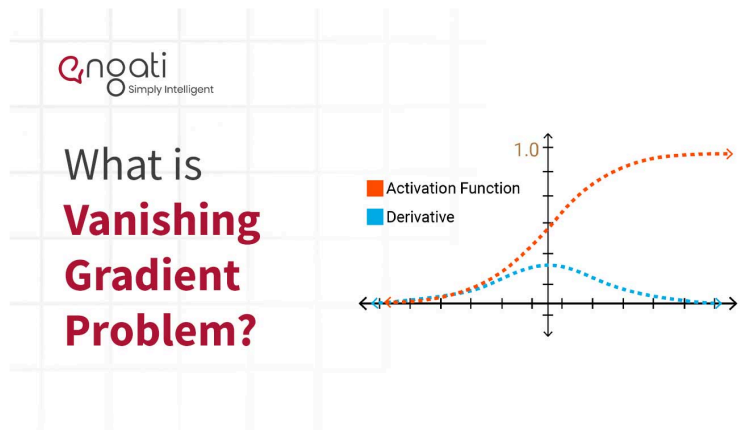


# Vanishing Gradient Problem : Everything you need to know

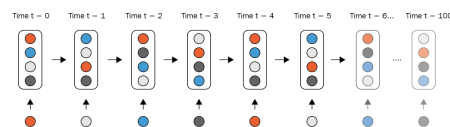


## What is vanishing gradient problem?

Vanishing gradient problem is a phenomenon that occurs during the training of deep neural networks, where the gradients that are used to update the network become extremely small or "vanish" as they are backpropagated from the output layers to the earlier layers.

During the training process of the neural network, the goal is to minimize a loss function by adjusting the weights of the network.

The **backpropagation algorithm** calculates these gradients by propagating the error from the output layer to the input layer.



Source: O'Reilly Media

The consequences of vanishing gradient problem include slow convergence, network getting stuck in low minima, and impaired learning of deep representations.

## What is the importance of gradient descent in training neural



3. Optimization landscape exploration

4. Efficiency & scalability

5. Generalization & learning

## Vanishing Gradient Problem In Deep Learning

The vanishing gradient problem in deep learning is a big issue that comes up when training neural networks in deep structures with lots of layers. It happens when the gradients used to update the weights during backpropagation get small disappearing. This makes it hard for the network to learn because the weights of the earlier layers change, which slows down or stops training altogether. Fixing the vanishing gradient problem is key to train deep neural networks . People have come up with ways to deal with this challenge, like setting up weights in a certain way at the start, using batch normalization, and trying out activation functions such as ReLU.

## What is the impact of the Vanishing Gradient Problem in Deep Learning Models?

After understanding what is vanishing gradient problem, here's how it impacts **Deep learning** models:

1. Limited learning capacity
2. Difficulty in capturing the long-term dependencies
3. Slow convergence & training instability
4. Preferential learning in shallow layers
5. Architectural design considerations

## How do you know if your model is suffering from the vanishing gradient problem?

Here are some signs that are indicators of your problem suffering from the vanishing gradient problem:

- The parameters of the higher layers change to a great extent, while the parameters of lower layers barely change (or, do not change at all).
- The model weights could become 0 during training.
- The model learns at a particularly slow pace and the training could stagnate at a very early phase after only a



## Vanishing Gradient Problem

The **vanishing gradient problem** is caused by the fact that while the process of backpropagation goes on, the gradient of the early layers (the layers that are nearest to the input layer are derived by multiplying the gradients of the later layers (the layers that are near the output layer). Therefore, if the gradients of later layers are less than one, their multiplication vanishes at a particularly rapid pace.

3x your revenue  
with Chatbots and  
Live Chat

[SCHEDULE A DEMO](#)

## How to calculate gradients in neural networks?

1. Forward pass - Input is propagated through the network, layer by layer, for computing the output predictions. It has 2 layers - linear transformation & activation function.
2. Loss calculation - After forward passing is done, the output predictions are compared to true labels to compute the loss.
3. Backpropagation - Used to calculate the gradients of the loss function wrt the weights of the network. It has 2 steps - backward pass & gradient calculation.
4. Weight update - After calculating the gradients, the total weights of the network are updated to minimize the loss function.

## Why is the vanishing gradient problem significant?

The vanishing gradient problem causes the gradients to shrink. But, if a gradient is small, it won't be possible to effectively update the weights and biases of the initial layers with each training session.

These initial layers are vital for recognizing the core elements of the input data, so, if their weights and biases are not properly updated, it is possible that the entire network could be inaccurate.

### TABLE OF CONTENTS

What is vanishing gradient problem?

What is the importance of gradient descent in training neural networks?

Vanishing Gradient Problem In Deep Learning

What is the impact of the Vanishing Gradient Problem in Deep Learning Models?


How do you know if your model is suffering from the vanishing gradient



shing gradient

adients in neural

gradient problem

Switch to Engati: Smarter choice for  
WhatsApp Campaigns 

TRY NOW

couple of activation functions:

1. **Sigmoid** & Tanh activation function
2. ReLU activation function
3. Leaky ReLU and Parametric ReLU activation function
4. Exponential linear units
5. Scaled exponential units
6. Swish activation functions

## How do you overcome the vanishing gradient problem?

Here are some methods that are proposed to overcome the vanishing gradient problem:

1. Residual neural networks (ResNets)
2. Multi-level hierarchy
3. Long short term memory (LSTM)
4. ReLU
5. Batch normalization
6. Weight initialization
7. Gradient clipping
8. Skip connections
9. Gated architectures
10. Reduced network depth
11. Faster hardware

### 1. Residual neural networks (ResNets)

Here's a better understanding of how does ResNet solve vanishing gradient problem by the use of skip connections to learn the residual mapping, enabling easier gradient flow & efficient training of deep neural networks.

### 2. Rectified linear unit (ReLU)

ReLU avoids the saturation issues of functions like sigmoid tangent or hyperbolic tangent, that can cause the gradients to vanish.

ReLU allows for better gradient flow, promoting more effective training of deep networks.

### 3. Multi-level hierarchy



**LSTMs** are designed to capture the long-term dependencies by allowing information to persist over many time steps by incorporating the memory cells & gating mechanisms.

This enables the effective training of deep recurrent networks & the modeling of complex temporal dependencies in sequential data.

## 5. Batch normalization

It is a technique that normalizes the activations within each min-batch during training.

It helps to reduce the internal shifts & ensures more stable gradient propagation.

## 6. Weight initialization

Initializing the weights with appropriate values minimizes the possibility of the gradients becoming too small or too large.

Techniques like Xavier/Glorot initialization set the initial weights based on the network's inputs & outputs, ensuring a more balanced & stable initialization.

## 7. Gradient Clipping

It is a technique where the gradients are rescaled to the maximum threshold during backpropagation.

If the gradients exceed the thresholds, they are scaled down to prevent them from exploding.

## 8. Skip connections

It provides direct connections between layers, allowing the gradients to bypass multiple layers during backpropagation.

## 9. Gated architectures

Architectures like Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) incorporate gating mechanisms that selectively control the flow of information within the network.

These architectures allow the network to retain & propagate relevant information over long sequences.

GRUs are designed to overcome the limitations of traditional RNNs by providing efficient solutions for modeling the sequential data.



Shallower networks have shorter paths for propagating gradients, reducing the likelihood of the gradients vanishing or exploding.

## What is exploding gradient problem?

It's a counterpart to the vanishing gradient problem that occurs when the gradients in a **neural network** grow exponentially during the training phase.

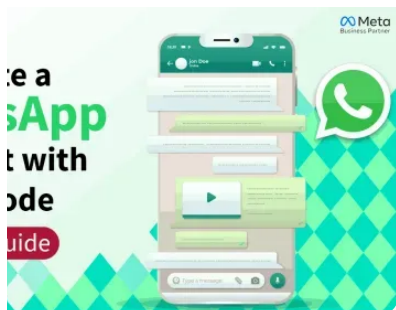
This can lead to unstable training dynamics, making it challenging for the network to converge into an optimal solution.

## Continue Reading

### MARKETING

#### WhatsApp Bot for Free With No Code (2024)

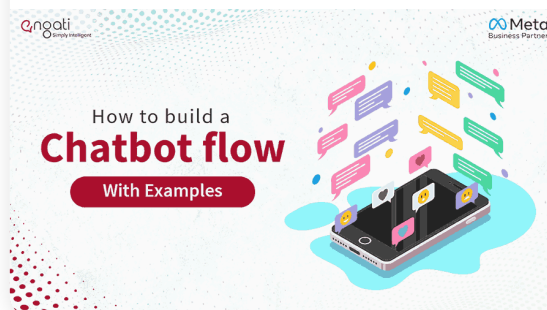
mins



### CHATBOT

#### How to Design Chatbot Conversational Flow with Examples (2024)

Aug 3, 20236-8 mins



### TECH CORNER

#### Contrastive Learning in

Nov 3, 20232-3 mins



Leverage the power of GPT and LLMs for customer experience

[Experience Now!](#)[Solutions](#) ▼[Features](#) ▼[Pricing](#)[Partners](#)[Resources](#) ▼[SIGN UP FOR FREE](#)[VIEW OUR BLOGS](#)



Solutions ▾

Features ▾

Pricing

Partners

Resources ▾

SIGN UP FOR FREE

Name\*

Email\*

+00\*

Phone Number\*

SUBMIT

About Engati

Engati enables transformative Customer eXperiences for businesses to connect & engage with customers, wherever they exist.

We activate Conversational Intelligence to scale customer experience and build deep and lasting connections.



SOLUTIONS

INDUSTRIES

TRANSFORMED

E-commerce & Retail  
Real Estate  
Edtech  
Travel & Hospitality  
Insurance & Banking

Government  
Logistics

USE CASES

WhatsApp Chatbot  
WhatsApp Marketing  
Instagram Chatbot  
AI Chatbot  
Omnichannel  
Engagement  
Shopify Chatbot  
E-Commerce Insights  
Assistant

PLATFORM

WHAT WE ENABLE

Low-code Bot Builder  
eSenseGPT  
Natural Language Processing  
Live Chat  
Integrations  
Multi-lingual  
Conversations

RESOURCES

LEARN MORE

Blogs  
Glossary  
eBooks  
Webinars  
Engatica  
Engati vs Yellow  
Engati vs Haptik  
Engati vs Wati.io  
Engati vs  
Chatbot.com

GET STARTED

Tutorials  
Developer Docs  
Bot Marketplace

COMPANY

About us  
Careers [We're Hiring](#)  
News  
Partners  
Contact Us

CUSTOMERS

PARTNERS

PRICING

● ALL SERVICES OPERATIONAL



(c) 2024 Engati Technologies Inc. All rights reserved.

GDPR

Term of Use

Privacy Policy