STEVEN DING,

Transforming Generic Coder LLMs to Effective Binary Code Embedding Models for Similarity Detection

Compiled programs appear as opaque binary instructions, yet many security tasks require deciding when two such binaries implement the same underlying computation. In this talk, I'll show how large language models can vectorize binary code in a way that captures this hidden structure: similar programs map to nearby points in embedding space, even when the binaries have been transformed by optimization, architecture changes, or obfuscation.

We introduce several simple training strategies—data augmentation, translation-style learning, improved embedding extraction, and a cumulative metric-learning loss—that greatly strengthen these representations. The result is a general-purpose model that learns stable, invariant embeddings of program behavior and outperforms specialized tools for binary similarity.