Gunjan Chhablani

chhablani.gunjan@gmail.com gchhablani gchhablani gchhablani.github.io +91-7719900640

Employment

Oracle India Pvt. Ltd. **Sep 2020 – Present**
Applications Engineer *Hyderabad, TG, India*

- Oracle Fusion Applications - HCM Cloud development.
- Actively involved in UI and Model(View/Entity) layer implementation - development and unit testing - of various modules in Global HR - Oracle HCM Cloud.
- Identifying process automation opportunities and automating various tasks including REST API testing.
- Worked as Sprint Master for Project Progress and Release Management. Managed Product and Sprint Backlog to achieve the Sprint objectives in agile methodology.
- Oracle ADF, Java, SQL, PL/SQL, ADE, Shell (csh, bash), RESTful Services, JDeveloper 11g/12c, Python, JavaScript.

Internships

University of Pittsburgh **Jan 2020 – Aug 2020**
Student Intern *Pittsburgh, PA, USA*

- Studied the performance of novel weight initialization technique in neural networks - Perceptually Inspired Deep Learning (PIDL) on MNIST/FashionMNIST and OCT segmentation datasets.
- Used U-Net/U-Net++ deep learning architectures with VGG-19 backbone on a home-grown Choroid segmentation dataset in OCT scans.
- Abstract accepted at ARVO, 2021. Another paper in the pipeline for publication at TVST, ARVO.

IBM India Pvt. Ltd. **May 2019 – Jul 2019**
Project Trainee *Gurgaon, HN, India*

- Actively contributed to the ongoing project at IBM Watson Services on Instance Segmentation.
- Annotated car parts and damages on 100+ insurance claims using VGG Image Annotator.
- Helped in developing a MaskRCNN model in TensorFlow for car parts segmentation and damage detection.
- Achieved 85% Mean Average Precision score on car parts segmentation.

L.V. Prasad Eye Institute - Winter Intern **Dec 2018 – Jan 2019**
Winter Intern *Hyderabad, TG, India*

- Segmented Hard-Exudates from Real-life OCT Scans on home-grown dataset with 100 images.
- Segmented choroids in 900+ OCT scans using Tensor Voting based segmentation tool in MATLAB.
- Built networks like U-Net/Deep U-Net in Keras for segmentation of hard-exudates with Dice Coefficient as evaluation metric.

Education

BITS Pilani, K.K. Birla Goa Campus **Aug 2016 – Sep 2020**
*B.E. Computer Science - **CGPA** - **9.65/10** - Batch Rank 3/617* *Goa, India*

Publications

- Victor Sanh, Albert Webson, Colin Raffel, Stephen H. Bach, Lintang Sutawika, Zaid Alyafeai, Antoine Chaffin, Arnaud Stiegler, Teven Le Scao, Arun Raja, Manan Dey, M Saiful Bari, Canwen Xu, Urmish Thakker, Shanya Sharma Sharma, Eliza Szczechla, Taewoon Kim, **Gunjan Chhablani**, Nihal Nayak, Debajyoti Datta, Jonathan Chang, Mike Tian-Jian Jiang, Han Wang, Matteo Manica, Sheng Shen, Zheng Xin Yong, Harshit Pandey, Rachel Bawden, Thomas Wang, Trishala Neeraj, Jos Rozen, Abheesht Sharma, Andrea Santilli, Thibault Fevry, Jason Alan Fries, Ryan Teehan, Stella Biderman, Leo Gao, Tali Bers, Thomas Wolf, Alexander M. Rush. **Multitask Prompted Training Enables Zero-Shot Task Generalization**. Accepted at ICLR 2022 (Spotlight).
- Quentin Lhoest, Albert Villanova del Moral, Yacine Jernite, Abhishek Thakur, Patrick von Platen, Suraj Patil, Julien Chaumond, Mariama Drame, Julien Plu, Lewis Tunstall, Joe Davison, Mario Šaško, **Gunjan Chhablani**, Bhavitvya Malik, Simon Brandeis, Teven Le Scao, Victor Sanh, Canwen Xu, Nicolas Patry, Angelina McMillan-Major, Philipp Schmid, Sylvain Gugger, Clément Delangue, Théo Matussière, Lysandre Debut, Stas Bekman, Pierric Cistac, Thibault Goehringer, Victor Mustar, François Lagunas, Alexander Rush, Thomas Wolf.

Datasets: A Community Library for Natural Language Processing. Published at EMNLP Demo Track 2021 (Best Demo Paper).

• Abheesht Sharma*, **Gunjan Chhablani***, Harshit Pandey*, Rajaswa Patil. **DRIFT: A Toolkit for Diachronic Analysis of Scientific Literature**. Published at EMNLP Demo Track 2021.

• **Gunjan Chhablani***, Abheesht Sharma*, Harshit Pandey*, Yash Bhartia, Shan Suthaharan. **NLRG at SemEval-2021 Task 5: Toxic Spans Detection Leveraging BERT-based Token Classification and Span Prediction Techniques**. Published at SemEval Workshop, ACL-2021.

• Abheesht Sharma*, Harshit Pandey*, **Gunjan Chhablani***, Yash Bhartia, Tirtharaj Dash. **LRG at SemEval-2021 Task 4: Improving Reading Comprehension with Abstract Words using Augmentation, Linguistic Features and Voting**. Published at SemEval Workshop, ACL-2021.

• Shan Suthaharan, **Gunjan Chhablani**, Kiran Kumar Vupparaboina, Jose-Alain Sahel, Kunal K. Dansingani, Jay Chhablani. **An automated choroid segmentation approach using transfer learning and encoder-decoder networks**. Extended abstract published at Invest. Ophthalmol. Vis. Sci., ARVO Special Issue 2021.

• Raymond Lai-Man Wong, Sumit Randhir Singh, Mohammed Abdul Rasheed, Abhilash Goud, **Gunjan Chhablani**, Srujanika Samantaray, Samantha AnkiReddy, Kiran Kumar Vupparaboina, and Jay Chhablani. **En-Face Choroidal Vascularity in Central Serous Chorioretinopathy**. Published at European Journal of Ophthalmology 31, no. 2 (March 2021).

Open Source Contributions

HuggingFace **Feb 2020 – Present**

Core Contributor

- Active core contributor to HuggingFace transformers and datasets libraries.
- Contributed and maintaining the VisualBERT and FNet models in PyTorch. Currently working on adding SMITH and PLBART models to the repository.
- Contributed over 11 datasets and worked on metadata validation for the datasets library. Currently working on automated metadata generation for datasets. Paper accepted at EMNLP Demo Track 2021.

Personal Projects

Multimodal Multilingual Models | JAX/Flax, PyTorch, Python, TPUv3-8 VM **Jul 2021 – Aug 2021**

- Participated with one other team member in the HuggingFace JAX/Flax sprint for multilingual image captioning and multilingual visual question answering (VQA).
- Created a novel combination of ViT encoder-BART decoder for image captioning and ViT encoder+BERT encoder in Flax for VQA.
- Worked on CC12M image captioning dataset translation using MarianMT and MBART50 models and trained the models on TPUv3 VMs provided by Google.
- Prepared a Streamlit demo for both shared on HuggingFace spaces. The project was ranked among top-15 out of 50 other teams.

DRIFT | PyTorch, Python, Streamlit **Jul 2021 – Aug 2021**

- Actively developed an open-source tool with Streamlit-backed GUI for diachronic analysis of scientific literature.
- Compiled and implemented 10 different kinds of analysis techniques including semantic drift, acceleration analysis, wordclouds, trend detection, etc.
- Performed extensive analysis on arxiv.cs.CL dataset. Paper accepted at EMNLP Demo Track 2021.

SemEval-2021 Task-5 Toxic Spans Detection | PyTorch, Python, GitHub **Sep 2020 – Jan 2021**

- Worked on toxic spans detection where the offsets of the toxic phrases are to be detected in a given text.
- Performed a thorough analysis of token classification, span prediction, and novel hybrid approaches involving LSTM, CRF and BERT.
- Achieved an offset-wise F1 score of 67.53% (rank 28/92) officially, and 68.95% during the post-evaluation period. Paper accepted at SemEval2021 workshop (ACL-IJCNLP 2021).

SemEval-2021 Task-4 ReCAM | PyTorch, Python, GitHub **Sep 2020 – Jan 2021**

- Worked on a cloze-style question answering task with abstract words as options.
- Analyzed the working of existing state-of-the-art techniques like BERTCloze, GAReader, and related variants in PyTorch.

- Improved the performance using engineering features for imperceptibility and using augmentation and pretraining and hyponyms for non-specificity.
- Achieved a score of 77.28% (rank 15/23) on imperceptibility, 78.98% (rank 17/28) on non-specificity subtasks. Paper accepted at SemEval2021 workshop (ACL-IJCNLP 2021).

Detecting Terrorism Promotion on Twitter | Python, Jupyter Notebook, SQL **Aug 2019 – Aug 2020**

- Extracted context-based and socio-psychological features from the data using LiWC, GrammarBot, and TextRazor API after pre-processing and training models (XGBoost, Neural Networks, Linear Regression) on the data.
- Used semi-supervised learning to make a mental-state model to identify promotion of terrorist sympathizers, and a categorization method to classify users based on their level of involvement.

Technical Skills

Languages: Python, Java, C, C++, SQL, Shell
Technologies/Frameworks: Git, PyTorch, Keras, Tensorflow, JAX/Flax, ADF, Scikit-learn, Streamlit

Additional Info

Scholarships: *Institute Merit Scholarship* - for ranking among top 1% of the batch for 8 semesters at BITS Goa, *National Talent Search Scholarship* 2014 – by NCERT for qualifying National Talent Search Examination.
Certificates: Deep Learning Specialization (Coursera), Machine Learning (Coursera), Intro to Data Science in Python (Coursera), Applied Machine Learning in Python (Coursera), Algorithmic Toolbox (Coursera).
Teaching Experience:
Instructor of ML/DL Course Course : Intro to ML and DL under the Center for Technical Education at BITS, Goa.
Teaching Assistant : Theory of Computation — Data Structures and Algorithms — Discrete Structures for Computer Science.