Cook County's Justice Advisory Council (JAC) partnered with the University of Chicago Data Science Clinic to improve how Illinois criminal charges are represented and analyzed. The JAC advances criminal justice reform to create safer, flourishing communities. JAC and practitioners, including judges and researchers, need accurate offense information to understand crime trends, charging patterns, and effects of reforms. The primary reference, the Administrative Office of the Illinois Courts' Offense Code Table (OCT), contains shortened descriptions that lack sufficient legal context.

This project enriched the OCT by linking offenses to their full statute text from the Illinois Compiled Statutes (ILCS), generating plain-language summaries, and creating meaningful offense groupings. The team web-scraped and validated ILCS text, matched statutes to OCT entries using regular expressions, and merged the datasets. They then produced summaries using large language models (LLMs) and used text embeddings to generate initial semantic clusters, refining them through principal component analysis (PCA) to visualize relationships among offenses.

The enriched OCT now includes matched ILCS text, LLM summaries of crimes, and cluster labels, providing a stronger foundation for data analysis. Future work includes incorporating National Incident-Based Reporting System (NIBRS) crime categories, imputing missing categories using embeddings, and evaluating alignment with the embedding-based clusters.

Full Statute Text	OCT Offense	Chapter	Act	Section	Subsection Summary
(a) A person commits theft when he or she knowingly: (1) obtains or exerts unauthorized control over property of the owner; or (2) obtains by deception control over property of the owner; or (3) obtains by threat control over property of the owner; or (5) obtains or exerts control over property in the custody of any law enforcement agency which any law enforcement officer or any individual acting on behalf of a law enforcement agency explicitly represents to the person as being stolen or represents to the person such circumstances as would reasonably induce the person to believe that the property was stolen.	THEFT/BY THREAT/> 100K	720	50	16-1(a)(3)	By threatening the owner, the offender obtains control over the owner's property.

Figure 1: Example of an enriched OCT entry. The ILCS linkage and LLM summary make the offense easier to interpret.
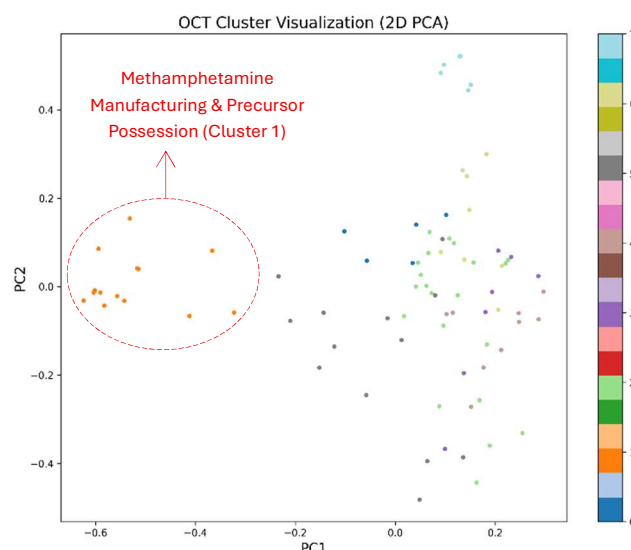

Figure 2: Two-dimensional PCA projection of enriched offense embeddings, revealing cluster structure that supports more interpretable and analytically useful groupings.