Bibliography

[AA+15] Martín Abadi, Ashish Agarwal, et al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[ADL+06] Howard Y Ando, Luc Dehaspe, Walter Luyten, Elke Van Craenenbroeck, Henk Vandecasteele, and Luc Van Meervelt. Discovering h-bonding rules in crystals with inductive logic programming. *Molecular pharmaceutics*, 3(6):665–674, 2006.

[ADRDS+20] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, et al. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information Fusion*, 58:82–115, 2020.

[Aiz99] Igor N Aizenberg. Neural networks based on multi-valued and universal binary neurons: theory, application to image processing and recognition. In *International Conference on Computational Intelligence*, pages 306–316. Springer, 1999.

[ARD05] Eric E. Altendorf, Angelo C. Restificar, and Thomas G. Dietterich. Learning from sparse data by exploiting monotonicity constraints. In *Proceedings of the Twenty-First Conference on Uncertainty in Artificial Intelligence*, UAI'05, page 18–26, Arlington, Virginia, USA, 2005. AUAI Press.

[BDBC+10] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine learning*, 79(1):151–175, 2010.

[BDR+20] Yoshua Bengio, Tristan Deleu, Nasim Rahaman, Nan Rosemary Ke, Sebastien Lachapelle, Olexa Bilaniuk, Anirudh Goyal, and Christopher Pal. A meta-transfer objective for learning to disentangle causal mechanisms. In *International Conference on Learning Representations*, 2020.

[BGB+17] Tarek R Besold, Artur d'Avila Garcez, Sebastian Bader, Howard Bowman, Pedro Domingos, Pascal Hitzler, Kai-Uwe Kühnberger, Luis C Lamb, Daniel Lowd, Priscila Machado Vieira Lima, et al. Neural-symbolic learning and reasoning: A survey and interpretation. *arXiv preprint arXiv:1711.03902*, 2017.

[BGC17] Yoshua Bengio, Ian Goodfellow, and Aaron Courville. *Deep learning*, volume 1. MIT press Cambridge, MA, USA, 2017.

[BGKP21] Julius Berner, Philipp Grohs, Gitta Kutyniok, and Philipp Petersen. The modern mathematics of deep learning. *arXiv preprint arXiv:2105.04026*, 2021.

[BGLA21] Filippo Maria Bianchi, Daniele Grattarola, Lorenzo Livi, and Cesare Alippi. Graph neural networks with convolutional arma filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1, 2021.

[BHB+18] Peter W Battaglia, Jessica B Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, et al. Relational inductive biases, deep learning, and graph networks. *arXiv preprint arXiv:1806.01261*, 2018.

[BKBR21] Navneet Bung, Sowmya Ramaswamy Krishnan, Gopalakrishnan Bulusu, and Arijit Roy. De novo design of new chemical entities for sars-cov-2 using artificial intelligence. *Future Medicinal Chemistry*, 13(6):575–585, 2021.

[BLPL07] Yoshua Bengio, Pascal Lamblin, Dan Popovici, and Hugo Larochelle. Greedy layer-wise training of deep networks. In *Advances in neural information processing systems*, pages 153–160, 2007.

[Blu92] Avrim Blum. Learning boolean functions in an infinite attribute space. *Machine Learning*, 9(4):373–386, 1992.

[BMO+21] Jannis Born, Matteo Manica, Ali Oskooei, Joris Cadow, Greta Markert, and María Rodríguez Martínez. Paccmannrl: De novo generation of hit-like anticancer molecules from transcriptomic data via reinforcement learning. *Iscience*, 24(4):102269, 2021.

[BMR+20] Tom B Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sas-

try, Amanda Askell, et al. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*, 2020.

[BPZ97] Igor I Baskin, Vladimir A Palyulin, and Nikolai S Zefirov. A neural device for searching direct correlations between structures and properties of chemical compounds. *Journal of chemical information and computer sciences*, 37(4):715–721, 1997.

[BVV+16] Samuel R. Bowman, L. Vilnis, Oriol Vinyals, Andrew M. Dai, R. Józefowicz, and S. Bengio. Generating sentences from a continuous space. In *CoNLL*, 2016.

[BW91] Wray L. Buntine and A. Weigend. Bayesian back-propagation. *Complex Syst.*, 5, 1991.

[C+15] François Chollet et al. Keras. https://keras.io, 2015.

[CD20] Andrew Cropper and Sebastijan Dumančić. Inductive logic programming at 30: a new introduction. *arXiv preprint arXiv:2008.07912*, 2020.

[CDEM22] Andrew Cropper, Sebastijan Dumančić, Richard Evans, and Stephen H. Muggleton. Inductive logic programming at 30. *Machine Learning*, 111(1):147–172, Jan 2022.

[CDM20] Andrew Cropper, Sebastijan Dumančić, and Stephen H. Muggleton. Turning 30: New ideas in inductive logic programming. In Christian Bessiere, editor, *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, pages 4833–4839, 2020.

[CHBP02] Anthony Coates, Yanmin Hu, Richard Bax, and Clive Page. The future challenges facing the development of new antimicrobial drugs. *Nature reviews Drug discovery*, 1(11):895–910, 2002.

[CL14] Chin-Liang Chang and Richard Char-Tung Lee. *Symbolic logic and mechanical theorem proving*. Academic press, 2014.

[CLZ+19] Qibin Chen, Junyang Lin, Yichang Zhang, Ming Ding, Yukuo Cen, Hongxia Yang, and Jie Tang. Towards knowledge-based recommender dialog system. *arXiv preprint arXiv:1908.05391*, 2019.

[CM21] Andrew Cropper and Rolf Morel. Learning programs by learning from failures. *Machine Learning*, 110(4):801–856, 2021.

[CS82] Brian Cohen and Claude Sammut. Object recognition and concept learning with confucius. *Pattern Recognition*, 15(4):309–316, 1982.

[CVJ+18] Cătălina Cangea, Petar Veličković, Nikola Jovanović, Thomas Kipf, and Pietro Liò. Towards sparse hierarchical graph classifiers. *ArXiv*, abs/1811.01287, 2018.

[CWPZ18] Peng Cui, Xiao Wang, Jian Pei, and Wenwu Zhu. A survey on network embedding. *IEEE Transactions on Knowledge and Data Engineering*, 31(5):833–852, 2018.

[CYK+18] Daniel Cer, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, et al. Universal sentence encoder for english. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 169–174, 2018.

[CYM20] William Cohen, Fan Yang, and Kathryn Rivard Mazaitis. Tensorlog: A probabilistic database implemented using deep-learning infrastructure. *Journal of Artificial Intelligence Research*, 67:285–325, 2020.

[DCLT19] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *NAACL-HLT (1)*, pages 4171–4186, 2019.

[DDS+09] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[Dec86] Rina Dechter. Learning while searching in constraint-satisfaction-problems. In Tom Kehler, editor, *Proceedings of the 5th National Conference on Artificial Intelligence. Philadelphia, PA, USA, August 11-15, 1986. Volume 1: Science*, pages 178–185. Morgan Kaufmann, 1986.

[dGL20] Artur d'Avila Garcez and Luis C. Lamb. Neurosymbolic ai: The 3rd wave, 2020.

[DGS17] Michelangelo Diligenti, Marco Gori, and Claudio Sacca. Semantic-based regularization for learning and inference. *Artificial Intelligence*, 244:143–165, 2017.

[DQW+21] Tianjian Dong, Qi Qi, Jingyu Wang, Alex X Liu, Haifeng Sun, Zirui Zhuang, and Jianxin Liao. Generative adversarial network-based transfer reinforcement learning for routing with prior knowledge. *IEEE Transactions on Network and Service Management*, 2021.

[DRG17] Michelangelo Diligenti, Soumali Roychowdhury, and Marco Gori. Integrating prior knowledge into deep learning. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 920–923. IEEE, 2017.

[DRKT07] Luc De Raedt, Angelika Kimmig, and Hannu Toivonen. Problog: A probabilistic prolog and its application in link discovery. In *IJCAI*, volume 7, pages 2462–2467. Hyderabad, 2007.

[DRR16] T. Demeester, Tim Rocktäschel, and S. Riedel. Lifted rule injection for relation embeddings. *ArXiv*, abs/1606.08359, 2016.

[DS13] Brian W Dymock and Cheng Shang See. Inhibitors of jak2 and jak3: an update on the patent literature 2010–2012. *Expert opinion on therapeutic patents*, 23(4):449–501, 2013.

[DSW+20] Chunning Du, Haifeng Sun, Jingyu Wang, Qi Qi, and Jianxin Liao. Adversarial and domain-aware bert for cross-domain sentiment analysis. In *Proceedings of the 58th annual meeting of the Association for Computational Linguistics*, pages 4019–4028, 2020.

[DYCFY14] Brian W Dymock, Eugene Guorong Yang, Yuyi Chu-Farseeva, and Lianbin Yao. Selective jak inhibitors. *Future medicinal chemistry*, 6(12):1439–1471, 2014.

[EG18] Richard Evans and Edward Grefenstette. Learning explanatory rules from noisy data. *J. Artif. Intell. Res.*, 61:1–64, 2018.

[EMM+18] Kevin Ellis, Lucas Morales, Mathias Sabl Meyer, Armando Solar-Lezama, and Joshua B Tenenbaum. Dreamcoder: Bootstrapping domain-specific languages for neurally-guided bayesian program learning. In *Proceedings of the 2nd Workshop on Neural Abstract Machines and Program Induction*, 2018.

[ES09] Peter Ertl and Ansgar Schuffenhauer. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *Journal of cheminformatics*, 1(1):1–11, 2009.

[EWN+20] Kevin Ellis, Catherine Wong, Maxwell Nye, Mathias Sable-Meyer, Luc Cary, Lucas Morales, Luke Hewitt, Armando Solar-Lezama, and Joshua B Tenenbaum. Dreamcoder: Growing generalizable, interpretable knowledge with wake-sleep bayesian program learning. *arXiv preprint arXiv:2006.08381*, 2020.

[FBDC+19] M. Fischer, Mislav Balunovic, Dana Drachsler-Cohen, Timon Gehr, Ce Zhang, and Martin T. Vechev. Dl2: Training and querying neural networks with logic. In *ICML*, 2019.

[FCDRDG14] Paolo Frasconi, Fabrizio Costa, Luc De Raedt, and Kurt De Grave. kLog: A language for logical and relational learning with kernels. *Artificial Intelligence*, 217:117–143, 2014.

[FGU+21] Hossein Rajaby Faghihi, Quan Guo, Andrzej Uszok, Aliakbar Nafar, Elaheh Raisi, and Parisa Kordjamshidi. Domiknows: A library for integration of symbolic domain knowledge in deep learning. *arXiv preprint arXiv:2108.12370*, 2021.

[FL19] Matthias Fey and Jan E. Lenssen. Fast graph representation learning with PyTorch Geometric. In *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019.

[FO93] Justin Fletcher and Zoran Obradovic. Combining prior symbolic knowledge and constructive neural network learning. *Connection Science*, 5(3-4):365–375, 1993.

[Fog08] Agner Fog. Sampling methods for wallenius' and fisher's noncentral hypergeometric distributions. *Communications in Statistics—Simulation and Computation®*, 37(2):241–257, 2008.

[FSK12] Tanveer A Faruquie, Ashwin Srinivasan, and Ross D King. Topic models with relational features for drug design. In *International conference on inductive logic programming*, pages 45–57. Springer, 2012.

[Fu93] L. M. Fu. Knowledge-based connectionism for revising domain theories. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(1):173–182, 1993.

[Fu95] Li Min Fu. Introduction to knowledge-based neural networks. *Knowledge-Based Systems*, 1995.

[FZG14] Manoel VM França, Gerson Zaverucha, and Artur S d'Avila Garcez. Fast relational learning using bottom clause propositionalization with artificial neural networks. *Machine learning*, 94(1):81–104, 2014.

[FZG15] Manoel Vitor Macedo França, Gerson Zaverucha, and ASD Garcez. Neural relational learning through semi-propositionalization of bottom clauses. In *AAAI Spring Symposium Series*, 2015.

[GBG12] Artur S d'Avila Garcez, Krysia B Broda, and Dov M Gabbay. *Neural-symbolic learning systems: foundations and applications*. Springer Science & Business Media, 2012.

[GC10] Kurt De Grave and Fabrizio Costa. Molecular graph augmentation with rings and functional groups. *Journal of chemical information and modeling*, 50(9):1660–1668, 2010.

[GC21] Victor Guimarães and Vítor Santos Costa. Neurallog: a neural logic language. *arXiv preprint arXiv:2105.01442*, 2021.

[GFS21] Manas Gaur, Keyur Faldu, and Amit Sheth. Semantics of the black-box: Can knowledge graphs help make deep learning systems more interpretable and explainable? *IEEE Internet Computing*, 25(1):51–59, 2021.

[GHN+17] Anna Gaulton, Anne Hersey, Michał Nowotka, A Patricia Bento, Jon Chambers, David Mendez, Prudence Mutowo, Francis Atkinson, Louisa J Bellis, Elena Cibrián-Uhalte, et al. The chembl database in 2017. *Nucleic acids research*, 45(D1):D945–D954, 2017.

[GMLS20] Francesca Grisoni, Michael Moret, Robin Lingwood, and Gisbert Schneider. Bidirectional molecule generation with recurrent neural networks. *Journal of chemical information and modeling*, 60(3):1175–1183, 2020.

[GMS05] Marco Gori, Gabriele Monfardini, and Franco Scarselli. A new model for learning in graph domains. In *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.*, volume 2, pages 729–734. IEEE, 2005.

[Gre21] Daria Grechishnikova. Transformer neural network for protein-specific de novo drug generation as a machine translation problem. *Scientific reports*, 11(1):1–13, 2021.

[GRS+21] Manas Gaur, Kaushik Roy, Aditya Sharma, Biplav Srivastava, and Amit Sheth. " who can help me?": Knowledge infused matching of support seekers and support providers during covid-19 on reddit. *arXiv preprint arXiv:2105.06398*, 2021.

[GSR+17] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *International conference on machine learning*, pages 1263–1272. PMLR, 2017.

[GWW+16] Shu Guo, Quan Wang, Lihong Wang, Bin Wang, and Li Guo. Jointly embedding knowledge graphs and logical rules. In *Proceedings of the 2016*

conference on empirical methods in natural language processing, pages 192–202, 2016.

[GZ99] Artur S Avila Garcez and Gerson Zaverucha. The connectionist inductive learning and logic programming system. *Applied Intelligence*, 11(1):59–77, 1999.

[Hai10] William N. Hait. Anticancer drug development: the grand challenges. *Nature Reviews Drug Discovery*, 9(4):253–254, Apr 2010.

[Háj13] Petr Hájek. *Metamathematics of fuzzy logic*, volume 4. Springer Science & Business Media, 2013.

[Ham20] William L Hamilton. Graph representation learning. *Synthesis Lectures on Artifical Intelligence and Machine Learning*, 14(3):1–159, 2020.

[HBC+20] Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard de Melo, Claudio Gutierrez, José Emilio Labra Gayo, Sabrina Kirrane, Sebastian Neumaier, Axel Polleres, et al. Knowledge graphs. *arXiv preprint arXiv:2003.02320*, 2020.

[HBZ+18] William L. Hamilton, P. Bajaj, M. Zitnik, Dan Jurafsky, and J. Leskovec. Embedding logical queries on knowledge graphs. In *NeurIPS*, 2018.

[HDFN95] Geoffrey E Hinton, Peter Dayan, Brendan J Frey, and Radford M Neal. The" wake-sleep" algorithm for unsupervised neural networks. *Science*, 268(5214):1158–1161, 1995.

[Heb49] Donald Olding Hebb. *The organisation of behaviour: a neuropsychological theory*. Science Editions New York, 1949.

[HH21] Hiroshi Honda and Masafumi Hagiwara. Analogical reasoning with deep learning-based symbolic processing. *IEEE Access*, 9:121859–121870, 2021.

[HKBG21] Nicholas Hoernle, Rafael Michael Karampatsis, Vaishak Belle, and Kobi Gal. Multiplexnet: Towards fully satisfied logical constraints in neural networks. *arXiv preprint arXiv:2111.01564*, 2021.

[HMK07] David Heckerman, Chris Meek, and Daphne Koller. Probabilistic entity-relationship models, prms, and plate models. *Introduction to statistical relational learning*, pages 201–238, 2007.

[HML+16] Zhiting Hu, Xuezhe Ma, Zhengzhong Liu, E. Hovy, and E. Xing. Harnessing deep neural networks with logic rules. *ArXiv*, abs/1603.06318, 2016.

[HOT06] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554, 2006.

[HS97] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[HVD15] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.

[HYL17] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Advances in neural information processing systems*, pages 1024–1034, 2017.

[HZJ07] Yu-Chi Ho, Qian-Chuan Zhao, and Qing-Shan Jia. *Ordinal Optimization: Soft Optimization for Hard Problems*. Springer, 2007.

[Jan20] Vince Jankovics. vakker/cilp. https://github.com/vakker/CILP, 2020.

[JRS08] Sachindra Joshi, Ganesh Ramakrishnan, and Ashwin Srinivasan. Feature construction using theory-guided sampling and randomised search. In *International Conference on Inductive Logic Programming*, pages 140–157. Springer, 2008.

[KB15] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR (Poster)*, 2015.

[KBBR21] Sowmya Ramaswamy Krishnan, Navneet Bung, Gopalakrishnan Bulusu, and Arijit Roy. Accelerating de novo drug design against novel proteins using deep learning. *Journal of Chemical Information and Modeling*, 61(2):621–630, 2021.

[KGC17] Jan Kukačka, Vladimir Golkov, and Daniel Cremers. Regularization for deep learning: A taxonomy. *arXiv preprint arXiv:1710.10686*, 2017.

[KGS19] Ugur Kursuncu, Manas Gaur, and Amit Sheth. Knowledge infused learning (k-il): Towards deep incorporation of knowledge in deep learning. *arXiv preprint arXiv:1912.00512*, 2019.

[Kit16] Hiroaki Kitano. Artificial intelligence to win the nobel prize and beyond: Creating the engine for scientific discovery. *AI magazine*, 37(1):39–49, 2016.

[KKM+16] Kristian Kersting, Nils M. Kriege, Christopher Morris, Petra Mutzel, and Marion Neumann. Benchmark data sets for graph kernels, 2016. http://graphkernels.cs.tu-dortmund.de.

[KLF01] Stefan Kramer, Nada Lavrač, and Peter Flach. *Propositionalization Approaches to Relational Data Mining*, pages 262–291. Springer Berlin Heidelberg, Berlin, Heidelberg, 2001.

[KMSS96] Ross D King, Stephen H Muggleton, Ashwin Srinivasan, and MJ Sternberg. Structure-activity relationships derived by machine learning: The use of atoms and their bond connectivities to predict mutagenicity by inductive logic programming. *Proceedings of the National Academy of Sciences*, 93(1):438–442, 1996.

[KSH12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.

[KT07] Eyal Krupka and Naftali Tishby. Incorporating prior knowledge on features into learning. In *AISTATS*, 2007.

[KW14] Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In *ICLR*, 2014.

[KW17] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*, 2017.

[KWJ+04] Ross D King, Kenneth E Whelan, Ffion M Jones, Philip GK Reiser, Christopher H Bryant, Stephen H Muggleton, Douglas B Kell, and Stephen G Oliver. Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature*, 427(6971):247–252, 2004.

[L+06] Greg Landrum et al. Rdkit: Open-source cheminformatics. https://www.rdkit.org/docs/index.html, 2006.

[Lav90] Nada Lavrac. *Principles of knowledge acquisition in expert systems*. PhD thesis, Ph. D. thesis, Faculty of Technical Sciences, University of Maribor, 1990.

[LBD+89] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989.

[LDG91] Nada Lavrač, Sašo Džeroski, and Marko Grobelnik. Learning nonrecursive definitions of relations with linus. In *European Working Session on Learning*, pages 265–281. Springer, 1991.

[LFJ+18] Lei Li, Min Feng, Lianwen Jin, Shenjin Chen, Lihong Ma, and Jiakai Gao. Domain knowledge embedding regularization neural networks for workload prediction and analysis in cloud computing. *J. Inf. Technol. Res.*, 11(4):137–154, October 2018.

[LIJ+15] Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick Van Kleef, Sören Auer, et al. Dbpedia–a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic web*, 6(2):167–195, 2015.

[Lip16] Zachary C. Lipton. The mythos of model interpretability. *arXiv preprint arXiv:1606.03490*, 2016.

[LLK19] Junhyun Lee, Inyeop Lee, and Jaewoo Kang. Self-attention graph pooling. In *International Conference on Machine Learning*, pages 3734–3743, 2019.

[Llo12] John W Lloyd. *Foundations of logic programming*. Springer Science & Business Media, 2012.

[Lod13] Huma Lodhi. Deep relational machines. In *International Conference on Neural Information Processing*, pages 212–219. Springer, 2013.

[LS20] Tao Li and Vivek Srikumar. Augmenting neural networks with first-order logic. In *ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, 2020.

[LSR20] Nada Lavrac, Blaz Skrlj, and Marko Robnik-Sikonja. Propositionalization and embeddings: two sides of the same coin. *Mach. Learn.*, 109(7):1465–1507, 2020.

[LWM18] Xuan Liu, Xiaoguang Wang, and Stan Matwin. Improving the interpretability of deep neural networks with knowledge distillation. *arXiv preprint arXiv:1812.10924*, 2018.

[LZZ21] Xing Luo, Dongxiao Zhang, and Xu Zhu. Deep learning based forecasting of photovoltaic power generation by incorporating domain knowledge. *Energy*, 225:120240, 2021.

[Mar18] Gary Marcus. Deep Learning: A Critical Appraisal. *arXiv*, jan 2018.

[Mar20] Gary Marcus. The next decade in ai: four steps towards robust artificial intelligence. *arXiv preprint arXiv:2002.06177*, 2020.

[MB88] Stephen Muggleton and Wray Buntine. Machine invention of first-order predicates by inverting resolution. In *Machine Learning Proceedings 1988*, pages 339–352. Elsevier, 1988.

[MCP+21] Kenneth Marino, Xinlei Chen, Devi Parikh, Abhinav Gupta, and Marcus Rohrbach. Krisp: Integrating implicit and symbolic knowledge for open-domain knowledge-based vqa. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14111–14121, 2021.

[Md94] Stephen Muggleton and Luc de Raedt. Inductive logic programming: Theory and methods. *The Journal of Logic Programming*, 19-20:629–679, 1994. Special Issue: Ten Years of Logic Programming.

[MDK+18] Robin Manhaeve, Sebastijan Dumancic, Angelika Kimmig, Thomas Demeester, and Luc De Raedt. Deepproblog: Neural probabilistic logic programming. *Advances in Neural Information Processing Systems*, 31:3749–3759, 2018.

[MDRP+12] Stephen Muggleton, Luc De Raedt, David Poole, Ivan Bratko, Peter Flach, Katsumi Inoue, and Ashwin Srinivasan. Ilp turns 20. *Machine learning*, 86(1):3–23, 2012.

[Mic73] Ryszard S. Michalski. Discovering classification rules using variable-valued logic system VL1. In Nils J. Nilsson, editor, *Proceedings of the 3rd International Joint Conference on Artificial Intelligence. Standford, CA, USA, August 20-23, 1973*, pages 162–172. William Kaufmann, 1973.

[Mic80] Ryszard S Michalski. Pattern recognition as rule-guided inductive inference. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-2(4):349–361, 1980.

[MIM+19] Nikhil Muralidhar, Mohammad Raihanul Islam, Manish Marwah, Anuj Karpatne, and Naren Ramakrishnan. Incorporating Prior Domain Knowledge into Deep Neural Networks. In *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018*, 2019.

[MMBC21] Omar Mahmood, Elman Mansimov, Richard Bonneau, and Kyunghyun Cho. Masked graph modeling for molecule generation. *Nature communications*, 12(1):1–12, 2021.

[MMPS94] Donald Michie, Stephen Muggleton, David Page, and Ashwin Srinivasan. To the international computing community: A new east-west challenge. *Distributed email document available from https://www.doc.ic.ac.uk/~shm/Papers/ml-chall.pdf*, 1994.

[MOHU03] Kenneth A Marx, Philip O'Neil, Patrick Hoffman, and ML Ujwal. Data mining the nci cancer cell line compound gi50 values: identifying quinone subtypes effective against melanoma and leukemia cell classes. *Journal of chemical information and computer sciences*, 43(5):1652–1667, 2003.

[MP43] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.

[MR19] Kit-Kay Mak and Pichika. Mallikarjuna Rao. Artificial intelligence in drug development: present status and future prospects. *Drug Discovery Today*, 24(3):773–780, 2019.

[MRF+19] Christopher Morris, Martin Ritzert, Matthias Fey, William L Hamilton, Jan Eric Lenssen, Gaurav Rattan, and Martin Grohe. Weisfeiler and leman go neural: Higher-order graph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 4602–4609, 2019.

[MS98] Eric McCreath and Arun Sharma. LIME: A System for Learning Relations. In *International Conference on Algorithmic Learning Theory*, pages 336–374. Springer, 1998.

[Mug87] Stephen Muggleton. Duce, an oracle-based approach to constructive induction. In *IJCAI*, pages 287–292. Citeseer, 1987.

[Mug91] Stephen Muggleton. Inductive logic programming. *New generation computing*, 8(4):295–318, 1991.

[Mug95] Stephen Muggleton. Inverse entailment and progol. *New generation computing*, 13(3-4):245–286, 1995.

[Mug96] Stephen Muggleton. Learning from positive data. In *International conference on inductive logic programming*, pages 358–376. Springer, 1996.

[MW+97] Alan D McNaught, Andrew Wilkinson, et al. *Compendium of chemical terminology*, volume 1669. Blackwell Science Oxford, 1997.

[MZB+21] Fanhe Ma, Faen Zhang, Shenglan Ben, Shuxin Qin, Pengcheng Zhou, Changsheng Zhou, and Fengyi Xu. Monotonic neural network: combining

deep learning with domain knowledge for chiller plants energy optimization. *arXiv preprint arXiv:2106.06143*, 2021.

[Nea95] Radford M. Neal. *Bayesian Learning for Neural Networks*. PhD thesis, University of Toronto, CAN, 1995. AAINN02676.

[Nil91] Nils J Nilsson. Logic and artificial intelligence. *Artificial intelligence*, 47(1-3):31–56, 1991.

[OOD+21] Ivan Olier, Oghenejokpeme I Orhobor, Tirtharaj Dash, Andy M Davis, Larisa N Soldatova, Joaquin Vanschoren, and Ross D King. Transformational machine learning: Learning how to learn from many related scientific problems. *Proceedings of the National Academy of Sciences*, 118(49), 2021.

[PGM+19] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, pages 8024–8035, 2019.

[PIT18] Mariya Popova, Olexandr Isayev, and Alexander Tropsha. Deep reinforcement learning for de novo drug design. *Science Advances*, 4(7):eaap7885, 2018.

[PKD+19] Namyong Park, Andrey Kan, Xin Luna Dong, Tong Zhao, and Christos Faloutsos. *Estimating Node Importance in Knowledge Graphs Using Graph Neural Networks*, page 596–606. Association for Computing Machinery, New York, NY, USA, 2019.

[Plo70] Gordon D Plotkin. A note on inductive generalization. *Machine intelligence*, 5(1):153–163, 1970.

[Plo72] Gordon Plotkin. Automatic methods of inductive inference. *PhD Thesis, The University of Edinburgh*, 1972.

[Pre98] Lutz Prechelt. Early stopping-but when? In *Neural Networks: Tricks of the trade*, pages 55–69. Springer, 1998.

[PSS20] Hemant Purohit, Valerie L Shalin, and Amit P Sheth. Knowledge graphs to empower humanity-inspired ai systems. *IEEE Internet Computing*, 24(4):48–54, 2020.

[PW80] Fernando CN Pereira and David HD Warren. Definite clause grammars for language analysis—a survey of the formalism and a comparison with augmented transition networks. *Artificial intelligence*, 13(3):231–278, 1980.

[Qui90] J. Ross Quinlan. Learning logical definitions from relations. *Machine learning*, 5(3):239–266, 1990.

[Rae10] Luc De Raedt. *Inductive Logic Programming*, pages 529–537. Springer US, Boston, MA, 2010.

[RBSR14] Tim Rocktäschel, Matko Bosnjak, Sameer Singh, and Sebastian Riedel. Low-dimensional embeddings of logic. In *Proceedings of the ACL 2014 Workshop on Semantic Parsing*, pages 45–49, Baltimore, MD, June 2014. Association for Computational Linguistics.

[RDMM20] Luc de Raedt, Sebastijan Dumančić, Robin Manhaeve, and Giuseppe Marra. From statistical relational to neuro-symbolic artificial intelligence. In Christian Bessiere, editor, *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, pages 4943–4950. International Joint Conferences on Artificial Intelligence Organization, 7 2020. Survey track.

[RGL+20] Ryan Riegel, Alexander G. Gray, Francois P. S. Luus, Naweed Khan, Ndivhuwo Makondo, Ismail Yunus Akhalwaya, Haifeng Qian, Ronald Fagin, Francisco Barahona, Udit Sharma, Shajith Ikbal, Hima Karanam, Sumit Neelam, Ankita Likhyani, and Santosh K. Srivastava. Logical neural networks. *CoRR*, abs/2006.13155, 2020.

[RHW86] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.

[RJBS07] Ganesh Ramakrishnan, Sachindra Joshi, Sreeram Balakrishnan, and Ashwin Srinivasan. Using ilp to construct features for information extraction from semi-structured text. In *International Conference on Inductive Logic Programming*, pages 211–224. Springer, 2007.

[Rob97] Sam Roberts. An introduction to progol. *Department of Computer Science, University of York*, 244, 1997.

[Ros57] Frank Rosenblatt. *The perceptron, a perceiving and recognizing automaton Project Para*. Cornell Aeronautical Laboratory, 1957.

[RP20] Alan Ramponi and Barbara Plank. Neural unsupervised domain adaptation in nlp—a survey. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 6838–6855, 2020.

[RS22] Davor Runje and Sharath M Shankaranarayana. Constrained monotonic neural networks. *arXiv preprint arXiv:2205.11775*, 2022.

[RSR15] Tim Rocktäschel, Sameer Singh, and Sebastian Riedel. Injecting logical background knowledge into embeddings for relation extraction. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1119–1129, Denver, Colorado, May–June 2015. Association for Computational Linguistics.

[RSSB05] Liva Ralaivola, Sanjay J Swamidass, Hiroto Saigo, and Pierre Baldi. Graph kernels for chemical informatics. *Neural networks*, 18(8):1093–1110, 2005.

[Ruc91] William H. Ruckle. A discrete search game. In *Theory and Decision Library*, pages 29–43. Springer Netherlands, 1991.

[RVBW06] Francesca Rossi, Peter Van Beek, and Toby Walsh. *Handbook of constraint programming*. Elsevier, 2006.

[RWC+19] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.

[SAZ+18] Gustav Sourek, Vojtech Aschenbrenner, Filip Zelezny, Steven Schockaert, and Ondrej Kuzelka. Lifted relational neural networks: Efficient learning of latent relational structures. *Journal of Artificial Intelligence Research*, 62:69–100, 2018.

[SB86] Claude Sammut and Ranan B Banerji. Learning concepts by asking questions. *Machine learning: An artificial intelligence approach*, 2:167–192, 1986.

[Sch15] Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural networks*, 61:85–117, 2015.

[SdCRG21] Prithviraj Sen, Breno W. S. R. de Carvalho, Ryan Riegel, and Alexander G. Gray. Neuro-symbolic inductive logic programming with logical neural networks. *CoRR*, abs/2112.03324, 2021.

[SFK+19] Niclas Stahl, Goran Falkman, Alexander Karlsson, Gunnar Mathiason, and Jonas Bostrom. Deep reinforcement learning for multiparameter optimization in de novo drug design. *Journal of Chemical Information and Modeling*, 59(7):621–630, 2019.

[SG16] Luciano Serafini and Artur d'Avila Garcez. Logic tensor networks: Deep learning and logical reasoning from data and knowledge. *arXiv preprint arXiv:1606.04422*, 2016.

[SGKW19] A. Sheth, M. Gaur, U. Kursuncu, and R. Wickramarachchi. Shades of knowledge-infused learning for enhancing deep learning. *IEEE Internet Computing*, 23(6):54–63, 2019.

[SGS15] Rupesh Kumar Srivastava, Klaus Greff, and Jürgen Schmidhuber. Highway networks. *arXiv preprint arXiv:1505.00387*, 2015.

[SGT+08] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1):61–80, 2008.

[SHK+14] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

[SK99] Ashwin Srinivasan and Ross D King. Feature construction with inductive logic programming: A study of quantitative predictions of biological activity aided by structural attributes. *Data Mining and Knowledge Discovery*, 3(1):37–57, 1999.

[SKB03] Ashwin Srinivasan, Ross D King, and Michael E Bain. An empirical study of the use of relevance information in inductive logic programming. *Journal of Machine Learning Research*, 4(Jul):369–383, 2003.

[SKB+18] Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In *European semantic web conference*, pages 593–607. Springer, 2018.

[SKTW17] Marwin H.S. Segler, Thierry Kogej, Christian Tyrchan, and Mark P. Waller. Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS Central Science*, 4(1):120–131, 2017.

[SLM20] Mattia Silvestri, Michele Lombardi, and Michela Milano. Injecting domain knowledge in neural networks: a controlled experiment on a constrained problem. *arXiv preprint arXiv:2002.10742*, 2020.

[Šou20] Gustav Šourek. *Deep Learning with Relational Logic Representations*. PhD thesis, Czech Technical University in Prague, 2020.

[SPG19] Amit Sheth, Swati Padhee, and Amelie Gyrard. Knowledge graphs and knowledge networks: the story in brief. *IEEE Internet Computing*, 23(4):67–75, 2019.

[SR11] Ashwin Srinivasan and Ganesh Ramakrishnan. Parameter screening and optimisation for ilp using designed experiments. *Journal of Machine Learning Research*, 12(2), 2011.

[Sri99a] Ashwin Srinivasan. A study of two probabilistic methods for searching large spaces with ilp, 1999.

[Sri99b] Ashwin Srinivasan. A study of two sampling methods for analyzing large datasets with ilp. *Data Mining and Knowledge Discovery*, 3(1):95–123, Mar 1999.

[Sri01] Ashwin Srinivasan. The Aleph Manual. https://www.cs.ox.ac.uk/activities/programinduction/Aleph/aleph.html, 2001.

[SS97] Alessandro Sperduti and Antonina Starita. Supervised neural networks for the classification of structures. *IEEE Transactions on Neural Networks*, 8(3):714–735, 1997.

[SSJ+09] Lucia Specia, Ashwin Srinivasan, Sachindra Joshi, Ganesh Ramakrishnan, and Maria das Graças Volpe Nunes. An investigation into feature construction to assist word sense disambiguation. *Machine Learning*, 76(1):109–136, 2009.

[SSR12] Amrita Saha, Ashwin Srinivasan, and Ganesh Ramakrishnan. What kinds of relational features are useful for statistical learning? In *International Conference on Inductive Logic Programming*, pages 209–224. Springer, 2012.

[SSRN06] Lucia Specia, Ashwin Srinivasan, Ganesh Ramakrishnan, and Maria das Graças Volpe Nunes. Word sense disambiguation using inductive logic programming. In *International Conference on Inductive Logic Programming*, pages 409–423. Springer, 2006.

[Stå21] Niclas Ståhl. *Integrating domain knowledge into deep learning: Increasing model performance through human expertise*. PhD thesis, Högskolan i Skövde, 2021.

[STE13] Christian Szegedy, Alexander Toshev, and Dumitru Erhan. Deep neural networks for object detection. In C. J. C. Burges, L. Bottou, M. Welling,

Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013.

[STN+20] Rick Stevens, Valerie Taylor, Jeff Nichols, Arthur Barney Maccabe, Katherine Yelick, and David Brown. Ai for science. Technical report, Argonne National Lab.(ANL), Argonne, IL (United States), 2020.

[Sto76] Lawrence D Stone. *Theory of optimal search*, volume 118. Elsevier, 1976.

[SWP+20] Petra Schneider, W Patrick Walters, Alleyn T Plowright, Norman Sieroka, Jennifer Listgarten, Robert A Goodnow, Jasmin Fisher, Johanna M Jansen, José S Duca, Thomas S Rush, et al. Rethinking drug design in the artificial intelligence era. *Nature Reviews Drug Discovery*, 19(5):353–364, 2020.

[SYS+20] Jonathan M Stokes, Kevin Yang, Kyle Swanson, Wengong Jin, Andres Cubillos-Ruiz, Nina M Donghia, Craig R MacNair, Shawn French, Lindsey A Carfrae, Zohar Bloom-Ackermann, et al. A deep learning approach to antibiotic discovery. *Cell*, 180(4):688–702, 2020.

[ŠŽK21] Gustav Šourek, Filip Železnỳ, and Ondřej Kuželka. Beyond graph neural networks with lifted relational neural networks. *Machine Learning*, pages 1–44, 2021.

[TA18] Naoya Takeishi and Kosuke Akimoto. Knowledge-based distant regularization in learning probabilistic models. *arXiv preprint arXiv:1806.11332*, 2018.

[Tan97] Ah Hwee Tan. Cascade ARTMAP: Integrating neural computation and symbolic knowledge processing. *IEEE Transactions on Neural Networks*, 1997.

[THDS15] Eric Tzeng, Judy Hoffman, Trevor Darrell, and Kate Saenko. Simultaneous deep transfer across domains and tasks. In *Proceedings of the IEEE international conference on computer vision*, pages 4068–4076, 2015.

[THM21] Efthymia Tsamoura, Timothy Hospedales, and Loizos Michael. Neural-symbolic integration: A compositional perspective. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(6):5051–5060, May 2021.

[TS93] Geoffrey G Towell and Jude W Shavlik. Extracting refined rules from knowledge-based neural networks. *Machine learning*, 13(1):71–101, 1993.

[TS94] Geoffrey G Towell and Jude W Shavlik. Knowledge-based artificial neural networks. *Artificial intelligence*, 70(1-2):119–165, 1994.

[TSN90] Geofrey G Towell, Jude W Shavlik, and Michiel O Noordewier. Refinement of approximate domain theories by knowledge-based neural networks. In *Proceedings of the eighth National conference on Artificial intelligence*, volume 861866. Boston, MA, 1990.

[Tur50] A. M. Turing. Computing machinery and intelligence. *Mind*, 59(236):433–460, 1950.

[TVdM18] Niket Tandon, Aparna S. Varde, and Gerard de Melo. Commonsense knowledge in machine intelligence. *SIGMOD Rec.*, 46:49–52, 2018.

[VCC+18] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations*, 2018.

[VCVD02] E. Van Craenenbroeck, H. Vandecasteele, and L. Dehaspe. Dmax's functional group and ring library. https://dtai.cs.kuleuven.be/software/dmax/, 2002.

[vRMB+21] Laura von Rueden, Sebastian Mayer, Katharina Beckh, Bogdan Georgiev, Sven Giesselbach, Raoul Heese, Birgit Kirsch, Michal Walczak, Julius Pfrommer, Annika Pick, et al. Informed machine learning-a taxonomy and survey of integrating prior knowledge into learning systems. *IEEE Transactions on Knowledge and Data Engineering*, 2021.

[VSBV17] Lovekesh Vig, Ashwin Srinivasan, Michael Bain, and Ankit Verma. An investigation into the role of domain-knowledge on the use of embeddings. In Nicolas Lachiche and Christel Vrain, editors, *Inductive Logic Programming - 27th International Conference, ILP 2017, Orléans, France, September 4-6, 2017, Revised Selected Papers*, volume 10759 of *Lecture Notes in Computer Science*, pages 169–183. Springer, 2017.

[VSKB10] S Vichy N Vishwanathan, Nicol N Schraudolph, Risi Kondor, and Karsten M Borgwardt. Graph kernels. *Journal of Machine Learning Research*, 11:1201–1242, 2010.

[VSP+17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.

[WBS+15] Kevin Williams, Elizabeth Bilsland, Andrew Sparkes, Wayne Aubrey, Michael Young, Larisa N Soldatova, Kurt De Grave, Jan Ramon, Michaela De Clare, Worachart Sirawaraporn, et al. Cheaper faster drug development validated by the repositioning of drugs against neglected tropical diseases. *Journal of the Royal society Interface*, 12(104):20141289, 2015.

[WD18] Mei Wang and Weihong Deng. Deep visual domain adaptation: A survey. *Neurocomputing*, 312:135–153, 2018.

[Wei88] David Weininger. SMILES, a chemical language and information system. 1. introduction to methodology and encoding rules. *Journal of chemical information and computer sciences*, 28(1):31–36, 1988.

[Wil89] Andrew F Wilks. Two putative protein-tyrosine kinases identified by application of the polymerase chain reaction. *Proceedings of the National Academy of Sciences*, 86(5):1603–1607, 1989.

[WMMR21] Thomas Winters, G. Marra, Robin Manhaeve, and L. D. Raedt. Deepstochlog: Neural stochastic logic programming. *ArXiv*, abs/2106.12574, 2021.

[WPC+20] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S Yu Philip. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 2020.

[WZX+19] Hongwei Wang, Miao Zhao, Xing Xie, Wenjie Li, and Minyi Guo. Knowledge graph convolutional networks for recommender systems. In *The World Wide Web Conference*, WWW '19, page 3307–3313, New York, NY, USA, 2019. Association for Computing Machinery.

[XHLJ19] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations*, 2019.

[XXK+19] Yaqi Xie, Ziwei Xu, Mohan S. Kankanhalli, Kuldeep S. Meel, and Harold Soh. Embedding symbolic knowledge into deep networks. In *Advances in Neural Information Processing Systems*, 2019.

[XZF+18] Jingyi Xu, Zilu Zhang, Tal Friedman, Yitao Liang, and Guy Van Den Broeck. A semantic loss function for deep learning with symbolic knowledge. In *35th International Conference on Machine Learning, ICML 2018*, 2018.

[YLD+21] Shweta Yadav, Usha Lokala, Raminta Daniulaityte, Krishnaprasad Thirunarayan, Francois Lamy, and Amit Sheth. "when they say weed causes depression, but it's your fav antidepressant": Knowledge-aware attention framework for relationship extraction. *PloS one*, 16(3):e0248299, 2021.

[YN13] Shuo Yang and Sriraam Natarajan. Knowledge intensive learning: Combining qualitative constraints with causal independence for parameter learning in probabilistic models. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 580–595. Springer, 2013.

[YZQ+19] Changchang Yin, Rongjian Zhao, Buyue Qian, Xin Lv, and Ping Zhang. Domain knowledge guided deep learning with electronic health records. In *2019 IEEE International Conference on Data Mining (ICDM)*, pages 738–747. IEEE, 2019.

[ZCH+20] Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. Graph neural networks: A review of methods and applications. *AI Open*, 1:57–81, 2020.

[ZKK+21] Jieyu Zhao, Daniel Khashabi, Tushar Khot, Ashish Sabharwal, and Kai-Wei Chang. Ethical-advice taker: Do language models understand natural language interventions? In *Findings of the Association for Computational Linguistics: ACL-IJCNLP*, pages 4158–4164, 2021.

[ZLLS21] Aston Zhang, Zachary C. Lipton, Mu Li, and Alexander J. Smola. Dive into deep learning. *arXiv preprint arXiv:2106.11342*, 2021.

[ZQD+20] Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1):43–76, 2020.

[ZWQ+19] Zirui Zhuang, Jingyu Wang, Qi Qi, Haifeng Sun, and Jianxin Liao. Toward greater intelligence in route planning: A graph-aware deep learning approach. *IEEE Systems Journal*, 14(2):1658–1669, 2019.

[ZYZZ18] Daokun Zhang, Jie Yin, Xingquan Zhu, and Chengqi Zhang. Network representation learning: A survey. *IEEE transactions on Big Data*, 2018.