**Github Applied-ML**

The Github “[applied-ml](https://github.com/eugeneyan/applied-ml)” is a repository with links to 500+ papers from 100+ companies.

This repository is described as “***Papers & tech blogs by companies sharing their work on data science & machine learning in production.***”

The papers are divided into 30 different categories that represent the state of AI at this time.

Over 200 (37%) of the papers were submitted by just 7 corporations: LinkedIn, Google, DoorDash, Alibaba, Uber, Facebook and Netflix.

***Recommendation (Systems) 74 entries Application***

Recommender systems are a type of machine learning based systems that are used to predict the ratings or preferences of items for a given user. There are three main types of Recommender Systems: collaborative filtering, content-based, and hybrid.

***Search & Ranking 58 entries Application***

AI ranking refers to a variety of machine learning algorithms used to optimize the order of search results. With any search query, there could be many relevant results. That's where search result ranking comes into play.

***Validation and A/B Testing 47 entries Application***

A/B Testing, also known as split testing, refers to a randomized experimentation process wherein two or more versions of a variable (web page, page element, etc.) are shown to different segments of website visitors at the same time to determine which version leaves the maximum impact and drives business metrics.

***Natural Language Processing 32 entries Process***

Natural language processing (NLP) is a branch of artificial intelligence within computer science that focuses on helping computers to understand the way that humans write and speak.

***Computer Vision 25 entries Process***

Computer vision is a field of artificial intelligence (AI) that enables computers and systems to derive meaningful information from digital images, videos and other visual inputs — and take actions or make recommendations based on that information.

***ML OPs Platforms 23 entries Process***

An MLOps platform provides data scientists and software engineers with a collaborative environment that facilitates iterative data exploration, real-time co-working capabilities for experiment tracking, feature engineering, and model management, as well as controlled model transitioning, deployment, and monitoring.

***Feature Stores 20 entries Process***

A feature store is an emerging, ML-specific data system used to centralize storage, processing, and access to frequently used features, making them available for reuse in the development of future machine learning models.

***Data Discovery 17 entries Process***

Data discovery is the business-user-oriented data science process of visually navigating data and applying advanced analytics in order to detect patterns, gain insight, answer highly specific business questions, and derive value from business data. See this link: <https://www.heavy.ai/technical-glossary/data-discovery>

***Classification 17 Entries Process/Application***

The first job for many artificial intelligence (AI) algorithms is to examine the data and find the best classification. An autonomous car, for example, may take an image of a street sign; the classification algorithm must interpret the street sign by reading any words and comparing it to a list of known shapes and sizes. A phone must listen to a sound and determine whether it is one of its wake-up commands (“Alexa,” “Siri,” “Hey Google”).

***Practices 17 Entries Process***

This might be a critical key to the whole course. Very important link: <https://developers.google.com/machine-learning/guides/rules-of-ml>

***Anomaly Detection 17 Entries Application***

Anomaly detection is a technique that uses AI to identify abnormal behavior as compared to an established pattern. Anything that deviates from an established baseline pattern is considered an anomaly.

***Data Engineering 13 Entries Process***

Managing data. Usually large, changing datasets.

***Graph 13 Entries Process***

Graph AI is the science of using Machine Learning on graphs to focus on the relationships between variables to achieve deeper insights. By using specific algorithms like clustering, partitioning, PageRank and shortest path, some problems. become easier to solve.

***Forecasting 12 Entries Application***

AI planning and forecasting is a field of artificial intelligence used to make scientific predictions about the future without requiring oversight. AI planning tools use time series data to estimate future developments for many industries, such as sales, healthcare, financial services, and manufacturing.

***Embeddings 11 Entries Process***

An embedding is a relatively low-dimensional space into which you can translate high-dimensional vectors. Embeddings make it easier to do machine learning on large inputs like sparse vectors representing words.

***Sequence Modeling 10 Entries Process***

Machine learning models that input or output data sequences are known as sequence models. Text streams, audio clips, video clips, time-series data, and other types of sequential data are examples of sequential data. Recurrent Neural Networks (RNNs) are a well-known method in sequence models.

***Reinforcement Learning 9 Entries Process***

Reinforcement learning is a machine learning training method based on rewarding desired behaviors and/or punishing undesired ones. In general, a reinforcement learning agent is able to perceive and interpret its environment, take actions and learn through trial and error.

***Optimization 8 Entries Process***

AI operations and optimization involves the application of Artificial Intelligence (AI) technologies, such as machine learning and advanced analytics. This is done to automate problem-solving and processes in network and IT operations, and to enhance network design and optimization capabilities.

***Team Structure 8 Entries Process***

Details on assembling teams to code AI projects.

***Data Quality 7 Entries Process***

There are five traits that you'll find within data quality:

accuracy, completeness, reliability, relevance, and timeliness.

***Information Extraction 6 Entries Application***

Information extraction is the process of extracting information from unstructured textual sources to enable finding entities as well as classifying and storing them in a database.

***Generation 5 Entries Application***

The ability of AI to generate text output that looks like it was written by a human. ChatGPT is a prime example of Generation.

***Model Management 5 Entries Process***

Model management is a part of MLOps. ML models should be consistent, and meet all business requirements at scale. To make this happen, a logical, easy-to-follow policy for model management is essential. ML model management is responsible for development, training, versioning and deployment of ML models.

<https://neptune.ai/blog/machine-learning-model-management>

***Ethics 5 Entries Process***

AI ethics is a set of guidelines that advise on the design and outcomes of artificial intelligence. Human beings come with all sorts of cognitive biases, such as recency and confirmation bias, and those inherent biases are exhibited in our behaviors and subsequently, our data.

<https://www.ibm.com/topics/ai-ethics>

***Regression 4 Entries Process***

AI Linear Regression is a supervised machine learning algorithm with a continuous and constant slope expected performance. Rather than attempting to classify values into groups (such as cats and dogs), it is used to estimate values within a continuous range (such as price and revenue).

***Weak Supervision 4 Entries Process***

Weak supervision is an approach to machine learning in which high-level and often noisier sources of supervision are used to create much larger training sets much more quickly than could otherwise be produced by manual supervision (i.e. labeling examples manually, one by one).

<https://snorkel.ai/weak-supervision/>

***Efficiency 3 Entries Process***

<https://openai.com/research/ai-and-efficiency>

***Audio 2 Entries Application***

<https://super.ai/products/super-audio>

***Infrastructure 2 Entries Process***

AI infrastructure encompasses almost every stage of the machine learning workflow. It enables data scientist, data engineers, software engineers and DevOps teams to access and manage the computing resources to test, train and deploy AI algorithms.