

Giant List of AI/Machine Learning Tools & Datasets

AI/machine learning technology is growing at a rapid pace. There is a great deal of active research & big tech is leading the way. Luckily there are also a lot of resources out there for the technologist to utilize. So many we had to cherry pick what look like the most legit & useful tools.

1. Accord Framework
<http://accord-framework.net>
2. Aligned Face Dataset from Pinterest (CCO)
<https://www.kaggle.com/frules11/pins-face-recognition>
3. Amazon Reviews Dataset
<https://snap.stanford.edu/data/web-Amazon.html>
4. Apache SystemML
<https://systemml.apache.org>
5. AWS Open Data
<https://registry.opendata.aws>
6. Baidu Apolloscapes
<http://apolloscape.auto>
7. Beijing Laboratory of Intelligent Information Technology Vehicle Dataset
<http://iitlab.bit.edu.cn/mcislabs/vehicledb>
8. Berkley Caffe
<http://caffe.berkeleyvision.org>
9. Berkley DeepDrive
<https://bdd-data.berkeley.edu>
10. Caltech Dataset
<http://www.vision.caltech.edu/html-files/archive.html>

11. Cats in Movies Dataset
<https://public.opendatasoft.com/explore/dataset/cats-in-movies/information>
12. Chinese Character Dataset
http://www.iapr-tc11.org/mediawiki/index.php?title=Harbin_Institute_of_Technology_Opening_Recognition_Corpus_for_Chinese_Character_Recognition_2011_CVPR_WOR3C
13. Chinese Text in the Wild Dataset (CC4.0)
<https://ctwdataset.github.io>
14. CelebA Dataset (research only)
<http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html>
15. Cityscapes Dataset
<https://www.cityscapes-dataset.com> | [License](#)
16. Clash of Clans User Comments Dataset (GPL 2)
<https://www.kaggle.com/moradnejad/clash-of-clans-50000-user-comments>
17. Core ML
<https://developer.apple.com/machine-learning>
18. Cornell Movie Dialogs Corpus
http://www.cs.cornell.edu/~cristian/Cornell_Movie-Dialogs_Corpus.html
19. Deep Learning for Java
<https://deeplearning4j.org>
20. Enron Email Dataset
<https://www.cs.cmu.edu/~./enron>
21. Facebook AI Tools
<https://ai.facebook.com/tools>
22. GitHub Deep Learning
<https://github.com/topics/deep-learning>
23. GitHub Machine Learning
<https://github.com/topics/machine-learning>

24. GitHub Natural Language Processing
<https://github.com/topics/nlp>
25. GitHub Tensorflow
<https://github.com/topics/tensorflow>
26. Google Dataset Search
<https://toolbox.google.com/datasetsearch>
27. Google Facial Expression Comparison Dataset (CC0 1.0)
<https://ai.google/tools/datasets/google-facial-expression>
28. Google Landmarks Dataset
<https://www.kaggle.com/google/google-landmarks-dataset>
29. Google ML Kit
<https://developers.google.com/ml-kit>
30. Google Open Images Dataset
<https://ai.googleblog.com/2016/09/introducing-open-images-dataset.html>
31. Google Teachable Machine
<https://teachablemachine.withgoogle.com>
32. H2O AI
<https://www.h2o.ai>
33. IBM Watson Starter Kits
<https://cloud.ibm.com/developer/watson/starter-kits>
34. IMDB Movie Review Dataset
<http://ai.stanford.edu/~amaas/data/sentiment>
35. Imagenet Image Database
<http://image-net.org>
36. JVC Video Game Reviews Dataset
<https://www.kaggle.com/floval/jvc-game-reviews>
37. Kaggle Datasets
<https://www.kaggle.com>

38. Labeled Faces in the Wild
<http://vis-www.cs.umass.edu/lfw>
39. LabelMe Dataset
<http://labelme.csail.mit.edu/Release3.0/browserTools/php/dataset.php>
40. LISA Traffic Light Dataset (CC BY-NC-SA 4.0)
<https://www.kaggle.com/mbornoe/lisa-traffic-light-dataset>
41. Machine Learning Playground
<http://ml-playground.com>
42. Machine Learning Showcase
<https://ml-showcase.com>
43. Mahout
<https://mahout.apache.org>
44. Microsoft Cognitive Toolkit
<https://docs.microsoft.com/en-us/cognitive-toolkit>
45. Microsoft Distributed Machine Learning Toolkit
<http://www.dmtk.io>
46. Million Song Dataset
<http://millionsongdataset.com>
47. MLlib
<https://spark.apache.org/mllib>
48. Movie Review Datasets
<http://www.cs.cornell.edu/people/pabo/movie-review-data>
49. MovieLens Datasets
<https://grouplens.org/datasets/movielens>
50. Mushroom Dataset
<https://archive.ics.uci.edu/ml/datasets/mushroom>
51. MXNet
<https://mxnet.apache.org>

52. Mycroft
<https://mycroft.ai>
53. Natural Earth Data
<http://www.naturearthdata.com/downloads>
54. Numenta
<https://numenta.com>
55. ONNX
<https://onnx.ai>
56. Open ML Datasets
<https://www.openml.org/search?type=data>
57. OpenCyc
<https://www.cyc.com/opencyc>
58. OpenNN
<http://www.opennn.net>
59. Oryx 2
<http://oryx.io>
60. Oxford Robotcar Dataset (CC4.0)
<https://robotcar-dataset.robots.ox.ac.uk>
61. PredictionIO
<http://predictionio.apache.org>
62. Price of Weed Dataset
<https://github.com/frankbi/price-of-weed>
63. PyTorch
<https://pytorch.org>
64. Real & Fake Face Detection
<https://www.kaggle.com/ciplab/real-and-fake-face-detection>
65. Scikit-learn
<https://scikit-learn.org>

66. Shogun
<https://www.shogun-toolbox.org>
67. Stanford Cars Dataset
http://ai.stanford.edu/~jkrause/cars/car_dataset.html
68. Stanford Dogs Dataset
<http://vision.stanford.edu/aditya86/ImageNetDogs>
69. Stanford Large Network Dataset Collection
<https://snap.stanford.edu/data>
70. Stanford Sentiment Treebank
<https://nlp.stanford.edu/sentiment/code.html>
71. The Blog Authorship Corpus (research only)
<http://u.cs.biu.ac.il/~koppel/BlogCorpus.htm>
72. The French Lexicon Project
<https://sites.google.com/site/frenchlexicon/results>
73. Theano
<http://www.deeplearning.net/software/theano>
74. Tensorflow
<https://www.tensorflow.org>
75. TME Motorway Dataset (research only)
<http://cmp.felk.cvut.cz/data/motorway>
76. Torch
<http://torch.ch>
77. Tufts Face Database (research only)
<http://tdface.ece.tufts.edu>
78. UCI Machine Learning Repository
<http://archive.ics.uci.edu/ml/index.php>
79. UFO Reports Dataset
<https://github.com/planetsig/ufo-reports>

80. Vandal Video Game Reviews Dataset
<https://www.kaggle.com/floval/12-000-video-game-reviews-from-vandal>
81. Visual Genome
<http://visualgenome.org>
82. Wacky Corpus (CC BY-NC-SA 4.0)
<https://wacky.sslmit.unibo.it/doku.php?id=corpora>
83. Wine Quality Dataset
<https://archive.ics.uci.edu/ml/datasets/wine+quality>
84. World Bank Open Data
<https://data.worldbank.org>
85. Yale Face Database (research only)
<http://cvc.cs.yale.edu/cvc/projects/yalefaces/yalefaces.html>
86. Yelp Open Dataset (research only)
<https://www.yelp.com/dataset>
87. YouTube-8M Segments Dataset
<https://research.google.com/youtube8m>

Big Tech R&D

1. AI2
<https://allenai.org>
2. AWS Machine Learning
<https://aws.amazon.com/machine-learning>
3. Baidu Research
<http://research.baidu.com/Blog>
4. Berkeley Artificial Intelligence Research (BAIR)
<https://bair.berkeley.edu>

5. DeepMind
<https://deepmind.com>
6. Duolingo AI
<https://ai.duolingo.com>
7. Energy.gov
<https://www.energy.gov/artificial-intelligence-and-machine-learning>
8. Facebook AI
<https://ai.facebook.com>
9. Facebook AI Research
<https://research.fb.com/category/facebook-ai-research>
10. GE Artificial Intelligence
<https://www.ge.com/research/technology-domains/artificial-intelligence>
11. Google AI
<https://ai.google>
12. Google AI & Machine Learning Products
<https://cloud.google.com/products/ai>
13. IBM Research AI
<https://www.research.ibm.com/artificial-intelligence>
14. Intel AI
<https://software.intel.com/en-us/ai>
15. Journal of Artificial Intelligence Research (JAIR)
<https://www.jair.org>
16. Microsoft Artificial Intelligence
<https://www.microsoft.com/en-us/research/research-area/artificial-intelligence>
17. OpenAI
<https://openai.com>
18. Partnership on AI
<https://www.partnershiponai.org>

19. TayTweets

<https://twitter.com/tayandyou>

Let us know if we missed your favorite AI/machine learning tool or dataset. Also be sure to check out places to [educate yourself about AI/machine learning](#) & [AI/machine learning events](#).

This data is from [Vuild's list of AI/machine learning tools & datasets](#). Please visit vuild.com for more.