Google DeepMind Research Ready Programme 2024

Research Projects

Constraining Large Language Model (LLM) for Creative Purposes
If you ask ChatGPT for a sonnet comprising only words beginning with the letter "s", it confidently says "of course" and produces a great sounding poem. However, the poem will contain words like "the", "and" and "of", as well as some non-S words which make the poem rhyme. This highlights that LLMs are probabilistic models - they cannot produce output which is guaranteed to have certain qualities. Working with transformers for music generation, we have successfully investigated a process where the logits produced by the model are constrained according to the user. This has enabled us to produce great sounding piano music with certain constraints guaranteed, e.g., arpeggios in the left hand. In this project, the students will do the same for a large language model, to see if the process generalises from applications to music to applications in generative text.

Automatic Annotations
Investigate what happens in a game using Large Language Models. LLMs have trouble thinking outside of the box and very large LLMs are needed to be used. In this project, students will investigate if we can use smaller LLMs as well as how we can increase and evaluate the diversity of the gameplay descriptions.

Situation Similarity
Using Large Language Models as gameplaying agents doesn't often work, as they fail to properly consider their options. Can we provide wikipedia data to these agents based on the situation they are in? However, how do we detect in what situation we are in? In this project, students will investigate how we can detect if a situation is like what we've seen before.

Text Diversity
This project is related to automatic annotations. Can we measure how diverse two sets of text are? Meaning we have two different processes (for example LLM prompts) that generate text. We are now interested in which process generates more diverse text. Of course, this isn't so simple to say, as they might generate distinct sets of text. So, this project could start simple by visualizing the datasets (sentence transformers + dimensionality reduction). The students could then move on to more complex topics.

Symbolic Music Segmentation – Phrase Border Identification
Symbolic music segmentation is the task of dividing symbolic melodies into smaller meaningful groups. Symbolic music segmentation is an important task in Music In-formation Retrieval (MIR) and it serves as a basis for other applicative tasks, such as melodic feature computation, melody indexing, and retrieval of melodic excerpts as well as for musicological

research and recently music generation research. In this project, students will build a transformer-encoder based (or similar) deep learning model to identify musical phrase boundaries in a supervised learning context with annotated data. We will also explore the benefits of pre-training under various settings and fine-tuning the phrase identifier model.

Style Transfer for Folk Melodies

Musical style transfer involves transforming a melody from a certain style/genre to a different one. Current research on music style transfer has some limitations. In this project, students will explore a corruption-refinement based technique for genre transfer. Several corruptions such as melodic note pitch, duration and bar masking along with incorrect transposition, fragmentation, etc. will be tested on multi-genre datasets. At generation time, these will be tested on a folk song dataset with known phrases to preserve the structural information. The goal would be to make the folk melodies sound more “jazzy” or “classical”. The melodies may later be re-harmonized in the target style by an accompaniment generation model.

Reducing Computational Complexity for Multi-Instrument Music Generation Models

Multi-track multi-instrument music generation in the symbolic domain has many challenges. Among them is the sequence length overflow problem in which the length of tokens increase when all instruments are added to the mix. Recent research on multi-instrument symbolic music generation has introduced various encoding techniques such as byte pair encoding to fit in longer tokens as part of the sequence. But this may not be optimal for maintaining long term structure. In this advanced project, we will explore a multi-modal approach that involves audio spectrograms and symbolic tokens. New instruments would be added to the audio mix as spectrograms would keep the memory complexity constant while the neural network would generate the accompanying symbolic tokens for the target instrument. At generation time, the instruments specified by the user would be added in to the mix iteratively to preserve memory.

Explainable Text-To-Music Generation

In this advanced project, our goal is to build a text-to-music generation model that is capable of explaining its own generation. This would involve a reciprocal loop where the model first generates a textual account detailing the next intended phrase's qualities, then produces a matching symbolic excerpt, before repeating the cycle to create an evolving track with paired descriptive justifications. The textual description of the phrase can include listening models, tonal tension algorithms, genre, pitch contour, note density, average pitch, key and time signature, among various other attributes. This would be a more advanced multi-modal AI problem that may take longer than a month to yield results.

Bias Correction in Seasonal Temperature Forecasts

Accurate seasonal temperature forecasts are crucial for various sectors, including agriculture, energy, and water resource management. However, these forecasts often suffer from biases and drift, leading to significant errors in decision-making. This project aims to address these issues by employing advanced deep learning techniques and transfer learning to correct biases in seasonal temperature forecasts from the Copernicus Climate Change Service (C3S).

Project 1: U-Net Architecture Design for Bias Correction
Project 2: Transfer Learning for Bias Correction Across Different Forecasts

Spatial Audio Generation through Visual Cues
In scenarios with multiple speakers, such as online conferences or live streaming, a common challenge is the difficulty in understanding the content due to competing voices and background noise. In this project, students will combine visual and audio cues to enhance speech in multi-speaker environments. Students will use face recognition to generate spatial audio, improving speech comprehension and clarity in such scenarios.