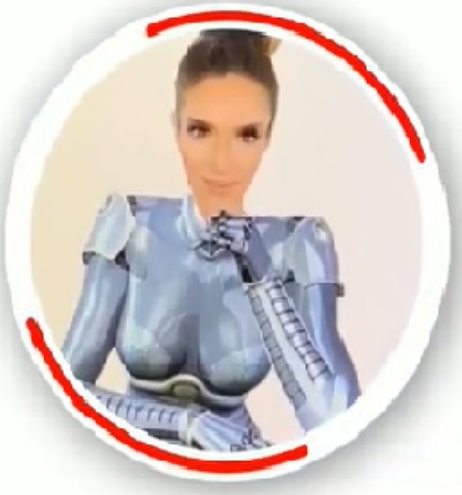
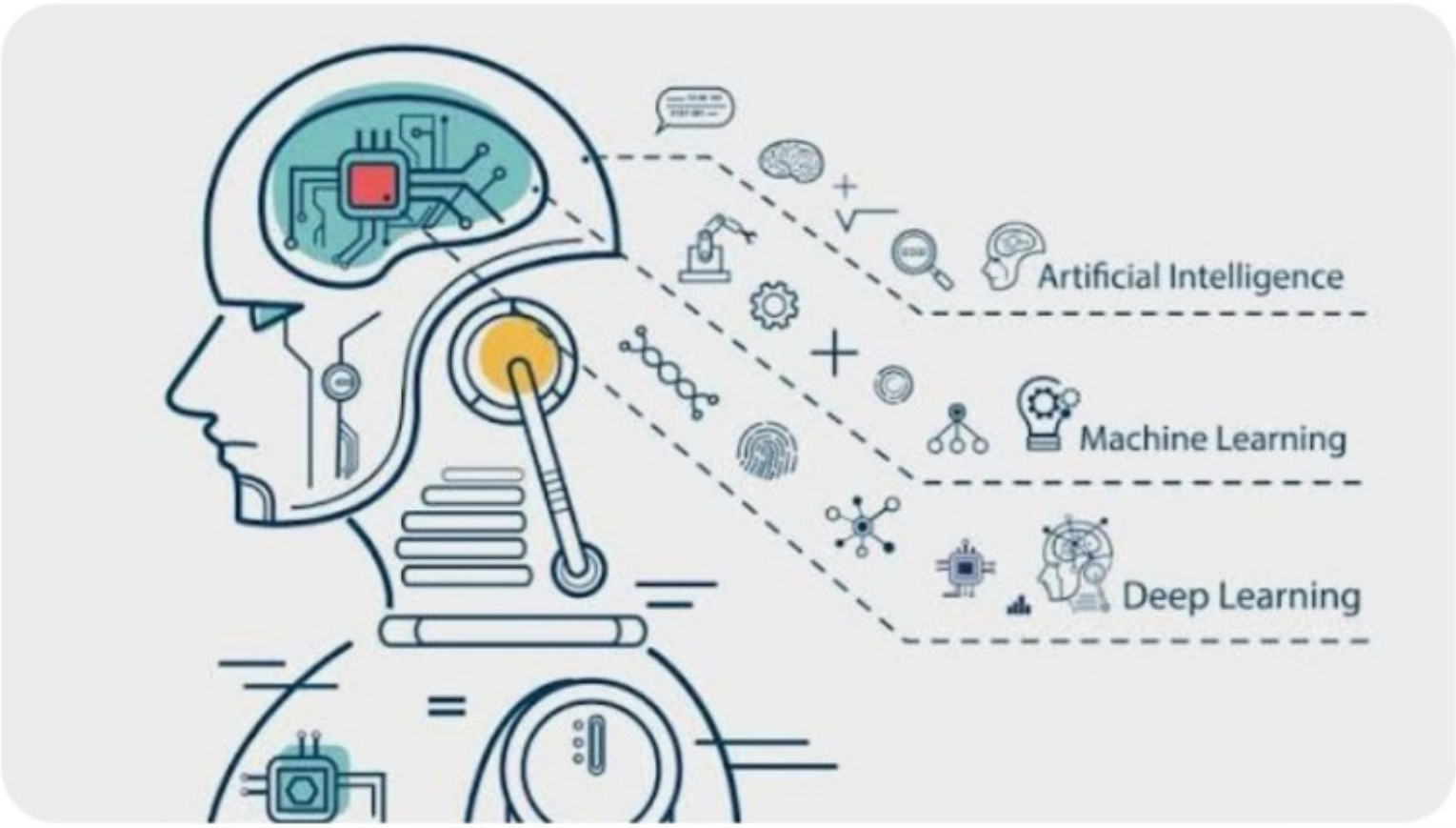
Do You Know
How to Manage and
Increase Performance of
Massive Volumes of Data?



The Democratization of Artificial Intelligence and Deep Learning



HANDBOOK FOR DATA / IT / ECOM TEAMS

By Dr.Keskin
August, 2021

TO UNDERSTAND BETTER WATCH BEFORE STARTING TO READ

Hilary Mason to explain machine learning to 5 different people; a child, a teen, a college student, a grad student, and an expert.

Computer Scientist Explains Machine Learning in 5 Levels of Difficulty | WIRED

WIRED has challenged computer scientist and Hidden Door cofounder and CEO Hilary Mason to explain machine learning to 5

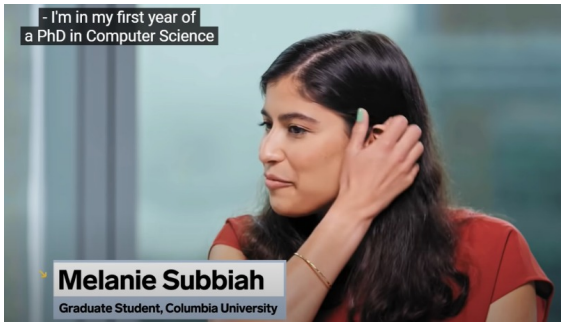
youtu.be

https://youtu.be/5q87K1WaoFI

1.Level For KIDS:
Watch this section with your kid 0:34

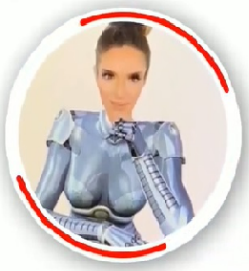
2.Level For TEENS:
Watch this section with your Teen 4:16

3.Level For COLLEGE STUDENTS:
Watch this section with College Students 8:50



4.Leve For Post Graduates:
Watch this section with your Grad Students 14:56

5.Level For C-LEVEL EXECUTIVES :
Watch this section with C-Level Executives 18:37

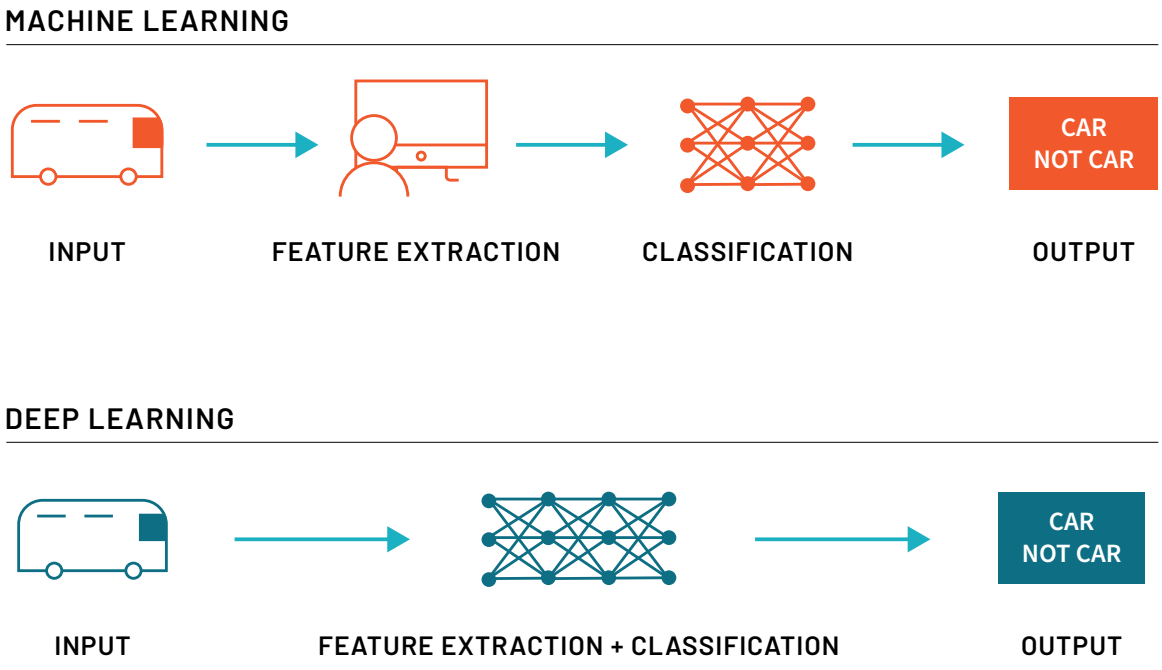


Abstract

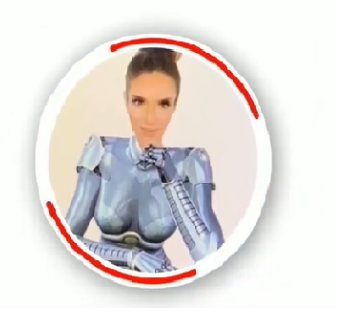
Deep learning is driving rapid innovations in artificial intelligence and influencing massive disruptions across all markets. This paper provides an understanding of the promise of deep learning, the challenges with leveraging this technology, how it is currently solving real-world problems, and more importantly, how deep learning can be made more accessible for data professionals as a whole.

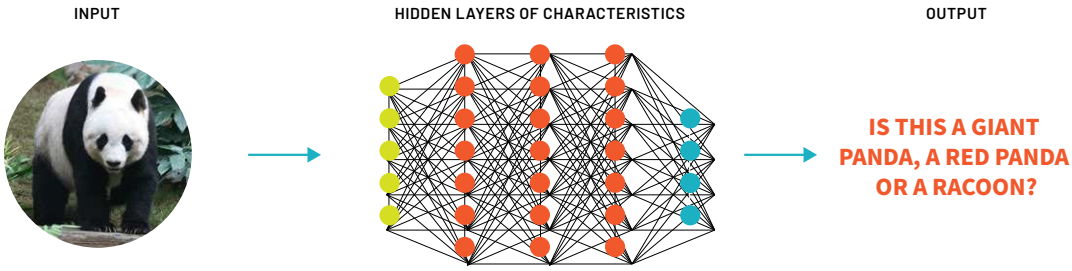
What is Deep Learning?

Deep learning, which is a specialized and advanced form of machine learning, performs what is considered "end-to-end learning". A deep learning algorithm is given massive volumes of data, typically unstructured and disparate, and a task to perform such as classification. The resulting model is then capable of solving complex tasks such as recognizing objects within an image and translating speech in real time.



Deep learning models can be trained to perform complicated tasks such as image or speech recognition and determine meaning from these inputs. A key advantage is that these models scale well with data and their performance will improve as the size of Office Depot data increases.





Deep learning not only performs best with larger volumes of data, but also requires powerful hardware such as graphical processing units (GPUs) in order to perform. The deep learning market is expected to be worth $1.75 billion by the year 2022[1]. Investment in this area is driven by the fact that 61% of enterprises with an innovation strategy are applying AI to their data to find previously missed opportunities such as process improvements or new revenue streams[2].

Innovative Deep Learning Uses

Deep learning has enabled innovation and transformation across a broad range of industries. From anomaly detection to video analysis, businesses have been able to leverage artificial intelligence to gain competitive advantage and even change the way their markets approach the customer experience.

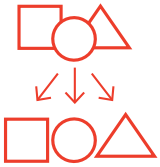
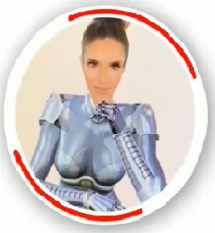


IMAGE CLASSIFICATION

This is the process of an AI application identifying and detecting an object or a feature in a digital image or video. Image classification has taken off in the retail vertical which is using deep learning models to quickly scan and analyze in-store imagery to intuitively determine inventory movement.
This has lead[3] to streamlined operations, reduced costs, and new sales opportunities.

VOICE RECOGNITION

This is the ability of a deep learning model to receive and interpret dictation or to understand and carry out spoken commands. Models are able to convert captured voice commands to text and then use natural language processing to understand what is being said and in what context. This has delivered massive benefits to industries like automotive which uses voice commands to enable drivers to make phone calls and adjust internal controls – all without taking their hands off the steering wheel, thereby improving safety.



1.https://www.marketsandmarkets.com/Market-Reports/deep-learning-market-107369271.html
2.https://www.forbes.com/sites/gilpress/2016/07/20/artificial-intelligence-rapidly-adopted-by-enterprises-survey-says/?sh=a5e060312dae
3.https://www.bandt.com.au/three-ways-image-recognition-technology-will-change-retail-marketing-2017/

ANOMALY DETECTION

This deep learning technique strives to recognize abnormal patterns which don't match the behaviors expected for a particular system, out of millions of different transactions. These applications can lead to the discovery of an attack on financial networks, fraud detection in insurance filings or credit card purchases, even isolating sensor data in industrial facilities signifying a safety issue[4].

RECOMMENDATION ENGINES FOR IT, DATA AND ECOM TEAMS

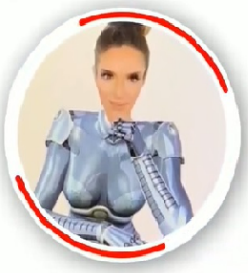
These models analyze user actions in order to provide recommendations based on user behavior. Recommendation engines are critical components of e-commerce sites such as Overstock.com which uses a recommendation engine to provide accurate suggestions of products to users for future purchases based on their shopping history. This massively reduces the friction for the user and provides efficient revenue streams for the company.

SENTIMENT ANALYSIS

This leverages deep learning-heavy techniques such as natural language processing, text analysis, and computational linguistics to gain clear insight into customer opinion, understanding of consumer sentiment, and measuring the impact of marketing strategies5. A real-world application of this particular type of AI is deployed by Riot Games to better understand and combat abusive language that can occur during in-game experiences in order to improve the user experience.

VIDEO ANALYSIS

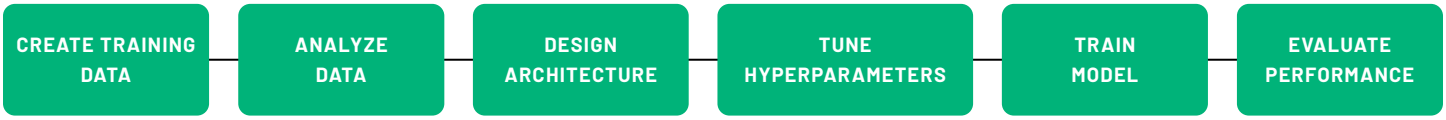
Deep learning models have made it possible to process and evaluate vast streams of video footage for a range of tasks including threat detection, which can be used in airport security, banks, and sporting events. Media companies like Viacom leverage video analysis to ensure that lag is eliminated and the user experience is maximized.



4.https://digitalcommons.unf.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1783&context=etd
5.https://www.oreilly.com/library/view/artificial-i

A Deep Learning Workflow

A generalized workflow for building and training a deep learning model consists of steps that vary in complexity. This spans the ingestion of data, through network architecture choice, to production.[6]



CREATE YOUR TRAINING DATA SET - This can include a wide variety of data types from a wide variety of sources needed to train a model, which may include additional effort to obtain labels or target variable values.

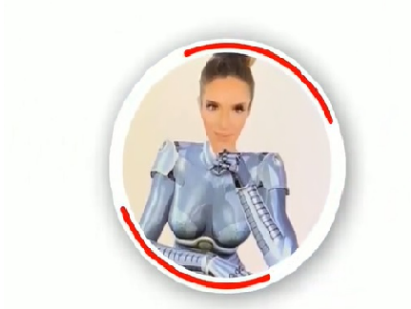
ANALYZE THE DATA - It is critical to clean and organize the data in order to eliminate errors and discrepancies.

DESIGN YOUR ARCHITECTURE - The key is to understand the type of problem you are trying to solve and then choosing the right architecture for the job.

TUNE YOUR HYPERPARAMETERS - Getting the best results in deep learning requires experimenting with different values for training hyperparameters.

TRAIN THE MODEL - In this step, you provide the data to the learning algorithm which performs optimization routines to produce the best model it can.

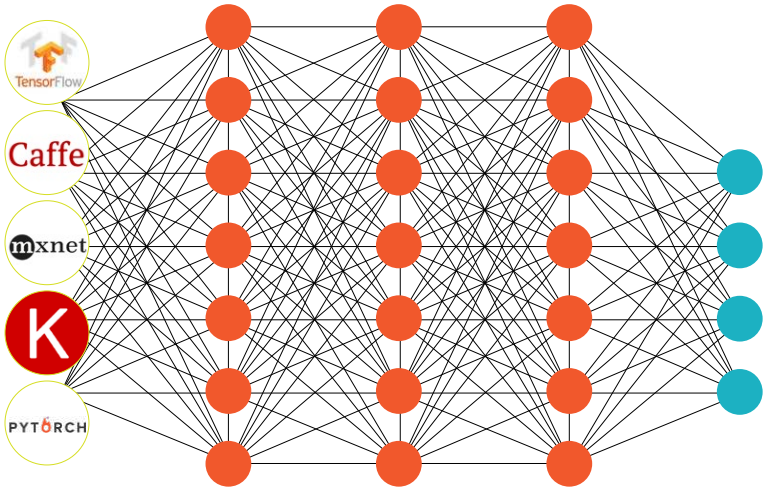
EVALUATE PERFORMANCE - This validates the ability of the model to confidently perform forecasting and estimation — the actual "thinking" of the model — against an unseen ("test") dataset.



6.https://databricks.com/blog/2017/06/06/databricks-vision-simplify-large-scale-deep-learning.html

What is Powering Deep Learning?

If you were to start doing deep learning, which framework would you use? That question can be answered by understanding the problem you are trying to solve. Most frameworks support different programming languages, offer varying levels of architectural complexity, different degrees of performance, and a variety of deep learning algorithms suitable for specific use cases. Here's an overview of some of the common deep learning frameworks available.



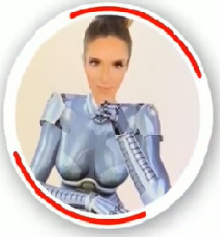
TENSORFLOW – is a powerful, open-source software library for numerical computation using data flow graphs.

CAFFE – is an open-source deep learning framework designed with expression, speed, and modularity in mind. It is often used for operationalizing models for prediction in production.

MXNET – is a deep learning framework designed for both efficiency and flexibility. It allows you to mix symbolic and imperative programming to maximize efficiency and productivity.

KERAS – is an open source neural network library written in Python designed to enable fast experimentation with deep neural networks, and focuses on being minimal, modular and extensible.

PYTORCH – is a Python based computing package that provides a wide range of algorithms for deep machine learning.



It needs to be noted that even with all of the available frameworks and greater understanding of how to pursue AI through machine learning and deep learning, there are still substantial limitations to current abilities of artificial intelligence. Trying to compare the capabilities of a deep learning application against the capabilities of an actual human is going to be disappointing to say the least – it's currently not an apples to apples comparison. A reason for this is that while a person may easily recognize the nuance associated with knowing for example how to clearly distinguish the differences between images of a toddler and an infant, a deep learning program will find this incredibly difficult to do with consistent accuracy. The performance of any artificial intelligence will only be as good as the data that it is being fed, and if the data itself is either incorrect or incomplete, the performance will be simply wrong. Processing a lot of data is easy, but feature learning is very difficult[7].

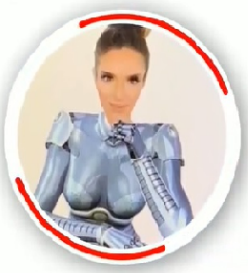
But, as deep learning algorithms evolve to recognize nuances and overcome their current limitations, the future of AI is bright. What's most critical is for big data analytics platforms to handle the growing complexities of these algorithms as well as the scale of the data that will be required for better model training.

The Challenges of Achieving Deep Learning

Leveraging the promise of deep learning today is as challenging as big data was yesterday. Deep learning frameworks have unique capabilities and steep learning curves. Scaling out over distributed hardware requires specialization and significant manual work; and even with the combination of time and resources, achieving success requires tedious fiddling and experimenting with parameters.

Additionally, each of these frameworks provide different strengths for different deep learning approaches and their own unique challenges. For example, Caffe is a strong option for image classification but can be resource intensive and the framework itself is difficult to deploy[8]. Things can get very hard, and very complicated, very quickly for Data Science and Engineering teams.

Also, the effectiveness of deep learning, or artificial intelligence of any type, rests on the quality of the infrastructure that powers it. The infrastructure should be viewed as a multiplier of the effectiveness of your AI. Neural networks come in various architectures, the performance of which is a function of the architecture, which can be challenging for traditional engineering teams to manage, let alone a data science team. Also, the processing requirements of an AI infrastructure can be massive, requiring specialized – and expensive – processors such as GPU's (graphical processing units) in order to perform the mathematical computations that power deep learning models[9].



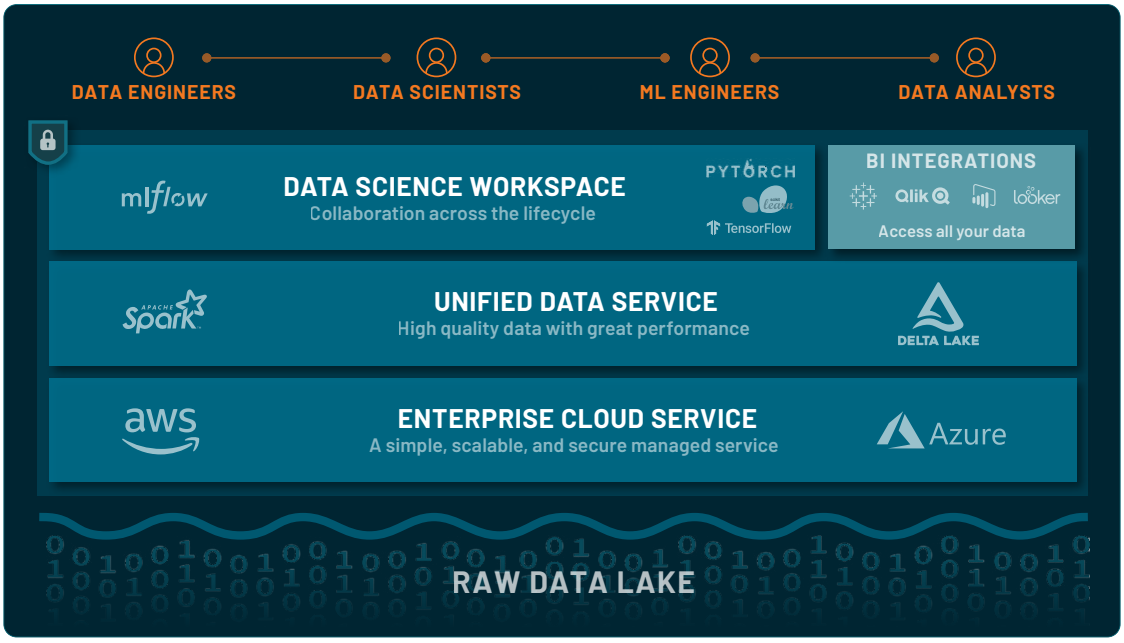
7.https://www.oreilly.com/data/free/files/what-is-artificial-intelligence.pdf
8.https://www.infoworld.com/article/3163525/the-best-machine-learning-and-deep-learning-libraries.html
9.https://www.forbes.com/sites/janakirammsv/2017/08/07/in-the-era-of-artificial-intelligence-gpus-are-the-new-cpus/?sh=3e64fd205d16

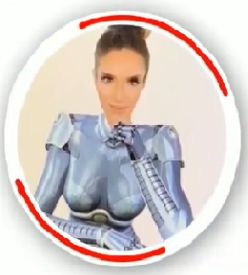
From a resource standpoint, training an accurate deep learning model can be extremely taxing. Parameters need to be tweaked to create a great model, and this step can be manually intensive. This puts a lot of pressure on a data science team as the number of decisions required to develop successful deep learning models can be incredibly time-consuming. Significant time and money can be wasted if poor decisions are made[10]. Also, since deep learning models are complex, it requires a significant amount of data to accurately train a model.

Make Deep Learning Easier and Accessible

While much of what has been described may make deep learning sound impossible to achieve, there is technology that offers to overcome these challenges. In 2010, the Spark project was launched as a way to democratize the power of big data. This was accomplished by offering high-level APIs and a unified engine to do machine learning, ETL, streaming and interactive SQL. The outcome of this project is that today, Apache Delta Lake, and MLflow, is helping make deep learning and AI accessible to everyone.

There are platforms unified data analytics platform that allows organizations to build reliable, fast, and scalable deep learning pipelines on Apache Spark that enable data science teams to build, train, and deploy deep learning applications with ease.





10.https://schibsted.com/blog/deep-learning-changing-data-science-paradigms/

- Unified Infrastructure: Platforms offer a fully managed, serverless, cloud infrastructure that simplifies operations, delivers elasticity, and enables greater cost control. Data engineers will benefit from unified workflows that simplify data preparation and ETL11 and an API that makes it easy to work with major frameworks such as TensorFlow, Keras, PyTorch, MXNet, Caffe, CNTK.

- Unified Workflows: A single platform for end-to-end workflow automation from data preparation to exploration, modeling, and large-scale prediction. Platforms also offer high-level APIs to leverage TensorFlow models more easily.

- Performance Optimized: Paltforms offer unparalleled processing performance at scale for deep learning Having the ability to crunch massive volumes of data and build highly.
  Performant data pipelines allows for more accurate deep learning models.

- Model Management: Platforms natively integrate with MLflow - an open source framework for managing the complete Machine Learning lifecycle - so that data scientists can easily track experiments, reproduce results, and deploy models virtually anywhere.

- Collaborative Data Science: For data science teams, a collaborative workspace helps teams work better together to interactively explore data and train deep learning models against real-time data sets with the flexibility of multi-programming language support. These are all critical elements for any organization that is serious about AI.

Through the use of the open source unified data analytics platforms, data teams can benefit from a fully managed, scalable, and secure cloud infrastructure that reduces operational complexity and total cost of ownership.

Conclusion

Artificial intelligence through deep learning will drive innovations in IT for the foreseeable future. The rise of big data has helped to bring this to reality and the benefits are just beginning to be realized. AI is already being explored across many industries, and as the technology improves, will have a major impact on how it can be used to solve real-world problems. The opportunity to bring AI to the mainstream is here, and Databricks has a vision for making deep learning more accessible to a broad range of users

The open source unified analytics platforms simplify the integration of scalable deep learning into organizational workflows, from machine learning practitioners to business analysts, and it helps fulfill the promise of AI.