Driving Manufacturing Best Practices Using Multimodal AI

By

Mark Zachary

B.S. Biomedical Engineering

University of Minnesota, 2021

Submitted to the MIT Sloan School of Management and
Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degrees of
Master of Business Administration
and
Master of Science in Electrical Engineering and Computer Science
in conjunction with the Leaders for Global Operations program
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
May 2025

Authored by: Mark Zachary
MIT Sloan School of Management
Department of Electrical Engineering and Computer Science
May 9th, 2025

Certified by: Duane Boning
Clarence J. LeBel Professor in Electrical Engineering and Computer Science
Thesis supervisor

Certified by: Stephen C. Graves
Abraham J. Siegel Professor of Management
Thesis supervisor

Accepted by: Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

Accepted by: Maura Herson
Assistant Dean, MBA Program
MIT Sloan School of Management

THIS PAGE IS INTENTIONALLY LEFT BLANK

Driving Manufacturing Best Practices Using Multimodal AI

by

Mark Zachary

MIT Sloan School of Management and
Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degrees of
Master of Business Administration
and
Master of Science in Electrical Engineering and Computer Science
in conjunction with the Leaders for Global Operations program

Abstract

Multimodal artificial intelligence offers promising solutions for enhancing operational excellence in contract manufacturing, where small job shops typically operate with limited standardization and high process variability. This research develops a part similarity tool that integrates geometric, material, and scale information to improve quoting accuracy and engineering efficiency in high-mix, low-volume production environments. After examining the fragmented manufacturing landscape and reviewing current AI applications in manufacturing, the study introduces an approach based on Variational Autoencoders for encoding 3D geometry alongside material properties and dimensional scale information. The technical implementation addresses challenges of multimodal fusion, missing data handling, and computational efficiency, while a qualitative ablation study demonstrates how this comprehensive approach outperforms single-modal methods in manufacturing relevance. Engineers benefit from improved insights for manufacturing planning, while estimators achieve more consistent cost predictions using the multimodal system. Reinforcement learning with human feedback provides a mechanism for continuous refinement, creating a framework that bridges geometric similarity with manufacturing context and reduces subjectivity in critical business processes. The research contributes both theoretical insights into multimodal learning and practical implementation strategies for standardizing operations in contract manufacturing environments.

Thesis supervisor: Duane Boning
Title: Clarence J. LeBel Professor in Electrical Engineering and Computer Science

Thesis supervisor: Stephen C. Graves
Title: Abraham J. Siegel Professor of Management

Acknowledgements

To Company X, thank you for embracing me into your world of manufacturing despite my unfamiliarity with the aerospace and defense industry. Your openness to new ideas and willingness to value the analytical experience I brought created an environment where innovation could flourish. The opportunities you provided have shaped my understanding of how small and medium enterprises can leverage advanced technologies to overcome unique challenges.

To Sean Holly, thank you for the countless hours of mentorship and the profound leadership lessons that extended far beyond manufacturing. I am grateful for your trust in including me in important discussions and decisions, and for those enlightening visits to other manufacturing facilities that broadened my perspective. Our time at George's house in Palos Verdes with Arabic music playing in the background remains one of my most cherished memories from this experience.

To Caitlin Braun, thank you for your remarkable responsiveness and steadfast support throughout this journey. Your ability to master complex manufacturing operations so quickly after transitioning from the medical device industry has been nothing short of inspirational. I've watched in admiration as you navigated challenging technical problems with grace and determination. Your enthusiasm for staying at the cutting edge of AI applications pushed me to deepen my research and explore new possibilities. Your mentorship has been invaluable, and I am genuinely grateful for your guidance.

To Professor Stephen Graves and Professor Duane Boning, thank you for your patient guidance and wisdom that helped shape this thesis from an ambitious concept into a focused contribution. Your thoughtful questions challenged me to think more critically about the intersection of theory and practice, while your extensive knowledge provided a foundation upon which to build.

This thesis journey has been both challenging and rewarding, made possible only through the generosity and support of everyone mentioned above as well as my friends and family. The lessons learned extend far beyond the technical aspects of manufacturing and artificial intelligence, touching on collaboration, leadership, and the value of diverse perspectives. For this holistic education, I am profoundly thankful.

THIS PAGE IS INTENTIONALLY LEFT BLANK

Table of Contents

List of Figures

Chapter 1

Introduction

Contract manufacturing in aerospace and defense has historically been characterized by its fragmented nature. Small to medium-sized enterprises, colloquially referred to as "mom-and-pop" shops, often operate with deeply specialized skill sets developed over years of experience but seldom backed by standardized processes. The result is a sector that struggles with inefficiencies across multiple operational domains, including quoting, engineering, and production. Particularly in job shop environments—where each project is treated as a unique endeavor—these inefficiencies become magnified, leading to high work-in-process (WIP), unbalanced workloads, and underutilized machinery. The problem extends beyond mere operational costs; it also impedes efforts to deliver consistently high-quality products on time. [1]

This thesis aims to address these inefficiencies by applying advanced multimodal AI techniques. Leveraging historical data, 3D geometric models, material properties, and part scale, the research goal is to develop robust AI models capable of learning part similarities. Successful implementation of such tools can standardize early-stage processes like quoting and engineering, thus reducing subjectivity, and variability. This not only promises to lower costs and shorten lead times but also enhances the overall ecosystem of contract manufacturing through better knowledge capture and operational excellence. Specifically, by identifying similarities among parts, manufacturers can more effectively improve quoting accuracy and consistency in engineering practices.

The following sections present the rationale, motivation, and strategic approach for employing AI-driven technology in contract manufacturing. Section 1.1 further defines the specific problem at hand; Section 1.2 details the broader motivation for this work in the context of both industry needs and academic research; and Section 1.3 provides an overview of the thesis goals and the structured plan to achieve them.

1.1 Problem Statement

The aerospace and defense manufacturing sector presents unique challenges arising from the highly specialized and varied nature of its components. While A&D serves as a prominent example, similar complexities occur across other low-volume, high-mix industries such as medical device manufacturing and custom automotive part fabrication. In these sectors, the quoting process heavily relies on the subjective expertise of long-tenured employees. As new complex parts are introduced—with unfamiliar geometries, tolerances, and materials—reliance on personal experience alone results in significant variations in quoting accuracy. These inaccuracies impact profitability, competitiveness, and the capability to secure and retain contracts. These variations negatively impact profitability, competitive positioning, and the ability to secure or retain contracts, further exacerbating the issue of inherently low anticipated production volumes.

Beyond quoting, the engineering process also suffers from a lack of standardization, especially when dealing with high-complexity parts. Even minor design changes can demand entirely new methods of manufacturing leading to repetitive engineering cycles that inflate lead times. These challenges are further complicated by limited data sharing across the organization, resulting in knowledge silos that prevent best practices from scaling effectively. The broader effect is that many shops continue to operate with outdated or ad hoc processes, reinforcing a vicious cycle of high work in progress (WIP), uncertain delivery times, and elevated operational costs.

1.2 Thesis Motivation

There is a pressing need to formalize operational best practices in contract manufacturing by harnessing data and digital tools. Recent developments in artificial intelligence—particularly in deep learning and multimodal representation—offer promising avenues for transforming how these shops quote and engineer parts.

1. **Operational Efficiency**

   By using AI to identify and cluster similar parts that have been manufactured by the firm, quoting teams can leverage historical data and built in knowledge to predict manufacturing times and costs with greater confidence. This data-driven approach reduces variability and the risk of misquoting, aligning customer expectations with actual production realities.

2. **Knowledge Sharing and Standardization**

   In many shops, specialized engineering knowledge is locked in the minds of a few experts. A robust AI-driven similarity model can help codify this expertise, making it accessible to less-experienced staff. As the industry grapples with an aging workforce, such tools become critical in capturing and transferring tacit knowledge.

3. **Competitive Advantage**

   With increased global competition, the ability to quote quickly and accurately is paramount. Companies capable of adopting AI-based methodologies may see a boost in both their market responsiveness and profitability. This competitive edge is further amplified by on-time delivery and improved quality control—two pivotal metrics for attracting high-value contracts.

From an academic standpoint, the intersection of 3D deep learning and manufacturing data presents a rich domain for exploring new techniques in multimodal learning. By bridging these disciplines, this thesis aims to advance the body of knowledge on how AI can be effectively deployed in small to medium-sized contract manufacturing environments—a context that has often been overlooked in mainstream Industry 4.0 research.

1.3 Thesis Goals and Approach

Building on the needs identified above, the central objective of this thesis is to develop and validate a multimodal AI model that predicts geometric and operational similarities between manufacturing components, thereby standardizing and streamlining both quoting and engineering processes in small-scale job shops. This objective can be broken down into the following goals:

1. Model Development
    - Expand on the Existing Part Similarity Tool: Enhance a variational autoencoder-based model initially trained on 3D geometry by incorporating additional data modalities such as material type and dimensional scale.
    - Improve Predictive Power: Evaluate different network architectures, data augmentation strategies, and loss functions to reduce model bias and improve overall prediction accuracy.
2. Model Validation and Evaluation
    - Real-World Pilot Testing: Deploy a prototype tool at Company X and track improvements in quoting accuracy and engineering lead times.
    - Human-in-the-Loop Reinforcement: Integrate user feedback from quoting and engineering teams to iteratively refine the model, thereby ensuring practicality and acceptance.
3. Dashboard and Stakeholder Engagement
    - Interactive Visualization: Create or enhance an existing dashboard to display model outputs in a clear and actionable format, facilitating widespread adoption and alignment across different levels of management.

- Change Management and Organizational Buy-In: Collaborate with management and shop-floor personnel to champion the tool's benefits, address concerns, and develop best practices for long-term sustainability.

The subsequent chapters of this thesis will discuss the background (Chapter 2) and relevant literature (Chapter 3), delve into technical aspects of 3D deep learning and multimodal modeling (Chapter 4), present the model validation framework (Chapter 5), and conclude with a discussion on broader applications and future directions (Chapter 6).

Chapter 2

Background

This chapter provides critical context for understanding the aerospace and defense (A&D) contract manufacturing landscape, with particular emphasis on the challenges faced by Tier 3/4 suppliers. We examine the industry structure, detail the contract manufacturing process flow, introduce Company X (where this research was conducted), and explore the technological developments in Industry 4.0 and artificial intelligence that are transforming traditional manufacturing operations. Throughout, we focus on the quoting and process steps of the manufacturing process—areas that directly benefit from the multimodal AI solutions presented later in this thesis.

2.1 Industry Background

The A&D supply chain spans multiple tiers, each specializing in different aspects of design, production, and assembly (Figure 1). At the apex, Original Equipment Manufacturers (OEMs) such as Boeing, Airbus, and Lockheed Martin handle the final integration and assembly of aircraft. Below them are Tier 1 suppliers (e.g., General Electric, Honeywell, United Technologies), which design and build large systems such as engines, avionics, and hydraulic assemblies for OEMs. These Tier 1 suppliers, in turn, source components from Tier 2 suppliers—entities like Parker and Eaton—that specialize in sub-systems and place a strong emphasis on performance, cost-effectiveness, and predictable delivery.

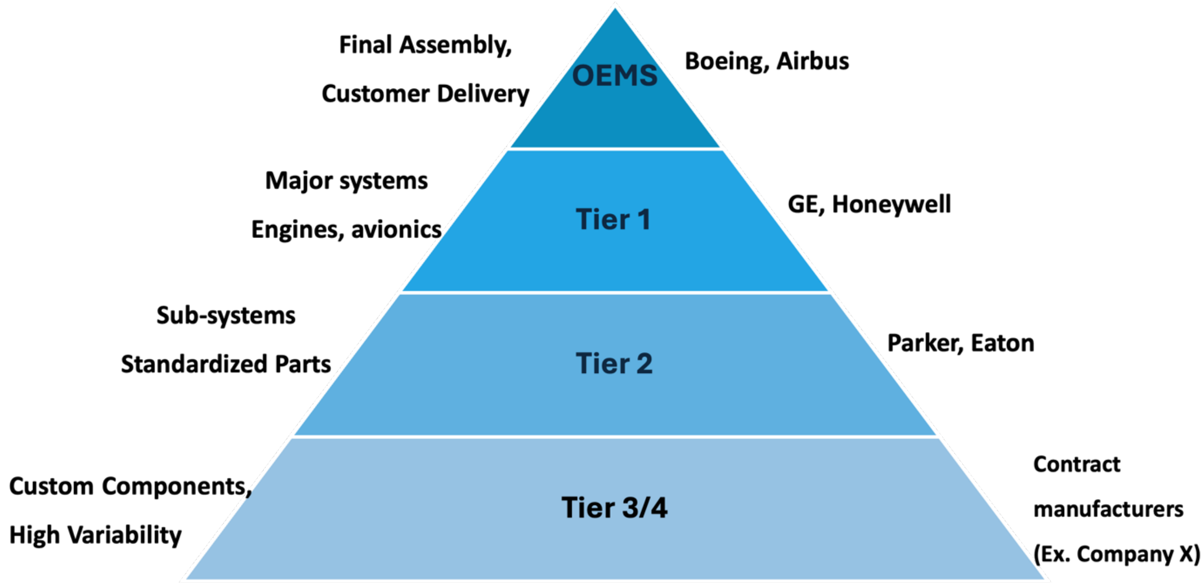

Figure 1: Aerospace and Defense (A&D) supply chain spans multiple tiers, each specializing in different aspects of design, production, and assembly.

Further downstream, Tier 3 and Tier 4 suppliers predominantly operate as contract manufacturers, often in the form of small job shops spread throughout the United States. With annual revenues typically well under $100 million, these shops have historically enjoyed some insulation from global competition due to the highly variable, low-volume, and classified nature of their work. However, this fragmented environment—coupled with the capital constraints many of these suppliers face—has often resulted in inconsistencies in quality, lengthy lead times, and unreliable delivery performance. OEMs and Tier 1 suppliers react to these uncertainties by ramping up inventories and building in excess buffer times into their quotes and engineering plans, leading to inefficiencies and higher overall costs across the supply chain [2].

Recent shocks—ranging from geopolitical conflicts and inflationary pressures to labor strikes and raw material shortages—have laid bare the fragility of the aerospace and defense supply network. Heightened tensions in Eastern Europe and sporadic export restrictions on critical metals (e.g., titanium, aluminum) have forced manufacturers to navigate sudden cost surges and unpredictable lead times [2]. Meanwhile, domestic labor disputes, coupled with a tightening labor market, have slowed production schedules and increased competition for skilled workers. Smaller suppliers with concentrated customer portfolios and limited contingency plans often find themselves unable to adapt, resulting in drastic restructuring or closure. By contrast, those that endure—generally companies with diversified client bases, robust operational processes, and a track record of reliable on-time performance—demonstrate that operational excellence is more than just a best practice; it is a vital strategic differentiator. As the industry continues to experience volatility in 2025 and beyond, contract manufacturers must increasingly adopt new methodologies and advanced technologies to remain resilient and competitive.

2.2 Contract Manufacturing Overview

The contract manufacturing process begins when a customer, typically a higher-tier supplier or OEM, submits a Request for Quote (RFQ). This initiates a complex workflow (Figure 2) that involves multiple functional departments and decision points before culminating in the delivery of finished components.

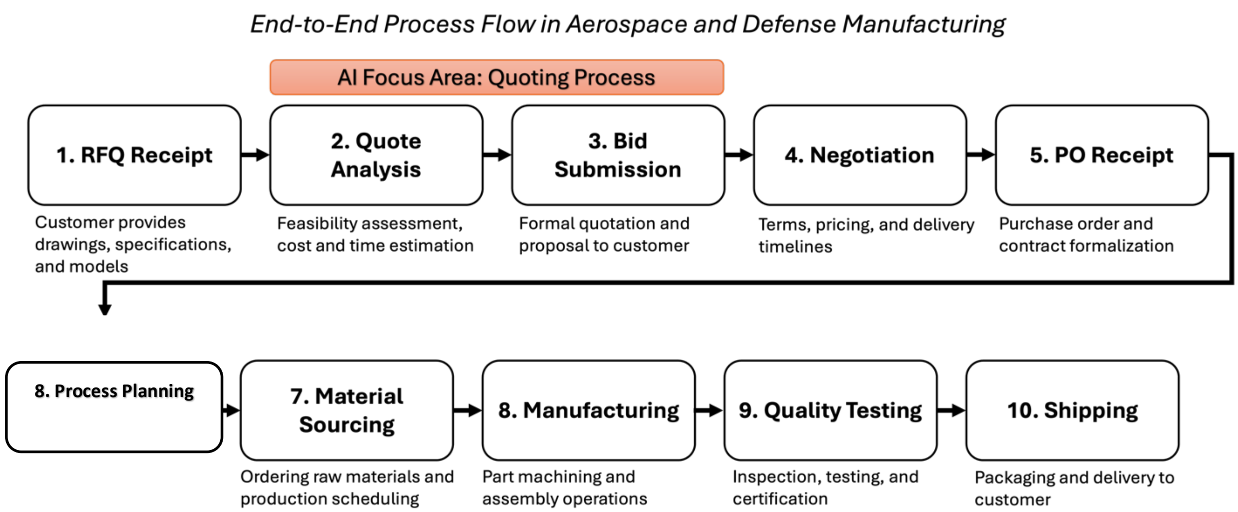

Figure 2: Contract Manufacturing Workflow

The following steps detail each phase of this process.

1. Request for Quote (RFQ) Receipt: The process begins when a customer approaches the contract manufacturer with an RFQ for a specific component or assembly. These requests vary significantly in format and completeness, but typically include:

- **Engineering Drawing Formats:** Customers provide drawings in diverse formats including paper blueprints, PDFs, AutoCAD (DWG/DXF), SolidWorks, CATIA, NX, and other CAD system exports (Figure 3). Each format presents different challenges in data extraction, with critical dimensions sometimes embedded in drawing notes rather than directly on the model.

- **3D Model Variations:** 3D models arrive in different formats:
    - STEP (Standard for the Exchange of Product Data) – also known as ISO 10303, is a widely used, standardized file format for 3D models and designs, facilitating easy sharing and editing across different CAD programs.
    - IGES (Initial Graphics Exchange Specification) – is a vendor-neutral file format used for exchanging 2D and 3D CAD data between different CAD systems, storing information like wireframes, surfaces, and solid models in an ASCII text format.
    - STL (Stereolithography) – is a file format used to represent 3D object surface geometry using a collection of triangles, commonly used for 3D printing and CAD

  Each format handles feature information differently—STEP files typically retain parametric features, while STL files only provide triangulated surfaces without feature intelligence. Estimators must often work with incomplete geometry that lacks manufacturing feature recognition.

- **GD&T Specification Inconsistencies:** Geometric dimensioning and tolerancing information appears in different locations and styles across customer documentation, sometimes following ASME Y14.5, ISO, or proprietary standards. Tolerance requirements may be scattered across notes, tables, or title blocks, requiring careful scrutiny.
- **Supplementary Documentation:** Material specifications, surface treatment requirements, and certification demands can be specified separately in documents like material condition reports, finishing requirements, or quality assurance provisions—each with customer-specific formatting and terminology.
- **Quantity and Delivery Specifications:** Production volume requirements range from one-time prototypes to ongoing programs with scheduled releases. These delivery expectations may be communicated in formal scheduling documents or simply embedded in email correspondence that accompanies the RFQ.

Aerospace and defense RFQs present additional challenges due to security and compliance requirements. ITAR (International Traffic in Arms Regulations) or EAR (Export Administration Regulations) controlled documents often cannot be processed through standard email channels, requiring secure data transfer protocols and restricted handling, which further complicates the information gathering process.

The heterogeneity of RFQ inputs forces estimators to develop expertise in "translating" between documentation styles—often manually extracting critical information and standardizing it internally before proceeding with quote development.

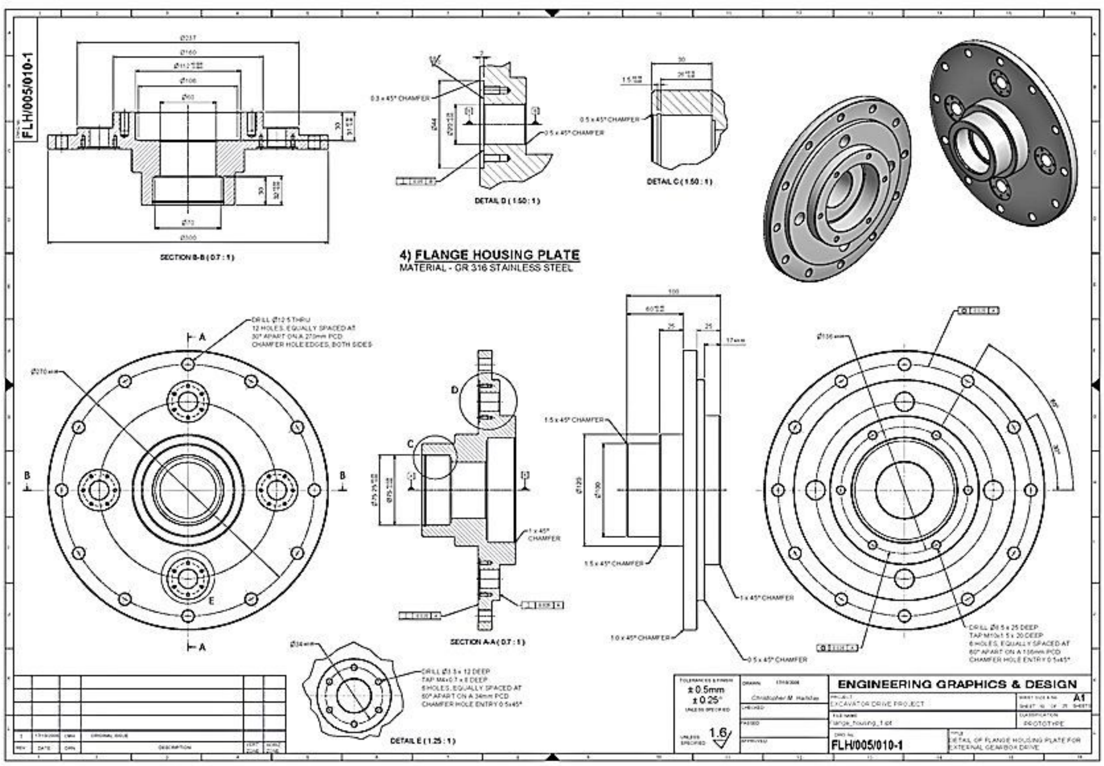

Figure 3: Example of an engineering drawing PDF. The basic components of an engineering drawing are the "Title block," "Dimensioning & Tolerances," "GD&T," "Block Diagrams and Schematics,"Detailed Views," "Parts & Components List," and "Assembly Diagrams." [3]

2. Quote Analysis and Development: Upon receipt of the RFQ, the contract manufacturer conducts a detailed technical and commercial analysis to determine feasibility and develop a comprehensive quote. This critical phase typically involves multiple stakeholders:

- Estimators evaluate part geometry, required tolerances, and machining complexity to predict setup and run times
- Engineers assess manufacturability, and determine process
- Procurement specialists research material costs and availability
- Production planners evaluate capacity constraints and scheduling implications

Historical knowledge plays a crucial role, as estimators and engineers draw on experience with similar parts to anticipate challenges, cycle times, and potential risks.

3. Bid Submission: Following internal analysis, the contract manufacturer prepares and submits a detailed proposal to the customer. This typically includes:

- Itemized pricing structure (material, labor, overhead, profit)
- Lead time estimates and production schedule
- Terms and conditions including payment schedule
- Clarifications or exceptions to the customer's requirements
- Value engineering proposals (if applicable)

Win rates for contract manufacturers in aerospace & defense average 20-30%, suggesting that quote preparation represents a significant operational cost with uncertain return on investment [4]. Manufacturers typically invest substantial engineering resources in quotes that may never convert to orders.

4. Negotiation and Contract Finalization: If the customer finds the quote compelling, a negotiation phase follows to finalize pricing, delivery schedules, and other contractual terms. For aerospace components, these discussions often include:

- Long-term agreements (LTAs) that secure volume commitments
- Material procurement responsibilities
- Quality requirements and inspection protocols
- Terms for potential design changes

Once these terms are agreed upon and the contract is awarded, the supplier performs a final contract review to verify alignment between the purchase order and the quoted assumptions. Any discrepancies are resolved with the customer before formal project kickoff.

5. Purchase Order Receipt: Once terms are agreed upon, the customer issues a formal purchase order (PO) that serves as the contractual basis for production. The manufacturer then creates internal work orders, assigns project numbers, and begins detailed planning for execution.

6. CAD/CAM Programming: During this phase, manufacturing engineers translate design specifications into a manufacturing process.

This process also heavily relies on tribal knowledge—experienced programmers apply techniques and approaches refined over years of practical experience, with limited standardization or knowledge capture.

7. Material Procurement: While the CAD/CAM programming are underway, procurement personnel begin sourcing the required raw materials and any standard hardware. A&D contracts often call for specific material grades or lot traceability, so the purchasing team must order from approved suppliers and allow lead time for material mill certifications. Lead times for aerospace-grade materials (special alloys, composites, etc.) can be substantial – recent industry data shows production material deliveries averaging 87 days in 2023, which is still higher than pre-pandemic norms [5]. This means procurement must anticipate needs well in advance to avoid delaying production. For long-lead or scarce items (for example, a particular forged billet or an electronic component for a defense assembly), buyers might place orders immediately upon contract award. They may also purchase extra to hedge against scrapped parts or future orders. In this stage, the team also orders any subcontracted services (like heat treating or non-destructive testing) and ensures those vendors meet the necessary certifications (for defense work, subcontractors might also need to be ITAR compliant or NADCAP accredited for special processes). Effective supply chain management is crucial here – any delay or error in material procurement can push out the overall project timeline. The outcome of Stage 7 is that all required materials and outside services are lined up to support the production schedule.

8. Manufacturing Operations: With programs complete and materials on hand, the manufacturing floor produces the parts. Machinists set up CNC machines, load the NC programs, and perform in-process inspections of critical dimensions. Complex parts may go through multiple operations—milling, turning, grinding, etc.—and the shop must follow the prescribed routing to ensure each feature is produced in the correct sequence. In many aerospace contracts, a first article part is fully inspected before the complete batch is run (per AS9102 guidelines). Assembly operations (if applicable) and additional processes (e.g., heat treat, surface finishing) are also performed here.

9. Quality Assurance and Testing: Aerospace and defense manufacturing demands extremely high quality and adherence to specifications. After production (and sometimes at intermediate steps), the QA team conducts thorough inspections. They use precision measurement tools—calipers, micrometers, coordinate measuring machines (CMMs)—to verify dimensions against the drawing tolerances, and they check surface finish, hardness, etc. as required. The QA team collects material and process certifications (e.g., plating certs, material certs) and compiles a formal inspection report or First Article Inspection (FAI). Because AS9100 and other regulations require robust documentation, all measurements and certifications must be traceable to the specific parts produced. If a non-conformance is identified, the parts may be scrapped or reworked, and a corrective action process is triggered.

10. Packaging and Shipment: In the final stage, the completed parts and their associated documentation are delivered to the customer. Proper packaging and labeling—sometimes along with shipping via ITAR-compliant carriers—is essential to protect and track the product. Once the customer receives the parts, the manufacturing contract is essentially fulfilled, although suppliers remain available for feedback or support. For successful projects, the cycle often repeats with new RFQs or follow-on orders, allowing the supplier to leverage refined workflows for greater efficiency and competitiveness.

Throughout this end-to-end process, the quoting and process development phase presents a significant opportunity for AI-driven improvement—precisely where the multimodal AI solution presented in this thesis focus.

2.3 Company Overview

Company X is a precision machining supplier occupying both Tier 2 and Tier 3 roles within the A&D industry. Established in 1979 by two brothers, the company has grown to over 100 employees and 30 machines, processing around 200 different custom components (jobs) at any given time. This growth has been driven by a steadfast commitment to quality—a hallmark that has consistently led to repeat business and deeper relationships with major defense and aerospace customers.

Following its acquisition by a private equity firm in 2019, Company X welcomed new executive leadership (CEO and CFO), who worked alongside the original owners. This leadership team set out to maintain the company's historic reputation for quality while pursuing aggressive growth targets and financial performance objectives. However, the influx of new customer orders also introduced increasing operational complexity, pushing existing processes and resources to their limits. In particular, rising customer demand, combined with pressure to diversify beyond a few key accounts, highlighted the need for improved on-time delivery and the ability to quote accurately and efficiently.

To meet these challenges, Company X initiated a variety of operational transformations focused on workflow redesign, capacity planning, and digital strategy. In this context, the company identified AI-driven solutions—such as part similarity modeling—as a potentially powerful means to streamline quoting and engineering processes, reduce lead times, and enhance resilience. This thesis research was conducted within Company X during this transformative period, focusing on developing and implementing a multimodal AI approach aimed at improving operational excellence.

2.4 Industry 4.0

Industry 4.0 is often presented as a convergence of cyber-physical systems, IoT-enabled devices, and advanced analytics in manufacturing. However, in the context of contract manufacturers—particularly in the aerospace and defense sectors—it is equally vital to highlight the data-rich environment that emerges from CAD /CAM processes. While conventional discussions of Industry 4.0 emphasize machine-level IoT sensors (e.g., capturing spindle speeds or vibration data), CAD data provides an earlier, design-focused vantage point into each part's geometry, features, and potential manufacturing path.

Crucially, these 3D geometries and part definitions can be archived and reused well beyond their original design intent, forming a digital asset that underpins analytics across quoting and manufacturing best practices. When combined with ERP systems that handle bills of materials, routing information, workforce allocations, and real-time production data, manufacturers obtain a comprehensive digital thread of each part's life cycle. This thread can support quoting and cost estimation (e.g., quickly deriving machining time from geometries), can enforce standardized manufacturing procedures, and can drive continuous improvement via data-driven insights.

Moreover, geometric part data offers additional opportunities for analytics that go beyond standard operational metrics—precisely the focus of this thesis. Storing 3D models in central repositories, along with metadata about materials and scale, opens the door to automated shape comparison, part similarity retrieval, and variant detection. As explored in the following chapters, these capabilities allow engineers and managers to identify functionally or geometrically related parts, leading to reduced design duplication, improved process planning, and refined cost predictions.

Within an Industry 4.0 framework, this design-driven perspective positions CAD environments as a principal source of geometric intelligence, and ERP systems as the operational backbone. Bringing them together creates an integrated ecosystem—one in which both real-time sensor data and rich geometric datasets fuel advanced analytics. By systematically merging design information with production and business data (ERP), contract manufacturers can leverage geometric insights—such as those presented in this thesis—to enhance quoting accuracy, optimize part reuse, and ultimately strengthen their responsiveness in a rapidly fluctuating market.

2.5 The AI Revolution

Artificial Intelligence (AI) has revolutionized numerous industries by offering capabilities that extend well beyond traditional statistical methods. From image recognition in healthcare to natural language processing in customer service, AI-driven insights and automation are changing how organizations operate. Within manufacturing, AI can analyze vast amounts of data—from machine logs to enterprise resource planning (ERP) records—to optimize processes, predict failures, and enhance product quality [6].

For contract manufacturers, AI appears to hold significant potential to transform core activities like quoting, scheduling, and quality control. In quoting processes, machine learning models could examine historical time studies, material costs, and geometric complexity to generate estimates with potentially greater accuracy than traditional methods. On the shop floor, AI-driven scheduling systems might dynamically allocate resources based on real-time data, which could reduce bottlenecks and improve throughput if successfully implemented. Furthermore, AI applications in defect detection and predictive maintenance represent promising avenues for improvement in areas that typically affect operational costs and delivery times. While these applications have been explored in large manufacturing enterprises, they remain largely untested and unevaluated in the context of small and medium-sized contract manufacturers, suggesting an opportunity to address persistent operational challenges in this underserved sector.

However, effectively deploying AI in smaller manufacturing environments requires grappling with several challenges: data quality is often poor or inconsistent, labeling processes can be time-consuming, and many job shops lack the necessary infrastructure (both technological and organizational) to embrace large-scale transformations. Nonetheless, ongoing advances in hardware and software—plus the rising availability of off-the-shelf AI solutions—are making it more feasible for smaller enterprises to experiment with and adopt these innovations.

2.6 Deep Learning

Within the broad field of AI, deep learning has emerged as a particularly powerful approach for tackling complex pattern-recognition tasks. Unlike classical machine learning methods, deep learning models can learn hierarchical representations of data, making them especially effective for high-dimensional inputs such as images, video, and 3D geometry.

Traditional convolutional neural networks (CNNs) have historically dominated computer vision applications by using sliding filters to detect features at multiple scales. These architectures process visual data through a series of convolutional layers that progressively extract higher-level features—from simple edges in early layers to complex shapes in deeper layers. When applied to 3D data (e.g., voxelized part models), CNNs require 3D convolutions that dramatically increase computational requirements but effectively capture volumetric patterns.

More recently, transformer-based architectures have revolutionized deep learning across multiple domains. Originally developed for natural language processing, transformers use self-attention mechanisms to weigh the importance of different input elements when constructing representations. In the 3D domain, models such as Point Transformer and 3D-DETR adapt these attention mechanisms to point clouds and voxel data, achieving state-of-the-art performance on shape classification and part segmentation tasks [7] [8].

In the context of contract manufacturing—especially in the precision machining field—deep learning holds promise in multiple domains:

1. **Part Similarity**

   Models trained on historical part geometries can group components by similarity, enabling more efficient quoting by automating the process of recognizing common features to drive manufacturing standardization.

2. **Defect Detection**

   Advanced image-based models can identify surface defects or dimensional anomalies far more consistently than manual inspections, which are prone to human error and fatigue.

3. **Predictive Maintenance**

   Recurrent or graph-based architectures can learn from machine sensor data to predict mechanical failures before they occur, minimizing downtime and associated costs.

Despite these advantages, applying deep learning in smaller shops comes with its own set of challenges: labeled data can be scarce, model interpretability may be limited, and retraining or updating models requires computational resources and expertise not always found in traditional manufacturing environments. However, as the technology matures and cloud-based services lower the barriers to entry, these hurdles are gradually becoming more manageable—particularly for forward-thinking job shops like Company X that seek to remain competitive through digital innovation.

Having established the background on the A&D industry, the specific context of Company X, and the digital transformation trends that shape modern manufacturing, we will next move to a detailed Literature Review. There, we will delve deeper into the existing body of research on data analytics in manufacturing, multimodal representation learning, and the role of generative and 3D deep learning techniques that form the foundation for this thesis.

Chapter 3

Literature Review

This chapter critically examines the evolving role of artificial intelligence and data-driven strategies in the context of contract manufacturing. First, it explores how small and mid-sized enterprises are embracing AI technologies for production optimization and discusses the unique challenges of integrating disparate data sources such as ERP systems, CAD files, and industry-specific standards. Next, it delves into current approaches to 3D deep learning, focusing on the merits and limitations of voxel-based methods, point clouds, multi-view projections, and graph representations—each holding particular value for different manufacturing tasks. Finally, the review highlights the emergence of multimodal learning in manufacturing, underscoring how researchers have begun unifying geometric, textual, and tabular data to create robust, end-to-end AI applications. By surveying these three major themes, this chapter offers a comprehensive view of existing research, setting the stage for subsequent chapters where the thesis proposes its own AI-driven solutions.

3.1 AI in Manufacturing – Trends and Infrastructure Challenges

The manufacturing sector has increasingly embraced artificial intelligence (AI) as part of the Industry 4.0 movement, leading to the rise of smart factories and smart manufacturing systems [9]. These AI-driven factories integrate cyber-physical systems, Internet of Things (IoT) sensors, and data analytics to enable real-time, autonomous decision-making on the shop floor. In practical terms, AI technologies are being applied to enhance production efficiency, quality control, and flexibility. For example, computer vision systems powered by deep learning now perform automated visual inspection for defect detection on assembly lines, significantly improving quality assurance in automotive manufacturing [10]. Predictive models can anticipate equipment failures for predictive maintenance, minimizing unplanned downtime and maintenance costs [11]. Likewise, AI optimization algorithms are used for process planning and scheduling, dynamically adjusting production schedules in response to real-time conditions [12]. Across multiple case studies, manufacturers have reported substantial gains – from reduced defect rates to improved throughput – by incorporating AI in these domains. Notably, predictive maintenance and process optimization rank among the top AI use cases in manufacturing, offering measurable ROI by extending machine life and optimizing resource use. These successes have driven a trend where over half of manufacturers plan to integrate AI-based quality control or maintenance solutions by the mid-2020s, underscoring the growing strategic importance of AI for competitiveness [13].

While large corporations can more readily adopt comprehensive “smart factory” solutions, small and medium-sized contract manufacturers often struggle with fragmented technology stacks, legacy equipment, and limited IT resources. Multiple studies suggest that data integration is one of the key roadblocks to AI deployment in manufacturing SMEs (small- to medium-sized enterprises) [14] [15]. In these settings, disparate data formats—ranging from enterprise resource planning (ERP) logs and scheduling records to design files—are rarely harmonized in a central system. ERP data often resides in proprietary databases or on-premise systems designed primarily for financial transactions and order management, while CAD/CAM tools generate geometry-rich files (STEP, STL) whose metadata is not trivially parsed. The result is a suite of disconnected “information islands,” making it cumbersome to train or deploy AI models that require consistent, high-quality datasets.

Several authors address these challenges by proposing either data pipelines that extract relevant parameters from CAD/CAM projects into a more standardized, machine-readable form, or domain-specific data lakes that unify ERP, shop-floor sensors, and design documentation for downstream analytics [16]. Even so, the data capture process can be complicated by older machinery that cannot automatically log process data, pushing SMEs toward manual data-entry or retrofitted sensor solutions. Moreover, aerospace and defense (A&D) manufacturers must navigate additional layers of compliance—such as ITAR restrictions—which limit how (and where) design files can be stored, complicating multi-tenant cloud-based analytics. Consequently, bridging these infrastructure and data challenges remains essential for any meaningful AI-driven transformation in contract manufacturing.

3.2 3D Deep Learning Methods for Manufacturing Tasks

Many manufacturing applications of AI involve 3D data, since products and components are inherently three-dimensional. Whether it is recognizing a mechanical part, inspecting a 3D scan of a manufactured item for defects, or retrieving similar designs from a database, the question arises of how to represent and learn from 3D geometry in a form that neural networks can process. In recent years, the field of 3D deep learning has produced a variety of approaches to represent 3D shapes for input to machine learning models. The major representations include voxel grids, point clouds, multi-view projections, and graph-based meshes. Each representation has strengths and limitations, and researchers have evaluated their suitability for different manufacturing-specific tasks such as part classification, metrology inspection, and defect detection.

3.2.1 Point Clouds

A point cloud represents a shape as an unstructured set of sample points in $\mathbb{R}^3$, typically the surface points of the object (Figure 4). Point clouds are a very natural output of 3D scanning technologies (laser scanners, structured light scanners, LIDAR, etc.), so they often arise in inspection and reverse-engineering tasks. However, point clouds lack any inherent connectivity information – they are simply a cloud of points. Traditional grid convolutional neural networks (CNNs) cannot be directly applied to point sets, but groundbreaking work by Qi *et al.* introduced **PointNet** and **PointNet++**, which are neural network architectures specifically designed to consume point clouds directly [17] [18]. These networks use symmetric functions (like pooling) to achieve permutation invariance (so the order of points does not matter) and can capture local geometric features (PointNet++ builds neighborhood hierarchies, akin to CNN filters on the point set).

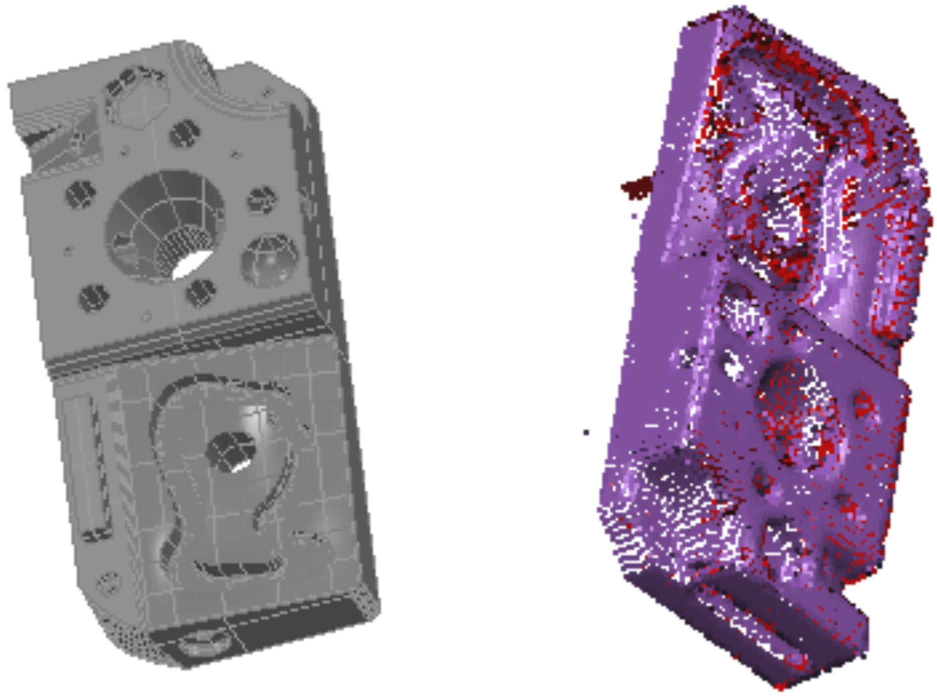

Figure 4: Illustration of a manufacturing component CAD model with its corresponding point cloud representation [19]

For manufacturing, point cloud-based deep learning has proven especially useful in quality inspection and defect detection. For example, detecting geometric deformities in a produced part can be done by scanning it into a point cloud and using a PointNet++ derived model to identify out-of-spec regions [20]. A recent review by Liu *et al.* (2025) highlights that point cloud deep learning methods have matured to meet many real-time industrial requirements, even for large-scale point sets [21]. The strengths of point cloud representation include flexibility (points can sample arbitrary geometry) and efficiency – one can adjust the density of points as needed, focusing high density in areas of interest.

However, challenges remain: point clouds from real sensors can be noisy or incomplete (with occlusions causing missing data), which requires the AI models to be robust to holes and outliers. In manufacturing metrology, aligning a point cloud of a produced part to the original CAD model for comparison is itself a non-trivial problem; once aligned, learning algorithms can flag deviations. Another challenge is scaling to millions of points if a very dense scan is taken – subsampling or hierarchical processing becomes necessary [22]. Despite these issues, point cloud learning is currently one of the most popular and successful 3D AI approaches in industrial applications due to its alignment with how 3D data is actually captured on the factory floor

3.2.2 Voxel Grids

A straightforward way to feed a 3D shape to a neural network is to discretize the 3D volume into a 3D grid of voxels (volume pixels). Each voxel is like a 3D pixel that can be marked as filled (inside the object) or empty (outside the object), or even contain values like distances or material properties. This regular grid structure allows the use of 3D CNNs much like 2D CNNs on images. Early pioneering work like VoxNet demonstrated real-time object recognition using voxelized representations [23]. In manufacturing, a voxel approach could be applied to tasks like classifying objects by shape or comparing a scanned part to its CAD model by voxel-by-voxel overlap. The advantage of voxels is their simplicity and compatibility with existing CNN architectures.

However, the resolution vs. memory trade-off is a significant challenge – capturing fine details requires a dense voxel grid (e.g. 256^3 voxels or more), which leads to high memory and computational costs [24] [25]. This makes naive voxel approaches less practical for large or detailed CAD models. For instance, a small defect like a scratch or a 0.5 mm misalignment might not alter many voxels in a coarse grid, potentially escaping detection. Researchers have developed techniques like octree-based CNNs (which adaptively refine voxels where needed) to mitigate this [26], but voxel methods still struggle with precision requirements in many manufacturing contexts. Additionally, voxel models tend to be invariant to object surface semantics – they only see occupied vs. empty space – which can be a drawback if distinguishing feature types (holes vs. bosses, etc.) is important. Despite these issues, voxel CNNs have been successfully used in some manufacturing studies (e.g., 3D convolutional models for detecting internal defects in castings via volumetric scans). Overall, voxels offer a generic but brute-force approach to 3D learning.

3.2.3 Multi-View 2D Projections

An alternative to directly processing 3D data is the multi-view approach, which leverages proven 2D image-based deep learning. The idea is to render or capture a series of 2D images of the 3D object from different viewpoints, then feed those images into a conventional CNN (or a fusion network) to glean 3D understanding. This approach was exemplified by Su *et al.* in their multi-view CNN for 3D shape recognition, which showed that aggregating features from multiple 2D views gave excellent performance on shape classification benchmarks [27]. For manufacturing tasks, multi-view methods are intuitive whenever multiple cameras inspect a part from different angles (common in automated optical inspection systems) [28] [29] [30]. For instance, an electronic part on a conveyor might be simultaneously imaged from the top and sides; a multi-view CNN can fuse these perspectives to decide if the part is correctly assembled. The multi-view approach benefits from the rich pretrained models available for 2D vision – one can use ImageNet-pretrained backbones for each view, for example – and from the fact that many defects (scratches, dents, misalignments) are visible on surfaces that a camera can capture [31]. It also tends to capture fine texture or appearance details better than raw 3D representations.

On the downside, multi-view methods can miss occluded or internal features (since the set of views might not see every nook of a complex shape), and the data processing scales with number of views per object (e.g., 12 images per part means 12 forward passes through a CNN backbone, although some architectures merge views earlier to reduce cost). Another challenge is the alignment of views – ensuring the network learns to register that, say, one view's left side corresponds to another view's right side of the same object. Still, multi-view networks have achieved state-of-the-art performance in shape retrieval tasks. In a manufacturing context, Mazzetto *et al.* used multiple camera views to automatically classify objects in an automotive assembly line, reporting high accuracy in identifying different part types on a fast-moving line [10]. Multi-view approaches are especially appealing when image data is readily available or when the model must integrate visual appearance cues (e.g., surface finish or color) with geometric form. Figure 5 conceptually shows a part being viewed from multiple angles for CNN analysis. As computing power grows, multi-view and point-cloud approaches are also being combined (e.g., a network might use both 3D point data and 2D images to improve robustness).

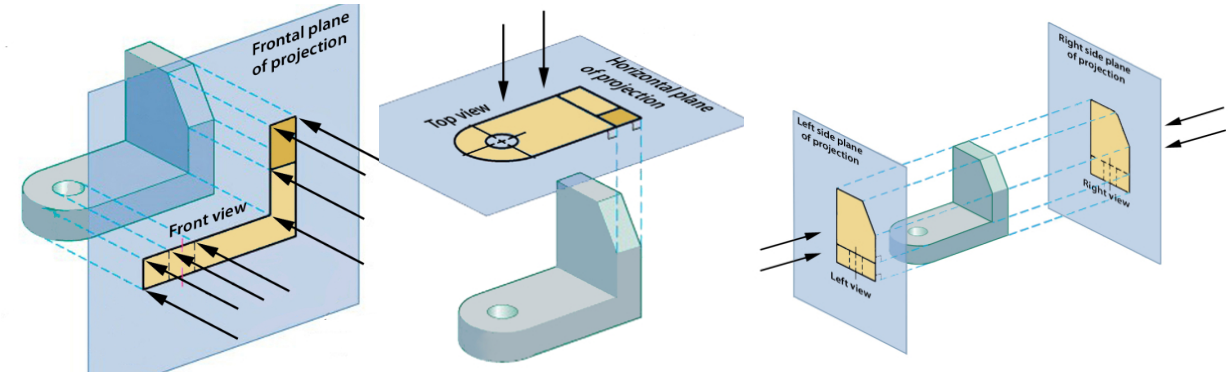

Figure 5: Multi-View 2D Projections of a manufacturing component [32]

3.2.4 Graph-Based and Mesh Representations

A more sophisticated class of 3D representation leverages the fact that CAD models often come in the form of Boundary-Representation (B-Rep) models or meshes, which explicitly encode the topology of faces, edges, and vertices of the object. In a B-Rep (common in solid CAD formats like STEP or Parasolid), the model consists of surfaces (like planes or cylinders) with explicit connectivity information (which edges bound a face, which faces meet at an edge, etc.). This structure can be naturally represented as a graph: faces as nodes, shared edges or vertices as connections. Similarly, a triangulated mesh can be seen as a graph of vertices connected by edges in a certain pattern. Graph neural networks (GNNs) have emerged as a powerful tool to handle such data. Recent work by Lambourne *et al.* introduced *BRepNet*, a message-passing neural network tailored to B-Rep data structures, showing superior accuracy in segmenting CAD model features compared to point-based or mesh-based methods [33]. The advantage of graph-based representation for manufacturing is the preservation of exact geometry and semantic structure. Unlike a point cloud or voxel grid, a B-Rep graph knows which surfaces are fillets, which edges are part of a hole, etc., if properly labeled. This makes it ideal for tasks like identifying machining features (pockets, holes, slots) on a part – an important problem in automated process planning. Indeed, several studies have applied GNNs to classify or segment CAD features for CAM purposes

[34] [35]. Graph representations are also compact – a complex engine block B-Rep might consist of a few thousand faces, far fewer data points than a dense point cloud of the same part.

However, the challenges are significant: constructing and inputting graphs requires more pre-processing (one must have the CAD model in B-Rep form, which may not be available if only a mesh or scan is given). Also, training data with ground-truth annotations (e.g. which faces correspond to which feature) is scarce, though efforts like the Fusion 360 Gallery dataset (with 35,000 CAD models) have started to alleviate this [36]. Graph neural nets can also be slower per sample and trickier to implement due to non-uniform graph sizes. Nonetheless, for manufacturing design automation, graph-based learning is very promising – for example, a graph network could learn to predict if a design is manufacturable by analyzing the configuration of features and flagging those violating design rules. Another related approach is to use mesh convolutional networks (e.g., spiral or spectral CNNs on meshes), which treat the mesh like an image defined on a manifold. These have seen use in 3D shape analysis but are slightly less popular in manufacturing contexts than B-Rep graphs, since most engineering CAD is B-Rep rather than raw mesh. In summary, graph-based 3D learning methods excel at tasks requiring understanding of topology and precise geometry, while point and voxel methods excel at tasks requiring geometry processing at scale (inspection, scanning) and multi-view excels when appearance or easy integration with vision is needed. Researchers often choose the representation that best matches the data available for the task at hand.

3.3 Multimodal Learning in Manufacturing

Modern manufacturing operations generate diverse data types across the product lifecycle—from 3D CAD models and engineering specifications to bills of materials and sensor readings. While traditional AI approaches have treated these data sources in isolation, recent research has increasingly focused on multimodal learning techniques that can integrate and reason across these heterogeneous information streams. This section examines the emerging field of multimodal AI in manufacturing contexts, with particular emphasis on recent research developments that combine geometric data with textual and tabular information to create more comprehensive manufacturing intelligence systems.

3.3.1 Cross-Modal Alignment of CAD Models and Text

One of the most active research areas in multimodal manufacturing applications involves establishing meaningful connections between 3D shapes and natural language descriptions. By learning joint embedding spaces for geometry and text, researchers have developed systems capable of cross-modal tasks such as text-to-shape retrieval and shape captioning. Early pioneering work by Chen et al. (2018) introduced the Text2Shape framework, which established a joint embedding space for 3D models and textual descriptions by training on paired data from the ShapeNet dataset [37]. This approach used voxel-based convolutional networks to encode 3D geometry while processing text descriptions through recurrent neural networks, creating a shared latent space where similar concepts aligned regardless of input modality.

Han et al. (2019) advanced this field with their View2Seq-to-Seq (Y2Seq2Seq) architecture, which improved cross-modal alignment by processing sequences of rendered views alongside word sequences from textual descriptions [38]. This approach demonstrated superior retrieval performance by capturing view-dependent information that enhanced the system's understanding of 3D geometry from multiple perspectives.

More recent work has pushed toward increasingly fine-grained understanding of the relationship between geometry and language. Tang et al. (2023) developed Parts2Words, a system that establishes detailed correspondences between specific geometric features and their textual descriptions [39]. This granular approach enables matching individual phrases like "cylindrical hole" to the corresponding regions in 3D models—particularly valuable in manufacturing contexts where technical documentation frequently references specific features using precise terminology.

Ruan et al. (2022) extended these principles further by introducing TriCoLo, a trimodal contrastive learning approach that incorporates images alongside 3D models and text [40]. This innovation creates a unified representation space spanning all three modalities, enabling more flexible queries and retrieval across visual, geometric, and textual information. Such capabilities are particularly valuable in manufacturing environments where components may be documented through a combination of 3D models, photographs, and written specifications.

While early multimodal research primarily demonstrated concepts using consumer product datasets (furniture, household objects), recent work has increasingly focused on industrial applications. The alignment techniques developed in academic research are gradually finding application in practical manufacturing contexts—creating systems that can, for instance, automatically associate O-ring grooves in a CAD model with relevant sections of a material specification document discussing sealing elements [41].

Despite these advances, data availability remains a primary obstacle for industrial applications—while academic datasets like ShapeNet provide paired geometry-text data for consumer products, equivalent resources for industrial components are scarce. Manufacturing companies rarely publish CAD models with corresponding textual descriptions due to intellectual property concerns, limiting the training data available for industrial applications.

3.3.2 Multimodal Generative Design and Synthesis

Beyond information retrieval and association, multimodal techniques enable novel approaches to generative design that integrate multiple input modalities to produce new designs. Whereas traditional generative design tools like Autodesk's topology optimization primarily focus on physics-based performance criteria, multimodal AI approaches can incorporate more diverse inputs including textual requirements, functional specifications, and manufacturing constraints. Wu et al. (2021) laid foundational work in this direction with DeepCAD, a transformer-based architecture that generates parametric CAD construction sequences from various input conditions [42]. Although DeepCAD did not directly incorporate textual inputs, it established the paradigm of treating CAD creation as a sequential process that could be conditioned on external parameters—opening the door for multimodal extensions.

A significant advancement in this field comes from Xu et al. (2025), who introduced CAD-MLLM, a multimodal large language model capable of generating parametric CAD models from combined text, image, and point cloud inputs [43]. Trained on the novel OmniCAD dataset, this system demonstrates remarkable capabilities for translating natural language descriptions into 3D designs. For example, a user can provide a text prompt describing a mechanical component and receive a fully realized CAD model that aligns with the specified requirements.

While still emerging, these approaches hint at future design environments where engineers interact with AI systems through natural conversation and sketches rather than through traditional CAD interfaces. A designer might describe functional requirements in text while providing partial geometric information through a sketch or scan, with the multimodal AI system synthesizing a complete design that satisfies all constraints.

Another dimension of multimodal generative design involves performance prediction and optimization. Regenwetter et al. (2022) comprehensively reviewed deep generative models for engineering design, highlighting techniques that incorporate physics-based simulation results alongside geometric information [44]. These approaches enable "performance-conditioned generation," where the AI learns to not only create shapes but also predict their functional characteristics, iteratively refining designs to meet specified performance targets.

In practice, such multimodal generative systems could dramatically accelerate the design process for contract manufacturers by automatically proposing manufacturing-ready designs based on textual requirements documents, material specifications, and performance criteria. Rather than manually translating customer requirements into CAD models, engineers could focus on refining and validating AI-generated designs—potentially reducing engineering lead times while improving design quality.

The emergence of large language models (LLMs) has further accelerated this field. Recent research has explored using LLMs to interpret technical standards and apply that knowledge to CAD models—for instance, suggesting appropriate tolerances based on a standard's requirements and a part's geometry. Work at MIT by Matusik et al. (2024) showcases the emerging capabilities where LLMs function as multimodal assistants that can answer queries about CAD models when provided with appropriate context [45].

3.3.3 Fusion of Geometry with Tabular and Sensor Data

While geometry-text integration has received significant research attention, the combination of geometric data with tabular information represents another crucial frontier for manufacturing intelligence. Tabular data in manufacturing encompasses diverse sources including bills of materials, manufacturing methods, cost structures, and time-series sensor readings—all containing vital information that complements geometric understanding.

Researchers have explored using graph neural networks to establish connections between CAD models and structured BOM data, enabling intelligent queries that span both domains. This capability is particularly valuable for complex assemblies where engineers need to understand the relationship between physical components and their documentation in product lifecycle management (PLM) systems. While academic literature on geometry-tabular fusion remains less developed than geometry-text integration, the application potential is substantial. Ghahramani et al. (2020) demonstrated how combining simulation data, sensor readings, and machine settings can create unified models for smart manufacturing process evaluation [46]. This work exemplifies the growing trend toward digital twins that integrate geometric models with operational data to create comprehensive virtual representations of physical manufacturing systems.

The application to manufacturing cost estimation is particularly relevant to this thesis. By learning relationships between geometric features, material specifications, and historical production costs, multimodal models can potentially generate more accurate quotes than traditional methods. For example, a system might learn that thin-walled features in certain materials disproportionately increase machining time and tool wear, allowing for more precise cost estimates based on both geometric and material considerations.

Wojciech Matusik's group at MIT has been particularly influential in advancing geometric learning and its integration with other modalities. Their work on JoinABLe (presented at CVPR 2022) demonstrates how geometric reasoning can infer functional relationships between parts in an assembly [47]. While primarily focused on geometry, this research illustrates how high-level semantic concepts (such as joint types) can be extracted from raw geometric data—a critical step toward bridging geometry with semantic descriptions.

Despite the promising advances in this field, evaluation methodologies present an ongoing challenge, particularly for generative and conversational multimodal systems. While retrieval tasks have clear accuracy metrics, evaluating the quality of generated designs or the correctness of multimodal question-answering remains more subjective. Researchers are actively working to develop standardized benchmarks that reflect realistic manufacturing scenarios for meaningful progress assessment.

In summary, multimodal AI in manufacturing represents an exciting frontier where geometry, language, and structured data converge. By leveraging the complementary information in each modality, integrated AI models can provide a more holistic understanding than any single-modality approach. The convergence of geometric understanding, natural language processing, and structured data analysis promises to transform key manufacturing processes, including rapid quoting systems, automated compliance verification, and intelligent design assistance that bridges the gap between geometry, specification, and production.

Chapter 4

Deep Learning in 3D and Beyond

Having established a comprehensive literature review of AI applications in manufacturing, we now transition from theoretical foundations to practical implementation. The previous chapter explored various approaches to leveraging 3D geometric data and the emerging potential of multimodal learning techniques in manufacturing contexts. Building on this foundation, Chapter 4 delves into the technical architecture and implementation details behind training these deep learning models for part similarity detection, a powerful tool for contract manufacturers to improve their quoting accuracy and CAD design lead times.

We begin with an examination of Variational Autoencoders (VAEs), which serve as the fundamental building blocks of the part similarity tool at the center of this thesis. The VAE architecture provides a powerful framework for encoding complex 3D geometries into compact, searchable latent representations while capturing key shape characteristics. This section explores the mathematical foundations, training considerations, and practical applications of VAEs specifically optimized for manufacturing data.

The chapter then progresses beyond purely geometric approaches to investigate technical implementations of multimodal models that integrate additional manufacturing-relevant attributes. By incorporating material properties and dimensional scale alongside geometric features, these enhanced models address a fundamental limitation of geometry-only approaches: the fact that manufacturing similarity depends on more than shape alone. We explore various fusion strategies, from simple feature concatenation to sophisticated embedding architectures, evaluating their relative merits for contract manufacturing applications.

4.1 Variational Autoencoders for 3D Geometric Data

4.1.1 Fundamentals of Autoencoders

Autoencoders represent a foundational class of neural networks designed for unsupervised learning of efficient data encodings [48]. Unlike supervised approaches that require labeled data, autoencoders learn directly from the data itself, making them particularly valuable for manufacturing applications where labeled datasets are often scarce or expensive to produce.

At its core, an autoencoder consists of two primary components: an encoder network and a decoder network. The encoder transforms an input $\boldsymbol{x}$ (such as a 3D model) into a compressed latent vector $\boldsymbol{z}$, while the decoder attempts to reconstruct the original input from this compressed representation, producing an output $\hat{\boldsymbol{x}}$ that ideally matches the input as closely as possible (Figure 6).

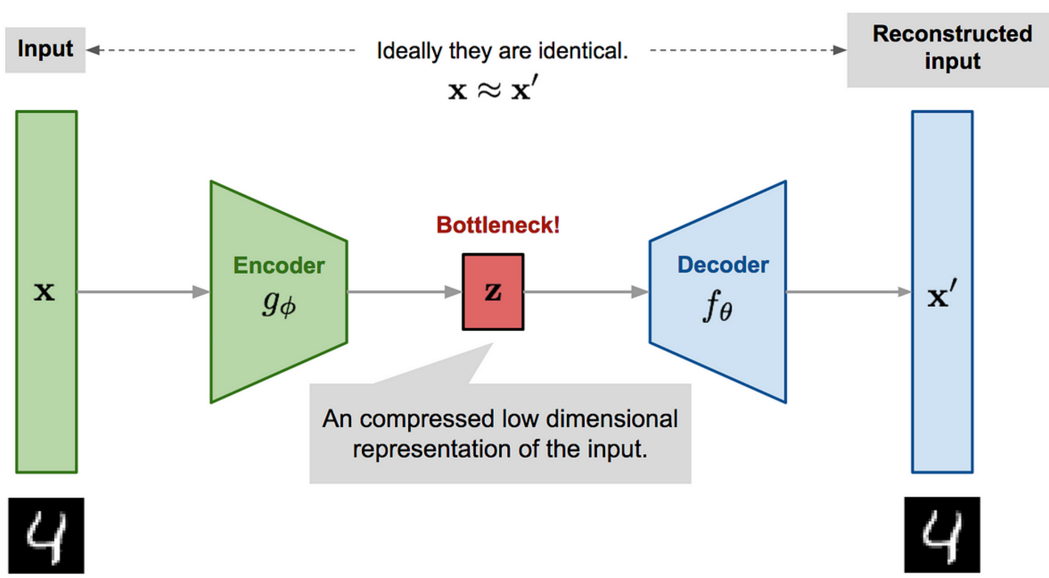

Figure 6: Autoencoder Architecture

This architecture is trained end-to-end by minimizing a reconstruction loss function $L_{recon}(x, \hat{x})$ that quantifies the difference between the original input and its reconstruction. Common loss functions include:

- **Mean Squared Error (MSE)**: $L_{MSE}(x, \hat{x}) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2$, which penalizes the squared difference between each element of the input and reconstruction.
- **Binary Cross-Entropy**: $L_{BCE}(x, \hat{x}) = \sum_{i=1}^{n}[x_i \log(\hat{x}_i) + (1 - x_i)\log(1 - \hat{x}_i)]$, often used when input values are normalized between 0 and 1.

The latent representation $\boldsymbol{z}$ typically has significantly lower dimensionality than the original input, forcing the network to learn an efficient encoding that captures the most important features while discarding redundant or noise-related information. This dimensionality reduction property makes autoencoders valuable for compressing high-dimensional data like 3D models into compact representations that can be efficiently stored, compared, or manipulated.

For 2D image data, convolutional autoencoders have become the standard architecture, using convolutional layers in both the encoder and decoder to exploit spatial relationships in the data. The encoder typically consists of multiple convolutional layers with increasing filter counts and decreasing spatial dimensions (often through strided convolutions or pooling operations), gradually compressing the input into a dense latent vector. The decoder mirrors this structure in reverse, using transposed convolutions (sometimes called deconvolutions) to progressively upsample the latent vector back to the original input dimensions.

When extending autoencoders to 3D geometric data, several adaptations are necessary:

1. **3D Convolutional Layers**: Rather than operating on 2D grids of pixels, these layers process 3D volumes of voxels, dramatically increasing computational requirements. Each 3D convolutional filter must move along three spatial dimensions instead of two.
2. **Memory Management**: A modest-resolution 3D model represented as voxels (e.g., $128 \times 128 \times 128$) already contains over 2 million elements, and the intermediate feature maps in deep networks can quickly exhaust available GPU memory. Careful architecture design with appropriate striding and pooling becomes essential.

These limitations prompted the development of Variational Autoencoders (VAEs), which introduce probabilistic constraints on the latent space to create more structured representations suited for generative modeling and smoother interpolation between encoded data points.

4.1.2 From Autoencoders to Variational Autoencoders (VAEs)

Unlike traditional autoencoders that map each input to a fixed point in latent space, VAEs encode inputs as probability distributions [49]. Specifically, the encoder in a VAE outputs parameters of a probability distribution—typically a multivariate Gaussian—rather than a direct latent representation. For an input $\boldsymbol{x}$, the encoder produces two vectors: **a mean vector $\boldsymbol{\mu}(\mathbf{x})$** and a **log-variance vector $\log \boldsymbol{\sigma}^2(\boldsymbol{x})$**. These parameters define a probability distribution in latent space from which the actual latent representation $\boldsymbol{z}$ is sampled (Figure 7).

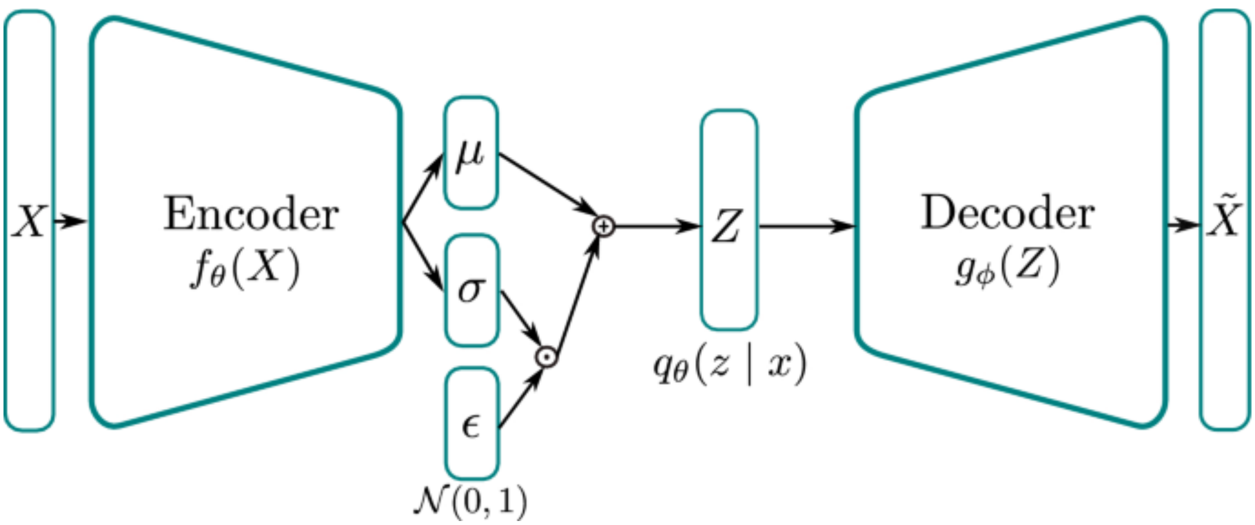

Figure 7: Variational Autoencoder Architecture

To enable backpropagation through this sampling process, VAEs employ a reparameterization trick:

$$z = \mu(x) + \sigma(x) \odot \epsilon$$

where $\epsilon \sim \mathcal{N}(0, I)$ is a random vector sampled from a standard normal distribution, and $\odot$ epresents element-wise multiplication. This transformation allows the network to backpropagate gradients through the sampling operation, as the randomness is isolated in the independent variable ϵ.

The reparameterization trick is a crucial innovation that enables VAEs to be trained effectively. To understand its importance, consider what the latent vector $\boldsymbol{z}$ actually represents: in a VAE, $\boldsymbol{z}$ is a specific sample drawn from the probability distribution $\mathcal{N}(\mu(x), \sigma^2(x))$ that the encoder predicts for input $\boldsymbol{x}$. Rather than encoding an input to a fixed point, the VAE encodes it to a cloud of likely points centered around $\mu(\mathrm{x})$ with spread determined by $\sigma(\mathrm{x})$.

The standard approach for sampling from this distribution would be to directly sample $\boldsymbol{z} \sim \mathcal{N}(\mu(\mathrm{x}), \sigma^2(x))$. However, this sampling operation is non-differentiable—we cannot compute gradients through a random sampling process—which would prevent backpropagation during training.

The reparameterization trick solves this problem by reformulating the sampling process. Instead of directly sampling from the predicted distribution, we:

1. Sample a random noise vector ϵ from a standard normal distribution $\mathcal{N}(0, \mathrm{I})$
2. Transform this sample using the predicted mean and standard deviation: $\mathrm{z} = \mu(\mathrm{x}) + \sigma(\mathrm{x}) \odot \epsilon$

Mathematically, this produces a sample from the same distribution $\mathcal{N}(\mu(\mathrm{x}), \sigma^2(x))$, but the sampling randomness is now isolated in ϵ, which is independent of the network parameters. This allows gradients to flow through $\mu(\mathrm{x})$ and $\sigma(\mathrm{x})$ during backpropogation.

The latent vector z serves multiple critical functions in a VAE:

1. **Representation of Part Characteristics**: Each dimension in z implicitly encodes some feature of the 3D part—possibly geometric attributes like overall proportions, presence of specific features (holes, pockets), or curvature characteristics. The VAE learns these features automatically during training.

2. **Specific Instance of Part Encoding**: Since z is sampled from a distribution, it represents one specific instance of how the part could be encoded. If we repeatedly encode the same part, we'll get slightly different z vectors each time, all clustered around $\mu(\mathrm{x})$.

3. **Input to the Decoder**: The decoder takes z as input and attempts to reconstruct the original part. The quality of this reconstruction depends on how well z captures the essential features of the part.

4. **Generative Applications**: The probabilistic nature of the latent space enables generative capabilities. We can sample random points from the latent space or interpolate between existing encodings to generate novel designs with a smooth transition of features.

The VAE's objective function contains two components:

$$\mathcal{L}_{\mathrm{VAE}} = \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x \mid z)] - \beta D_{KL}\left(q_\phi(z \mid x) \,\|\mathcal{N}\,(0, I)\right)$$

Where:

- $q_\phi(z \mid x)$ represents the encoder network (with parameters ϕ) that produces the approximate posterior distribution of latent vectors given the input
- $p_\theta(x \mid z)$ denotes the decoder network (with parameters θ) that generates reconstructions given latent vectors
- D_{KL} is the Kullback-Leibler divergence, measuring the difference between the encoded distribution and a standard normal distribution
- β is a hyperparameter controlling the balance between reconstruction quality and latent space regularization

The first term is the reconstruction loss, similar to standard autoencoders, encouraging the model to accurately reconstruct the input. The second term is the KL divergence, which acts as a regularizer by pushing the encoded distributions toward a standard normal distribution. This regularization creates a structured, continuous latent space with useful properties for generative applications and smooth interpolations.

When working with 3D geometric data represented as voxel grids or point clouds, VAEs face significant computational challenges due to the high dimensionality of the data. A typical 3D convolutional VAE processing voxelized parts must handle orders of magnitude more parameters than its 2D counterpart. Successful implementations require careful consideration of:

1. **Voxel Resolution**: The choice of resolution (e.g., $128 \times 128 \times 128$ voxels) directly impacts the model's ability to capture geometric details. Higher resolutions preserve fine features but increase memory requirements exponentially.

2. **Network Architecture**: The depth and stride configuration of 3D convolutional layers determine both model capacity and computational efficiency. Aggressive downsampling through strided convolutions can reduce memory usage but may sacrifice geometric details important for manufacturing applications.

3. **Latent Dimension**: The dimensionality of the latent space (typically ranging from 64 to 256) represents a balance between compression rate and information preservation. Too small a dimension may fail to capture important geometric variations, while too large a dimension may lead to less structured latent spaces and higher computational costs.

The probabilistic framework of VAEs provides several advantages over standard autoencoders for 3D applications:

1. **Regularized Representations**: The KL divergence term encourages latent vectors to distribute more evenly, preventing the "dead zones" that can occur in standard autoencoder latent spaces.

2. **Controlled Sampling**: The latent space becomes a well-structured probability distribution, allowing for principled sampling of new shapes that smoothly interpolate between existing ones.

3. **Uncertainty Quantification**: The stochastic encoding process implicitly captures uncertainty about the representation, which can be valuable in engineering contexts where confidence in predictions matters.

For 3D manufacturing data, VAEs create a foundation for working with complex geometric information in a computationally tractable manner. The structured latent space enables numerous downstream applications, from retrieval tasks to generative design, all while accommodating the inherent variability and complexity of real-world manufacturing parts.

4.1.3 Latent Space Indexing for Similar Part Retrieval

Once a Variational Autoencoder has been trained to encode 3D part geometries into a lower-dimensional latent space, the resulting embeddings can be leveraged to create a powerful similarity search system. This section explores the technical implementation of latent space indexing for part retrieval—a core component of a part similarity tool that enhances quoting accuracy and reduces engineering lead times.

After training, each part in the manufacturing database can be encoded into a latent vector $\mathbf{z}_i \in \mathbb{R}^d$, where d typically ranges from 64 to 256 dimensions. This process involves passing each part's 3D representation (voxel grid, point cloud, or mesh) through the encoder network to obtain the distribution parameters $\mu(\mathbf{x})$ and $\sigma(\mathbf{x})$. While the VAE's stochastic nature means that multiple encodings of the same part will yield slightly different latent vectors through the sampling process $\mathbf{z}_i = \mu(\mathbf{x}) + \sigma(\mathbf{x}) \odot \boldsymbol{\epsilon}$, for indexing purposes we typically use the mean vector $\mu(\mathbf{x})$ directly to ensure consistency.

4.1.4 Relevance for the Manufacturing Context

From a design standpoint, latent vector indexing helps in the reuse of existing engineering work. When a newly proposed component strongly resembles a part already in the catalog, the user can retrieve proven CAD Files for previously manufactured parts, furthermore, standardizing manufacturing processes through a part family commonization.

Beyond part family grouping quoting stands to gain significantly from similarity-based retrieval. Instead of relying on ad hoc or purely experience-driven estimates, a system can automatically retrieve historical parts with similar geometry and glean real-world data on actual production performance. This retrieval process incorporates both successful and unsuccessful quotes—a critical distinction that enhances predictive power. Data from successful quotes provides proven cost baselines and production parameters, while unsuccessful quotes (those that didn't win business) offer valuable insights into competitive positioning and potentially overestimated costs. By analyzing why similar parts were quoted successfully or unsuccessfully, estimators can calibrate their approach to the current market conditions. This historical baseline provides a more informed quote, especially in job-shop or contract-manufacturing contexts where the next RFQ might involve radically different geometric demands. By incorporating not only geometry but also scale and material tags into the latent representation—or as side-channel metadata—the quoting system can approximate how minor dimensional changes or different material compositions could shift the cost and lead time.

Quality control and traceability are also enhanced via part family identification. If a defect emerges in a batch of manufactured components, engineers can query the latent database for part families, possibly uncovering a systemic machining parameter or tool choice that triggers the same flaw. This approach can be invaluable in regulated industries—such as aerospace or medical—where process traceability and corrective action must be rigorously documented.

In sum, the introduction of a latent-vector-driven similarity retrieval system brings multiple manufacturing workflows into closer alignment. By tying design, production, quoting, and quality feedback loops together, a robust VAE indexing structure effectively transforms previously unstructured geometric data into a knowledge repository of shape relationships and best practices. This synergy is especially impactful in high-mix, low-volume operations, as it mitigates the overhead of starting from first principles for each new job. This provides a scalable foundation for data-driven decision-making that can reduce lead times, enhance consistency, and ultimately confer a competitive edge in the contract-manufacturing marketplace.

4.2 Improving Predictive Power with Multimodal Approaches

While geometric similarity alone can accelerate part reuse and streamline quoting processes, practical manufacturing considerations often hinge on more than just shape. In aerospace and defense contexts, two parts may share nearly identical geometry but differ in their material composition or overall scale—factors that dramatically affect machining time, tool choice, and production cost. Recognizing this gap, the present work extends the original VAE-based latent space for geometry by integrating additional data modalities, most notably material and scale, to improve predictive power in real-world manufacturing scenarios.

4.2.1 Motivations for Multimodal Integration

Manufacturing engineers routinely emphasize how material and dimensional scale can overshadow geometric similarity when evaluating production feasibility. Exotic metals such as titanium or Inconel require specialized tooling and higher machine rigidity, even for relatively simple geometries, while a large-scale aluminum part might demand completely different manufacturing strategies. Merging these attributes—geometry, material, and scale—into a unified representation ensures that any similarity measure more accurately reflects actual production requirements.

4.2.2 Data Modalities and Preprocessing

To create a multimodal part similarity system, we must first establish appropriate representations for each data modality. This section summarizes how geometric, scale, and material information is processed to create compatible numerical vectors that can be integrated into a unified similarity measure. Each modality presents unique preprocessing challenges that must be addressed to ensure meaningful comparisons across diverse manufacturing parts.

1. **Geometry (Existing VAE Latent Space):** The foundation remains the Variational Autoencoder trained on voxelized 3D parts. Each part's geometry is encoded as a latent vector $\boldsymbol{z_g} \in \mathbb{R}^{\mathrm{d}}$. Prior sections outlined how the VAE learns to compress shape information, making these vectors effective for shape-based retrieval. To draw a distinction between the latent space vector representing a part's geometry and that of its scale and material, the notation $\boldsymbol{z_g}$ is used where the subscript "g" refers to geometry.
2. **Scale:** Many parts differ primarily in their bounding box dimensions along $\mathrm{x}, \mathrm{y}, \mathrm{z}$. The bounding box represents the smallest rectangular prism that fully contains the part, providing a simple measure of overall part size. These dimensions can span multiple orders of magnitude in a typical manufacturing database, so a common approach is to normalize or log-transform the raw dimensional values (e.g., $\log{(\mathrm{x}+1)}$) to reduce skew and ensure that larger parts don't disproportionately influence similarity calculations. The resulting processed scale vector $\mathrm{z_s}$ typically has three dimensions corresponding to length, width, and height, though some applications might augment this with derived values such as aspect ratios or volume.
3. **Material:** Material identity often appears in textual or categorical form, e.g., "AISI 4140 Steel" or "Grade 5 Titanium." One approach is one-hot encoding, generating a binary

vector with length equal to the number of known materials. Alternatively, if the material field is free-text or highly diverse, NLP-based embeddings (e.g., TF–IDF or a lightweight language model) may better capture nuances. The resulting vector z_m thus encodes material information in a numeric format suitable for similarity computations.

4.2.3 Strategies for Multimodal Fusion

After preprocessing each data modality into compatible vector representations, the next challenge is determining how to effectively combine these disparate information sources into a unified similarity measure. Several integration approaches exist, each with distinct advantages and trade-offs in terms of implementation complexity, interpretability, and predictive performance. The following methods represent a spectrum of techniques for multimodal fusion, ranging from straightforward concatenation to sophisticated joint embedding architectures designed specifically for manufacturing applications.

a) **Simple Feature Concatenation:** A straightforward approach appends material and scale encodings directly to the geometric latent vector. The final vector might look like

$$\mathbf{z} = [\, \mathbf{z}_g \;\|\; \mathbf{z}_m \;\|\; \mathbf{z}_s \,],$$

where "‖" indicates concatenation. The system can then compute cosine similarity or Euclidean distance on z. A weighting scheme helps balance the different modalities; one might rescale or multiply each sub-vector by a coefficient $\alpha_g, \alpha_m, \alpha_s$ to modulate their relative importance. This method is simple to implement and interpret, although care must be taken so that large or high-dimensional sub-vectors do not drown out the others.

b) **Multimodal Autoencoder:** More advanced architectures revise the original VAE to jointly encode geometry, material, and scale within a single latent space. Here, the VAE encoder ingests both 3D voxel data and structured attributes. The decoder may attempt to reconstruct the part's geometry and optionally predict or re-generate the associated scale and material labels. By training a single network end-to-end, correlations among geometry, size, and material can be captured more holistically. However, integrating these diverse data types often demands careful hyperparameter tuning, a well-designed loss function that balances reconstruction across modalities, and larger training sets to ensure generalizability.

c) **Metric Learning (Siamese Networks):** Another alternative uses labeled pairs or triplets of parts—labeled "similar" or "dissimilar"—to train a Siamese or triplet-loss network [50]. This architecture ingests geometry, material, and scale data and directly learns an embedding space in which distance corresponds to actual similarity judgments. If robust training labels exist (e.g., curated by manufacturing experts who label which parts are "equivalent enough" for particular processes), metric learning can yield strong performance. Still, this method relies on obtaining or generating the necessary similarity labels, which can be time-intensive.

d) **Hybrid or Ensemble Methods:** A pragmatic approach might keep geometry separate from a "scale + material" vector, compute separate similarity scores for each, and then combine them (e.g., via a weighted sum). This yields a final similarity between parts A and B:

$$S(A, B) = \alpha_g \, \mathrm{sim}_g(A, B) + \alpha_s \, \mathrm{sim}_s(A, B) + \alpha_m \, \mathrm{sim}_m(A, B),$$

where each sim_* is chosen (e.g., cosine or inverse Euclidean). Ensemble methods can be tuned with domain expertise or automatic validation, ensuring each modality's impact matches its real-world importance.

4.2.4 Handling Missing Data

In practice, material or scale data can be incomplete. Some shops may not have consistent records for older parts, or they might omit bounding box metadata for certain assemblies. One common remedy is imputation, wherein missing scale values are replaced by means or medians, and missing materials default to a placeholder category. Alternatively, advanced model-based or probabilistic approaches can be employed, though these often require consistent patterns of missingness. An alternative "masking" strategy is to adjust weighting or skip certain modalities if data for them is unavailable, ensuring partial data does not unduly skew the similarity measure.

4.2.5 Practical Implementation Steps

Implementing multimodal integration follows a logical pipeline that transforms raw part data into meaningful similarity comparisons. The process combines preprocessing, feature extraction, and modality fusion into a cohesive workflow that balances computational efficiency with retrieval accuracy (Figure 9).

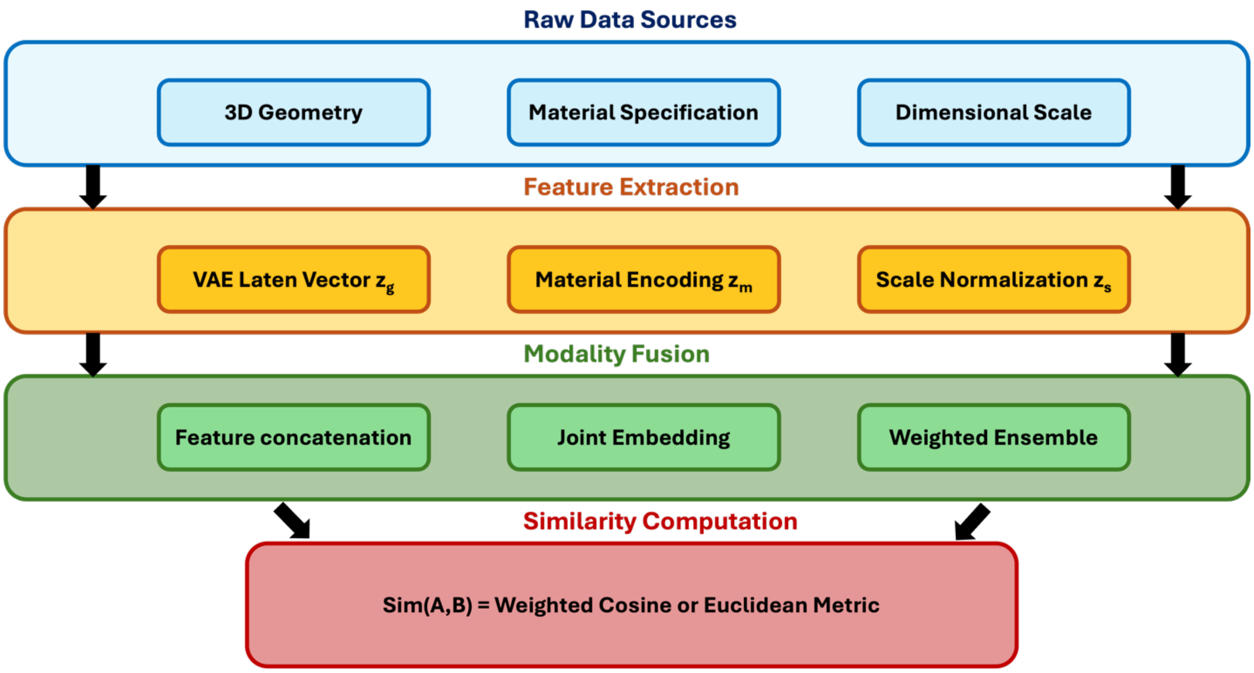

Figure 8: Multimodal Integration Pipeline

The implementation process follows several key steps:

1. **Feature Extraction**: Each part's geometry latent vector $\mathbf{z}_\mathrm{g}$ is either precomputed (if continuing to use a geometry-only VAE) or generated on the fly (if using a unified multimodal VAE). Scale vectors $\mathbf{z}_\mathrm{s}$ are derived by extracting bounding box dimensions and applying appropriate normalization. Material vectors $\mathbf{z}_\mathrm{m}$ are generated through one-hot encoding or more sophisticated embeddings of the material specification.

2. **Modality Fusion**: The system then merges or fuses these features using one of the approaches discussed in Section 4.2.3, optionally introducing user-configurable weights to balance the influence of each modality.

3. **Similarity Calculation**: For comparing parts, a weighted cosine similarity measure provides an effective and computationally efficient approach:

$$\text{Similarity}(\mathrm{A},\mathrm{B}) = \frac{(\alpha_g\,\mathbf{z}_{g,A} + \alpha_m\,\mathbf{z}_{m,A} + \alpha_s\,\mathbf{z}_{s,A}) \cdot (\alpha_g\,\mathbf{z}_{g,B} + \alpha_m\,\mathbf{z}_{m,B} + \alpha_s\,\mathbf{z}_{s,B})}{\|\,\alpha_g\,\mathbf{z}_{g,A} + \alpha_m\,\mathbf{z}_{m,A} + \alpha_s\,\mathbf{z}_{s,A}\,\| \;\; \|\,\alpha_g\,\mathbf{z}_{g,B} + \alpha_m\,\mathbf{z}_{m,B} + \alpha_s\,\mathbf{z}_{s,B}\,\|},$$

where $\mathbf{z}_g, \mathbf{z}_m, \mathbf{z}_s$ represent geometry, material, and scale embeddings respectively, and $\alpha_g, \alpha_m, \alpha_s$ are modality weights. The dot product in the numerator measures the alignment between the weighted combined vectors, while the denominator normalizes by the product of their magnitudes (Euclidean norms), yielding a similarity score between -1 and 1.

This weighted similarity measure can be adapted to accommodate domain-specific priorities. These weights can be adjusted based on expert knowledge or systematically optimized through validation against known similar part pairs.

The implementation balances theoretical rigor with practical utility. Rather than pursuing complex fusion architectures that might offer marginal improvements at the cost of interpretability, the system prioritizes a transparent approach where domain experts can understand and adjust the influence of each modality. This transparency is particularly important for manufacturing applications, where engineers and estimators must trust the system's recommendations to incorporate them into critical business processes.

4.2.6 Evaluation and Weight Tuning

Assessing the impact of multimodal integration typically involves both quantitative and qualitative measures. For the former, one may compute metrics such as Precision at K, Recall at K, or Mean Average Precision (mAP) by comparing retrieved similar parts to a ground truth set. Alternatively, if the organization systematically tracks “acceptable alternatives” in quoting or production records, those labels can validate how well the integrated similarity measure aligns with actual manufacturing outcomes. Qualitative evaluation, however, remains essential. Manufacturing engineers can inspect top retrieval results to confirm whether the system’s “similar” suggestions truly a part in the same family.

Tuning the weight parameters $\alpha_g, \alpha_m, \alpha_s$ likewise relies on some combination of heuristic knowledge and validation data. A partial grid or random search over weight triplets can quickly converge on a region with acceptable performance. More advanced techniques, such as Bayesian optimization or gradient-based methods, may be applied if the search space is high-dimensional or if new modalities are added [51].

4.2.7 Balancing Complexity and Real-World Utility

A recurring concern is the trade-off between complexity and interpretability. Simple concatenation plus weighting is relatively transparent: domain experts easily see how geometry, material, and scale each factor into the final similarity. In contrast, a multimodal autoencoder or Siamese network might achieve higher accuracy but can obscure the exact role each modality plays. Moreover, extended architectures demand larger training datasets, and the overhead of capturing or cleaning additional data cannot be neglected.

From a manufacturing operations perspective, the ultimate goal is to reduce lead time and improve quoting accuracy, while ensuring that staff trust the system enough to adopt it. If a complex method outperforms simpler approaches only marginally, the additional engineering overhead may not be justified in certain contract manufacturing settings. Conversely, in highly specialized environments (e.g., advanced aerospace shops dealing with exotic alloys), a more sophisticated approach might deliver substantial cost savings by accurately capturing the interplay of geometry, material, and scale.

4.3 Conclusion

In this chapter, we explored the limitations of pure geometry-based part retrieval methods and established the need for multimodal approaches to improve predictive power in industrial settings. Variational Autoencoders (VAEs), particularly those designed for 3D volumes, provide a structured, low-dimensional latent space conducive to shape-based similarity searches. The extension of these VAEs to multimodal embeddings—through augmentation, specialized loss functions, and tailored training routines—enables the incorporation of material and scale data.

Ultimately, these approaches support more accurate and contextually relevant part matching, laying the groundwork for next-generation manufacturing intelligence systems. However, they also introduce new challenges in terms of model complexity, interpretability, and data management. The next chapter (Chapter 5) will examine how to validate such deep neural models in practice, discussing issues such as data leakage, model interpretability, and real-world metrics for success in an industrial setting.

In sum, multimodal integration enhances the utility of latent-space part embeddings by reflecting the true constraints of manufacturing. Geometry alone, while powerful, overlooks critical production drivers like scale and material type. By fusing these data sources—via simple concatenation, specialized VAE designs, metric learning frameworks, or hybrid ensembles—manufacturers gain a more faithful measure of part similarity. This leads to more robust part retrieval, quoting, and scheduling decisions, ultimately boosting efficiency in complex, high-mix, low-volume production environments.

Chapter 5

Model Learning and Validation

This chapter details the methods and challenges involved in training and validating the deep neural networks described earlier, focusing on how to ensure practical reliability in a manufacturing setting. After discussing the general background and pitfalls of validating deep learning in industrial contexts, we highlight several tools and procedures used to test the utility of a part-similarity tool. We then present a case study comparing the performance of a model trained solely on 3D geometry against a multimodal variant that integrates scale and material data. Finally, we explore additional strategies for future validation, including reinforcement learning with human feedback.

5.1 Background

Model validation in the domain of 3D part similarity must balance theoretical correctness—ensuring that learned embeddings align with ground truth geometric relationships—and practical utility, meaning that the tool genuinely aids in quoting and lead time reduction. Unlike purely academic tasks where curated datasets and standardized benchmarks dominate, manufacturers typically deal with heterogeneous, incomplete, and noisy production data. Additionally, specialized domain constraints—for instance, the need to maintain traceability in regulated industries—place extra emphasis on interpretability and verifiability.

With the Variational Autoencoder (VAE) at the center of the geometry-based approach, validation requires assessing both reconstruction fidelity (how well the model captures shape details) and latent consistency (whether similar shapes embed nearby in latent space). As we integrate material and scale features (Chapter 4.3), validation further expands to assess whether the model appropriately reflects real-world manufacturing constraints—i.e., do parts that differ in material or bounding dimensions show appropriately reduced similarity scores?

5.2 Challenges in Model Validation

Validating deep learning models for manufacturing applications presents unique challenges that extend beyond those encountered in typical academic machine learning research. In contract manufacturing environments, where data quality varies significantly and ground truth can be subjective, traditional validation approaches must be adapted and extended. The following sections outline the key challenges faced during the development and validation of the proposed part similarity tool, and the strategies implemented to address each obstacle. These foundational considerations informed the design of the comparative case study, which quantitatively evaluates the performance differential between geometry-only and multimodal approaches.

5.2.1 Data Quality and Completeness

One of the earliest challenges arises from missing or inconsistent data in manufacturing environments. Several specific issues may be encountered during model development:

- **Inconsistent Material Documentation**: Approximately 50% of historical parts lack precise material specifications, with entries ranging from detailed alloy designations (Ex: "6061-T6 Aluminum") to vague descriptions ("Aluminum Alloy") or completely missing values. This variation complicats the creation of standardized material embeddings.
- **Dimensional Inconsistencies**: Bounding box dimensions are often recorded in different units (inches vs. millimeters) or coordinate systems, requiring extensive preprocessing to normalize. In some cases, parts could have been modeled at non-unity scale factors, further complicating dimensional analysis.
- **Revision Control Challenges**: In an SME manufacturing context, many parts exist in multiple versions with minor geometric modifications, creating ambiguity about which version should be considered the "reference" for similarity calculations.

- **Temporal Data Gaps**: Production data (cycle times, setup hours) can also be more consistently available for recent parts (post-2021 for Company X) but spotty for older components, creating potential biases in validation metrics that relied on historical manufacturing outcomes.

To address these issues, a multi-faceted data preparation pipeline was implemented at Company X. This included rigorous cleaning protocols for detecting and correcting common unit inconsistencies, with anomalous values flagged for human review. For missing material data, a hierarchical imputation strategy was employed that first attempted to infer material from part naming conventions, then from part family associations, and finally from geometric characteristics that correlate with certain materials.

Before splitting training and testing sets, part revisions were identified and grouped to ensure that all versions of a part remained in the same split, preventing data leakage. These strategies improved data quality substantially, but incomplete information remained an inherent limitation of real-world manufacturing datasets that any practical model must accommodate.

5.2.2 Overfitting and Data Leakage

Deep neural networks processing complex 3D geometric data are inherently susceptible to overfitting and data leakage issues. These challenges are particularly acute in manufacturing contexts due to the high-dimensional nature of 3D part representations and the limited size of available training datasets. Overfitting occurs when models effectively memorize training examples rather than learning generalizable features, while data leakage happens when information from the test set implicitly influences model training, creating artificially inflated performance metrics that fail to translate to real-world applications.

At Company X, these general challenges manifested in several specific ways during model development. Initial VAE implementations with large latent spaces (200 dimensions) showed signs of memorization, particularly for uncommon geometries. The models achieved a decent reconstruction of training examples but performed worse on validation parts with similar but nonidentical features. This discrepancy revealed that the models were learning to encode specific parts rather than meaningful geometric characteristics that would generalize to novel designs.

Many components existed in multiple revisions with minor geometric modifications, stored as separate files within the product data management system. When these revisions were randomly distributed between training and testing sets during early experiments, the model achieved misleadingly high performance by essentially recognizing minor variations of already-seen parts rather than learning true similarity principles. This phenomenon can create an illusion of accuracy that disappears when the model encounters genuinely new designs.

To address these issues, several technical countermeasures can be implemented. Both L2 regularization applied to network weights and KL-divergence constraints on the latent space can effectively limit the model's capacity to memorize specific examples. Training can be augmented with controlled geometric variations including rotations, minor scaling adjustments, and small perturbations to vertex positions, encouraging robustness to variations.

Rather than simple random splits, manufacturing family-aware stratification can be employed, ensuring that each fold contained representative samples from all major part categories. For certain experiments, time-based validation can be used, training on older parts and testing on newer designs, which better reflects the model's intended use in production. These strategies can significantly reduce overfitting.

5.2.3 Domain Complexity and Lack of Standard Benchmarks

Unlike image recognition tasks with established benchmarks such as ImageNet, 3D manufacturing data lacks widely accepted test sets and evaluation protocols. This absence of standardization created significant validation challenges for the part similarity system. No publicly available dataset exists that combines 3D manufacturing geometries with corresponding material specifications, manufacturing process parameters, and production outcomes—the key elements needed for comprehensive validation.

Company X's part library, while extensive, contains proprietary designs that cannot be shared with the broader research community, limiting opportunities for external validation or comparison with other techniques. Standard metrics like nearest-neighbor accuracy or clustering purity fail to capture manufacturing-specific requirements such as process compatibility or setup similarity. Additionally, different stakeholders value different aspects of similarity; estimators prioritize picing opportunities, while production is heavily focused on efficiency opportunities, requiring a nuanced evaluation framework that can accommodate multiple perspectives.

To address these benchmark limitations, a multi-faceted validation framework should be explored. This includes the creation of an expert-annotated test set where manufacturing engineers identify groups of parts with similar production requirements, providing ground truth for retrieval evaluation. A process validation matrix can be constructed mapping combinations of materials, geometries, and scales to true manufacturing part families allowing automated assessment of whether model-suggested similarities align with reality. Rather than relying solely on absolute performance measures, the evaluation should emphasize comparative improvements between baseline and enhanced models, which provides more meaningful assessment given the lack of established benchmarks. Beyond standard machine learning metrics, practical outcomes should be tracked, such as reduction in development time when engineers used model-suggested similar parts as starting points. This framework can enable meaningful validation despite the lack of standardized benchmarks, but highlights the need for community-wide efforts to develop shared resources for manufacturing AI evaluation.

5.2.4 Interpretability and Trust

Perhaps the most significant challenge for practical adoption of deep learning in manufacturing environments is the need for interpretability and trust. Engineers and managers responsible for critical business decisions must understand why the model makes specific similarity suggestions. The VAE's learned latent space, while mathematically powerful, lacks inherent interpretability—it is not immediately clear what each dimension represents or how dimensions interact to determine similarity. When combining geometry, material, and scale, the relative influence of each factor on the final similarity score is not naturally transparent to end-users. Similarity recommendations that contradict experienced manufacturers intuition, even if theoretically justified, face significant adoption barriers without clear explanations. In regulated industries like aerospace, decisions must be traceable and auditable, requiring explanations for why specific manufacturing approaches were selected.

To enhance interpretability and build user trust, several targeted approaches can be implemented. SHAP (SHapley Additive exPlanations) techniques can be used provide visual explanations of which factors most strongly influenced each similarity recommendation [52]. Each similarity suggestion is accompanied by a confidence score and specific indicators highlighting potential concerns, such as similar geometry but different material class or significant scale differences.

A user interface can be utilized to allow engineers to adjust the weights of different modalities in real-time, observing how changes affect recommendations and developing intuition about the model's behavior. Rather than deploying the complete system at once, capabilities should be introduced gradually, starting with the most interpretable features and adding complexity as user confidence grew.

These interpretability enhancements proved critical for practical adoption at Company X. In conversation with engineers, they reported that understanding why parts were considered similar was often as important as the accuracy of the recommendations themselves. The transparent approach also facilitates iterative refinement, as users could provide more precise feedback when they understood the reasoning behind recommendations.

The challenges outlined in this section shape both the technical approach and evaluation methodology that can be implemented at a contract manufacturing firm like Company X. By addressing data quality issues, preventing overfitting, developing appropriate validation frameworks, and prioritizing interpretability, a foundation was established for meaningful assessment of the multimodal part similarity system. The comparative case study in the following section builds directly on these considerations, qualitatively evaluating how effectively the enhancements address the limitations of geometry-only approaches.

5.3 Comparative Case Study: Geometry-Only vs. Multimodal Approaches

To assess the impact of incorporating material and scale information alongside geometric features, a comparative study was conducted at Company X. This evaluation analyzed how different components of the similarity metric affected retrieval results through a systematic ablation approach. Rather than creating separate training and test datasets, the study focused on comparing the output of the same model with different combinations of modality weights enabled or disabled, providing direct insight into each modality's contribution to manufacturing-relevant similarity detection.

5.3.1 Ablation Study Design

The study employed a controlled ablation methodology to isolate the contribution of each modality to the similarity measure. Using the multimodal similarity function developed in Chapter 4, the relative weights of geometric, material, and scale components were systematically manipulated to create several experimental conditions: $\alpha_g, \alpha_m, \alpha_s$

1. **Geometry-Only**: Setting $\alpha_g = 1.0,\ \alpha_m = 0.0,\ \alpha_s = 0.0$, effectively using only the VAE's geometric latent representation
2. **Geometry+Material**: Setting $\alpha_g = 0.7,\ \alpha_m = 0.3,\ \alpha_s = 0.0$
3. **Geometry+Scale**: Setting $\alpha_g = 0.7,\ \alpha_m = 0.0,\ \alpha_s = 0.3$
4. **Full Multimodal**: Setting $\alpha_g = 0.6,\ \alpha_m = 0.2,\ \alpha_s = 0.2$, engaging all modalities with a geometry-dominant weighting

This approach allowed direct observation of how each modality affected similarity results when incorporated into or removed from the similarity calculation, without requiring retraining of separate models for each condition

5.3.2 Query Selection and Evaluation Process

The engineering team at Company X selected 30 representative query parts spanning diverse manufacturing characteristics. The selection included simple components with standard features, complex assemblies with intricate geometric details, parts machined from both standard and exotic materials, and components representing various scale extremes. Particular attention was paid to including parts with unusual combinations of geometry and material that might highlight the limitations of geometry-only approaches.

For each query part, the system generated similarity rankings under each ablation condition, retrieving the top 10 most similar parts from Company X's historical database of over 14,000 components. These results were then evaluated through a structured qualitative assessment process.

The evaluation team assessed each set of results based on manufacturing process similarity, tooling compatibility, fixturing requirements, expected cost similarity, and overall usefulness for engineering and quoting purposes. Rather than using formal numerical ratings, the team engaged in detailed discussions about the practical implications of each result set, documenting specific observations about manufacturing relevance and limitations.

5.3.3 Qualitative Findings

The ablation study revealed significant differences in the nature and quality of similarity results across the different modality configurations. When using only geometric similarity, the system frequently returned parts that appeared visually similar but would require fundamentally different manufacturing approaches. Engineers at Company X particularly noted this limitation when examining results for complex components made from specialized materials. In several cases, the geometry-only configuration returned visually similar components that would be machined using entirely different parameters due to material property differences.

For example, when querying with a complex aluminum part, the geometry-only configuration returned visually similar components made from stainless steel and titanium. Despite their geometric resemblance, these parts would have significantly different costs associated with the manufacturing process. An experienced programmer commented that using the toolpath from the stainless steel component as a starting point for the aluminum part would result in excessive tool wear and potential quality issues, despite their geometric similarity.

Adding material weighting dramatically improved manufacturing relevance. It was noted that the Geometry+Material configuration consistently returned parts that could be machined with similar tooling and process parameters. The impact was particularly evident with different grades of steel, where material properties fundamentally dictate machining approach regardless of geometric characteristics.

The Geometry+Scale configuration revealed how dimensional factors influenced manufacturing approach independently of shape. For large structural components, scale often dictated process requirements and machine selection that transcended pure geometric similarity. It was noted that certain large components required machinery that differed from that of smaller parts. Similarly, very small precision components sometimes needed specialized approaches regardless of their geometric similarity to larger parts. The scale weighting helped capture these critical manufacturing considerations that would otherwise be missed by pure shape analysis.

The full multimodal configuration consistently received the strongest positive feedback for manufacturing relevance and practical utility. It was noted that the combined weighting created more balanced results that considered all factors affecting manufacturing similarity. The evaluation team reported that these results most closely matched how they would manually group parts for process planning or cost estimation purposes.

5.3.4 Implications for Practice

The ablation study conducted at Company X yielded several practical insights for implementing part similarity tools in manufacturing environments. The study revealed that different manufacturing roles benefited from different modality emphases. Estimators found material and scale information particularly valuable for costing. One estimator noted that the multimodal system allowed them to generate quotes that were more consistent with actual production costs by identifying truly comparable historical parts rather than merely similar-looking components.

The incorporation of scale information was especially valuable. Estimators noted that scale directly impacts machine selection, cycle time, and material costs—factors that significantly affect price calculations. The ability to find similar parts at comparable scales allow for more accurate cost extrapolation and reduce the variance in quoted prices. Meanwhile, CAD/CAM programmers emphasized that scale often dictates process paramaters regardless of of geometric similarity. By considering scale alongside geometry, the system can return results that required similar machining approaches, further streamlining the manufacturing process.

Based on these findings, the deployed system should be designed with user-adjustable modality weights, allowing manufacturing teams to emphasize different factors depending on their specific goals. This flexibility is critical for adoption, as it allows each individidual to tailor the similarity calculations to their specific manufacturing concerns.

The comparative case study clearly demonstrated the limitations of purely geometric approaches to part similarity in manufacturing contexts. While geometry provides an essential foundation, the incorporation of material and scale information substantially improves the alignment between computational similarity measures and practical manufacturing considerations. This finding validates the multimodal approach developed in Chapter 4 and establishes it as a useful tool for manufacturing applications where cost estimation and efficient manufacturing processes are primary concerns.

5.4 Reinforcement Learning with Human Feedback in Manufacturing

While the ablation study demonstrated the value of integrating multiple modalities for part similarity assessment, a fixed weighting scheme cannot fully capture the nuanced decision-making of experienced manufacturing professionals. The case study revealed that different experts placed varying emphasis on different aspects of similarity based on their domain knowledge and specific manufacturing constraints. To address this limitation and further refine the similarity model, reinforcement learning with human feedback presents a promising approach for continuous improvement. This section explores how human expertise can be systematically incorporated into the similarity assessment system through feedback mechanisms that allow the model to adapt over time to the specific priorities at Company X.

5.4.1 Motivation for Human-in-the-Loop Systems

In many industrial AI deployments, purely data-driven models fail to fully capture the expert heuristics of seasoned machinists or programmers. Reinforcement learning (RL) with human feedback offers a hybrid approach: after the system proposes similar parts or cost estimates, domain experts can provide signals indicating “good” or “bad” matches [53]. Over time, the model tunes its similarity metric or cost predictions to align better with these real-world judgments.

5.4.2 Mechanisms of Feedback and Reward

A straightforward mechanism for implementing RL is to treat each user decision (e.g., “Accept” or “Reject” a recommended part) as a reward signal. The system’s embedding or similarity function updates accordingly, effectively learning which latent-space distinctions matter most. More advanced setups might track partial matches, weighting certain attributes—like hardness or machining history—more heavily if experts routinely prioritize these factors in their acceptance or rejections.

5.4.3 Practical Considerations for Deployment

Reinforcement learning solutions in manufacturing must be carefully introduced so as not to overburden staff. A user-friendly interface that captures minimal, low-friction feedback is critical. Also, the data can be highly unbalanced if “Accept” signals are rare or if certain sets of parts dominate daily operations. In these cases, online or semi-supervised RL algorithms must maintain stability despite skewed reward distributions.

5.5 Future Validation Methods

One avenue for future work is the development of open, standardized part libraries that reflect real manufacturing diversity. Such shared resources would enable more robust external benchmarking and facilitate academic–industry collaboration. Given the complexity of both geometry-based embeddings and multimodal expansions, interpretability remains a priority.

Techniques like SHAP (SHapley Additive exPlanations) can apportion how geometry vs. scale vs. material factors drive the similarity score. This helps build trust among engineers and provides a basis for identifying erroneous model behaviors.

A pragmatic next step involves deploying the part-similarity tool in a real shop environment, using A/B testing to compare production lines or quoting teams that have access to the tool versus those that do not. Metrics such as quoting error, or average lead time can measure the direct impact on business outcomes, lending further credibility to the model's practical value. Beyond the pilot-level demonstration, future systems could incorporate online reinforcement learning and active feedback. Over time, the system would refine its similarity metrics, quote estimates, and part reusability suggestions based on direct acceptance or rejection feedback.

In conclusion, this chapter underscored the multiple facets of deep learning model validation in a real-world manufacturing context. It illustrated how a naive geometry-only approach, though capable of capturing 3D shapes, can fall short when material and scale are pivotal. Incorporating these additional modalities significantly improves the retrieval's practicality, as shown by the simple but revealing case study. By further refining the validation pipeline—through advanced benchmarking, user feedback, and interpretability mechanisms—this system can continue to evolve and make substantial contributions to data-driven manufacturing.

Chapter 6

Applications and Conclusion

This final chapter explores broader opportunities for multimodal AI in manufacturing and concludes the thesis by reflecting on the challenges, limitations, and key insights that emerged from our research. By examining the concept of Large Manufacturing Models (LMMs)—unified systems that learn from data across an entire production lifecycle—this chapter underscores both the exciting possibilities and the pragmatic barriers that remain.

6.1 Large Manufacturing Models and Applications

The notion of a Large Manufacturing Model (LMM) builds on the idea of consolidating multiple data modalities—geometric shapes, material characteristics, machine-sensor logs, scheduling records, and even textual design annotations—into a single analytical framework. Inspired by large-scale neural networks used in other domains, an LMM in manufacturing would capture comprehensive process knowledge, improving tasks like quoting and manufacturing process development. For instance, engineers could upload a new 3D design, and an LMM would not only identify geometrically similar parts but also factor in cost data, manufacturing contraints, and process capabilities. This reduces rework while fostering greater process standardization.

Real-world applications of such models might include automated reallocation of jobs across machines based on real-time utilization metrics. In the long term, LMMs could serve as "digital brains" for factories, orchestrating everything from vendor selection to adaptive process corrections when sensor readings deviate. These scenarios highlight a future in which advanced analytics and AI become integral to everyday manufacturing decisions.

6.2 Challenges, Limitations, and Key Takeaways

Implementing large-scale multimodal systems faces significant obstacles. Data fragmentation remains common: CAD/CAM files, ERP entries, and sensor logs are often sequestered in different software silos, lacking standardized naming conventions or timestamps. Beyond data availability, computational costs can be immense, especially when merging high-resolution 3D geometries with streaming IoT data and unstructured text. Researchers and engineers must consider trade-offs between model depth, training time, and inference speed, balancing accuracy with practical deployment requirements.

In addition, explainability and domain trust remain critical. A deep model that recommends a particular toolpath or cost estimate must provide sufficient rationale for machinists, production managers, and quality engineers to adopt it in daily practice. Human expertise, honed over years, offers nuanced insight not always captured by even the most comprehensive datasets. Effective deployments likely require user interfaces that let practitioners see how geometry, material, or machine data jointly influence a recommendation, bridging any gap between data-driven suggestions and shop-floor realities.

Nonetheless, the potential benefits are considerable. By unifying geometry with operational features, like material cost trends, companies can develop more informed quotes, better job scheduling, and improved traceability of production anomalies. As demonstrated in the earlier chapters, integrating non-geometric features (scale and material) into a shared latent space can immediately boost the practical relevance of part-similarity retrieval. Looking further ahead, a fully realized LMM might track end-to-end life cycles of complex assemblies, identifying common design bottlenecks or anticipating supply chain disruptions based on historical patterns.

6.3 Conclusion and Industry Implications

The work presented in this thesis underscores the transformative potential of multimodal AI for manufacturing. We have shown that geometry-only methods, though valuable, fail to capture essential real-world factors such as material choice and part scale. By broadening the latent representation to include these attributes, retrieval and cost-estimation systems become more aligned with actual shop-floor constraints, making them far more useful to engineers and planners. This multimodal vision naturally extends toward large, unified models that, when properly integrated into existing infrastructures, could reshape the way companies design parts, plan production, and respond to shifting operational demands.

In practical terms, any successful path forward will hinge on robust data engineering, capable computing infrastructures, and organizational willingness to trust (and refine) AI-driven insights. While many obstacles stand in the way—particularly around data silos, interpretability, and domain knowledge capture—the steady convergence of digitalization, sensor technologies, and advanced analytics is driving manufacturers closer to a future where intelligent systems support nearly every stage of production. Embracing these developments can yield not only efficiency gains but also foster higher innovation, providing a strategic edge in competitive markets.

Thus, multimodal AI—culminating in the prospect of full-scale LMMs—offers a compelling route to modernize manufacturing, uniting geometric, material, and operational data in a single tapestry of learning. As this field evolves, the main challenge is to balance sophistication with implementability, ensuring that even complex models provide actionable, transparent value in daily industrial contexts.

Bibliography

[1] W. Hopp and M. Spearman, Factory Physics, Waveland Press, 2011.

[2] Deloitte, 2024 Manufacturing Industry Outlook, 2024.

[3] Tik, "Fiverr," [Online]. Available: https://www.fiverr.com/taklasawiris/redraw-technical-drawings-from-pdf-or-pictures-to-cad. [Accessed April 2025].

[4] Greenwood Aerospace, Government Contract Solutions for Modern Aerospace Challenges, 2024.

[5] Deloitte, 2024 Aerospace and Defense Industry Outlook, Deloitte Insights, 2023.

[6] V. D'Silva and B. Lawler, What Makes a Company Successful at Using AI, 2022.

[7] H. Zhao and et al., Point Transformer, 2021.

[8] I. Midra, R. Girdhar and A. Joulin, An End-to-End Transformer Model for 3D Object Detection, ICCV, 2021.

[9] A. Alquod, D. Schaefer and J. Millisavljevic-Syed, "Industry 4.0: a systematic review of legacy manufacturing system digital retrofitting," *Manufacturing Review,* vol. 9, no. 32, 2022.

[10] M. Mazzetto and e. al., "Deep Learning Models for Visual Inspection on Automotive Assembling Line," *International Journal of Advanced Engineering Research and Science ,* 2020.

[11] D. Wu and e. al., "Cloud-based machine learning for predictive analytics: Tool wear prediction in milling," in *2016 IEEE International Conference on Big Data (Big Data)*, Washington DC, USA, 2016.

[12] P. Denno, C. Dickerson and J. A. Harding, "Dynamic production system identification for smart manufacturing systems," *Journal of Manufacturing Systems,* vol. 48, pp. 192-203, 2018.

[13] MarketsAndMarkets, "Artificial Intelligence in Manufacturing Market Size, Share & Industry Growth Analysis Report by Offering (Hardware, Software, Services), Technology (Machine Learning, Natural Language Processing), Application(Predictive Maintenance & Machinery Inspection," MarketsAndMarkets, September 2023. [Online]. Available: https://www.marketsandmarkets.com/Market-Reports/artificial-intelligence-manufacturing-market-72679105.html. [Accessed 2025].

[14] A. Rahimi, M. Anvaripour and K. Hayat, "Object Detection using Deep Learning in a Manufacturing Plant to Improve Manual Inspection," in *2021 IEEE International Conference on Prognostics and Health Management (ICPHM).*, Detroit (Romulus), MI, USA, 2021.

[15] J. Vater, M. Kirschning and A. Knoll, "Closing the loop: Real-time Error Detection and Correction in automotive production using Edge-/Cloud-Architecture and a CNN," in *2020 International Conference on Omni-layer Intelligent Systems (COINS)*, Barcelona, Spain, 2020.

[16] E. Peretz-Andersson, S. Tabares and P. Mikalef, "Artificial intelligence implementation in manufacturing SMEs: A resource orchestration approach," *International Journal of Information Management,* vol. 77, no. 1, 2024.

[17] C. R. Qi, H. Su, K. Mo and L. J. Guibas, "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation," in *Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 2017.

[18] C. R. Qi, H. Su, K. Mo and L. J. Guibas, "PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space," in *Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 2017.

[19] A. P. 2021, "To Align Parts Using Point Clouds," AutoDesk, 2021. [Online]. Available: https://help.autodesk.com/view/PWRI/2021/CHS/?guid=GUID-754BE9BB-D8EE-4E40-9D12-7A6D700560EB.

[20] R.-S. Mei, C. H. Conway, M. V. Bimrose, W. P. King and C. Shao, "Deep learning of 3D point clouds for detecting geometric defects in gears," *Manufacturing Letters,* vol. 41, pp. 1324-1333, 2024.

[21] Y. Guo, H. Wang, Q. Hu, H. Liu and M. Bennanmoun, "Deep Learning for 3D Point Clouds: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 43, no. 12, 2021.

[22] A. Rani, D. Ortiz-Arroyo and P. Dirdevic, "Advancements in Point Cloud-Based 3D Defect Detection and Classification for Industrial Systems: A Comprehensive Survey," in *Computer Vision and Pattern Recognition*, Seattle, Washington, USA, 2024.

[23] D. Maturana and S. Scherer, "VoxNet: A 3D Convolutional Neural Network for real-time object recognition," in *EEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Hamburg, Germany, 2015.

[24] Z. Zhang, P. Jaiswal and R. Rai, "FeatureNet: Machining feature recognition based on 3D Convolution Neural Network," *Computer-Aided Design ,* vol. 101, pp. 12-22, 2018.

[25] P. Shi and e. al., "Highly interacting machining feature recognition via small sample learning," *Robotics and Computer-Integrated Manufacturing,* vol. 73, pp. 102-260, 2022.

[26] J. Chen, Y. Zeng, S. Wang, S. L. Min and G.-B. Huang, "Octree-based Convolutional Autoencoder Extreme Learning Machine for 3D Shape Classification," in *2018 International Joint Conference on Neural Networks (IJCNN)*, Rio de Janeiro, Brazil, 2018.

[27] H. Su, S. Maji, E. Kalogerakis and E. Learned-Miller, "Multi-View Convolutional Neural Networks for 3D Shape Recognition," in *IEEE International Conference on Computer Vision (ICCV)*, 2015.

[28] P. Shi and e. al., "A novel learning-based feature recognition method using multiple sectional view representation," *Journal of Intelligent Manufacturing,* vol. 31, no. 5, p. 1291–1309, June 2020.

[29] Y. Wu and e. al., "Realtime Single-Shot Refinement Neural Network With Adaptive Receptive Field for 3D Object Detection From LiDAR Point Cloud," *IEEE Sensors Journal,* vol. 21, no. 21, p. 24505–24519, 2021.

[30] D.-Y. Chen and e. al., "On Visual Similarity Based 3D Model Retrieval," *Computer Graphics Forum,* vol. 22, no. 3, p. 223–232, 2003.

[31] ImageNet, "https://www.image-net.org/index.php (visited on 02/21/2025).," [Online].

[32] McGill, "Projections and Views," McGill, 2023. [Online]. Available: https://www.mcgill.ca/engineeringdesign/step-step-design-process/basics-graphics-communication/projections-and-views.

[33] . G. Lambourne, . D. Willis, . K. Jayaraman, . Sanghi, P. Meltzer and H. Shayani, "BRepNet: A topological message passing system for solid models," in *Computer Vision and Pattern Recognition* , 2021.

[34] P. K. Jayaraman, A. Sanghi, J. Lambourne, T. Davies, H. Shayani and N. Morris, "Uv-net: Learning from curve-networks and solids".

[35] W. Cao, T. Robinson, Y. Hua, F. Boussuge, A. R. Colligan and W. Pan, "Graph representation of 3d cad models for machining feature recognition with deep learning," in *ASME 2020 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, 2020.

[36] K. D. D. Willis, Y. Pu, J. Luo, H. Chu, T. Du, J. G. Lambourne, A. Solar-Lezama and W. Matusik, "Fusion 360 Gallery: A Dataset and Environment for Programmatic CAD Construction from Human Design Sequences," *ACM Transactions on Graphics (TOG),* vol. 40, no. 4, 2021.

[37] K. Chen, C. Zheng, Y. Jiang and e. al., "Text2Shape: Generating Shapes from Natural Language by Learning Joint Embeddings," in *Asian Conference on Computer Vision (ACCV)*, 2019.

[38] Z. Han, X. Zhu, Z. Li and e. al., "View2Seq: 3D Shape and Text Joint Embedding Learning for Cross-Modal Retrieval," in *AAAI Conference on Artificial Intelligence*, 2019.

[39] C. Tang, X. Yang, B. Wu, Z. Han and Y. Chang, "Parts2Words: Learning Joint Embedding of Point Clouds and Texts by Bidirectional Matching Between Parts and Words," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2023.

[40] Y. Liu, C. Zhang, X. Dong and J. Ning, "Point Cloud-Based Deep Learning in Industrial Production: A Survey," *ACM Computing Surveys,* vol. 57, no. 12, 2025.

[41] M. T. Khan, W. Feng, L. Chen, Y. H. Ng, N. Y. J. Tan and S. K. Moon, "AUTOMATIC FEATURE RECOGNITION AND DIMENSIONAL ATTRIBUTES EXTRACTION FROM CAD MODELS FOR HYBRID ADDITIVE-SUBTRACTIVE MANUFACTURING," in *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference IDETC/CIE2024*, Washington, DC, August 25-28, 2024.

[42] R. Wu, C. Xiao and C. Zheng, "DeepCAD: A Deep Generative Network for Computer-Aided Design Models," in *IEEE/CVF Int. Conf. on Computer Vision (ICCV)*, 2021.

[43] J. Xu, C. Wang, Z. Zhao, W. Liu, Y. Ma and S. Gao, "CAD-MLLM: Unifying Multimodality-Conditioned CAD Generation With MLLM," 2025.

[44] M. Regenwetter, T. Tomczak, D. Cormier and A. Piper, "Deep Generative Models in Engineering Design: A Review,," *arXiv:2206.07699,* 2022.

[45] L. Makatura, M. Foshey, B. Wang, F. Hähnlein, P. Ma, B. Deng, M. Tjandrasuwita, A. Spielberg, C. E. Owens, P. Y. Chen, A. Zhao, A. Zhu, W. J. Norton, E. Gu, J. Jacob, Y. Li and Sc, "Large Language Models for Design and Manufacturing," *MIT Exploration of Generative AI,* 2024.

[46] M. Ghahramani, Y. Qiao, M. Zhou, A. OHagan and J. Sweeney, "AI-based Modeling and Data-driven Evaluation for Smart Manufacturing Processes," *arXiv preprint,* 2020.

[47] K. D. D. Willis, P. K. Jayaraman, H. Chu, Y. Tian, Y. Li, D. Grandi, A. Sanghi, L. Tran, J. G. Lambourne, A. Solar-Lezama and W. Matusik, "JoinABLe: Learning Bottom-up Assembly of Parametric CAD Joints," *arXiv,* 2022.

[48] R. Rastogi, "Papers Explained Review 11: Auto Encoders," Medium, 30 December 2024. [Online]. Available: https://ritvik19.medium.com/papers-explained-review-11-auto-encoders-3b8f08b4eac0.

[49] D. Bregmann and C. Stryker, "What is a variational autoencoder?," IBM, 12 June 2024. [Online]. Available: https://www.ibm.com/think/topics/variational-autoencoder.

[50] G. H. de Rosa and J. P. Papa, "Optimum-Path Forest," in *Learning to weight similarity measures with Siamese networks: a case study on optimum-path forest☆* , The Associated Press, 2022.

[51] P. I. Frazier, "A Tutorial on Bayesian Optimization," *arXiv:1807.02811,* 2018.

[52] MikeSuperman, "SHapley Additive exPlanations or SHAP : What is it ?," Data Scientest , 9 March 2023. [Online]. Available: https://datascientest.com/en/shap-what-is-it.

[53] D. Bregmann, "What is reinforcement learning from human feedback (RLHF)?," IBM, 10 November 2023. [Online]. Available: https://www.ibm.com/think/topics/rlhf.

[54] E. Trabelsi, "Comprehensive Guide To Approximate Nearest Neighbors Algorithms," Medium, 14 February 2020. [Online]. Available: https://medium.com/data-

science/comprehensive-guide-to-approximate-nearest-neighbors-algorithms-8b94f057d6b6.

[55] M. Douze, A. Guzhva, C. Deng, J. Johnson, G. Szilvasy, P.-E. Mazaré, M. Lomeli, L. Hosseini and H. Jégou, "The Faiss library," *arXiv2401.08281,* 2024.

[56] F. Feng and S. R, "What is Annoy?," Zilliz, 6 January 2025. [Online]. Available: https://zilliz.com/learn/what-is-annoy.

[57] R. Winastwan, "Understanding HNSWlib: A Graph-based Library for Fast Approximate Nearest Neighbor Search," Zilliz, 29 October 2024. [Online]. Available: https://zilliz.com/learn/learn-hnswlib-graph-based-library-for-fast-anns.