



(https://www.intuit.com/careers/oa/technology/?cid=dis_aim_clicks_in_ttt-global_aw_round5TechCareerRhyme|alltechaudience_gif|980x90_intuit_talent)



(<https://analyticsindiamag.com/>).



(https://praxis.ac.in/data-science-course-in-bangalore/?utm_source=AIM&utm_medium=banner&utm_campaign=DS_PAT4June22)

PUBLISHED ON AUGUST 28, 2021

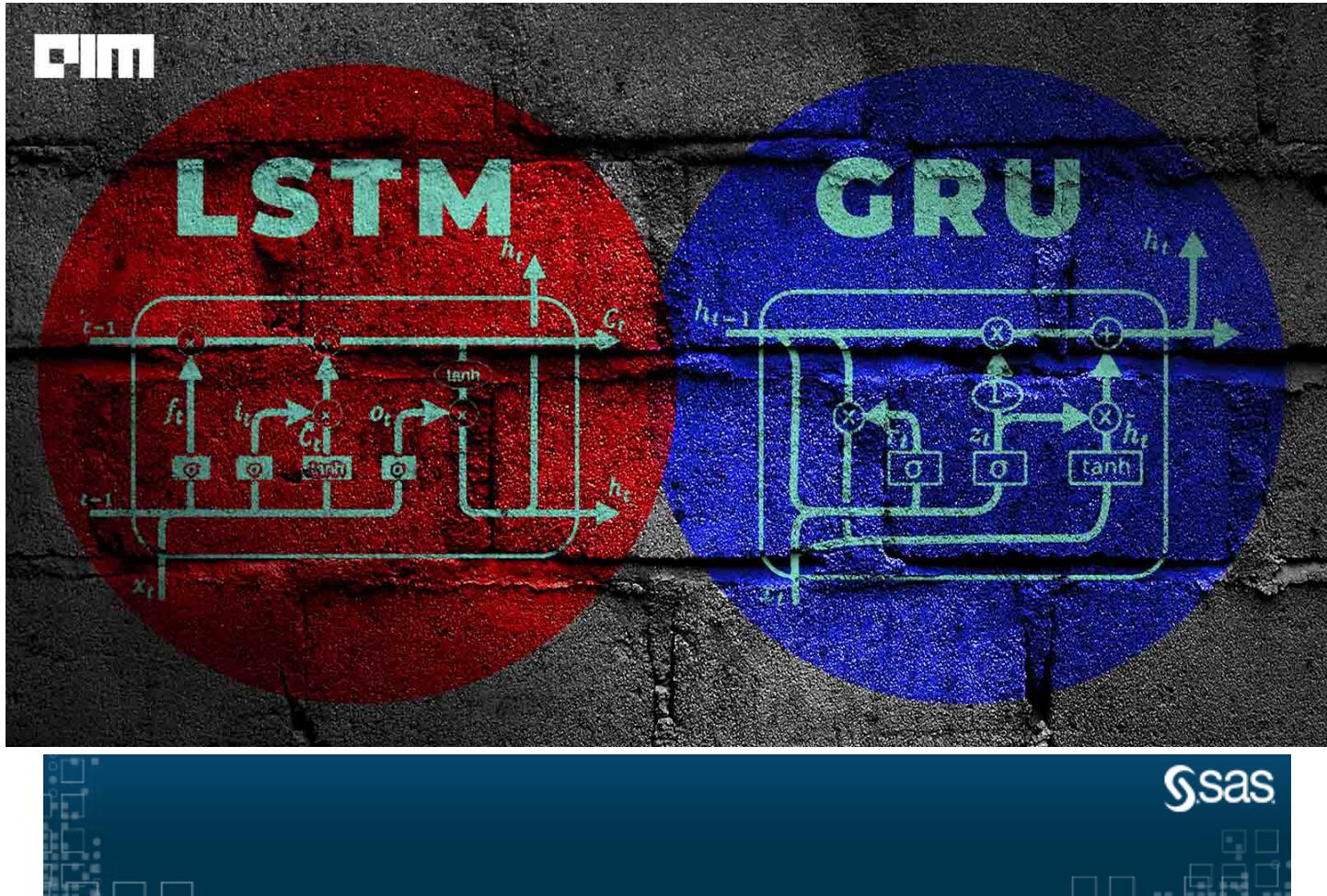
IN DEVELOPERS CORNER

(https://analyticsindiamag.com/category/developers_corner/)

LSTM Vs GRU in Recurrent Neural Network: A Comparative Study

Long Short Term Memory in short LSTM is a special kind of RNN capable of learning long term sequences. They were introduced by Schmidhuber and Hochreiter in 1997. It is explicitly designed to avoid long term dependency problems. Remembering the long sequences for a long period of time is its way of working.

By Vijaysinh Lendave(<https://analyticsindiamag.com/author/vijaysinh-lendaveanalyticsindiamag-com/>)



(https://www.sas.com/gms/redirect.jsp?detail=GMS224019_309942)

A recurrent neural network (<https://analyticsindiamag.com/implementing-a-recurrent-neural-network-rnn-from-scratch/>) is a type of ANN (<https://analyticsindiamag.com/ann-with-linear-regression/>) that is used when users want to perform predictive operations on sequential or time-series (<https://analyticsindiamag.com/guide-to-implementing-time-series-analysis-predicting-bitcoin-price-with-rnn/>) based data. These Deep learning layers are commonly used for ordinal or temporal problems such as Natural Language Processing (<https://analyticsindiamag.com/how-to-identify-entities-in-nlp/>), Neural Machine Translation, automated image captioning tasks (<https://analyticsindiamag.com/hands-on-guide-to-effective-image-captioning-using-attention-mechanism/>) and likewise. Today's modern voice assistance devices such as Google Assistance, Alexa, Siri are incorporated with these layers to fulfil hassle-free experiences for users.

What is the difference between RNN and CNN?

The main difference between the RNN and CNN is that RNN is incorporated with memory to take any information from prior inputs to influence the Current input and output. Training methods for this network are the same. While traditional neural networks assume that both input and output are independent of each other, RNN gives the output based on previous input and its context.

THE BELAMY

Sign up for your weekly dose of what's up in emerging technology.

Enter your email

SIGN UP

Another distinguishing parameter is that RNN shares parameters across each layer of the network. While feedforward networks have different weights across each node, recurrent neural networks share the same weight parameter within each layer of the network.

Below at left one is a representation of standard RNN and the right one is a representation of Feed-Forward Network.

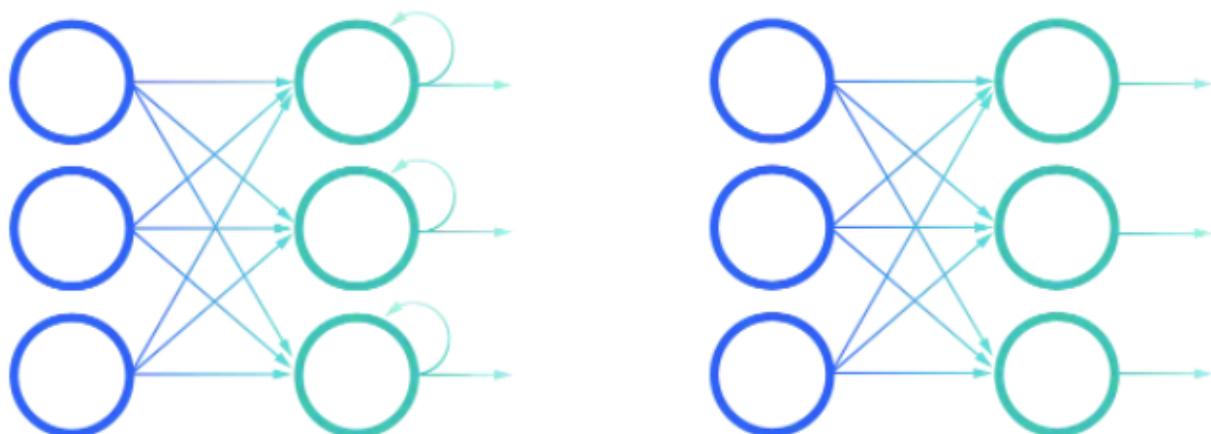
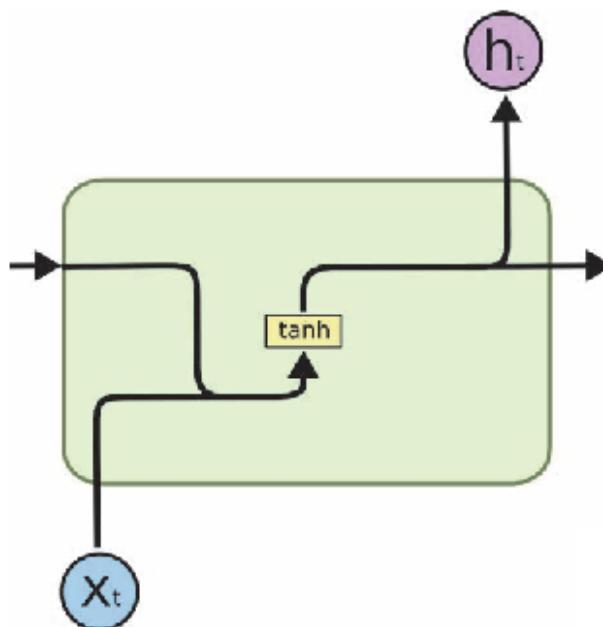


Image Source (<https://www.ibm.com/cloud/learn/recurrent-neural-networks>)

In the last few years for RNN's, there has been an incredible success in a variety of problems such as speech recognition, language modelling, translation, image captioning and list goes on.

Let's take a look at a single unit of RNN architecture. Where it takes input from the previous step and current state X_t and incorporated with Tanh as an activation function, here we can explicitly change the activation function.



Sometimes we only need to look at recent information to perform a present task. But this is not the case we face all the time. When a standard RNN network is exposed to long sequences or phrases it tends to lose the information because it can not store the long sequences and as the methodology is concerned it focuses only on the latest information available at the node. This problem is commonly referred to as Vanishing gradients.

In RNN to train networks, we backpropagate through time and at each time step or loop operation gradient is being calculated and the gradient is used to update the weights in the networks. Now if the effect of the previous sequence on the layer is small then the relative gradient is calculated small. Then if the gradient of the previous layer is smaller then this makes weights to be assigned to the context smaller and this effect is observed when we deal with longer sequences. Due to this network does not learn the effect of earlier inputs and thus causing the short term memory problem.

To overcome this problem specialized versions of RNN are created like LSTM, GRU, Time Distributed layer, ConvLSTM2D layer.

- Gated Recurrent Unit – What Is It And How To Learn
(<https://analyticsindiamag.com/gated-recurrent-unit-what-is-it-and-how-to-learn/>)
- How To Code Your First LSTM Network In Keras
(<https://analyticsindiamag.com/how-to-code-your-first-lstm-network-in-keras/>)
- How To Implement LSTM RNN Network For Sentiment Analysis
(<https://analyticsindiamag.com/how-to-implement-lstm-rnn-network-for-sentiment-analysis/>)
- Salesforce Launches AI-Economist: A Complete Guide With Python Codes
(<https://analyticsindiamag.com/salesforce-launches-ai-economist-a-complete-guide-with-python-codes/>)
- Bill Gates Says No To Sharing Covid Vaccine Recipe With India
(<https://analyticsindiamag.com/bill-gates-says-no-to-sharing-covid-vaccine-recipe-with-india/>)

Working of LSTM

What is Long Short Term Memory or LSTM?

Long Short Term Memory in short LSTM is a special kind of RNN capable of learning long term sequences. They were introduced by Schmidhuber *and* Hochreiter in 1997. It is explicitly designed to avoid long term dependency problems. Remembering the long sequences for a long period of time is its way of working.

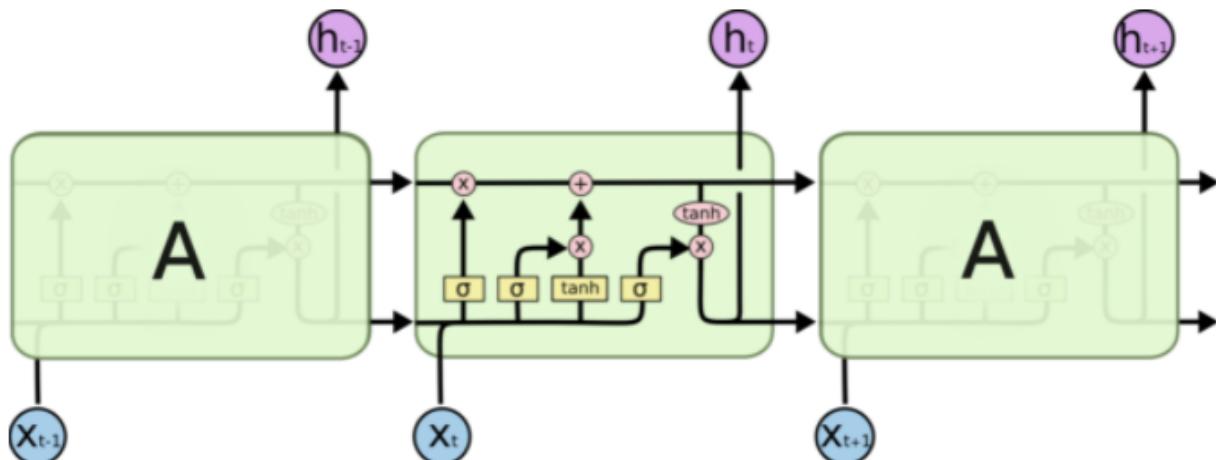


Image Source (<https://colah.github.io/posts/2015-08-Understanding-LSTMs/>)

The popularity of LSTM is due to the Getting mechanism involved with each LSTM cell. In a normal RNN cell, the input at the time stamp and hidden state from the previous time step is passed through the activation layer to obtain a new state. Whereas in LSTM the process is slightly complex, as you can see in the above architecture at each time it takes input from three different states like the current input state, the short term memory from the previous cell and lastly the long term memory.

These cells use the gates to regulate the information to be kept or discarded at loop operation before passing on the long term and short term information to the next cell. We can imagine these gates as Filters that remove unwanted selected and irrelevant information. There are a total of three gates that LSTM uses as *Input Gate*, *Forget Gate*, and *Output Gate*.

Input Gate

The input gate decides what information will be stored in long term memory. It only works with the information from the current input and short term memory from the previous step. At this gate, it filters out the information from variables that are not useful.

Forget Gate

The forget decides which information from long term memory be kept or discarded and this is done by multiplying the incoming long term memory by a forget vector generated by the current input and incoming short memory.

Output Gate

The output gate will take the current input, the previous short term memory and newly computed long term memory to produce new short term memory which will be passed on to the cell in the next time step. The output of the current time step can also be drawn from this hidden state.

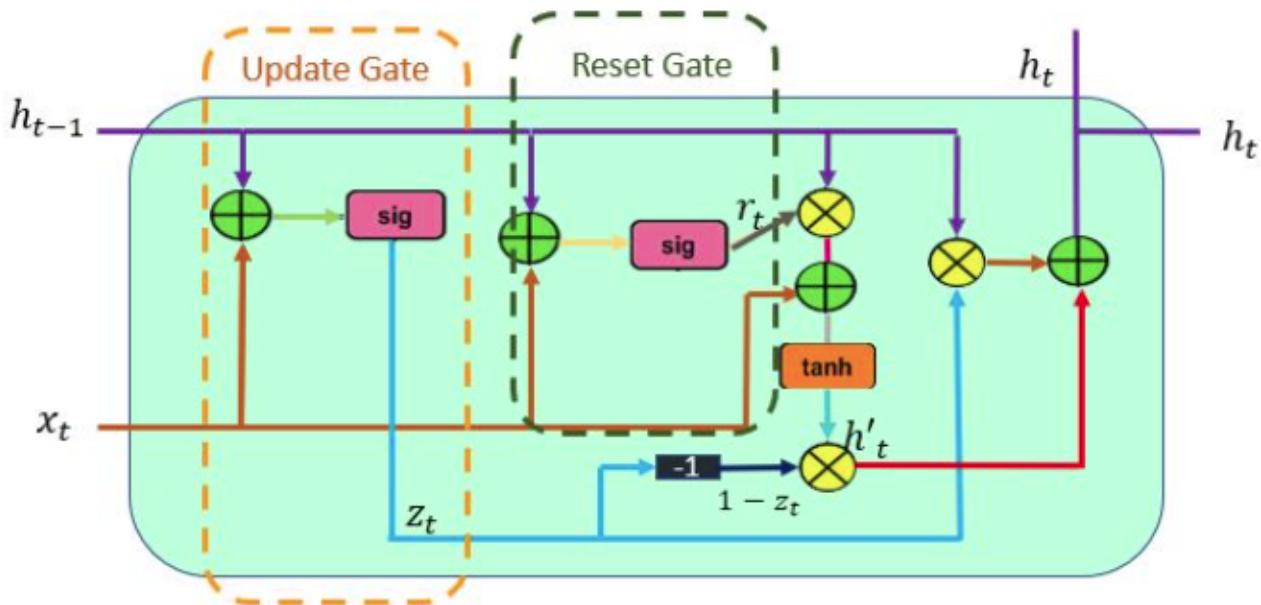
So this is all about the mechanism of LSTM to realise this with practical implementation. [Here](https://colab.research.google.com/drive/1-c36k29ptuMUXUrQzH5SrQ-24t5pFZZL?usp=sharing) (<https://colab.research.google.com/drive/1-c36k29ptuMUXUrQzH5SrQ-24t5pFZZL?usp=sharing>) I have demonstrated the

LSTM use case in which you can check input and output sequences with their shape.

Working of GRU

What is Gated Recurrent Unit or GRU?

The workflow of the Gated Recurrent Unit, in short GRU, is the same as the RNN but the difference is in the operation and gates associated with each GRU unit. To solve the problem faced by standard RNN, GRU incorporates the two gate operating mechanisms called Update gate and Reset gate.



Update gate

The update gate is responsible for determining the amount of previous information that needs to pass along the next state. This is really powerful because the model can decide to copy all the information from the past and eliminate the risk of vanishing gradient.

Reset gate

The reset gate is used from the model to decide how much of the past information is needed to neglect; in short, it decides whether the previous cell state is important or not.

First, the reset gate comes into action it stores relevant information from the past time step into new memory content. Then it multiplies the input vector and hidden state with their weights. Next, it calculates element-wise multiplication between the reset gate and previously hidden state multiple. After summing up the above steps the non-linear activation function is applied and the next sequence is generated.

This is all about the operation of GRU, the practical examples are included in the notebooks.

LSTM Vs GRU

Now we have seen the operation of both the layers to combat the problem of vanishing gradient. So you might wonder which one is to use? As GRU is relatively approaching its tradeoffs haven't been discussed yet.

According to empirical evaluation, there is not a clear winner. The basic idea of using a getting mechanism to learn long term dependencies is the same as in LSTM.

What is the difference between GRU & LSTM?

The few differencing points are as follows:

The GRU has two gates, LSTM has three gates

GRU does not possess any internal memory, they don't have an output gate that is present in LSTM

In LSTM the input gate and target gate are coupled by an update gate and in GRU reset gate is applied directly to the previous hidden state. In LSTM the responsibility of reset gate is taken by the two gates i.e., input and target.

Conclusion

Through this article, we have understood the basic difference between the RNN, LSTM and GRU units. From working of both layers i.e., LSTM and GRU, GRU uses less training parameter and therefore uses less memory and executes faster than LSTM whereas LSTM is more accurate on a larger dataset. One can choose LSTM if you are dealing with large sequences and accuracy is concerned, GRU is used when you have less memory consumption and want faster results.

More Great AIM Stories

[Top Cybersecurity Budgets Around The World](https://analyticsindiamag.com/top-cybersecurity-budgets-around-the-world/) (<https://analyticsindiamag.com/top-cybersecurity-budgets-around-the-world/>)

[How Did Airbnb Achieve Model Metric Consistency At Scale](https://analyticsindiamag.com/how-did-airbnb-achieve-model-metric-consistency-at-scale/)
(<https://analyticsindiamag.com/how-did-airbnb-achieve-model-metric-consistency-at-scale/>)

[Andrew Ng To Kickstart A New Generation Of AI](https://analyticsindiamag.com/andrew-ng-to-kickstart-a-new-generation-of-ai/) (<https://analyticsindiamag.com/andrew-ng-to-kickstart-a-new-generation-of-ai/>)

[What Are DQN Reinforcement Learning Models](https://analyticsindiamag.com/what-are-dqn-reinforcement-learning-models/) (<https://analyticsindiamag.com/what-are-dqn-reinforcement-learning-models/>)

[Tech Behind Locobuzz's COVID Relief Platform](https://analyticsindiamag.com/tech-behind-locobuzzs-covid-relief-platform/) (<https://analyticsindiamag.com/tech-behind-locobuzzs-covid-relief-platform/>)

[Virtual Round Table Discussion On Analytics As A Driver Of Talent Transformation – 30th June](https://analyticsindiamag.com/register-for-a-virtual-round-table-discussion-on-analytics-as-a-driver-of-talent-transformation/) (<https://analyticsindiamag.com/register-for-a-virtual-round-table-discussion-on-analytics-as-a-driver-of-talent-transformation/>)



[\(https://analyticsindiamag.com/author/vijaysinh-lendaveanalyticsindiamag-com/\)](https://analyticsindiamag.com/author/vijaysinh-lendaveanalyticsindiamag-com/)

Vijaysinh is an enthusiast in machine learning and deep learning. He is skilled in ML algorithms, data manipulation, handling and visualization, model building.



(https://www.spjain.org/master-of-artificial-intelligence-in-business?utm_campaign=AIM-Banner-june22&utm_medium=Banner&utm_source=AIM).



(<https://business.louisville.edu/learnmore/>

[utm_campaign=MSBA-INDIA&utm_source=analyticsindia&utm_medium=display&utm_keyword=analyticsindia&](https://business.louisville.edu/learnmore/)

Our Upcoming Events

Masterclass, Virtual

AI Application in Semiconductor Design

16th Jun

Register
[\(https://attendee.gotowebinar.com/register/824430734203312400\)](https://attendee.gotowebinar.com/register/824430734203312400)

Conference, in-person (Bangalore)

MachineCon 2022

24th Jun

Register
[\(https://machinecon.analyticsindiamag.com/tickets/\)](https://machinecon.analyticsindiamag.com/tickets/)

Workshop, Virtual

Advanced SYCL Concepts for Heterogenous Computing

24th Jun

Register
[\(https://attendee.gotowebinar.com/register/8182474697530800652\)](https://attendee.gotowebinar.com/register/8182474697530800652)

Conference, Virtual

Deep Learning DevCon 2022

30th Jul

Register
[\(https://dldc.adasci.org/get-the-tickets/\)](https://dldc.adasci.org/get-the-tickets/)

Conference, in-person (Bangalore)

Cypher 2022

21-23rd Sep

Register
[\(https://www.analyticsindiasummit.com/cypher-2022/register/\)](https://www.analyticsindiasummit.com/cypher-2022/register/)

3 Ways to Join our Community

Discord Server

Stay Connected with a larger ecosystem of data science and ML Professionals

**JOIN DISCORD COMMUNITY
([HTTPS://DISCORD.GG/SBTJ3JDEAZ](https://discord.gg/SBTJ3JDEAZ))**

Telegram Channel

Discover special offers, top stories, upcoming events, and more.

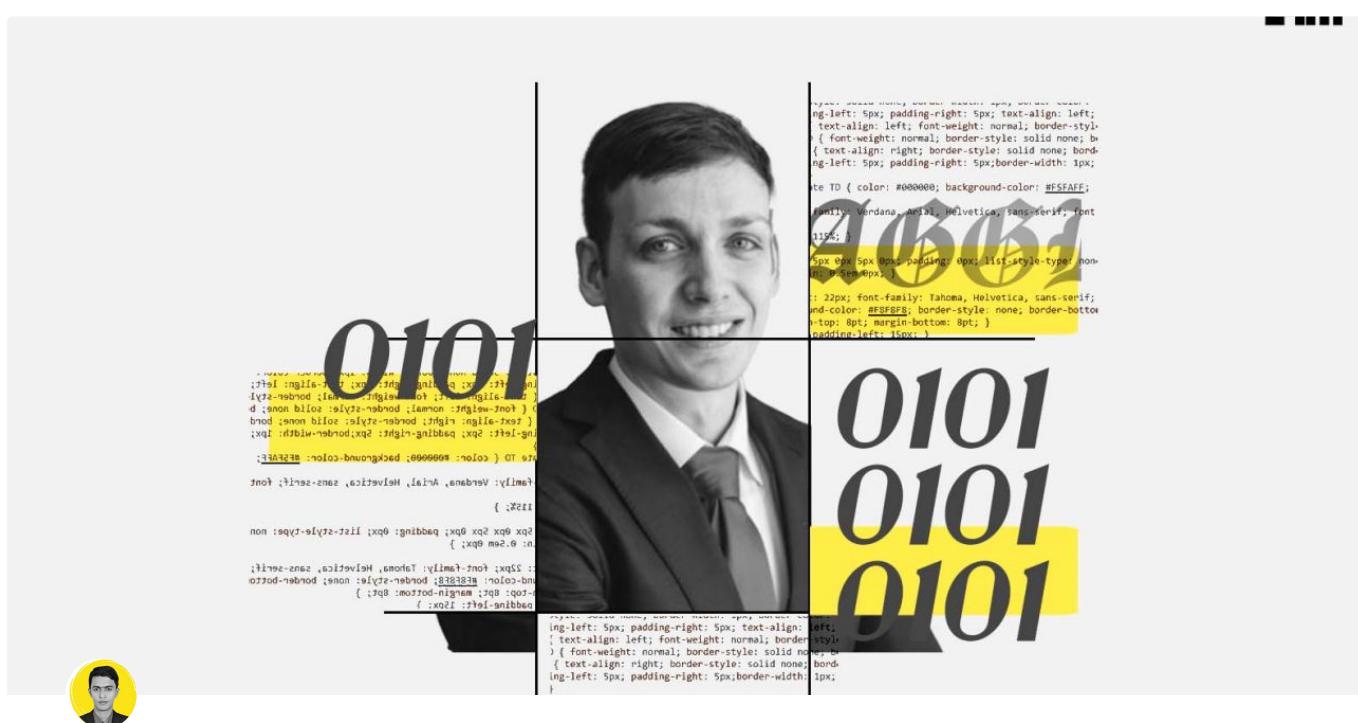
**JOIN TELEGRAM
([HTTPS://T.ME/+TRPAPV7GNN2OZ1AZ](https://t.me/+TRPAPV7GNN2OZ1AZ))**

Subscribe to our newsletter

Get the latest updates from AIM

[SUBSCRIBE](#)

MORE FROM AIM



The story of Kaggle Grandmaster Janio Martinez Bachmann
[\(https://analyticsindiamag.com/the-story-of-kaggle-grandmaster-janio-martinez-bachmann/\)](https://analyticsindiamag.com/the-story-of-kaggle-grandmaster-janio-martinez-bachmann/)

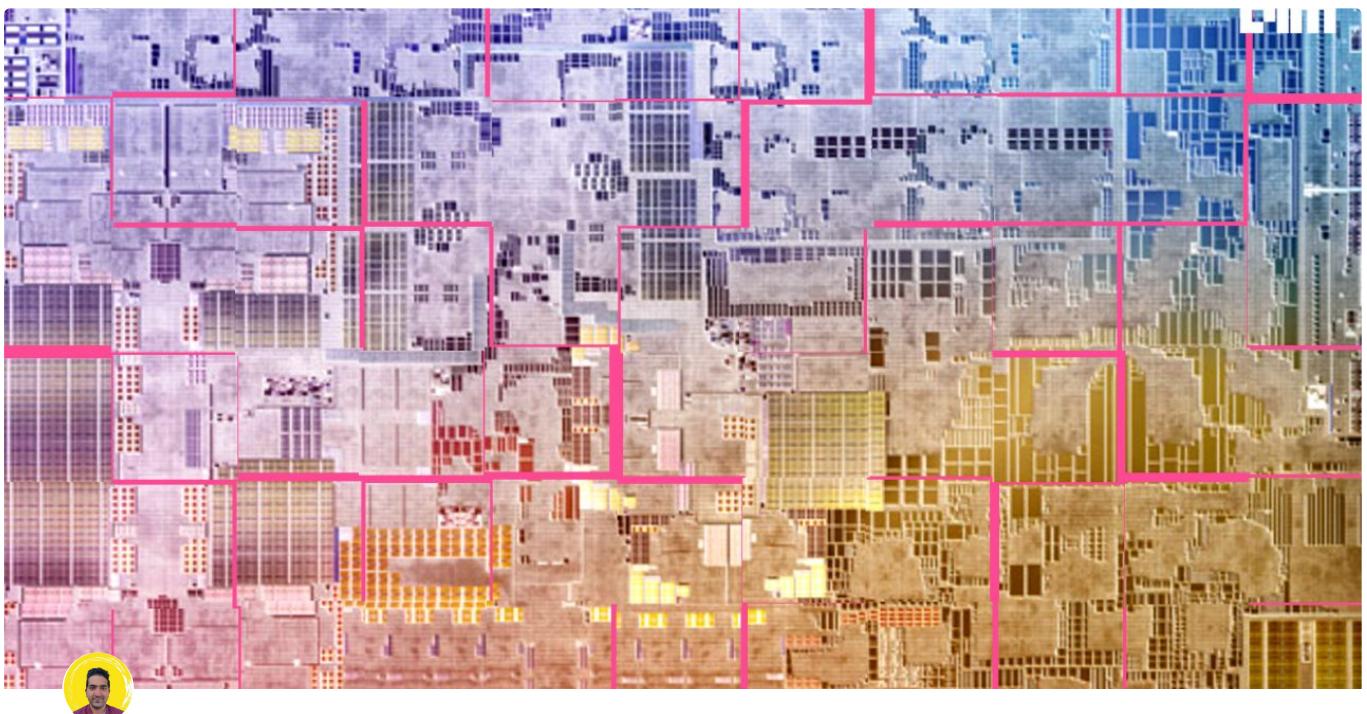
I tend to get nervous when I must code next to a person





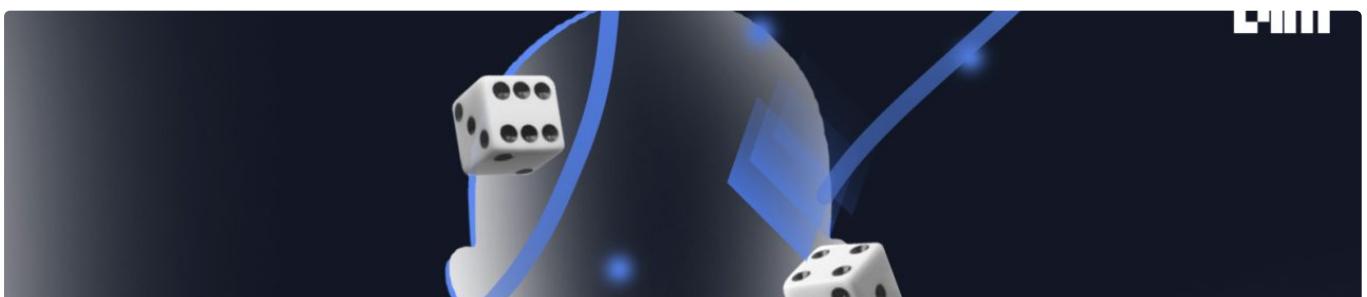
India needs dedicated cybersecurity laws: Kanwaljeet Kaur (<https://analyticsindiamag.com/india-needs-dedicated-cybersecurity-laws-kanwaljeet-kaur/>)

The government needs to inspire enough confidence that user data be protected from potential misuse.



Apple WWDC 2022 highlights: M2 chip with neural engine, search updates, APIs and more! (<https://analyticsindiamag.com/apple-wwdc-2022-highlights-m2-chip-with-neural-engine-search-updates-apis-and-more/>)

M2 incorporates the enhanced second-generation 5-nanometer technology.





How to learn from uncertainty using probabilistic machine learning? (<https://analyticsindiamag.com/how-to-learn-from-uncertainty-using-probabilistic-machine-learning/>)

This article briefs on the various ways a machine learning model learns from the data with uncertainty and why is probabilistic machine learning the best.

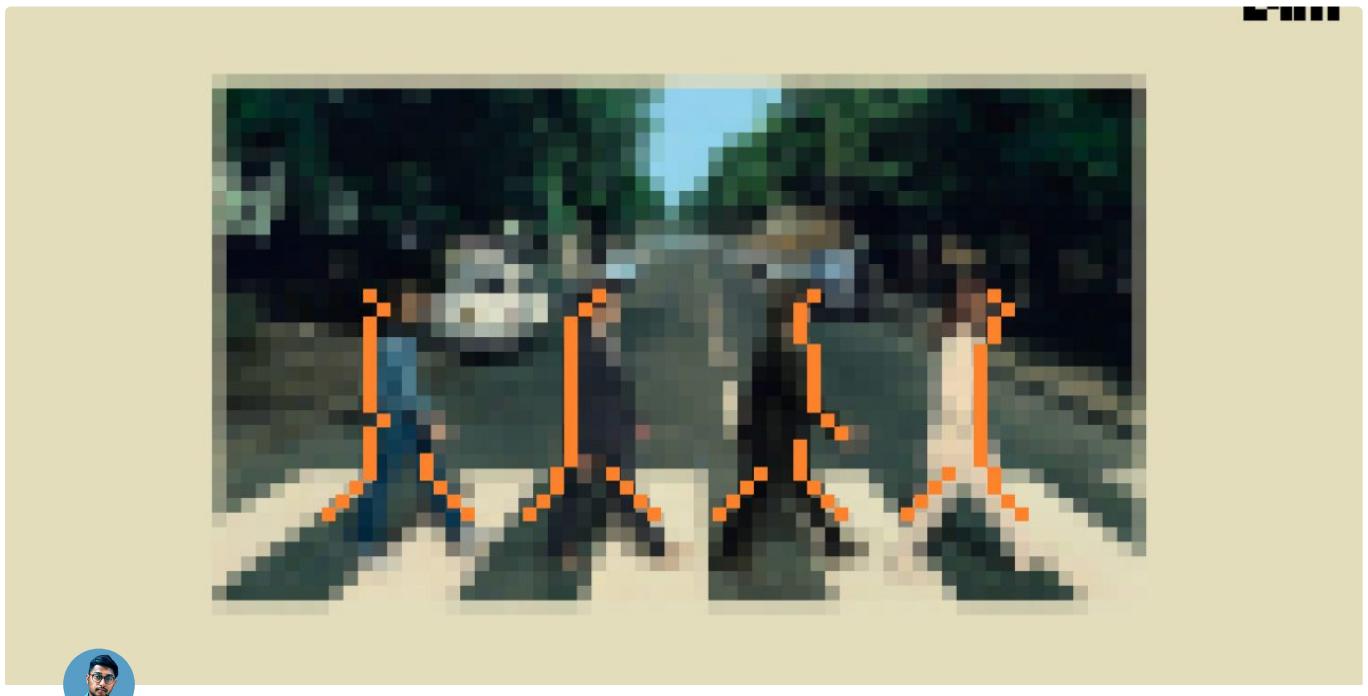
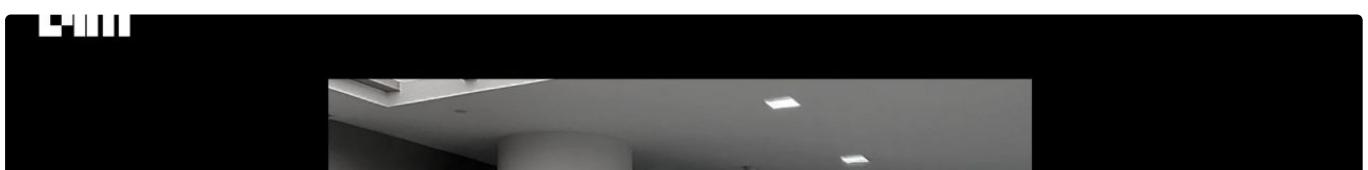


Image annotation techniques with implementation in OpenCV (<https://analyticsindiamag.com/image-annotation-techniques-with-implementation-in-opencv/>)

Image annotation is labelling of objects in an image





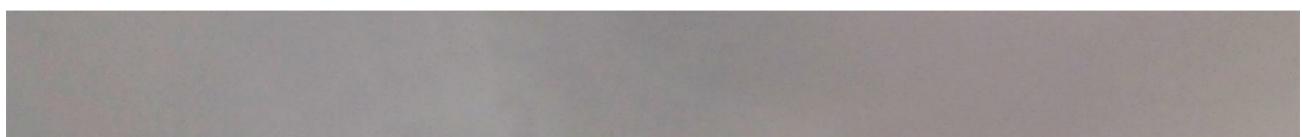
Medical tech giant Stryker opens captive in Gurgaon (<https://analyticsindiamag.com/medical-tech-giant-stryker-opens-captive-in-gurgaon/>)

SGTC strengthens our ability to innovate and develop new products and solutions that help improve and save lives around the world.



Why Maruti invested in Dave AI (<https://analyticsindiamag.com/why-maruti-invested-in-dave-ai/>)

Ananthakrishnan GopalAshok Balasundaram and Sriram P H founded SSPL in 2016.

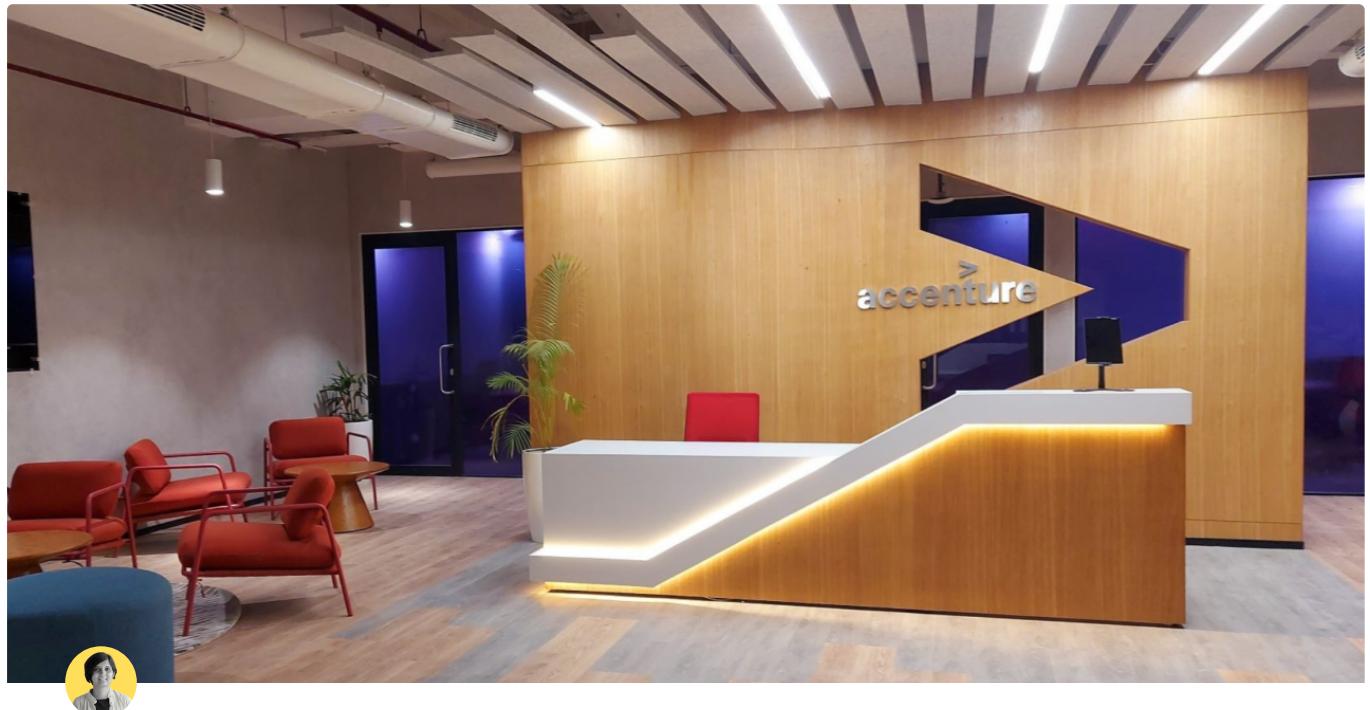




IIT Jodhpur launches MTech in AR/VR

(<https://analyticsindiamag.com/iit-jodhpur-launches-mtech-in-ar-vr/>)

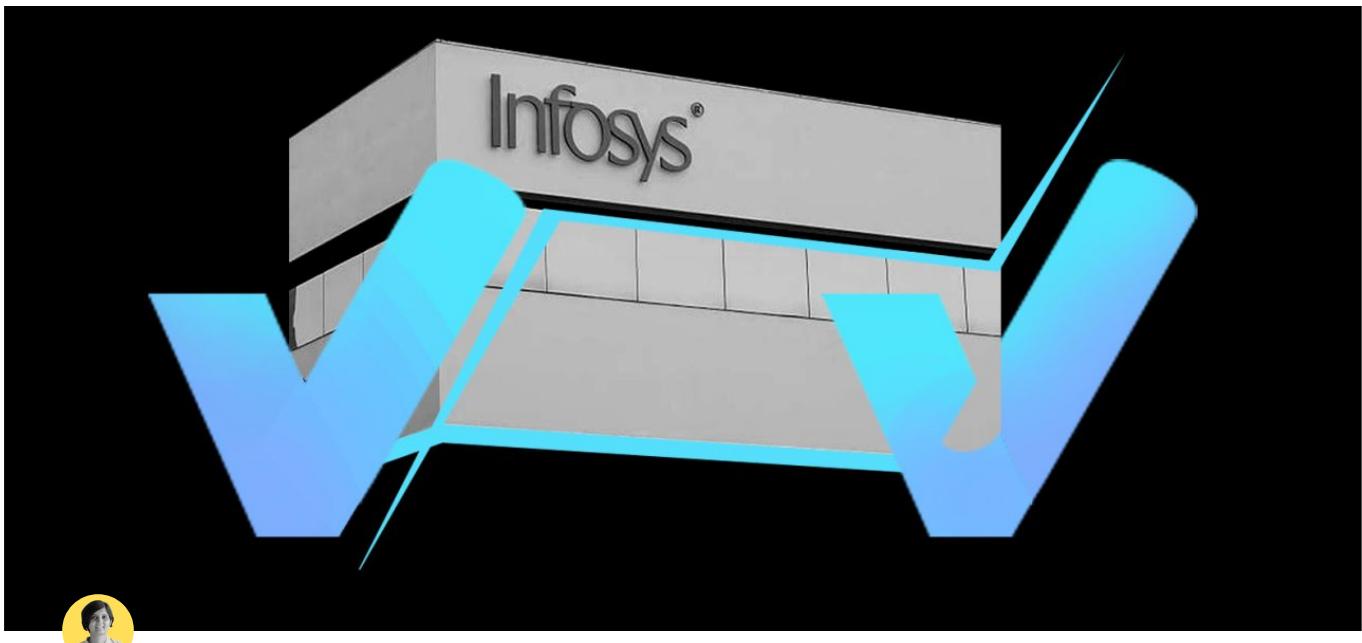
Open to all working professionals, registration for the course will be open until 17th June.



Accenture opens a Technology Center in Indore

(<https://analyticsindiamag.com/accenture-opens-a-technology-center-in-indore/>)

Accenture has Advanced Technology Centers in Mumbai, Bangalore, Chennai, Hyderabad, Pune, Kolkata, and Delhi.



CredAvenue + Infosys: When tech firms reimagine co-lending (<https://analyticsindiamag.com/credavenue-infosys-when-tech-firms-reimagine-co-lending/>)

CredAvenue streamlines the entire co-lending process for lenders and originators by bringing banks and NBFCs in a unified marketplace.

Our Mission Is To Bring About Better-Informed And More Conscious Decisions About Technology Through Authoritative, Influential, And Trustworthy Journalism.

SHAPE THE FUTURE OF TECH

CONTACT US →
([HTTPS://ANALYTICSINDIAMAG.COM/CONTACT-US/](https://analyticsindiamag.com/contact-us/))



(<https://analyticsindiamag.com>)

(<https://www.linkedin.com/company/analytics-india-magazine/>)

About Us

Advertise

Weekly Newsletter

Write for us

Careers

Contact Us

RANKINGS & LISTS

Academic Rankings

Best Firms To Work For

PeMa Quadrant

OUR CONFERENCES

Cypher

The MachineCon

Machine Learning Developers Summit

The Rising

Data Engineering Summit

OUR BRANDS

MachineHack
AIM Recruits
AIM Leaders Council
Best Firm Certification
AIM Research

VIDEOS

Documentary – The Transition Cost
Web Series – The Dating Scientists
Podcasts – Simulated Reality
Analytics India Guru
The Pretentious Geek
Deeper Insights with Leaders
Curiosum – AI Storytelling

AWARDS

AI50
40 under 40 Data Scientists
Women in AI Leadership

EVENTS

AIM Custom Events
AIM Virtual

FOR ML DEVELOPERS

Hackathons
Discussion Forum
Job Portal
Mock Assessments
Practice ML
Free AI Courses

NEWSLETTER

Stay up to date with our latest news, receive exclusive deals, and more.

<https://analyticsindiamag.com/lstm-vs-gru-in-recurrent-neural-network-a-comparative-study/>

Enter Your Email Address

SUBSCRIBE →

© Analytics India Magazine Pvt Ltd 2022

[Terms of use](https://analyticsindiamag.com/terms-use/)(<https://analyticsindiamag.com/terms-use/>)

[Privacy Policy](https://analyticsindiamag.com/privacy-policy/)(<https://analyticsindiamag.com/privacy-policy/>)

[Copyright](https://analyticsindiamag.com/copyright-trademarks/)(<https://analyticsindiamag.com/copyright-trademarks/>)