

# Index

- 3CosAdd, 2173
- 3CosMul, 2173
- 5-V paradigm, 1955, 1963
- 7-V paradigm, 1963
- ablations, 2174
- access (transport), 1912
- access control, 1975, 1983, 1993
- actions, 2004
- ad targeting, 1904
- AdaDelta, 2052
- AdaGrad, 2051–2053
- Adam, 2052, 2053, 2058, 2070
- adapters, 2164
- adaptive learning rates, 2053
- adversarial training, 2077
- aggregation, 1963, 1972, 1976, 1980, 1981, 1984, 1986, 1987, 2243
  - tree reduction, 1985
- AIC, 2191
- airline networks, 1910
- airline scheduling, 1910
- algorithmic bias, 1895
- algorithms
  - machine learning, 2029
- Alzheimer's disease, 1903
- ambiguity, 2132
- ambiguity (natural language), 2112
- Amdahl's Law, 1977
- American Mathematical Society (AMS), 1935
- anisotropy, 2173
- anomaly detection, 2020
- anonymization, 1895
- antonym, 2134
- antonymy, 1903
- Apache Hadoop, 1969, 1986
- Apache Software Foundation, 1904
- Apache Spark, 1904, 1956, 1969, 1973, 1980, 1981, 1983, 1986, 2004, 2012
  - DataFrames, 2007
  - RDDs, 2007
- applications, 1875, 1905
- architectures
  - specialized, 2076
- arcplot, 1892
- arithmetic intensity, 1978
- Arrow, 1980, 1983
- artificial intelligence (AI), 1876, 2029, 2109
- Asimov Institute, 2076
- assignment step, 2004
- Assist Score, 1916
- assistive technology, 2134
- assists, 1916
  - primary assists, 1916
  - secondary assists, 1916
- associativity, 1985
- attention, 2176, 2177
  - multi-head, 2082
  - scaled dot-product, 2082
- attributes (networks), 1885
  - edge attributes, 1885
  - node attributes, 1885
- AUC-ROC, 2023
- author-topic model, 2198
- autoencoders, 2030, 2053, 2079, 2099
  - de-noising, 2080, 2099
  - regularized, 2080
  - undercomplete, 2079
- automatic speech recognition, 2133, 2160
- automatic summarization, 2254
  - abstractive, 2254
  - extractive, 2254
- automation bias, 1966
- autoregressive generation, 2179
- back-off, 2161
- backpropagation, 2030, 2111
  - through time, 2064
- Bacon number, 1877, 1878
- Bacon, Kevin, 1877
- bag-of-words, 2153, 2156, 2204
- bag-of-words (BoW), 2235, 2240
- bagging, 2049
- balance theory, 1886
- Bayes' rule, 2194
- Bayesian nonparametrics, 2198
- beam search, 2083
- benchmarking, 1976, 1978, 1993
  - best practices, 1976
  - pitfalls, 1977
- benchmarks
  - gaming, 1978
- beneficence, 1895
- Bernoulli mixture model (BMM), 2190
- BERT, 2064, 2083
- bias, 2170
- bias-variance trade-off, 2114
- BIC, 2191
- big data, 1875, 1877, 1904, 1955, 1956, 1960, 1962, 1968
- R
  - ladder of escalation, 1979
- constraints, 1962, 1968
- debugging, 1967
- ethics, 1966
- Five Vs, 1904
- historical data, 1965
- monitoring, 1973
- moving target, 1966
- questions, 1966
- software, 1956, 1978, 1979, 1986
- stakes, 1965
- tools, 1904
- validation, 1967
  - provenance, 1967
  - robustness, 1968
  - sanity checks, 1967, 1983, 1994

time series data, 1967  
 value, 1904  
 variety, 1904  
 velocity, 1904  
 veracity, 1904  
 volume, 1904  
 workflow, 1962  
 bigram, 2245  
 bigrams, 2114, 2148  
     signature, 2124  
 binary min-heap, 1899  
 biodiversity conservation, 1909  
 biological networks, 1885  
 BlueDot, 1873, 1875, 1876  
 BookCrossing, 1914  
 Boston Housing dataset, 2089  
 bot detection, 1901, 1902  
 botnets, 1907  
 bots, 1901, 1902  
 bottleneck, 1968, 1970, 1976–1978, 1982, 1987, 1992  
 broadcast, 1973, 1982  
 brokers (networks), 1893, 1894  
     brokerage, 1893  
 bubble graph, 1892  
 buffer scheduling, 1912  
  
 C-style order, 2037  
 cache, 1974, 1980–1982, 1986, 1988, 1992  
 CANDECOMP, 2035  
 canonical polyadic decomposition, 2035  
 capacity limits, 1911  
 cascading failures, 1874  
 causal delay networks, 1912  
 centrality, 1874, 1875, 1878, 1887, 1889, 1893, 1894, 1901, 1902, 1905, 1907, 1908, 1910, 1916  
     betweenness centrality, 1887, 1895, 1902, 1905, 1910  
     temporal, 1901  
     closeness centrality, 1887, 1902  
     degree centrality, 1887, 1905  
     eigenvector centrality, 1902, 1905  
     centroid, 2004, 2009  
     CERN, 1969  
     character n-grams, 2163  
     citation networks, 1885, 1902  
     civic engagement, 1907  
     class imbalance, 2020, 2023  
     classification and regression tree (CART), 2123  
     classifier, 2086  
     clickstream analysis, 1973  
     cliques, 1889  
     clock speed, 1977  
     closed groups (social media), 1907  
     closeness centrality  
         temporal closeness, 1901  
     cloud, 1975  
     cloud computing, 1975, 1992, 1997  
     cluster, 1975  
     cluster computing, 1975, 1978, 1992  
     cluster purity, 2011, 2012  
     clustering, 1874, 1878, 1889, 1901, 1903, 1905, 2004, 2009  
         centroid, 2009  
         clustering coefficient, 1881, 1886, 1888  
         global clustering coefficient, 1888  
         graph clustering, 1903  
     clustering coefficient, 1902  
     clusters, 1914  
     co-liked books, 1914  
     co-occurrence networks, 1875, 1903  
     co-reference, 2111  
     cohesion (networks), 1886  
     collaboration networks, 1873  
     collocation, 2236  
     colonization rates, 1908  
     columnar storage, 1988  
     communication, 1970, 1981  
     communication networks, 1884, 1901  
     community detection, 1874, 1875, 1896, 1903, 1905, 1914  
     aggregation, 1897  
     community structure, 1873–1875, 1889, 1893, 1896, 1897, 1902, 1905, 1914  
         Louvain method, 1874, 1875, 1896, 1897, 1903, 1914  
     modularity, 1909  
     modularity maximization, 1905  
 community structure definitions, 1896  
 hierarchical communities, 1897  
 strong community, 1896  
 weak community, 1896  
 community-driven mapping, 1912  
 commutativity, 1985  
 compiler, 1976  
 complete graphs, 1900  
 composite transit networks, 1912  
 compositionality, 2112  
 computational feasibility, 1955, 1964  
 compute-bound, 1978  
 computer vision, 2054  
 computing near the data, 1972, 1980, 1981, 1988, 1992  
 concept drift, 1965  
 conceptual semantics, 2134  
 concordance, 2228  
 conflict (networks), 1886  
 confusion matrix, 2122  
 congestion, 1910  
 connected graphs, 1882  
 connections (networks), 1877, 1907  
 connectivity, 1875, 1881, 1882, 1896, 1909, 1912  
     algebraic connectivity, 1908  
     connected components, 1875, 1882, 1889, 1904, 1908  
     edge connectivity, 1900  
     robustness, 1900  
     strong connectivity, 1883

vertex connectivity, 1900  
 weak connectivity, 1883  
 consent, 1894  
 constituent order, 2132  
 constraints, 1955, 1979  
 consumer analytics, 1904  
 content determination, 2243  
 content moderation, 2054  
 context window, 2168  
 contextual embeddings, 2111  
 Continuous Bag-of-Words (CBoW), 2167, 2168  
 contradiction, 2175  
 convolution, 2054  
 convolutional neural networks (CNN), 2030, 2053, 2054, 2058, 2097, 2111, 2224  
 coordinated inauthentic behaviour, 1907  
 coordination, 1970, 1978, 1979, 1981  
 corpus-based learning, 2111  
 correlation and causation, 1965  
 corridors (networks), 1889  
 cosine similarity, 2164, 2165, 2171, 2255  
 cost (networks), 1910  
 cost and performance, 1975, 1997  
 cost model, 1975, 1994  
 cost-performance trade-off, 1978  
 covariance matrix, 2189  
 covariate shift, 1965  
 COVID-19, 1877  
 CP decomposition, 2035, 2036  
 CPU, 1974  
   cores, 2004  
 CPU core, 1974, 1975  
 critical routes, 1910  
 cross-community interaction, 1905  
 cross-correlation, 2055  
 cross-entropy loss, 2070, 2083, 2162, 2176  
 cross-validation, 2049, 2089, 2090  
 CSV, 1987, 1997  
 Dalian, 1912  
 damping factor, 2255  
 data cleaning, 1904, 1981  
 data collection, 1884, 1893  
 data engineering, 1962, 1981  
 data ingestion, 1963  
 data locality, 1972, 1976, 1978, 1980, 1982, 1988  
 data movement, 1972, 1976, 1978, 1980, 1982, 1986–1988  
 data ownership, 1895, 1983  
 data parallelism, 1981  
 data quality, 1963  
 data reduction, 1980, 1981, 1992, 2009, 2235  
 data sources, 1876  
 data transfer, 1974  
 data-dictated decisions, 1966  
 data-informed decisions, 1966  
 data.table, 1980, 1983  
 datacenter, 1975  
 DataFrames, 1987  
 Davidson–Harel layout, 1889  
 de-duplication, 1973  
 de-noising autoencoder, 2099  
 debugging, 1973, 1975, 1991  
 decay factor, 2009  
 decision trees, 2016  
 decision-making, 1966, 1983  
   data-dictated decisions, 1966  
   data-informed decisions, 1966  
 decoder, 2082, 2177  
 deduplication, 1904  
 deep learning  
   examples, 2083  
   expressive power, 2084  
   frameworks, 2083  
   networks, 2083  
 deep learning (DL), 1981, 2029, 2048, 2109, 2111  
 deep networks, 2049, 2050  
 degree, 1875, 1881, 1897, 1901  
   in-degree, 1881, 1887, 1902, 1916  
   out-degree, 1881, 1887, 1902, 1916  
 degree distribution, 1874  
 delays (transport), 1910, 1911  
 dendrogram, 1890, 1892  
 Dennard scaling, 1977  
 denotation, 2132  
 dense graphs, 1884  
 dependence, 1873  
 dependency grammar, 2133, 2152  
 derivation, 2163  
 design matrix, 2121  
 device-level parallelism, 1974  
 diacritics, 2114  
 diameter (networks), 1888  
 diffusion on networks, 1873, 1875, 1876, 1885, 1903  
 diffusion processes, 1903  
 Digital Matatus Project, 1913  
 Dijkstra’s algorithm, 1874  
   unvisited nodes, 1898  
   visited nodes, 1898  
 dimension reduction, 2079  
 directed acyclic graph (DAG), 1986  
 Dirichlet distribution, 2185, 2196  
 Dirichlet prior, 2196  
 disambiguation, 1904, 2151  
 disconnected graphs, 1882  
 discrete symbols, 2132  
 discrete variables, 2112  
 discriminator (GAN), 2077  
 disinformation campaigns, 1907  
 dispersal corridors, 1908  
 dispersal kernel, 1908  
 dispersal pathways, 1907  
 disruptions (transport), 1910, 1911  
 distance (networks), 1910  
 distance matrix, 1899  
 distributed computing, 1955, 1960, 1968, 1969, 1979–1981, 1988  
   bottleneck, 1968  
   distributed datasets, 2004  
   distributed memory, 1969  
   workers, 1968, 1975, 1982, 1988, 1993

distributed graph processing, 1904

distributed optimization, 1972, 1981

distributed representations, 2165, 2166

distributed systems, 1904, 1982, 2007

distributed training, 1981

distributional hypothesis, 2164

distributional semantics, 2163, 2165

diverse viewpoints, 1905

diversity, 1895

divide-and-aggregate, 1970, 1976, 2004

document structuring, 2243

driver, 1988, 1993

dropout, 2049, 2050, 2070, 2097, 2157, 2217

dyads, 1875, 1886

- directed dyads, 1886
- undirected dyads, 1886

dynamic networks

- signal vs noise, 1901
- tie dynamics, 1900
- time-dependent measures, 1901

dynamic routing, 1910

dynamic topic model, 2198

early stopping, 2049, 2059, 2072, 2094

Ebola, 1877

echo chambers, 1905

ecological networks, 1873, 1905, 1907

economic graphs, 1904

edge, 1873, 1879, 1897, 1905, 1907, 1910

edge cases, 2112

ego networks, 1889

eigenvalues and eigenvectors, 1903

elastic computing, 1975

elastic net regularization, 2049

ELIZA, 2109

ELIZA effect, 2109

embarrassingly parallel, 1976

embedding layer, 2199

embedding matrix, 2162

embeddings, 1902, 2148, 2154, 2163, 2178

encoder, 2082, 2177

encoder-decoder models, 2064, 2066

energy efficiency, 1977

ensemble clustering for graphs (ECG), 1874

entailment, 2175

entities, 1873, 1879, 1904

entropy loss, 2126

epochs, 2049, 2059, 2061, 2085

equity (transport), 1912

Erdős collaboration network, 1875

Erdős number, 1935

estimators, 2012

ethics, 1874, 1875, 1893

Euclidean distance, 2004

evaluation metrics

- BLEU, 2161
- F1 score, 2256
- precision, 2256
- recall, 2256
- ROUGE, 2256

event data, 1962

event logs, 1962

evidence lower bound (ELBO), 2081

examples, 1875

exclusion (networks), 1894

executors, 1988, 1994

Expectation-Maximization (EM) algorithm, 2189

exploding gradients, 2064

exploratory data analysis, 1964

F1 score, 2023

Facebook, 1904, 1905

facial recognition, 2054

fact-checking, 1907

failure mode

- organizational, 1967
- statistical, 1967

failure modes, 1975, 1979, 1983

fairness, 1895, 1966

false positives, 1967

fault tolerance, 1975, 1982, 1987

feature extraction, 1902, 2111, 2132, 2148

- edge-level features, 1902
- graph-level features, 1902

feature maps, 2054, 2055

feature matching, 2078

feature scaling, 2016, 2093

feature selection, 2049, 2079

feature vector, 2132, 2153

feed-forward network

- position-wise, 2082

feed-forward networks (FFNN), 2064

feed-forward neural networks (FFNN), 2064, 2066, 2067, 2084, 2087, 2089, 2093, 2104, 2175, 2176, 2201, 2220

filtered graph, 1914

filters (image processing), 2054

fine-tuning, 2164

flight duration, 1910

flight frequency, 1910

FLOPS, 1976, 1978

flows on networks, 1877, 1887, 1909

Floyd–Warshall algorithm, 1874, 1898

- intermediate node, 1899

folds, 2090

follows (social media), 1905

Folsomia candida, 1908

food webs, 1908

forecasting, 1877

formal linguistic frameworks, 2112

formal semantics, 2134

formal system, 2133

fragmented landscapes, 1908

frameworks, 1982

Fruchterman–Reingold layout, 1889

fully connected network, 2087

functionalist grammar, 2133

GAN, 2077

|                                                         |                                           |                                                                            |
|---------------------------------------------------------|-------------------------------------------|----------------------------------------------------------------------------|
| Gaussian mixture model<br>(GMM), 2185, 2188,<br>2189    | matrix visualizations,<br>1889            | hyponymy, 1903                                                             |
| Gaussian noise, 2099                                    | node-link diagrams, 1889,<br>1891         | hypothesis class, 2113                                                     |
| generative adversarial<br>networks (GAN),<br>2053, 2077 | radial layouts, 1889                      | hypothesis space, 2030                                                     |
| generative AI, 2177                                     | spectral layouts, 1889                    | I/O, 1964, 1968, 1974, 1976,<br>1978–1983, 1986, 1987,<br>1992, 1993, 1997 |
| generative grammar, 2133                                | techniques, 1889                          | I/O-bound, 1978                                                            |
| generative models, 2081                                 | graph-structured data, 1901               | ICL, 2191                                                                  |
| generator (GAN), 2077                                   | Graphlet kernels, 1902                    | identifiers, 1962                                                          |
| gensim, 2250                                            | GraphX, 1874, 1904                        | igraph, 1922                                                               |
| Georgetown-IBM experiment,<br>2109                      | grid, 1975                                | image classification, 2054,<br>2058                                        |
| Gibbs sampling, 2198                                    | grid computing, 1969, 1975                | image data, 2054, 2083                                                     |
| Giraph, 1874, 1904                                      | group privacy, 1895                       | image segmentation, 2054                                                   |
| GitHub, 1943                                            | group structures, 1886                    | image synthesis, 2054                                                      |
| global health, 1876                                     | group-by, 1964, 1980, 1987,<br>1992       | IMDB dataset, 2199                                                         |
| GloVe, 2171, 2210                                       | Gustafson’s Law, 1977                     | inclusivity, 1912                                                          |
| Goal Score, 1916                                        | habitat patches, 1907                     | independence, 1972                                                         |
| Goodhart’s Law, 1978                                    | habitat restoration, 1909                 | inductive bias, 2113                                                       |
| governance, 1895                                        | Hadamard product, 2053                    | infectious disease, 1873, 1876,<br>1896, 1900, 1901                        |
| government shutdown, 1911                               | ham, 2065                                 | inflection, 2163                                                           |
| GPT, 2064, 2083, 2177, 2179                             | handwritten digits, 2092                  | influence (networks), 1901,<br>1905                                        |
| GPU, 1974, 1975, 1982                                   | hapax, 2236                               | influencer marketing, 1905                                                 |
| gradient, 1972, 1981                                    | hardware, 1955, 1973                      | influential individuals, 1905                                              |
| gradient boosting, 2016                                 | clusters, 1978                            | informal signals, 1876                                                     |
| grammar, 2132                                           | cost trends, 1978                         | information extraction, 2133                                               |
| graph, 1873, 1879, 1905, 1907,<br>2254                  | economics, 1978                           | information flow, 1907                                                     |
| directed graphs,<br>1879–1881, 1883                     | hardware thread, 1974                     | information retrieval, 1875,<br>1903                                       |
| undirected graphs,<br>1879–1881                         | HDF5, 2097                                | initialization                                                             |
| unweighted graphs,<br>1880                              | HDFS, 1986                                | k-means, 2004, 2007                                                        |
| weighted graphs, 1880                                   | heavy-tailed distributions,<br>1874, 1900 | input, 2132                                                                |
| graph embedding, 1903                                   | heterogeneity, 1975                       | input (NLP), 2112, 2153                                                    |
| graph kernels, 1902                                     | heterogeneous data, 1963                  | input data, 2054                                                           |
| graph neural networks (GNN),<br>1874, 1875, 1901        | hidden state, 2064                        | input projection matrices,<br>2082                                         |
| message passing, 1902                                   | hierarchical softmax, 2162                | input-output pairs, 2029                                                   |
| graph processing, 1904                                  | high-dimensional data, 2036,<br>2154      | Instagram, 1904, 1905                                                      |
| graph similarity, 1902                                  | hinge loss, 2126                          | instruction set, 1976                                                      |
| graph theory, 1874, 1875, 1879,<br>1909                 | hiring patterns, 1904                     | intelligent agents, 2029                                                   |
| graph visualization, 1875,<br>1889                      | hockey, 1916                              | interactions (networks), 1884,<br>1886                                     |
| circular layouts, 1889                                  | holonym, 2134                             | interpolation, 2161                                                        |
| force-directed layouts,<br>1889, 1891                   | homonym, 2134                             | invasive-species management,<br>1909                                       |
| geographical layouts,<br>1889                           | hop length, 1881                          | Iris dataset, 2086                                                         |
|                                                         | horizontal scaling, 1975                  | isolated nodes, 1916                                                       |
|                                                         | hubness, 2173                             | iteration, 1969, 1983                                                      |
|                                                         | hubs, 1874, 1889, 1893, 1900,<br>1910     | iterative algorithms, 2004                                                 |
|                                                         | hyperlinks, 1919                          | iterative analysis, 1895, 1981,<br>1983, 1986                              |
|                                                         | hypernym, 2134                            |                                                                            |
|                                                         | hyperparameter, 2009, 2049                |                                                                            |
|                                                         | hyperthreading, 1974                      |                                                                            |
|                                                         | hyponym, 2134                             |                                                                            |

iterative methods, 1972  
 iterative optimization, 1972  
  
 Jaccard similarity, 2165  
 Jensen's inequality, 2193  
 job recommendations, 1904  
 job runtime, 1977  
 job search, 1904  
 join, 1964, 1976, 1980, 1981, 1987, 1992  
 justice, 1895  
  
 k-fold cross-validation, 2089  
 K-means, 2189  
 k-means, 1903, 2004, 2005, 2007, 2009, 2011, 2012  
     k-means parallel, 2004, 2005  
     k-means++, 2004, 2007  
 Kamada–Kawai layout, 1889  
 Keras, 2058, 2084, 2199  
 kernel (convolution), 2054  
 kernel size, 2057  
 key matrix, 2082  
 keystone species, 1908  
 keyword spotting, 2133  
 Khan, Kamran, 1876  
 Kneser–Ney smoothing, 2161  
 knowledge base, 1904  
 knowledge graph, 2134  
 knowledge graphs, 1873, 1875, 1904  
 Koomey's Law, 1977  
 Kronecker delta, 1897  
 Kullback–Leibler divergence (KL divergence), 2080  
  
 $L_0$  regularization, 2049  
 $L_1$  regularization, 2214  
 $L_1$  regularization, 2048, 2049, 2080, 2096  
 $L_2$  regularization, 2214  
 $L_2$  regularization, 2048, 2049  
 label propagation, 1905  
 label shift, 1965  
 label smoothing, 2083  
 labeled data, 2112  
 Lagrange multiplier, 2189, 2193, 2194  
 LaGuardia Airport, 1911  
  
 language identification, 2114, 2148  
 language modeling, 2160  
 language models, 2160  
 Laplace smoothing, 2161  
 Laplacian (graph), 1889, 1903, 1908  
 large language model, 2175, 2177  
 large language models (LLM), 2112  
 LASSO, 2048  
 latency, 1974, 1977  
 latency (networks), 1901  
 latent Dirichlet allocation (LDA), 2185, 2196, 2250  
 latent space, 2077  
 latent variable, 2175  
 latent variables, 2077  
 law of truly large numbers, 1964, 1967  
 layers (neural networks), 2030  
 lazy evaluation, 1986  
 learning curves, 2061  
 learning rate, 2050–2053, 2128  
 LEC, 2191  
 lemma, 2242  
 lemmatization, 2133, 2148, 2242  
 lexical choice, 2243  
 lexical diversity, 2233  
 lexical semantics, 2134  
 Lidstone smoothing, 2161  
 Lightning, 2069  
 likes (social media), 1905  
 linear algebra, 1976  
 linear models, 2148  
     limitations, 2084  
 linear regression, 2012  
     normal equations, 1972  
 linear separability, 2084  
 linguistic properties, 2132  
 link prediction, 1901, 1902  
 linked-document topic model, 2198  
 LinkedIn, 1902, 1904  
     Economic Graph, 1904  
     Knowledge Graph, 1904  
 Little's Law, 1977  
  
 LLM, 2175, 2177  
 load (networks), 1910  
 local computation, 1972  
 log loss, 2126  
 log-likelihood, 2190  
 logistic regression, 1902, 2022  
 logistics, 1910, 1911  
 long distance links, 1878  
 loss functions, 2030, 2084  
 Louvain method  
     aggregation phase, 1897  
     local moving phase, 1897  
  
 machine learning, 1974  
 machine learning (ML), 1874–1876, 1901, 1904, 1981, 2029, 2109, 2148  
 machine translation, 1903, 2064, 2109, 2111, 2133, 2134, 2160  
 macroscale, 1875  
 map phase, 2004, 2007, 2009  
 MapReduce, 1955, 1960, 1970, 1984, 1987, 2004, 2009, 2012  
     map, 1970, 1984, 1987  
     reduce, 1970, 1984, 1987  
 marginalized individuals, 1894  
 Markov chain, 2243  
     text generation, 2245  
 Markov clustering (MCL), 1903  
 Markov hypothesis, 2160  
 matatus, 1912  
 MathSciNet, 1935  
 MATLAB-style order, 2037  
 matrix factorization, 2035  
 matrix multiplication, 2036  
 max pooling, 2057, 2058  
 maximum a posteriori estimation, 2187  
 maximum likelihood estimation, 2161, 2187  
 MDL, 2191  
 mean  
     responsibility-weighted mean, 2189

mean absolute error (MAE), 2091  
 mean squared error (MSE), 2070, 2084  
 meaning, 2132  
 measurement error, 1955, 1963  
 medical imaging, 2054  
 memes, 2113  
 memory, 1974, 1975  
 memory (RNN), 2064  
 memory bandwidth, 1974, 1976–1978  
 memory limits, 1964, 1968, 1979–1981, 1983, 1992, 1993  
 memory wall, 1977  
 memory-bound, 1978  
 Menlo Report, 1895  
 mentions (social media), 1905  
 meronym, 2134  
 meronymy, 1903  
 mesoscale structure, 1874, 1875  
 messages (social media), 1905  
 micro-influencers, 1905  
 microscale, 1875  
 MIDI, 2068  
 Milgram, Stanley, 1877  
 mini-batch, 1981, 2009, 2049, 2053, 2103  
     discrimination, 2078  
 minimax optimization, 2077  
 MIPS, 1976, 1978  
 misinformation, 1907  
 missed connections, 1910  
 missing values, 1955, 1963  
 mixture model, 2185  
 MLP, 2176  
 MML, 2191  
 MNIST dataset, 2092  
 mode collapse, 2078  
 mode hopping, 2078  
 model class, 2113  
 model evaluation, 2023  
 model parallelism, 1981  
 model parameters, 2163  
 model-based clustering, 2190  
 modularity, 1874, 1896  
     modularity value, 1897  
 molecular graphs, 1901  
 momentum, 2050, 2051, 2053  
 monitoring, 1964, 1975, 1982, 1983, 1997  
 Moore’s Law, 1977  
 morpheme, 2133  
 morphological segmentation, 2133  
 morphologically rich  
     languages, 2112  
 morphology, 2163  
 multi-core processor, 1974  
 multi-token expressions, 2170  
 multicollinearity, 2121  
 MultiDiGraph, 1930  
 multilayer perceptron (MLP), 2022  
 multilayer perceptrons, 2030  
 multilingual contexts, 2125  
 multiple comparisons, 1967  
 music generation, 2068  
 mutual neighbours, 1902  
  
 n-gram language models, 2161  
 Nairobi, 1912  
 named-entity recognition, 2150  
 named-entity recognition (NER), 2150  
 National Hockey League (NHL), 1916  
 natural language data, 2132  
 natural language generation (NLG), 2243  
 natural language inference, 2175  
 natural language processing (NLP), 1875, 1876, 1903, 2064, 2083, 2109, 2148, 2153, 2175, 2199, 2226  
 Natural Language Toolkit (NLTK), 2226  
 nearest neighbours, 1903, 2171  
 NEC, 2191  
 negative cycles, 1899  
 negative sampling, 2168  
 neologisms, 2113  
 Nesterov momentum, 2051  
  
 network analysis, 1876, 1877, 1883, 1893, 1904  
     case studies, 1875  
     methods, 1896  
 network centric methods, 1877  
 network data, 1876, 1883  
     adjacency list, 1884  
     adjacency matrix, 1884, 1897  
     edge list, 1883, 1884  
     machine learning, 1901  
     representations, 1883  
 network data analysis, 1873, 1875–1877  
 network effects, 1912  
 network measures, 1887  
 network models, 1877, 1878  
 Network Repository, 1935  
 network snapshot, 1885  
 network structure, 1911  
 networks, 1873, 1876–1878  
     big data, 1904  
     neural networks, 2030  
 NetworkX, 1922  
 neural language models, 2162  
 neural networks, 2030, 2100, 2111, 2148  
 neutral relation, 2175  
 NHL public API, 1941  
 NLI, 2175  
 node, 1873, 1879, 1904, 1905, 1907, 1910  
 node classification, 1902  
 node deletion, 1908  
 node strength, 1914  
 noise-contrastive estimation, 2162  
 nonlinear clusters, 1903  
 nonlinearity, 2084  
 nonsaturating loss (GAN), 2077  
 normalization, 2089, 2101, 2164  
 null dyads, 1934  
 null models (dynamic graphs), 1901  
  
 object detection, 2054  
 object recognition, 2054  
 one-hot encoding, 2016, 2093, 2153, 2204  
 online, 2009

online communities, 1901  
 online harassment, 1907  
 online learning, 2009  
 ontology, 2134  
 opinion dynamics, 1874  
 optical character recognition, 2134  
 optical character recognition (OCR), 2134  
 optimal routes, 1910  
 optimization, 1981, 2052, 2111  
 optimizers, 2030, 2093  
 Order of the Stick (OOTS), 1922  
 ordering constraint, 1973, 1986  
 organizational cost, 1975, 1980, 1983  
 organizational research, 1901  
 Ottawa Senators, 1916  
 out-of-vocabulary, 2157  
 out-of-vocabulary (OOV), 2170  
 outbreak detection, 1876, 1877  
 outer product, 2035  
 output (NLP), 2112  
 output projection matrix, 2082  
 overfitting, 1965, 2048, 2049, 2059, 2060, 2122, 2157, 2201  
 overload, 1910  
 oversampling, 2020  
  
 p-hacking, 1967  
 padding, 2157  
 padding (convolution), 2056  
 PageRank, 1904, 1919, 2254  
 Palmer Penguins dataset, 2005, 2009, 2012  
 PARAFAC, 2035  
 parallel computing, 1969, 1970, 1974  
     concurrency, 1969  
     limits, 1973, 1977  
     scaled workloads, 1977  
 parallel efficiency, 1978  
 parallelism, 2004  
 Parquet, 1988, 1997, 2016  
 parse tree, 2133  
 parsing, 2133, 2152, 2158  
  
 part-of-speech (POS), 2133, 2157, 2239  
     tagging, 2133, 2148, 2157, 2240  
 partitioning, 1973, 1988  
 path length, 1874, 1878, 1881  
     average path length, 1887  
 paths, 1881, 1887, 1896, 1897  
 performance, 1992  
     achieved, 1976  
     peak, 1976  
 perplexity, 2161  
 persistence (ecosystems), 1907, 1908  
 pipeline, 1976, 2007, 2017  
 pipelines, 2012  
 player-interaction networks, 1916  
 PLSA, 2185  
 pointwise mutual information (PMI), 2165  
 political polarization, 1905  
 polysemy, 2134, 2170  
 pooling, 2057  
 Porter stemmer, 2242  
 position (networks), 1875, 1893  
 post-processing, 2134  
 posterior probability, 2189  
 power, 1977  
 power dynamics, 1894  
 power method, 2255  
 PPMI, 2165  
 pre-trained embeddings, 2164  
 pre-training, 2163  
 preconditioner, 2053  
 predicate pushdown, 1987, 1988  
 prediction, 1877  
 prediction tasks (graphs), 1901  
     edge-level tasks, 1902  
     graph-level tasks, 1901, 1902  
     node-level tasks, 1902  
 preferential attachment, 1900  
     link-centric preferential attachment, 1900  
 Pregel, 1874, 1904  
 preliminaries, 1878  
  
 premature scaling, 1983, 1997  
 principal component analysis (PCA), 2079  
 privacy, 1874, 1894, 1895, 1966  
 probabilistic latent semantic analysis (PLSA), 2191, 2195  
 probability, 1874  
 probability distribution, 2133  
 probability simplex, 2162  
 professional sports, 1916  
 projection matrix  
     input, 2082  
     output, 2082  
 provenance, 1962, 1983  
 proxy, 1965, 1978  
 proxy variables, 1966  
 pruning, 2049  
 public health, 1901  
 public interest, 1895  
 public transport, 1910  
 PySpark, 1981, 1987, 2004, 2007, 2012  
 Python, 1922, 2226  
  
 query matrix, 2082  
 question answering, 2111, 2134  
 queueing, 1975, 1977, 1982, 1986  
  
 R (programming language), 1956, 1979  
 R-squared, 2013  
 RAG, 2179  
 random forests, 1902, 2016  
 random graphs, 1874, 1888, 1896  
 random initialization, 2163  
 random seed, 2091  
 random walks, 1903  
 random-walk distance, 1881  
 RDD, 1987  
 re-identification, 1895  
 re-routing, 1910  
 re-sharing (social media), 1907  
 reachability, 1874  
     over intervals, 1901  
 real time data, 1876, 1910  
     travel, 1876

realization, 2243  
recommendation systems, 1905, 1907  
recruiter search, 1904  
recurrent neural networks (RNN), 2030, 2053, 2064, 2111, 2199  
    bidirectional, 2066  
    GRU, 2064, 2067, 2201  
    LSTM, 2064, 2067, 2104, 2220  
reduce, 1973, 1981  
reduce phase, 2005, 2007, 2009  
referring expression generation, 2243  
reflexivity, 1895  
regime change, 1965  
    concept drift, 1965  
    covariate shift, 1965  
    label shift, 1965  
regression, 2012, 2089  
regularization, 2012, 2013, 2048–2050, 2070, 2080, 2094, 2157  
     $L_1$ , 2096  
Reingold–Tilford layout, 1889  
relational data, 1904  
relationships (networks), 1873, 1877, 1879, 1883, 1884, 1904, 1905  
reliability, 1968, 1982  
ReLU, 2057, 2058, 2176  
reparameterization trick, 2081  
representations (machine learning), 2029, 2111  
reproducibility, 1983  
residual connections, 2082  
resilience (ecosystems), 1907, 1909  
resilience planning, 1909  
resilience strategies, 1911  
Resilient Distributed Datasets (RDDs), 1904  
resistance distance, 1881  
resolution (reporting), 1895  
resources (networks), 1893  
respect for persons, 1895  
responsibility, 2186, 2188  
responsible data science, 1966  
retrieval-augmented generation, 2179  
ride-share algorithms, 1910  
ridge regression, 2048  
RMSE, 2013  
RMSprop, 2052, 2053  
robustness, 1896, 1911, 1912  
    articulation points, 1900  
    bottlenecks, 1874, 1887, 1889, 1894, 1908, 1910  
    bridges, 1878, 1887, 1889, 1893, 1894, 1905  
    cascading failures, 1908, 1909  
roles (networks), 1874, 1875, 1889, 1893, 1895  
Roofline model, 1978  
routing, 1911  
row-major order, 2037  
sampled softmax, 2162  
scale free networks, 1874, 1900  
scaling, 1974, 1975  
scaling laws, 1977  
scaling out, 1975  
scaling up, 1975  
scheduled sampling, 2083  
scheduling, 1911, 1982, 1986  
    buffers, 1912  
schema, 1987  
schema drift, 1963, 1983  
scikit-learn, 2004  
security, 1993  
selection bias, 1955  
self-normalizing objective, 2162  
self-supervised learning, 2164  
semantic analysis, 2111, 2134  
semantic coherence, 2134  
semantic networks, 1875, 1903  
semantics, 2132, 2235, 2241  
sense-annotated corpus, 2134  
sentence boundary detection, 2133  
sentence tokenizer, 2239  
sentiment analysis, 2111, 2199  
sepsis, 2020  
sequence-to-sequence models, 2064, 2065  
sequence-to-vector models, 2064, 2065  
sequential bottleneck, 1973, 1975, 1977, 1978  
sequential data, 2064  
sequential generation, 2064  
sessionization, 1973  
sharing (social media), 1907  
shortest path, 1874, 1875, 1878, 1881, 1887, 1898, 1902, 1910  
    A\* algorithm, 1910  
    Dijkstra’s algorithm, 1898, 1910  
SHRDLU, 2110  
shuffle, 1972, 1973, 1975, 1976, 1978, 1979, 1982–1984, 1986, 1987, 1992, 1993, 1997  
shuffling, 2103  
sigmoid function, 2123  
similarity graphs, 1903  
SimpleRNN, 2199  
single machine, 1968  
singular value decomposition (SVD), 2166  
SIR model, 1874  
Six Degrees of Kevin Bacon, 1873, 1875, 1877  
six degrees of separation, 1873, 1877  
skew, 1987, 1992  
skills, 1904  
skip-gram, 2167, 2168  
slang, 2113  
small data, 1955, 1962  
small world experiment, 1877  
small world network, 1873, 1874, 1878, 1888  
smoothing, 2161, 2245  
SMOTE, 2020  
SNLI, 2175  
social graphs, 1873  
social media, 1904, 1905  
social network analysis (SNA), 1873, 1886, 1905  
social networks, 1873, 1880, 1884, 1896, 1904, 1905

softmax, 2162, 2166, 2176  
sorting, 1984  
spam, 2065  
sparklyr, 1980, 1987, 1988, 1994  
sparse graphs, 1883, 1884  
sparsity, 2049, 2080, 2153  
spatial configuration (habitats), 1908  
species extinction, 1908  
spectral clustering, 1903  
spectral graph theory, 1908  
speech recognition, 2064  
speech transcripts, 1903  
speed (transport), 1912  
sports analytics, 1873, 1905  
spurious correlations, 1967  
stability (ecosystems), 1907  
standardization, 2005, 2007  
state space, 2243  
static networks, 1885  
stationarity, 1965  
stationary distribution, 2244  
statistical NLP, 2111  
statistical physics, 1874  
statistics, 1874  
stemming, 2133, 2242  
stochastic gradient descent (SGD), 2050, 2053, 2128  
stochastic grammar, 2133  
stochastic processes on networks, 1873  
stopwords, 2134, 2148, 2173, 2233, 2240, 2250  
storage, 1975  
storage layout, 1963, 1988  
strategic advantage (networks), 1893  
streaming, 2009  
streaming k-means, 2009  
stride (convolution), 2055  
Strogatz, Steven H., 1878  
structural holes, 1893, 1894  
structure, 2132  
style transfer, 2054  
subgraph, 1914  
subway systems, 1910  
subword features, 2163, 2168, 2173  
sufficient statistics, 1972  
sunspots, 2100  
support vector machines

(SVM), 1902  
kernel SVM, 1903  
surveillance, 1895  
surveillance systems, 1876  
sustainability, 1911  
swapping, 1964, 1968  
symbolic AI, 2029, 2110  
synonym, 2134  
synsets, 1903  
syntactic ambiguity, 2112  
syntactic chunk, 2133  
syntax, 2132  
synthetic data, 2005, 2022  
systemic issues, 1894  
systems  
dynamic, 1965  
systems biology, 1901  
systems thinking, 1877  
São Paulo, 1912  
T5, 2064, 2083  
tagging (social media), 1907  
talent migration, 1904  
target, 1965, 1978  
teacher forcing, 2083  
temperature (sampling), 2074  
temporal distance, 1901  
temporal networks, 1874, 1875, 1885, 1900, 1901  
time-respecting paths, 1901  
tensor, 1976  
tensor contraction, 2036, 2039  
tensor train, 2036  
decomposition, 2036  
TensorFlow, 2058  
tensors, 2030, 2035, 2036  
order, 2035, 2036  
rank, 2035  
text classification, 2111  
text generation, 2064, 2243  
text statistics, 2235  
TextRank, 2254  
textual entailment, 2175  
tf-idf, 2254  
throughput, 1968, 1974, 1977, 1978  
time series  
forecasting, 2100  
time series data, 1962, 1983  
time-series forecasting, 2064  
token, 2232, 2233, 2239

tokenization, 2178, 2239  
topic model, 2175, 2185, 2191  
topic modeling, 2250  
topology (networks), 1910  
toxicity prediction, 1901  
train-test split, 2020  
training  
epochs, 2085  
loss, 2085  
machine learning, 2029  
validation accuracy, 2088  
transactions (networks), 1884  
transformations, 2004  
transformers, 2064, 2082, 2111, 2177  
cross-attention, 2178  
layer normalization, 2082, 2177  
multi-head attention, 2082, 2178  
scaled dot-product, 2082  
positional encoding, 2082, 2178  
self-attention, 2082  
transformer block, 2082  
transition matrix, 2244  
transition probability, 2243  
transitivity, 1881, 1886, 1888  
transportation networks, 1873, 1880, 1889, 1905, 1910  
airline networks, 1873, 1876  
public transport, 1873  
travel time, 1910  
tree-based methods, 2016  
trends, 1904  
triadic census, 1934  
triadic closure, 1886, 1888  
triads, 1875, 1886  
closed triads, 1886  
open triads, 1886  
triangle counting, 1904  
TT core, 2036  
TT decomposition, 2036  
TT rank, 2036  
Tucker decomposition, 2036  
tweet classification, 2214  
type, 2232, 2233  
UCI Machine Learning

Repository, 2020  
uncertainty quantification,  
1968  
universal approximation,  
2114  
unrolling (RNN), 2064  
unsupervised learning,  
2030  
urban planning, 1911  
  
validation data, 2049  
validity, 1955, 1964  
external, 1966  
value, 1963, 1983  
value matrix, 2082  
vanishing gradients, 2064,  
2067  
variability, 1964, 1966, 1993  
variational autoencoders  
(VAE), 2079  
variational inference, 2198  
variety, 1963  
vector representations, 2163  
vector-to-sequence models,  
2064, 2065  
vectorization, 1976  
velocity, 1963  
veracity, 1963, 1964, 1966,

1983  
vertex, 1873, 1879  
vertical scaling, 1975  
virtual machine (VM), 1975  
virtualization, 1975  
visibility (networks), 1893  
Vision Transformers (ViT),  
2083  
visualization, 1964  
volume, 1963  
vulnerability (networks),  
1894  
  
Watts, Duncan J., 1878  
web graph, 1919  
web pages, 1919  
weight decay, 2048, 2049,  
2070  
weighted graphs, 1910  
weights, 1880, 1883, 1884, 1910,  
1914  
Weisfeiler–Lehman kernel,  
1902  
WhatsApp, 1907  
Whitney inequalities, 1900  
window size, 2168, 2169  
within-cluster variation,  
2004

word analogies, 2173  
word embedding, 2176  
word embedding algorithms,  
2165  
word embeddings, 2111, 2163,  
2207  
word error rate, 2161  
word sense disambiguation,  
1903, 2134  
word tokenizer, 2239  
word–context matrix, 2165  
Word2Vec, 2167  
WordNet, 1875, 1903, 2242  
workflow, 1955, 1962, 1979,  
1980, 1982, 1983,  
1987, 1988  
World Health Organization  
(WHO), 1876  
  
X (Twitter), 1905  
XOR function, 2084  
linear non-separability,  
2084  
  
YARN, 1994  
Yogi, 2053  
  
Zika, 1877