



## References

- [1] Donald E. Knuth. *The T<sub>E</sub>Xbook*. Addison-Wesley, 1984.
- [2] Leslie Lamport. *L<sub>A</sub>T<sub>E</sub>X: A Document Preparation System*. Addison-Wesley, 2nd edition, 1994.
- [3] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.
- [4] Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Pearson, 3rd edition, 2016.
- [5] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- [6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *\*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)\**, 2016.
- [7] Ashish Vaswani et al. Attention is all you need. In *\*Advances in Neural Information Processing Systems (NeurIPS)\**, 2017.
- [8] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *\*Nature\**, 521(7553):436–444, 2015.
- [9] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *\*Proceedings of NAACL-HLT\**, 2019.
- [10] François Chollet. Keras. <https://keras.io>, 2015.