ABHISHEK RAI

Senior AI Engineer (Vidyo.ai)

CONTACT

+91-9818289547

abhishekrai.tech@gmail.com

Varanasi

linkedin.com/in/abhishek-rai-286502113

SOFT SKILLS

- Analytical
- Communicator
- Collaborator
- Decisive
- Team Management
- Planner

EDUCATION & CERTIFICATIONS

- **2019:** Bachelor of Technology (Computer Science Engineering) from Galgotia College of Engineering & Technology with 7.09 GPA
- **2014:** 12th from Sunbeam School, Varanasi with 85.8%
- NLP Fundamentals Nano Degree - Udacity
- Java Core Technology - INCAPP S18/4896/1026
- Introduction to Machine Learning in Production
- Improving Deep Neural Networks: Hyperparameter Tuning, Regularization and Optimization - Coursera EBA6TV7GX3AZ
- Neural Networks And Deep Learning - Coursera G7YJ5NFHKXTP
- Structuring Machine Learning Projects - Coursera L8UYVCBWFZJH
- Introduction to cloud computing - Udemy UC-V7R0880F
- GPT-3 Full Guide - Udemy
- The Complete Database Design and Modelling Beginners Tutorial - Udemy
- Transformers for Natural Language Processing - Udemy

Dynamic, visionary, & achiever *with an unmatched track record in conceptualizing & implementing effective ideas & strategies; utilizing extensive knowledge in* ***Natural Language Processing (NLP) and Generative AI Models*** *to foster innovation and deliver significant outcomes*

PROFILE SUMMARY

- Result-oriented professional with **over 4 years of experience in Artificial Intelligence and Data Science**
- Skilled in **designing, developing, and optimizing prompts for LLMs, such as GPT-3, GPT-4, or similar models.**
- Resourceful in working closely with cross-functional teams, including machine learning engineers, product managers, and domain experts to **leverage LLMs for different use cases, such as natural language understanding, text generation, and content creation.**
- Expert in **conducting experiments and A/B testing to improve prompt effectiveness.**
- Possess an understanding of **deep learning, reinforcement learning, natural language processing, computer vision, or other subfields of AI and ML**
- **Resourceful in managing projects from conceptualization, and visualization to technology mapping** and final execution of projects
- **Spearheaded the establishment of vidyo.ai's core AI value** proposition as the Founding AI Engineer, while also serving as the head of the AI department, resulting in significant achievements
- **Played a key role in propelling vidyo.ai's remarkable growth**, transforming it from zero-dollar annual recurring revenue to over $1.5 million in just one year
- **Demonstrated a track record of swiftly adapting to dynamic business environments** and effectively solving complex problems through a practical and innovative approach
- Possess a wealth of knowledge in **deep learning technologies, utilizing them to create prototypes and analytical models** that effectively tackle challenges faced by both new and existing clients

CORE COMPETENCIES

- RAG
- Attention Mechanisms
- Transformers
- Langchain
- Large Language Models
- Computer Vision
- Generative AI
- Prompt Engineering
- Natural Language Processing

TECHNICAL SKILLS

Statistics/ Machine-Learning	GenerativeAI, Prompt Engineering, Statistical Analysis, LangChain, R Linear/Logistic Regression, Clustering, Computer Vision, NLP, Deep Learning, ML for audio
Other Skills	Transformers, Hugging face, Large Language Models(LLMs) , Tensorflow, PyTorch, Keras, Numpy, Pandas, SciKit Learn, Matplotlib, PySpark, Jupyter Notebook, Sentiment Analysis, Image Classification, Image Identification, Image Verification, Image Detection, Exploratory Data Analysis(EDA), Multi-Layer Perceptron (MLP), Neural Networks, Support Vector Machines (SVM), Landmark Detection, Image Segmentation, Transfer Learning, ORB (Oriented Fast and Rotated Brief), Attention Mechanism, Variational Autoencoders (VAE), GAN, Residual Network, GRU, LSTM, RNN, CNN, YOLO, OpenCV, CUDA, Nnabla and Keras

WORK EXPERIENCE

Sept'22 – Present| Senior Artificial Intelligence Engineer (ML Division) | Vidyo.ai, Bangalore

Key Result Areas:

- Spearheading the development of vidyo.ai's core AI value proposition as the Founding AI Engineer and head of the AI department, resulting in an impressive annual revenue from **0 $ to over $1.5 million within a year**.
- Evaluating model performance using appropriate metrics and testing methodologies
- Implementing monitoring systems to track model performance and make necessary updates or improvements as data distributions change or models degrade over time.
- Deploying machine learning models into production environments, ensuring they are scalable, maintainable, and performant & integrating models into software systems and applications
- Innovating and implementing cutting-edge AI technologies, with a focus on NLP and Generative AI models
- Researching and developing state-of-the-art AI technologies, including transformers and generative AI models
- Developing and optimizing AI models for performance and scalability, ensuring seamless integration into real-world applications
- Collaborating with back-end and front-end engineers to integrate AI models into the product pipeline, ensuring seamless integration and enhancing the overall functionality and user experience
- Communicating AI concepts to stakeholders and producing technical documentation, including system architecture and model evaluations
- Ensuring compliance with data protection regulations and maintaining high standards of reliability and accuracy through rigorous testing

Nov'19 – Sept'22 | Software Engineer/ Senior Software Engineer (ML Division) | Sony India Software Centre, Bangalore

Key Result Areas:

- Developed and implemented a Machine Learning Operations (MLOps) framework, driving scalable full-stack solutions for model versioning and retrieval
- Presented OpenCV machine learning solutions to senior management, facilitating informed decision-making
- Applied statistical machine learning solutions to various business problems, leveraging complex, high-dimensional datasets to extract valuable insights and drive informed decision-making
- Researched and implemented appropriate ML algorithms and tools, manipulating and analyzing complex, high-dimensional datasets
- Collaborated in a cross-functional environment to deliver scalable solutions and meet project requirements

INTERNSHIP

Jan'19 - Mar'19 | Machine Learning Intern at Sony India Software Centre
Highlights: – Completed Deep Learning models in the SONY machine learning framework called Neural Network Libraries or Nnabla

Jul'18 - Aug'18 | Machine Learning and NLP Intern at Adfluence Hub
Highlights: – Sentiment analysis using NLP on collected comments

PERSONAL DETAILS

Date of Birth	21st November 1996
Languages Known	English and Hindi

ANNEXURE – PROJECTS

Making AI-based Content Repurposed clips using Generative AI and Large Language Models (LLMs)

- Content Repurposing using Generative AI and Large Language Models (LLMs) (for Vidyo.ai)
- Content repurposing is the process of generating various short content from a large pillar of content
- The technology used are Generative AI, Large Language Models (LLMs) and various other NLP techniques
- Most important sections of large clip is selected to create short meaningful content
- These sections selected can be either – continuous or discontinuous based on our decision which is later concatenated to give two different types of output clips:
    - Continuous Clips
    - Discontinuous clips
- Along with Content Repurposing, did many tasks such as segmenting content into various topics

- Used Prompt Engineering to get keywords for each short content created
- Used various prompt engineering techniques such as in-context learning, and one-shot learning to get a comparative virality score for each
- Used RAG with different pre retrieval (similarity search, mmr, query based retrieval) and post retrieval combinations (Map reduce, refine and rerank)

AI Emoji Predictor
- Developed an AI algorithm which predicts and inserts emoji in subtitles string both linewise and position-wise based on our choice
- The emojis when used makes the content more engaging

Long Pause and filler word removal from a content
- Drive this project from idea creation phase to deployment stage
- The long pauses between speech is removed and filler words like umm, uhm, etc are removed from a video and subtitle by identifying their timestamps
- The subtitle sync function is made to sync the subtitle with new trimmed video from the old larger video
- The video becomes a lot more engaging when it has no long pauses and filler words which helps the content creators in making better content

AI/ML-based MDR personalization - Personalized Sound
- The idea generation was done by my team and accepted by Sony Home Entertainment Sound Products Inc., Tokyo for Sony's MDR
- The idea is to provide a Personalized Sound Experience using unique Hearing Profiles of users
- Tech Used: - Pytorch, PyQt, Numpy, Pandas and DNN
- ML-based gain predictor was used which predicts gains for different users based on their hearing capacity (and loss). These gains can be used to perform personalized amplification at different frequencies

Analyze user's hearing profile and Provide Recommendations
- A method to analyze the user's hearing profile was made and warn the user to take medical assistance if required. This was done by implementing the Pure Tone Audiometry test in Python. A GUI was made using Tkinter
- PTA test is completed using GUI and it produces audiogram of the user which shows the user's hearing loss at different frequencies

Singing Voice Conversion using Deep Learning- R&D project with Sony's Tokyo Laboratory 21
- Voice conversion means transforming the speech of the source speaker to that of the target speaker preserving linguistic information
- In this R&D we achieved the state of the art result and then improved the already SOTA results of voice conversion
- Worked on 2 main approaches and used VQVAE, StarGAN and Scyclone for these approaches
- VQVAE and StarGAN were used for many to many voice conversions while Scyclone was used for one to one

Source Separation using Deep Learning- R&D project with Sony's Tokyo Laboratory 21
- Worked on Speech source separation and performed POC of SOTA papers such as that of ConvTasnet and ConvTasNet V-2
- Devised new novel training algorithm called OR-PIT (one and rest permutation invariant training) used for Recursive Speech Separation

Cube Image Detection and localization
- Used DNNs to detect cube and perform image localization

Academic Projects:
- Image Captioning using CNN as Encoder and LSTM as decoder
- Machine Translation using GRU to translate English to French
- Face Recognition system: - Haar Cascade forface detection and used pretrained CNN to compute face encodings to perform face recognition
- YOLO object detection using filter boxes and Non-max suppression