Milim Lee

Redmond, WA, 98052 | 312.292.6129 | milimlee.3d@gmail.com
linkedin.com/in/milimlee | milimlee.com

TECHNICAL ARTIST | DATA ENGINEER

Technical Artist specializing in large-scale 3D synthetic data pipelines for machine learning and generative AI. Experienced in building and scaling preprocessing systems across 1M+ assets, improving geometry, materials, and lighting to produce high-quality, train-ready datasets that measurably enhance model performance. Brings 12+ years of cross-disciplinary experience spanning software engineering, real-time graphics, digital content creation, and ML/GenAI workflows. Combines deep expertise in DCC tools (Blender, Houdini, Unreal), Python-based automation, and data-centric pipeline design to deliver robust solutions for normalization, decomposition, and dataset quality optimization. Proven ability to translate research requirements into scalable production pipelines, collaborating closely with research scientists, ML engineers, and data teams to align data quality with model objectives. Core strengths include large-scale 3D dataset analysis, pipeline architecture, dataset curation, and distributed processing.

AREAS OF EXPERTISE

3D Synthetic Data Pipelines & Large-Scale Dataset Generation | Data Processing & Quality Optimization
Scalable Pipeline Architecture & Distributed Processing Systems | ML / Generative AI Data Workflows & Train-Ready Dataset Development | Research-to-Production Translation & Cross-Functional Collaboration

TECHNICAL SKILLS

Programming: Python, Java, C#, C, SQL, HTML
3D DCC(Digital Content Creation) Tools: Blender, Houdini, Unreal Engine
Rendering: Blender (Cycles), Unreal Engine (Lumen, Path Tracer)
Scripting & APIs: Python (Blender, Unreal, Houdini), Blueprints (Unreal), VEX (Houdini), HScript (Houdini)
Data & Pipelines: ETL workflows, distributed processing, metadata systems, schema design, validation
Cloud, Infra & Tools: AWS, Docker, OpenCue, Git, GitHub, Perforce, SVN, Visual Studio Code, PyCharm
File Formats & Pipelines: FBX, OBJ, glTF, USD, JSON, metadata
MR / VR: Meta Quest, Oculus Rift

PROFESSIONAL EXPERIENCE

Meta, Redmond, WA **June 2021 - March 2026**
Technical Artist (3D Data pipeline for ML/AI)
Built and scaled 3D synthetic data pipelines for machine learning and generative AI, processing **1M+ assets** and enabling high-quality, train-ready datasets across research and production teams.

- Designed and implemented 3D data preprocessing systems to normalize, validate, and prepare **1M+ assets** for ML training and evaluation
- Built automation frameworks using Python and DCC APIs (Blender, Unreal) to streamline normalization and synthetic data generation workflows
- Improved dataset usability by **54%** through large-scale filtering, recovery, and data quality enhancement systems
- Built geometry and material validation systems ensuring consistency and reliability for downstream ML tasks
- Developed physically based synthetic data pipelines (Blender + path tracing) that generated **4,500+ high-quality indoor scenes** for model training; originally developed for research and later adopted into production, helping establish lighting quality standards for downstream content and datasets.
- Operationalized data pipelines across teams, enabling consistent dataset production for cross-org research and model development
- Built procedural asset generation systems in Houdini with parameterized controls for scalable dataset creation and controlled variation

- Partnered with research scientists and ML engineers to translate ambiguous requirements into technical solutions; analyzed large-scale 3D datasets to identify patterns and drive data quality improvements
- Led cross-team collaboration, coordinating between internal technical art teams and external vendors to drive adoption of USD-based asset pipelines and real-time content ingestion workflows for scalable production

Production & Research Impact

- Contributed large-scale 3D dataset preprocessing for the launch of **Mesh Gen** and **Texture Gen** in **Meta Horizon Studio** (June 2025)
- Contributed 3D dataset curation and pipeline development for a **CVPR 2025** publication (see PUBLICATIONS & PROJECTS section)

Lucasfilm, ILM Immersive, San Francisco, CA **February 2019 - May 2021**
Technical Artist (Real-Time 3D Pipelines, VFX & Content Production)
Worked across real-time content creation and pipeline development for VR production, building VFX assets and scalable tooling to optimize 3D environments and automate production workflows.

- Created and optimized real-time VFX and environment assets, including scene-level optimization of VFX-heavy content, balancing visual fidelity with strict runtime performance constraints
- Built pipeline tools for environment mesh optimization, enabling large-scale assets to meet real-time performance budgets on target hardware
- Developed artist-facing tools for batch processing, debugging, and performance validation through interfaces
- Partnered with artists and engineers to translate creative goals into scalable, production-ready workflows

Production Impact

- Contributed to the delivery of multiple large-scale VR titles, including *Star Wars: Tales from the Galaxy's Edge* and the *Vader Immortal series*, on Meta Quest

CCC Intelligent Solutions, Chicago, IL **February 2012 - August 2016**
Software Engineer
Contributed to a multi-year enterprise modernization effort, migrating a 30+ year legacy COBOL-based vehicle valuation system to a scalable Java web services architecture.

- Developed end-to-end backend services spanning database design, API development, and integration with user-facing applications
- Translated complex business logic from COBOL to Java, ensuring functional parity through large-scale parallel validation and testing
- Authored technical design specifications based on business and functional requirements
- Built automated testing frameworks using JUnit, Cucumber, and jBehave to improve system reliability and maintainability
- Supported the full software development lifecycle, including integration, QA, production deployment, and on-call operations

PUBLICATIONS & PROJECTS

Objaverse Data Intelligence (Personal Project / Public Dataset)
Hugging Face: https://huggingface.co/datasets/milimlee-synth3d/objaverse.data.intelligence

- Designed and deployed a distributed AWS pipeline (Docker, t3.small instances — no GPU) processing **~680K** 3D scenes to extract geometry, material, and topology metadata at scale; architecture mirrors production ETL systems for large unstructured data
- Built a derived intelligence layer on top of Objaverse — per-scene and aggregated metadata tables enabling ML-ready filtering, scene understanding, and large-scale 3D curation
- Published dataset publicly on Hugging Face; **500+** downloads within 2 weeks of release

CVPR 2025 (IEEE/CVF Conference on Computer Vision and Pattern Recognition)
"LIRM: Large Inverse Rendering Model for Progressive Reconstruction of Shape, Materials and View-dependent Radiance Fields."
Li, Z., Wang, D., Chen, K., Lv, Z., Nguyen-Phuoc, T., Lee, M., Huang, J.-B., Xiao, L., Zhang, C., Zhu, Y., Marshall, C. S., Ren, Y., Newcombe, R., & Dong, Z.

- Contributed large-scale 3D dataset curation and preprocessing pipelines supporting model training and evaluation

EDUCATION

Master of Fine Arts (MFA) in Visual Effects, Savannah College of Art and Design, Savannah, GA

Master of Science (MS) in Information Technology, Northwestern University, Evanston, IL

Bachelor of Science (BS) in Computer Science, Purdue University Fort Wayne, Fort Wayne, IN

Bachelor of Engineering (BE) in Computer Engineering and Science, Seoul Women's University, Seoul, South Korea