# Research Statement

Vignesh Gokul

## 1 Research Overview

With the development of intelligent chatbots, humans have found a method to communicate with artificial digital assistants. However, human beings are able to communicate an enormous amount of information without ever saying a word. The expressions on our faces, the way we stand, and the gestures we use are all forms of non-verbal communication that can often express information better and faster than anything we say. In fact, non-verbal communication is such an important and integral part of conveying the true meaning of verbal communication that any interaction with a virtual person will not seem natural or human without it. We also communicate and express ourselves with other forms of expressions such as music, art and dance.

My research objective is to enable a non-verbal communication between humans and artificial agents, a problem I call co-creativity. For instance, in the problem of music improvisation, two expert musicians communicate with each other via their music. Similarly, we have other art forms of expression such as dance. Can we simulate such interactive non-verbal communication amongst humans and learning systems? My work also spans to deploying such creative agents for real-time use cases. I am very fortunate to have successfully collaborated with multiple research groups in Europe, industry and startups to contribute more in this research space.

### 1.1 Co-Creativity

Enabling creativity in artificial agents is a challenging problem. Let us consider the problem of music improvisation. Given an input musical signal, the system has to respond by generating "creative" music that is also in musical sync with the input. How do we evaluate whether a particular musical output is creative? How can we verify if the output is also in musical sync with the given musical signal? These are important questions that are still an active part of research. Moreover, we need a powerful generative model to synthesize good musical outputs. In response, I developed two works: A generative model based on graph based data structures and an entropy based measure to evaluate the correctness of any generative model.

Music/video content creators usually have large datasets that showcase their talent and reflects their unique style of composition/dancing. However, in an improvisation technique, there is a lack of paired supervised data, i.e given one dance move from user A, how would user B react to the same dance move? To solve this, I developed a framework called Variable Markov Oracle (VMO) that can learn mappings between two styles of data in an unsupervised manner. Assuming we have large datasets of unpaired data of both user A and user B, I use a pre-trained model to get strong representations of the data. VMO then constructs a graph where an edge represents that two nodes are similar in embeddings based on a threshold. Now traversing the graph using these edges can give new improvisations of how the would respond. This simple technique works for both audio (music) and video data and has been utilized in a variety of live performances and demos in conferences. Currently, I am working on extending this framework with large language models (LLMs).

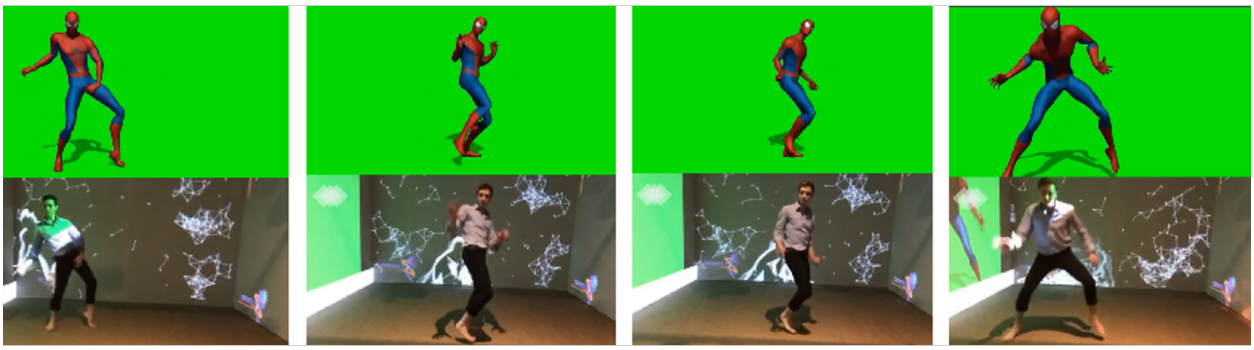

Figure 1: Human and agent showcasing co-creativity. The agent "spiderman" expresses itself via dance in response to changes in human motion.

How do we evaluate such models? As shown in figure 2, given an input musical signal, different generative models such as VMO can output a variety of cohesive musical signals. But how do we check the correctness of the output? This is a tricky question, as in a creative space such as dance or music, there is no one "right" answer. I developed an evaluative measure called SymTE to quantitatively measure how "good" an output is compared to the input. SymTE computes the transfer entropy between representations of the generated(output) and input control signals and provides a quantitative measure of compatibility.

## 1.2 Distributed Co-Creativity

Even if we find ways to solve the problem of co-creativity, there are many challenges in using the models in real-life applications: 1) Lack of data: generative models need massive amounts of data to train. Often in niche problems like dance or music improvisation, it is hard to acquire a clean annotated paired dataset. 2) Complexity of models: we need to reduce both training and inference time of models to efficiently train and deploy large scale models 3) Privacy of data: typical datasets would include private data from content creators that should not be leaked. An example real-life scenario would be a case where we have content creators distributed across the globe. Each creator has access to their own private, small dataset and want to train models on them (either discriminative/generative models). How can we facilitate such training? I introduced a framework for training models using distributed/federated learning and a means of protecting the privacy of the users' datasets when training. I created Differentially Private Distributed-InfoGAN, a framework to train generative models where multiple parties share their data together to train a stronger model. Moreover the privacy of each user's data is protected and cannot be inferred when the gradient information is compromised during network transmissions. Federated/Distributed of very large models can be quite tricky because of the complexity. The communication cost of sending/receiving model parameters drastically increases with the number of users participating in the training and the model size. My research aims to alleviate this by reducing the workload in terms of communication cost of the training. I developed FedLTN, a federated learning framework based on the Lottery Ticket Hypothesis for users to learn personalized lower complexity models that reduces the communication cost of training.

## 1.3 Ethical AI

I understand the ethical risks of my research and want to work on mitigating some of the harmful side-effects of my work. In federated learning, each user's data is kept private and not seen by other users. Only

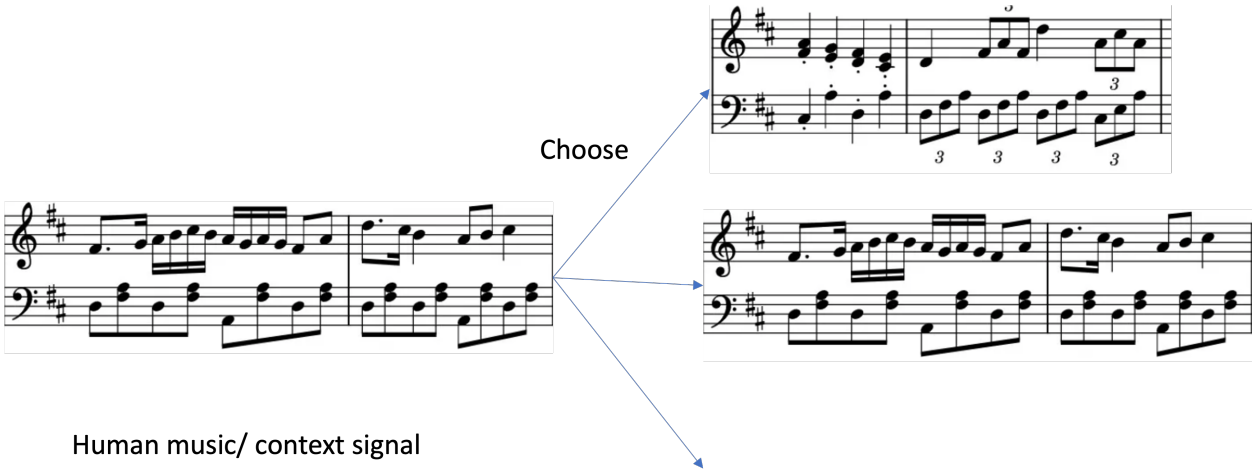


Figure 2: In the creative process of music improvisation, there exists multiple "right" answers. How do we choose?

the shared model is used by all the participants. During my experiments in training federated models for generating facial images, I found a concerning result, almost 98% of the images generated were that of white people. On further investigation, the reason behind this was that all the users had the same biases unknown to other users. For example, each user had only majority of white people images in the dataset thinking other users might have more data, which was not the case. This effect of propagating all the biases of the users' data can have dangerous effects in real-life applications. I proposed a bias-free federated learning paradigm to alleviate these issues.

Another important issue in AI research is acquiring data. Recently many research labs and organizations have solved this problem by scraping data from the web. However, this raises significant privacy concerns on the potential misuse of personal data for model training without authorisation. Can we protect our data from being used from model training? I created a framework PosCUDA by which we can make any audio dataset "unlearnable". Unlearnability refers to the fact that anyone can train any learning model architectures (CNN, RNN, Transformers) and can also use any feature processing methods, but the model would not be able to learn any useful information from the data, due to certain characteristics of the data.

## 2 Future Research Agenda

I would like to continue my research on co-creativity and distributed co-creativity. With the recent advances of Large Language Models (LLMs), I would like to collect massive datasets of human computer interactions either in the visual domain or in the audio domain.

Although my existing research is not focused on computing education, I would like to extend the same concept of "co-creativity" to the education space. With the advent of new tools such as ChatGPT, I envision a future where learning could happen along with these AI frameworks. More specifically, I would like to do more research on how AI can assist students in their learning and the potential of AI to accelerate students' learning.