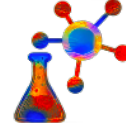


# Graph-Based Data Science: Hybrid AI meets data science process

TL;DR: *Integration of several disparate graph-based technologies available as Python open source, made to work efficiently with popular tooling for Data Science and Data Engineering practices, leading toward Hybrid AI solutions*



Python offers excellent libraries for working with graphs: semantic technologies, graph queries, interactive visualizations, graph algorithms, probabilistic graph inference, as well as embedding and other integrations with deep learning. However, most of these approaches share little common ground, nor do many of them integrate effectively with popular data science tools (pandas, scikit-learn, spaCy, PyTorch), nor efficiently with popular data engineering infrastructure such as Spark, RAPIDS, Ray, Parquet, fspect, etc. The library has use cases in large enterprise firms in industry and is also used as a teaching tool.

This talk reviews **kglab** – an open source project focused on integrating the priorities described above, and moreover providing ways to leverage disparate graph technologies in ways that complement each other, to produce Hybrid AI solutions for industry use cases. At its core, this effort is about self-supervised learning in graph-based data science workflows, leading toward Hybrid AI solutions.

keywords: *knowledge graph, data science, graph algorithms, probabilistic graph inference, deep learning, hardware accelerators, hybrid ai, artificial intelligence, design patterns*

curated public superset of slides: <https://derwen.ai/s/kcgh>

public repository: <https://github.com/DerwenAI/kglab>

DOI: **10.5281/zenodo.4602403**

