



# CDW Documentation

## Sample Workbook to Monitor Azure OpenAI

---

# Sample Workbook to Monitor Azure OpenAI

## Azure Workbook: Monitor Azure OpenAI

### How to Set It Up

1. Go to the Azure Portal.
2. Navigate to Monitor > Workbooks.
3. Click + New to create a new workbook.
4. Click the “</> Advanced Editor” tab.
5. Paste the following JSON:

```
{
  "version": "Notebook/1.0",
  "items": [
    {
      "type": 1,
      "content": {
        "json": "## Azure OpenAI API Monitoring Workbook\nThis workbook provides an overview of Azure OpenAI metrics such as total calls, success rate, latency, and throttling."
      }
    },
    {
      "type": 9,
      "content": {
        "version": "KqlParameterItem/1.0",
        "parameters": [
          {
            "id": "subscription",
            "type": 2,
            "value": "",
            "name": "Subscription",
            "query": "ResourceContainers | where type == 'microsoft.resources/subscriptions' | project id, name",
            "isRequired": true
          },
          {
            "id": "resource",
            "type": 2,
            "value": "",
            "name": "OpenAI Resource",
            "query": "Resources | where type == 'microsoft.cognitiveservices/accounts' and name has 'openai' | project id, name",
            "isRequired": true
          }
        ]
      }
    }
  ]
}
```

```
    }
  },
  {
    "type": 3,
    "content": {
      "chartSettings": {
        "chartType": 1,
        "title": "Total API Calls"
      },
      "query": {
        "resourceType": "microsoft.insights/components",
        "query": "AzureMetrics\n| where Resource contains '{resource}'\n|
where MetricName == 'TotalCalls'\n| summarize Total=sum(Total) by
bin(TimeGenerated, 1h)\n| order by TimeGenerated asc"
      }
    }
  },
  {
    "type": 3,
    "content": {
      "chartSettings": {
        "chartType": 1,
        "title": "Success vs Failed Calls"
      },
      "query": {
        "resourceType": "microsoft.insights/components",
        "query": "AzureMetrics\n| where Resource contains '{resource}'\n|
where MetricName in ('SuccessfulCalls', 'FailedCalls')\n| summarize
Total=sum(Total) by bin(TimeGenerated, 1h), MetricName\n| render timechart"
      }
    }
  },
  {
    "type": 3,
    "content": {
      "chartSettings": {
        "chartType": 1,
        "title": "Throttled Requests"
      },
      "query": {
        "resourceType": "microsoft.insights/components",
        "query": "AzureMetrics\n| where Resource contains '{resource}'\n|
where MetricName == 'ThrottledCalls'\n| summarize Total=sum(Total) by
bin(TimeGenerated, 1h)"
      }
    }
  },
  {
    "type": 3,
    "content": {
      "chartSettings": {
```

```
    "chartType": 1,  
    "title": "Latency Percentiles"  
  },  
  "query": {  
    "resourceType": "microsoft.insights/components",  
    "query": "AzureMetrics\n| where Resource contains '{resource}'\n|  
where MetricName startswith 'Latency'\n| summarize avg(Total) by  
bin(TimeGenerated, 1h), MetricName\n| render timechart"  
  }  
}  
]  
}
```

[AI Knowledge](#)