

# Bonan Yang

(618)-803-2588 | [yangbonan11@gmail.com](mailto:yangbonan11@gmail.com) | [LinkedIn](#) | [Scholar](#)

## Technical Skills

- **Specializations:** Graph Neural Networks (GCN, GraphSAGE, GAT, Graph Autoencoder, Spatio temporal GNNs), Knowledge Graphs, Complex Network Analysis, NLP & LLM, GraphRAG, Vector Databases, End-to-End ML Deployment
- **Languages:** Python, C/C++, R, SQL, Cypher
- **Databases & Storage:** Neo4j, Milvus, Pinecone, HDFS
- **Platforms & Tools:** Linux, Docker, AWS, GCP, Apache Spark, PyTorch, PyG (Pytorch Geometry), Transformers, vLLM/SGLang CUDA, Git, NetworkX, Gephi, yEd Graph Editor

## Education & School Activity

• <b>Southern Illinois University Carbondale</b> <i>Ph.D. (ABD) in Computer Science</i>	Carbondale, IL 2022 – Present
• <b>Southern Illinois University Edwardsville</b> <i>M.S. in Computer Science</i>	Edwardsville, IL 2019 – 2021
• <b>Shanghai Institute of Technology</b> <i>B.S. in Computer Science</i>	Shanghai, China 2014 – 2018

## Selected Projects

- **Large-scale Wikipedia Knowledge Graph** 2025 – Present  
*Built infrastructure for grounding LLM outputs in structured world knowledge.*
  - Processed Wikipedia, parsed 18.8M page nodes and 747M link edges, built large-scale knowledge graph with **HDFS**.
  - Implemented distributed graph processing with **Spark** for namespace filtering, redirect resolution, entity disambiguation.
  - Generated entity embeddings with **FastText** and indexed in **Milvus** for semantic similarity search.
  - Applied to concept alignment and knowledge path discovery, supporting downstream GraphRAG and QA systems.
- **Curriculum-Industry Skill Gap Analysis** 2025 – Present  
*Extending knowledge flow analysis from classroom to workplace through course-skill-job alignment.*
  - Built a knowledge graph (**neo4j**) integrating occupational skill ontology with STEM course syllabus text, enabling multi-source fusion of course-skill-job entities across 462 courses, 3556 skills and 770 jobs.
  - Applied semantic encoding for entity vectorization and automatically constructed cross-layer relationships via semantic similarity, supporting multi-hop reasoning and path queries for courses-skills-jobs alignment.
  - Designed graph feature engineering to compute skill coverage and job reachability metrics with explainable outputs
  - Applied the model to 5 regional public universities' educational data to quantify regional curriculum-industry skill gaps, providing data-driven insights for Illinois State Board of Education (ISBE)
- **Course Graph Visualization Platform** 2023 – Present  
*Structuring and visualizing knowledge dependencies in university course systems.*
  - Built full-stack web application: **Flask** + **Neo4j** backend with **RESTfulAPI**, **Cytoscape.js** frontend
  - Deployed to production with Docker containerization, Nginx reverse proxy, and SSL configuration
- **Course Graph Student Progression Analysis** 2023 – 2024  
*Tracking how students flow through knowledge structures and identifying where they get stuck.*
  - Constructed course prerequisite graph with NetworkX, integrating 14 years of transcript data for path analysis.
  - Applied Bayesian inference and survival analysis to model path completion probability and expected time-to-graduation.
  - Designed Markov-style grade transition matrix and graph centrality metrics to capture performance decay and locate systematic bottlenecks.
- **Self-supervised Network Qualification of 3D Objects for Personalized Manufacturing** 2024 – present  
*Enabling similarity retrieval and quality comparison of 3D printed objects via self-supervised graph learning.*
  - Transformed 3D printing point cloud data into graph structures to capture geometric topology of manufactured objects.
  - Designed **Graph Autoencoder** with contrastive learning ( Anchor-Positive-Negative) to learn object embeddings.
  - Built embedding database of printed objects, enabling similarity retrieval and deviation localization for new prints.
  - Applied to quality assessment and process parameter optimization in personalized manufacturing.
- **GNN-Based Molecular Toxicity Prediction System** 2025 – 2025  
*Building GNN pipeline from scratch for graph-level classification on molecular structures.*
  - Implemented and compared GCN and GAT architectures in **PyTorch** and **PyG** for multi-label toxicity prediction across 12 biological assays (Tox21 dataset, 7.8K compounds).
  - Parsed SMILES into molecular graphs with **RDKit**, extracting atom features (element, degree, charge, aromaticity) and building adjacency matrices from bond structures.
  - Implemented block-diagonal batching for variable-size graphs and masked loss for sparse labels.
  - Achieved 0.78 (GCN) and 0.81 (GAT) average ROC-AUC across 12 toxicity endpoints under severe class imbalance (1:22)

## Awards & Professional Activities

- Research Grants for Doctoral Students (RGRDS) Award, SIUE ([SIUE News](#))
- Outstanding Teaching Assistant Award, SIUE ([Illinois Business Journal](#))
- Conference Presentation, Complex Network and Application 2025 ([Collinsville Daily News](#))
- Conference Presentation, American Society for Engineering Education (ASEE) 2025 ([SIUE News](#))