**Report on Author Identification Using Chaos Game Representation and Deep Learning**

**Aleksandr Pitkin 333879013**

**Ilya Vorotyntsev 324775735**

**Link to Google Disk with materials:**

[Press on Me 🖱️](https://drive.google.com/drive/folders/11JY-d5g0fVaWp4km1SjwBIWvyh5Xj9wc?usp=sharing)

Introduction

Our team, passionate about the potential of machine learning, received a project to see if a computer could identify authors from their writing. Our task was to dive into deep learning, and to understand technique which teaches computers to identification text author.

Data Collection

We started by finding the right data. The popular IMDb62 dataset was easy to find. The CCAT10 dataset was harder; after a month of searching through various academic sources, we found CCAT50 and tailored it down to CCAT10, making it fit for our project.

First Attempts

With our datasets in hand, we tried to run existing code from GitHub. We first worked with the text data of CCAT10 but hit a roadblock—it wasn't the right format. We needed images, not just text.

Creating Images

We learned that we needed CGR images—pictures made from text. We found new code in a .nb file that could create these images. Running this file through Wolfram Alpha, after much trial and error, we finally got our CGR images from the CCAT10 text data.

Improving Our Data

Back to GitHub, we followed the lead of past researchers to make more images. We built a GAN, which created new, unique images. This helped us expand our dataset with additional images for the computer to learn from.

Building and Testing the Model

The focus shifted to the neural network. We carefully constructed the model based on the research, then compiled, trained, and tested it. This took time, adjustments, and learning. Eventually, we reached a 77% accuracy rate in identifying authors. The performance and learning process of the model were captured in various graphs.

Understanding the Process

We spent time understanding how the model worked—how each part of the code contributed to the goal of author identification. This deep understanding helped us see how each layer of the neural network played a role in the outcome.

Presentation of Findings

We wrapped up our work in a presentation that outlined our research process, the questions we addressed, and the insights we gained.