

# 580 Artificial Intelligence - Individual AI Talk Proposal

**TOPIC:** UNSUPERVISED TRANSLATION OF PROGRAMMING LANGUAGES SNEHITH JONNAIKODE

**PRESENTER:** SNEHITH JONNAIKODE

**DATE:** 11/22/2022 TUESDAY

## **PROPOSAL:**

Migrating a codebase from an old programming language such as COBOL to a modern alternative like Java or C++ is a difficult, resource-intensive task that requires expertise in both the source and target languages. COBOL, for example, is still widely used today in mainframe systems around the world, so companies, governments, and others often must choose whether to manually translate their code bases or commit to maintaining code written in a language that dates to the 1950s.

Transcompilers, with handwritten rules to convert source syntax tree, are typically used to solve such problems. However, writing these tools is very expensive. AI researchers at Facebook proposed an entirely self-supervised neural system [1] that can make code migration far easier and more efficient. Their model is built on top of recent advances in neural machine translation (NMT) models. They built a sequence-to-sequence (seq2seq) model with attention, composed of an encoder and a decoder with a transformer architecture and trained it on opensource GitHub codebase.

In my talk, I will briefly debrief the manual transcompiler tools first, and then head on to the model used in the paper. Facebook has also made the pre-trained model weights publicly available [2]. I also will try to play around with the model and test it with certain edge cases in Java to Python translation and present the results to the class.

## **REFERENCES:**

1. Marie-Anne Lachaux, Baptiste Roziere, Lowik Chanasot, and Guillaume Lample. 2020. [Unsupervised translation of programming languages](#).
2. Facebook Research pre-trained model. 2021. <https://github.com/facebookresearch/CodeGen>