The Herzberg Lecture, Ottawa.
1993: The Broadbent Lecture, London, UK.
1992: The Benjamin Meaker Lectures, Bristol University (5 lectures).
1991: The St Andrews Easter Lectures, St Andrews (6 lectures).
1989: The fourth annual Hebb lecture, Dalhousie University, Halifax.
1989: The Sun Annual Lectures, University of Manchester (8 lectures)
1987: The Weigand Lecture, University of Toronto
1986: The David Marr Memorial Lecture, Kings College Cambridge

Recent Media Appearances

Since early in 2023, I have made many television appearances warning about the various risks of AI. Here is a sample of them.

```
CBS 60 Minutes, Sept 2023
https://www.youtube.com/watch?v=qrvK_KuIeJk

CNN Amanpour and Company, May 2023
 https://www.youtube.com/watch?v=Y6Sgp7y178k&t=935s

PBS, May 2023
https://www.youtube.com/watch?v=yAgQWnD31nE

BBC News, May 2023
href="https://www.youtube.com/watch?v=DsBGaHywRhs

CNN Jake Tapper, May 2023
 https://www.youtube.com/watch?v=FAbsoxQtUwM&t=3s

CBC The National, May 2023
href="https://www.youtube.com/watch?v=CkTUgOOa3n8&t=99s

CBS Morning News, March 2023
 https://www.youtube.com/watch?v=qpoRO378qRY&t=13s
```

PUBLICATIONS

Refereed Journal Papers

1. Yoshua Bengio, Geoffrey Hinton, Andrew Yao, Dawn Song, Pieter Abbeel, Trevor Darrell, Yuval Noah Harari, Ya-Qin Zhang, Lan Xue, Shai Shalev-Shwartz, Gillian Hadfield, Jeff Clune, Tegan Maharaj, Frank Hutter, Atilim Gne Baydin, Sheila Mcilraith, Qiqi Gao, Ashwin Acharya, David Krueger, Anca Dragan, Philip Torr, Stuart Russell, Daniel Kahneman, Jan Brauner , and Sren Mindermann (2024) Managing extreme AI risks amid rapid progress *Science*, **384**(6698), 842-845

2. Hinton, G. E. (2022) How to represent part-whole hierarchies in a neural network *Neural Computation*, 1-40.

3. Bengio, Y., LeCun, Y. and Hinton, G. E. (2021) Deep Learning for AI *Communications of the ACM*, 64(7), 58-65.

4. Lillicrap, T. P., Santoro, A., Marris, C. J,. Akerman, C., and Hinton, G. E. (2020) Backpropagation and the Brain *Nature Reviews Neuroscience*, **21**, pp 335–346.

5. LeCun, Y., Bengio, Y. and Hinton, G. E. (2015) Deep Learning *Nature*, **521**, pp 436-444.

6. Srivastava, N., Hinton, G. E., Krizhevsky, K., Sutskever, I. and Salakhutdinov. R. (2014) Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, **15(1)**, pp 1929-1958

7. Sarikaya, R., Hinton, G. E. and Deoras, P. (2014) Application of Deep Belief Networks for Natural Language Understanding. *IEEE/ACM Transactions on Audio, Speech & Language Processing*, **22.4**, pp 778-784.

8. Hinton, G. E. (2014) Where do features come from? *Cognitive Science*, **38(6)**, 1078-1101.

9. Ranzato, M., Mnih, V., Susskind, J. and Hinton, G. E. (2013) Modeling Natural Images Using Gated MRFs. *IEEE Trans. Pattern Analysis and Machine Intelligence*, **35:9**, pp 2206-2222.

10. Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T., and Kingsbury, B. (2012) Deep Neural Networks for Acoustic Modeling in Speech Recognition. *IEEE Signal Processing Magazine*, **29:6**, pp 82-97.

11. Salakhutdinov, R. R. and Hinton, G. E. (2012) An Efficient Learning Procedure for Deep Boltzmann Machines. *Neural Computation*, **24**, pp 1967-2006.

12. van der Maaten, L. J. P. and Hinton, G. E. (2012) Visualizing Non-Metric Similarities in Multiple Maps. *Machine Learning*, **87**, pp 33-55.

13. Mohamed, A., Dahl, G. and Hinton, G. E. (2012) Acoustic Modeling using Deep Belief Networks. *IEEE Transactions on Audio, Speech, and Language Processing*, **20**, pp 14-22.

14. Taylor, G. W, Hinton, G. E., and Roweis, S. (2011) Two distributed-state models for generating high-dimensional time series *Journal of Machine Learning Research*, **12**, pp 863-907.

15. Hinton, G. E. and Salakhutdinov, R. (2011) Discovering Binary Codes for Fast Document Retrieval by Learning Deep Generative Models. *Topics in Cognitive Science*, **3:1**, pp 74-91.

16. Schmah, T., Yourganov, G., Zemel, R. S., Hinton, G. E., Small, S. l., and Strother, S. C. (2010) Comparing Classification Methods for Longitudinal fMRI Studies *Neural Computation*, **22**, pp 2729-2762.

17. Memisevic, R. and Hinton, G. E. (2010) Learning to represent spatial transformations with factored higher-order Boltzmann machines. *Neural Computation*, **22**, pp 1473-1492.

18. Hinton, G. E. (2010) Learning to represent visual input. *Philosophical Transactions of the Royal Society, B.* **365**, pp 177-184.

19. Sutskever, I. and Hinton, G. E. (2010) Temporal Kernel Recurrent Neural Networks *Neural Networks*, **23**, pp 239-243

20. Salakhutdinov, R. and Hinton, G. E. (2009) Semantic Hashing. *International Journal of Approximate Reasoning*, **50**, pp 969-978.

21. Mnih, A., Yuecheng, Z., and Hinton, G. E. (2009) Improving a statistical language model through non-linear prediction. *NeuroComputing*, **72**, pp 1414-1418.

22. van der Maaten, L. J. P. and Hinton, G. E. (2008) Visualizing Data using t-SNE. *Journal of Machine Learning Research*, **9**(Nov) pp 2579-2605.

23. Sutskever, I. and Hinton, G. E. (2008) Deep Narrow Sigmoid Belief Networks are Universal Approximators *Neural Computation*, **20**, pp 2629-2636.

24. Hinton, G. E. (2007) Learning multiple layers of representation. *Trends in Cognitive Science*, **11**, pp 428-434.

25. Hinton, G. E. and Salakhutdinov, R. (2006) Non-linear dimensionality reduction using neural networks. *Science*, **313**, pp 504-507, July 28 2006.

26. Hinton, G. E., Osindero, S., Welling, M. and Teh, Y. (2006) Unsupervised discovery of non-linear structure using contrastive back-propagation. *Cognitive Science*, **30**, (4), pp 725-731.

27. Hinton, G. E., Osindero, S. and Teh, Y. (2006) A fast learning algorithm for deep belief nets. *Neural Computation*, **18**, pp 1527-1554.

28. Osindero, S., Welling, M. and Hinton G. E. (2006) Topographic Product Models Applied To Natural Scene Statistics. *Neural Computation*, **18**, pp 381-414.

29. Memisevic, R. and Hinton, G. E. (2005) Improving dimensionality reduction with spectral gradient descent. *Neural Networks*, **18**, pp 702-710.

30. Sallans, B and Hinton, G. E. (2004) Reinforcement Learning with Factored States and Actions. *Journal of Machine Learning Research*, **5** pp 1063–1088.

31. Welling, M., Zemel, R. and Hinton, G. E. (2004) Probabilistic sequential independent components analysis. *IEEE Transactions on Neural Networks*, **15**, pp 838-849.

32. Teh, Y. W, Welling, M., Osindero, S. and Hinton G. E. (2003) Energy-Based Models for Sparse Overcomplete Representations. *Journal of Machine Learning Research*, **4**, pp 1235-1260.

33. Friston, K.J., Penny, W., Phillips, C., Kiebel, S., Hinton, G. E., and Ashburner, J. (2002) Classical and Bayesian Inference in Neuroimaging: Theory. *NeuroImage*, **16**, pp 465-483.

34. Hinton, G. E.(2002) Training Products of Experts by Minimizing Contrastive Divergence. *Neural Computation*, **14**, pp 1771-1800.

35. Mayraz, G. and Hinton, G. E. (2001) Recognizing hand-written digits using hierarchical products of experts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24**, pp 189-197.

36. Paccanaro, A., and Hinton, G. E. (2000) Learning distributed representations of concepts from relational data using linear relational embedding. *IEEE Transactions on Knowledge and Data Engineering*, **13**, 232-245.

37. Ueda, N. Nakano, R., Ghahramani, Z. and Hinton, G. E. (2000) SMEM Algorithm for Mixture Models. *Neural Computation*, **12**, 2109-2128.

38. Ghahramani, Z. and Hinton, G.E. (2000) Variational Learning for Switching State-space Models. *Neural Computation*, **12**, 831-864.

39. Ueda, N. Nakano, R., Ghahramani, Z. and Hinton, G. E. (1999) Split and Merge EM Algorithm for Improving Gaussian Mixture Density Estimates. *Journal of VLSI Signal Processing Systems*, **26**, 133-140.

40. Frey, B. J., and Hinton, G. E. (1999) Variational Learning in Non-linear Gaussian Belief Networks. *Neural Computation*, **11**, 193-214.

41. Ennis M, Hinton G, Naylor D, Revow M, Tibshirani R. (1998) A comparison of statistical learning methods on the GUSTO database. *Statistics in Medicine*, **17** 2501-2508.

42. Tibshirani, R. and Hinton, G.E. (1998) Coaching variables for regression and classification. *Statistics and Computing*, **8**, 25-33.

43. de Sa, V. R. and Hinton, G. E. (1998) Cascaded Redundancy Reduction. *Network: Computation in Neural Systems*, **9**, 73-84.

44. Fels, S. S. and Hinton, G. E. (1997) Glove-TalkII: A neural network interface which maps gestures to parallel formant speech synthesizer controls. *IEEE Transactions on Neural Networks*, **8**, 977-984.

45. Hinton, G. E. and Ghahramani, Z. (1997) Generative Models for Discovering Sparse Distributed Representations. Philosophical Transactions of the Royal Society, B. **352**, 1177-1190.

46. Frey, B. J., and Hinton, G. E. (1997) Efficient stochastic source coding and an Application to a Bayesian Network Source Model. *The Computer Journal*, **40 (2)**.

47. Hinton, G. E., Dayan, P. and Revow M. (1997) Modeling the manifold of images of handwritten digits. *IEEE Transactions on Neural Networks*, **8**, 65-74.

48. Williams, C. K. I., Revow, M. and Hinton, G. E. (1997) Instantiating deformable models with a neural net. *Computer Vision and Image Understanding*. **68**, 120-126

49. Dayan, P. and Hinton, G. E. (1997) Using Expectation-Maximization for Reinforcement Learning. *Neural Computation*, **9**, 271-278.

50. Oore, S., Hinton, G. E. and Dudek, G. (1997) A mobile robot that learns its place. *Neural Computation*, **9**, 683-699.

51. Dayan, P. and Hinton, G. E. (1996) Varieties of Helmholtz Machine. *Neural Networks*, **9**, 1385-1403.

52. Revow, M., Williams, C. K. I. and Hinton, G. E. (1996) Using Generative Models for Handwritten Digit Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**, 592-606.

53. Dayan, P., Hinton, G. E., Neal, R., and Zemel, R. S. (1995) Helmholtz Machines. *Neural Computation*, bf 7, 1022-1037.

54. Hinton, G. E., Dayan, P., Frey, B. J. and Neal, R. (1995) The wake-sleep algorithm for self-organizing neural networks. *Science*, **268**, pp 1158-1161.

55. Zemel, R. S. and Hinton, G. E. (1995) Learning Population Codes by Minimizing Description Length *Neural Computation*, **7**, 549-564.

56. Becker, S. and Hinton, G. E. (1993) Learning mixture models of spatial coherence. *Neural Computation*, **5**, 267-277.

57. Nowlan. S. J. and Hinton, G. E. (1993) A soft decision-directed LMS algorithm for blind equalization. *IEEE Transactions on Communications*, **41**, 275-279.

58. Fels, S. S. and Hinton, G. E. (1992) Glove-Talk: A neural network interface between a data-glove and a speech synthesizer. *IEEE Transactions on Neural Networks*, **3**.

59. Becker, S. and Hinton, G. E. (1992) A self-organizing neural network that discovers surfaces in random-dot stereograms. *Nature*, **355:6356**, 161-163.

60. Nowlan. S. J. and Hinton, G. E. (1992) Simplifying neural networks by soft weight sharing. *Neural Computation*, **4**, 173-193.

61. Jacobs, R., Jordan, M. I., Nowlan. S. J. and Hinton, G. E. (1991) Adaptive mixtures of local experts. *Neural Computation*, **3**, 79-87.

62. Hinton, G. E. and Shallice, T. (1991) Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review* **98**, 74-95.

63. Hinton, G. E. (1990) Mapping part-whole hierarchies into connectionist networks. *Artificial Intelligence*, **46**, 47-75.

64. Hinton, G. E. and Nowlan, S. J. (1990) The bootstrap Widrow-Hoff rule as a cluster-formation algorithm. *Neural Computation*, **2**, 355-362.

65. Lang, K., Waibel, A. and Hinton, G. E. (1990) A Time-Delay Neural Network Architecture for Isolated Word Recognition. *Neural Networks*, **3**, 23-43.

66. Hinton, G. E. (1989) Connectionist learning procedures. *Artificial Intelligence*, **40**, 185-234.

67. Waibel, A. Hanazawa, T. Hinton, G. Shikano, K. and Lang, K. (1989) Phoneme Recognition Using Time-Delay Neural Networks. *IEEE Acoustics Speech and Signal Processing*, **37**, 328-339.

68. Hinton, G. E. (1989) Deterministic Boltzmann learning performs steepest descent in weight-space. *Neural Computation*, **1**, 143-150.

69. Touretzky, D. S. and Hinton, G. E. (1988) A distributed connectionist production system. *Cognitive Science*, **12**, 423-466.

70. Hinton, G. E. and Parsons, L. A. (1988) Scene-based and viewer-centered representations for comparing shapes. *Cognition*, **30**, 1–35.

71. Hinton, G. E. (1987) The horizontal-vertical delusion. *Perception*, **16**.

72. Plaut, D. C. and Hinton, G. E. (1987) Learning sets of filters using back-propagation. *Computer Speech and Language*, **2**, 35–61.

73. Hinton, G. E. and Nowlan, S. J. (1987) How learning can guide evolution. *Complex Systems*, **1**, 495–502.

74. Fahlman, S. E. and Hinton, G. E. (1987) Connectionist architectures for Artificial Intelligence. *IEEE Computer*, **20**, 100–109.

75. Sejnowski, T. J., Kienker, P. K., and Hinton, G. E. (1986) Learning symmetry groups with hidden units: Beyond the perceptron. *Physica D*, **22**, 260–275.

76. Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986) Learning representations by back-propagating errors. *Nature*, **323**, 533–536.

77. Kienker, P. K., Sejnowski, T. J., Hinton, G. E., and Schumacher, L. E. (1986) Separating figure from ground with a parallel network. *Perception*, **15**, 197–216.

78. Ackley, D. H., Hinton, G. E., and Sejnowski, T. J. (1985) A learning algorithm for Boltzmann machines. *Cognitive Science*, **9**, 147–169.

79. Hutchins, E. L. and Hinton, G. E. (1984) Why the islands move. *Perception*, **13**, 629–632.

80. Hinton, G. E. (1984) Parallel computations for controlling an arm. *The Journal of Motor Behavior*, **16**, 171–194.

81. Ballard, D. H., Hinton, G. E., and Sejnowski, T. J. (1983) Parallel visual computation. *Nature*, **306**, 21–26.

82. Hinton, G. E. (1979) Some demonstrations of the effects of structural descriptions in mental imagery. *Cognitive Science*, **3**, 231-250.

83. Hinton, G. E. (1978) Respectively reconsidered. *Pragmatics Microfiche*, May issue.

Refereed Conference Papers

84. Ren, M., Kornblith, S., Liao, R., and Hinton, G. (2023) Scaling Forward Gradient With Local Losses *ICLR* arXiv preprint arXiv:2210.03310

85. Agarwal, R., Melnick, L., Frosst, N., Zhang, X., Lengerich, B., Caruana, R., and Hinton, G. E. (2021) Neural additive models:Interpretable machine learning with neural nets *Advances in Neural Information Processing Systems, 34*, 4699-4711.

86. Sun, W., Tagliasacchi, A., Deng, B., Sabour, S., Yazdani, S., Hinton, G. E., Yi, K. M. (2021) Canonical Capsules: Unsupervised Capsules in Canonical Pose *Advances in Neural Information Processing Systems, 34*. arXiv preprint arXiv:2012.04718

87. Deng, B., Genova, K., Soroosh Yazdani, S., Sofien Bouaziz, S., Geoffrey Hinton, G. and Tagliasacchi, A. (2020) CvxNet: Learnable Convex Decomposition *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 31-44

88. Deng, B., Lewis, J. P., Jeruzalski, T., Pons-Moll, G., Hinton, G. E., Norouzi, M., Tagliasacchi, A. (2020) NASA: Neural Articulated Shape Approximation *ECCV*

89. Chen, T., Kornblith, S., Swersky, K., Norouzi, M., and Hinton, G. E. (2020) Big Self-Supervised Models are Strong Semi-Supervised Learners *Advances in Neural Information Processing Systems 34*

90. Chen, T., Kornblith, S., Norouzi, M., and Hinton, G. E. (2020) A Simple Framework for Contrastive Learning of Visual Representations *Proceedings of the 37th International Conference on Machine Learning* Eds. Hal Daume III and Aarti Singh, pp 1597–1607.

91. Qin, Y., Frosst, N., Sabour, S., Raffel, C., Cottrell, C. and Hinton, G. (2020) Detecting and Diagnosing Adversarial Images with Class-Conditional Capsule Reconstructions *ICLR-2020*

92. Kosiorek, A. R., Sabour, S., Teh, Y. W. and Hinton, G. E. (2019) Stacked Capsule Autoencoders *Advances in Neural Information Processing Systems 32*

93. Zhang, M., Lucas, J., Ba, J., and Hinton, G. E. (2019) Lookahead Optimizer: k steps forward, 1 step back *Advances in Neural Information Processing Systems 32*

94. Muller, R., Kornblith, S. and Hinton G. (2019) When Does Label Smoothing Help? *Advances in Neural Information Processing Systems 32*

95. Deng, B., Kornblith, S. and Hinton, G. (2019) Cerberus: A multi-headed derenderer. *3D Scene Understanding Workshop, CVPR 2019* arXiv preprint arXiv:1905.11940

96. Deng, B., Genova, K., Yazdani, S., Bouaziz, S., Hinton, G. and Tagliasacchi, A. (2019) Cvxnet: Learnable convex decomposition. *Perception as Generative Reasoning Workshop, NeurIPS 2019* arXiv preprint arXiv:1909.05736

97. Kornblith, S., Norouzi, M., Lee, H. and Hinton, G. (2019) Similarity of neural network representations revisited *ICML-2019*

98. Hinton, G. E., Sabour, S. and Frosst, N. (2018) Matrix Capsules with EM Routing *ICLR-2018*

99. Kiros, J. R., Chan, W. and Hinton, G. E. (2018) Illustrative Language Understanding: Large-Scale Visual Grounding with Image Search *ACL-2018*

100. Anil, R., Pereyra, G., Passos, A., Ormandi, R., Dahl, G. and Hinton, G. E. (2018) Large scale distributed neural network training through online distillation *ICLR-2018*

101. Guan, M. Y., Gulshan, V., Dai, A. M. and Hinton, G. E. (2018) Who Said What: Modeling Individual Labelers Improves Classification *AAAI-2018*

102. Sabour, S., Frosst, N. and Hinton, G. E. (2017) Dynamic Routing between Capsules *NIPS-2017*

103. Frosst, N. and Hinton, G. E. (2017) Distilling a Neural Network Into a Soft Decision Tree. Preprint at arXiv:1711.09784

104. Pereyra, G., Tucker, T., Chorowski, J., Kaiser, L. and Hinton, G. E. (2017) Regularizing neural networks by penalizing confident output distributions. Preprint at arXiv:1701.06548

105. Shazeer, N., Mirhoseini, A., Maziarz, K., Davis, A., Le, Q., Hinton, G., and Dean, J. (2017) Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. *NIPS-2017*, Preprint at arXiv:1701.06538

106. Ba, J. L., Hinton, G. E., Mnih, V., Leibo, J. Z. and Ionescu, C. (2016) Using Fast Weights to Attend to the Recent Past. *NIPS-2016*, Preprint at arXiv:1610.06258v2

107. Ba, J. L., Kiros, J. R. and Hinton, G. E. (2016) Layer normalization. *Deep Learning Symposium, NIPS-2016*, Preprint at arXiv:1607.06450

108. Ali Eslami, S. M., Nicolas Heess, N., Theophane Weber, T., Tassa, Y., Szepesvari, D., Kavukcuoglu, K. and Hinton, G. E. (2016) Attend, Infer, Repeat: Fast Scene Understanding with Generative Models. *NIPS-2016*, Preprint at arXiv:1603.08575v3

109. Hinton, G. E., Vinyals, O., and Dean, J. (2015) Distilling the knowledge in a neural network. *Workshop on Deep Learning, NIPS-2014*, Preprint at arXiv:1503.02531

110. Jaitly, N., Vanhoucke, V. and Hinton, G. E. (2014) Autoregressive product of multi-frame predictions can improve the accuracy of hybrid models. *Fifteenth Annual Conference of the International Speech Communication Association.*

111. Jaitly, N., and Hinton, G. E. (2013) Vocal Tract Length Perturbation (VTLP) improves speech recognition. *Proc. ICML Workshop on Deep Learning for Audio,Speech and Language Processing*, Atlanta, USA.

112. Srivastava, N., Salakhutdinov, R. R. and Hinton, G. E. (2013) Modeling Documents with a Deep Boltzmann Machine. *Uncertainty in Artificial Intelligence (UAI 2013)*

113. Graves, A., Mohamed, A. and Hinton, G. E. (2013) Speech Recognition with Deep Recurrent Neural Networks. *IEEE International Conference on Acoustic Speech and Signal Processing (ICASSP 2013)*, Vancouver.

114. Dahl, G. E., Sainath, T. N. and Hinton, G. E. (2013) Improving Deep Neural Networks for LVCSR Using Rectified Linear Units and Dropout. *IEEE International Conference on Acoustic Speech and Signal Processing (ICASSP 2013)*, Vancouver.

115. Zeiler, M. D., Ranzato, M., Monga, R., Mao, M., Yang, K., Le, Q.V., Nguyen, P., Senior, A., Vanhoucke, V., Dean, J. and Hinton, G. E. (2013) On Rectified Linear Units for Speech Processing. *IEEE International Conference on Acoustic Speech and Signal Processing (ICASSP 2013)*, Vancouver.

116. Deng, L., Hinton, G. E. and Kingsbury, B. (2013) New types of deep neural network learning for speech recognition and related applications: An overview *IEEE International Conference on Acoustic Speech and Signal Processing (ICASSP 2013)*, Vancouver.

117. Sutskever, I., Martens, J., Dahl, G. and Hinton, G. E. (2013) On the importance of momentum and initialization in deep learning. *International Conference on Machine Learning*, Atlanta, USA

118. Tang, Y., Salakhutdinov, R. R. and Hinton, G. E. (2013) Tensor Analyzers. *International Conference on Machine Learning*, Atlanta, USA

119. Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. R. (2012) Improving neural networks by preventing co-adaptation of feature detectors. http://arxiv.org/abs/1207.0580

120. Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012) ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing 25*, MIT Press, Cambridge, MA

121. Salakhutdinov, R. R. and Hinton, G. E. (2012) A Better Way to Pretrain Deep Boltzmann Machines. *Advances in Neural Information Processing 25*, MIT Press, Cambridge, MA

122. Tang, Y., Salakhutdinov, R. R. and Hinton, G. E. (2012) Deep Lambertian Networks. *International Conference on Machine Learning*,

123. Mnih, V. and Hinton, G. E. (2012) Learning to Label Aerial Images from Noisy Data. *International Conference on Machine Learning*,

124. Tang, Y., Salakhutdinov, R. R. and Hinton, G. E. (2012) Deep Mixtures of Factor Analysers. *International Conference on Machine Learning*,

125. Tang, Y., Salakhutdinov, R. R. and Hinton, G. E. (2012) Robust Boltzmann Machines for Recognition and Denoising. *IEEE Conference on Computer Vision and Pattern Recognition*,

126. Mohamed,A., Hinton, G. E. and Penn, G. (2012) Understanding how Deep Belief Networks perform acoustic modelling *ICASSP 2012*, Kyoto.

127. Jaitly, N. and Hinton, G. E. (2011) A new way to learn acoustic events. *Advances in Neural Information Processing Systems 24, Deep Learning workshop*, Grenada, Spain.

128. Mnih, V., Larochelle, H. and Hinton, G. (2011) Conditional Restricted Boltzmann Machines for Structured Output Prediction *Uncertainty in Artificial Intelligence, 2011.*

129. Hinton, G.E., Krizhevsky, A. and Wang, S. (2011) Transforming Auto-encoders. *ICANN-11: International Conference on Artificial Neural Networks*, Helsinki.

130. Suskever, I., Martens, J. and Hinton, G. E. (2011) Generating Text with Recurrent Neural Networks. *Proc. 28th International Conference on Machine Learning*, Seattle.

131. Ranzato, M., Susskind, J., Mnih, V. and Hinton, G. (2011) On deep generative models with applications to recognition. *IEEE Conference on Computer Vision and Pattern Recognition*

132. Susskind,J., Memisevic, R., Hinton, G. and Pollefeys, M. (2011) Modeling the joint density of two images under a variety of transformations. *IEEE Conference on Computer Vision and Pattern Recognition*

133. Hinton, G. E., Krizhevsky, A. and Wang, S. (2011) Transforming Auto-encoders. In T. Honkela *et. al.* (Eds.): *ICANN 2011, Part I*, LNCS 6791, pp. 44-51.

134. Jaitly, N. and Hinton, G. E. (2011) Learning a better Representation of Speech Sound Waves using Restricted Boltzmann Machines. *ICASSP 2011*, Prague.

135. Mohamed,A., Sainath, T., Dahl, G., Ramabhadran, B., Hinton, G. and Picheny, M. (2011) Deep Belief Networks using Discriminative Features for Phone Recognition. *ICASSP 2011*, Prague.

136. Sarikaya, R., Hinton, G. and Ramabhadran, B. (2011) Deep Belief Nets for Natural Language Call-Routing. *ICASSP 2011*, Prague.

137. Krizhevsky, A. and Hinton, G.E. (2011) Using Very Deep Autoencoders for Content-Based Image Retrieval In *European Symposium on Artificial Neural Networks ESANN-2011)*, Bruges, Belgium.

138. Deng, L., Seltzer, M., Yu, D., Acero, A., Mohamed A., and Hinton, G. E. (2010) Binary Coding of Speech Spectrograms Using a Deep Auto-encoder. *Interspeech 2010*, Makuhari, Chiba, Japan.

139. Ranzato, M., Mnih, V., and Hinton, G. E. (2010) How to generate realistic images using gated MRF's. *Advances in Neural Information Processing Systems 23*.

140. Dahl, G., Ranzato, M., Mohamed, A., Hinton, G. E. (2010) Phone Recognition with the Mean-Covariance Restricted Boltzmann Machine. *Advances in Neural Information Processing Systems 23*.

141. Larochelle, H. and Hinton, G. E. (2010) Learning to combine foveal glimpses with a third-order Boltzmann machine. *Advances in Neural Information Processing Systems 23*.

142. Memisevic, R., Zach, C., Hinton, G. E. and Pollefeys M. (2010) Gated Softmax Classification. *Advances in Neural Information Processing Systems 23*.

143. Ranzato, M. and Hinton, G. E. (2010) Modeling pixel means and covariances using factored third-order Boltzmann machines. *IEEE Conference on Computer Vision and Pattern Recognition*.

144. Taylor, G., Sigal, L., Fleet, D. and Hinton, G. E. (2010) Dynamic binary latent variable models for 3D human pose tracking. *IEEE Conference on Computer Vision and Pattern Recognition*.

145. Nair, V. and Hinton, G. E. (2010) Rectified linear units improve restricted Boltzmann machines. *Proc. 27th International Conference on Machine Learning*, Israel.

146. Ranzato, M., Krizhevsky, A. and Hinton, G. E. (2010) Factored 3-way restricted Boltzmann machines for modeling natural images. *Proc. Thirteenth International Conference on Artificial Intelligence and Statistics*, Sardinia.

147. Mnih, V. and Hinton, G. E. (2010) Learning to detect roads in high-resolution aerial images. *To appear in European Conference on Computer Vision.*

148. Mohamed, A. R. and Hinton, G. E. (2010) Phone recognition using restricted Boltzmann machines. *ICASSP-2010*

149. Mohamed, A. R., Dahl, G. and Hinton, G. E. (2009) Deep belief networks for phone recognition. *NIPS 22 workshop on deep learning for speech recognition*

150. Salakhutdinov, R. and Hinton, G. E. (2009) Replicated Softmax: An Undirected Topic Model. *Advances in Neural Information Processing Systems 22*, Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, and A. Culotta (Eds.), pp 1607-1614.

151. Nair, V. and Hinton, G. E. (2009) 3-D Object recognition with deep belief nets. *Advances in Neural Information Processing Systems 22*, Y. Bengio, D. Schuurmans, J. lafferty, C. K. I. Williams, and A. Culotta (Eds.), pp 1339-1347.

152. Palatucci, M, Pomerleau, D. A., Hinton, G. E. and Mitchell, T. (2009) Zero-Shot Learning with Semantic Output Codes. *Advances in Neural Information Processing Systems 22*, Y. Bengio, D. Schuurmans, J. lafferty, C. K. I. Williams, and A. Culotta (Eds.), pp 1410-1418.

153. Heess, N., Williams, C. K. I. and Hinton, G. E. (2009) Learning generative texture models with extended Fields-of-Experts. In *Proc. British Machine Vision Conf.*.

154. Taylor, G. W. and Hinton, G. E. (2009) Products of Hidden Markov Models: It Takes N¿1 to Tango. *Proc. of the 25th Conference on Uncertainty in Artificial Intelligence.*

155. Taylor, G. W. and Hinton, G. E. (2009) Factored Conditional Restricted Boltzmann Machines for Modeling Motion Style. *Proc. 26th International Conference on Machine Learning*, pp 1025-1032. Omnipress, Montreal, Quebec.

156. Tieleman, T. and Hinton, G. E. (2009) Using Fast Weights to Improve Persistent Contrastive Divergence. *Proc. 26th International Conference on Machine Learning*, pp 1033-1040. Omnipress, Montreal, Quebec.

157. Zeiler, M.D., Taylor, G.W., Troje, N.F. and Hinton, G.E. (2009) Modeling pigeon behaviour using a Conditional Restricted Boltzmann Machine. In *European Symposium on Artificial Neural Networks ESANN-2009)*.

158. Salakhutdinov, R. and Hinton, G. E. (2009) Deep Boltzmann Machines. In D. van Dyk and M. Welling (Eds.), *Proc. Twelfth International Conference on Artificial Intelligence and Statistics*, JMLR: W&CP 5, pp 448-455, Clearwater Beach, Florida, April 2009.

159. Mnih, A. and Hinton, G. E. (2009) A Scalable Hierarchical Distributed Language Model. *Advances in Neural Information Processing Systems 21*, MIT Press, Cambridge, MA
PLEASE NOTE: In 2009 NIPS changed from publishing in the year after the conference to publishing in the same year. So both NIPS22 and NIPS 21 were published in 2009

160. Nair, V. and Hinton, G. E. (2009) Implicit Mixtures of Restricted Boltzmann Machines. *Advances in Neural Information Processing Systems 21*, MIT Press, Cambridge, MA

161. Sutskever, I. and Hinton, G. E. (2009) Using matrices to model symbolic relationships. *Advances in Neural Information Processing Systems 21*, MIT Press, Cambridge, MA

162. Sutskever, I., Hinton, G. E. and Taylor, G. W. (2009) The Recurrent Temporal Restricted Boltzmann Machine. *Advances in Neural Information Processing Systems 21*, MIT Press, Cambridge, MA

163. Schmah, T., Hinton, G. E., Zemel, R., Small, S. and Strother, S. (2009) Generative versus Discriminative Training of RBM's for classification of fMRI images. *Advances in Neural Information Processing Systems 21*, MIT Press, Cambridge, MA

164. Nair, V., Susskind, J., and Hinton, G.E. (2008) Analysis-by-Synthesis by Learning to Invert Generative Black Boxes. *ICANN-08: International conference on Artificial Neural Networks*, Prague.

165. Yuecheng, Z, Mnih, A, and Hinton, G (2008) Improving a statistical language model by modulating the effects of context words. 16th *European Symposium on Artificial Neural Networks*, pages 493–498.

166. Osindero, S. and Hinton, G. E. (2008) Modeling image patches with a directed hierarchy of Markov random fields. *Advances in Neural Information Processing Systems 20*, J.C. Platt and D. Koller and Y. Singer and S. Roweis (eds.), MIT Press, Cambridge, MA

167. Salakhutdinov, R. and Hinton, G. E. (2008) Using Deep Belief Nets to Learn Covariance Kernels for Gaussian Processes. *Advances in Neural Information Processing Systems 20*, J.C. Platt and D. Koller and Y. Singer and S. Roweis (eds.), MIT Press, Cambridge, MA

168. Salakhutdinov R. R, and Hinton, G. E. (2007) Semantic Hashing. *Proceedings of the SIGIR Workshop on Information Retrieval and Applications of Graphical Models*, Amsterdam.

169. Memisevic R. F. and Hinton, G. E. (2007) Unsupervised learning of image transformations. *IEEE Conference on Computer Vision and Pattern Recognition* Pages: 508-515.

170. Mnih, A. and Hinton, G. E. (2007) Three New Graphical Models for Statistical Language Modelling *International Conference on Machine Learning*, Corvallis, Oregon.

171. Salakhutdinov, R., Mnih, A. and Hinton, G. E. (2007) Restricted Boltzmann Machines for Collaborative Filtering *International Conference on Machine Learning*, Corvallis, Oregon.

172. Salakhutdinov R.R, and Hinton, G. E. (2007) Learning a non-linear embedding by preserving class neighbourhood structure. (Meila, M. and Shen, X. eds), pp 409-416, *Proc. Eleventh International Conference on Artificial Intelligence and Statistics*, The Society for AI and Statistics, Puerto Rico.

173. Sutskever, I. and Hinton, G. E. (2007) Learning multilevel distributed representations for high-dimensional sequences. (Meila, M. and Shen, X. eds), *Proc. Eleventh International Conference on Artificial Intelligence and Statistics*, pp 544-551, The Society for AI and Statistics, Puerto Rico.

174. Cook, J. A., Sutskever, I., Mnih, A. and Hinton , G. E. (2007) Visualizing similarity data with a mixture of maps. (Meila, M. and Shen, X. eds), *Proc. Eleventh International Conference on Artificial Intelligence and Statistics*, pp 65-72, The Society for AI and Statistics, Puerto Rico.

175. Taylor, G. W., Hinton, G. E. and Roweis, S. (2007) Modeling human motion using binary latent variables. *Advances in Neural Information Processing Systems* **19** MIT Press, Cambridge, MA

176. Hinton, G. E. and Nair, V. (2006) Inferring motor programs from images of handwritten digits. *Advances in Neural Information Processing Systems* **18** pp 515-522, MIT Press, Cambridge, MA

177. Memisevic, R. and Hinton, G. E. (2005) Embedding via clustering: Using spectral information to guide dimensionality reduction. *IEEE International Joint Conference on Neural Networks (IJCNN 2005)* Pages: 3198-3203

178. Mnih, A. and Hinton. G. E. (2005) Learning Unreliable Constraints using Contrastive Divergence. *IJCNN 2005*

179. Carreira-Perpignan, M. A. and Hinton. G. E. (2005) On Contrastive Divergence Learning. *Artificial Intelligence and Statistics, 2005*, Barbados.

180. Hinton, G. E., Osindero, S. and Bao, K. (2005) Learning Causally Linked Markov Random Fields. *Artificial Intelligence and Statistics, 2005*, Barbados.

181. Welling, M,, Rosen-Zvi, M. and Hinton, G. E. (2005) Exponential Family Harmoniums with an Application to Information Retrieval. *Advances in Neural Information Processing Systems* **17** MIT Press, Cambridge, MA

182. Memisevic, R. and Hinton, G. E. (2005) Multiple Relational Embedding. *Advances in Neural Information Processing Systems* **17** MIT Press, Cambridge, MA

183. Goldberger, J., Roweis, S., Salakhutdinov, R and Hinton, G. E. (2005) Neighborhood Components Analysis *Advances in Neural Information Processing Systems* **17** MIT Press, Cambridge, MA

184. Bishop, C. M. Svensen, M. and Hinton, G. E. (2004) Distinguishing Text from Graphics in On-line Handwritten Ink. In Kimura, F. and Fujisawa, H. (eds.), *Proceedings Ninth International Workshop on Frontiers in Handwriting Recognition, IWFHR-9*, Tokyo, Japan, pp. 142147.

185. Hinton, G. E., Welling, M. and Mnih, A. (2004) Wormholes Improve Contrastive Divergence. *Advances in Neural Information Processing Systems* **16** pages 417-424. MIT Press, Cambridge, MA

186. Welling, M., Zemel, R. S., and Hinton, G. E. (2003) Efficient parametric projection pursuit density estimation. In *UAI-2003: 19th Conference on Uncertainty in Artificial Intelligence.*

187. Welling, M., Hinton, G. E. and Osindero, S. (2003) Learning Sparse Topographic Representations with Products of Student-t Distributions. *Advances in Neural Information Processing Systems* **15** MIT Press, Cambridge, MA

188. Welling, M., Zemel, R. and Hinton, G. E. (2003) Self-Supervised Boosting. *Advances in Neural Information Processing Systems* **15** MIT Press, Cambridge, MA

189. Hinton, G. E. and Roweis, S. (2003) Stochastic Neighbor Embedding. *Advances in Neural Information Processing Systems* **15** MIT Press, Cambridge, MA

190. Welling, M. and Hinton, G. E. (2002) A New Learning Algorithm for Mean Field Boltzmann Machines. *International Joint Conference on Neural Networks*, Madrid.

191. Oore, S., Terzopoulos, D. and Hinton, G. E. (2002) Local Physical Models for Interactive Character Animation. *Eurographics 2002*, **21**, Blackwell Publishers, Oxford.

192. Oore, S., Terzopoulos, D. and Hinton, G. E. (2002) A Desktop Input Device and Interface for Interactive 3D Character Animation. *Graphics Interface, to appear*

193. Roweis, S., Saul, L. and Hinton, G. E. (2002) Global Coordination of Local Linear Models *Advances in Neural Information Processing Systems* **14** MIT Press, Cambridge, MA

194. Paccanaro, A., and Hinton, G. E. (2002) Learning Hierarchical Structures with Linear Relational Embedding. *Advances in Neural Information Processing Systems* **14** MIT Press, Cambridge, MA

195. Brown, A. D. and Hinton, G. E. (2002) Relative Density Nets: A New Way to Combine Backpropagation with HMM's. *Advances in Neural Information Processing Systems* **14** MIT Press, Cambridge, MA

196. Paccanaro, A. and Hinton, G. E. (2001) Learning distributed representations of relational data using linear relational embedding. Proceedings of the 12th Italian Workshop on Neural Nets. WIRN VIETRI-2001.

197. Brown, A. D. and Hinton, G. E. (2001). Products of Hidden Markov Models. *Proceedings of Artificial Intelligence and Statistics 2001*

198. Mayraz, G. and Hinton, G. E. (2001) Recognizing Hand-Written Digits Using Hierarchical Products of Experts. *Advances in Neural Information Processing Systems* **13**. MIT Press, Cambridge, MA

199. Sallans, B. and Hinton, G. E. (2001) Using Free Energies to Represent Q-values in a Multiagent Reinforcement learning Task. *Advances in Neural Information Processing Systems* **13**. MIT Press, Cambridge, MA

200. Teh, Y. and Hinton, G. E. (2001) Rate-coded Restricted Boltzmann Machines for Face Recognition. *Advances in Neural Information Processing Systems* **13**. MIT Press, Cambridge, MA

201. Hinton, G. E. and Teh, Y. (2001) Discovering multiple constraints that are frequently approximately satisfied. In *Proceedings of the Seventh Conference on Uncertainty in Artificial Intelligence*, pp 227-234.

202. Paccanaro, A. and Hinton, G. E. (2000) Extracting Distributed Representations of Concepts and Relations from Positive and Negative Propositions. In *Proceedings of the International Joint Conference on Neural Networks, IJCNN 2000.*

203. Paccanaro, A. and Hinton, G. E. (2000) Learning Distributed Representations by Mapping Concepts and Relations into a Linear Space. In P. Langley (Ed.) *Proceedings of the Seventeenth International Conference on Machine Learning, ICML2000*, pp 711-718, Morgan Kaufmann Publishers, San Francisco.

204. Hinton, G.E., Ghahramani, Z. and Teh, Y.W. (2000) Learning to Parse Images. In S. A. Solla, T. K. Leen, K.-R. Muller, (Eds.) *Advances in Neural Information Processing Systems* **12**. Cambridge, MA: MIT Press.

205. Hinton, G. E. and Brown, A. (2000) Spiking Boltzmann Machines. In S. A. Solla, T. K. Leen, K.-R. Muller, (Eds.) *Advances in Neural Information Processing Systems* **12**. Cambridge, MA: MIT Press.

206. Ghahramani, Z., Korenberg, A., and Hinton, G.E. (1999) Scaling in a Hierarchical Unsupervised Network. *ICANN 99: Ninth international conference on Artificial Neural Networks*, Edinburgh.

207. Ueda, N., Nakano, R., Ghahramani, Z. and Hinton, G.E. (1999) SMEM Algorithm for Mixture Models. In M. S. Kearns, S. A. Solla, D. A. Cohn, (eds.) *Advances in Neural Information Processing Systems* **11**. Cambridge, MA: MIT Press.

208. Grzeszczuk, R., Terzopoulos, D., and Hinton, G. E. (1999) Fast Neural Network Emulation of Dynamical Systems for Computer Animation. In M. S. Kearns, S. A. Solla, D. A. Cohn, (eds.) *Advances in Neural Information Processing Systems* **11**. Cambridge, MA: MIT Press, pp 882-889.

209. Ueda, N., Nakano, R., Ghahramani, Z. and Hinton, G.E. (1999) Pattern classification using a mixture of factor analyzers. *IEEE Neural Networks for Signal Processing (NNSP99)*, pp. 525-533.

210. Ueda, N., Nakano, R., Ghahramani, Z. and Hinton, G.E. (1998) Split and Merge EM Algorithm for Improving Gaussian Mixture Density Estimates. *IEEE Neural Networks for Signal Processing (NNSP98)*, pp. 274-283.

211. Ghahramani, Z. and Hinton, G. E. (1998) Hierarchical Non-linear Factor Analysis and Topographic Maps. *Advances in Neural Information Processing Systems 10.* M. I. Jordan, M. J. Kearns, and S. A. Solla (Eds.) MIT Press: Cambridge, MA.

212. Grzeszczuk, R., Terzopoulos, D., and Hinton, G. E. (1998) NeuroAnimator: Fast Neural Network Emulation and Control of Physics-Based Models. *Proc. ACM SIGGRAPH-98, Computer Graphics Proceedings, Annual Conference Series*, pp 9-20.

213. Bishop, C. M., Hinton, G. E. and Strachan, I. D. G. (1997) GTM through time. *Proceedings IEE Fifth International Conference on Artificial Neural Networks.* pp 111–116. IEE, London.

214. Frey, B. J., and Hinton, G. E. (1996) Free energy coding. J. A. Storer and M. Cohn (Eds.), *Proceedings of the Data Compression Conference 1996*, IEEE Computer Society Press, Los Alamitos, CA.

215. Hinton, G. E. and Revow, M. (1996) Using Pairs of Data-Points to Define Splits for Decision Trees. *Advances in Neural Information Processing Systems 8.* D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo (Eds), pp 507-514. MIT Press, Cambridge MA.

216. Frey, B. J., Hinton, G. E. and Dayan, P. (1996) Does the wake-sleep algorithm learn good density estimators? *Advances in Neural Information Processing Systems 8.* D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo (Eds), pp 661-668. MIT Press, Cambridge MA.

217. Fels, S. S. and Hinton, G. E. (1995) GloveTalk: An adaptive interface that uses neural networks. *Proceedings of Computer Human Interface Conference.*

218. Xu, L., Jordan, M. I. and Hinton, G. E. (1995) An alternative model for mixtures of experts. *Advances in Neural Information Processing Systems 7.* G. Tesauro, D. S. Touretzky and T. K. Leen (Eds), pp 633-640 MIT Press, Cambridge MA.

219. Hinton, G. E., Revow, M. and Dayan P. (1995) Recognizing handwritten digits using mixtures of local models. *Advances in Neural Information Processing Systems 7.* G. Tesauro, D. S. Touretzky and T. K. Leen (Eds), pp 1015-1022 MIT Press, Cambridge MA.

220. Fels, S. S. and Hinton, G. E. (1995) GloveTalkII: Mapping hand gestures to speech using neural networks. *Advances in Neural Information Processing Systems 7.* G. Tesauro, D. S. Touretzky and T. K. Leen (Eds), pp 843-850 MIT Press, Cambridge MA.

221. Williams, C. K. I., Hinton, G. E. and Revow, M. (1995) Using a neural net to instantiate a deformable model. *Advances in Neural Information Processing Systems 7.* G. Tesauro, D. S. Touretzky and T. K. Leen (Eds), pp 965-972 MIT Press, Cambridge MA.

222. Zemel, R. S. and Hinton, G. E (1994) Developing Population Codes by Minimizing Description Length. *Advances in Neural Information Processing Systems 6.* J. D. Cowan, G. Tesauro and J. Alspector (Eds.), Morgan Kaufmann: San Mateo, CA.

223. Hinton, G. E. and Zemel, R. S. (1994) Autoencoders, Minimum Description Length, and Helmholtz Free Energy. *Advances in Neural Information Processing Systems 6.* J. D. Cowan, G. Tesauro and J. Alspector (Eds.), Morgan Kaufmann: San Mateo, CA.

224. Xu, L. Jordan, M. I. and Hinton, G. E. (1994) A modified gating network for the mixtures of experts architectures. *Proc. WCNN94*, San Diego, CA. vol. 2, pp. 405410.

225. Zemel, R. S. and Hinton, G. E. (1993) Developing Population Codes for Object Instantiation Parameters. *AAAI Fall Symposium Series: Machine Learning in Computer Vision* Raleigh, North Carolina USA.

226. Hinton, G. E. and van Camp, D. (1993) Keeping Neural Networks Simple by Minimizing the Description Length of the Weights. In: *Proceedings of COLT-93.*

227. Revow, M., Williams, C. K. I., and Hinton, G. E. (1993) Using mixtures of deformable models to capture variations in the shapes of hand-printed digits. *Third International Workshop on Frontiers of Handwriting Recognition.*

228. Dayan, P. and Hinton, G. E. (1993) Feudal reinforcement learning. *Advances in Neural Information Processing Systems 5.* S. J. Hanson, J. D. Cowan and C. L. Giles (Eds.), Morgan Kaufmann: San Mateo, CA.

229. Hinton, G. E., Williams, C. K. I., and Revow, M. (1992) Adaptive Elastic Models for Character Recognition. *Advances in Neural Information Processing Systems 4.* J. E. Moody, S. J. Hanson and R. P. Lippmann (Eds.), Morgan Kaufmann: San Mateo, CA.

230. Nowlan, S. J. and Hinton, G. E. (1992) Adaptive Soft Weight Tying Using Gaussian Mixtures. *Advances in Neural Information Processing Systems 4.* J. E. Moody, S. J. Hanson and R. P. Lippmann (Eds.), Morgan Kaufmann: San Mateo, CA.

231. Becker, S. and Hinton, G. E. (1992) Learning to make coherent predictions in domains with discontinuities. *Advances in Neural Information Processing Systems 4.* J. E. Moody, S. J. Hanson and R. P. Lippmann (Eds.), Morgan Kaufmann: San Mateo, CA.

232. Nowlan, S. J. and Hinton, G. E. (1991) Evaluation of a system of competing experts on a vowel recognition task. *Advances in Neural Information Processing Systems 3.* R. P. Lippmann, J. E. Moody, and D. S. Touretzky (Eds.), Morgan Kaufmann: San Mateo, CA.

233. Zemel, R. and Hinton, G. E. (1991) Discovering viewpoint-invariant relationships that characterize objects. *Advances in Neural Information Processing Systems 3.* R. P. Lippmann, J. E. Moody, and D. S. Touretzky (Eds.), Morgan Kaufmann: San Mateo, CA.

234. Galland, C. G. and Hinton, G. E. (1990) Deterministic Boltzmann Learning in Networks with Asymmetric Connectivity. In Touretzky, D. S., Elman, J. L., Sejnowski, T. J. and Hinton, G. E. (Eds.) Connectionist Models: Proceedings of the 1990 Connectionist Summer School. Morgan Kauffman: San Mateo, CA.

235. Williams, C. K. I. and Hinton, G. E. (1990) Mean field networks that learn to discriminate temporally distorted strings. In Touretzky, D. S., Elman, J. L., Sejnowski, T. J. and Hinton, G. E. (Eds.) Connectionist Models: Proceedings of the 1990 Connectionist Summer School. Morgan Kauffman: San Mateo, CA.

236. Fels, S. S. and Hinton, G. E. (1990) Building adaptive interfaces with neural networks: The Glove-Talk pilot study. In D. Daiper, D. Gilmore, G. Cockton and B Shackel (Eds.) *Proceedings of the IFIP TC 13 Third International Conference on Human-Computer Interaction*, pages 683-688, North-Holland: Amsterdam.

237. Lang, K. J. and Hinton, G. E. (1990) Dimensionality reduction and prior knowledge in E-set recognition. In Touretzky, D. S., (Ed.) *Advances in Neural Information Processing Systems 2*, Morgan Kaufmann: San Mateo, CA.

238. Zemel, R., Mozer, M. and Hinton, G. E. (1990) TRAFFIC: Recognizing objects using local reference frame transformations. In Touretzky, D. S., (Ed.) *Advances in Neural Information Processing Systems 2*, Morgan Kaufmann: San Mateo, CA.

239. Galland, C. C. and Hinton, G. E. (1990) Discovering higher-order features with mean field networks. In Touretzky, D. S., (Ed.) *Neural Information Processing Systems 2*, Morgan Kaufmann: San Mateo, CA.

240. Hinton, G. E. and Becker, S. (1990) An unsupervised learning procedure that discovers surfaces in random-dot stereograms. *Proceedings of the International Joint Conference on Neural Networks*, **Vol 1**, 218-222, Lawrence Erlbaum Associates, Hillsdale, NJ.

241. LeCun, Y., Galland, C. C., and Hinton, G. E. (1989) GEMINI: Gradient Estimation by Matrix Inversion after Noise Injection. In Touretzky, D. S., (Ed.) *Neural Information Processing Systems 1*, Morgan Kaufmann: San Mateo, CA.

242. Zemel, R. S., Mozer, M. C. and Hinton, G. E. (1988) TRAFFIC: A model of object recognition based on transformations of feature instances. In Touretzky, D. S., Hinton, G. E. and Sejnowski, T. J., editors, Proceedings of the 1988 Connectionist Summer School, Morgan Kauffman: Los Altos, CA.

243. Hinton, G. E. (1988) Representing part-whole hierarchies in connectionist networks. *Proceedings of the Tenth Annual Conference of the Cognitive Science Society*, Montreal, Canada.

244. Hinton, G. E. and McClelland, J. L. (1988) Learning representations by recirculation. In D. Z. Anderson, editor, *Neural Information Processing Systems*, pages 358–366, American Institute of Physics: New York.

245. Hinton, G. E. and Plaut, D. C. (1987) Using fast weights to deblur old memories. *Proceedings of the Ninth Annual Conference of the Cognitive Science Society*, Seattle, WA.

246. Hinton, G. E. (1987) Learning translation invariant recognition in a massively parallel network. In Goos, G. and Hartmanis, J., editors, *PARLE: Parallel Architectures and Languages Europe*, pages 1–13, Lecture Notes in Computer Science, Springer-Verlag, Berlin.

247. Hinton, G. E. (1986) Learning distributed representations of concepts. *Proceedings of the Eighth Annual Conference of the Cognitive Science Society*, Amherst, Mass.
Reprinted in Morris, R. G. M. editor, Parallel Distributed Processing: Implications for Psychology and Neurobiology, Oxford University Press, Oxford, UK.

248. Pearlmutter, B. A. and Hinton, G. E. (1986) G-maximization: An unsupervised learning procedure for discovering regularities. In Denker, J., editor, *Neural Networks for Computing: American Institute of Physics Conference Proceedings* **151**,pp. 333-338

249. Touretzky, D. S. and Hinton, G. E. (1985) Symbols among the neurons: Details of a connectionist inference architecture. *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, Los Angeles.

250. Hinton, G. E. and Lang, K. J. (1985) Shape recognition and illusory conjunctions. *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, Los Angeles, pp 252-259.

251. Szeliski, R. and Hinton, G. E. (1985) Solving random-dot stereograms using the heat equation. *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, San Francisco.

252. Hammond, N., Hinton, G., Barnard, P., Long, J., and Whitefield, A. (1984) Evaluating the interface of a document processor: A comparison of expert judgement and user observation. *Proceedings of the First IFIP Conference on Human-Computer Interaction*, North-Holland.

253. Hinton, G. E. and Sejnowski, T. J. (1983) Analyzing cooperative computation. *Proceedings of the Fifth Annual Conference of the Cognitive Science Society*, Rochester NY.

254. Hinton, G. E. and Sejnowski, T. J. (1983) Optimal perceptual inference. *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, Washington DC.

255. Fahlman, S. E., Hinton, G. E., and Sejnowski, T. J. (1983) Massively parallel architectures for A.I.: Netl, Thistle, and Boltzmann machines. *Proceedings of the National Conference on Artificial Intelligence*, Washington DC.

256. Hinton, G. E. (1981) Shape representation in parallel systems. *Proceedings of the Seventh International Joint Conference on Artificial Intelligence Vol 2*, Vancouver BC, Canada.

257. Hinton, G. E. (1981) A parallel computation that assigns canonical object-based frames of reference. *Proceedings of the Seventh International Joint Conference on Artificial Intelligence Vol 2*, Vancouver BC, Canada.

258. Hinton, G. E. (1981) The role of spatial working memory in shape perception. *Proceedings of the Third Annual Conference of the Cognitive Science Society*, Berkeley CA.

259. Sloman, A., Owen, D., Hinton, G., Birch, F., and O'Gorman, F. (1978) Representation and control in vision. *Proceedings of the A.I.S.B. Summer Conference*, Hamburg.

260. Hinton, G. E. (1976) Using relaxation to find a puppet. *Proceedings of the A.I.S.B. Summer Conference*, University of Edinburgh.

Invited Papers

261. Hinton, G. E. (2018) Deep learning: a technology with the potential to transform health care. *Journal of the American Medical Association* [published online August 30, 2018].JAMA. doi:10.1001/jama.2018.11100

262. Vinyals, O., Kaiser, L., Koo, T., Petrov, S., Sutskever, I. and Hinton, G.E. (2015) Grammar as a foreign language. http://arxiv.org/abs/1412.7449v3

263. Lee, Q. L., Jaitly, N. and Hinton, G. E. (2015) A simple way to initialize recurrent networks of rectified linear units. http://arxiv.org/abs/1504.00941

264. Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. R. (2012) Improving neural networks by preventing co-adaptation of feature detectors. http://arxiv.org/abs/1207.0580

265. Hinton, G. E. (2005) What kind of a graphical model is the brain? *International Joint Conference on Artificial Intelligence 2005*.

266. Hinton, G. E. (2003) The ups and downs of Hebb synapses. *Canadian Psychology*, **44**, pp 10-13.

267. Hinton, G. E., Welling, M., Teh, Y. W., and Osindero, S. (2001) A new view of ICA. In *ICA-2001*, San Diego, CA.

268. Hinton, G. E. and Brown, A. D. (2001) Training many small hidden markov models. WISP-2001 Workshop on Innovation in Speech Processing. *Proceeding of the Institute of Acoustics*, **23**, Part 3.

269. Hinton, G. E. (2000) Modelling High-Dimensional Data by Combining Simple Experts. *AAAI-2000: Seventeenth National Conference on Artificial Intelligence*, Austin, Texas.

270. Hinton, G. E. (1999) Products of Experts. *ICANN 99: Ninth International Conference on Artificial Neural Networks*, Edinburgh. 1-6. Institution of Electrical Engineers, London, UK.

271. R. Grzeszczuk, R. D. Terzopoulos, D., and Hinton, G. (1999) Fast Neural Network Emulation and Control of Dynamical Systems. *Proc. AAAI 1999 Spring Symposium Series: Hybrid Systems and AI: Modeling, Analysis and Control of Discrete + Continuous Systems*, Stanford, CA, March, 1999, 83–88.

272. Hinton, G. E. and Ghahramani, Z. (1997) Towards Neurally Plausible Bayesian Networks. *Proceedings of the 1997 International Conference on Neural Networks* Houston, Texas. *The paper was accidentally ommitted from the proceedings but the conference organizers distributed copies to the attendees.*

273. Hinton, G. E. and Frey, B. J. (1995) Using neural networks to monitor for rare failures. *Proceedings of the 37th Mechanical Working and Steel Processing Conference*, Hamilton, Ontario.

274. Hinton, G. E., Dayan, P., To, A. and Neal R. M. (1995) The Helmholtz Machine Through Time. *Artificial Neural Networks V: Proceedings of ICANN-95*, pp 483-490. Elsevier North-Holland.

275. Hinton, G. E., Dayan, P., Neal, R. M., and Zemel, R. S. (1994) Using Neural Networks to Learn Intractable Generative Models. *American Statistical Association, 1994 Proceedings of the Statistical Computing Section.*, American Statistical Association, Alexandria, VA.

276. Hinton, G. E., Plaut, D. C. and Shallice, T. (1993) Simulating Brain Damage *Scientific American, October Issue*

277. Hinton, G. E. and van Camp, D. (1993) Keeping Neural Networks Simple. In: *Artificial Neural Networks III: Proceedings of ICANN-93.* Elsevier North-Holland.

278. Hinton, G. E., Williams, C. K. I., and Revow, M. (1992) Combining Two Methods of Recognizing Hand-Printed Digits. *Artificial Neural Networks II: Proceedings of ICANN-92.* I. Aleksander and J. Taylor (Eds.), Elsevier North-Holland.

279. Hinton, G. E. (1992) How Neural Networks Learn from Experience. *Scientific American, September Issue*

280. McClelland, J. L., Rumelhart, D. E., and Hinton, G. E. (1987) Une nouvelle approche de la cognition: Le connexionnisme. *Débat*, **47**, Novembre - Decembre.

281. Hinton, G. E. (1987) Learning procedures that construct representations in neural networks. *Kagaku*, **57**, 228–237.

282. Hinton, G. E. and Sejnowski, T. J. (1985) Learning in Boltzmann Machines. *Cognitiva - Colloque Scientifique* Paris.

283. Hinton, G. E. (1985) Learning in parallel networks. *Byte*, April issue.

284. Hinton, G. E. and Sejnowski, T. J. (1984) Learning semantic features. In *Proceedings of the Sixth Annual Conference of the Cognitive Science Society*, Boulder, CO.

Book Chapters

285. Hinton, G. E. (2010) Deep Belief Networks In C. Sammut and G. Webb (eds.), *Encyclopedia of Machine Learning*, Springer.

286. Hinton, G. E. (2010) Boltzmann Machines In C. Sammut and G. Webb (eds.), *Encyclopedia of Machine Learning*, Springer.

287. Susskind, J.M., Hinton, G. E., Movellan, J.R., and Anderson, A.K.(2008) Generating Facial Expressions with Deep Belief Nets. In V. Kordic (ed.) *Affective Computing, Emotion Modelling, Synthesis and Recognition.* ARS Publishers.

288. Hinton, G. E. (2007) To recognize shapes, first learn to generate images. In P. Cisek, T. Drew and J. Kalaska (Eds.) *Computational Neuroscience: Theoretical insights into brain function.* Elsevier.

289. Hinton, G. E. and Brown, A. D. (2002) Learning to Use Spike Timing in a Restricted Boltzmann Machine. In R. P. N. Rao, B. A. Olshausen, and M. S. Lewicki (Eds.) *Probabilistic Models of the Brain.* MIT Press.

290. Hinton, G. E. Sallans, B. and Ghahramani, Z. (1998) Hierarchical Communities of Experts. In M. I. Jordan (Ed.) *Learning in Graphical Models.* Kluwer Academic Press.

291. Neal, R., and Hinton, G. E. (1998) A new view of the EM algorithm that justifies incremental and other variants. In M. I. Jordan (Ed.) *Learning in Graphical Models.* Kluwer Academic Press.

292. Frey, B. J. and Hinton, G. E. (1996) A simple algorithm that discovers efficient perceptual codes. In L. Harris and M. Jenkin (Eds) *Computational and Biological Mechanisms of Visual Coding*, Cambridge University press, New York.

293. Becker, S. and Hinton, G. E. (1995) Using Spatial Coherence as an Internal Teacher for a Neural Network. In Y. Chauvin and D. E. Rumelhart (Eds) Advances in back-propagation. Erlbaum, Hillsdale, NJ.

294. Williams, C. K. I., Revow, M. and Hinton, G. E. (1993) Hand-printed digit recognition using deformable models. In L. Harris and M. Jenkin (Eds) *Spatial Vision in Humans and Robots*, Cambridge University press, New York.

295. Hinton, G. E. and Becker, S. (1992) Using coherence assumptions to discover the underlying causes of the sensory input. In S. Davis (Ed.) *Connectionism: Theory and practice*, Oxford University Press, New York.

296. Hinton, G. E. (1991) The unity of consciousness: A connectionist account. In Kessen, Ortony & Craik (Eds.) Festschrift in honor of George Mandler. Erlbaum, Hillsdale, NJ.

297. Hinton, G. E. and Anderson, J. A. (1989) Introduction to the second edition. In Hinton, G. E. and Anderson, J. A, editors, *Parallel Models of Associative Memory* (second edition), Erlbaum, Hillsdale, NJ.

298. Touretzky, D. S. and Hinton, G. E. (1987) Pattern matching and variable binding in a stochastic neural network. In Davis, L., editor, *Genetic Algorithms and Simulated Annealing*, Pitman, London.

299. Hinton, G. E. Learning to recognize shapes in a parallel network. In Imbert, M., editor, *Proceedings of the 1986 Fyssen Conference*, (Since I sent the finished manuscript in 1986, I have been unable to discover what has happened to this proceedings).

300. Sejnowski, T. J. and Hinton, G. E. (1987) Separating figure from ground using a Boltzmann machine. In Arbib, M. and Hanson, A. R., editors, *Vision, Brain and Cooperative Computation*, MIT Press, Cambridge, MA.

301. Rumelhart, D. E., Smolensky, P., McClelland, J. L., and Hinton, G. E. (1986) Parallel distributed models of schemata and sequential thought processes. In McClelland, J. L. and Rumelhart, D. E., editors, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 2: Applications*, MIT Press, Cambridge, MA.

302. Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986) Learning internal representations by error propagation. In Rumelhart, D. E. and McClelland, J. L., editors, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations*, MIT Press, Cambridge, MA.

303. Rumelhart, D. E., Hinton, G. E., and McClelland, J. L. (1986) A general framework. In Rumelhart, D. E. and McClelland, J. L., editors, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations*, MIT Press, Cambridge, MA.

304. McClelland, J. L., Rumelhart, D. E., and Hinton, G. E. (1986) The appeal of parallel distributed processing. In Rumelhart, D. E. and McClelland, J. L., editors, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations*, MIT Press, Cambridge, MA.

305. Hinton, G. E. and Sejnowski, T. J. (1986) Learning and relearning in Boltzmann machines. In Rumelhart, D. E. and McClelland, J. L., editors, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations*, MIT Press, Cambridge, MA.

306. Hinton, G. E., McClelland, J. L., and Rumelhart, D. E. (1986) Distributed representations. In Rumelhart, D. E. and McClelland, J. L., editors, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations*, MIT Press, Cambridge, MA.

307. Hinton, G. E. (1984) Some computational solutions to Bernstein's problems. In Whiting, H., editor, *Human Motor Actions: Bernstein Reassessed*, North-Holland, New York.

308. Hinton, G. E. and Parsons, L. A. (1981) Frames of reference and mental imagery. In Long, J. and Baddeley, A., editors, *Attention and Performance IX*, Erlbaum, Hillsdale, NJ.

309. Hinton, G. E. (1981) Implementing semantic networks in parallel hardware. In Hinton, G. E. and Anderson, J. A., editors, *Parallel Models of Associative Memory*, Erlbaum, Hillsdale, NJ.

310. Anderson, J. A. and Hinton, G. E. (1981) Models of information processing in the brain. In Hinton, G. E. and Anderson, J. A, editors, *Parallel Models of Associative Memory*, Erlbaum, Hillsdale, NJ.

Books

311. Hinton, G. E. and Sejnowski, T. J. (Editors) *Unsupervised Learning: Foundations of Neural Computation*. MIT Press, Cambridge, Massachusetts, 1999.

312. Hinton G. E. (Ed.) Connectionist Symbol Processing. 1990 Special issue of the journal *Artificial Intelligence* issued as a book by MIT press in 1991.

313. Touretzky, D. S., Elman, J., Sejnowski, T. J. and Hinton, G. E. (Eds.) Proceedings of the 1990 Connectionist Models Summer School, Morgan Kauffman: Los Altos, CA, 1990.

314. Touretzky, D. S., Hinton, G. E. and Sejnowski, T. J. (Eds.) Proceedings of the 1988 Connectionist Models Summer School, Morgan Kauffman: Los Altos, CA, 1988.

315. Hinton, G. E. and Anderson, J. A. (Eds.) *Parallel Models of Associative Memory* Hillsdale, NJ: Erlbaum, 1981.
(Updated second edition, 1989).

Technical Reports

*Technical reports that were subsequently published as papers or chapters are marked with a * and are not numbered.*

316. Chen, T., Zhang, R., and Hinton, G. (2023) Analog bits: Generating discrete data using diffusion models with self-conditioning arXiv preprint arXiv:2208.04202

317. Chen, T., Saxena, S., Li, L., Lin, T. Y., Fleet, D. J., and Hinton, G. (2022) A unified sequence interface for vision tasks arXiv preprint arXiv:2206.07669

318. Chen, T., Li, L., Saxena, S., Hinton, G., and Fleet, D. J. (2022) A generalist framework for panoptic segmentation of images and videos arXiv preprint arXiv:2210.06366

319. Hinton, G. E. (2022) The Forward-Forward Algorithm: Some Preliminary Investigations arXiv preprint arXiv:2212.13345

320. Liao, R., Kornblith, S., Ren, M., Fleet, D. J., and Hinton, G. (2021) Gaussian-Bernoulli RBMs Without Tears arXiv preprint arXiv:2210.10318

321. Culp, L., Sabour, S., and Hinton, G. E. (2021) Testing GLOM's ability to infer wholes from ambiguous parts arXiv preprint arXiv: 2211.16564

322. Sabour, S., Tagliasacchi, A., Yazdani, S., Hinton, G. E., Fleet, D. J. (2021) Unsupervised part representation by Flow Capsules arXiv preprint arXiv:2011.13920

323. Mller, R., Kornblith, S., Hinton, G. E. (2020) Subclass distillation arXiv preprint arXiv:2002.03936

324. Susskind, J., Anderson, A. and Hinton, G. E. (2010) The Toronto Face Database. Technical Report UTML TR 2010-001, University of Toronto.

325. Sminchisescu, C., Welling, M., and Hinton, G.E. (2003) A Mode-Hopping MCMC Sampler Technical Report CSRG-478, University of Toronto.

*   Hinton, G. E. (2000) Training Products of Experts by Minimizing Contrastive Divergence. Technical Report GCNU 2000-004, Gatsby Computational Neuroscience Unit, University College London.

*   Paccanaro, A and Hinton, G. E. (2000) Learning Distributed Representation of Concepts using Linear Relational Embedding. Technical Report GCNU 2000-002, Gatsby Computational Neuroscience Unit, University College London.

326. Hinton, G. E. and Revow, M. (1997) Using Mixtures of Factor Analyzers for Segmentation and Pose Estimation. (available at http://www.cs.toronto.edu/ hinton/papers.html)

327. Rasmussen, C. E., Neal, R. M., Hinton, G. E., van Camp, D, Revow, M.. Ghahramani, Z., Kustra, R, and Tibshirani, R. (1996) The DELVE Manual. Department of Computer Science, University of Toronto

*   Ghahramani, Z. and Hinton, G. E. (1996) Switching Mixtures of State space Models. Technical Report CRG-TR-96-3, University of Toronto.

328. Ghahramani, Z. and Hinton, G. E. (1996) Parameter Estimation for Linear Dynamical Systems. Technical Report CRG-TR-96-2, University of Toronto.

329. Ghahramani, Z. and Hinton, G. E. (1996) The EM algorithm for Mixtures of Factor Analyzers. Technical Report CRG-TR-96-1, University of Toronto.

*   Galland, C. C. and Hinton, G. E. (1990) Experiments on discovering higher-order features with mean field networks. Technical Report CRG-TR-90-3, Department of Computer Science, University of Toronto, Toronto, Canada.

*   Nowlan, S. J. and Hinton, G. E. (1989) Maximum Likelihood Decision-Directed Adaptive Equalization. Technical Report CRG-TR-89-8, Department of Computer Science, University of Toronto, Toronto, Canada.

*   Becker, S. and Hinton, G. E. (1990) Using Spatial Coherence as an Internal Teacher for a Neural Network. Technical Report CRG-TR-89-7, Department of Computer Science, University of Toronto, Toronto, Canada.

*   Galland, C. G. and Hinton, G. E. (1989) Deterministic Boltzmann Learning in Networks with Asymmetric Connectivity. Technical Report CRG-TR-89-6, Department of Computer Science, University of Toronto, Toronto, Canada.

*   Lang, K. and Hinton, G. E. (1988) A time-delay neural network architecture for speech recognition. Technical Report CMU-CS-88-152. Department of Computer Science, Carnegie-Mellon University, Pittsburgh PA.

*   Hinton, G. E. (1988) Representing part-whole hierarchies in connectionist networks. Technical Report CRG-TR-88-2, Department of Computer Science, University of Toronto, Toronto, Canada.

*   Waibel, A. Hanazawa, T. Hinton, G. Shikano, K. and Lang, K. (1987) Phoneme Recognition Using Time-Delay Neural Networks. Technical Report TR-1-0006. ATR Interpreting Telephony Research Laboratories, Japan.

*   Hinton, G. E. and Nowlan, S. J. (1986) How learning can guide evolution. Technical Report CMU-CS-86-128. Department of Computer Science, Carnegie-Mellon University, Pittsburgh PA.

330. Plaut, D., Nowlan, S. and Hinton, G. E. (1986) Experiments on learning by back-propagation. Technical Report CMU-CS-86-126. Department of Computer Science, Carnegie-Mellon University, Pittsburgh PA.

*   Hinton, G. E. (1984) Distributed Representations. Technical Report CMU-CS-84-157. Department of Computer Science, Carnegie-Mellon University, Pittsburgh PA.

331. Hinton, G. E., and Smolensky, P. (1984) Parallel computation and the mass-spring model of motor control. Technical Report. Center for Human Information Processing, University of California, San Diego.

*   Hinton, G. E., Sejnowski, T. J., and Ackley, D. H. (1984) Boltzmann Machines: Constraint satisfaction networks that learn. Technical Report CMU-CS-84-119, Carnegie-Mellon University.

332. Hammond, N., MacLean, A., Hinton, G., Long, J., Barnard, P., and Clark, I. (1983) Novice use of an interactive graph-plotting system. *Human factors report HF083* IBM (UK) Laboratories, Hursley Park.

333. Hinton, G. E. (1978) Relaxation and its role in vision. PhD Thesis, University of Edinburgh.

Working Papers

334. Dudek, G. and Hinton, G. E. (1993) Navigating without a map by directly transforming sensory inputs into location.

335. Hinton, G. E. (1982) Displays for network management. Applied Psychology Unit report for British Telecom.

336. Hinton, G. E. (1981) Some examples of novices problems with the CHART utility. MRC Applied Psychology Unit internal working paper.

337. Hinton, G. E. (1980) Larger receptive fields give more accurate representations. Program in Cognitive Science internal paper, University of California, San Diego.

338. Hinton, G. E. (1980) Self-tuning feature detectors. Program in Cognitive Science internal paper, University of California, San Diego.

339. Hinton, G. E. (1979) Are mental images like 2-D arrays? Program in Cognitive Science internal paper, University of California, San Diego.

Commentaries, Book Reviews and other Minor Publications

340. Hinton, G. E. (2011) Machine learning for neuroscience. *Neural Systems and Circuits*, **1**, Aug 2011.

341. Hinton, G. E. (2011) A better way to learn features: technical perspective. *Communications of the ACM*, **54**, No. 10, p 94.

342. Hinton, G. E. (2010) Deep Belief Networks In C. Sammut and G. Webb (eds.), *Encyclopedia of Machine Learning*, Springer.(3 pages) *An almost identical entry appears in Scholarpedia*

343. Hinton, G. E. (2010) Boltzmann Machines In C. Sammut and G. Webb (eds.), *Encyclopedia of Machine Learning*, Springer. (4 pages) *An almost identical entry appears in Scholarpedia*

344. Taylor, G.W., Hinton, G. E. and Roweis, S. (2008) Deep Generative Models for Modeling Animate Motion. Proc. 4th Int. Symp. Adaptive Motion of Animals and Machines.

345. Hinton, G. E. (2003) Neural Networks *Van Nostrand's Scientific Encyclopedia*

346. Hinton, G. E. (2000) Computation by Neural Networks. *Nature Neuroscience Supplement*, **3**, p1170

347. Hinton, G. E. (1999) Supervised Learning in Multilayer Neural Networks In The MIT Encyclopedia of the Cognitive Sciences Edited by Robert Wilson and Frank Keil The MIT Press, Cambridge, Mass.

348. R. Grzeszczuk, D. Terzopoulos, G. Hinton (1997) Learning fast neural network emulators for physics-based models (technical sketch) *Proc. ACM SIGGRAPH 97 Conference*, Los Angeles, CA, August, 1997, in *Computer Graphics Visual Proceedings, Annual Conference Series, 1997*, 167.

349. Hinton, G. E. (1995) Foreword to the book "Neural Networks for Pattern Recognition" by Chris Bishop. Oxford University Press, Oxford.

350. Hinton, G. E. and Nowlan, S. J. (1994) Preface to "Simplifying neural networks by soft weight-sharing". In D. H. Wolpert (Ed.) *The Mathematics of Generalization*. Santa Fe Institute Studies in the Sciences of Complexity.

351. Hinton, G. E. (1990) Review of Aleksander and Morton *Introduction to Neuro-Computing*, In *Nature*, **347**, 627-628.

352. Hinton, G. E., and LeCun, Y. (1988) Review of: R. K. Miller *Neural Networks: Implementing associative memory models in neurocomputers*, In Canadian Artificial Intelligence, **41**.

353. Hinton, G. E. (1987) Models of human inference. Invited commentary on a paper by D. McDermott. *Computational Intelligence*, **3**, 189-190.

354. Hinton, G. E. (1987) Boltzmann Machines. In S. Shapiro (Ed.) *The Encyclopedia of Artificial Intelligence* , New York: Wiley and Sons.

355. Hinton, G. E. (1982) Review of: S. E. Fahlman *NETL: A system for representing and using real-world knowledge.* In *A.I.S.B. Quarterly*, **42/43**.

356. Hinton, G. E. (1985) Three frames suffice. Invited commentary on a paper by J. Feldman. *The Behavioral and Brain Sciences*,

357. Hinton, G. E. (1980) Inferring the meaning of direct perception. Invited commentary on a paper by Ullman, S. *The Behavioral and Brain Sciences*, **3**, 387-388.

358. Hinton, G. E. (1979) Imagery without arrays. Invited commentary on a paper by S. M. Kosslyn, S. Pinker, G.E. Smith, and S. P. Shwartz. *The Behavioral and Brain Sciences*, **2**, 555-556

359. Hinton, G. E. (1979) Report on The La Jolla Conference on Cognitive Science, In *A.I.S.B. Quarterly*, **35**.

360. Hinton, G. E. (1979) Review of: D. C. Dennett *Brainstorms.* In *Contemporary Psychology*, **24**, 746-748.

361. Hinton, G. E. (1979) Review of: E. L. J. Leeuwenberg and H. F. J. M. Buffart (Eds.) *Formal theories of visual perception.* In *Journal of the Optical Society of America*, **69**, p.1492.

362. Hinton, G. E. (1978) Review of: J. Metzler (Ed.) *Systems Neuroscience.* In *Perception*, **7**, 364-365.

Graduate students

I have been the adviser for 22 completed MSc's and the following 37 completed PhD's:

Peter Brown (1987)
The Acoustic-Modeling Problem in Automatic Speech Recognition.

David Ackley (1987)
Stochastic Iterated Genetic Hillclimbing.

Mark Derthick (1988)
Mundane Reasoning by Parallel Constraint Satisfaction.

Richard Szeliski (1988)
Bayesian Modeling of Uncertainty in Low-Level Vision.

Kevin Lang (1989)
Phoneme Recognition Using Time-Delay Neural Nets.

Steven Nowlan (1991)
Soft Competitive Adaptation.

David Plaut (1991)
Connectionist Neuropsychology.

Conrad Galland (1991)
Learning in Deterministic Boltzmann Machine Networks.

S. Becker (1992)
An Information Theoretic Unsupervised Learning Algorithm for Neural Networks.

Richard Zemel (1994)
A Minimum Description Length Framework for Unsupervised Learning.

Tony Plate (1994)
Distributed Representations and Nested Compositional Structure.

Sidney Fels (1994)
Glove-TalkII: Mapping Hand Gestures to Speech Using Neural Networks.

Christopher Williams (1994)
Combining Deformable Models and Neural Networks for Handprinted Digit Recognition.

Radford Neal (1994)
Bayesian Learning in Neural Networks

Carl Rasmussen (1996)
Evaluation of Gaussian Processes and Other Methods for Non-linear Regression.

Brendan Frey (1997)
Bayesian Networks for Pattern Classification, Data Compression and Channel Coding

Evan Steeg (1997)
Automated Motif Discovery in Protein Structure Prediction.

Radek Grzeszczuk (1998) (co-advised by Demitri Terzopoulos)
NeuroAnimator: Fast neural network emulation and control of physics-based models.

Brian Sallans (2002)
Reinforcement Learning for Factored Markov Decision Processes.

Sageev Oore (2002)
Digital Marionette: Augmenting Kinematics with Physics for Multi-Track Desktop Performance Animation.

Andrew Brown (2002)
Product Models for Sequences.

Alberto Paccanaro (2002)
Learning Distributed Representations of Relational Data using Linear Relational Embedding.

Yee-Whye Teh (2003)
Bethe Free Energy and Contrastive Divergence Approximations for Undirected Graphical Models.

Simon Osindero (2004)
Contrastive Topographic Models: Energy-based density models applied to the understanding of sensory

coding and cortical topography.

Roland Memisevic (2007)
Non-linear Latent Factor Models for Revealing Structure in High-dimensional Data.

Ruslan Salakhutdinov (2009)
Learning deep generative models.

Graham Taylor (2009)
Composable, distributed-state models for high-dimensional time-series.

Andriy Mnih (2009)
Learning distributed representations for language modeling and collaborative filtering.

Vinod Nair (2010)
Visual object recognition using generative models of images.

Josh Susskind (2011)
Interpreting faces with neurally inspired generative models.

Ilya Sutskever (2012)
Training Recurrent Neural Networks.

Abdel-rahman Mohamed (2013)
Deep Neural Network Acoustic Models for ASR.

Vlad Mnih (2013)
Machine learning for aerial image labeling.

Navdeep Jaitly (2014)
Exploring Deep Learning Methods for Discovering Features in Speech Signals.

Tijmen Tieleman (2014)
Optimizing Neural Networks that Generate Images.

George Dahl (2015)
Deep Learning Approaches to Problems in Speech Recognition, Computational Chemistry and Natural Language Processing.

Charlie) Yichuan Tang (2015)
Learning Generative Models using Structured Latent Variables.

Nitish Srivastava (2016)
Deep Learning Models for Unsupervised and Transfer Learning.

Jimmy Lei Ba (2018)
Learning to Attend with Neural Networks.