Hack into Text: NLP powered by PyTorch

Description

Are you curious about how the cool AI generated images are made? Or how German is translated into French without losing the meaning? Or how chatbots work? Or how summarisation works? Or how autocomplete works? Or how keyword extraction works? Or how Alexa understands what we say? Or how fake news is detected?

This workshop will teach you how to train, adapt and optimize a NLP model. We will be going over all of these tasks and explaining how these work. The workshop will feature a hands on tutorial on Pytorch and ONNX Runtime for deployment.

Learning Outcomes

After this workshop, you will be able to:
- Understand what the terms deep learning and NLP mean
- Use Pytorch in building and training machine learning models
- Understand how NLP pipelines work and how downstream tasks are used

Prerequisite Knowledge

In order to get the most out of this workshop you should be comfortable with the following concepts:
- Python Programming, including Object Oriented Programming (OOP)
- Basic Machine Learning and statistics
- Jupyter Notebooks and Python Scripting

Pre-Workshop Checklist

Before the workshop, please make sure you complete the following items:

- Get enough sleep
- We will be using Colab so please create a Google account if you don't have one already
- Please create a Huggingface.co account
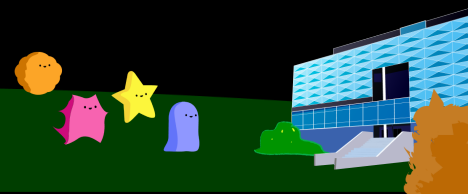
Technical Jargon and Definitions

- **NLP**: Natural language processing (NLP) refers to the branch of computer science—and more specifically, the branch of artificial intelligence or AI—concerned with giving computers the ability to understand text and spoken words in much the same way human beings can.
- **PyTorch**: PyTorch is an open source machine learning framework based on the Torch library, used for applications such as computer vision and natural language processing, primarily developed by Meta AI.
- **ASR**: Automatic Speech Recognition, as it's known in short, is the technology that allows human beings to use their voices to speak with a computer interface in a way that, in its most sophisticated variations, resembles normal human conversation.
- **Text Generation**: Generating text is the task of producing new text. These models can, for example, fill in incomplete text or paraphrase.
- **Question Answering**: Question Answering models can retrieve the answer to a question from a given text, which is useful for searching for an answer in a document. Some question answering models can generate answers without context!
- **Fill-mask**: Masked language modeling is the task of masking some of the words in a sentence and predicting which words should replace those masks. These models are useful when we want to get a statistical understanding of the language in which the model is trained in.
- **MT**: Machine Translation is the task of automatically converting one natural language into another, preserving the meaning of the input text, and producing fluent text in the output language.
- **Summarization:** Summarization is the task of producing a shorter version of a document while preserving its important information. Some models can extract text from the original input, while other models can generate entirely new text.
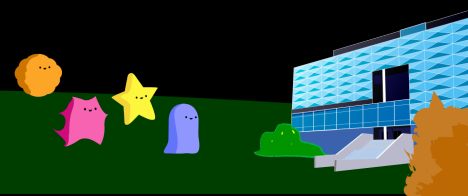
Timeline (60 Minutes)

| Time	Module	Description
2 Mins	NLP Introduction	What is NLP?
3 Mins	Why NLP?	Why do we need NLP? Why should you care about NLP?
5 Mins	What is Pytorch?	Pytorch Basics
10 Mins	torch.nn	Pytorch model building
5 Mins	Transformers is all you need	Intro to Huggingface transformers and how to use them
30 Mins	Build a better google translate	Machine Translation example
5 Mins	Export to ONNX Runtime	What is ONNX runtime and how does it speed up inference?
5 Mins	Question and Answers	Get your doubts solved

Workshop Lead Contact

Gagan Bhatia
Email: gbhatia880@gmail.com Linkedin: https://www.linkedin.com/in/gaganbhatiaml/ GitHub: https://github.com/gagan3012
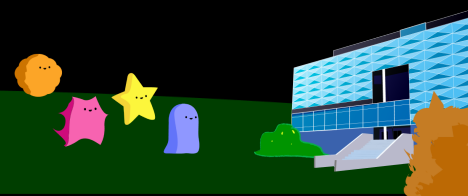
Additional Resources

Hack the North Resources

Hack the North 2022 Event Schedule

Check this out to stay up-to-date on activities, workshops, and other key happenings this weekend.

Workshop-Specific Resources

1. Python Resources: https://www.fullstackpython.com/best-python-resources.html
2. Python Get started: https://www.python.org/about/gettingstarted/
3. Pytorch Basics: https://pytorch.org/tutorials/beginner/basics/intro.html
4. HuggingFace Basics: https://huggingface.co/docs