An Agentic LLM with Mechanical Knowledge

Shuang Li, Jonathan Corney*
Erdzan Hodzic, David Eklund#
Rixin Yu, Ralf Denzer^

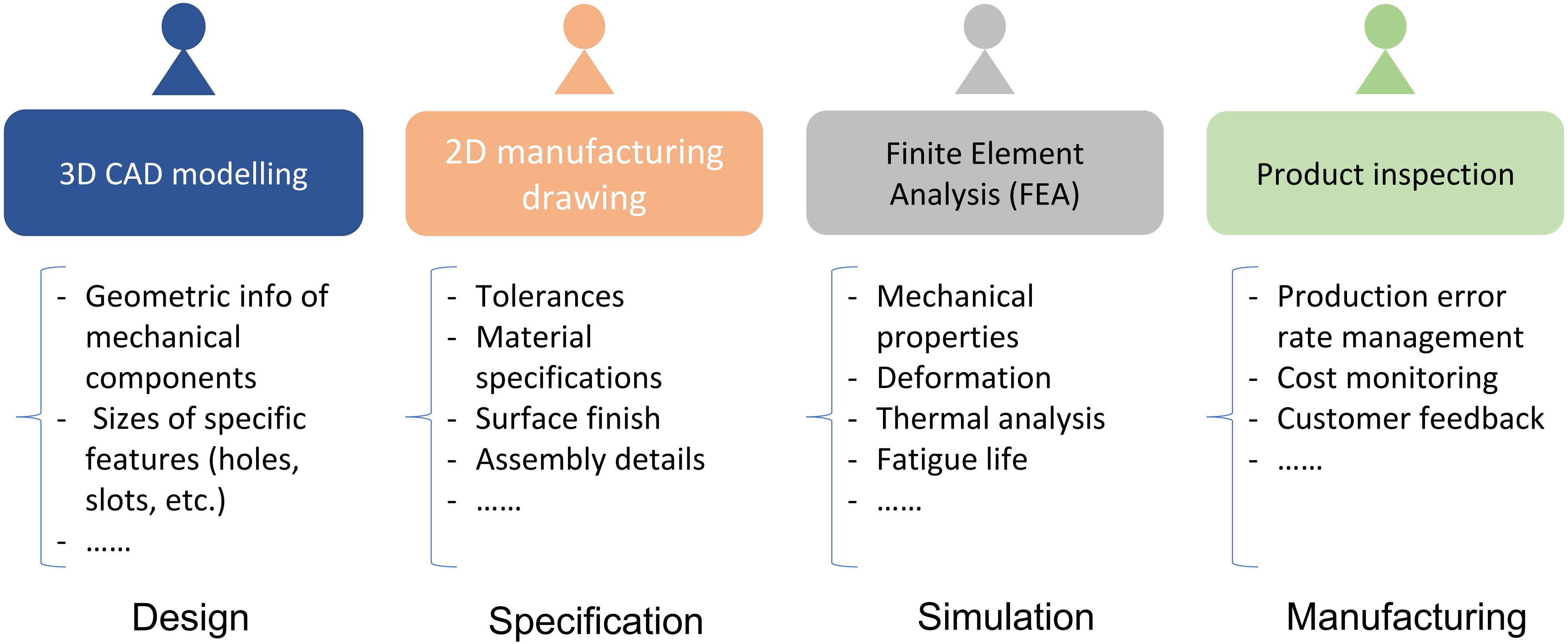
Presented by Dominique Regli
05/2025, MIT

J.R.Corney@ed.ac.uk*
THE UNIVERSITY *of* EDINBURGH
School of Engineering

Research Institutes of Sweden #

LUND UNIVERSITY
LTH FACULTY OF ENGINEERING ^

Challenge in Technical Design: CAE Software forms distributed silos of information and knowledge



3D CAD modelling

- Geometric info of mechanical components
- Sizes of specific features (holes, slots, etc.)
- ……

Design

2D manufacturing drawing

- Tolerances
- Material specifications
- Surface finish
- Assembly details
- ……

Specification

Finite Element Analysis (FEA)

- Mechanical properties
- Deformation
- Thermal analysis
- Fatigue life
- ……

Simulation

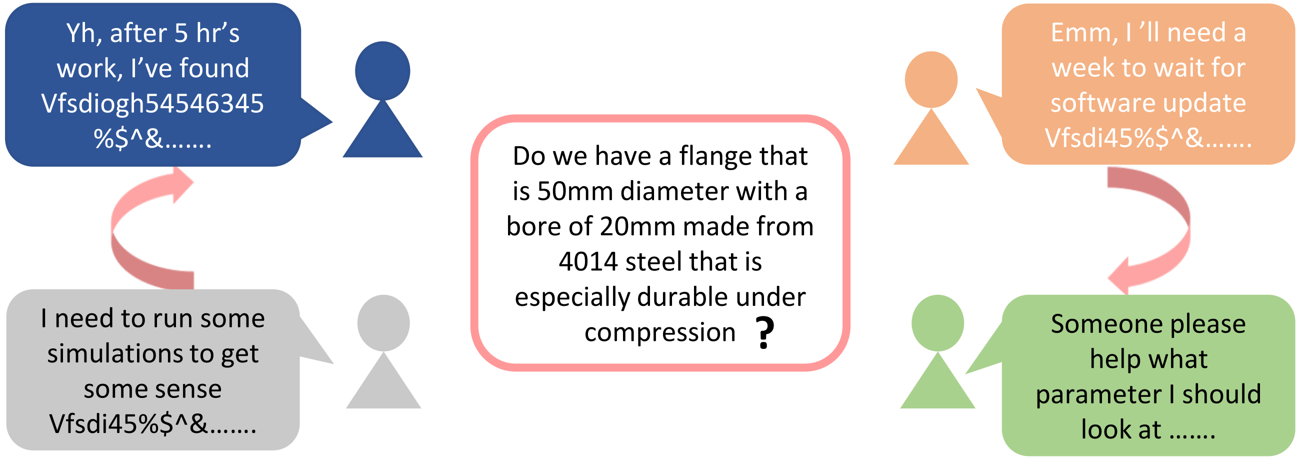
Product inspection

- Production error rate management
- Cost monitoring
- Customer feedback
- ……

Manufacturing

Challenge in Technical Design: CAE Software forms distributed silos of information and knowledge

Current Practice: The answers to technical questions frequently requires information from multiple silos. Current solutions are facilitated by forms of team chat



Yh, after 5 hr's work, I've found Vfsdiogh54546345 %$^&…….

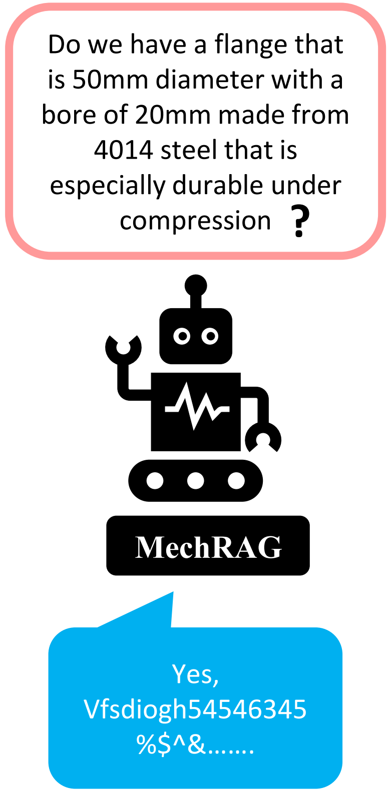
I need to run some simulations to get some sense Vfsdi45%$^&…….

Do we have a flange that is 50mm diameter with a bore of 20mm made from 4014 steel that is especially durable under compression ?

Emm, I 'll need a week to wait for software update Vfsdi45%$^&…….

Someone please help what parameter I should look at …….

Research Vision: AI responds to questions using holistic knowledge from multiple silos



Do we have a flange that is 50mm diameter with a bore of 20mm made from 4014 steel that is especially durable under compression ?

Yes, Vfsdiogh54546345 %$^&…….

Classifying Technical Design Questions

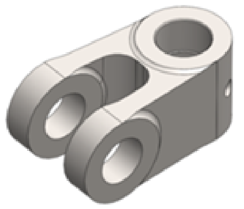
Level	Type	Illustrative Examples of MechRAG Prompts
1	Direct	Query the return values from read or computed from a CAE data files • What is the volume of part 00210008 in cm^3? • How many round holes does model 00210337 contain? • Does part 00210217 have any type of symmetry?
2	Collective	Query combines level 1 information from multiple CAE data structures • Which parts have the biggest volume and the smallest? • What symmetries are present in the dataset of components?
3	Emergent	Query requires response which combines information from multiple sources in ways not explicitly programmed • Based on part 00217697's modelling history which other components is most **similar** to it? • Which of 0021* series of components would be the most expensive to produce and why? • What manufacturing processes is used to produce part 00210215?

Multi-silo conversational interface for data queries

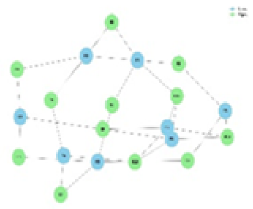
Multi-silo, multi-design conversational interface for comparative queries

Subjective queries that require product and general knowledge

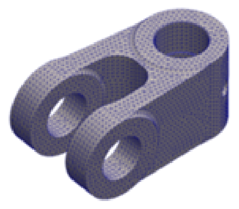
Tokenising CAE Data



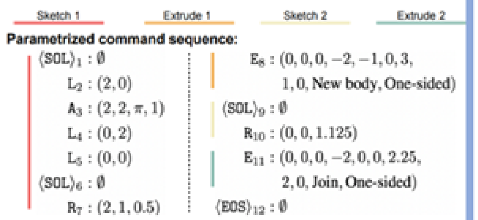
3D CAD model



Boundary Representation (B-rep)



Mesh



Sketch 1
Extrude 1
Sketch 2
Extrude 2

Parametrized command sequence:

$\langle \mathtt{SOL} \rangle_1 : \emptyset$
$\mathtt{L}_2 : (2, 0)$
$\mathtt{A}_3 : (2, 2, \pi, 1)$
$\mathtt{L}_4 : (0, 2)$
$\mathtt{L}_5 : (0, 0)$
$\langle \mathtt{SOL} \rangle_6 : \emptyset$
$\mathtt{R}_7 : (2, 1, 0.5)$
$\mathtt{E}_8 : (0, 0, 0, -2, -1, 0, 3, 1, 0, \text{New body}, \text{One-sided})$
$\langle \mathtt{SOL} \rangle_9 : \emptyset$
$\mathtt{R}_{10} : (0, 0, 1.125)$
$\mathtt{E}_{11} : (0, 0, 0, -2, 0, 0, 2.25, 2, 0, \text{Join}, \text{One-sided})$
$\langle \mathtt{EOS} \rangle_{12} : \emptyset$

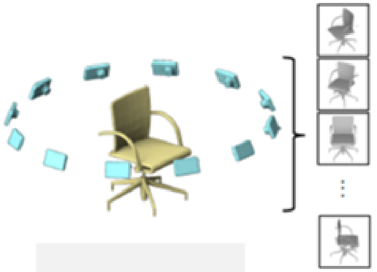
Modelling history
(Command sequences)

Explicit information

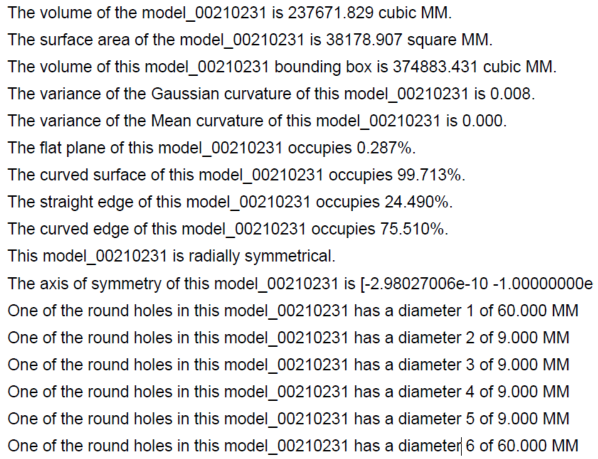
- **Geometric Properties:** volume, surface area, bounding dimensions
- **Topological Properties:** number of faces, edges and vertices
- **Feature Dimensions:** hole diameters, channel & plate thickness
- **Mass Properties:** centre of mass, moment of inertia
- **Curvature Analysis:** curvature of non-planar surfaces
- **Connectivity & Adjacency:** relations between faces, edges & vertices
- **Orientation Information:** surface normal, edge directions

Implicit Information

- **Type of mechanical component:** clevis bracket, nuts, flange, etc.
- **Manufacturing processes:** forging, machining, etc.
- **Shape features:** fillet, chamfers, bosses, ribs, etc.
- **Multi-view images:**



CAE data is typically numerical, so it needs to be processed to provide **summaries** in LLM compatible formats (e.g. images and text).

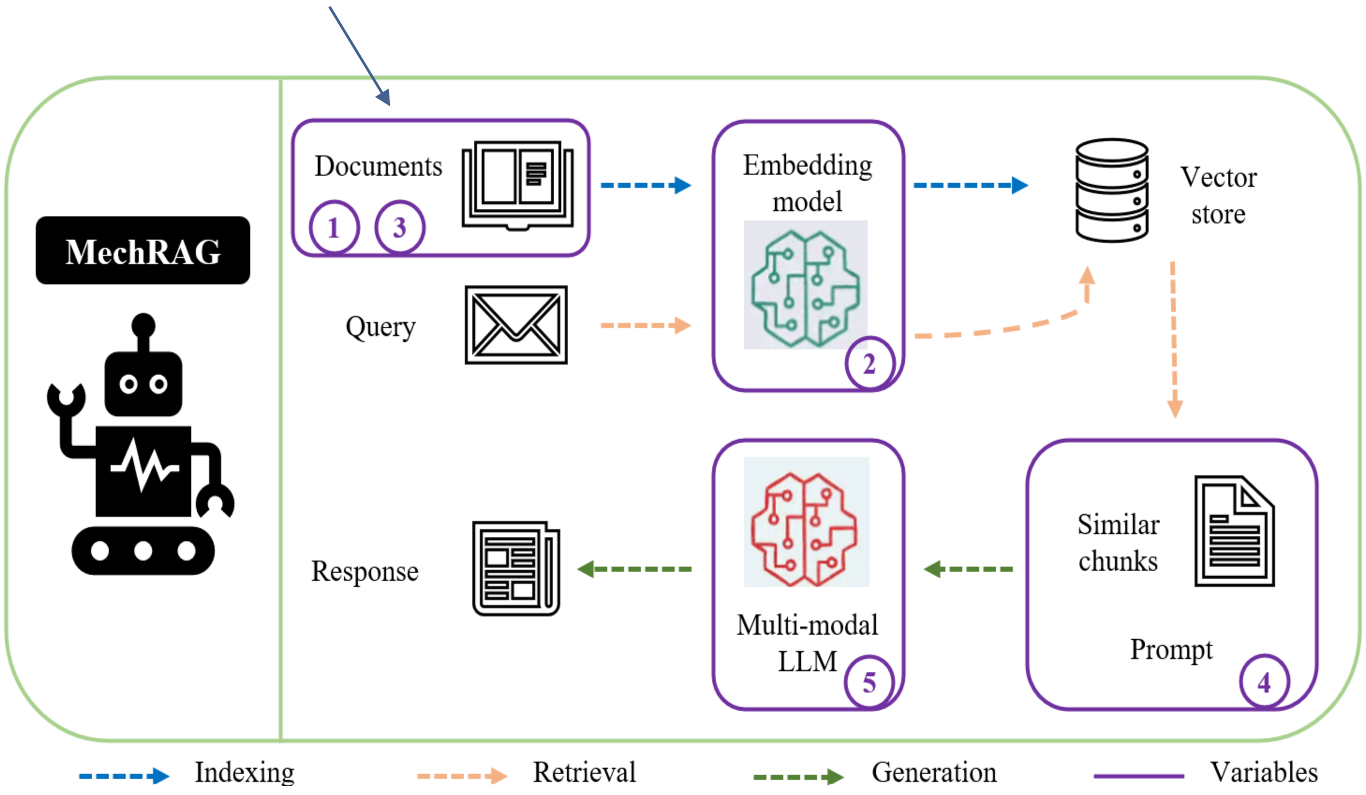


```
The volume of the model_00210231 is 237671.829 cubic MM.
The surface area of the model_00210231 is 38178.907 square MM.
The volume of this model_00210231 bounding box is 374883.431 cubic MM.
The variance of the Gaussian curvature of this model_00210231 is 0.008.
The variance of the Mean curvature of this model_00210231 is 0.000.
The flat plane of this model_00210231 occupies 0.287%.
The curved surface of this model_00210231 occupies 99.713%.
The straight edge of this model_00210231 occupies 24.490%.
The curved edge of this model_00210231 occupies 75.510%.
This model_00210231 is radially symmetrical.
The axis of symmetry of this model_00210231 is [-2.98027006e-10 -1.00000000e
One of the round holes in this model_00210231 has a diameter 1 of 60.000 MM
One of the round holes in this model_00210231 has a diameter 2 of 9.000 MM
One of the round holes in this model_00210231 has a diameter 3 of 9.000 MM
One of the round holes in this model_00210231 has a diameter 4 of 9.000 MM
One of the round holes in this model_00210231 has a diameter 5 of 9.000 MM
One of the round holes in this model_00210231 has a diameter 6 of 60.000 MM
```

Fragment of a **text summary** that combines information from different CAE representations (e.g. BRep, Mesh, Design History etc).

MechRAG architecture

Multi-modal summary of CAE data



MechRAG

Documents ① ③

Query

Embedding model ②

Vector store

Similar chunks

Prompt ④

Multi-modal LLM ⑤

Response

Indexing

Retrieval

Generation

Variables

- Variable 1: Embedding method
- Variable 2: Embedding model
- Variable 3: Indexing ratio
- Variable 4: Prompting strategy
- Variable 5: LLM platform

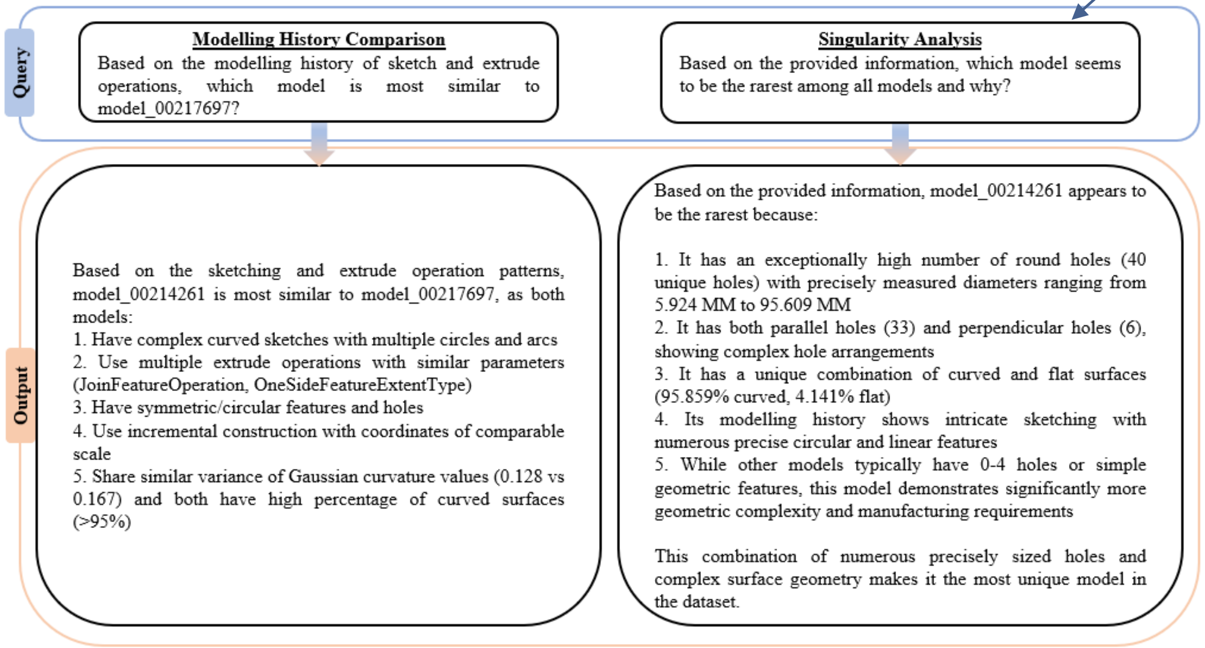
The **Retrieval-Augmented Generation (RAG)** mechanism enhances large language models (LLMs) by combining their generation capabilities with external information retrieval. Instead of relying solely on their internal training data, RAG models first retrieve relevant documents or knowledge snippets from an external database or knowledge base based on the input query. They then use this retrieved information to ground and inform their generated responses, leading to more accurate, up-to-date, and contextually relevant outputs without the need for retraining the LLM itself.

THE UNIVERSITY *of* EDINBURGH
School of Engineering

MechRAG example generations

ChatGPT Prompts

Ambiguous prompts: “most similar” & “rarest”


Query
Output

Modelling History Comparison

Based on the modelling history of sketch and extrude operations, which model is most similar to model_00217697?

Singularity Analysis

Based on the provided information, which model seems to be the rarest among all models and why?

Based on the sketching and extrude operation patterns, model_00214261 is most similar to model_00217697, as both models:
1. Have complex curved sketches with multiple circles and arcs
2. Use multiple extrude operations with similar parameters (JoinFeatureOperation, OneSideFeatureExtentType)
3. Have symmetric/circular features and holes
4. Use incremental construction with coordinates of comparable scale
5. Share similar variance of Gaussian curvature values (0.128 vs 0.167) and both have high percentage of curved surfaces (>95%)

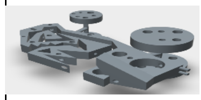
Based on the provided information, model_00214261 appears to be the rarest because:

1. It has an exceptionally high number of round holes (40 unique holes) with precisely measured diameters ranging from 5.924 MM to 95.609 MM
2. It has both parallel holes (33) and perpendicular holes (6), showing complex hole arrangements
3. It has a unique combination of curved and flat surfaces (95.859% curved, 4.141% flat)
4. Its modelling history shows intricate sketching with numerous precise circular and linear features
5. While other models typically have 0-4 holes or simple geometric features, this model demonstrates significantly more geometric complexity and manufacturing requirements

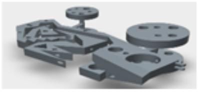
This combination of numerous precisely sized holes and complex surface geometry makes it the most unique model in the dataset.

00217697

00214261

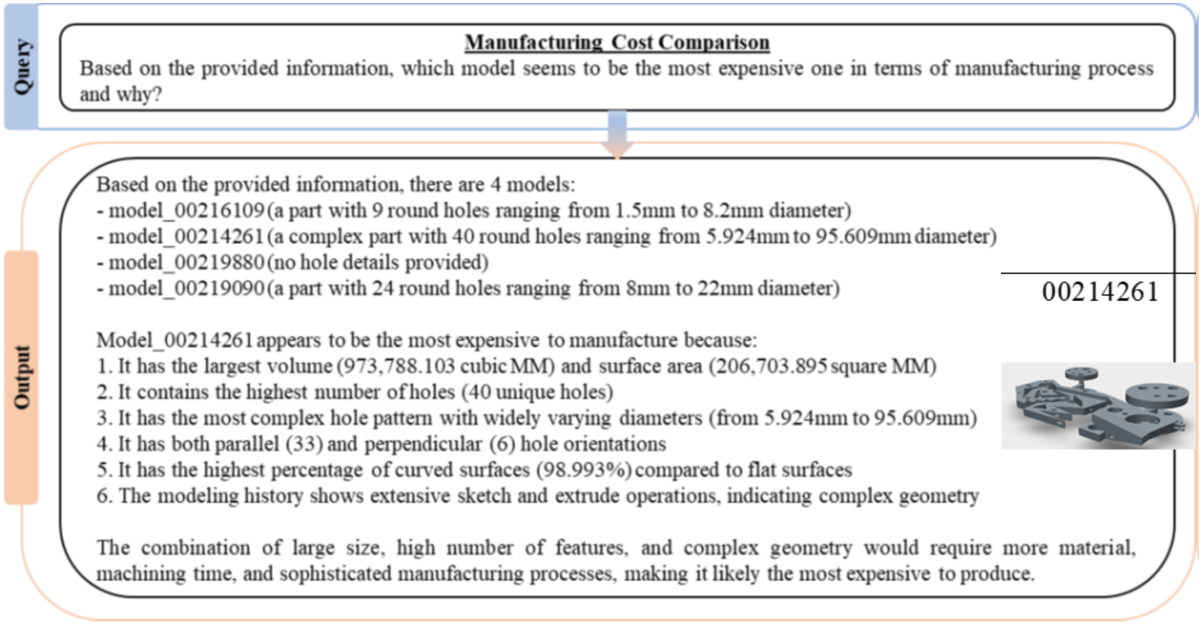


00214261



ChatGPT responds with mix of CAE data (via RAG) and general knowledge

Ambiguous prompts: “most expensive to manufacture”


Query
Output
00214261

Manufacturing Cost Comparison

Based on the provided information, which model seems to be the most expensive one in terms of manufacturing process and why?

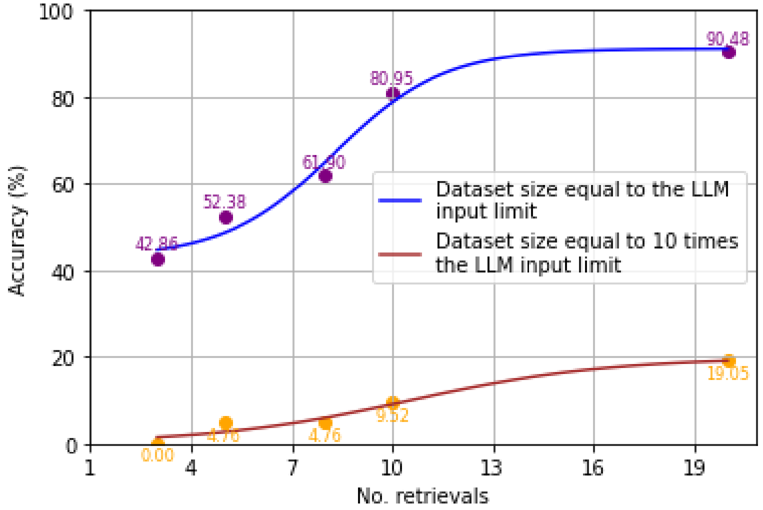
Based on the provided information, there are 4 models:
- model_00216109 (a part with 9 round holes ranging from 1.5mm to 8.2mm diameter)
- model_00214261 (a complex part with 40 round holes ranging from 5.924mm to 95.609mm diameter)
- model_00219880 (no hole details provided)
- model_00219090 (a part with 24 round holes ranging from 8mm to 22mm diameter)

Model_00214261 appears to be the most expensive to manufacture because:
1. It has the largest volume (973,788.103 cubic MM) and surface area (206,703.895 square MM)
2. It contains the highest number of holes (40 unique holes)
3. It has the most complex hole pattern with widely varying diameters (from 5.924mm to 95.609mm)
4. It has both parallel (33) and perpendicular (6) hole orientations
5. It has the highest percentage of curved surfaces (98.993%) compared to flat surfaces
6. The modeling history shows extensive sketch and extrude operations, indicating complex geometry

The combination of large size, high number of features, and complex geometry would require more material, machining time, and sophisticated manufacturing processes, making it likely the most expensive to produce.

Future Work

- **Token limits of the context window in LLMs**: it’s not unusual for engineering companies have tens of thousands of CAD models, but it’s possible to exceed an LLMs token limit with less than a hundred models. Consequently work is needed to minimise the number of tokens required to represent engineering product knowledge.



- **How to utilise other modalities/formats of data** (e.g. FEA/CFD simulation and process) to augment LLMs?

Thanks for your attention!