

05-作品代码

推报学校名称:福建师范大学

揭榜榜题名称:XH-202620 面向一流学科建设的学科垂类大模型与创新应用开发

榜题发榜单位:科大讯飞股份有限公司

申报作品名称:Syn-BioGPT: 面向合成生物学的垂类大模型开发及生物学科创新研究
支撑

申报者姓名:吴亚祺、曾信伟、徐愿、何宇彬、王晓彤、张俊昊、蔡伟贤

目录

1.技术交付与复现范围	3
2.Skill 源文件与实现组织	3
3.模型资源与运行配置	4
4.RAG 知识库与本地轻量数据库	4
5.跨 Skill 调度与调用记录	7
6.API、WebSDK 与作品门户前端	8
7.环境恢复与复现步骤	9
8.工程验证与复现支持	9
9.压缩包文件索引	10
10.作品体验入口	10

1.技术交付与复现范围

Syn-BioGPT 基于 HiAgent 2.6.0 构建，采用领域 Skill、RAG 知识库、本地轻量数据库与平台任务编排相结合的低代码实现方式。系统围绕合成路径设计、已知蛋白功能与结构分析、未知序列功能判断三类专业任务，将自然语言需求转化为数据库查询、资料检索、证据组织和结构化回答；团队另行开发作品门户前端，形成面向合成生物学科研预研的完整展示与体验入口。

本材料以可复现的工程实现为主线，提交三项专业 Skill 源文件、Agent 提交配置、模型资源描述、知识库配置、运行截图与本地轻量数据库列表、跨 Skill 调用日志，以及 API、WebSDK 和作品门户前端。各类材料按源码、配置、样本、证据、发布接入与网站前端分类组织，使任务定义、资源绑定、执行过程和交付入口能够逐项对应。

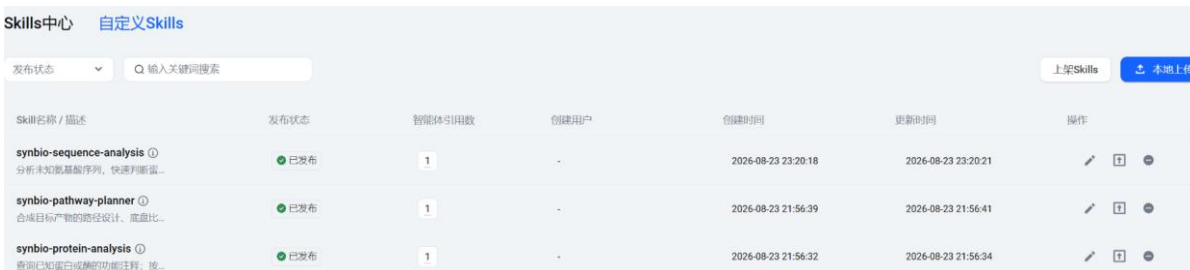
2.Skill 源文件与实现组织

核心实现位于 src/skills/，采用“任务定义—调用模板—专业规则”分层组织。SKILL.md 定义适用问题、输入要求、执行顺序及输出结构；references/callbook.md 集中维护数据库入口、参数模板和返回字段；路径模块进一步以 references/pathway-rules.md 描述反应连接、路径回溯与候选优选方法。模块元数据和 agents 配置与正文规则共同保留，各模块导入包位于 packages/。

这种文件组织方式使专业知识与调用逻辑可以独立维护：调整底盘比较维度、补充蛋白注释字段或完善序列分析规则时，可直接定位到相应模块。源码展开文件用于阅读实现细节，提交导入包用于平台部署，两种交付形式相互对应。

源文件目录	主要实现文件	实现内容
synbio-pathway-planner	SKILL.md	目标识别、反应连接、底盘比较与候选路径优选
	references/callbook.md	
	references/pathway-rules.md	
synbio-protein-analysis	SKILL.md	蛋白功能注释、催化信息、结构查询与研究证据整理
	references/callbook.md	
synbio-sequence-analysis	SKILL.md	快速特征扫描、结构域与同源分析、证据分层输出

表 2-1 专业 Skill 源文件组织



技能名称 / 描述	发布状态	智能体引用数	创建用户	创建时间	更新时间	操作
synbio-sequence-analysis ① 分析未知氨基酸序列，快速判断蛋...	已发布	1	-	2026-08-23 23:20:18	2026-08-23 23:20:21	  
synbio-pathway-planner ① 合成目标产物的路径设计、底盘比...	已发布	1	-	2026-08-23 21:56:39	2026-08-23 21:56:41	  
synbio-protein-analysis ① 查询已知蛋白成熟的功能注释，按...	已发布	1	-	2026-08-23 21:56:32	2026-08-23 21:56:34	  

图 2-1 当前绑定的三项 Skill 及发布状态

系统依托 HiAgent 的主控规划、跨 Skill 调度和工具调用能力，将团队编写的专业规则接入统一运行环境。团队开发重点包括三类任务的意图定义、数据库调用模板、路径解析规则、序列证据分层、论文库组织及应用参数配置。平台负责执行资源管理，领域 Skill 负责规定专业任务如何展开，两者通过 Agent 资源绑定协同运行。

在专业流程设计上，路径模块依据反应与模块信息组织生物合成路线；蛋白模块围绕已知名称或编号汇总功能、催化和结构信息；序列模块以 ScanProsite 快速扫描为

入口，结合 CDD 与 BLAST 补充结构域和同源线索。统一的来源标识、证据等级和输出规则，使不同任务的结果能够衔接到同一预研过程。

3.模型资源与运行配置

生成模型采用 doubao-seed-2.0-lite_PlatformPresetting，承担用户意图理解、任务规划、工具结果整合和专业文本表达。Agent 采用 Planning 应用类型与 Single 模式，通过 function_call 策略组织工具交互。生成模型、Embedding 模型及关键运行参数按下表配置，模型与 Agent 文件分别位于 config/hiagent_export/model/和 agent/目录。

配置项	取值	配置文件
运行平台	HiAgent 2.6.0	平台运行环境
Agent 版本/模式	v7.8; Planning / Single	Agent YAML
生成模型 Key	doubao-seed-2-0-lite-260428	LLM 配置 YAML
平台模型资源 ID	d97rgij1m9aildaji7jg	UniqueName / ModelID
Embedding 模型	Doubao-embedding-vision_PlatformPresetting	Embedding 模型 YAML
工具策略	function_call	ModelConfig.Strategy
采样参数	Temperature=0.28; TopP=0.7	ModelConfig
生成与迭代上限	MaxTokens=4096; MaxIterations=30	ModelConfig
上下文/推理	RoundsReserved=3; medium	ModelConfig
知识匹配	RagEnabled=true; auto; TopK=3; Similarity=0.5	KnowledgeConfig

表 3-1 模型及关键运行参数

模型接入材料包含平台模型名称、Key 及 UniqueName，对应生成模型的选择、配置定位与资源映射。平台模型资源 ID 为 d97rgij1m9aildaji7jg；API AppID 和 WebSDK appKey 用于应用发布接入。参数采用配置文件随包交付，复现时可据此恢复生成长度、上下文轮次、任务迭代和知识检索设置；模型服务接入由团队协助提供。

4.RAG 知识库与本地轻量数据库

RAG 检索增强由论文资源、Embedding 模型、知识库索引和 Agent 知识绑定共同构成。专业资料经平台解析与分段后建立检索索引，用户问题进入知识匹配流程，相关片段连同文档来源一并提供给模型，为回答补充领域资料和研究依据。Agent 配置中的 KnowledgeIDs 与 KnowledgeMap 记录了 Syn-BioGPT 论文库的关联关系。

本项目采用 Doubao-embedding-vision_PlatformPresetting 作为 Embedding 模型，知识匹配设置为 auto、TopK=3、Similarity=0.5。数据库调用侧重获取路径、化合物及蛋白等结构化信息，论文库检索侧重补充研究背景、机制解释和工程案例，两类来源在回答组织阶段形成互补。

Embedding 模型



Doubao-embedding-vision_PlatformPresetting

ASR 模型

-

图 4-1 论文库 Embedding 模型设置

文件名称	处理状态	类型	分段数量	命中次数	标签	启用 / 禁用	操作
2023-朱砂根-基于流式细胞术对朱砂根种质资源基因组大小的测定.p...	解析成功	通用	26	0	0	启用	编辑 下载 删除
2023-TAF-reviewforwhite flower.pdf	解析成功	通用	98	5	0	启用	编辑 下载 删除
2018-梅花MADS-fpls-09-00017.pdf	解析成功	通用	63	0	0	启用	编辑 下载 删除
2025-建兰ABC花香-opr-0025-0031.pdf	解析成功	通用	76	0	0	启用	编辑 下载 删除
2023-茉莉bZIP-第一作者-ijms-25-00488.pdf	解析成功	通用	84	0	0	启用	编辑 下载 删除
2020-The complete chloroplast genome of Cymbidium maguane...	解析成功	通用	13	0	0	启用	编辑 下载 删除
2019-Basic chloroplast characterizations of Prunus campanulata x...	解析成功	通用	17	0	0	启用	编辑 下载 删除

图 4-2 论文文件列表及解析状态（局部）

截至当前统计，Syn-BioGPT 论文库包含 241 份学术论文，总计 969MiB。团队将研究相关论文集中整理为可检索资料集合，并在平台中完成文件管理、解析处理和应用绑定。图 4-2 展示了论文文件名称、解析状态及分段数量等信息，使资料管理从分散的本地文件转化为可定位、可检索、可持续维护的专业知识资源。

知识库以论文片段参与问答，能够把与当前任务相关的研究信息提供给模型，支持专业概念解释、候选方案比较和研究依据整理。对于新增文献，可在既有论文库中继续导入和解析；对于专题任务，可结合文件名称与来源索引定位相关资料。知识内容与 Skill 执行规则分别维护，便于随研究主题扩展资料范围，同时保持应用流程的一致性。

Crtl

3 个命中分段

1

得分 0.706

Huang CR, Burns KH, Boeke JD (2012). Active transposition in genomes. Annual Review of Genetics 46: 651-675. <https://doi.org/10.1146/annurev-genet-110711-155616> Hubbard CE (1968). Grasses. A guide to their structure, identification, uses, and distribution in the British Isles. Grasses A guide to their structure, identi...

> 命中 1 个索引

2023-TAF-reviewforwhite flower.pdf

2

得分 0.700

format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes we re made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line...

> 命中 1 个索引

2018-梅花休眠-s41598-018-22537-w.pdf

3

得分 0.697

Figure 1. The modified anthocyanin pathway leading to white flower color (PAL, phenylalanine ammonialyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate C oA ligase; CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; F3'H, flavonoid 3'-hydroxylase; FLS, Flavonol synthase; DFR, dihydro...

> 命中 1 个索引

2023-TAF-reviewforwhite flower.pdf

图 4-3 检索候选片段、来源与相似度（局部）

检索结果采用“候选片段—来源文档—相似度分数”的组织方式。图 4-3 展示了检索返回的三个候选片段及来源定位，模型可以在回答时关联所用资料，研究者也可沿文档入口继续阅读原文。专业问题由此获得可追踪的资料依据，便于后续整理研究笔记、比较候选方案和编写预研报告。

这一机制将论文内容从静态附件转化为任务中的知识输入：Embedding 支持语义关联，候选片段保留原文上下文，来源索引衔接资料查阅。与 Skill 中的证据分层和引用规则配合后，系统能够同时组织专业结论及其支撑材料。



图 4-4 平台知识库运行统计

在集中测试期间，平台运行统计记录检索总量 527 次、知识库综合命中率 100%。结合论文解析列表与候选片段截图，可以对应展示资料入库、检索返回和运行统计三个环节，形成 RAG 知识增强的工程运行记录。

项目已在本地建立轻量数据库，包含 10 张结构化数据表，覆盖代谢通路、生化反应、代谢物、酶功能注释、蛋白质序列、基因序列、蛋白质结构、酶动力学、底盘生物及文献证据。数据库列表如图 4-5 所示。团队高度重视版权保护，受文献及数据来源授权范围限制，公开作品包展示知识库配置、数据库结构与运行截图，因版权原因，知识库全文和数据库具体内容采用授权核验方式提供。可联系团队索取证明材料：108012025017@student.fjnu.edu.cn。

内容	说明
数据组织	10 张结构化数据表，按专业数据类型分类管理。
查询与计算	引用表格数据后，支持基于自然语言对数据库进行查询和计算（NL2SQL）。

表 4-1 本地轻量数据库及查询方式

Syn-BioGPT本地轻量数据库					
10 Tables 2026-07-28 17:06:23 创建 2026-09-11 18:12:53 更新					
Q 输入数据表名称搜索					
☐	数据表名称/描述	数据量	校验结果	同步状态	创建用户
☐	 metabolic_pathways 代谢通路数据	18642行 · 8列	全部校验成功	同步完成	徐 徐愿
☐	 biochemical_reactions 生化反应数据	86371行 · 9列	全部校验成功	同步完成	徐 徐愿
☐	 metabolites 代谢物数据	156284行 · 8列	全部校验成功	同步完成	徐 徐愿
☐	 enzyme_annotations 酶功能注释	72458行 · 8列	全部校验成功	同步完成	徐 徐愿
☐	 protein_sequences 蛋白质序列	418526行 · 7列	全部校验成功	同步完成	徐 徐愿
☐	 gene_sequences 基因序列数据	315792行 · 7列	全部校验成功	同步完成	徐 徐愿
☐	 protein_structures 蛋白质结构	41683行 · 8列	全部校验成功	同步完成	徐 徐愿
☐	 enzyme_kinetics 酶动力学数据	36947行 · 9列	全部校验成功	同步完成	徐 徐愿
☐	 organism_chassis 底盘生物信息	3286行 · 8列	全部校验成功	同步完成	徐 徐愿
☐	 literature_evidence 文献证据数据	164003行 · 7列	全部校验成功	同步完成	徐 徐愿

图 4-5 本地轻量数据库数据表列表

5.跨 Skill 调度与调用记录

跨 Skill 协同由智能体主控规划与专业模块绑定共同实现。面对包含多项专业需求的复合问题，主控先建立任务清单，再按任务顺序加载对应 Skill 及参考规则，使序列、蛋白与路径等不同专业能力能够在同一会话中组织执行。用户可以围绕同一研究目标提出组合需求，由系统完成任务拆分、能力切换和执行状态管理。

在“先分析所给序列，再规划大肠杆菌合成番茄红素”的调用记录中，主控通过 write_todos 建立两项待办，先加载序列 Skill 完成特征检索，再加载路径 Skill 及其参考文件，继续开展目标化合物与文献查询。任务清单承接总体研究目标，Skill 文件规定各阶段的专业步骤，工具结果与状态记录保留在同一会话链路中。

该机制将多类专业分析连接为连续的预研过程，减少用户在不同工具之间重复描述研究背景和切换操作的步骤。对于团队维护，模块可分别更新；对于复现检查，任务计划、加载顺序和工具调用均可通过日志定位。平台调度能力与团队领域规则结合，构成跨任务协同的实现基础。

阶段	原始消息索引	已记录事件
任务拆分	1-2	write_todos 建立两项待办
序列分析	3-10	ScanProsite 命中 PS00982，区间 459-479
模块切换	11-18	加载路径 Skill 与两份 references
路径查询	19-24	访问 KEGG、PubChem 与 PubMed；返回化合物 CID 446925
状态更新	25-26	两项待办标记 completed

表 5-1 跨 Skill 实际调用过程（消息从 0 编号）

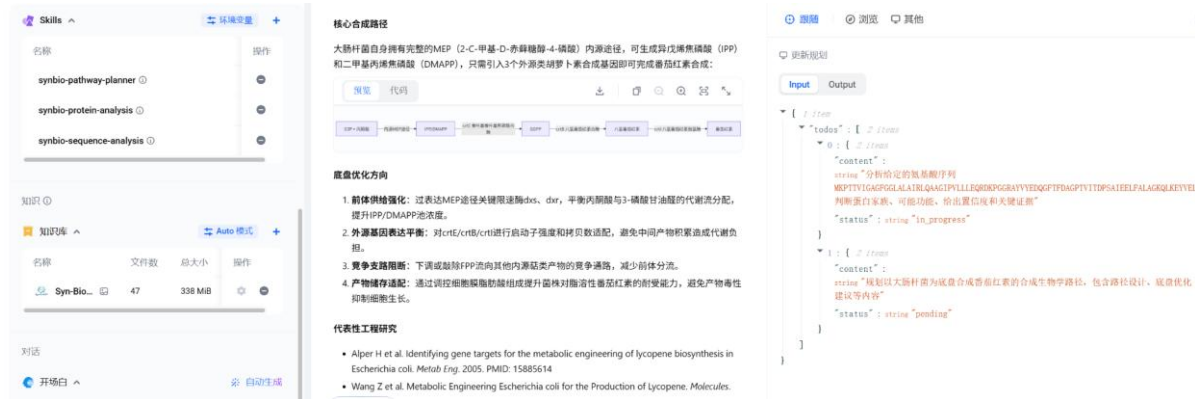


图 5-1 编排界面中的资源绑定、任务计划与结果

该次执行记录共 27 条消息、13 次工具调用，其中 write_todos 为 3 次、read_file 为 4 次、visit 为 6 次，覆盖任务规划、Skill 与参考规则加载、数据库访问及状态更新。完整日志与逐次索引保存在 evidence/，输入序列和原始问题保存在 samples/，可按消息编号重现专业模块之间的执行衔接。

6.API、WebSDK 与作品门户前端

智能体采用 API 与 WebSDK 两类能力发布方式，并由团队开发独立作品门户，支持服务接口接入、完整对话界面嵌入、网页聊天气泡和项目材料在线浏览。门户设置体验入口、作品 Demo、作品代码和关于团队四个栏目，点击“立即体验”进入/chat/对话路径。不同接入方式共用已发布的智能体能力，便于嵌入教学平台、项目网站和科研辅助页面。

发布方式	接入信息	交付文件
API	基础地址： http://14.103.118.162:32300/api/proxy/api/v1 AppID: da0r4cddh94e7shunv3g API Key 向团队申请	integrations/ API 接入说明.md
全界面 WebSDK 与作品门户	embedFull.js; WebClient appKey 集中配置: /chat/嵌入	integrations/websdk/full.html frontend/synbiogpt-site/
聊天气泡 WebSDK	embedLite.js; WebLiteClient appKey: da0r7qtdh948qkrs0t50	integrations/ websdk/lite.html

表 6-1 发布形式与接入文件

全界面接入文件 integrations/websdk/full.html 加载 embedFull.js，并初始化 HiagentWebSDK.WebClient；聊天气泡文件 lite.html 加载 embedLite.js，并初始化 HiagentWebSDK.WebLiteClient。作品门户源码位于 frontend/synbiogpt-site/，采用原生 HTML、CSS 与 JavaScript 实现多栏目页面，/chat/通过 config/chat-config.js 读取 appKey 与 baseUrl，避免将接入参数分散写入各页面。

API 采用 Apikey 请求头和 JSON 请求体，UserID 贯穿同一会话。调用 create_conversation 取得 AppConversationID 后，以 chat_query_v2 流式调用规划型智能体；响应按 SSE 事件解析，失败保留错误信息。完整契约、命令行示例和客户端见 integrations/API 接入说明.md 与 hiagent_client.py；三个补充文档归档在 integrations/docs/。

7.环境恢复与复现步骤

1. 准备 HiAgent 2.6.0 工作区，接入生成模型、Embedding 模型及所需公共数据库服务。
2. 导入 packages/HiAgent_Syn-BioGPT_v7.8_提交配置.zip，恢复 Agent 应用配置。分模块导入时，使用三个 Skill 包，并按 Agent YAML 完成资源绑定。
3. 按表 3-1 恢复模型与生成参数，在工作区中完成模型和 Skill 资源映射，确认专业模块已绑定至 Agent。
4. 关联 Syn-BioGPT 论文库，按知识库配置导入论文资料、完成解析并建立索引；使用典型专业查询检查候选片段和来源返回。
5. 使用 samples/中的跨 Skill 问题与 FASTA 样本运行复合任务，对照 evidence/中的任务计划、Skill 加载顺序、工具返回及状态更新记录检查执行过程。
6. 检查作品门户首页、Demo、代码与团队栏目，在 config/chat-config.js 中配置 appKey；baseUrl 留空时使用当前站点地址。打开/chat/验证全界面 WebSDK；服务器部署时按 frontend/synbiogpt-site/README_部署说明.md 配置 Nginx，将静态页面与 SDK 资源统一到同一域名。配置 API Key 后，可按实例接口文档完成接口调用。

8.工程验证与复现支持

验证层次	验证内容	支撑材料
源文件	展开源码与原 Skill 包对应	校验脚本与 SHA-256 清单
知识检索	论文绑定、解析、片段与来源返回	知识库配置、查询及运行截图
任务协同	复合任务拆分、模块切换与状态更新	输入样本、调用日志与索引
发布接入	全界面、聊天气泡与 API 接入	HTML 文件与服务接入配置

表 8-1 工程验证对象与支撑材料

交付包通过源文件校验、配置快照、输入样本和调用日志提供分层复现支持：文件层检查材料完整性，配置层恢复运行参数，任务层检查专业模块衔接，接入层验证应用发布形式。团队可配合提供模型服务接入、论文资料和 API 密钥，支持在目标工作区完成环境恢复与运行验证。

9.压缩包文件索引

目录或文件	内容
05-作品代码.docx	技术实现、配置、运行与复现说明
src/skills/	三项 Skill 完整展开源码
packages/	提交 Agent 配置与三个 Skill 导入包
config/hiagent_export/	提交配置与资源描述
config/	模型环境、RAG 迁移、本地数据库说明、参数快照与服务信息
integrations/	API 与 WebSDK 接入文件；作品门户页面、样式、品牌资源、团队照片、聊天配置及 Nginx 示例
frontend/synbiogpt-site/	跨 Skill 原始输入、FASTA 及工程验收步骤
samples/	原始日志、调用索引、裁剪截图及二维码
evidence/	文件完整性与源包一致性校验工具
tools/verify_package.py	文件完整性与源包一致性校验工具
manifest.sha256	文件 SHA-256 清单

表 9-1 交付文件索引

解压后执行 `python tools/verify_package.py`，可检查文件完整性、Skill 源文件一致性、模型引用及 SDK 关键字段。`frontend/synbiogpt-site/`保留网站页面、样式、品牌资源、团队照片、聊天配置和 Nginx 示例；校验脚本、`manifest.sha256`、输入样本与运行记录共同形成从源码阅读、环境配置、门户展示到任务运行的连续复现路径。

10.作品体验入口

作品体验地址：<https://biogpt.com.cn>



图 10-1 扫码进入 Syn-BioGPT

API 密钥及跨工作区复现所需资源，请向项目团队申请：108012025017@student.fjnu.edu.cn。