

AI+CAD 评测系列技术白皮书

通用大模型的 CAD 能力评测报告 ——以 GPT-6 Astra 为例

邹强 沈佳军 杨智杰 王禹皓 冷石天一 尹东硕

浙江大学 DMA 实验室

2026 年 10 月

编写说明

本报告由浙江大学 DMA（Design and Manufacturing Automation Lab）实验室编写。

团队简介

DMA 实验室隶属于浙江大学 CAD&CG 全国重点实验室，主要从事几何计算、智能 CAD 研究及 CAD/CAM 软件研发。

主页（中文）：<https://person.zju.edu.cn/qiangzou>

主页（英文）：<https://qiang-zou.github.io/>

作者贡献

邹强：报告大纲与测试大纲设计、测试结果综合分析、报告撰写及最终审定。

沈佳军：实验实施、报告撰写及统稿。

杨智杰：报告大纲设计、实验实施与报告撰写。

王禹皓、冷石天一、尹东硕：实验实施与报告撰写。

前言

人工智能技术的快速发展正在重塑多个领域的技术范式。在软件开发、科学研究等领域，大模型开始承担原本依赖专门方法与流程的工作，改变既有的人机分工。在 CAD 领域，三维理解与建模能力的不足，曾被视为通用大模型难以深入专业任务的重要障碍。然而，随着 GPT-6 Astra 等新一代大模型与智能体技术的发展，一些此前被认为超出通用大模型能力范围的 CAD 任务，已经出现成功实例。这一变化将如何影响 CAD 软件的技术体系与使用方式，需要系统考察。

对这一变化的认识，需要建立在 CAD 自身的专业要求之上。CAD 模型承载设计意图，支撑后续编辑、分析与制造，其质量直接关系到设计方案能否被准确表达和持续使用。通用模型在具体任务中的表现展示了新的可能，但这些能力能够覆盖多大范围、在何种条件下成立、能否可靠满足工程要求，仍需通过实际任务加以辨明。

为此，本报告以 GPT-6 Astra 为评测对象，围绕 CAD 模型从理解、构建到后续修改的主要环节，组织六类典型任务，包括零件分类、特征识别、工程图到 CAD 模型生成、文本到 CAD 模型生成、图像到 CAD 模型生成，以及 B-rep 模型参数化重建。测试在统一的基本环境下开展，并设计了不同复杂程度的单零件与装配体测试案例。能力判断以实际输出的核验为依据：对识别结果检查其正确性与几何精度，对生成结果检查几何与尺寸、实体有效性和结构关系，对参数化重建结果进一步开展编辑验证。报告结合结果验证与执行过程分析，逐项考察各类任务的完成情况，再综合不同任务中的表现，分析 GPT-6 Astra 可能的推理过程，并总结已经表现出的能力、仍然存在的困难及其可能原因。

本报告面向 CAD 研究者、软件开发者和工程技术人员，旨在为认识通用大模型的 CAD 能力及其局限、审视 CAD 技术体系的变化提供实证依据并结合能力进展与失败案例，重新审视 CAD 研究问题价值。报告反映特定系统、样本与测试条件下的阶段性认识，相关结论仍需在更广泛的任务与实际应用中检验。

摘要

本报告以 GPT-6 Astra（下称 GPT-6）为评测对象，围绕零件分类、特征识别、工程图到 CAD 模型生成、文本到 CAD 模型生成、图像到 CAD 模型生成及 B-rep 模型参数化重建六类任务，考察其在不同复杂度和输入条件下的表现。能力判断以实际结果的专业核验为基础，并结合具体案例与可观察的执行过程展开。

一、核心结论

测评表明，通用大模型的 CAD 能力已开始显现工程应用价值，处于“奇点”阶段。与现有研究中“玩具”数据集和模型不同，GPT-6 在一系列真实场景中取得成功：能够为概念设计提供初步方案，为详细设计提供初始模型。然而，GPT-6 在实体有效性、功能完整性和参数可编辑性仍需专业核验，必要时修正或重建，同时较长的任务执行时间也制约着实际可用性。

（1）通用大模型已经能够承担多类 CAD 任务

GPT-6 能够识别已有 B-rep 模型中的零件类别与几何/加工特征，根据工程图、图像或文字描述生成三维模型，也能在部分任务中依据功能需求选择设计方案、确定尺寸并开展分析校核。本次测试已出现较复杂零件特征识别、中等复杂模型生成与重建，以及较复杂装配体构建的成功案例。

（2）生成模型的实体有效性和功能可行性仍不稳定

生成的 CAD 模型大多在外形上接近目标要求，但内部结构、关键尺寸和特征布局仍可能存在错误。这些问题既出现在单个零件中，也出现在装配体中：孔道可能未按要求连通或隔离，零件之间也可能缺少必要的固定和传动连接。部分错误对外形影响很小，却直接影响关键功能的实现。这类错误不能仅归因于形状复杂或输入信息不足：简单模型同样发现错误，即使提供完整的目标几何信息，也不能保证重建结果正确。

（3）建立三维模型后，仍难保证后续参数化修改正确

对于形状简单、特征数量较少、相互依赖关系简单的模型，GPT-6 已经能够建立可编辑的参数化模型，较好地支持尺寸修改及相关特征的联动。对于中等和复杂模型，所建立的特征依赖关系往往不完整，修改尺寸后可能出现孔口封闭、局部特征消失或零件错位，目前尚不能稳定生成具有良好可编辑性的参数化模型。

（4）GPT-6 已能自主检查和修改模型，但纠错能力仍不稳定

GPT-6 能够在建模过程中检查结果、发现部分问题并反复修改，这正是其完成一些复杂任务的核心原因。然而，测试结果发现，迭代过程中误差曲线经常出现波动，部分修改

反而使误差增大。这表明，GPT-6 对误差来源、修改对象及相应操作的判断仍不够准确。同时，几何特征与参数相互依赖，一处修改可能影响多个结构和设计要求。GPT-6 尚不能充分识别这些关联关系，也难以预判和控制修改的连带影响，因而可能在修正原有问题时引入新的偏差。

（5）建模耗时过长，制约实际可用性

GPT-6 在建模过程中往往需要反复生成、检查和修正模型，带来较大的时间开销。测试中，多类建模与重建任务的平均用时达到数十分钟，部分复杂案例超过 1 小时，最长达 3 小时 41 分钟。较长的执行时间使其难以满足设计过程中快速反馈和频繁修改的需求，尽管建模效果已有进展，当前实际可用性仍然有限。

表 1 GPT-6 在 CAD 任务中的当前能力与主要不足

测评内容	当前能力	主要不足
零件分类	能够结合几何信息与工程知识判断零件类别	细分类型判断仍不够可靠，可能将局部结构的功能误认为零件的整体用途
特征识别	能够识别孔、槽、型腔等加工特征，划分特征实例	复杂区域仍存在实例划分和类型判断错误，特别是对于嵌套、相交等复杂特征
工程图到 CAD 模型重建/生成	能够综合多视图、剖视与尺寸标注，恢复主要外形和内部结构，并完成部分多零件模型的装配	复杂结构层次与关键尺寸仍有误判，典型的有：遗漏局部特征或造成装配干涉
文本到 CAD 模型生成	能够按几何描述构造模型，依据功能需求选择设计方案，并完成建模与分析计算	复杂几何特征、孔道连通、局部过渡、零件固定及传动连接仍经常出错，功能要求与实体有效性尚不能稳定保证
图像到 CAD 模型重建/生成	能够根据多视图图像恢复主体轮廓、主要特征及零件布局	复杂曲面与连接过渡常被简化，实体连接及装配接口仍有偏差，无尺寸参照时难以确定实际尺度
B-rep 模型参数化重建	能够恢复简单和中等复杂模型的主体几何，建立参数化模型并支持部分尺寸修改	复杂曲面与内部孔道等几何的重建存在严重偏差，特征依赖关系经常不完整，修改后相关特征无法正确联动
Agentic 迭代优化	能够利用反馈补齐特征、调整尺寸与位置，降低几何误差，提高要求满足程度	仍难准确确定修改对象与范围，多轮迭代后仍可能遗留无效实体、功能缺失和参数修改失败等问题；迭代耗时较长，限制实际可用性

二、分项测评结论

(1) 零件分类

零件分类任务考察 GPT-6 能否依据三维 CAD 模型的几何与拓扑信息，正确判断零件类别。测试选取 12 个模型，覆盖 10 类零件，简单、中等和复杂样本各 4 个，代表样本见图 1。

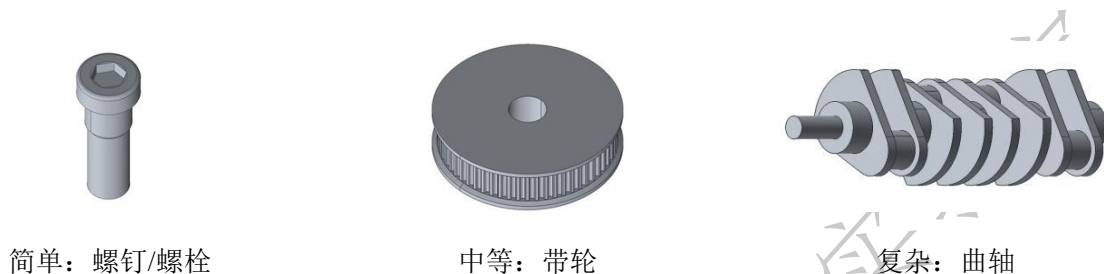
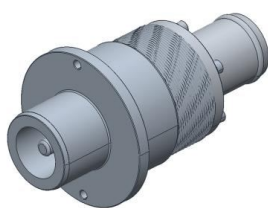


图 1 零件分类任务的不同复杂度样本

测试结果显示，在给定 10 个候选零件类别名称且默认开启网络搜索功能的条件下，GPT-6 对 12 个样本全部分类正确，准确率为 100%。简单、中等和复杂样本的平均用时分别为 2 分 30 秒、2 分钟和 2 分 30 秒，单个样本最长用时为 4 分钟。在不提供零件类别名称或关闭网络搜索功能的条件下，各自选取了 2 个代表性复杂样本进行测试，分类结果仍全部正确。

测试中出现的典型错误是：在不提供零件类别名称且关闭网络搜索的条件下，GPT-6 将管路快速接头误判为止回阀（单向阀），见图 2。GPT-6 虽识别出了内部的弹簧、阀杆和阀瓣，但将阀结构的功能等同于整个接头的用途，导致整体类别误判。



实际零件：管路快速接头
GPT-6 判断：止回阀（单向阀）

图 2 管路快速接头的类别误判

总体而言，GPT-6 目前的能力是：

- 对于典型零件，GPT-6 能够正确识别零件类别，还能指出具体细分类型和功能用途；
- 对于外形不典型或与其他类别相近的零件，GPT-6 能正确识别零件类别，但不一定能判断零件的具体细分类型。
- 对于装配体，GPT-6 可能将局部结构功能误认为整体用途，导致零件类别误判。

（2）特征识别

特征识别任务考察 GPT-6 能否从 B-rep 模型中自主划分加工特征实例、判断类别与子类型，并识别特征之间的相互关系（测试时不预设特征类型的候选列表）。测试选取简单、中等和复杂模型各 2 个。代表样本见图 3。

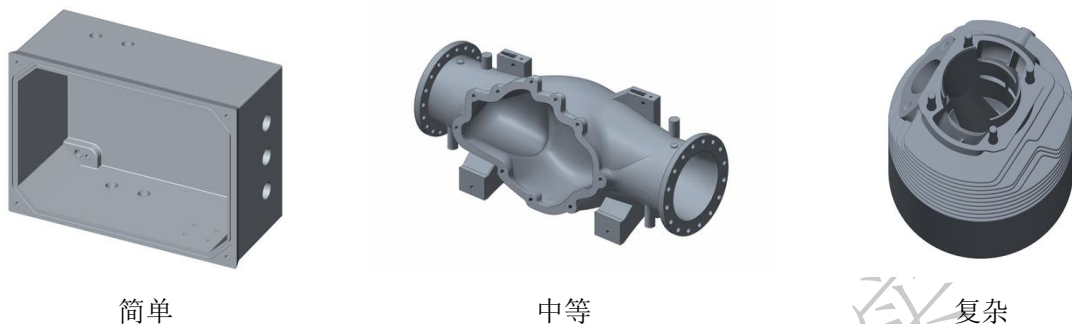
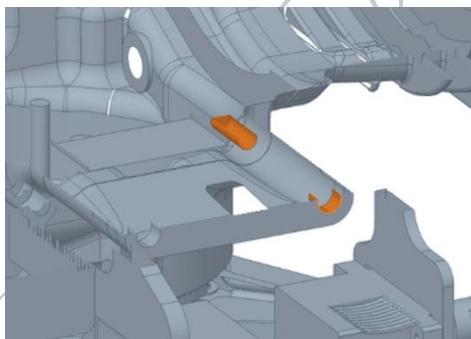


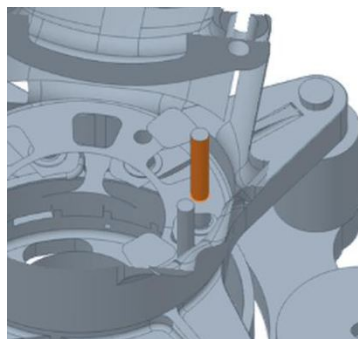
图 3 特征识别任务的不同复杂度样本

测试结果显示，在 6 个测试样例中，3 个样例的特征识别结果正确；2 个样例能正确识别出特征，但还会额外将一些非特征面组也标记为特征；剩下 1 个样例大部分特征识别正确，但存在少量错误。简单、中等和复杂样本的平均用时分别为 30 分 30 秒、40 分 30 秒和 1 小时 25 分钟，单个样本最长用时为 2 小时 1 分钟。在有、无网络搜索两种条件下，同一样本的识别结果接近，在特征类型细分情况上（比如孔特征的通盲判断）各有优劣。

测试中出现的典型错误是：GPT-6 会将部分模型的非特征面组误识别为加工特征，导致错误识别了过多特征。对于复杂模型，GPT-6 存在将两个圆孔识别为同一个特征和将圆柱面误判为圆孔等问题，见图 4。



两个圆孔被识别为同一个特征
橙色区域在空间上彼此分离



圆柱面被误识别为圆孔
橙色区域实际位于实体外侧

图 4 复杂模型中的特征实例划分与类别判断错误

总体而言，GPT-6 目前的能力是：

- 对于特征相互独立或仅存在简单嵌套关系的模型，GPT-6 能够准确识别孔、槽、型腔、圆角等全部加工特征。

- 对于包含多层嵌套、相交或截断结构的模型，GPT-6能正确识别大部分加工特征，但存在少量错误：包括（1）对于整体外形复杂的模型，GPT-6容易将非特征的表面面组误识别为特征；（2）对于结构密集、遮挡明显的模型，GPT-6对其内部特征的识别仍存在问题。

（3）工程图到 CAD 模型重建/生成

工程图到 CAD 模型重建/生成任务考察 GPT-6 能否综合多视图、剖视图和尺寸标注，重建零件的三维几何，并使分别生成的零件满足装配要求。测试包含 5 个单零件样本和 3 个多零件样本，简单、中等和复杂样本分别为 2 个、3 个和 3 个，代表样本的参考模型见图 5。

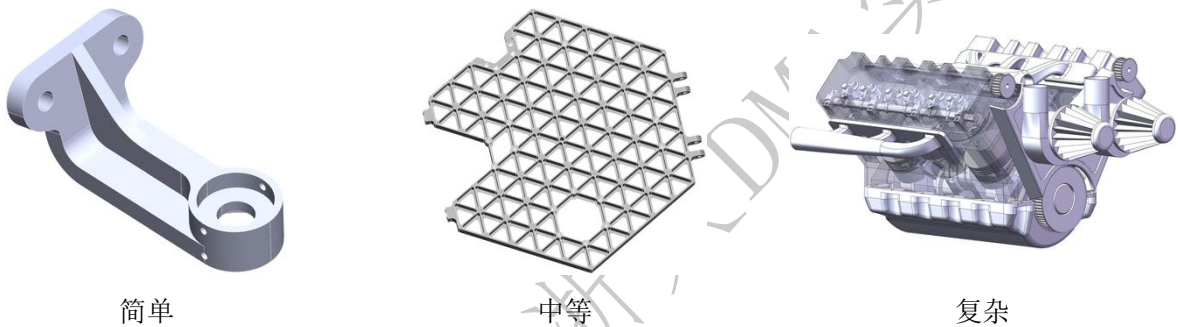


图 5 工程图到 CAD 模型重建/生成任务的不同复杂度样本

测试结果显示，实验的 8 个样本中，6 个样本的全部零件通过实体有效性检测；其余 2 个样本未通过，存在自交、非流形连接等错误。3 个多零件样本中，1 个能够完成装配，1 个存在零件干涉，1 个因结构缺失和尺寸错误无法完成装配。简单、中等和复杂样本的平均用时分别为 26 分 30 秒、43 分 40 秒和 31 分 20 秒，单个样本最长用时为 1 小时 8 分钟。

测试中出现的典型错误是：GPT-6 进行局部结构重建时会出现偏差，以及多零件配合任务存在参数偏差，使得部分配合尺寸不一致，造成零件实体重叠等问题，见图 6。

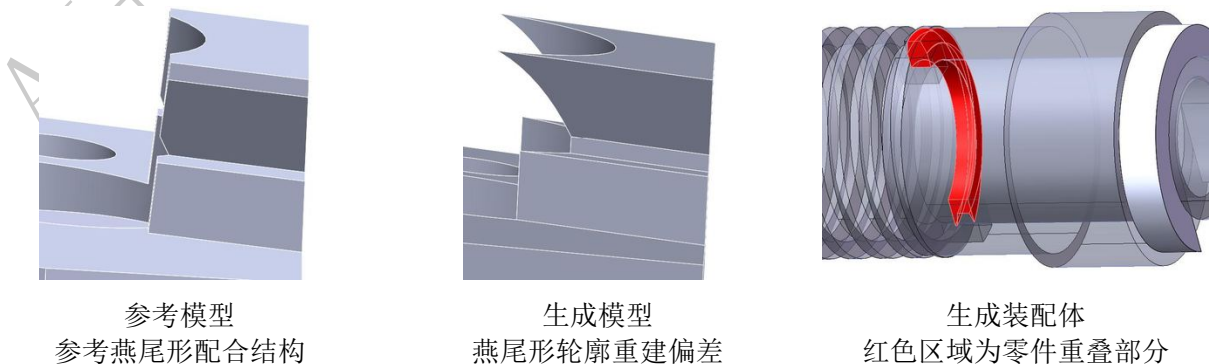


图 6 燕尾形配合结构重建偏差与装配干涉

总体而言，GPT-6 目前的能力是：

- 对于视图关系清晰、以常规孔槽和规则阵列为主的零件，GPT-6能够较好地恢复主体几何；但对于包含多层台阶、嵌套型腔及复杂薄壁过渡的零件，GPT-6仍容易误判结构层次、遗漏局部特征，有时还存在实体无效问题。
- 对于零件较少、配合关系明确的装配体，能够生成可完成装配的零件模型；但对于零件较多、配合要求复杂的装配体，GPT-6难以准确还原关键配合尺寸，容易造成干涉或无法装配。
- 对于拔模、抽壳、放样等CAD操作，GPT-6只会用简单平面去近似，在这方面与传统CAD建模仍存在差距。

（4）文本到 CAD 模型生成

文本到 CAD 模型生成任务考察 GPT-6 能否依据几何要求或功能需求，确定结构方案与尺寸参数，并生成相应的 CAD 模型。测试选取简单、中等和复杂样本各 4 个，另以 2 个复杂装配体进行压力测试，并对 4 个复杂样本开展无网络搜索测试，代表样本的参考模型见图 7。

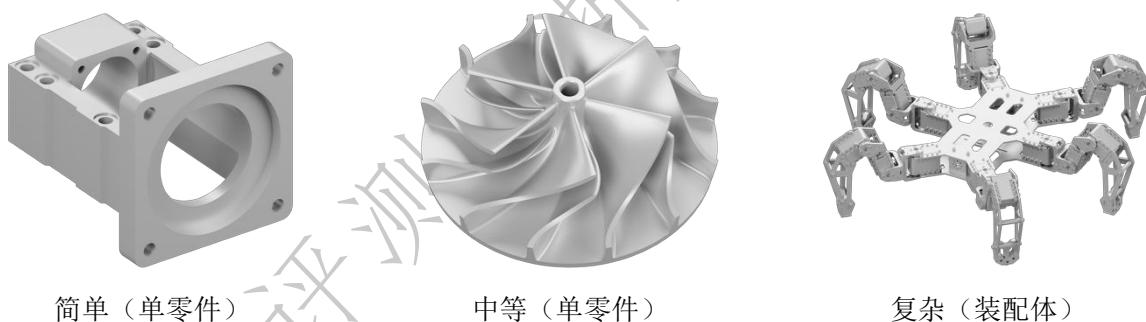
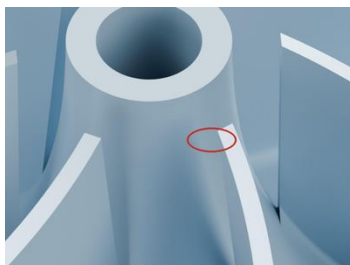


图 7 文本到 CAD 模型生成任务的不同复杂度样本

测试结果显示，12 个样本均满足全部的文本要求，11 个通过实体有效性检测。简单、中等和复杂样本平均用时分别为 42 分钟、1 小时 0 分 15 秒和 1 小时 37 分 30 秒，最长为 3 小时 17 分钟。2 个样本的压力测试中，1 个样本满足全部要求，2 个均未通过实体有效性检测；关闭网络搜索后的 4 个样本测试中，4 个均满足全部要求，2 个通过有效性检测。

测试中出现的典型错误是：GPT-6 会在本应连续连接的部位留下间隙，也会添加不承担支撑、连接或传动功能的多余构件，见图 8。



局部连接不完整
红圈处存在连接间隙



生成多余构件
橙色构件不承担实际功能

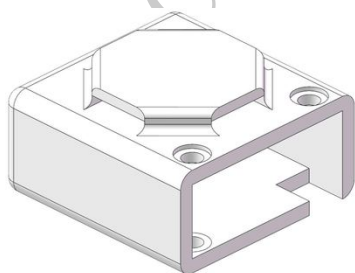
图 8 文本建模中的典型错误

总体而言，GPT-6 目前的能力是：

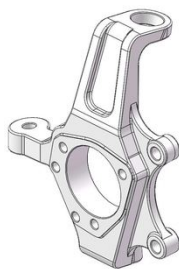
- 对于尺寸和特征位置明确的零件，或载荷、运动及安装要求清晰的机构，GPT-6能够依据文本确定结构方案和尺寸参数，生成单零件及装配体，并在迭代中补齐特征、调整部件位置（12个样本平均迭代4次）。
- 对于需要曲面连续过渡或零件协同运动的模型，GPT-6仍会在应接合处留下间隙，或使传动部件缺少固定、彼此分离，导致运动无法传递。GPT-6还会生成没有实际用途的多余构件，表明其对部件配合和构件用途的判断仍有不足，不能可靠落实文本中隐含的工程要求。
- 对于具有薄壁结构，以及内部孔腔交汇、孔壁贴近外壁的零件，GPT-6仍不能可靠处理曲面形状与相邻面的边界关系，生成结果会出现曲面自交以及非流形连接等问题。

（5）图像到 CAD 模型重建/生成

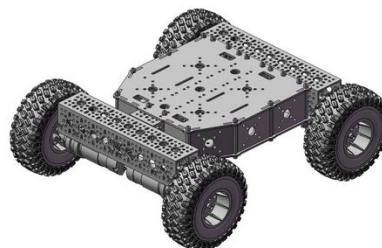
图像到 CAD 模型重建/生成任务以无尺寸标注的多视图渲染图为输入，考察 GPT-6 对三维形状、特征位置及连接关系的恢复能力。测试选取简单、中等和复杂样本各 2 个，另以 1 个复杂壳体开展压力测试，并设置不同视图组合和无网络搜索条件，代表样本见图 9。



简单



中等



复杂

图 9 图像到 CAD 模型重建/生成任务的不同复杂度样本

实验结果显示，21 次测试均生成 STEP 模型，输出中的各个实体均通过实体有效性检查。全部实验中，简单、中等和复杂样本的平均用时分别为 12.5 分钟、约 16.33 分钟和 43 分钟。采用六视图加轴测图时，6 个样本的多视图轮廓平均交并比（IoU）为 0.8790～0.9915，GPT-6 较好地恢复了主体轮廓和部件布局，但局部结构与实际尺寸仍有明显偏差。

测试中出现的典型错误是：对于包含自由曲面、薄壁及筋板交汇的零件，GPT-6 容易将复杂曲面简化为平面或分段曲面，并误判局部特征的方向和连接关系。例如，图 10 中的控制臂重建结果虽保留了总体轮廓，但侧壁的光顺曲面被分段曲面替代，根部圆角缺失，局部形成台阶。对于单实体模型，GPT-6 有时会生成多个相互穿插、体积重叠的实体，连接方式与参考零件不符。

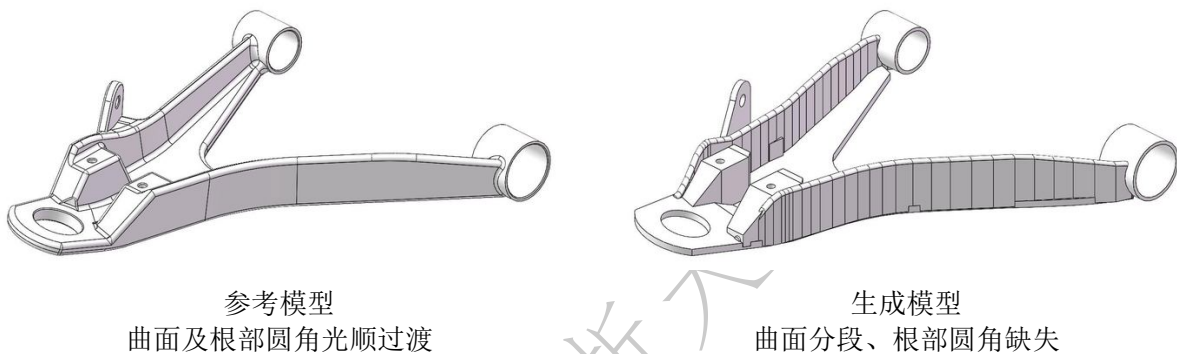


图 10 控制臂曲面及根部过渡的重建偏差

总体而言，GPT-6 目前的能力是：

- 对于轮廓清晰、主要特征充分可见的规则板件和壳体，GPT-6能够较好恢复主体形状与孔槽布局。
- 对于自由曲面、薄壁及密集筋板交汇的结构，GPT-6仍容易简化曲面和连接过渡，难以准确还原曲面与圆角。
- 对于复杂装配体，GPT-6生成的模型实体间连接、孔槽及装配接口会有偏差。
- 对于无尺寸参照的建模情况，GPT-6难以准确确定实际尺度甚至比例。

(6) B-rep 模型参数化重建

B-rep 模型参数化重建任务考察 GPT-6 能否依据已有模型的几何与拓扑信息，重建草图、建模特征及参数依赖关系，生成几何一致、可编辑的 CAD 模型。测试选取简单样本 5 个、中等样本 3 个和复杂样本 3 个，并对 3 个复杂样本进行无网络搜索测试，代表样本见图 11。

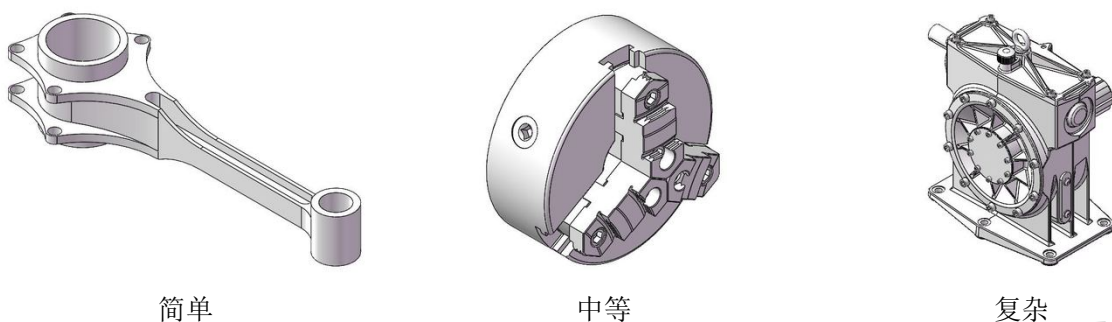


图 11 B-rep 模型参数化重建任务的不同复杂度样本

实验结果显示，实验的 11 个样本均生成了参数化模型（SolidWorks 格式），其中 6 个（占比 54.5%，5 个简单、1 个中等）通过实体有效性检查，具备参数编辑测试条件。其余 2 个中等模型和 3 个复杂模型中均存在无效实体，记为“无法编辑”，不进行参数编辑测试。对通过实体有效性检查的 6 个模型，各进行 10 次单参数修改，每次只改变一个尺寸参数。当模型参数修改生效、相关结构正确联动、修改后的模型通过实体有效性检查且能恢复原状态时，判定为通过。60 次单参数修改中，52 次通过，按样本统计，4 个样本通过全部参数编辑测试，占全部 11 个样本的 36.4%，其中 3 个简单、1 个中等。简单、中等和复杂样本的平均用时分别为 7 分 36 秒、13 分 20 秒和 50 分钟，单个样本最长用时为 1 小时 52 分钟。关闭网络搜索后的 3 个样本测试中，3 个均未通过实体有效性检查。

测试中出现的典型错误是：GPT-6 重建的模型虽然具有草图、特征和尺寸参数，外形也与参考模型接近，但仍然存在参数依赖关系不完整、可编辑性不足等问题。例如，壳体深度增加后，相关孔特征未随之正确更新，原有孔口被新增材料封闭。对于含复杂曲面的模型，GPT-6 常用多段拉伸形成的平面近似原有曲面，使光滑表面出现密集分段和台阶，造成较大的几何误差，所构建的特征结构不符合此类零件的常规建模方式，见图 12。

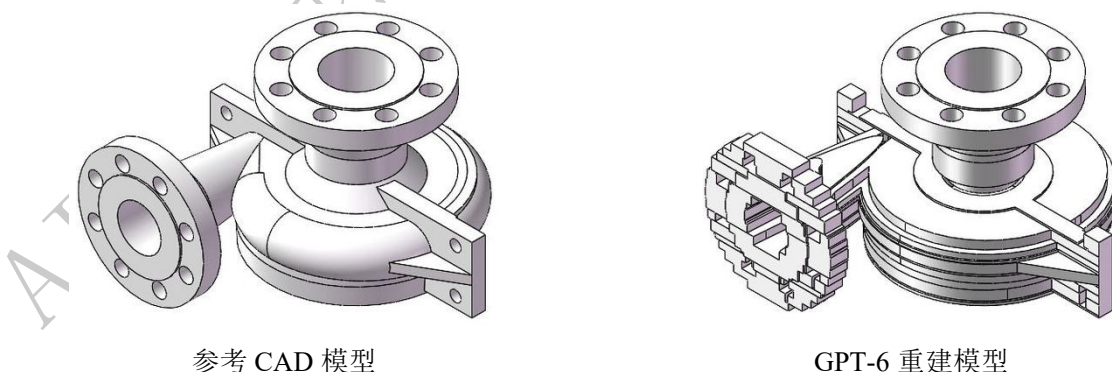


图 12 GPT-6 利用多段拉伸平面近似原有曲面

总体而言，GPT-6 目前的能力是：

- 对于以拉伸、旋转和孔槽等常规特征为主、尺寸关系较简单的模型，GPT-6 能够较准确地恢复主体几何，并支持部分尺寸修改。

- 对于孔槽与主体尺寸相互关联的模型，GPT-6建立的特征依赖关系仍可能不完整，修改主体尺寸后会出现孔口被封闭、安装孔消失等问题。
- 对于包含复杂曲面的模型，GPT-6常用多段拉伸的简单平面近似原有曲面，原本光滑的表面出现分段和台阶，重建误差较大。

AI+CAD评测，浙大DMA实验室