

API 接口文档

1、创建网络搜索

POST <https://api.deyunai.com/v1/responses>

请求头

字段	值	描述
Content-Type	application/json	数据交换格式
Authorization	Bearer {your api key}	将 {your api key} 替换为您的 token
Accept	string	必选
X-Forwarded-Host	string	可选

请求体

参数名称	类型	必填	参数描述
model	String	可选	选择模型
tools	array	可选	模型在生成响应时可能调用的工具数组
type	String	是	网络搜索工具的类型
input	String	可选	输入文本

示例

代码块

```
1  {
2    "model": "gpt-4.1-2025-04-14",
3    "tools": [
4      {
5        "type": "web_search_preview"
6      }
7    ],
8    "input": "what was a positive news story from today?"
9  }
```

响应体

参数名称	类型	必填	参数描述
id	String	是	用户唯一标识
object	String	是	固定返回 chat.completion
created	integer	是	响应生成时的 Unix 时间戳
choices	array	是	模型给出的候选答案数组
index	integer	可选	当前候选答案在数组中的序号
message	object	可选	模型返回的单条消息对象
role	string	是	消息作者角色
content	string	是	模型实际输出的文本内容
finish_reason	string	可选	停止原因
usage	object	是	本次请求 token 用量汇总
prompt_tokens	integer	是	用户输入 + 系统提示消耗的总 token 数
completion_tokens	integer	是	模型生成文本消耗的总 token 数
total_tokens	integer	是	两者之和，即本次请求总共消耗的 token 量

示例

代码块

```
1  {
2      "id": "chatcpl-123",
3      "object": "chat.completion",
4      "created": 1677652288,
5      "choices": [
6          {
7              "index": 0,
8              "message": {
9                  "role": "assistant",
10                 "content": "\n\nHello there, how may I assist you today?"
11             },
12             "finish_reason": "stop"
13         }
14     ],
15     "usage": {
16         "prompt_tokens": 9,
17         "completion_tokens": 12,
18         "total_tokens": 21
19     }
20 }
```

2、创建模型响应 gpt-5启用思考

POST <https://api.deyunai.com/v1/responses>

请求头

字段	值	描述
Content-Type	application/json	数据交换格式
Authorization	Bearer {your api key}	将 {your api key} 替换为您的 token
Accept	string	必选
X-Forwarded-Host	string	可选

请求体

参数名称	类型	必填	参数描述
model	String	可选	选择模型
tools	array	可选	模型在生成响应时可能调用的工具数组
type	String	是	网络搜索工具的类型
input	String	可选	输入文本
role	String	是	消息作者角色
content	array	是	模型实际输出的文本内容
text	object	可选	纯文本内容字段
format	object	可选	强制模型按指定结构
verbosity	String	可选	控制回答详尽度
reasoning	object	可选	配置模型是否展示推理过程及其详尽程度
effort	String	可选	推理工作量档位
summary	String	可选	推理摘要级别
stream	boolean	可选	是否流式逐字返回
store	boolean	可选	是否把本次响应存档，后续可通过 API 拉取

示例

代码块

```
1  {
2    "model": "gpt-5-2025-08-07",
3    "input": [
4      {
```

```
5         "role": "user",
6         "content": [
7             {
8                 "type": "input_text",
9                 "text": "1+2+3+4+5....9985"
10            }
11        ]
12    }
13],
14    "tools": [],
15    "text": {
16        "format": {
17            "type": "text"
18        },
19        "verbosity": "medium"
20    },
21    "reasoning": {
22        "effort": "medium",
23        "summary": "auto"
24    },
25    "stream": true,
26    "store": true
27 }
```

响应体

参数名称	类型	必填	参数描述
id	String	是	用户唯一标识
object	String	是	固定返回 chat.completion
created	integer	是	响应生成时的 Unix 时间戳
choices	array	是	模型给出的候选答案数组
index	integer	可选	当前候选答案在数组中的序号
message	object	可选	模型返回的单条消息对象
role	string	是	消息作者角色
content	string	是	模型实际输出的文本内容
finish_reason	string	可选	停止原因
usage	object	是	本次请求 token 用量汇总
prompt_tokens	integer	是	用户输入 + 系统提示消耗的总 token 数
completion_tokens	integer	是	模型生成文本消耗的总 token 数
total_tokens	integer	是	两者之和，即本次请求总共消耗的 token 量

示例

代码块

```
1  {
2      "id": "chatcpl-123",
3      "object": "chat.completion",
4      "created": 1677652288,
5      "choices": [
6          {
7              "index": 0,
8              "message": {
9                  "role": "assistant",
10                 "content": "\n\nHello there, how may I assist you today?"
11             },
12             "finish_reason": "stop"
13         }
14     ],
15     "usage": {
16         "prompt_tokens": 9,
17         "completion_tokens": 12,
18         "total_tokens": 21
19     }
20 }
```

3、创建函数调用

POST <https://api.deyunai.com/v1/responses>

请求头

字段	值	描述
Content-Type	application/json	数据交换格式
Authorization	Bearer {your api key}	将 {your api key} 替换为您的 token
Accept	string	必选
X-Forwarded-Host	string	可选

请求体


```
33         "type": "integer"
34     }
35 }
36 }
37 },
38 {
39     "type": "function",
40     "name": "system_info",
41     "description": "获取当前系统的基本信息，包括操作系统、Java版本、内存使用情
况等",
42     "parameters": {
43         "type": "object",
44         "properties": {}
45     }
46 },
47 {
48     "type": "function",
49     "name": "math_calculator",
50     "description": "执行基本数学运算，支持加减乘除和幂运算",
51     "parameters": {
52         "type": "object",
53         "properties": {
54             "a": {
55                 "description": "第一个数字",
56                 "type": "number"
57             },
58             "b": {
59                 "description": "第二个数字",
60                 "type": "number"
61             },
62             "operation": {
63                 "description": "运算类型：add, subtract, multiply,
divide, power",
64                 "type": "string"
65             }
66         },
67         "required": [
68             "operation",
69             "a",
70             "b"
71         ]
72     }
73 },
74 {
75     "type": "function",
76     "name": "current_time",
77     "description": "获取当前系统时间",
```

```
78         "parameters": {
79             "type": "object",
80             "properties": {}
81         }
82     }
83 ]
84 }
```

响应体

参数名称	类型	必填	参数描述
id	String	是	用户唯一标识
object	String	是	固定返回 chat.completion
created	integer	是	响应生成时的 Unix 时间戳
choices	array	是	模型给出的候选答案数组
index	integer	可选	当前候选答案在数组中的序号
message	object	可选	模型返回的单条消息对象
role	string	是	消息作者角色
content	string	是	模型实际输出的文本内容
finish_reason	string	可选	停止原因
usage	object	是	本次请求 token 用量汇总
prompt_tokens	integer	是	用户输入 + 系统提示消耗的总 token 数
completion_tokens	integer	是	模型生成文本消耗的总 token 数
total_tokens	integer	是	两者之和，即本次请求总共消耗的 token 量

示例

代码块

```
1  {
2      "id": "resp_68ac1558e9488190bcc2cbb94c7d7a140b8064b15a091a77",
3      "object": "response",
4      "created_at": 1756108120,
5      "status": "completed",
6      "background": false,
7      "content_filters": null,
8      "error": null,
9      "incomplete_details": null,
10     "instructions": null,
11     "max_output_tokens": null,
12     "max_tool_calls": null,
```



```
13  "model": "gpt-4.1-data",
14  "output": [
15    {
16      "id": "fc_68ac15594be88190a58217f75a6a54cc0b8064b15a091a77",
17      "type": "function_call",
18      "status": "completed",
19      "arguments": "{\"sign\": \"Aquarius\"}",
20      "call_id": "call_6kbB2GTk62PPUNdGuBJPx9KQ",
21      "name": "get_horoscope"
22    }
23  ],
24  "parallel_tool_calls": true,
25  "previous_response_id": null,
26  "prompt_cache_key": null,
27  "reasoning": {
28    "effort": null,
29    "summary": null
30  },
31  "safety_identifier": null,
32  "service_tier": "default",
33  "store": true,
34  "temperature": 1.0,
35  "text": {
36    "format": {
37      "type": "text"
38    }
39  },
40  "tool_choice": "auto",
41  "tools": [
42    {
43      "type": "function",
44      "description": "Get today's horoscope for an astrological sign.",
45      "name": "get_horoscope",
46      "parameters": {
47        "properties": {
48          "sign": {
49            "description": "An astrological sign like Taurus or Aquarius",
50            "type": "string"
51          }
52        },
53        "required": [
54          "sign"
55        ],
56        "type": "object"
57      },
58      "strict": true
59    }
```

```
60 ],
61 "top_p": 1.0,
62 "truncation": "disabled",
63 "usage": {
64   "input_tokens": 62,
65   "input_tokens_details": {
66     "cached_tokens": 0
67   },
68   "output_tokens": 18,
69   "output_tokens_details": {
70     "reasoning_tokens": 0
71   },
72   "total_tokens": 80
73 },
74 "user": null,
75 "metadata": {}
76 }
```

4、创建模型响应

POST <https://api.deyunai.com/v1/responses>

请求头

字段	值	描述
Content-Type	application/json	数据交换格式
Authorization	Bearer {your api key}	将 {your api key} 替换为您的 token
Accept	string	必选
X-Forwarded-Host	string	可选

请求体

参数名称	类型	必填	参数描述
model	String	可选	选择模型
input	String	可选	输入文本
role	String	可选	消息作者角色
content	array	可选	模型实际输出的文本内容

示例

```
1  // 请求体
2  {
3      "model": "gpt-5.1",
4      "input": [
5          {
6              "role": "user",
7              "content": "Write a one-sentence bedtime story about a unicorn."
8          }
9      ]
10 }
```

响应体

参数名称	类型	必填	参数描述
id	String	是	用户唯一标识
object	String	是	固定返回 chat.completion
created	integer	是	响应生成时的 Unix 时间戳
choices	array	是	模型给出的候选答案数组
index	integer	可选	当前候选答案在数组中的序号
message	object	可选	模型返回的单条消息对象
role	string	是	消息作者角色
content	string	是	模型实际输出的文本内容
finish_reason	string	可选	停止原因
usage	object	是	本次请求 token 用量汇总
prompt_tokens	integer	是	用户输入 + 系统提示消耗的总 token 数
completion_tokens	integer	是	模型生成文本消耗的总 token 数
total_tokens	integer	是	两者之和，即本次请求总共消耗的 token 量

示例

```
代码块
1  {
2      "id": "chatcpl-123",
3      "object": "chat.completion",
4      "created": 1677652288,
5      "choices": [
6          {
7              "index": 0,
8              "message": {
9                  "role": "assistant",
```

```
10         "content": "\n\nHello there, how may I assist you today?"
11     },
12     "finish_reason": "stop"
13 }
14 ],
15 "usage": {
16     "prompt_tokens": 9,
17     "completion_tokens": 12,
18     "total_tokens": 21
19 }
20 }
```

5、创建函数调用

POST <https://api.deyunai.com/v1/responses>

请求头

字段	值	描述
Content-Type	application/json	数据交换格式
Authorization	Bearer {your api key}	将 {your api key} 替换为您的 token
Accept	string	必选
X-Forwarded-Host	string	可选

请求体

参数名称	类型	必填	参数描述
model	String	可选	选择模型
input	String	可选	输入文本
role	String	可选	消息作者角色
content	array	可选	模型实际输出的文本内容

示例

代码块

```
1
2 {
3     "input": [
4         {
5             "role": "user",
```

```
6         "content": "请同时帮我做以下几件事：\n1. 获取当前系统时间\n2. 查看系统信息\n（操作系统、内存等）\n3. 帮我计算 123.5 + 456.7 的结果\n4. 生成3个1-100之间的随机数\n\n这是一个并行工具调用测试，请同时执行这些任务。"
7     }
8 ],
9     "metadata": {
10         "model_id": "32"
11     },
12     "model": "gpt-4.1",
13
14     "tool_choice": "auto",
15     "tools": [
16         {
17             "type": "function",
18             "name": "random_generator",
19             "description": "生成指定范围内的随机数，支持批量生成",
20             "parameters": {
21                 "type": "object",
22                 "properties": {
23                     "min": {
24                         "description": "最小值（默认1）",
25                         "type": "integer"
26                     },
27                     "max": {
28                         "description": "最大值（默认100）",
29                         "type": "integer"
30                     },
31                     "count": {
32                         "description": "生成数量（默认1，最大10）",
33                         "type": "integer"
34                     }
35                 }
36             }
37         },
38         {
39             "type": "function",
40             "name": "system_info",
41             "description": "获取当前系统的基本信息，包括操作系统、Java版本、内存使用情
况等",
42             "parameters": {
43                 "type": "object",
44                 "properties": {}
45             }
46         },
47         {
48             "type": "function",
49             "name": "math_calculator",
```

```

50         "description": "执行基本数学运算，支持加减乘除和幂运算",
51         "parameters": {
52             "type": "object",
53             "properties": {
54                 "a": {
55                     "description": "第一个数字",
56                     "type": "number"
57                 },
58                 "b": {
59                     "description": "第二个数字",
60                     "type": "number"
61                 },
62                 "operation": {
63                     "description": "运算类型: add, subtract, multiply,
divide, power",
64                     "type": "string"
65                 }
66             },
67             "required": [
68                 "operation",
69                 "a",
70                 "b"
71             ]
72         }
73     },
74     {
75         "type": "function",
76         "name": "current_time",
77         "description": "获取当前系统时间",
78         "parameters": {
79             "type": "object",
80             "properties": {}
81         }
82     }
83 ]
84 }

```

响应体

参数名称	类型	必填	参数描述
id	String	是	用户唯一标识
object	String	是	固定返回 chat.completion
created	integer	是	响应生成时的 Unix 时间戳
choices	array	是	模型给出的候选答案数组
index	integer	可选	当前候选答案在数组中的序号
message	object	可选	模型返回的单条消息对象
role	string	是	消息作者角色
content	string	是	模型实际输出的文本内容
finish_reason	string	可选	停止原因
usage	object	是	本次请求 token 用量汇总
prompt_tokens	integer	是	用户输入 + 系统提示消耗的总 token 数
completion_tokens	integer	是	模型生成文本消耗的总 token 数
total_tokens	integer	是	两者之和，即本次请求总共消耗的 token 量

示例

代码块

```
1  {
2    "id": "resp_68ac1558e9488190bcc2cbb94c7d7a140b8064b15a091a77",
3    "object": "response",
4    "created_at": 1756108120,
5    "status": "completed",
6    "background": false,
7    "content_filters": null,
8    "error": null,
9    "incomplete_details": null,
10   "instructions": null,
11   "max_output_tokens": null,
12   "max_tool_calls": null,
13   "model": "gpt-4.1-data",
14   "output": [
15     {
16       "id": "fc_68ac15594be88190a58217f75a6a54cc0b8064b15a091a77",
17       "type": "function_call",
18       "status": "completed",
19       "arguments": "{\"sign\": \"Aquarius\"}",
20       "call_id": "call_6kbB2GTk62PPUNdGuBJPx9KQ",
21       "name": "get_horoscope"
22     }
23   ],
```

```
24     "parallel_tool_calls": true,
25     "previous_response_id": null,
26     "prompt_cache_key": null,
27     "reasoning": {
28         "effort": null,
29         "summary": null
30     },
31     "safety_identifier": null,
32     "service_tier": "default",
33     "store": true,
34     "temperature": 1.0,
35     "text": {
36         "format": {
37             "type": "text"
38         }
39     },
40     "tool_choice": "auto",
41     "tools": [
42         {
43             "type": "function",
44             "description": "Get today's horoscope for an astrological sign.",
45             "name": "get_horoscope",
46             "parameters": {
47                 "properties": {
48                     "sign": {
49                         "description": "An astrological sign like Taurus or Aquarius",
50                         "type": "string"
51                     }
52                 },
53                 "required": [
54                     "sign"
55                 ],
56                 "type": "object"
57             },
58             "strict": true
59         }
60     ],
61     "top_p": 1.0,
62     "truncation": "disabled",
63     "usage": {
64         "input_tokens": 62,
65         "input_tokens_details": {
66             "cached_tokens": 0
67         },
68         "output_tokens": 18,
69         "output_tokens_details": {
70             "reasoning_tokens": 0
```



```
71     },
72     "total_tokens": 80
73   },
74   "user": null,
75   "metadata": {}
76 }
```

6、创建模型响应(流式返回)

POST <https://api.deyunai.com/v1/responses>

请求头

字段	值	描述
Content-Type	application/json	数据交换格式
Authorization	Bearer {your api key}	将 {your api key} 替换为您的 token
Accept	string	必选
X-Forwarded-Host	string	可选

请求体

参数名称	类型	必填	参数描述
model	String	可选	选择模型
input	String	可选	输入文本
stream	boolean	可选	是否流式逐字返回
role	String	必选	消息作者角色
content	array	必选	模型实际输出的文本内容

示例

代码块

```
1  {
2    "model": "gpt-4.1",
3    "stream": true,
4    "input": [
5      {
6        "role": "user",
7        "content": "你好啊"
8      }
9    ]
10 }
```

```
9     ]
10 }
```

响应体

参数名称	类型	必填	参数描述
id	String	是	用户唯一标识
object	String	是	固定返回 chat.completion
created	integer	是	响应生成时的 Unix 时间戳
choices	array	是	模型给出的候选答案数组
index	integer	可选	当前候选答案在数组中的序号
message	object	可选	模型返回的单条消息对象
role	string	是	消息作者角色
content	string	是	模型实际输出的文本内容
finish_reason	string	可选	停止原因
usage	object	是	本次请求 token 用量汇总
prompt_tokens	integer	是	用户输入 + 系统提示消耗的总 token 数
completion_tokens	integer	是	模型生成文本消耗的总 token 数
total_tokens	integer	是	两者之和，即本次请求总共消耗的 token 量

示例

代码块

```
1  {
2      "id": "chatcml-123",
3      "object": "chat.completion",
4      "created": 1677652288,
5      "choices": [
6          {
7              "index": 0,
8              "message": {
9                  "role": "assistant",
10                 "content": "\n\nHello there, how may I assist you today?"
11             },
12             "finish_reason": "stop"
13         }
14     ],
15     "usage": {
16         "prompt_tokens": 9,
17         "completion_tokens": 12,
```

```
18         "total_tokens": 21
19     }
20 }
```

7、创建模型响应 (控制思考长度)

POST <https://api.deyunai.com/v1/responses>

请求头

字段	值	描述
Content-Type	application/json	数据交换格式
Authorization	Bearer {your api key}	将 {your api key} 替换为您的 token
Accept	string	必选
X-Forwarded-Host	string	可选

请求体

参数名称	类型	必填	参数描述
model	String	可选	选择模型
input	String	可选	输入文本
reasoning	object	可选	推理模型在生成回复时使用的思路链的描述
effort	String	可选	推理工作量档位

示例

代码块

```
1  {
2      "model": "gpt-5.1",
3      "input": [
4          {
5              "role": "user",
6              "content": "Write a one-sentence bedtime story about a unicorn."
7          }
8      ],
9      "reasoning": {
10         "effort": "high"
11     }
12 }
```

响应体

参数名称	类型	必填	参数描述
id	String	是	用户唯一标识
object	String	是	固定返回 chat.completion
created	integer	是	响应生成时的 Unix 时间戳
choices	array	是	模型给出的候选答案数组
index	integer	可选	当前候选答案在数组中的序号
message	object	可选	模型返回的单条消息对象
role	string	是	消息作者角色
content	string	是	模型实际输出的文本内容
finish_reason	string	可选	停止原因
usage	object	是	本次请求 token 用量汇总
prompt_tokens	integer	是	用户输入 + 系统提示消耗的总 token 数
completion_tokens	integer	是	模型生成文本消耗的总 token 数
total_tokens	integer	是	两者之和，即本次请求总共消耗的 token 量

示例

代码块

```
1  {
2      "id": "chatcpl-123",
3      "object": "chat.completion",
4      "created": 1677652288,
5      "choices": [
6          {
7              "index": 0,
8              "message": {
9                  "role": "assistant",
10                 "content": "\n\nHello there, how may I assist you today?"
11             },
12             "finish_reason": "stop"
13         }
14     ],
15     "usage": {
16         "prompt_tokens": 9,
17         "completion_tokens": 12,
18         "total_tokens": 21
19     }
20 }
```

