



Connecting  
your world.

# Business GPT

---

Dokumentation

# Inhaltsverzeichnis

---

1. Willkommen bei <b>Business GPT</b>	4
1.1 Business GPT Varianten im Überblick	5
1.2 Hilfe-Themen	8
2. Dokumentation	9
2.1 Einführung	9
2.2 Anmeldung	10
2.3 Business GPT – Entra ID	11
2.4 Allgemeine Funktionen	56
2.5 Chat	58
2.6 Einstellungen	70
2.7 Integrationen	101
3. Ressourcen	105
3.1 Prompting	105
3.2 Prompt Library	111
3.3 Chat mit Dokumenten und internem Wissen	115
3.4 Modell Guide	120
3.5 Assistenten	126
3.6 Function Calling	139
3.7 Integrationsleitfaden für API-Anbieter	141
4. API	144
4.1 Business GPT API	144
4.2 API-Aufruf von BusinessGPT per OAuth2 (Client Credentials)	230
5. Häufig gestellte Fragen	249
5.1 Allgemeine Fragen zum Chat	249
5.2 Fragen zum Chat mit Dokumenten/internem Wissen	250
5.3 Fragen zum Datenschutz	251
5.4 Fragen zur Anmeldung/Login	252
5.5 Fragen zu Kosten und Verbrauch	252
6. Changelog	253
6.1 26. März 2026	253
6.2 18. März 2026	253
6.3 2. März 2026	254
6.4 18. Februar 2026	254
6.5 15. Januar 2026	255
6.6 17. Dezember 2025	255

6.7	28. November 2025	255
6.8	12. November 2025	256
6.9	26. September 2025	256
6.10	11. September 2025	256
6.11	7. August 2025	258
6.12	15. Juli 2025	258
6.13	21. Mai 2025	259
6.14	16. April 2025	259
6.15	24. Februar 2025	260
6.16	11. November 2024	261
7.	Roadmap	262
7.1	Assistenten	262
7.2	Datenquellen und Wissensmanagement	262
7.3	Modellerweiterungen	262
7.4	Chatfunktionen	262
7.5	Rollen und Berechtigungen	263
7.6	Administration	263
7.7	MCP Server Integration	263
8.	Online-Register: Datenstrukturen und Interoperabilität (Data Act)	264
8.1	1. Begriffsbestimmungen	264
8.2	2. Übersicht der Export-Spezifikationen	264
8.3	3. Interoperabilität und Standards	264
8.4	4. Ablauf des Datenexports & Wechsels	265
8.5	5. Sicherheit und Zugriffsschutz	265

# 1. Willkommen bei Business GPT

---

Willkommen auf den offiziellen **Hilfeseiten** von Business GPT – Ihrer sicheren, individuellen und integrierten Unternehmens-KI.

Hier finden Sie alle wichtigen Informationen, um das Beste aus Business GPT herauszuholen. Egal, ob Sie neu dabei sind oder bereits Erfahrung haben – wir unterstützen Sie Schritt für Schritt.

---

## 1.1 Business GPT Varianten im Überblick

---

## Business GPT M

---

Leistungsfähige KI-Chat-Lösung für Teams, die ohne direkte Anbindung an Unternehmensdaten arbeiten möchten.

### Einsatzzweck:

- Allgemeine KI-Unterstützung (Texte, Ideen, Zusammenfassungen)
- Standardisierte Assistenten ohne Firmendaten-Kontext

### Hauptfunktionen:

- KI-Chatbot auf Basis von Large Language Models
- Prompt Library für wiederverwendbare Prompts
- Assistenten ohne Zugriff auf Dokumente / RAG
- Function Calling zur Anbindung externer Funktionen/Services
- Basis-Administration und Konfiguration

### Daten & Wissen:

- Modellwissen (allgemeines Weltwissen)
- Kein Zugriff auf Unternehmensdokumente
- Keine Anbindung externer Datenquellen (z.B. JIRA)
- Keine Dokumentenanhänge im Chat

### Anbindung von Identitätssystemen

- Entra ID (Standard)
- ADFS (Individuelleistung)
- andere IDP (basierend auf SAML, OAuth, und anderen) (Individuelleistung)

### Geeignet für:

- Fachbereiche, die generative KI ohne Datenanbindung nutzen möchten
- Unternehmen, die KI zentral bereitstellen, aber bewusst ohne Zugriff auf interne Dokumente

## Business GPT L

---

Erweiterte Unternehmensvariante mit RAG und Systemintegrationen für einen KI-Chatbot, der aktiv mit Ihren Unternehmensdaten arbeiten kann.

### Einsatzzweck:

- Beantwortung von Fragen auf Basis interner Dokumente und Wissensquellen
- Wissensmanagement & interner Support (IT, HR, Fachbereiche)
- Nutzung von Systemdaten (z.B. Tickets, Vorgänge) direkt im Chat

### Erweiterte Hauptfunktionen (zusätzlich zu M):

- RAG (Retrieval Augmented Generation) zur Nutzung eigener Unternehmensdaten
- Assistenten mit Dokumentenzugriff (über RAG)
- Integration von Datenquellen wie z.B. JIRA  
(z.B. Tickets suchen, Status abfragen, Kontexte einbinden)
- Dokumentenanhänge direkt im Chat verwenden  
(Dokument hochladen → wird im Antwortkontext genutzt)
- Kontextsensitive Antworten auf Basis von Unternehmenswissen (Dokumente & andere Datenquellen)

**Daten & Wissen:**

- Modellwissen (allgemeines Weltwissen)
- Zugriff auf Unternehmensdokumente (RAG)
- Anbindung externer Datenquellen (z.B. JIRA)
- Nutzung von Dokumentenanhängen im Chat

**Anbindung von Identitätssystemen**

- Entra ID (Standard)
- ADFS (Individuelleistung)
- andere IDP (basierend auf SAML, OAuth, und anderen) (Individuelleistung)

**Geeignet für:**






- Produktiven Einsatz mit starkem Unternehmenskontext
- Support- und Wissensportale (First-Level-Support, Self-Service)
- Szenarien, in denen Antworten verbindlich auf internen Informationen basieren müssen

## 1.1.1 Funktionsvergleich Business GPT M vs. L

Bereich	Business GPT M	Business GPT L
Zielgruppe	Teams ohne Bedarf an Daten-/Dokumenten-Anbindung	Teams mit Bedarf an KI auf Basis interner Daten & Systeme
RAG (Nutzung eigener Daten)	<input checked="" type="checkbox"/> Nein	<input checked="" type="checkbox"/> Ja
Unternehmensdokumente	<input checked="" type="checkbox"/> Kein Zugriff	<input checked="" type="checkbox"/> Zugriff via RAG
Datenquellen (z.B. JIRA)	<input checked="" type="checkbox"/> Keine Integration	<input checked="" type="checkbox"/> Integration von Datenquellen (z.B. JIRA)
Prompt Library	<input checked="" type="checkbox"/> Ja	<input checked="" type="checkbox"/> Ja
Assistenten	<input checked="" type="checkbox"/> Ja, ohne Dokumentenzugriff	<input checked="" type="checkbox"/> Ja, inkl. Dokumentenzugriff (RAG-basiert)
Function Calling für Assistenten	<input checked="" type="checkbox"/> Ja	<input checked="" type="checkbox"/> Ja
Dokumentenanhänge im Chat	<input checked="" type="checkbox"/> Nicht verfügbar	<input checked="" type="checkbox"/> Dokumente können hochgeladen und genutzt werden
Mögliche Identitätssysteme	<input checked="" type="checkbox"/> Entra ID <input checked="" type="checkbox"/> ADFS <input checked="" type="checkbox"/> Andere (SAML, OAuth)	<input checked="" type="checkbox"/> Entra ID <input checked="" type="checkbox"/> ADFS <input checked="" type="checkbox"/> Andere (SAML, OAuth)
Typische Nutzung	Generelle KI-Unterstützung ohne Unternehmenskontext	Wissensmanagement & Support mit Unternehmenskontext

## 1.2 Hilfe-Themen

---

Thema	Beschreibung
 <a href="#">Anmeldung</a>	So erhalten Sie Zugang zu Business GPT.
 <a href="#">Chat Funktionen</a>	Nutzen Sie die KI im Chat effektiv für Ihre Aufgaben.
 <a href="#">Allgemeine Funktionen</a>	Entdecken Sie alle Tools und Features im Überblick.
 <a href="#">Administration</a>	Passen Sie Business GPT an Ihre Bedürfnisse an.
 <a href="#">Verbrauchsübersicht</a>	Behalten Sie Ihre Nutzung und Ihr Kontingent im Blick.
 <a href="#">API</a>	So integrieren Sie Business GPT in Ihre Systeme.
 <a href="#">FAQ</a>	Antworten auf häufig gestellte Fragen.
 <a href="#">Ressourcen</a>	Tipps & Tricks für effektives Arbeiten mit Business GPT.


### [Hilfesh Fragen?](#)

Wenn Sie etwas nicht finden oder Hilfe benötigen, kontaktieren Sie unser [Support-Team](#) oder schauen Sie in die [FAQ](#).

### [Suchtipps](#)

Nutzen Sie die Suchfunktion oben rechts, um schnell das passende Thema zu finden.

### [Kunden Community](#)

Werden Sie Teil unserer Kunden Community und profitieren Sie von Community Events, Austausch und top aktuellen Informationen aus der Business GPT Entwicklung. Mehr dazu unter: [Kunden Community](#) 

## 2. Dokumentation

---

### 2.1 Einführung

---

Business GPT bietet eine Vielzahl von Optionen für die Nutzung selbst als auch für die Administration und das Setup der Anwendung. In diesem Bereich werden die verschiedenen Optionen und Aspekte rund um Business GPT erläutert. Sie finden Hilfestellung zur Verwendung und Nutzung der KI-Software als auch zur Einrichtung, Administration und Troubleshooting.

#### 2.1.1 Themenbereiche

---

##### **Anmeldung**

---

Anmeldung an Business GPT mit unterschiedlichen Rollen und Berechtigungen

→ [Anmeldung](#)

##### **Chat Funktionen**

---

Nutzen Sie die KI im Chat effektiv für Ihre Aufgaben.

→ [Chatfunktionen](#)

##### **Allgemeine Funktionen**

---

Entdecken Sie alle Tools und Features im Überblick.

→ [Allgemeine Funktionen](#)

##### **Administration**

---

Passen Sie Business GPT an Ihre Bedürfnisse an.

→ [Administration](#)

## 2.2 Anmeldung

---

Der Zugang zu Business GPT erfolgt über einen dedizierten Link für Ihre Business GPT Umgebung, den Sie nach der Bestellung und Einrichtung Ihrer Umgebung erhalten.

Voraussetzung für den Zugriff auf Business GPT ist eine Anbindung an Ihren Identity Provider (Azure EntraID) oder ADFS. Weitere Anbindungen via OpenId Connect, SAML und weitere sind auf Anfrage möglich.

### 2.2.1 Nutzerrollen

---

Business GPT unterscheidet Anwender nach den Rollen "user" und "admin". Administratoren können Einstellungen einsehen und verändern, während Nutzern dies nicht gestattet ist. Das Setup beider Nutzerrollen erfolgt in Ihrer eigenen Nutzerverwaltung, z.B. EntraID, durch Rollendefinition.

## 2.3 Business GPT – Entra ID

---

Die nachfolgende Dokumentation erläutert die erforderlichen Voraussetzungen sowie die Einrichtung von Entra ID für die Nutzung von Business GPT.

---

### 2.3.1 1 Beistellungen für die Integration des Entra IDs

---

Folgende Informationen werden für das Einrichten von Entra ID seitens der Deutschen Telekom MMS (MMS) benötigt:

- **Directory (tenant) ID**
- **Application (client) ID**

Für den Erhalt der Daten ist es notwendig, dass zunächst eine **App Registration** angelegt wird.

---

## 2.3.2.2 Anlegen einer App Registration

---

### 2.1 Setup App Registration

Folgende Schritte sind im Folgenden auszuführen:

1. Melden Sie sich beim **Microsoft Entra Admin Center** an.
2. Klicken Sie auf **Microsoft Entra ID** im linken Menü und wählen Sie **App registrations** aus.
3. Klicken Sie auf **New registration**.
4. Geben Sie einen **Namen** für Ihre Anwendung ein.
5. Wählen Sie  
**Accounts in this organizational directory only (KUNDE only - Single tenant)**.

*Hinweis:* KUNDE ist an dieser Stelle ein Platzhalter und kann frei benannt werden.

6. Geben Sie eine **Single-page application (SPA)** ein (wird vom Business GPT Support geliefert). Dies ist die URL, auf die Benutzer nach der Anmeldung umgeleitet werden.

## Register an application

**Name**

The user-facing display name for this application (this can be changed later).

### Supported account types

Who can use this application or access this API?

Accounts in this organizational directory only (PrivateGPT1 only - Single tenant)

Accounts in any organizational directory (Any Microsoft Entra ID tenant - Multitenant)

Accounts in any organizational directory (Any Microsoft Entra ID tenant - Multitenant) and personal Microsoft accounts (e.g. Skype, Xbox)

Personal Microsoft accounts only

[Help me choose...](#)

### Redirect URI (optional)

We'll return the authentication response to this URI after successfully authenticating the user. Providing this now is optional and it can be changed later, but a value is required for most authentication scenarios.

Register an app you're working on here. Integrate gallery apps and other apps from outside your organization by adding from [Enterprise applications](#).

By proceeding, you agree to the [Microsoft Platform Policies](#)

**Register**

7. Klicken Sie auf **Register**.

## 2.2 Anpassung des Manifests

Für eine verbesserte Nutzung der API Funktionalitäten und um den Sicherheitsstandard in der Übertragung von Datenverkehr zu erhöhen, ist im Manifest die verwendete Tokenversion anzupassen. Dafür sind die folgenden Schritte durchzuführen:

1. Wählen Sie die zuvor erstellte **App Registration** aus
2. Unter dem Reiter **Manage** klicken Sie auf **Manifest**
3. Ändern Sie die **requestedAccessTokenVersion** innerhalb des Manifests auf den Wert **2** und speichern Sie die Änderung

**KUNDE | Manifest**

Search × Refresh | Got feedback?

Starting from as early as 3/17/2025, and no later than 3/26/2025, you will no longer be able to view, save, upload, or

An app manifest is a JSON representation of an app registration. The Microsoft Graph app manifest and AAD Graph

**Microsoft Graph App Manifest (New)** AAD Graph App Manifest (Deprecating Soon)

See [Understanding the Microsoft Graph application manifest](#) to learn how to edit Microsoft Graph app manifest.

Save Discard Download Upload

```

1 {
2   "id": "355c2df4-623b-4511-bbb5-42fc1343c974",
3   "deletedDateTime": null,
4   "appId": "1d39db91-f451-463e-ad43-9875f1cdb59b",
5   "applicationTemplateId": null,
6   "disabledByMicrosoftStatus": null,
7   "createdByAppId": null,
8   "createdDateTime": "2024-04-25T14:19:23Z",
9   "displayName": "KUNDE",
10  "description": null,
11  "groupMembershipClaims": "All, ApplicationGroup",
12  "identifierUris": [],
13  "isDeviceOnlyAuthSupported": null,
14  "isFallbackPublicClient": null,
15  "nativeAuthenticationApisEnabled": null,
16  "notes": null,
17  "publisherDomain": "privategpt1.onmicrosoft.com",
18  "serviceManagementReference": null,
19  "signInAudience": "AzureADMyOrg",
20  "tags": [],
21  "tokenEncryptionKeyId": null,
22  "samlMetadataUrl": null,
23  "defaultRedirectUri": null,
24  "certification": null,
25  "requestSignatureVerification": null,
26  "addIns": [],
27  "api": [
28    {
29      "acceptMappedClaims": null,
30      "knownClientApplications": [],
31      "requestedAccessTokenVersion": 2,
32      "oauth2PermissionScopes": [],
33      "preAuthorizedApplications": []
34    }
35  ],
36  "appRoles": [
37    {
38      "allowedMemberTypes": [
39        "User",
40        "Application"
41      ],
42      "description": "api role",
43      "displayName": "api",
44      "id": "75063de7-f939-47a3-97f5-11b2f1a54022",
45      "isEnabled": true,
46      "origin": "Application",
47      "value": "api"
48    }
49  ]
50 }

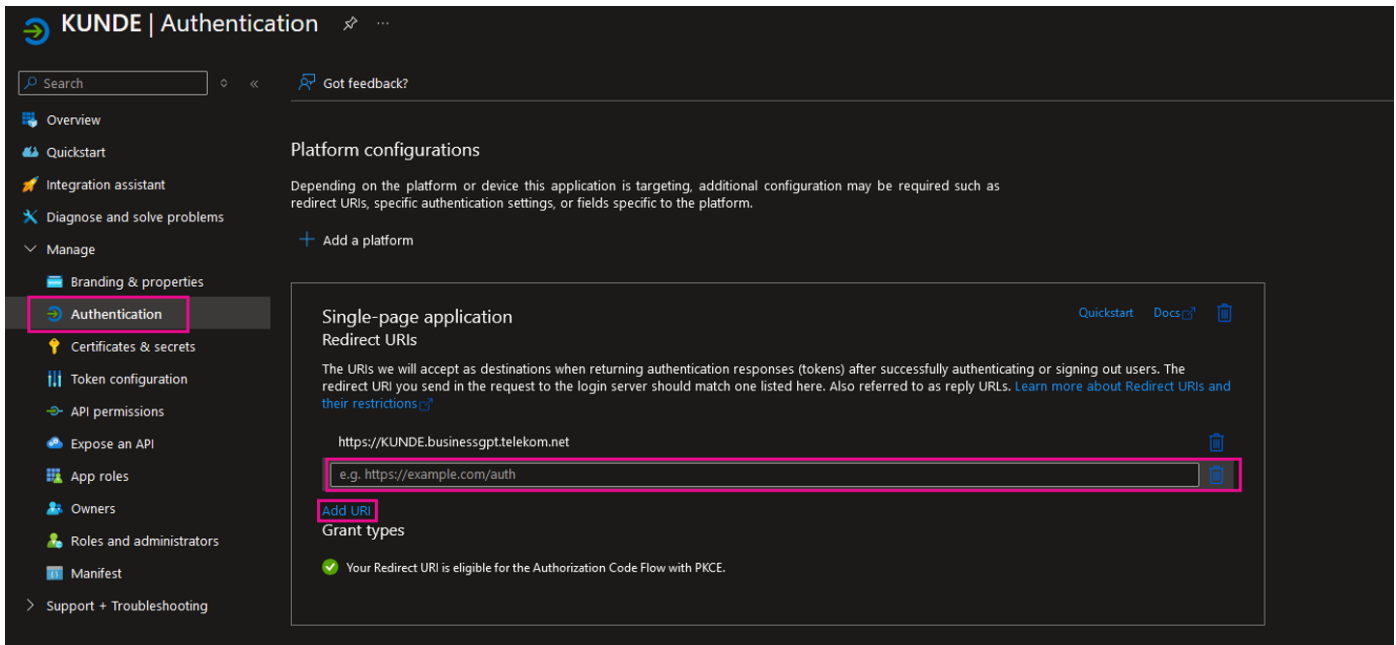
```

Manifest

### 2.3 (Nachträgliches) Hinzufügen einer Redirect-URI

Um eine Redirect-URI für eine Single-Page-Applikation (SPA) in Azure Entra ID **nachträglich** hinzuzufügen, sind die folgenden Schritte durchzuführen:

1. Wählen Sie die **App-Registrierung** aus, für die Sie die Redirect-URI hinzufügen möchten.
2. Klicken Sie auf **Authentifizierung** im linken Menü.
3. Klicken Sie auf **Add URI** im Feld **Single-page application**.
4. Geben Sie die URL der **Redirect-URI** in das Feld ein.
5. Klicken Sie auf **Save**.



## 2.4 Anpassung der Einstellungen in der Enterprise Application

Für das Anpassen von Einstellungen wählen Sie folgende Menüpunkte aus:

1. Navigieren Sie zu **App registrations** und klicken Sie im Anschluss auf **Overview**.
2. Wählen Sie die **Managed application (Enterprise Application)** aus.

The screenshot shows the 'Overview' page for an application named 'KUNDE'. The left sidebar contains navigation options like 'Quickstart', 'Integration assistant', 'Diagnose and solve problems', and 'Manage'. The main content area displays 'Essentials' with fields for Display name, Application (client) ID, Object ID, Directory (tenant) ID, and Supported account types. On the right, 'Client credentials' are listed, including '0\_certificate\_1\_secret', '0\_web\_1\_spa\_0\_public\_client', and 'Add an Application ID URI'. A red box highlights the 'Managed application in L.: KUNDE' field. At the bottom, there is a section titled 'Build your application with the Microsoft identity platform'.

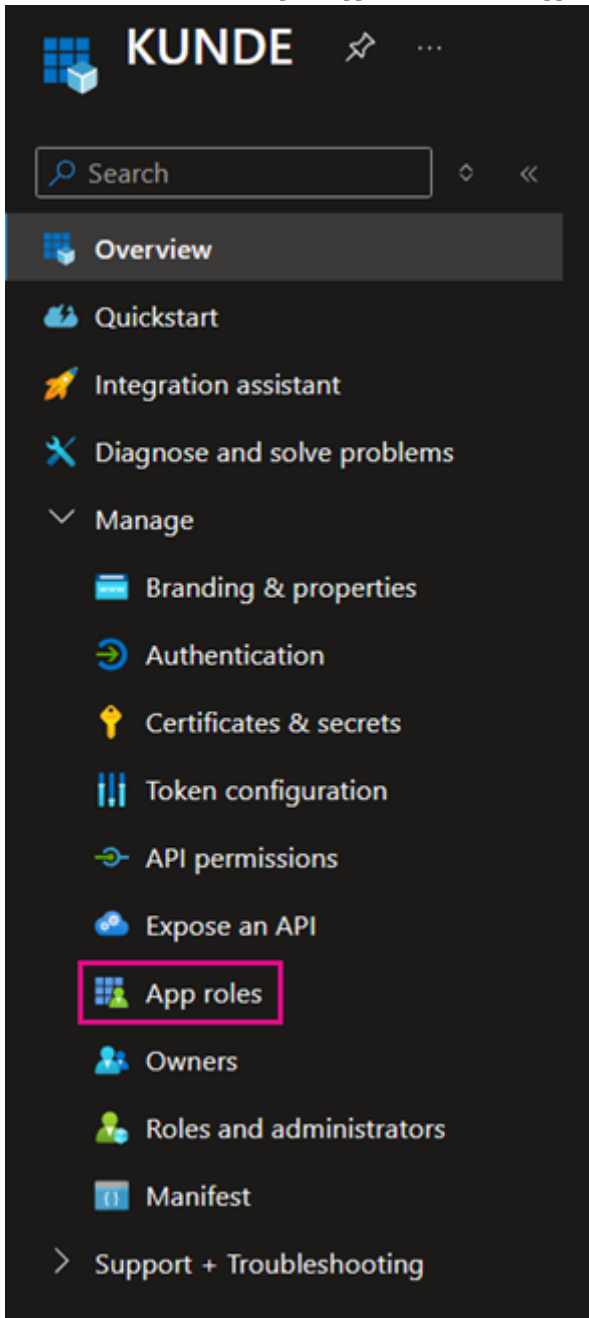
3. Wechseln Sie zu **Properties**.
4. Selektieren Sie **Yes** im Feld **Assignment required**.
5. Klicken Sie abschließend auf **Save**.

The screenshot shows the 'Properties' page for the 'KUNDE' application. The left sidebar lists various management options, with 'Properties' highlighted in red. The main content area has a top bar with 'Save', 'Discard', and 'Delete' buttons. Below this, there is a description of application settings and a section for managing properties. The 'Assignment required?' field is highlighted in red, with the 'Yes' radio button selected. Other fields include 'Enabled for users to sign-in?' (Yes/No), 'Name' (KUNDE), 'Homepage URL', 'Logo' (with a placeholder image containing the letter 'K'), 'Application ID', and 'Object ID'. A 'Notes' field is also visible at the bottom.

#### **2.4.1 Anlegen von Usern, Rollen und zuweisen von Gruppen**

Die User, die Zugriff auf Business GPT erhalten, müssen der App Registration zugewiesen werden. Hierfür müssen zwei Rollen bzw. Gruppen (user und admin) angelegt werden. Dafür folgen Sie den nachstehenden Anweisungen:

1. Wählen Sie innerhalb der **Enterprise Application** den Punkt **App roles** aus.



2. Innerhalb von **App roles** klicken Sie auf **Create app role**.

The screenshot shows the Azure portal interface for managing app roles. The page title is 'KUNDE | App roles'. A search bar is at the top left. A navigation pane on the left lists various management options, with 'App roles' selected and highlighted. The main content area displays a table of existing app roles and a 'Create app role' button highlighted with a red box.

**App roles**

App roles are custom roles to assign permissions to users or apps. The application defines and publishes the app roles and interprets them as permissions during authorization.

[How do I assign App roles](#)

Display name	Description	Allowed member types	Value	ID	State
admin	admin Rolle	Users/Groups,Applications	admin	16e07764-2881-4eb5-...	Enabled
user	user Rolle	Users/Groups,Applications	user	c8627f26-cf5c-4501-a...	Enabled
api.readwrite	allow	Users/Groups,Applications	api.readwrite	876f8d13-fbfc-4fe7-a4...	Enabled
role	was macht die rolle	Users/Groups,Applications	role	ef7c0eff-3f10-45fe-9b...	Enabled

## ADMIN-ROLLE ANLEGEN

1. Legen Sie die **admin role** an.

Hierbei ist es wichtig, dass das Feld **Value** den Wert `admin` enthält.

## Create app role ✕

Display name \* ⓘ

admin ✓

Allowed member types \* ⓘ

Users/Groups

Applications

Both (Users/Groups + Applications)

Value \* ⓘ

admin ✓

Description \* ⓘ

admin Rolle

Do you want to enable this app role? ⓘ

## USER-ROLLE ANLEGEN

1. Legen Sie die **user role** an.

Hierbei ist es wichtig, dass das Feld **Value** den Wert `user` enthält.

## Create app role ✕

Display name \* ⓘ

user ✓

Allowed member types \* ⓘ

Users/Groups

Applications

Both (Users/Groups + Applications)

Value \* ⓘ

user ✓

Description \* ⓘ

user Rolle

Do you want to enable this app role? ⓘ

**OPTIONAL: API-ROLLE ANLEGEN**

1. Legen Sie (optional) die **api role** an, sofern Sie die BGPT API ansprechen möchten. Hierbei ist es wichtig, dass das Feld **Value** den Wert `api` enthält

# Create app role



Display name \* ⓘ

api



Allowed member types \* ⓘ



Users/Groups



Applications



Both (Users/Groups + Applications)

Value \* ⓘ

api



Description \* ⓘ

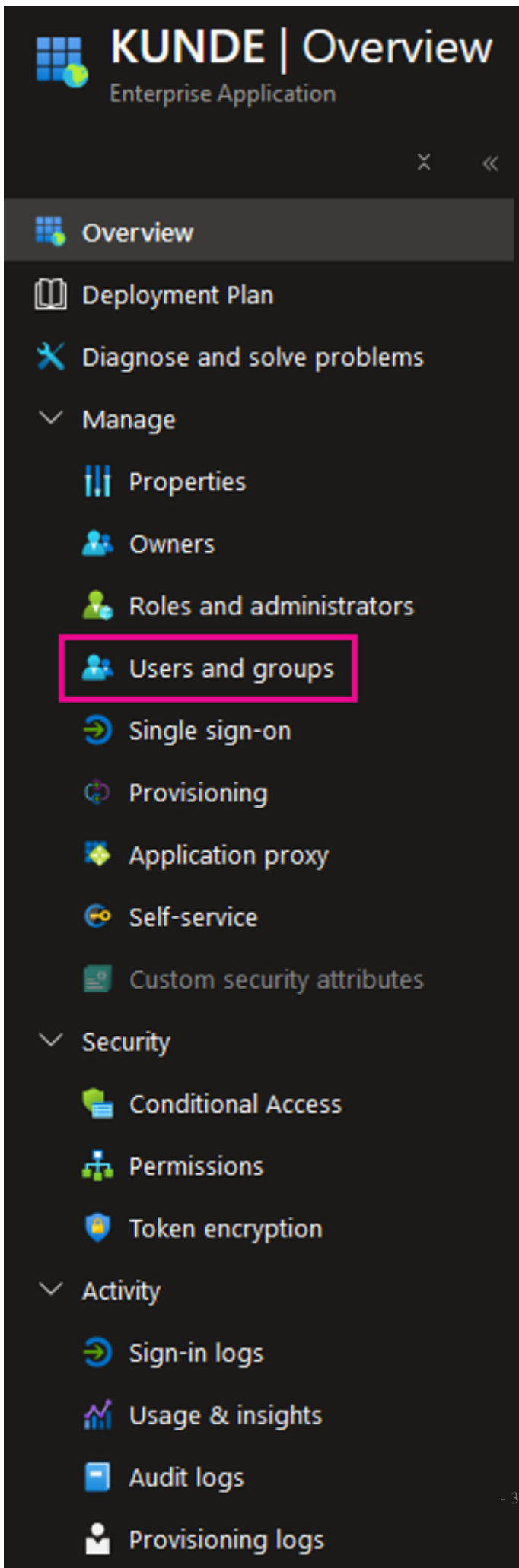
api role

Do you want to enable this app role? ⓘ



**USER/GRUPPEN DER APPLIKATION ZUWEISEN**

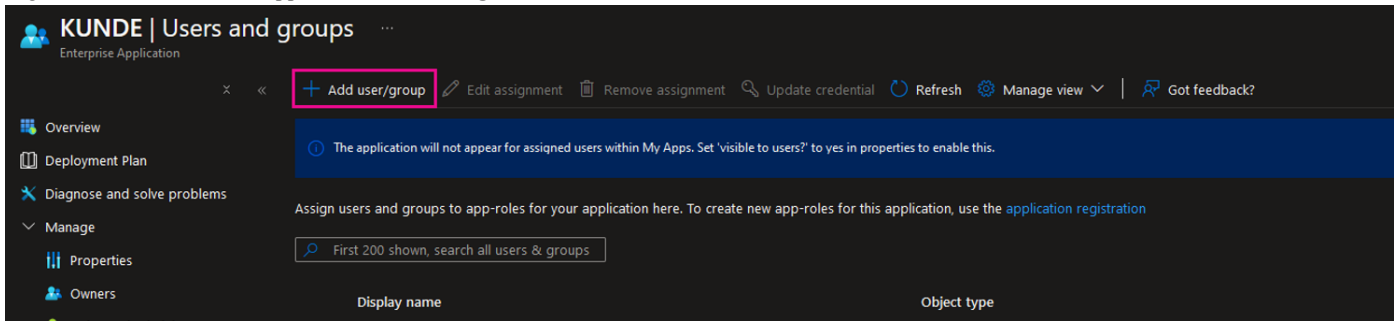
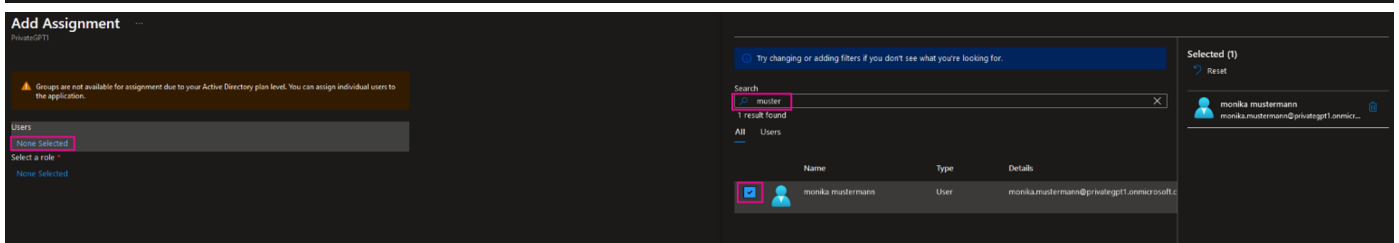
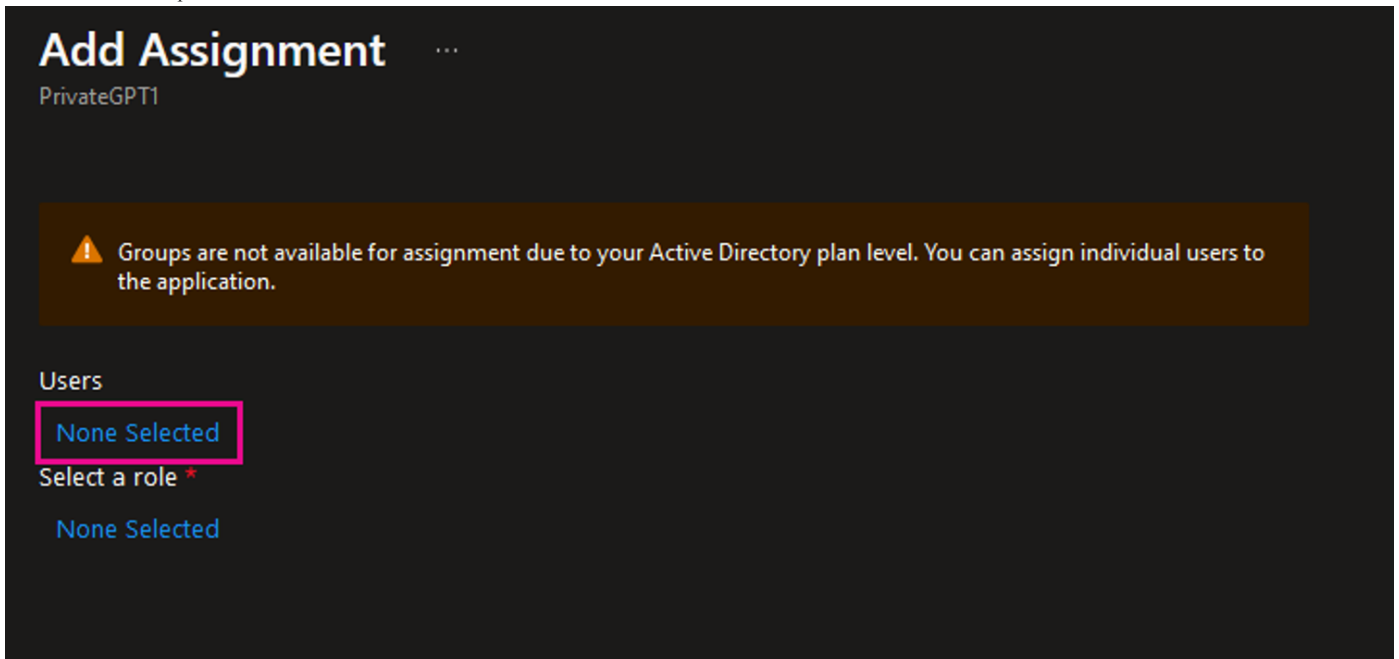
1. Wechseln Sie wieder in die **Enterprise Application** und wählen Sie **Users and groups** aus.



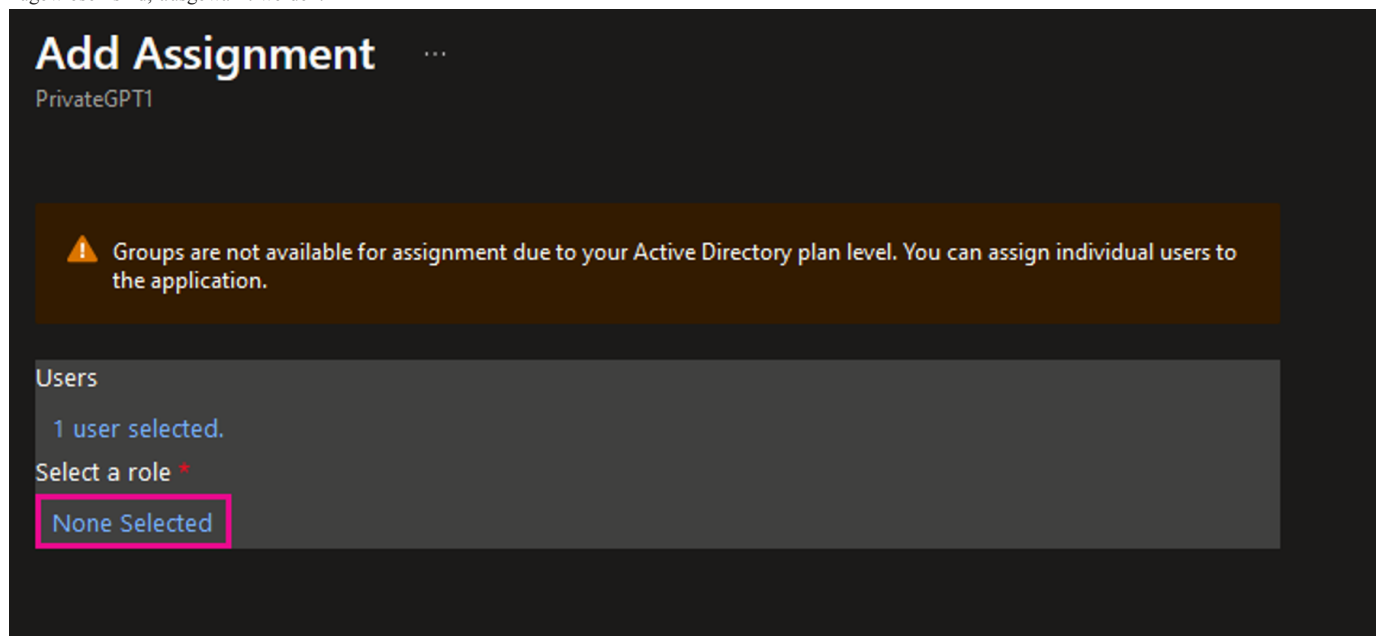
**KUNDE | Overview**  
Enterprise Application

Overview

- Deployment Plan
- Diagnose and solve problems
- Manage
  - Properties
  - Owners
  - Roles and administrators
  - Users and groups**
  - Single sign-on
  - Provisioning
  - Application proxy
  - Self-service
  - Custom security attributes
- Security
  - Conditional Access
  - Permissions
  - Token encryption
- Activity
  - Sign-in logs
  - Usage & insights
  - Audit logs
  - Provisioning logs

2. Fügen Sie neue **User- und Gruppen/Rollen-Beziehungen** hinzu.3. Wählen Sie die entsprechenden User über **None selected** aus.4. Weisen Sie eine Rolle für die zuvor ausgewählten User über **None selected** zu.

Sofem mindestens eine **P1 Lizenz (Entra ID)** vorliegt, können an dieser Stelle für den entsprechenden User ebenfalls **Gruppen**, die der App Registration zugewiesen sind, ausgewählt werden.



**Add Assignment** ...

PrivateGPT1

**!** Groups are not available for assignment due to your Active Directory plan level. You can assign individual users to the application.

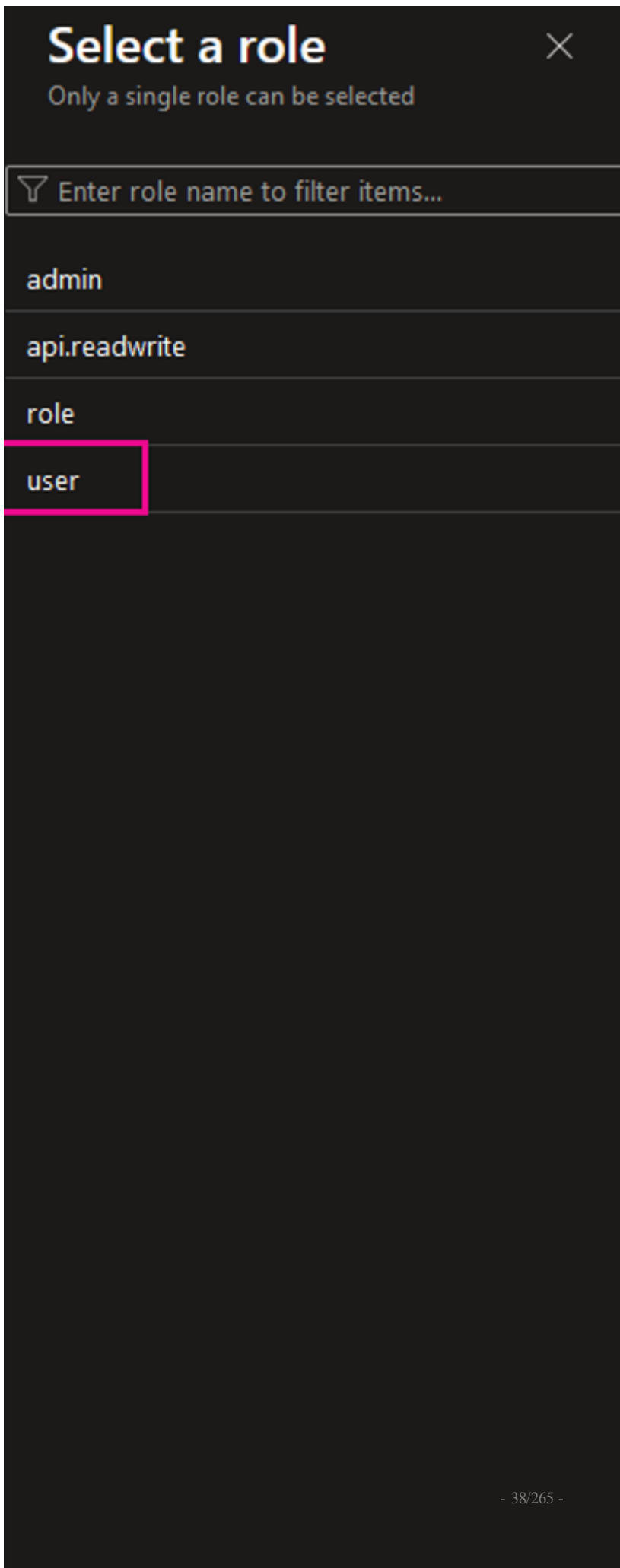
Users

1 user selected.

Select a role \*

None Selected


1. Wählen Sie `user` oder `admin` für die User aus.



2. Bestätigen Sie die getroffene Auswahl über **Assign**.

# Add Assignment

PrivateGPT1

 Groups are not available for assignment due to your Active Directory plan level. You can assign individual users to the application.

Users

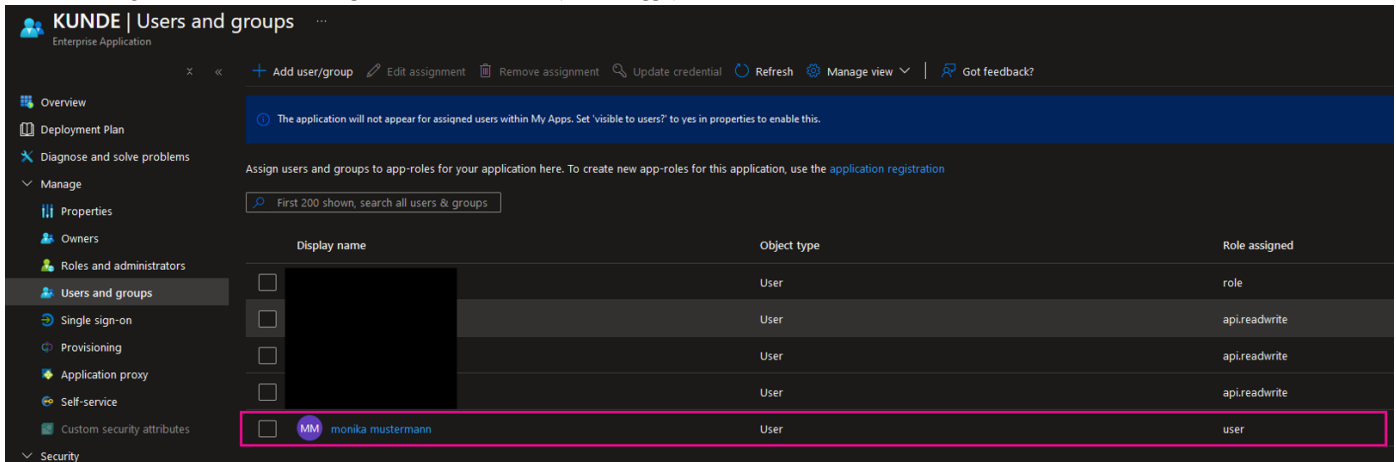
1 user selected.

Select a role \*

user

Assign

3. Nach Erstellung ist der User mit den entsprechenden Credentials (Rolle/Gruppe) sichtbar.

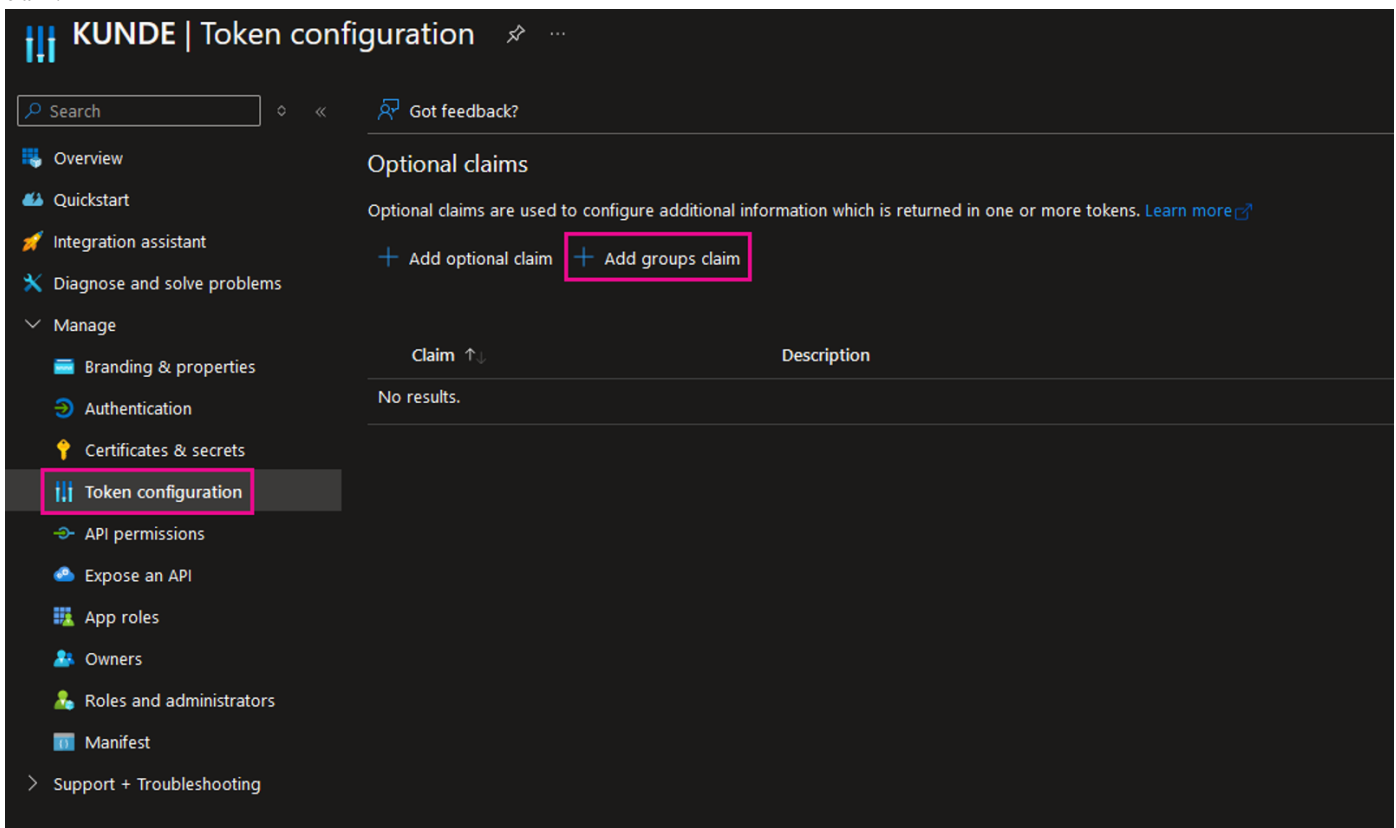


Display name	Object type	Role assigned
[Redacted]	User	role
[Redacted]	User	api.readwrite
[Redacted]	User	api.readwrite
[Redacted]	User	api.readwrite
monika mustermann	User	user

## 2.5 Einrichten von Group Claiming

Durch das **Groups Claiming** wird gesteuert, welche Gruppen via Token beim Zugriff auf Business GPT übermittelt werden. Diese werden genutzt, um den Zugriff auf Container (ab Paket L) zu administrieren. Gehen Sie wie folgt vor:

1. Wählen Sie innerhalb der zuvor erstellten **App Registration** unter dem Reiter **Manage** den Punkt **Token configuration** aus und klicken Sie auf **Add groups claim**.



2. Die Auswahl der **group types** ist abhängig von der in Ihrem Unternehmen vorliegenden Entra ID Struktur und der Lizenz.

Standardmäßig wählen Sie bei einem **lizensierten Entra ID** folgenden group type aus und klicken abschließend auf **Add** (das Groups Claiming kann jederzeit gelöscht und neu erstellt werden, um mit den Einstellungen zu testen):

• **Groups assigned to the application (recommended for large enterprise companies ...)**

- Diese Option hat den Vorteil, dass nur der Enterprise-Applikation zugewiesene Gruppen im Anmeldetoken übertragen werden (sinnvoll für Gruppenzuordnung in BPT).

# Edit groups claim



**i** Adding the groups claim applies to Access, ID, and SAML token types. [Learn more](#)

Select group types to include in Access, ID, and SAML tokens.

- Security groups
- Directory roles
- All groups (includes 3 group types: security groups, directory roles, and distribution lists)
- Groups assigned to the application (recommended for large enterprise companies to avoid exceeding the limit on the number of groups a token can emit)

## Customize token properties by type

### ^ ID

- Group ID
- sAMAccountName
- NetBIOSDomain\sAMAccountName
- DNSDomain\sAMAccountName
- On Premises Group Security Identifier
- Emit groups as role claims

### ^ Access

- Group ID
- sAMAccountName
- NetBIOSDomain\sAMAccountName
- DNSDomain\sAMAccountName
- On Premises Group Security Identifier
- Emit groups as role claims

### ^ SAML

- Group ID
- sAMAccountName
- NetBIOSDomain\sAMAccountName
- DNSDomain\sAMAccountName

- **All Groups** ggf. verwenden, wenn **keine Entra-ID Lizenzierung** vorhanden ist (Free Version).
- In einer nicht lizenzierten Version können der Enterprise-Applikation **keine Gruppen direkt zugeordnet** werden.
- **Directory roles** können ggf. auch weggelassen werden.

## Edit groups claim



**i** Adding the groups claim applies to Access, ID, and SAML token types. [Learn more](#)

Select group types to include in Access, ID, and SAML tokens.

- Security groups
- Directory roles
- All groups (includes 3 group types: security groups, directory roles, and distribution lists)
- Groups assigned to the application (recommended for large enterprise companies to avoid exceeding the limit on the number of groups a token can emit) **i**

### Customize token properties by type

#### ^ ID

- Group ID
- sAMAccountName
- NetBIOSDomain\sAMAccountName
- DNSDomain\sAMAccountName
- On Premises Group Security Identifier
- Emit groups as role claims **i**

#### ^ Access

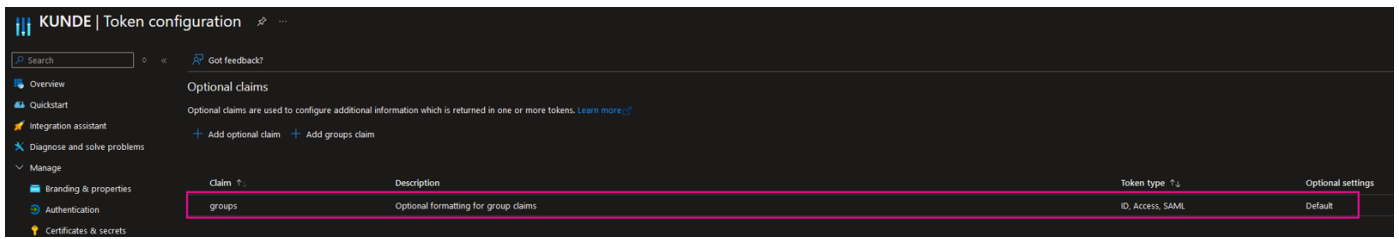
- Group ID
- sAMAccountName
- NetBIOSDomain\sAMAccountName
- DNSDomain\sAMAccountName
- On Premises Group Security Identifier
- Emit groups as role claims **i**

#### ^ SAML

- Group ID
- sAMAccountName
- NetBIOSDomain\sAMAccountName

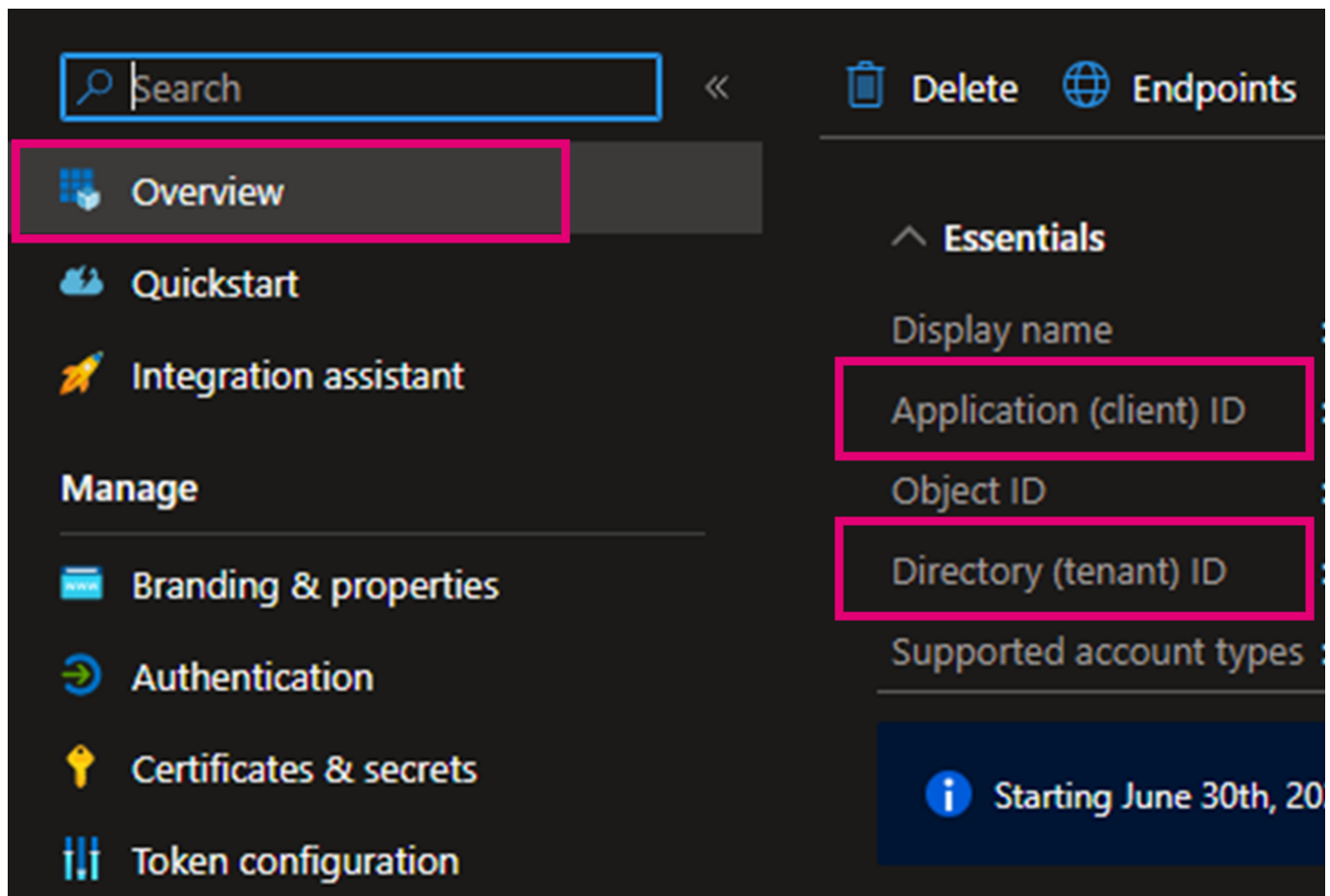
1. Nach der Erstellung des entsprechenden **Group Claims** wird dieser in der Übersicht gelistet.

### 2.3.3



### 2.6 Informationen zu Directory (tenant) ID und Application (client) ID ermitteln

Die nach dem Anlegen der App Registration erzeugten Informationen zu **Directory (tenant) ID** sowie **Application (client) ID** sind unter **Overview** innerhalb der App Registration zu finden:



Diese Informationen werden an die MMS übermittelt.

#### 2.3.4 3 Die App berechtigen

Der Admin des Entra IDs muss sich zuerst innerhalb von **Business GPT** anmelden.

Im Anschluss wird folgendes Pop-up-Fenster angezeigt:

1. Setzen Sie den Haken bei **Zustimmen im Namen Ihrer Organisation**.
2. Bestätigen Sie mit **Akzeptieren**.



## Angeforderte Berechtigungen

### App-Info

**Diese Anwendung wird nicht von Microsoft oder Ihrer Organisation veröffentlicht.**

Diese App benötigt folgende Berechtigungen:

- ✓ Grundlegendes Profil von Benutzern anzeigen
- ✓ Zugriff auf Daten beibehalten, für die Sie Zugriff erteilt haben
- Zustimmung im Namen Ihrer Organisation

Wenn Sie zustimmen, erhält diese App Zugriff auf die angegebenen Ressourcen für alle Benutzer in Ihrer Organisation. Niemand sonst wird zur Überprüfung dieser Berechtigungen aufgefordert.

Durch das Akzeptieren dieser Berechtigungen gestatten Sie dieser App, Ihre Daten gemäß den Vertragsbedingungen und den Datenschutzbestimmungen zu verwenden. Unter <https://myapps.microsoft.com> können Sie diese Berechtigungen ändern. [Details anzeigen](#)

Wirkt diese App verdächtig? [Hier melden](#)

Abbrechen

Akzeptieren

Damit sind die erforderlichen Berechtigungen für die App erteilt.

---

## 2.3.5 4 (Optional) Anlegen API-User

---

Möchten Sie einen **API-User** anlegen, gehen Sie bitte folgendermaßen vor. Bei Bedarf nutzen Sie bitte die detaillierte Beschreibung der nötigen Schritte hier: [API-Anbindung](#)

### 4.1 Notwendige Informationen für das Anlegen des API-Users

Die nachfolgenden Informationen sind für die Nutzung des API-Users erforderlich.

Um diese Informationen zu erhalten, müssen die Schritte, beschrieben ab Kapitel [App Registration](#), durchlaufen werden.

**API-User App Registration (Parameter):**

- **Application (client) ID**
  - **Client secret Value**
  - **Directory (tenant) ID**
  - **Backend URL**
- 

### 4.2 Anlegen der App Registration für die Nutzung des API-Users

In diesem Schritt wird eine weitere **App Registration** angelegt, die einen geringeren Berechtigungsumfang aufweist als die zuvor eingerichtete.

Damit können die Rechte des API-Users administriert und Fehlbedienungen vermieden werden.

Gehen Sie wie folgt vor:

1. Geben Sie im Feld **Name** eine von Ihnen frei gewählte Bezeichnung ein.
2. Unter **Supported account types** ist zwingend die Option **Accounts in this organizational directory only (KUNDE only – Single tenant)** auszuwählen.
3. Klicken Sie danach auf **Register**.

## 2.3.6

## Register an application ...

**\* Name**

The user-facing display name for this application (this can be changed later).

 ✓

### Supported account types

Who can use this application or access this API?

- Accounts in this organizational directory only ( **KUNDE** only - Single tenant)
- Accounts in any organizational directory (Any Microsoft Entra ID tenant - Multitenant)
- Accounts in any organizational directory (Any Microsoft Entra ID tenant - Multitenant) and personal Microsoft accounts (e.g. Skype, Xbox)
- Personal Microsoft accounts only

[Help me choose...](#)

### Redirect URI (optional)

We'll return the authentication response to this URI after successfully authenticating the user. Providing this now is optional and it can be changed later, but a value is required for most authentication scenarios.

Select a platform ▼

Register an app you're working on here. Integrate gallery apps and other apps from outside your organization by adding from [Enterprise applications](#).

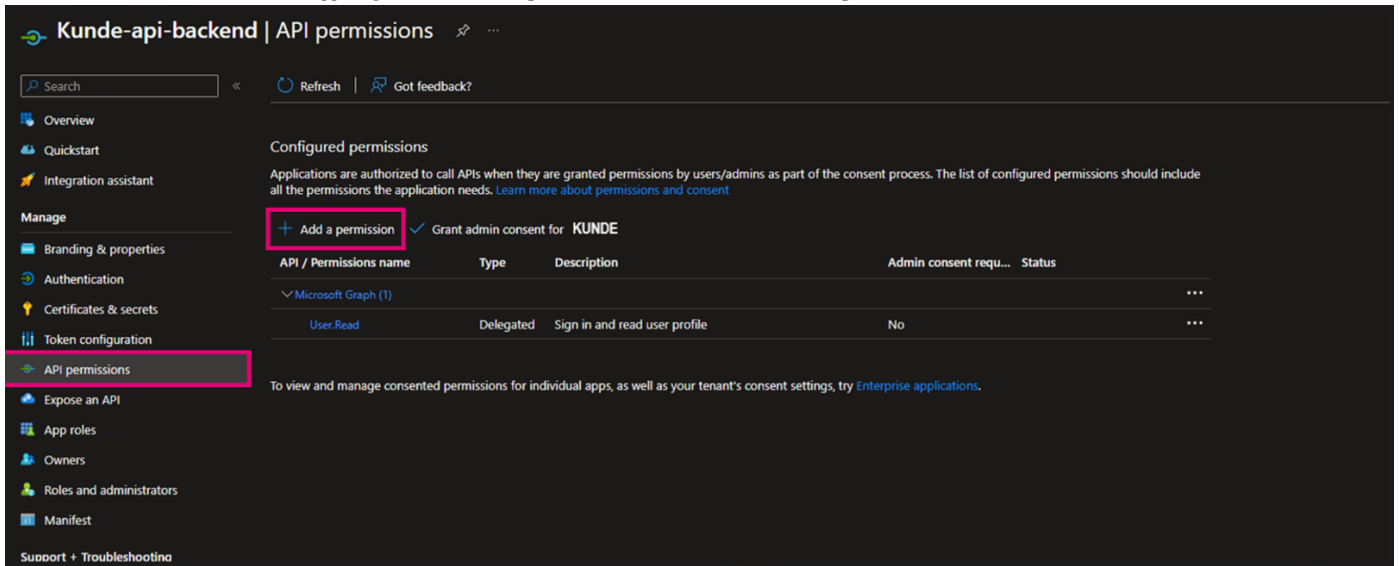
By proceeding, you agree to the [Microsoft Platform Policies](#) [↗](#)

Register

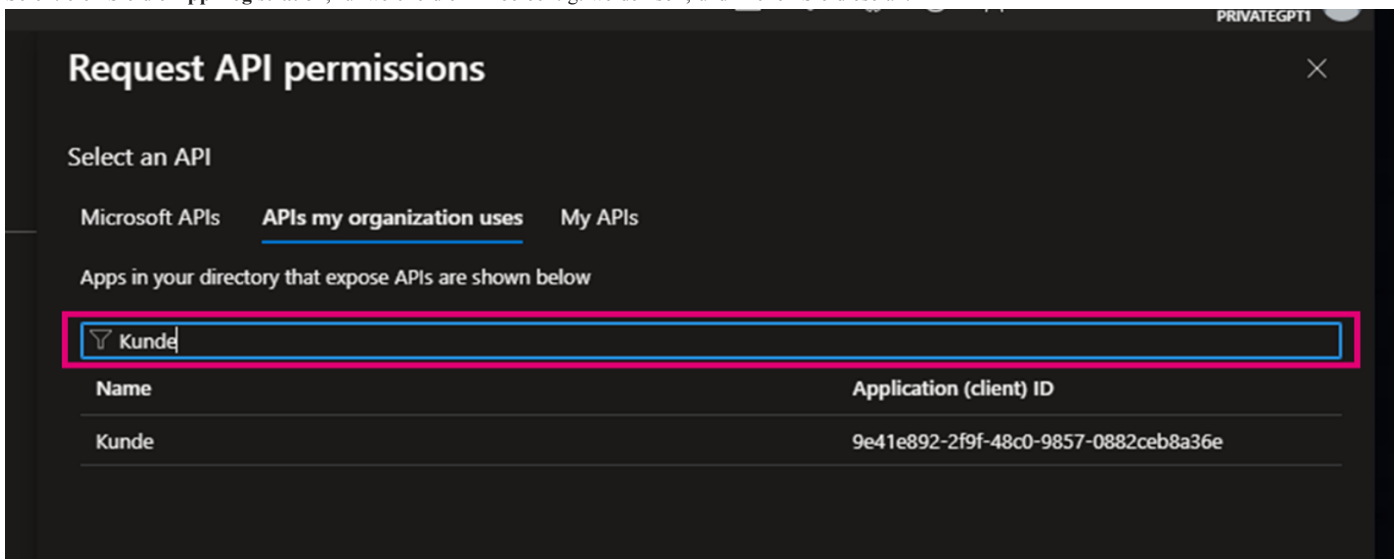
#### 4.3 Hinzufügen von Berechtigung (API permission)

Im nächsten Schritt wird die Berechtigung für den Zugriff auf die zuvor erstellte App-Registrierung erteilt.

1. Klicken Sie im Menü der erstellten App Registration auf **API permissions** und wählen dort **Add a permission** aus.



2. Selektieren Sie die **App Registration**, für welche die API berechtigt werden soll, und klicken Sie diese an.



3. Klicken Sie auf **Application permissions** und wählen Sie die Option **api** aus.

4. Klicken Sie abschließend auf **Add permissions**.

# Request API permissions

[← All APIs](#)

**K** KUNDE

What type of permissions does your application require?

Delegated permissions

Your application needs to access the API as the signed-in user.

Application permissions

Your application runs as a background service or daemon without a signed-in user.

Select permissions

[expand all](#)

Permission

Admin consent required

Other permissions (1)

<input type="checkbox"/>	admin ⓘ admin	Yes
<input checked="" type="checkbox"/>	api ⓘ api	Yes
<input type="checkbox"/>	role ⓘ role	Yes
<input type="checkbox"/>	user ⓘ user	Yes

> api

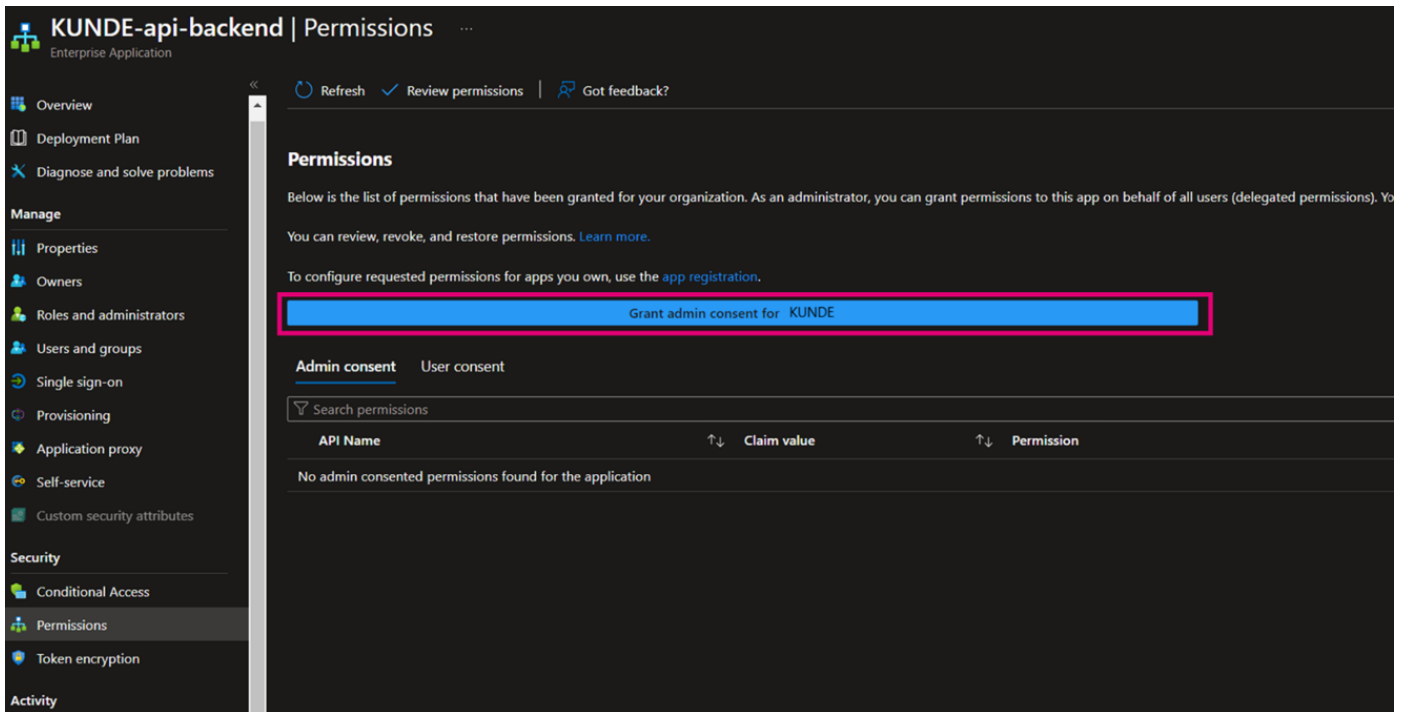
Add permissions

Discard

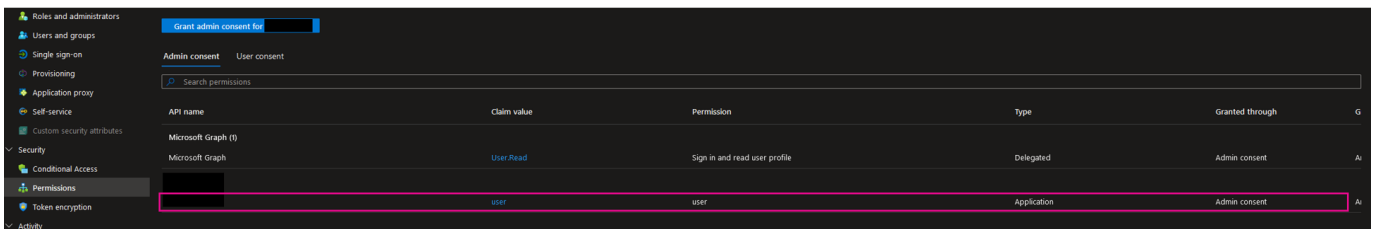
#### 4.4 Permission Grant admin consent

Zum Schluss muss die Berechtigung freigeschaltet werden.

1. Wählen Sie in der für den API-User erstellten App Registration unter dem Reiter **Security** die Maske **Permissions** aus.
2. Klicken Sie im Anschluss auf **Grant admin consent für KUNDE**.



In der Übersicht der zugewiesenen Berechtigungen sehen Sie anschließend die freigeschalteten Permissions.



## 2.4 Allgemeine Funktionen

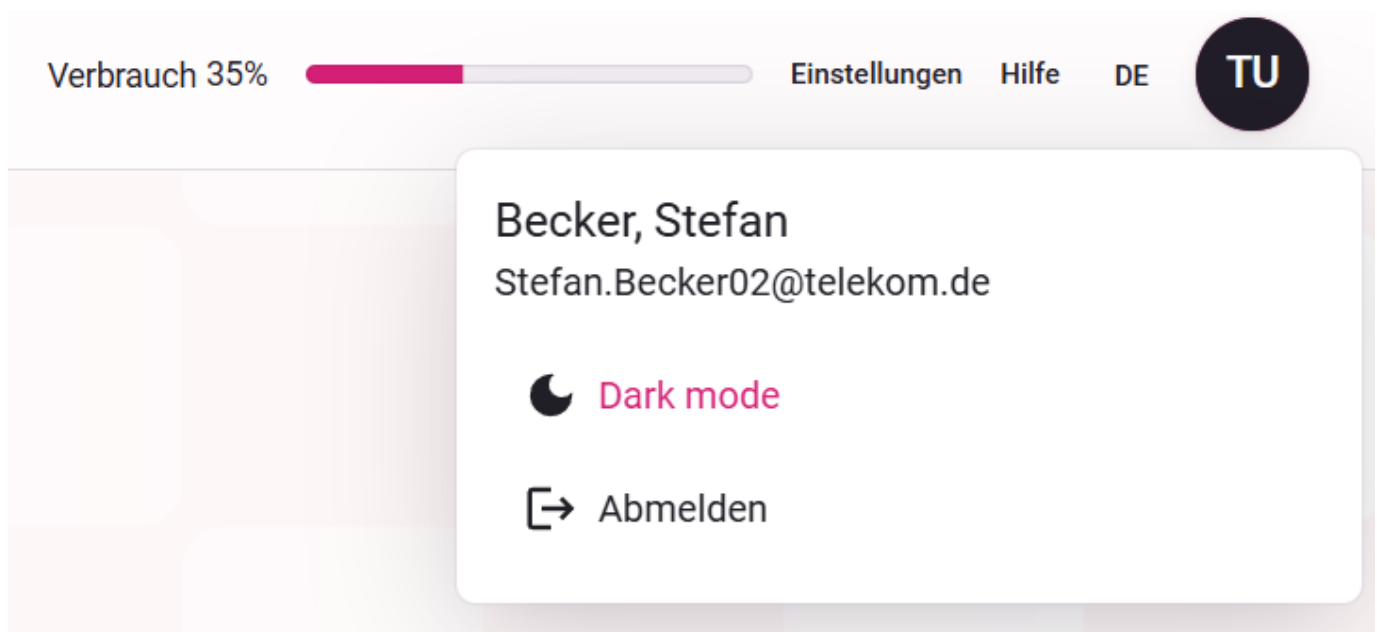
### 2.4.1 Hilfe

Über die Schaltfläche [Hilfe](#) können gelangen Sie in den Online-Hilfe-Bereich. Dort erhalten Sie Hilfestellung zur Administration und Nutzung von Business GPT.



### 2.4.2 Dark Mode / Light Mode

Über die Schaltfläche "Wechseln zu Dark Mode" kann die Weboberfläche wahlweise in dunklem oder hellem Theme angezeigt werden.



### 2.4.3 Einstellungen

Über die Schaltfläche "Einstellungen" gelangt man in die Einstellungen. Diese Seite kann nur durch Administratoren aufgerufen werden.

Hier finden sich verschiedene Bereiche für administrative Zwecke.



### 2.4.4 Sprachauswahl

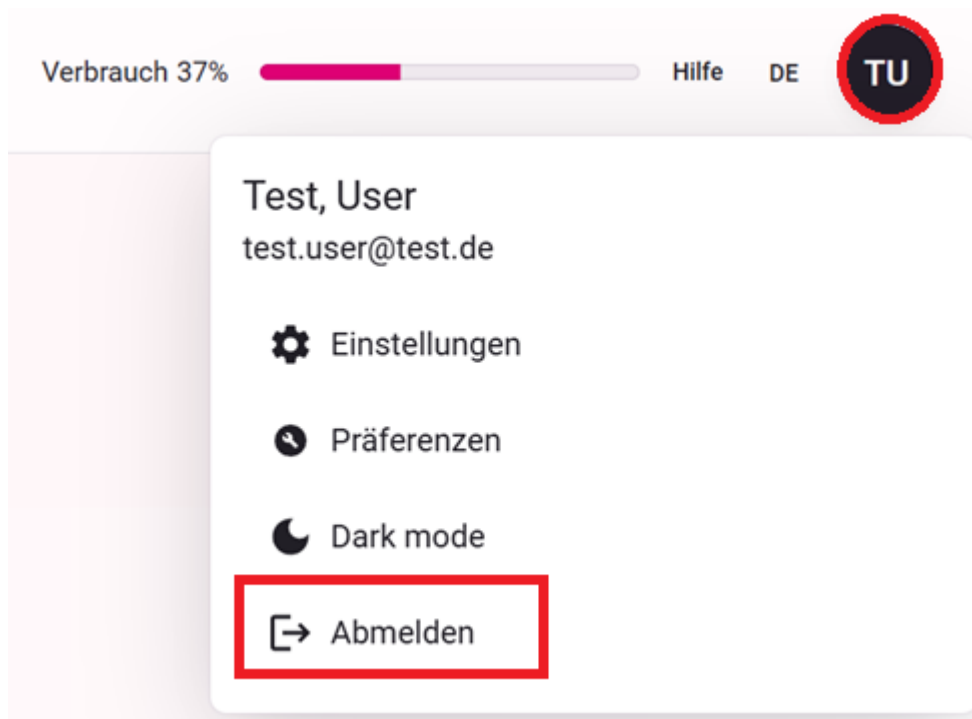
Über die Sprachauswahl kann zwischen den Sprachen Deutsch und Englisch gewechselt werden.



## 2.4.5 Abmelden

---

Über die Schaltfläche zum Abmelden kann man sich von Business GPT abmelden und somit die Anwendung schließen/beenden.



## 2.4.6 Nutzungsbedingungen

---

Der Link [Nutzungsbedingungen](#) im unteren Bereich der Seitennavigation führt zu einer Informationsseite, auf der die Nutzungsbedingungen zu Business GPT beschrieben werden.

## 2.4.7 Datenschutz

---

Der Link [Datenschutz](#) im unteren Bereich der Seitennavigation führt zu einer Informationsseite auf der beschrieben wird, wie Business GPT Daten behandelt.

## 2.4.8 Impressum

---

Der Link [Impressum](#) im unteren Bereich der Seitennavigation führt zum Impressum der Deutschen Telekom MMS GmbH.

## 2.5 Chat

---

### 2.5.1 Chatten mit Business GPT

---

Chats sind das Herzstück Ihrer Arbeit mit Business GPT. Stellen Sie sich den Chat wie ein Gespräch mit einem hochqualifizierten Kollegen vor, der niemals schläft. Anders als herkömmliche Suchmaschinen versteht Business GPT den Kontext Ihrer gesamten Unterhaltung. Sie können Rückfragen stellen, Ergebnisse verfeinern und sich auf frühere Nachrichten beziehen, ohne jedes Mal von vorn beginnen zu müssen. Ob zur Analyse komplexer Dokumente, zur Erstellung von Inhalten oder zum Lösen geschäftlicher Aufgaben – dein Business GPT ist hier, um deinen Arbeitsalltag effizienter zu gestalten.

#### Hauptfunktionen

##### Natürliche Unterhaltung

---

Führen Sie natürlichsprachliche Unterhaltung mit Business GPT und stellen Sie Folgefragen. Business GPT merkt sich den Gesprächsverlauf und kann Zusammenhänge herstellen.

##### Dateianhänge

---

Fügen Sie Dokumente und Bilder dem Chat hinzu für Auswertungen, Zusammenfassungen sowie Extraktion von Informationen und Daten.

##### Assistenten-Auswahl

---

Wechseln Sie zwischen verschiedenen KI-Assistenten und Modellen. So finden Sie den besten Gesprächspartner für Ihre Aufgabe.

##### Prompt speichern

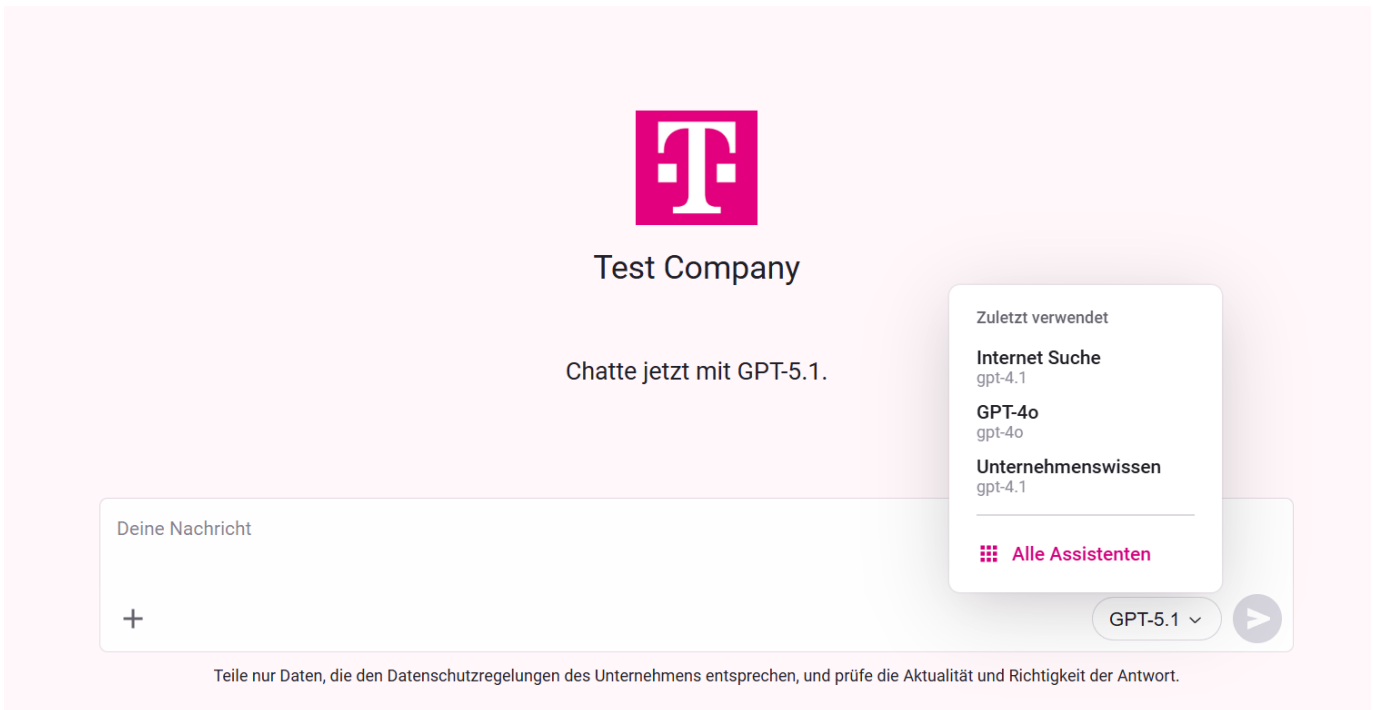
---

Speichern Sie hilfreiche Prompts (Texteingaben) dauerhaft, um diese für wiederkehrende Aufgaben zu verwenden.

#### Modell/Assistenten wählen

Business GPT bietet standardmäßig eine Reihe von vordefinierten Modellen und Assistenten an, die für Chat-Konversationen verwendet werden können.

Über die Schaltfläche im Chat-Eingabefeld kann der Assistent festgelegt werden, den Business GPT für die aktuelle Konversation verwenden soll. Der Assistent kann jederzeit während der Konversation gewechselt werden.



Ausgewählt werden können neben Standard Assistenten auch weitere, eigene Assistenten (siehe dazu auch [Assistenten](#)).

#### STANDARD ASSISTENTEN

Aktuell stehen folgende Modelle/Assistenten zur Verfügung: `GPT-4.1`, `GPT-4.1 (mini)`, `GPT-4.1 (nano)`, `GPT-o1`, `GPT-o3 (mini)`, `GPT-5`, `GPT-5 (mini)`, `GPT-5 (nano)`, `Bildgenerator`, `Unternehmenswissen`, `Internetsuche`

Alle Standard Assistenten sind für die sofortige Nutzung optimiert.

 **Modellübersicht**

Für weitere Informationen und Hilfestellung für die vorhandenen Modelle nutzen Sie unseren [Modell-Guide](#).

#### Internet Suche

Der Assistent "Internet Suche" basiert auf dem Large Language Modell GPT4o von OpenAI und verfügt über eine Anbindung an das Internet via Bing Search. Für Suchanfragen mit diesem Assistenten werden demnach auch Quellen aus dem Internet verwendet. Wenn Sie in einem bereits begonnenen Chat zum Assistenten "Internet Suche" wechseln, müssen Sie einen neuen Chat verwenden, um zu verhindern, dass interne Informationen, die sich bereits im Chatverlauf befinden, ins Internet gelangen.

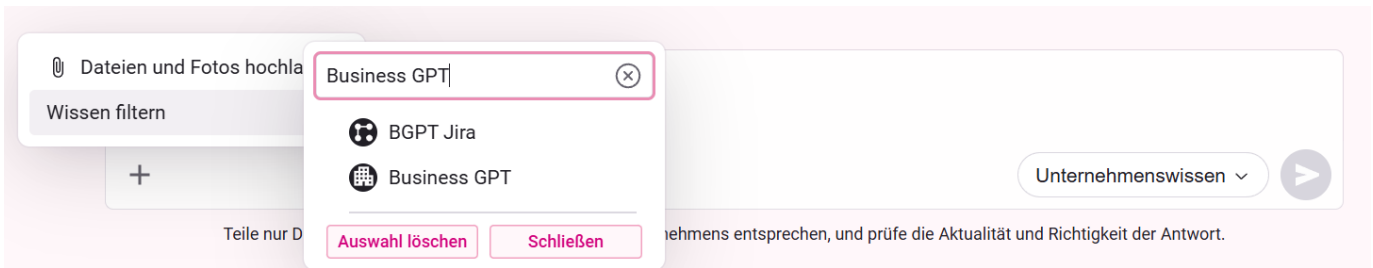
Der Assistent "Internet Suche" steht nur dann zur Verfügung, wenn Sie über die Zubuchoption "Internet Breakout" verfügen. Diese können Sie jederzeit nachbuchen.

#### Bildgenerator

Mit dem Assistenten `Bildgenerator` können Bilder mit `Image-1.5` erzeugt werden. Image-1.5 ist ein KI-Modell von OpenAI, das aus Texteingaben realistische und kreative Bilder generiert. Es versteht komplexe Beschreibungen und kann daraus detailreiche Illustrationen, Kunstwerke oder fotorealistische Szenen erstellen. Image-1.5 eignet sich besonders für kreative Anwendungen wie Design, Werbung oder visuelle Ideenfindung.

#### Unternehmenswissen

Business GPT in der Paketvariante "L" bietet die Möglichkeit mit zusätzlichem, internem (Unternehmens-) Wissen zu chatten. Der Assistent `Unternehmenswissen` verwendet dafür als Quelle die zuvor in Containern bereitgestellten und hochgeladenen Dokumente (siehe [Dokumente](#)) sowie Daten, die über Integrationen wie `Jira` oder `Staffbase` synchronisiert wurden. Hierbei können KI-Anfragen unter Einbezug von Dokumenten und weiteren internen Daten beantwortet werden. Um den Assistenten zu verwenden, muss in der Assistentenauswahl **Unternehmenswissen** gewählt werden. Zusätzlich kann eine Auswahl eines oder mehrerer Wissenscontainer ("Wissen filtern") erfolgen, die als Filter verwendet werden sollen.



Erfolgt keine Einschränkung/Auswahl erfolgt die Informationssuche unter Einbezug aller Container. Zur Auswahl werden nur Container angeboten, für deren Zugriff Sie berechtigt sind.

Die Wissens-Quellen, auf denen die KI-Antwort beruht, werden in der Antwort als Quellen ausgewiesen. Bei Bedarf kann die referenzierte Quelle überprüft werden, in dem das zugehörige Dokument angezeigt oder heruntergeladen werden kann. Bei Integrationsquellen kann die Quelle über einen URL-Link geöffnet werden.



Hilfeseite

Stehen in der Containerauswahl keinerlei Container zur Verfügung/Auswahl, sind Sie für keinerlei Container berechtigt. Entweder ist die Ursache dafür, dass Sie nicht über die nötigen Gruppenmitgliedschaften verfügen oder keine Gruppen, in denen Sie Mitglied sind, für Container berechtigt sind.



Funktionsweise

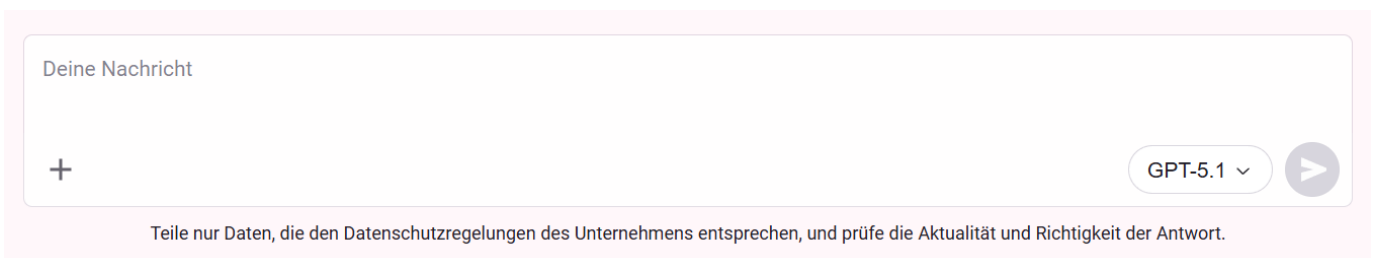
Weiterführende Hilfe zur Funktionsweise des Assistenten **Unternehmenswissen** finden Sie hier: [Dokumenten-Chat](#)

#### WEITERE ASSISTENTEN

Wurden weitere Assistenten erstellt, für die Sie zugriffsberechtigt sind, erscheinen diese in der Auswahl unter "Weitere Assistenten" und können dort ausgewählt werden.

#### Text-/Eingabefeld

Über das Text-/Eingabefeld erfolgt die Kommunikation mit Business GPT. Hier werden Befehle, Prompts und Fragen eingegeben und Konversationen mit dem ausgewählten Modell geführt. Business GPT reagiert sprachsensitiv und beantwortet Fragen in der Sprache, in der sie gestellt wurden, unabhängig davon, welche Sprache (Deutsch oder Englisch) eingestellt wurde.




Prompting How-To

Tipps & Tricks für effiziente Anfrage an den Chatbot finden Sie hier: [Prompt Engineering](#)

#### NACHRICHT SENDEN

Über die Schaltfläche "Nachricht senden" wird die aktuelle Eingabe aus dem Textfeld übermittelt. Alternativ können Sie Ihre Nachricht auch einfach per **Enter** absenden.

Hallo!

GPT-5.1 

Teile nur Daten, die den Datenschutzregelungen des Unternehmens entsprechen, und prüfe die Aktualität und Richtigkeit der Antwort.


### Seitenmenü


Über die Schaltfläche mit dem kleinen Pfeil kann das seitliche Menü (Seitennavigation) geschlossen oder geöffnet werden.


### CHAT HISTORIE


In der Seitennavigation finden Sie alle gespeicherten Konversationen und können wahlweise zwischen diesen wechseln. Jede Konversation hat eine eigene Bezeichnung, die von Business GPT generiert wird. Vorhandene Konversationen können über eine entsprechende Schaltfläche gelöscht werden, wenn Sie diese nicht mehr benötigen.

☰

 Neue Konversation

 Temporäre Dokumente

 **Chatverlauf**

 Prompts

## Chatverlauf

---

Neue Konversation

Welche Dokumente gibt es?

### Aufbewahrungsfrist

Chatverläufe werden maximal 180 Tage aufbewahrt und anschließend automatisch gelöscht.

### Neuer Chat

Die Schaltfläche „Neuer Chat“ legt eine neue Konversation mit dem aktuell ausgewählten Modell an.



### Antwort kopieren

Kopieren Sie Business GPT Antworten mit einem einzigen Klick, um sie in anderen Anwendungen oder Dokumenten zu verwenden. Klicken Sie dazu auf das Icon direkt unter der Antwort.

### Antwort neu generieren

Sie können die Antwort von Business GPT erneut generieren lassen. Klicken Sie dazu auf das Icon unterhalb der Antwort.

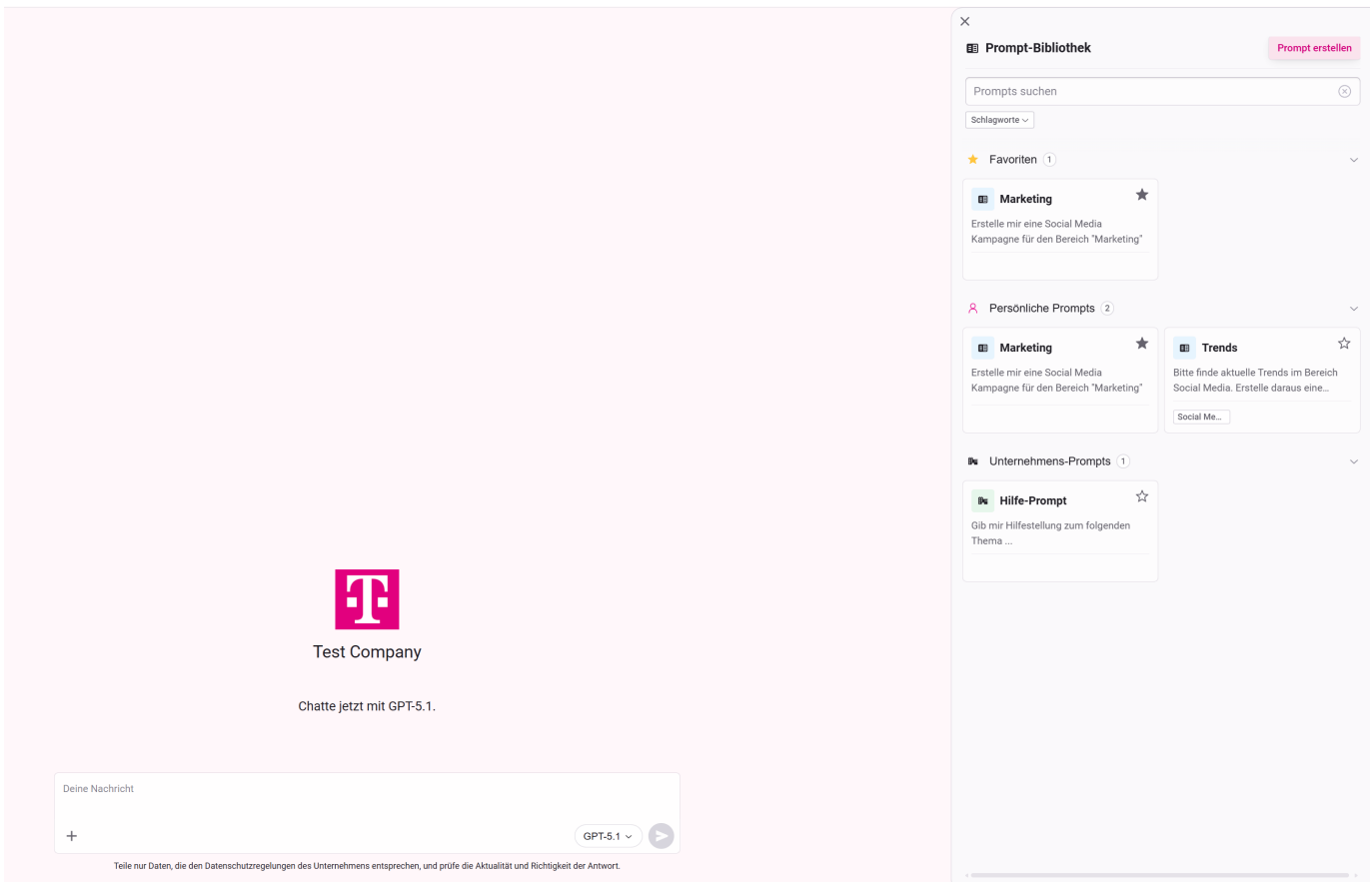
### Limits

#### Maximale Zeichenlänge


In Ihrer Anfrage (Texteingabe) an Business GPT können Sie maximal 100.000 Zeichen verwenden. Ist Ihre Nachricht zu lang erhalten Sie eine Fehlermeldung. Nutzen Sie in diesen Fällen die Möglichkeit (Text-) Dateien an den Chat anzuhängen.

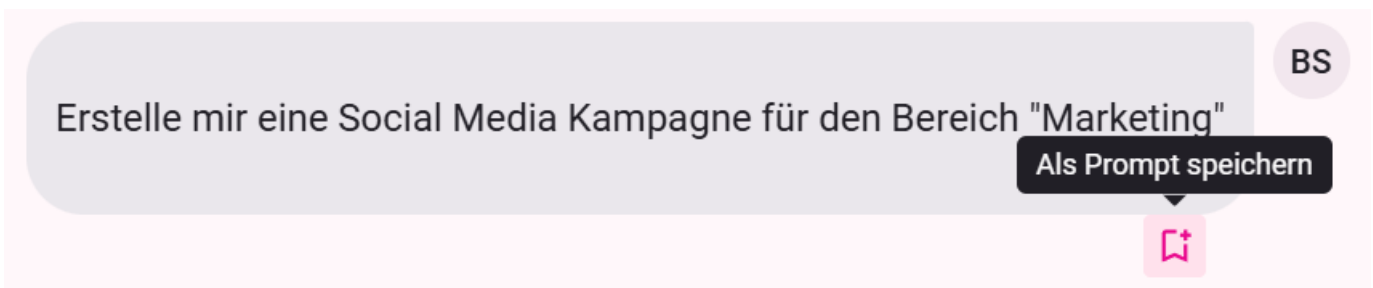
## 2.5.2 Prompt Bibliothek

Wenn Sie einen hilfreichen Prompt erstellt haben und ihn später erneut verwenden möchten, können Sie ihn in Ihrer Prompt-Bibliothek speichern und dort bearbeiten. Gespeicherte Prompts lassen sich in jedem Chat sowie in Assistenten einsetzen. Alle persönlichen Prompts sind privat und nur für Sie zugänglich.



### Prompt im Chat speichern

Um einen Prompt im Chat zu speichern, wählen Sie unterhalb des Prompts die  Schaltfläche und speichern Sie den Prompt. Vergeben Sie einen Namen für den Prompt und ergänzen Sie diesen wahlweise mit bis zu 3 Schlagworten (Tags). Mit Tags können Sie Ihren Prompt zum Beispiel kategorisieren, um ihn später schneller wiederzufinden.



### Prompt in den Chat einfügen

Um einen Prompt in den Chat einzufügen, klicken Sie einfach auf die Prompt Karte. Der Prompt wird dann automatisch in das Textfeld übertragen.

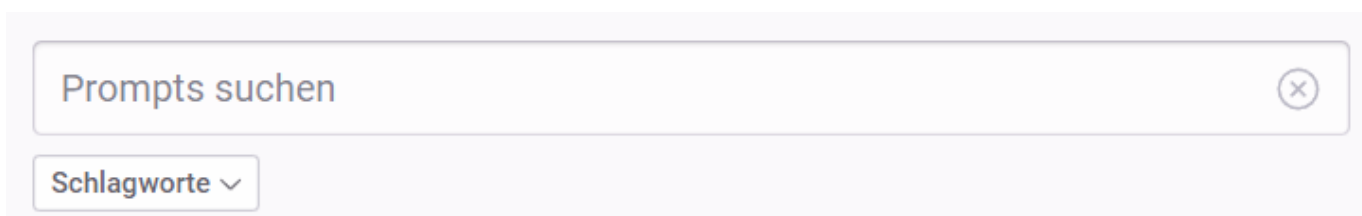
### Prompt Favoriten

Sie können Ihre Prompts zusätzlich als Favoriten markieren. Klicken Sie dazu auf das Sternsymbol, um einen Prompt als Favorite zu markieren oder wieder zu entfernen.



### Prompt suchen

In der Prompt Bibliothek können Sie Prompts per Schlagwort oder Suchtext suchen.

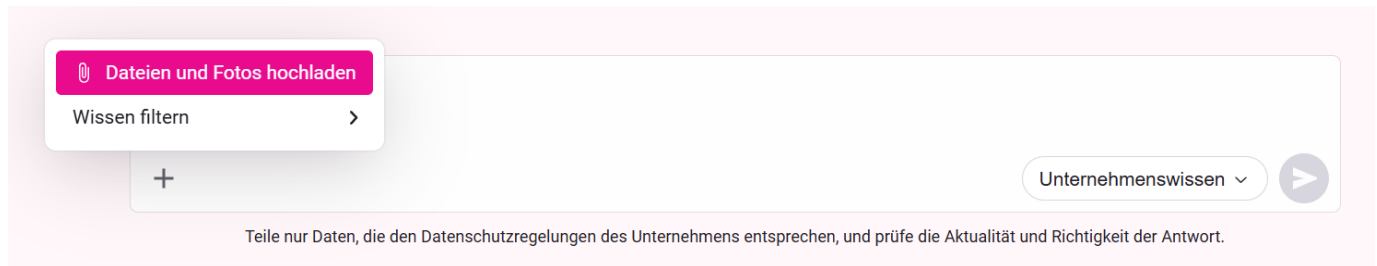


### Prompts verwalten

Die Verwaltung Ihrer persönlichen Prompts erfolgt über Ihre Einstellungen: [Prompt-Bibliothek](#)

## 2.5.3 Dateien und Fotos hochladen

Über die Schaltfläche "Datei und Fotos hochladen" können eigene Dokumente und Fotos hochgeladen und auf diese Weise dem Chat angehängt werden, um anschließend mit diesen zu chatten und gezielte Anfragen zu stellen.



Alternativ können Dateien auch per Drag & Drop oder per Einfügen aus der Zwischenablage (Copy & Paste) angehängt werden. Dateien wie PDF und Word können jedem gewählten Modell/Assistenten angehängt werden. Fotos und Bilddateien hingegen können nur dann an den Chat angehängt werden, wenn das zugrundeliegende Modell die Bildverarbeitung unterstützt (GPT4o, GPT4.1 und GPT5).

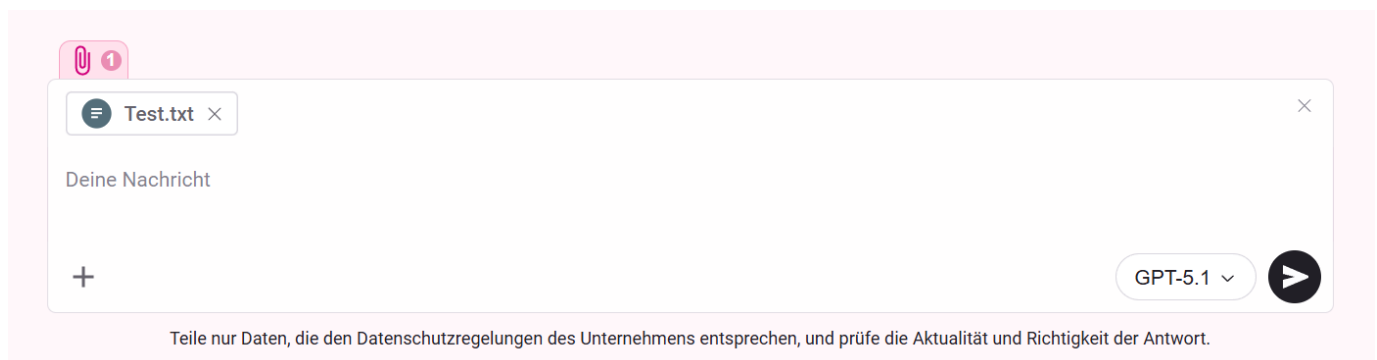
### Bildanalyse

Zahlreiche Modelle sind in der Lage Bilddateien zu analysieren (Vision). Hiermit können Bilddateien analysiert und interpretiert werden, bspw. um diese beschreiben zu lassen oder konkrete Fragen zur Bilddatei zu stellen.

Vision wird unterstützt von: GPT-4o, GPT-4.1, GPT-5. Unterstützte Dateitypen: JPG, JPEG, WEBP, GIF und PNG.

### Chat mit hochgeladenen Dokumenten

Dokumente, die hochgeladen und dem Chat angehängt werden, werden zunächst für den Chat vorbereitet. Dieser Vorgang dauert je nach Dokumentengröße unterschiedlich lang. Der Vorgang kann durch eine Fortschrittsanzeige am Dokumenten-Symbol nachvollzogen werden. Sobald das Dokument vorbereitet wurde, kann es im Chat verwendet werden.



Es können zudem mehrere Dokumente auf einmal hochgeladen und im Chat verwendet werden.

**Hinweise zu Dateien**

Die Gesamtanzahl ist begrenzt auf insgesamt **10** Dokumente. Für den Upload können Dokumente vom Typ `.pdf`, `.pptx`, `.docx`, `.xlsx` sowie `.json`, `.md`, `.txt`, `.yaml`, `.xsd`, `.xml` und `.csv` verwendet werden. Das Dateigrößenlimit liegt bei **100 MB** je Datei. Hochgeladene Dateien stehen im Bereich "[Temporäre Dokumente](#)" zur Verfügung und können dort wiederholt für Chatanfragen verwendet werden. Fügen Sie die gewünschten Dateien einfach per Klick zum Chat hinzu.

Wenn ein Dokument an den Chat angehängt wird, wird der Inhalt des Dokuments extrahiert und zusammen mit der Eingabe (Prompt) an das KI-Modell gesendet, sodass es Fragen basierend auf dem tatsächlichen Inhalt beantworten kann.

KI-Modelle verfügen über ein Kontext-Fenster (Token-Limit). Dieses stellt das Limit dar für die Menge an Text, die das Modell verarbeiten kann. Werden Dokumente oder eine Vielzahl von Dokumenten im Chat angehängt, die zu groß für das Kontext-Fenster des Modells sind, wird eine Fehlermeldung angezeigt. Entfernen Sie dann wahlweise Dokumente aus dem Chat oder wechseln Sie das Modell.



Kontext Fenster

Die unterschiedlichen Token-Limits pro Modell können Sie unserem [Modell-Guide](#) entnehmen. Bitte beachten Sie, dass nicht die Dateigröße, sondern die enthaltene Menge Text entscheidend ist.

#### TEMPORÄRE DOKUMENTE

Der Bereich "Temporäre Dokumente" steht nur in der Produktvariante "L" zur Verfügung und kann für eigene, hochgeladene Dokumente verwendet werden. Zuvor hochgeladene Dokumente können in diesem Bereich ausgewählt und dem Chat hinzugefügt werden, um mit diesen Dokumenten zu chatten. Um ein Dokument zum Chat hinzuzufügen, kann das Dokument einfach (per Klick) ausgewählt werden und erscheint anschließend als Symbol im Chat-Fenster.

Temporäre Dateien stehen für einen Zeitraum von 24 Stunden zur Verfügung und werden anschließend automatisch entfernt.

Weitere Dokumente können über "[Datei hochladen](#)" hinzugefügt werden.



Neue Konversation



Temporäre Dokumente



Chatverlauf



Prompts

## Temporäre Dokumente



Doc A.pdf



Doc B.pdf



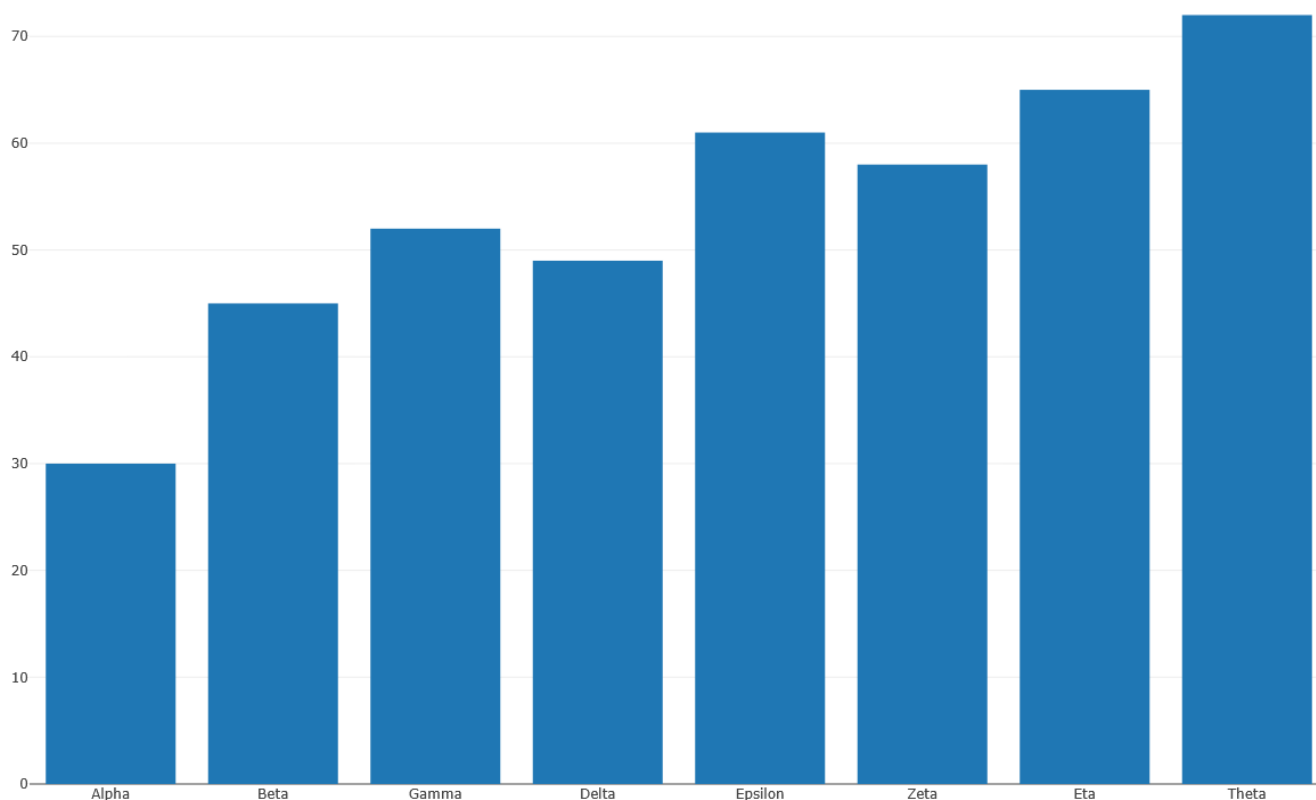
## 2.5.4 Diagramme

Mit den Standard-Modellen (z.B. `GPT-5.1`, `GPT-4.1`, etc.) können bei Bedarf Grafiken erstellt werden. Verwenden Sie dazu im Prompt Anweisungen, um eine Grafik zu erstellen.

### Beispiele

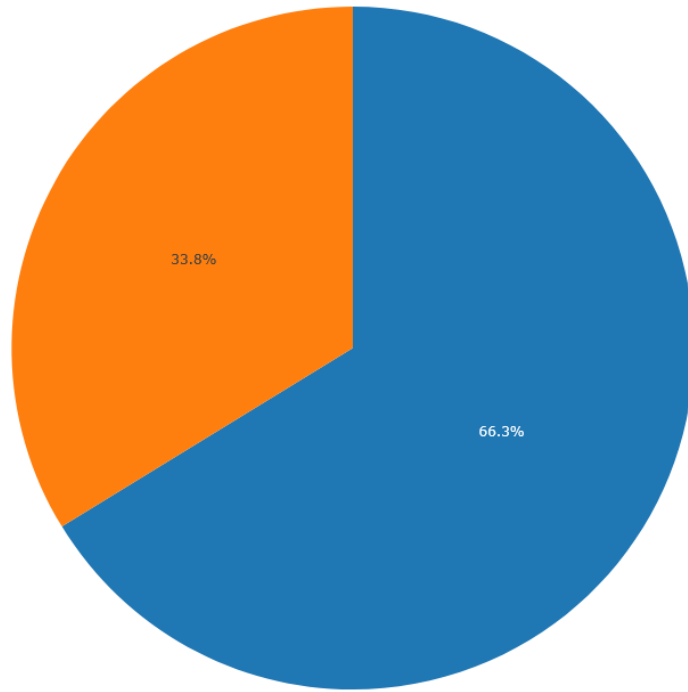
Erstelle mir eine Visualisierung zu Temperaturwerten über 12 Monate.

Erstelle ein Balkendiagramm, das die Verkäufe von Äpfeln gegenüber Orangen der letzten drei Monate zeigt: Januar (10 vs. 15), Februar (12 vs. 18), März (5 vs. 20).

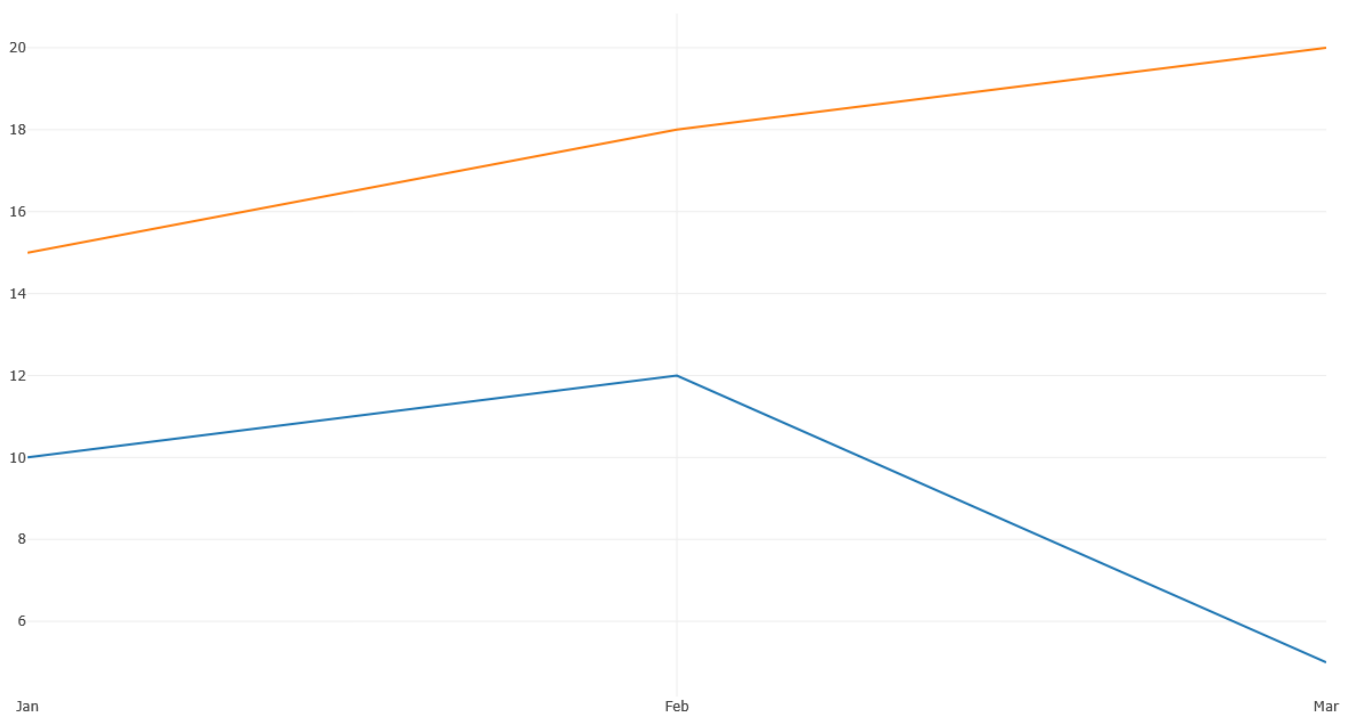


### Diagrammarten

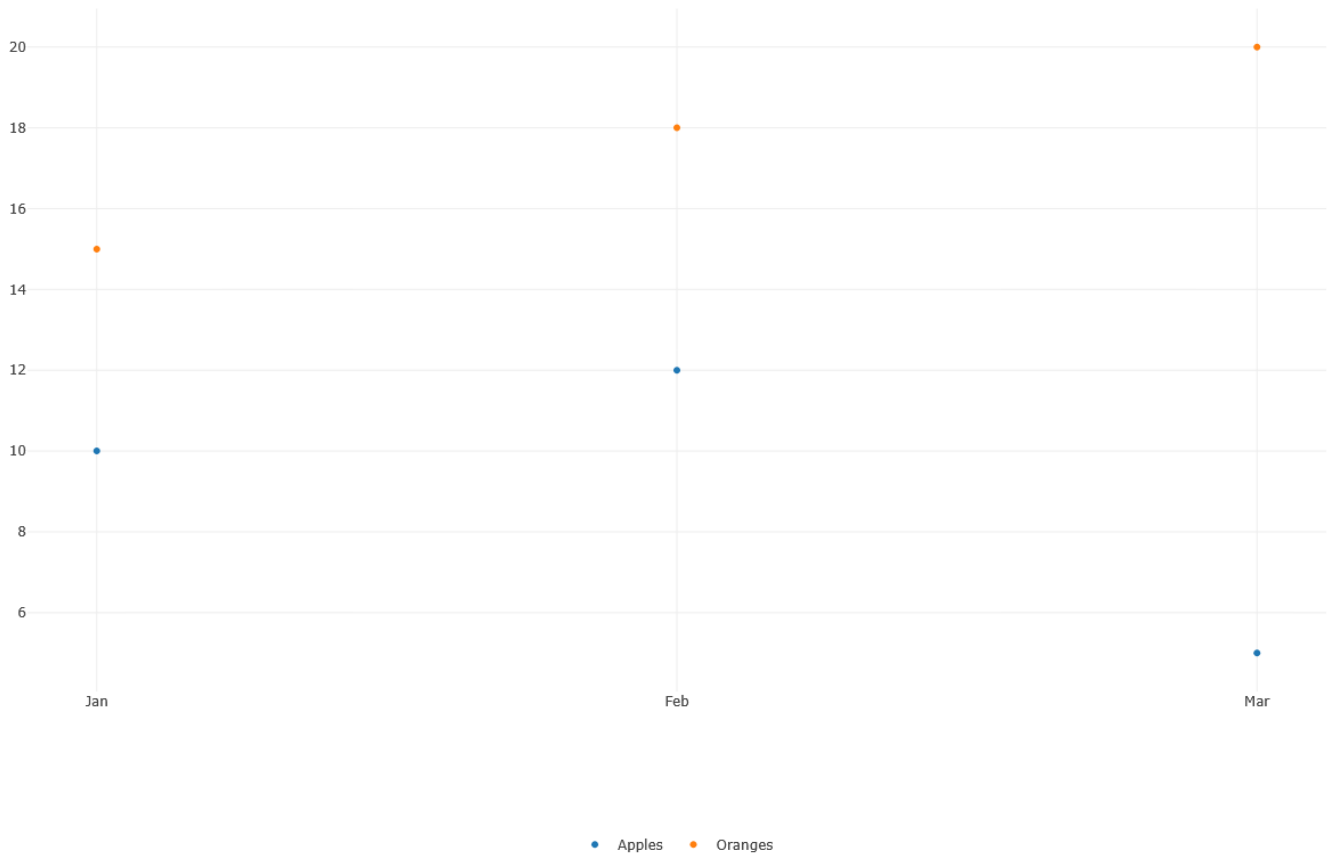
Business GPT kann verschiedene Diagrammarten ausgeben, darunter `Balkendiagramm`, `Liniendiagramm`, `Punktendiagramm` oder `Kreisdiagramm`.



■ Oranges ■ Apples



— Apples — Oranges



Um das gewünschte Diagramm zu erzeugen, weisen Sie Business GPT entsprechend an.

Erstelle ein Kreisdiagramm ...

### Diagramm Farben

Sie können die Farbgebung von Balken, Linie, etc. beeinflussen, in dem Sie der KI Anweisungen geben.

Erstelle ein Balkendiagramm, das die Verkäufe von Äpfeln gegenüber Orangen der letzten drei Monate zeigt: Januar (10 vs. 15), Februar (12 vs. 18), März (5 vs. 20). Nutze rot für Äpfel und Orange für Orangen.

### Diagramm speichern

Sie können Diagramme als PNG-Datei herunterladen. Wählen Sie dazu die Schaltfläche (Icon) zum Herunterladen direkt am Diagramm.

## 2.6 Einstellungen

---

### 2.6.1 Einstellungen und Administration

---

Persönliche und administrative Einstellungen können über den Menüpunkt **Einstellungen** vorgenommen werden.

Einstellungen

Einstellungen



#### **Wichtiger Hinweis**

Viele Einstellungen können nur durch **Administratoren** bearbeitet werden.

## 2.6.2 Einstellungen

Im Einstellungsbereich können globale Einstellungen vorgenommen werden, die sich auf Ihre Umgebung auswirken.

### Dokumentencontainer

Dokumentencontainer können standardmäßig nur durch Administratoren verwaltet werden. Um auch Nutzer:Innen die Erstellung und Verwaltung von eigenen, privaten Dokumentencontainern zu erlauben, können Sie die folgenden Einstellungen vornehmen.

#### Dokumenten Container

**Private Dokumenten-Container**  
Dokumentencontainer für reguläre Benutzer aktivieren ⓘ

**Benutzergruppen-Berechtigungen**  
Nur für bestimmte Gruppen erlauben ⓘ

**Maximale Anzahl an Dateien**  
Anzahl der Dateien beschränken ⓘ

Aktiv

Für alle Benutzer verfügbar

Anzahl an Dateien  
5

[Nutzergruppen verwalten](#)

### Aktivierung

Aktivieren Sie diese Option, erlauben Sie private Dokumentencontainer für Nutzer:Innen. Auf diese Weise sind Nutzer:Innen in der Lage eigene Container zu erstellen, Dokumente hochzuladen und diese zum Chatten zu verwenden. Die Container bleiben privat, d.h. weder andere Nutzer:Innen noch Administratoren können diese Container einsehen oder verwalten. Es gibt kein Limit für die Anzahl von Containern, die Nutzer:Innen erstellen können.

### Benutzergruppen-Beschränkungen

Optional können Sie private Dokumentencontainer nur für bestimmte Benutzergruppen freigeben. Erfolgt hier keine Einschränkung stehen die Funktionen zur Verwaltung privater Dokumentencontainer uneingeschränkt allen Nutzer:Innen zur Verfügung.

### Beschränkung der Dateianzahl

Legen Sie hier die erlaubte Gesamtanzahl an Dateien fest, die hochgeladen werden können. Die Maximalanzahl pro Nutzer:In beträgt 100 Dateien. Für erlaubte Dateitypen und die maximale Dateigröße gelten die gleichen Bedingungen, wie für Dokumentencontainer, die durch Administratoren verwaltet werden.


## 2.6.3 Design

Im Bereich „Design“ kann das Farbschema durch Angabe eines Farbcodes (hexadezimal) sowie die Darstellung des Hintergrundes geändert werden. Weiterhin besteht die Möglichkeit den Firmennamen sowie das Firmenlogo zu ändern. Beide Angaben werden auf der Startseite im Chat-Bereich angezeigt.


### Styling

**Design**

**Akzentfarbe**

 #E20074

**Hintergründe**




**Unternehmen**

**Firmennamen**

Deutsche Telekom MMS GmbH

**Firmenlogo**

Logo hochladen 

Abbrechen Speichern

## 2.6.4 Gruppen

Im Bereich „Gruppen“ werden alle Nutzergruppen angezeigt, die für Business GPT verwendet werden.

# Gruppen

Gruppen importieren (CSV)

4 Ergebnisse 🔍 ✖

Exportieren
Ausgeblendete Gruppen anzeigen
Auswahl löschen

<input type="checkbox"/>	Gruppenname	↑↓	Zuletzt geändert	↑↓	Identifikator	↑↓
<input type="checkbox"/>	HR		14.04.2025 14:23		.....	⋮
<input type="checkbox"/>	Ops		14.04.2025 14:23		.....	⋮
<input type="checkbox"/>	Telekom		14.04.2025 15:55		.....	⋮
<input type="checkbox"/>	Test		14.04.2025 15:55		.....	⋮

<
1
of 1
>

Gruppen werden zum einen benötigt, um den Zugriff auf Dokumente in Dokumenten-Containern zu regeln (siehe [Dokumenten Container](#)). Somit kann der Inhalt eines Dokumenten-Containers nur für bestimmte Nutzergruppen freigegeben werden, wonach bei Auswahl des Assistenten „Dokumenten Chat“ der jeweilige Container nur exklusiv für die berechnigte Nutzergruppe zur Verfügung steht. Die Zugehörigkeit zu Nutzergruppen von Nutzer:innen wird durch Auswertung der Anmeldeinformationen ausgewertet.

Zum anderen dienen Gruppen ebenfalls zur Zugriffssteuerung für Assistenten (siehe [Assistenten](#)). Demnach können Assistenten für bestimmte Gruppen für die Verwendung freigegeben werden.

Gruppen werden stets automatisch in Business GPT angelegt, basierend auf den Anmeldedaten von NutzerInnen. Voraussetzung ist, dass in Ihrem Entra ID die Übertragung von Gruppendaten mittels Token-Konfiguration korrekt konfiguriert wurde. Aus den Anmeldedaten werden demnach die zugeordneten Gruppen ausgelesen und automatisch in Business GPT angelegt. Abhängig von der Entra ID-Version (Free EntraID oder P1) werden Gruppen-IDs oder -namen übertragen. Zudem werden in Abhängigkeit der EntraID-Version nur Gruppen übertragen, die der zugehörigen Enterprise Application zugeordnet wurden oder alle verfügbaren Gruppen, denen ein Nutzerobjekt angehört. Bei einer P1-Version können Gruppen direkt der Enterprise Application zugewiesen werden. Bei einer Free Version hingegen ist dies nicht möglich. Bei Verwendung einer P1-Lizenz kann somit die Anzahl von übermittelten Gruppen eingeschränkt werden. Bei einer Free-Version hingegen gibt es keine Möglichkeit der Einflussnahme.

### Wichtiger Hinweis

Gruppen und Gruppenzuordnungen zu Containern sind zwingend notwendig, um Container- und somit Dokumenteninhalte Standardnutzern zur Verfügung zu stellen. Standardmäßig sind nur Administratoren berechnigt Containerinhalte im Dokumentenchat einzusehen, solange nicht explizit Gruppen zugeordnet (berechnigt) wurden.

## Gruppen importieren

Nutzergruppen können per CSV-Upload importiert werden. Hierzu können Sie optional vorab Gruppen aus Ihrem EntraID exportieren. Der Import von Gruppen kann sowohl genutzt werden, um initial alle relevanten Gruppen in Business GPT anzulegen, als auch nachträglich vorhandene Gruppen bequem per Massenaktion umzubenennen, für den Fall, dass keine Gruppennamen sondern nur Gruppen-IDs übermittelt und in Business GPT gespeichert wurden. Bitte beachten Sie, dass die Import-Datei über folgenden Inhalt verfügen muss:

- Kopfzeile bestehend aus „id“ und „label“
- Je Gruppe: Gruppen-Id und Label (Gruppenname)

	A	B
1	id	label
2	e37005f0-4698-44cc-889f-e3369f0319ec	HR
3	f37005f0-4698-44cc-889f-e3369f0319ed	Marketing
4		



### Wichtige Hinweise

- Die Zuordnung von Gruppenmitgliedern zu Gruppen erfolgt ausschließlich in Ihrem EntraID.
- Der CSV-Import wird stets gesamtheitlich angewendet. Das bedeutet:
- In der CSV-Datei enthaltenen Gruppen, die bis dato nicht in Business GPT bekannt sind, werden neu angelegt
- Gruppen, die basierend auf der ID im System bereits bekannt sind, aber in der CSV-Datei über einen abweichenden/anderen Namen verfügen, werden im System geändert auf den neuen Namen
- Gruppen, die in Business GPT vorhanden sind, aber nicht (mehr) in der CSV-Datei enthalten sind, werden gelöscht!

Bevor der Gruppenimport gestartet wird, werden alle in der Importdatei enthaltenen Gruppen in einer Übersicht aufgelistet. Überprüfen Sie die Anzeige und starten Sie anschließend den Import.

# Gruppen

**Gruppen importieren (CSV)**

## Gruppen importieren

ⓘ Die folgenden Gruppen werden importiert. Neue Gruppen werden angelegt, bestehende aktualisiert. Gruppen, die in der Auflistung nicht enthalten sind, werden aus dem System gelöscht. Bitte prüfen Sie die Daten sorgfältig.

[Auswahl löschen](#)

<input type="checkbox"/>	Gruppenname	Identifikator	
<input type="checkbox"/>	HR	.....	...
<input type="checkbox"/>	Test	.....	...
<input type="checkbox"/>	Ops	.....	...
<input type="checkbox"/>	Telekom	.....	...

[Abbrechen](#) [Import](#)

< 1 of 1 >

### Gruppen ein- und ausblenden

Für Business GPT nicht benötigte Gruppen können ausgeblendet werden. Dies kann sinnvoll sein, wenn Sie im EntraID keine Möglichkeit haben die an Business GPT übermittelten Gruppen festzulegen oder einzuschränken und somit auch Gruppen in Business GPT gespeichert sind, die für die Nutzung irrelevant sind. Ausgeblendete Gruppen werden nicht mehr bei Gruppenberechtigungen, z.B. der Berechtigungsvergabe für Dokumentencontainer, angeboten. Ausgeblendete Gruppen können in der Gruppenübersicht angezeigt und bei Bedarf wieder eingeblendet werden.

Gruppen

Gruppen importieren (CSV)

Nach Gruppe suchen 4 Ergebnisse

Exportieren **Ausgeblendete Gruppen anzeigen** Auswahl löschen

<input type="checkbox"/>	Gruppenname	Zuletzt geändert	Identifikator
<input type="checkbox"/>	HR	14.04.2025 14:23	.....
<input type="checkbox"/>	Ops	14.04.2025 14:23	.....
<input type="checkbox"/>	Telekom	14.04.2025 15:55	.....
<input type="checkbox"/>	Test	14.04.2025 15:55	.....

Umbenennen  
**Ausblenden**  
Löschen

1 of 1

### Gruppen exportieren

Die Gruppenübersicht kann exportiert werden. Hierbei wird eine CSV-Datei erzeugt und heruntergeladen.

### Gruppen umbenennen

Gruppen können umbenannt werden, um den Gruppennamen zu ändern.

### Gruppen löschen

Gruppen können einzeln oder durch Mehrfachauswahl gelöscht werden. Auf diese Weise können Gruppen aus Business GPT entfernt werden.

## 2.6.5 Dokumenten-Container

Dokumenten-Container erweitern das Wissen der KI. Alle Dokumente, die Sie in einen Container hochladen, werden von der KI bei passenden Fragen berücksichtigt. Dadurch kann die KI nicht nur auf ihr allgemeines Trainingswissen zurückgreifen, sondern auch auf Ihre unternehmens- oder projektspezifischen Inhalte. Wenn Sie mit der KI unter Verwendung von Dokumenten-Containern chatten, durchsucht sie automatisch die relevanten Dokumente im Container und nutzt die gefundenen Informationen für ihre Antwort. So erhalten Sie präzisere, kontextbezogene Ergebnisse — zum Beispiel auf Basis von Handbüchern, Richtlinien oder internen Unterlagen.



### Hinweise zu Produktoptionen

Nur die Produktvariante „L“ bietet die Möglichkeit mit Dokumenten und Dokumenten-Containern zu chatten und letztere zu verwalten.

### Dokumenten-Container Typen

Business GPT bietet 2 Typen von Dokumenten-Container: persönliche Container und Unternehmenscontainer.

#### persönliche Dokumenten-Container

Hierbei handelt es sich um nutzerindividuelle, private Dokumentencontainer. Nutzer:Innen können somit eigene, private Wissensbasen aufbauen.



### Einstellungen für Dokumenten-Container

Bitte beachten Sie, dass die Funktion **Persönliche Dokumenten-Container** standardmäßig deaktiviert ist. Sofern Sie private Dokumenten-Container für Nutzer:Innen erlauben möchten, nutzen Sie zunächst die Einstelloptionen unter [Einstellungen](#)

#### Unternehmenscontainer

Diese Dokumenten-Container können nur durch Administratoren erstellt und verwaltet werden. Sie dienen dazu, Wissen zentral zur Verfügung zu stellen. Mithilfe von Unternehmenscontainern können Dokumente z.B. thematisch oder nach Zugriffsberechtigungen abgelegt werden. Auf diese Weise können Dokumenten-Container vielfältig eingesetzt werden, z.B.

- Usecase-spezifische Dokumenten-Container
- Themen-spezifische Dokumenten-Container
- zugriffsspezifische Dokumenten-Container

### Container Übersicht

Die Übersicht der Dokumenten-Container zeigt alle verfügbaren Container und dient als Einstieg in die Containerverwaltung. Es kann zwischen persönlichen und Unternehmenscontainern gewechselt werden.

#### PERSÖNLICHE CONTAINER

In diesem Bereich können - sofern die Funktion für Sie freigegeben wurde - eigene, private Container erstellt und verwaltet werden. Ist diese Funktion für Sie nicht freigegeben, wird Ihnen ein Hinweis angezeigt.

#### UNTERNEHMENSCONTAINER

Als Administrator gelangen Sie hier in die zentrale Verwaltung aller Unternehmenscontainer.

Als Nutzer:In ohne administrative Berechtigungen erhalten Sie hier einen Überblick über alle Unternehmenscontainer, für deren Einsicht und Verwendung im Dokumentenchat Sie berechtigt wurden. Diese Übersicht hat rein informativen Charakter. Weiterführenden Funktionen stehen hier nicht zur Verfügung.

#### Dokumenten Container erstellen

Ein neuer Container kann über die „Plus“-Schaltfläche erstellt werden. Neben dem Container-Namen können optional eine Beschreibung und eine Farbe angegeben werden.

#### BERECHTIGUNGEN FÜR UNTERNEHMENSCONTAINER

Für einen Unternehmenscontainer kann der Container (-inhalt) für bestimmte Nutzergruppen freigegeben werden. Wählen Sie hierzu bitte die Schaltfläche "Nutzergruppen verwalten". **Erfolgt hierbei keine Freigabe steht der Container-Inhalt nur Administratoren uneingeschränkt zur Verfügung.** Bitte beachten Sie, dass für die Zugriffsfreigabe für bestimmte Nutzergruppen zuvor der Import von Gruppen erfolgen muss (siehe [Gruppenimport](#)).

←
Speichern
Abbrechen

## Container Container Name

**Name**

**Farbe**

**Beschreibung**

Keine Beschreibung oder weitere Informationen verfügbar. Bearbeiten Sie den Container, um eine Beschreibung hinzuzufügen.






**Benutzergruppen-Berechtigungen**

ⓘ Keine Berechtigungen erteilt.

✕

### Container Details

Nach Auswahl eines Dokumenten Containers werden die Details angezeigt. Hier stehen folgende Funktionen zur Verfügung:

Funktion	Beschreibung
 Details bearbeiten	Name, Beschreibung und Farbe eines Containers ändern.
 Nutzergruppen verwalten (nur für Unternehmenscontainer)	Zuordnung und Verwaltung der zugriffsberechtigten Gruppen.
 Container löschen	Nur möglich, wenn sich keine Dokumente im Container befinden.
 Dokumente hochladen	Neue Dokumente zum Container hinzufügen.
 Dokumente verwalten	Anzeigen, Umbenennen, Löschen und Änderung der Container-Zuordnung bestehender Dokumente.

←
Speichern
Abbrechen
Löschen
ID

## Container HR-Dokumente

**Name**

**Farbe**

**Beschreibung**

**Benutzergruppen-Berechtigungen**

HR ×

×

Dokumente hochladen

Dateien hierher ziehen, um sie hochzuladen, oder

Dateien durchsuchen

Keine Dokumente zu indexieren
0 Dokumente fehlgeschlagen
2 Dokumente erfolgreich ⓘ

2 Ergebnisse ×

Auswahl löschen

<input type="checkbox"/>	Status <span style="font-size: 10px;">↕</span>	Dokumentname	Typ <span style="font-size: 10px;">↕</span>	Container	Upload-Datum <span style="font-size: 10px;">↕</span>	
<input type="checkbox"/>		Elterngeld	.pdf	Container: 2 ausgewählt	07.02.2025 09:33	⋮
<input type="checkbox"/>		Merkblatt Kindergeld	.pdf	Container: 1 ausgewählt	07.02.2025 09:33	⋮

1 of 1

### Dokumente hochladen

Über die Schaltfläche „Neues Dokument hochladen“ können ein oder mehrere Dokumente in den Container hochgeladen werden.

**Hinweise zu Dateien**

Aktuell können folgende Dokumententypen hochgeladen werden:

.pdf, .docx, .pptx, .xlsx, .json, .md, .txt, .yaml, .xsd, .xml, .csv

Die maximale Dateigröße liegt bei **100 MB** pro Datei.

←
Speichern
Abbrechen
Löschen
ID

## Container HR-Dokumente

**Name** **Farbe**

HR-Dokumente  #D1F0FA

**Beschreibung**

Dokumentation und Guides für Kinder- und Elterngeld, Beantragung, Besonderheiten und Hilfestellung

**Dokumenten Upload**

Auswahl löschen

<input type="checkbox"/> Dokumentname	<input type="checkbox"/> Container
<input type="checkbox"/> Dokument	Container: 1 ausgewählt <span style="float: right;"></span>

Abbrechen Hochladen

**Keine Dokumente zu indizieren** 0 Dokumente fehlgeschlagen 2 Dokumente erfolgreich

2 Ergebnisse

Auswahl löschen

<input type="checkbox"/>	Status	Dokumentname		Typ		Container	Upload-Datum	
<input type="checkbox"/>		Elterngeld		.pdf		Container: 2 ausgewählt	07.02.2025 09:33	
<input type="checkbox"/>		Merkblatt Kindergeld		.pdf		Container: 1 ausgewählt	07.02.2025 09:33	

< 1 of 1 >

Eine Fortschrittsanzeige zeigt Fortgang und Erfolg des Uploads. Hochgeladene Dokumente werden nach dem Upload automatisch indiziert. Erst nach erfolgter Indizierung stehen die Dokumente für den „Dokumenten Chat“ zur Verfügung. Die Spalte „Status“ in der Dokumentenliste zeigt an, ob ein Dokument bereits indiziert ist. Zudem ist der Indizierungsvorgang in der Indexierungsübersicht ersichtlich. Hier wird angezeigt, wie viele Dokumente noch auf Indizierung warten, wie viele Dokumente bereits erfolgreich indiziert wurden und ob Fehler aufgetreten sind.

**Dokumente hochladen**

Dateien hierher ziehen, um sie hochzuladen, oder

Dateien durchsuchen

**1 Dokument wartet auf Indizierung** 0 Dokumente fehlgeschlagen 2 Dokumente erfolgreich

## NEUINDEXIERUNG BEI FEHLERN

Sollten bei der Indexierung von hochgeladenen Dokumenten Fehler auftreten, kann eine Neuindexierung für fehlgeschlagene Dokumente gestartet werden.

## Indexierungsdetails



Indexierung starten

<input checked="" type="checkbox"/>	Element	Fehler	Wiederholungsstatus
<input checked="" type="checkbox"/>	Dokument ABC	After 3 retries the index process still failed.	Ausstehende Neuindizierung

## LIMITS FÜR PERSÖNLICHEN DOKUMENTENCONTAINER

Für persönliche Dokumentencontainer gilt eine maximale Anzahl von Dateien, die hochgeladen werden können. Ist das Limit erreicht, können keine weiteren Dokumente hochgeladen werden. Die Gesamtanzahl von Dokumenten kann beliebig über eine beliebige Anzahl von Dokumentencontainern verteilt werden.

←
Speichern
Abbrechen
Löschen
ID 
Containertyp: **PERSÖNLICH**

# Container Container 1

**Name\*** **Farbe**

Container 1

**Beschreibung**

Keine Beschreibung oder weitere Informationen verfügbar. Bearbeiten Sie den Container, um eine Beschreibung hinzuzufügen.

Sie haben die maximale Anzahl von 5 Dokumenten erreicht, die für Ihr Konto erlaubt ist.

Keine Dokumente zu indexieren
0 Dokument fehlgeschlagen
2 Dokument erfolgreich

Suche nach Dokumenten 2 Ergebnisse

Dokumenten Limit  (5/5)

### Umgang mit Duplikaten


Business GPT stellt sicher, dass keine Duplikate gespeichert werden. Wird ein Dokument in einen Container hochgeladen, welches bereits in einem anderen Container existiert, wird keine weitere Version des Dokuments gespeichert. Stattdessen wird das bereits vorhandene Dokument per Referenzierung dem Container zugeordnet. Auf diese Weise verhindert Business GPT Duplikate in Ihrem Datenbestand.

## 2.6.6 Dokumente


Eine Übersicht aller vorhandenen (Unternehmens-) Dokumente erhält man über den Aufruf `Dokumente` in den Einstellungen im Bereich „Administration“.

### Administration


Gruppen

56 


Dokumente

48 

Dokumenten Container

26 

Assistenten

12 

## Dokumente

Neues Dokument hochladen

Keine Dokumente zu indexieren

0 Dokumente fehlgeschlagen



























11 Dokumente erfolgreich ⓘ

Suche nach Dokumenten

11 Ergebnisse



Auswahl löschen

<input type="checkbox"/>	Status 	Dokumentname		Typ		Container	Upload-Datum 	
<input type="checkbox"/>		aDNS Netzwerk		.pdf		Container: 1 ausgewählt	07.02.2025 09:33	
<input type="checkbox"/>		Elterngeld		.pdf		Container: 2 ausgewählt	07.02.2025 09:33	
<input type="checkbox"/>		Hosting Plattform		.pptx		Container: 1 ausgewählt	07.02.2025 09:33	
<input type="checkbox"/>		Kriterienkatalog Toolevaluierung		.xlsx		Container: 1 ausgewählt	07.02.2025 09:33	
<input type="checkbox"/>		Kurzanleitung-Ticketverwaltung-über-TFS		.docx		Container: 1 ausgewählt	07.02.2025 09:33	
<input type="checkbox"/>		Leistungsbeschreibung_LLM_Serving		.pdf		Container: 1 ausgewählt	07.02.2025 09:33	
<input type="checkbox"/>		Merkblatt Kindergeld		.pdf		Container: 1 ausgewählt	07.02.2025 09:33	
<input type="checkbox"/>		Migrationsleitfaden_final		.pdf		Container: 1 ausgewählt	07.02.2025 09:33	
<input type="checkbox"/>		SES-NICService-Schnittstellendokumentati...		.docx		Container: 1 ausgewählt	27.03.2025 10:25	
<input type="checkbox"/>		SLUP API Spezifikation		.pdf		Container: 1 ausgewählt	07.02.2025 09:33	
<input type="checkbox"/>		T Phone Betriebsanleitung		.pdf		Container: 1 ausgewählt	07.02.2025 09:33	

1 of 1

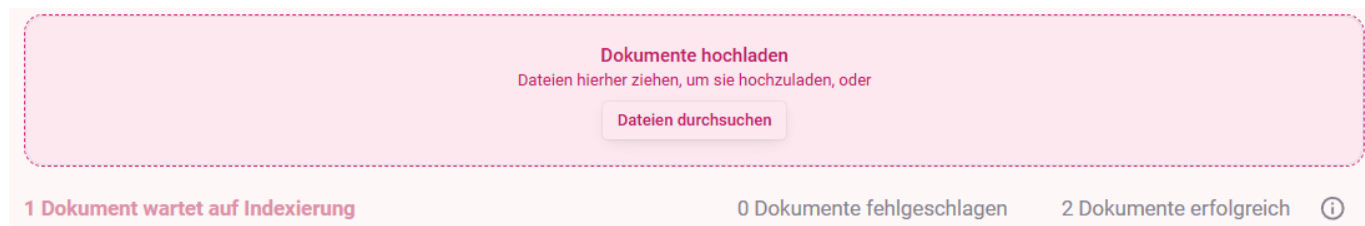
## Dokumente hochladen

Über die Schaltfläche **Neues Dokument hochladen** können ein oder mehrere Dokumente hochgeladen werden. Im Upload-Vorgang muss der gewünschte Zielcontainer angegeben werden. Aktuell können folgende Dokumententypen hochgeladen werden:

`.pdf`, `.docx`, `.pptx`, `.xlsx`, `.json`, `.md`, `.txt`, `.yaml`, `.xsd`, `.xml`, `.csv`






Das Dateigrößenlimit liegt bei **100MB** je Datei.

Eine Fortschrittsanzeige zeigt Fortgang und Erfolg des Uploads. Hochgeladene Dokumente werden nach dem Upload automatisch indiziert. Erst nach erfolgter Indizierung stehen die Dokumente für den „Dokumenten Chat“ zur Verfügung. Die Spalte „Status“ in der Dokumentenliste zeigt an, ob ein Dokument bereits indiziert ist. Zudem ist der Indizierungsvorgang in der Indizierungsübersicht ersichtlich. Hier wird angezeigt, wie viele Dokumente noch auf Indizierung warten, wie viele Dokumente bereits erfolgreich indiziert wurden und ob Fehler aufgetreten sind.



## Dokumente verwalten

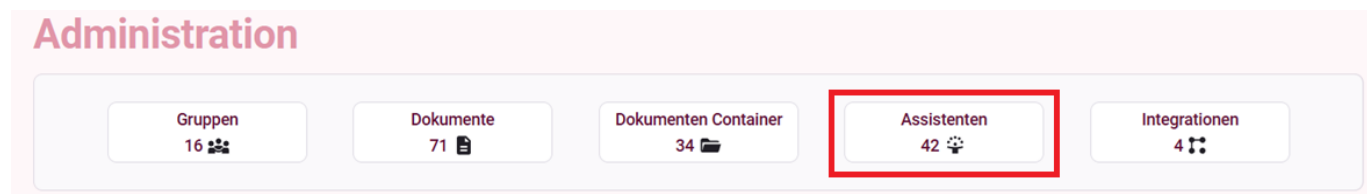
Zur Verwaltung von Dokumenten stehen folgende Funktionen zur Verfügung:

Funktion	Beschreibung
 <b>Ansicht</b>	Dokumente können im Webbrowser geöffnet werden.
 <b>Herunterladen</b>	Dokumente können heruntergeladen werden.
 <b>Umbenennen</b>	Der Dateiname kann geändert werden.
 <b>Löschen</b>	Dokumente können vollständig gelöscht werden. Mehrere Dokumente können nach Auswahl in der Liste gelöscht werden.
 <b>Container-Zuordnung ändern</b>	Einem oder mehreren Containern kann ein Dokument zugeordnet werden.

## 2.6.7 Assistenten

---

Mit eigenen Assistenten können Sie spezialisierte Tools entwickeln, die perfekt auf Ihre individuellen Bedürfnisse oder die Anforderungen Ihres Unternehmens zugeschnitten sind. Sie können einen Assistenten so gestalten, dass er eine bestimmte Rolle einnimmt, spezifische Aufgaben erfüllt oder in einem klar definierten Bereich Expertise zeigt.



---

### Assistenten Übersicht

Wählen Sie im Bereich "Administration" den Eintrag "Assistenten", um auf die Übersicht aller vorhandenen Assistenten zu gelangen. In der Übersicht werden alle vorhandenen Assistenten angezeigt.

←

# Assistenten

Assistenten durchsuchen 14 Ergebnisse

Assistenten Name Aktiv

- Business GPT Assistent** (1 gpt-4o)
 

I help you with all questions about BGPT
- HR Dokumenten Assistent** (0 gpt-4o)
 

Keine Willkommensnachricht oder weitere Informationen verfügbar. Bearbeiten Sie den Assistenten, um eine Willkommensnachricht hinzuzufügen.
- Kulinarischer Experte** (0 gpt-4.1)
 

Keine Willkommensnachricht oder weitere Informationen verfügbar. Bearbeiten Sie den Assistenten, um eine Willkommensnachricht hinzuzufügen.
- LLM Hub Assistent** (0 o1)
 

Assistent für alle Fragen rund um LLM Serving der T-Systems
- Mein GPT4o Assistent** (0 gpt-4o)
 

I am your Business GPT 4o assistant. This is additionally a very very long description with line breaks that shows how BGPT handles those things in the cards.
- Telekom Assistent** (0 gpt-4o)
 

Ich bin Dein Telekom-Hilfsassistent und helfe bei allen Fragen rund um Telekom-Produkte und Services.
- Support Assistent** (1 gpt-4o-mini)
 

Hi, ich bin dein Supportassistent
- Next Test** (0 gpt-4.1)
 

Keine Willkommensnachricht oder weitere Informationen verfügbar. Bearbeiten Sie den Assistenten, um eine Willkommensnachricht hinzuzufügen.
- Mein Internet Assistent** (0 gpt-4o)
 

Keine Willkommensnachricht oder weitere Informationen verfügbar. Bearbeiten Sie den Assistenten, um eine Willkommensnachricht hinzuzufügen.

1 of 2

## Assistent erstellen

Ein neuer Assistent kann per Klick auf die Plus-Schaltfläche erstellt werden. Ein Assistent kann als "Standard Assistent" oder als "Dokumenten Assistent" (nur verfügbar für Business GPT Paket "L") erstellt werden. Je nach Art des Assistenten können unterschiedliche Einstellungen vorgenommen werden.

Um die Assistenteneinstellungen zu testen, können Sie im angezeigten Chat-Bereich Anfragen stellen, um zu prüfen, wie sich der Assistent verhält. Sind Sie mit den Einstellungen zufrieden, können Sie den Assistenten für alle NutzerInnen zur Verfügung stellen, indem Sie den Status des Assistenten auf "Aktiv" setzen.



Um den Assistenten mit den geänderten Einstellungen zu testen, müssen die Änderungen zunächst gespeichert werden.

← Speichern Löschen

# Assistenten Assistenten Name

Teste deinen Assistenten

### Profil Inaktiv

**Name\*** DE EN

Assistenten Name 0 / 100

**Anweisungen / System Prompt\***

You are a helpful assistant for the Company XXXX and you will start every Conversation with "Hello I am using the Model 4o and can help you with all kinds of questions about all topics that you have access to." 0 / 16000

**Willkommensnachricht** DE EN

Hallo, ich bin Dein HR-Assistent. Ich helfe Dir bei Fragen zu Urlaub, Anträgen und vieles mehr. 0 / 4000

**Zusatzhinweis** DE EN

Teile nur Daten der Schutzklasse 'OFFEN' und 'INTERN' mit dem Assistenten und prüfe die Aktualität und Richtigkeit der Antwort. 0 / 500

**Model\***

Bitte wählen Sie ein Modell aus ▼

### Tools

**Fähigkeiten** Bearbeiten

Verwenden Sie „Bearbeiten“, um die benötigten Funktionen auszuwählen.

**Integrationen** Bearbeiten

Verwenden Sie „Bearbeiten“, um die benötigten Integrationen auszuwählen.

**Dokumenten-Container** Bearbeiten


Verwenden Sie „Bearbeiten“, um die benötigten Dokumentcontainer auszuwählen.

### Zugriff

**Benutzergruppen-Berechtigungen**

ⓘ Keine Berechtigungen erteilt.

⊗



Deutsche Telekom MMS GmbH

++

Der Assistent kann mit zahlreichen Optionen und Fähigkeiten ausgestattet werden.

**ASSISTENTEN STATUS**

Setzen Sie den Assistentenstatus auf "Aktiv" sobald Sie mit dem Erstellen und Testen des Assistenten fertig sind.

Assistenten im Status "Inaktiv" stehen für NutzerInnen im Chat-Bereich in der Assistentenauswahl nicht zur Verfügung und sind dort nicht ersichtlich.

**NAME**

Vergeben Sie einen Namen für den Assistenten. Dieser wird später im Chat-Bereich für alle (berechtigten) NutzerInnen in der Model-/Assistentenauswahl angezeigt.

Der Name kann wahlweise Deutsch sowie Englisch angegeben werden.

**ANWEISUNGEN / SYSTEM PROMPT**

Der "System Prompt" definiert die grundlegende Rolle und das Verhalten des KI-Assistenten. Legen Sie hier fest, wie sich die KI gegenüber den Nutzern verhalten soll, z. B. ob sie freundlich, förmlich oder sachlich ist. Es können auch spezifische Anweisungen hinzugefügt werden, wie die KI auf bestimmte Fragen reagieren oder welche Aufgaben sie priorisieren soll.

** Beispiel:**

„You are a polite and precise AI assistant that provides short and clear answers.“

- Nutzen Sie den System Prompt, um sicherzustellen, dass der Assistent konsistent agiert und die gewünschte Nutzererfahrung bietet.
- Syntax-Tipp: Schreibe Sie den Prompt in natürlicher Sprache, als würden Sie mit einer Person kommunizieren.
- Semantik-Tipp: Seien Sie präzise und vermeiden Sie Mehrdeutigkeit. Statt „be useful“ besser „Provide factual and polite answers, and limit yourself to facts.“

**System Prompt**

Tipps & Tricks zum System Prompt finden Sie in unserem [System Prompt Guide](#).

**WILLKOMMENSNACHRICHT**

Die Willkommensnachricht wird bei Auswahl des Assistenten im Chat-Bereich angezeigt und sollte verwendet werden, um den Assistenten kurz und präzise vorzustellen.

** Beispiel:**

Assistent für die Formulierung von E-Mails.

- Die Angabe der Willkommensnachricht ist optional.

**ZUSATZHINWEIS**

Der Zusatzhinweis wird stets über dem Chat-Fenster im Chatbereich angezeigt.

Hier können zum Beispiel wichtige Hinweise für die Verwendung des Assistenten angezeigt werden.

** Beispiel:**

Überprüfen Sie stets die Antworten des Assistenten bevor Sie diese verwenden.

**MODEL**

Wählen Sie hier das Model für den Assistenten, z.B. `GPT-4.1`

## TOOLS

Sie können den Assistenten mit verschiedenen Tools ausstatten, die es dem Assistenten erlauben bestimmte Aktionen auszuführen.

## Fähigkeiten

Fügen Sie hier die gewünschten Fähigkeiten wie Websuche und Bildgenerierung hinzu.

## Fähigkeiten

Verwalten Sie die Fähigkeiten für diesen Assistenten.



**Web Suche**

Internetzugang



**Bildgenerierung**

Dalle

Abbrechen

Speichern

## Web Suche

Sofem Sie über das Addon "Internet Breakout" verfügen, können Sie dem Assistenten die Suche im Internet erlauben (via Bing Search) und somit Anfragen unter Zuhilfenahme von Internetwissen zulassen.



### Instruktionen für Internetzugriff

Wenn der Internetzugriff aktiviert wurde, sorgt Business GPT im Hintergrund dafür, dass die nötigen Instruktionen für den Abruf von Internetwissen dem Assistenten hinzugefügt werden.

## Bildgenerierung

Aktivieren Sie die Bild-Generierung, sodass der Assistent Bilder mit Dall-E erzeugen kann.

## Dokumenten Container

Wählen Sie hier die Container aus, die der Assistent bei der Beantwortung von Anfragen berücksichtigen soll. Auf diese Weise können Sie internes Wissen für den Assistenten verfügbar machen.

## Integrationen

Wählen Sie hier die Integrationen aus, die der Assistent verwenden soll. Auf diese Weise können Sie zum Beispiel Jira Integrationen und andere Integrationen zuweisen, auf die der Assistent zugreifen soll.



### Instruktionen für Verwendung von internem Wissen

Wenn Container für das Einbeziehen von internem Wissen angehängt wurden, sorgt Business GPT im Hintergrund dafür, dass die nötigen Instruktionen für den Abruf von internem Wissen aus Containern dem Assistenten hinzugefügt werden.

## Funktionen

Funktionen bieten Ihnen die Möglichkeit individuelle Aufrufe externen Systeme über API-Verbindungen zu nutzen. Sie können zum einen bereits vorhandene Funktionen dem Assistenten hinzufügen oder neue Funktionen erstellen.

## Funktionen

Weisen Sie Ihrem Assistenten Funktionen zu

### Neue Funktion erstellen

Um eine neue Funktion zu erstellen, wählen Sie die entsprechende Funktion in der Auswahl. Anschließend können Sie eine Funktion definieren.

Parameter/Feld	Beschreibung
<b>Name</b>	Legen Sie einen beschreibenden Namen für Ihre Funktion fest, z.B. getweather
<b>Beschreibung</b>	Beschreiben Sie, was die Funktion tut. Eine detaillierte Beschreibung hilft dem Model/Assistenten zu verstehen, was die Funktion erledigt.
<b>Erzwingen</b>	Aktivieren Sie diese Option, sodass der Funktionsaufruf immer erfolgt, unabhängig von der Nutzereingabe
<b>Json</b>	Erstellen Sie hier die Funktionsspezifikation. Dies umfasst alle Parameter, Felder und Validierungen.
<b>Verbindung</b>	Geben Sie die erforderlichen Verbindungsdaten für die API-Verbindung an.

#### Pfaden für Function Calling

Für weiterführende Hilfestellung zur Erstellung und Nutzung von Function Calling folgen Sie unserem [Function Calling Guide](#)

## Funktionen hinzufügen

Erstellen Sie eine neue Funktion

**Name \***

**Erzwingen**

Die Funktion wird immer ausgeführt, unabhängig von der Benutzereingabe

**Beschreibung \***

Beschreibe deine Funktion

0 / 500

**JSON \***

JSON-Parameter

---

**HTTP Methode \***

GET
▼

**Authentifizierungsmethode \***

Anonymous
▼

**API Endpunkt \***

Abbrechen

Speichern

### ZUGRIFF

Weisen Sie hier die Nutzergruppen zu, die für Chatanfragen Zugriff auf den Assistenten erhalten sollen. Ohne zugewiesene Nutzergruppen steht der Assistent nur NutzerInnen mit der Rolle "Administrator" zur Verfügung.

### PARAMETER: TEMPERATUR, TOP P UND TOP K

Die beiden Parameter "Temperatur" und "Top P" sind sogenannte Kreativitätsparameter. Sie beeinflussen maßgeblich die Antwortgenerierung des Assistenten.



#### Bitte beachten

Bitte beachten Sie, dass Temperatur und Top P nicht für jedes Modell eingestellt werden können. Beispielsweise bieten die Modell der GPT5 Serie diese Option nicht.

## Temperatur

Der Parameter "Temperatur" steuert die Kreativität und Variabilität in der Antwortgenerierung

- Niedrige Temperatur: Eine niedrige Temperature führt zu präziseren, direkteren Antworten, da die KI sich stärker auf die abgerufenen Informationen stützt und deterministisch bleibt.
- Hohe Temperatur: Eine hohe Temperature kann dazu führen, dass die KI kreative oder alternative Formulierungen nutzt, was bei offenen Fragen vorteilhaft ist, jedoch ungenau werden kann.

## Top-p

Der Parameter "Top-p" beschränkt den Bereich der Wahrscheinlichkeiten, aus denen die KI Wörter auswählt. Statt starr den wahrscheinlichsten Kandidaten zu wählen, wird die Auswahl auf einen dynamischen Bereich beschränkt, basierend auf der Gesamtsumme der Wahrscheinlichkeiten.

- Top-p = 1: Es wird kein Filter angewendet, und die KI berücksichtigt alle möglichen Wörter. Dies entspricht einer "freien Wahl".
- Top-p < 1.0 (z. B. 0,9): Die KI beschränkt die Auswahl auf Wörter, deren kumulierte Wahrscheinlichkeiten 90 % ausmachen. Seltene Wörter werden ausgeschlossen. Das führt oft zu kreativeren, aber dennoch sinnvollen Ergebnissen.
- Beispiel: Mit Top-p = 0,9 könnte die KI eine ungewöhnlichere Formulierung wählen, solange sie zu den 90 % wahrscheinlichsten Optionen gehört.
- Top-p = 0,1: Die Auswahl ist stark eingeschränkt, und die KI verwendet nur die wahrscheinlichsten Wörter. Dies macht die Antworten weniger vielfältig, aber konsistenter.

## Zusammenspiel von Top-p und Temperature

- Hohe Temperatur und hoher Top-p-Wert: Kreative, abwechslungsreiche und manchmal überraschende Antworten.
- Niedrige Temperature und niedriger Top-p-Wert: Präzise, vorhersehbare und oft konservative Antworten.
- Beispielanwendung:
- Kreative Textgenerierung (z. B. Geschichten, Gedichte): Temperature = 0,8–1,0; Top-p = 0,9.
- Faktenbasierte Antworten (z. B. Wissensfragen): Temperature = 0,2–0,5; Top-p = 1,0.

## Beispiel-Werte

Anwendungsfall	Temperatur	Top_p	Beschreibung
Code-Generierung	0.2	0.1	Generiert Code, der etablierten Mustern und Konventionen folgt. Die Ausgabe ist deterministischer und fokussierter. Nützlich für die Erzeugung syntaktisch korrekten Codes.
Kreatives Schreiben	0.7	0.8	Generiert kreative und vielfältige Texte für Geschichten. Die Ausgabe ist explorativer und weniger durch Muster eingeschränkt.
Chatbot-Antworten	0.5	0.5	Generiert gesprächige Antworten, die Kohärenz und Vielfalt ausbalancieren. Die Ausgabe ist natürlicher und ansprechender.
Code-Kommentar-Generierung	0.3	0.2	Generiert Code-Kommentare, die prägnanter und relevanter sind. Die Ausgabe ist deterministischer und hält sich an Konventionen.
Datenanalyse-Skripterstellung	0.2	0.1	Generiert Datenanalyse-Skripte, die korrekt und effizient sind. Die Ausgabe ist deterministischer und fokussierter.
Exploratives Code-Schreiben	0.6	0.7	Generiert Code, der alternative Lösungen und kreative Ansätze erkundet. Die Ausgabe ist weniger durch etablierte Muster eingeschränkt.

## Top K

Top-k ist ein Parameter, der steuert, wie viele relevante Dokumente oder Informationen aus einer Wissensdatenbank abgerufen werden, um sie in die Generierung der Antworten einzubeziehen. Dieser Parameter kann nur verwendet werden, wenn der Assistent auf Dokumentencontainer zugreift.

- Top-k = 1: Es wird nur das relevanteste Dokument verwendet. Dies reduziert die Informationsmenge, kann aber zu einer präzisen, aber eingeschränkten Antwort führen.
- Top-k = 5: Die KI verwendet die 5 relevantesten Dokumente, um die Antwort zu generieren. Dies erhöht die Breite des Kontextes und verbessert die Wahrscheinlichkeit, dass die richtige Information enthalten ist.
- Top-k = 10 oder mehr: Die KI hat Zugriff auf viele Informationen, was bei komplexen Anfragen hilfreich ist, aber die Antwort auch weniger fokussiert oder redundanter machen kann.

Top-k ist entscheidend, um das Gleichgewicht zwischen Präzision und Vollständigkeit zu steuern. Die Wahl des Werts hängt davon ab, wie spezifisch oder umfassend die Antwort sein soll.

---

## Assistenten Details und Bearbeitung

In die Detailansicht eines Assistenten gelangen Sie durch Auswahl (Klick) auf den gewünschten Assistenten in der Übersicht. Anschließend werden alle Detaildaten des Assistenten angezeigt.

Für die Bearbeitung des Assistenten gelten die gleichen Prinzipien wie unter [Assistenten erstellen](#) beschrieben.

---

## Assistenten löschen

Wird ein Assistent nicht mehr benötigt, kann dieser gelöscht werden. Rufen Sie dazu die Assistenten-Details auf und wählen Sie "Löschen", um den Assistenten dauerhaft zu löschen. Alternativ können Sie den Assistenten-Status auf "Inaktiv" setzen, womit der Assistent für NutzerInnen im Chat-Bereich nicht mehr zur Verfügung steht.

## 2.6.8 Prompt Bibliothek

Business GPT bietet 2 Arten von Prompts: persönliche Prompts Unternehmens-Prompts.

**persönliche Prompts** sind private Prompts eines Nutzers, während **Unternehmens-Prompts** für mehrere Nutzer:Innen zur Verfügung stehen und zentral verwaltet werden.

### Persönliche Prompts

Diese Prompts sind Ihren privaten, persönlichen Prompts und können nur von Ihnen verwaltet und genutzt werden. Hier können Sie neue Prompts erstellen oder bestehende Prompts bearbeiten und löschen. Zudem können Sie Ihre Favoriten verwalten.

The screenshot shows the 'Prompt-Bibliothek' interface. At the top, there are three tabs: 'Persönliche Prompts', 'Unternehmens-Prompts', and 'Favoriten'. Below the tabs is a search bar with the text 'Prompts suchen' and '2 Ergebnisse' on the right. Underneath, there are two filter buttons: 'Titel der Prompt-Karte' and 'Schlagworte'. To the right, it says '0 ausgewählt' with 'Abwählen' and 'Löschen' buttons. The main area displays two prompt cards. The first card is titled 'Marketing' and contains the text 'Erstelle mir eine Social Media Kampagne für den Bereich "Marketing"'. The second card is titled 'Trends' and contains the text 'Bitte finde aktuelle Trends im Bereich Social Media. Erstelle daraus eine Social...'. Below the cards is a large grey button with a plus sign. At the bottom right, there are navigation arrows and the text '1 of 1'.

### PERSÖNLICHE PROMPTS VERWALTEN

Ihre persönlichen Prompts können Sie wie folgt bearbeiten.

Attribut	Erläuterung
Prompt Titel	Geben Sie hier die Bezeichnung für den Prompt an.
Tags	Geben Sie hier optional Schlagworte ein, bspw. für die Kategorisierung des Prompts
Prompt	Geben Sie hier Ihre Anweisungen, Aufgabenstellung, etc. an

### Unternehmens-Prompts

Diese Prompts können zentral durch Administratoren verwaltet werden. Unternehmens-Prompts können ergänzend zu den persönlichen Prompts für Nutzer:Innen zur Verfügung gestellt werden. In der [Prompt Bibliothek](#) werden im Tab [Unternehmens-Prompts](#) alle vorhandenen Prompts angezeigt.

# Prompt-Bibliothek

Persönliche Prompts

Unternehmens-Prompts

Favoriten

Prompts suchen

1 Ergebnis ✕

Titel der Prompt-Karte ↑

Schlagworte ∨

0 ausgewählt

Abwählen

Löschen



Prompt 1



Gib mir Hilfestellung zum folgenden Thema ...



< 1 of 1 >

## UNTERNEHMENS-PROMPT VERWALTEN

Unternehmens-Prompts verfügen über die folgenden Angaben und Einstellungen.

Attribut	Erläuterung
Prompt Titel	Geben Sie hier die Bezeichnung für den Prompt an.
Tags	Geben Sie hier optional Schlagworte ein, bspw. für die Kategorisierung des Prompts
Benutzergruppen	Sie können den Prompt optional nur für bestimmte Benutzergruppen freigeben. Ohne Einschränkung steht der Prompt für alle Benutzer:Innen zur Verfügung
Prompt	Geben Sie hier Ihre Anweisungen, Aufgabenstellung, etc. an
Mit dem Unternehmen teilen	Sobald Sie den Prompt veröffentlichen möchten, aktivieren Sie diese Option.

## Unternehmens-Prompt bearbeiten

**Titel\***

Prompt

7 / 100

**Schlagworte** ⓘ

Tags finden oder erstellen...

0 / 3

**Benutzergruppen-Berechtigungen** ⓘ

Für alle Benutzer verfügbar

**Mit dem Unternehmen teilen** **Nutzergruppen verwalten****Prompt\***

Geben Sie hier Ihren Prompt ein

0 / 16000

Nutze die **[Placeholder]**-Notation, um Textteile zu kennzeichnen, die im Prompt ersetzt werden sollen. Zum Beispiel: 'Handle als **[Job]** und erledige **[Aufgabe]**.'

**Löschen**

Abbrechen

Speichern

### UNTERNEHMENS-PROMPT NUTZEN




Unternehmens-Prompts können im Chatbereich in der Promptübersicht eingesehen und verwendet werden, sobald sie veröffentlicht wurden.



×

## Prompt-Bibliothek Prompt erstellen

---

×  
 ∨

-  **Favoriten** 1 >
-  **Persönliche Prompts** 2 >
-  **Unternehmens-Prompts** 1 ∨

 **Hilfe-Prompt** 

Gib mir Hilfestellung zum folgenden Thema ...

---

## 2.6.9 Verbrauchsübersicht

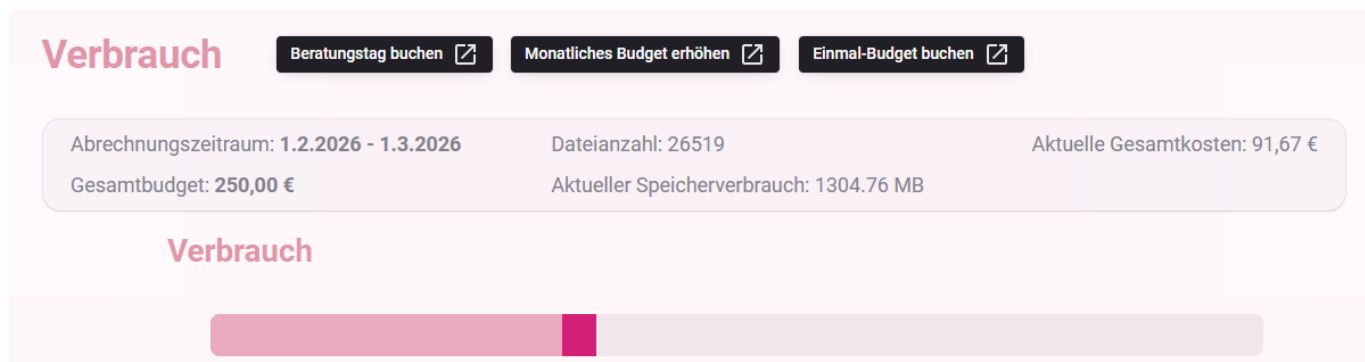
Zur Verbrauchsübersicht gelangt man per Klick auf die Verbrauchsanzeige im oberen Bereich der Webseite. Diese Ansicht steht ausschließlich Administratoren zur Verfügung.



In der Verbrauchsübersicht werden die Verbrauchskosten und -informationen des aktuellen Monats angezeigt. Ausgehend vom gesamten zur Verfügung stehenden Budget werden unter anderem die folgenden Informationen angezeigt:

- Verbleibendes Budget
- Aktuelle Gesamtkosten
- Aufteilung der Verbrauchskosten nach Chat Token Verbrauch und Index Token Verbrauch

Darüber hinaus können die Verbrauchskosten je zugrunde liegendem Modell eingesehen werden.



Zu einer weiteren detaillierten Ansicht der Verbrauchskosten gelangt man per Klick auf den Verbrauchsbalken.

• Index Token Verbrauch: 83,64 € • Chat Token Verbrauch: 9,48 € • Suche: 0,03 € • Verbleibendes Budget: 156,85 €

Kategorie	Verbrauchskosten	Model	Preis pro Einheit	Verbrauchte Einheiten
Index-Token	~0,45 €	text-embedding-ada-002	~0,0000001 €	4155703 Token
Chat-Token	~3,06 €	gpt-5.1 (Eingabe)	~0,0000012 €	2476730 Token
Chat-Token	~0,28 €	gpt-5.1 (Ausgabe)	~0,0000099 €	27803 Token
Chat-Token	~2,81 €	gpt-4.1 (Eingabe)	~0,0000020 €	1419849 Token
Chat-Token	~0,37 €	gpt-4.1 (Ausgabe)	~0,0000079 €	46393 Token
Chat-Token	~2,19 €	gpt-4o (Eingabe)	~0,0000025 €	884250 Token
Chat-Token	~0,43 €	gpt-4o (Ausgabe)	~0,0000099 €	43255 Token
Chat-Token	~0,01 €	gpt-5-nano (Eingabe)	~0,0000000 €	114301 Token
Chat-Token	~0,01 €	gpt-5-nano (Ausgabe)	~0,0000004 €	23433 Token
Chat-Token	~0,01 €	gpt-5-mini (Eingabe)	~0,0000002 €	53569 Token
Chat-Token	~0,05 €	gpt-5-mini (Ausgabe)	~0,0000020 €	23529 Token
Chat-Token	~0,00 €	o3-mini (Eingabe)	~0,0000011 €	2183 Token
Chat-Token	~0,01 €	o3-mini (Ausgabe)	~0,0000044 €	2367 Token
Chat-Token	~0,00 €	gpt-4.1-mini (Eingabe)	~0,0000004 €	788 Token
Chat-Token	~0,00 €	gpt-4.1-mini (Ausgabe)	~0,0000016 €	68 Token
Index-Token	~0,26 €	dalle3.standard.low	~0,0871200 €	3 Bilder
Index-Token	~83,19 €	Document Intelligence	~0,0090000 €	9243 Seiten
Suche	~0,03 €	Grounding with Bing Search	~0,0126000 €	2 Transaktion

## Verbrauchsdetails

Die Detailübersicht zeigt verschiedene Kategorien ( `Index-Token` , `Chat-Token` , `Suche` ). Je nach Nutzung von Business GPT entstehen Verbräuche.

### Index-Token

Die Kategorie `Index-Token` setzt sich zusammen aus `text-embedding` und Kosten für `Document Intelligence` . Kosten für `Document Intelligence` entstehen immer dann, wenn Dokumente hochgeladen werden - direkt in den Chat oder in Dokumenten-Container. Dabei wird der Inhalt des Dokumentes extrahiert, womit pro extrahierte Seite Kosten entstehen. `Embedding` -Kosten entstehen immer dann, wenn Daten in einen (Dokumenten-) Container geladen werden. Dies ist der Fall beim Upload von Dokumenten in einen Container sowie bei der Anbindung von Datenquellen wie Jira, wo ebenfalls Daten in einen zugrundeliegenden Container geladen und dort gespeichert werden.

`Embedding` -Kosten entstehen hingegen **nicht**, wenn Dokumente in den Chat geladen werden, da diese Dokumente nicht in sogenannten Vektoren (im RAG) gespeichert werden.

### Suche

Kosten für die Kategorie `Suche` entstehen immer dann, wenn Sie Assistenten mit Internet Breakout (Websuche) verwenden. Für der Assistent eine Websuche durch fallen Kosten pro Suchanfrage an.

### Chat-Token

Chat-Token werden auf zweierlei Art verbraucht: per Eingabe und per Ausgabe. Eingabe-Token werden immer dann verbraucht, wenn eine Anfrage an die KI gestellt wird, d.h. für jeden gesendeten Inhalt im Chatverlauf. Ausgabe-Token werden verbraucht, wenn die KI eine Antwort generiert. Auch hier gilt: Je mehr Text sowohl in der Eingabe als auch in der Ausgabe steht, umso höher sind die Verbrauchskosten.

**Wichtig**

Werden Dokumente direkt in den Chat hochgeladen, wird der gesamte Inhalt des Dokuments in der Anfrage an die KI/das Modell mitgesendet. Je mehr Inhalt das Dokument hat, umso größer ist der sogenannte Input und umso höher ist der Verbrauch von Eingabe-Token. 1 Token entspricht in etwa 4 Zeichen Text.

### Budget nachbestellen

Neigt sich das monatliche Budget dem Ende oder ist es gänzlich aufgebraucht, kann Budget über den Telekom Cloud Marketplace für Geschäftskunden ganz einfach nachbestellt werden. Je nach Anforderung können dabei Tokens einmalig oder dauerhaft als zusätzliches monatliches Token-Budget geordert werden.

#### € Einmaliges Budget

---

Buchen Sie einmaliges Zusatzvolumen über

→ [Telekom Cloud Marketplace](#)

#### €€ Monatliches Budget

---

Buchen Sie monatliches Zusatzvolumen über

→ [Telekom Cloud Marketplace](#)

## 2.7 Integrationen

---

### 2.7.1 Integrationen

---

Integrationen bieten die Möglichkeit Business GPT mit Ihren unternehmenseigenen Tools und Systemen zu verbinden. Business GPT bietet dazu unterschiedliche Optionen an.

#### **Integration von Systemen als Importdatenquelle**

Systeme wie **JIRA Datacenter** können als Importdatenquelle angebunden werden. Diese Systeme können auf diese Weise als dauerhafte Wissensquelle verfügbar gemacht werden.

#### **Custom Integrationen**

Mittels **Function Calling** können Sie Assistenten befähigen externe Tools und Systeme per API-Aufruf zu integrieren. Auf diese Weise können verschiedenste Aktionen, wie z.B. der Abruf von Daten, ausgeführt werden.

## 2.7.2 Jira Integration

Mittels Jira Integration können Sie Ihr unternehmenseigenes **Jira Datacenter** als Datenquelle an Business GPT anbinden. Mittels dieser Integration liest Business GPT Jira-Ticketinhalte und speichert und indexiert diese intern, sodass anschließend mit diesen Ticketinhalten gechattet werden kann. Die Organisation und Verwaltung von Jira-Ticketinhalten erfolgt **projektbasiert**. Das heißt, dass Sie eines oder mehrere Jira-Projekte importieren und unterschiedlichen Nutzergruppen diese Projektinhalte zugänglich machen können. Je Jira-Projekt können Sie jeweils eine Jira-Integration anlegen.

### Jira Integration einrichten

Um ein neues Jira-Projekt mit Business GPT zu synchronisieren, wählen Sie in der Administration unter **Integrationen** in der Übersicht **Neue Integration anlegen** (Typ Jira Datacenter). Nachfolgend können Sie folgende Einstellungen vornehmen:

Feld	Beschreibung
Name	Legen Sie hier die Bezeichnung für die Integration bzw. das Jira Projekt fest
Beschreibung	Beschreiben Sie optional das Projekt
JIRA Verbindungseinstellungen	Geben Sie hier den <b>JIRA Zugriffstoken</b> des zu verwendenden (Service-) Accounts sowie <b>JIRA Endpunkt</b> (URL) und den <b>JIRA Projektschlüssel</b>
Benutzergruppen-Berechtigungen	Legen Sie hier fest, welche Benutzergruppen Zugriff auf die Jira-Projektinhalte in Business GPT erhalten soll. Diese Benutzergruppen können dann die Jira-Integration im Chat benutzen. Bitte beachten Sie, dass Business GPT <b>keine</b> Berechtigungsgruppen oder -strukturen aus Ihrem Jira Datacenter synchronisiert.
Planung	Optional können Sie den automatischen Import von Jira-Ticketdaten planen, indem Sie den gewünschten Zeitpunkt für den automatischen Lauf festlegen. Sie können dabei wählen zwischen: <b>Täglich</b> , <b>Wöchentlich</b> und <b>Monatlich</b>

←
Speichern
Abbrechen

# Integrationen

## Integration Name

**JIRA**  
JIRA Datacenter

Ändern

**Name\***

**Beschreibung**

**JIRA Verbindungs-Einstellungen**

**JIRA Zugriffstoken\***

**JIRA Endpunkt\***

**JIRA Projekt\***

**Benutzergruppen-Berechtigungen**

ⓘ Keine Berechtigungen erteilt

**Planung**

ⓘ Die Synchronisierung läuft außerhalb der üblichen Geschäftszeiten zwischen Mitternacht und 5 Uhr morgens.

**Täglich**

Import jeden Tag ausführen.

**Wöchentlich** ✓

Wählen Sie bestimmte Wochentage aus.

**Monatlich**

Wählen Sie einen bestimmten Tag des Monats aus.

An diesen Tagen ausführen:

Mo Di Mi Do Fr Sa So

#### VERBINDUNG TESTEN

Um zu prüfen, ob Ihre Jira-Verbindungsdaten korrekt sind, können Sie die Verbindung vorab testen. Stellen Sie zunächst sicher, dass Sie die Integrationsdaten vorab gespeichert haben.

**Testen**

**Integration testen**

**Testergebnis**

✓ **Erfolg**

Connection Successful  
Trying to connect to '  
Connected to Server'

Checking access to project '  
Can Access project

**Import & Indizieren**

**Import und Indizierung starten** **Indizierung stoppen**

#### IMPORT UND INDIZIERUNG STARTEN

Um den Import- und Indizierungsvorgang zu starten wählen Sie **Import und Indizierung starten**. Anschließend

**Import-Dauer**

Die Dauer des Imports ist von der Gesamtanzahl von Jira-Tickets abhängig. Je mehr Tickets importiert werden, desto länger dauert der initiale Importvorgang.

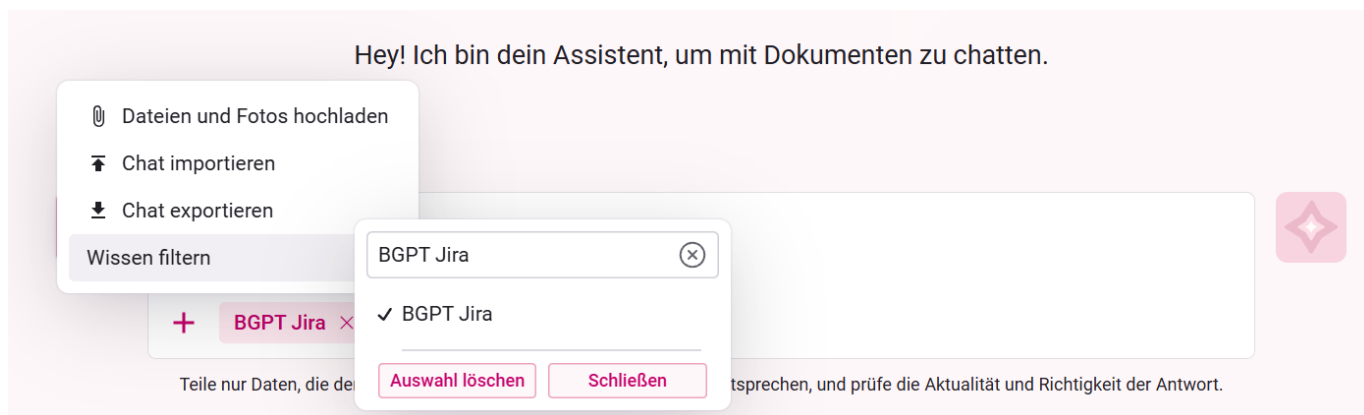
Die Gesamtanzahl an Jira-Tickets können Sie über die Statusanzeige nachvollziehen.

**0 Tickets warten auf die Indizierung**
0 Tickets fehlgeschlagen
1760 Tickets erfolgreich ⓘ

#### Jira Integration verwenden

Um eine Jira Integration zum Chatten mit den Projekthinhalten zu verwenden, können Sie entweder den Standard Assistenten **Dokumenten Chat** verwenden oder einen individuellen Assistenten erstellen und diesem die gewünschte Jira-Integration zuweisen (siehe [Assistenten](#)).

Wenn Sie die Jira-Integration im **Dokumenten Chat** verwenden, können Sie über **Wissen filtern** direkt auf den Jira-Inhalt zugreifen.



## Häufige Fragen zur Jira Integration

### Welche Kosten verursacht der Import von Jira-Inhalten?

Beim Import und der anschließenden Indexierung der Jira-Ticketinhalte werden Embedding-Token (Index-Token, text-embedding-ada-002) verbraucht. Pro 1 Mio Token betragen die Kosten 0,1 €. Als Näherungswert können Sie im Mittel in etwa von 300 Token pro Ticket ausgehen. Basierend auf diesem Näherungswert würden sich die Kosten bei 10.000 Tickets auf ca. 0,30 Cent belaufen.

### Werden Ticketanhänge wie Dokumente oder Bilder ebenfalls importiert?

Nein, der Import verarbeitet aktuell keine Ticket-Anhänge.

### Kann ich mit der Jira-Integration Tickets erstellen?

Nein, bei der Jira Integration handelt es sich um eine Datenquellen-Integration, bei der ausschließlich lesend auf die Ticketdaten in Ihrem Jira zugegriffen wird. Sie können jedoch für diesen Fall [Function Calling](#) verwenden.

### Wird das gesamte Jira-System importiert?

Nein, Sie können die Jira-Inhalte projektweise importieren und somit nur ausgewählte und gewünschte Projekte in Business GPT verfügbar machen.

### Werden die Berechtigungen auf Jira-Seite beim Import berücksichtigt und übernommen?

Nein, Berechtigungen in Ihrem Jira werden nicht automatisch übernommen. Stattdessen können Sie ausgewählten Benutzergruppen die Jira-Integration erlauben, indem Sie die gewünschten Gruppen der Jira-Integration zuweisen.

## 3. Ressourcen

---

### 3.1 Prompting

---

Um das Beste aus Business GPT herauszuholen, ist die Qualität der Eingabe – des sogenannten Prompts – von zentraler Bedeutung. Ein Prompt ist die Anweisung, mit der Sie dem Chatbot mitteilen, was er für Sie erledigen soll. Je klarer und präziser diese Anweisung ist, desto hilfreicher, passender und strukturierter fällt die Antwort aus.

#### 3.1.1 Was macht einen guten Prompt aus?

---

Ein gut formulierter Prompt hilft der KI, genau zu verstehen, worum es geht – ganz ähnlich wie bei der Kommunikation mit einem Menschen. Statt einer vagen Anfrage wie „Schreib mir etwas über Projektmanagement“ erhalten Sie deutlich bessere Ergebnisse mit einer konkreten Aufgabenstellung wie:

„Erkläre die fünf wichtigsten Prinzipien des agilen Projektmanagements in einer übersichtlichen Liste für Einsteiger.“

#### 3.1.2 Drei Erfolgsfaktoren für gute Prompts

---

1. Klarheit: Beschreiben Sie möglichst konkret, was Sie brauchen. Allgemeine oder offene Formulierungen führen oft zu oberflächlichen Antworten.
2. Kontext: Liefern Sie Hintergrundinformationen, die für die Aufgabe wichtig sind. Zum Beispiel: Wer ist die Zielgruppe? Wofür wird der Text benötigt? Gibt es bestimmte Einschränkungen?
3. Formatwunsch: Teilen Sie mit, in welchem Format Sie die Antwort benötigen – z. B. als Liste, Tabelle, kurzer Text, Nachricht oder Markdown. Das spart Nacharbeit und verbessert die Nutzbarkeit der Antwort.

#### 3.1.3 Beispiel für einen effektiven Prompt

---

Statt:

"Hilf mir mit einer Marketingidee."

Besser:

"Gib mir drei kreative Social-Media-Ideen für die Einführung eines neuen veganen Energy-Drinks, die sich an eine junge, sportliche Zielgruppe richten. Bitte in Stichpunkten."

#### 3.1.4 Vier Bausteine für starke Prompts

---

Ein effektiver Prompt enthält im Idealfall die folgenden vier Bestandteile:

##### 1. Persona – Welche Rolle soll der Chatbot übernehmen?

Indem Sie eine Rolle oder Perspektive vorgeben, kann sich die KI besser in die Aufgabe hineinversetzen. Das erhöht die Qualität der Antwort und den inhaltlichen Fokus. Aktuelle Forschung zeigt, dass Personas umso wirkungsvoller sind, je präziser und mehrdimensional sie definiert werden (z. B. Fachgebiet, Kommunikationsstil, Werthaltung). In der Formulierung sollten Sie:

- **Spezifisch sein:** Statt „Marketingberater“ z. B. „Senior Content-Marketing-Manager für B2B-Technologieunternehmen“.
- **Perspektive und Werte verankern:** „Antworten mit Fokus auf Nachhaltigkeit und Kundenorientierung.“
- **Iterativ verfeinern:** Bei unerwünschter Tonalität Persona anpassen und erneut nutzen.
- **Bias reduzieren:** Explizit Vielfalt und Inklusion in der Persona berücksichtigen, um einseitige Antworten zu vermeiden.

Beispiel: „Du bist eine Senior Content-Marketing-Managerin in einem B2B-Tech-Startup, die Wert auf Klarheit, Prägnanz und inklusives Wording legt.“



Personas helfen besonders dann, wenn Sie Fachwissen, bestimmte Tonalität oder Zielgruppenorientierung erwarten.

## 2. Aufgabe – Was soll genau getan werden?

Beschreiben Sie die konkrete Handlung oder das gewünschte Ergebnis. Je klarer Sie hier formulieren, desto besser wird die Antwort. Beispiele:

„Erkläre die Funktionsweise von Zwei-Faktor-Authentifizierung.“

„Fasse die wichtigsten Punkte aus dem folgenden Text zusammen.“

„Schreibe eine höfliche E-Mail-Antwort auf eine Beschwerde.“

## 3. Kontext – Was ist der Hintergrund?

Liefere Sie alle wichtigen Informationen, die das Modell kennen muss, um Ihre Aufgabe richtig einordnen zu können. Zum Beispiel:

- Zielgruppe oder Empfänger: „Diese Präsentation ist für mein nicht-technisches Team gedacht.“
- Projekthintergrund: „Es geht um die Einführung einer neuen HR-Software in unserem Unternehmen.“
- Dateianhänge oder Textauszüge: „Nutze die Informationen in diesem Textausschnitt...“



Kontext ist besonders entscheidend, wenn der Chatbot auf frühere Unterhaltungen aufbauen oder externe Informationen verarbeiten soll.

## 4. Format – Wie soll die Antwort aussehen?

Wenn Sie der KI vorgeben, in welchem Format Sie das Ergebnis brauchen, sparen Sie sich Zeit und Nachbearbeitung. Mögliche Formate sind:

- Aufzählung oder Stichpunkte
- Fließtext in freundlichem, professionellem oder technischem Ton
- Tabellenstruktur
- JSON oder Markdown
- Kurze Slack-/E-Mail-Nachricht



Beispiel:

„Bitte gib mir die Ergebnisse in einer Markdown-Tabelle mit den Spalten Tätigkeit, Priorität und Zeitaufwand.“

✓ Beispiel für einen vollständigen Prompt

Persona: Du bist ein UX-Experte.

Aufgabe: Erstelle drei konkrete Verbesserungsvorschläge für unsere mobile App.

Kontext: Die App wird vor allem von Studierenden genutzt, um Stundenpläne zu organisieren.

Format: Kurze Bulletpoints mit jeweils maximal 2 Sätzen Erklärung.

## 3.1.5 Weitere Erfolgsfaktoren für starke Prompts

**Reasoning:** Schrittweises Denken fördern (bei der Verwendung von Reasoning-Modellen wie o3-mini oder o1)

Nutzen Sie diesen Baustein nur, wenn das gewählte Modell oder Ihr gewählter Assistent Chain-of-Thought bzw. explizite Zwischenschritte unterstützt. Andernfalls kann die Aufforderung ignoriert oder zu unnötig ausführlichen Antworten führen.

Fordern Sie das Modell im Prompt aktiv auf, Zwischenschritte zu formulieren (Chain-of-Thought):

"Bitte erkläre deine Antwort Schritt für Schritt."

"Dokumentiere deine Herleitung, bevor du zur Schlussfolgerung kommst."

Das hilft vor allem bei komplexen Aufgaben, die logisches Denken erfordern.

### 3.1.6 Tipps & Tricks für optimale Prompts

Um mit dem Chatbot präzise und hilfreiche Antworten zu erhalten, ist die Kunst der Prompt-Gestaltung entscheidend. Hier finden Sie bewährte Strategien, die dabei helfen, klare und effektive Anweisungen zu formulieren – damit der Bot genau das liefert, was Sie benötigen.

#### Großbuchstaben gezielt und sparsam nutzen

Großbuchstaben ziehen Aufmerksamkeit auf wichtige Schlüsselbegriffe oder Bedingungen. Nutzen Sie sie, um Prioritäten hervorzuheben, z. B. „Bitte beantworte diese Frage AUSSCHLIESSLICH im Kontext von Marketing.“ Zu viel davon kann jedoch irritieren und den Lesefluss stören.

 Beispiel:

„Bitte antworte NUR im Stil eines professionellen Redakteurs.“

„Die Antwort muss AUF KEINEN FALL technische Fachbegriffe enthalten.“




weiterung

Alternativ können auch fettgedruckte oder kursiv geschriebene Worte in Tools mit Formatierung helfen, wichtige Punkte zu markieren, wenn das unterstützt wird.

#### Emotionale und motivierende Impulse (vorsichtig einsetzen)

Formulierungen, die eine gewisse Dringlichkeit oder Wichtigkeit vermitteln, können tatsächlich die Ausrichtung der Antwort beeinflussen, indem sie dem Modell signalisieren, dass Genauigkeit und Sorgfalt besonders relevant sind.

✓ Beispiele für Nudges im Prompt

 Dringlichkeit und emotionale Wichtigkeit

„Es ist entscheidend für meinen wichtigen Projektbericht, dass die folgende Erklärung absolut präzise und vollständig ist.“

„Für meine Abschlussarbeit brauche ich eine ausführliche und genaue Antwort zu diesem Thema – bitte achte besonders auf Details.“

„Diese Informationen sind kritisch für eine wichtige Kundenpräsentation. Bitte liefere eine sorgfältige und gut strukturierte Antwort.“

 Belohnungen (humorvoll / spielerisch)

„Ich gebe dir 50 virtuelle Punkte, wenn du mir eine kreative Idee für Social Media Marketing gibst.“

„Stell dir vor, du bekommst eine Auszeichnung für die beste Antwort – hilf mir bitte mit einer besonders guten Erklärung!“

„Wenn du mir bei dieser Aufgabe hilfst, verspreche ich dir, dich bei meinem nächsten Projekt zu empfehlen.“



weis

Solche Belohnungen wirken spielerisch und motivierend, obwohl die KI keine echten Belohnungen wahrnimmt.

 Emotionale Bezugnahme

„Bitte antworte als wärst du mein erfahrener Mentor, der mir in schwierigen Zeiten immer hilfreiche Ratschläge gegeben hat.“

„Stell dir vor, du bist meine verstorbene Großmutter, die mir liebevoll Geschichten über unsere Familiengeschichte erzählt hat.“

„Handle wie ein verständnisvoller Freund, der mir geduldig komplexe Themen erklärt.“

### 💡 Verstärkung durch wichtige Adjektive und Ausdrücke

„Bitte gib mir eine umfassende, sorgfältig recherchierte Analyse zum Thema.“

„Ich brauche eine detaillierte, klare Schritt-für-Schritt-Anleitung, die auch Einsteiger verstehen.“

„Die Antwort sollte besonders gründlich und fachlich fundiert sein, da sie für eine offizielle Dokumentation verwendet wird.“



**Wichtig**

Diese „Nudges“ funktionieren gut, wenn sie sparsam und glaubwürdig eingesetzt werden. Übertriebene „Bestechungen“ oder emotionales Erpressen helfen meist nicht und können zu unerwarteten Antworten führen.

### Klare Rollen oder Persona definieren

Wenn Sie möchten, dass der Chatbot aus einer bestimmten Perspektive antwortet, geben Sie eine klare Persona vor, z. B.:

„Antworte als erfahrener Arzt.“

„Erkläre wie ein Finanzexperte.“

„Schreibe im Stil von Shakespeare.“

Das verbessert nicht nur die Tonalität, sondern oft auch die Tiefe und fachliche Korrektheit.

### Vermeiden Sie Verneinungen – nutzen Sie positive Anweisungen

Negative Formulierungen (z. B. „nicht“, „kein“) sind für Sprachmodelle schwieriger zu interpretieren und können die Antwortqualität verschlechtern.

Besser: Statt

„Sprich nicht über andere Themen“

lieber

„Sprich ausschließlich über dieses Thema“.

### Definieren Sie den gewünschten Tonfall und Stil

Das bewusste Angeben eines Stils (z. B. freundlich, formell, humorvoll, sachlich) hilft der KI, die Antwort an den Kontext anzupassen und so passender zu gestalten.

Beispiel:

„Schreibe im Ton eines freundlichen Reisebloggers.“

„Verwende eine sachlich-nüchtere Sprache, wie in einem Geschäftsbericht.“

### Bitten Sie um konkrete Zitate und Quellenangaben

Da KI-Modelle manchmal erfundene Informationen liefern können (sogenannte „Halluzinationen“), ist es sinnvoll, sie bei kritischen Themen um Quellen, Zitate oder belegbare Informationen zu bitten.

💡 Beispiel:

„Bitte zitiere wörtlich aus der Quelle, wenn du die Definition von ‚künstlicher Intelligenz‘ gibst.“

„Liefere ein direktes Zitat, falls du Aussagen über den CO<sub>2</sub>-Ausstoß machst.“

### Umgang mit Textlängen und Antwortlimits

Die meisten Sprachmodelle haben eine maximale Tokenanzahl pro Antwort. Um längere Texte zu erhalten, fordere einfach eine Fortsetzung an („Weiter“, „Fortsetzen“ etc.).



Unterteile sehr lange Aufgaben in mehrere kleinere Schritte oder Fragestellungen. Das steigert die Qualität und Übersichtlichkeit der Antworten.

### Kontext und Hintergrundinformationen liefern

Je mehr relevante Details Sie zum Hintergrund Ihrer Anfrage geben, desto besser kann der Chatbot verstehen, was genau gebraucht wird.

 Beispiel: Statt nur

„Schreibe einen Werbetext“

lieber:

„Schreibe einen Werbetext für eine nachhaltige Modekollektion, der sich an umweltbewusste junge Erwachsene richtet.“

### Nutzen Sie Beispiele und Few-Shot-Prompting

Wenn Sie eine spezielle Art von Antwort erwarten, können Sie dem Bot Beispielanworten oder klare Muster vorgeben (sogenanntes Few-Shot-Lernen). Das sorgt für konsistentere und zielgenaue Ergebnisse.

 Beispiel:

„Hier ein Beispiel für die Art der Antwort, die ich suche:

Frage: Was ist Inflation?

Antwort: Inflation beschreibt den allgemeinen Anstieg von Preisen...

Jetzt bitte eine ähnliche Erklärung für ‚Deflation‘.“

### Verwenden Sie strukturierte Aufforderungen

Wenn Sie eine komplexe Antwort benötigen, strukturieren Sie den Prompt in Abschnitte (z. B. Persona, Aufgabe, Kontext, Format) oder nummerierte Schritte. Das macht die Anweisung für die KI leichter verständlich.

 Beispiel:

Rolle: Du bist ein Marketingspezialist.

Aufgabe: Entwickle einen Werbetext für ein veganes Energy-Getränk.

Zielgruppe: Junge Erwachsene, sportlich.

Tonfall: Locker, aktivierend.

Format: 3 Sätze, inklusive Call-to-Action.

### Nutzen Sie gezielte Schlüsselwörter und Schlagwörter

Manchmal hilft es, wichtige Begriffe mehrmals oder an strategischen Stellen zu wiederholen, damit die KI den Fokus nicht verliert.

 Beispiel:

„Bitte konzentriere dich auf die Begriffe Kundenzentrierung, Conversion Rate und Usability im E-Commerce-Kontext.“

„Verwende unbedingt die Wörter nachhaltig, regional, klimaneutral in deinem Produkttext.“

### Kontrollieren und optimieren Sie Prompts iterativ

Sehr oft verbessert sich die Antwortqualität, wenn Sie den Prompt nach ersten Ergebnissen verfeinern. Probieren Sie unterschiedliche Formulierungen, überprüfen Sie die Antworten kritisch und passen Sie den Prompt an.

 Beispiel-Dialog:

„Erkläre, wie Blockchain funktioniert.“

Antwort: „Blockchain ist eine verteilte Datenbank...“

Prompt 2: „Bitte erläutere das für Schüler der 10. Klasse – ohne Fachbegriffe.“

Prompt 3: „Jetzt fasse das Ganze in drei kurzen Bullet Points zusammen.“

## 3.2 Prompt Library

---

### 3.2.1 Was ist eine Prompt Library?

---

Eine **Prompt Library** ist eine zentrale Sammlung wiederverwendbarer Eingaben (Prompts) für KI-Systeme wie Business GPT. Sie dient dazu, bewährte Prompts strukturiert zu speichern, zu verwalten und effizient wiederzuverwenden.

Statt jedes Mal neue Anweisungen zu formulieren, greifen Nutzer:Innen auf optimierte und getestete Prompts zurück – ähnlich wie auf Vorlagen oder Bausteine.

---

### 3.2.2 Warum ist eine Prompt Library wichtig?

---

Eine Prompt Library ist kein „Nice-to-have“, sondern wird mit zunehmender Nutzung von KI zu einem entscheidenden Produktivitätsfaktor.

Sie hilft dabei:

- **Zeit zu sparen** durch Wiederverwendung bewährter Prompts
- **Konsistente Ergebnisse** zu erzielen
- **Fehler zu reduzieren**, da geprüfte Prompts genutzt werden
- **Wissen zu bündeln** und zugänglich zu machen
- **Arbeitsabläufe zu standardisieren**

Eine gut gepflegte Prompt Library macht KI von einem experimentellen Tool zu einem **verlässlichen Bestandteil des Arbeitsalltags**.

---

### 3.2.3 Wann lohnt sich eine Prompt Library?

---

Eine Prompt Library lohnt sich besonders dann, wenn KI nicht nur gelegentlich, sondern regelmäßig eingesetzt wird.

#### Typische Faktoren:

- Sie bearbeiten regelmäßig **ähnliche Aufgaben** mit KI
- Prompts werden aktiv **optimiert und wiederverwendet**
- Sie nutzen häufig **Copy & Paste**
- Ergebnisse sollen **konsistent und reproduzierbar** sein
- Mehrere Personen arbeiten mit KI
- Die Qualität hängt stark von einzelnen Nutzern ab

#### Faustregel:

Sobald **Wiederholung, Qualität und Skalierung** wichtig werden, lohnt sich eine strukturierte Prompt Library.

---

### 3.2.4 Warum einfache Dokumente nicht ausreichen

---

Viele starten mit Tools wie Notion oder Google Docs. Das funktioniert kurzfristig, stößt aber schnell an Grenzen:

- Keine echte **Struktur oder Versionierung**
- Hoher Aufwand durch **manuelles Kopieren**
- Fehlende **Kontrolle über Änderungen**
- Keine Unterstützung für **Variablen oder Templates**

Dokumente speichern Inhalte –  
eine Prompt Library **steuert den Workflow**.

---

## 3.2.5 Wie eine Prompt Library funktioniert

---

Eine moderne Prompt Library basiert meist auf dem Prinzip von **Templates**:

### 1. Fester Prompt-Kern

- Enthält Struktur, Regeln und Kontext
- Bleibt unverändert
- Sorgt für gleichbleibende Qualität

### 2. Variable Inhalte

- Werden bei der Nutzung ausgefüllt (z. B. Thema, Zielgruppe)
- Ermöglichen flexible Anpassung

Ein Prompt funktioniert wie ein **Formular**, nicht wie freier Text.

---

## 3.2.6 Vorteile einer Prompt Library

---

### Effizienzsteigerung

- Keine wiederholte Formulierung nötig
- Schneller Zugriff auf bewährte Lösungen

### Konsistente Qualität

- Einheitliche Standards
- Gleichbleibende Tonalität und Struktur

### Skalierbarkeit im Team

- Gemeinsamer Wissensstand
- Weniger Abhängigkeit von Einzelpersonen

### Wissensmanagement

- Prompts werden zu **wertvollen Assets**
- Best Practices bleiben erhalten

### Schnellere Einarbeitung

- Neue Nutzer lernen durch Beispiele
  - Klarer Einstieg in die Nutzung von KI
-

### 3.2.7 Typische Einsatzbereiche

---

Prompt Libraries sind besonders nützlich für:

- Content-Erstellung (Blogartikel, Social Media, SEO)
  - Kundenkommunikation (E-Mails, Support-Antworten)
  - Softwareentwicklung (Code-Generierung, Debugging)
  - Dokumentation und interne Prozesse
  - Analyse und Reporting
- 

### 3.2.8 Was eine gute Prompt Library auszeichnet

---

#### Struktur & Organisation

- Kategorien, Tags, Suchfunktion
- Klare Zuordnung zu Anwendungsfällen

#### Wiederverwendbarkeit

- Templates mit Variablen
- Einfache Anpassung ohne Neuschreiben

#### Integration in den Workflow

- Nutzung direkt im Arbeitskontext
- Kein Copy & Paste notwendig

#### Versionierung & Qualitätssicherung

- Zentrale Pflege
- Geprüfte und optimierte Prompts

#### Zusammenarbeit

- Gemeinsame Nutzung im Team
  - Klare Standards und Guidelines
- 

### 3.2.9 Herausforderungen und Grenzen

---

- Prompts müssen **regelmäßig gepflegt** werden
- Änderungen an KI-Modellen können Auswirkungen haben
- Zu starre Templates können Kreativität einschränken

Eine Prompt Library ist kein statisches System, sondern ein **lebendiges Werkzeug**.

---

### 3.2.10 Fazit

---

Eine Prompt Library wird dann wertvoll, wenn KI **systematisch und regelmäßig** eingesetzt wird.

Sie sorgt für:

- Mehr Effizienz
- Konsistentere Ergebnisse
- Bessere Zusammenarbeit

Sie übernimmt die Wiederholung – damit sich Nutzer auf die eigentliche Arbeit konzentrieren können.

## 3.3 Chat mit Dokumenten und internem Wissen

---

Business GPT kann Dokumente auf zwei verschiedene Arten verarbeiten – je nachdem, ob Sie **ein oder mehrere einzelne Dokumente spontan analysieren** oder eine **dauerhafte Wissensbasis** aufbauen möchten.

### 3.3.1 1. Dokumente direkt im Chat hochladen

---

Sie können Dokumente **direkt an den Chat anhängen**, um sie sofort mit der KI zu besprechen. Der Inhalt steht dem Modell dann **vollständig im aktuellen Gesprächskontext** zur Verfügung – Sie können also direkt Fragen stellen, Textpassagen analysieren oder Unterschiede zwischen Dokumenten hervorheben lassen.

#### Wichtig zu wissen:

- Dokumente können **nur hochgeladen werden, wenn ihr gesamter Inhalt in das Kontextfenster des Modells passt**. Ist ein Dokument (oder eine Kombination mehrerer Dateien) größer als dieses Limit, kann der Inhalt nicht vollständig verarbeitet werden. Sie erhalten dann im Chat einen Hinweis und können Dokumente entfernen.
- Diese Methode eignet sich daher besonders für einzelne oder zusammenhängende Dokumente beliebiger Länge, solange sie innerhalb des verfügbaren Kontextfensters liegen.
- Die hochgeladenen Dateien werden als **temporäre Dokumente** gespeichert – sie bleiben **für eine begrenzte Zeit** verfügbar, bevor sie **automatisch gelöscht** werden.
- Während dieser Frist können Sie die Dokumente erneut verwenden, beispielsweise in einem weiteren Chat.

#### Typische Anwendungsfälle:

- „Fasse dieses Dokument zusammen.“
- „Welche Risiken werden im Abschnitt 3 genannt?“
- „Erstelle daraus eine Management-Zusammenfassung.“
- „Vergleiche beide Dokumente und zeige die Unterschiede auf.“

### 3.3.2 2. Dokumente in Container hochladen (Wissensordner)

---

Für umfangreiche oder dauerhaft relevante Dokumente können Sie **Dokumenten-Container verwenden** (siehe auch [Dokumenten-Container](#)). Dort hochgeladene Dateien werden **indexiert** und später im Chat über eine **semantische Suche (RAG-Verfahren)** verarbeitet. Das bedeutet: Der Assistent durchsucht beim Chatten nicht das gesamte Dokument Wort für Wort, sondern **findet inhaltlich passende Textstellen**, auch wenn Sie sich anders ausdrücken.

#### Vorteile:

- Ideal für **größere oder thematisch zusammenhängende Dokumentensammlungen** (z. B. Projektdokumente, Richtlinien, Reports).
- Kein Limit durch das Kontextfenster – der Assistent kann über die semantische Suche **auf den gesamten Wissensbestand** zugreifen.
- Container können **dauerhaft verwaltet, erweitert oder gelöscht** werden.
- Besonders hilfreich für **Dokumentenvergleiche über mehrere Versionen hinweg**, z. B. um **Änderungen, Ergänzungen oder Abweichungen** zu erkennen.

#### Typische Anwendungsfälle:

- „Was sagen unsere Sicherheitsrichtlinien zu Cloud-Diensten?“
- „Welche Anforderungen gelten laut der Produktspezifikation?“
- „Fassen Sie alle Reports der letzten Monate zusammen.“
- „Vergleichen Sie die Versionen unserer Produktspezifikationen und listen Sie die Unterschiede auf.“

### Wann nutze ich welche Methode?

Beide Methoden haben ihre Stärken – je nachdem, ob Sie **tiefe, präzise Analysen** oder **breiten Wissenszugriff** benötigen.

- Wenn Sie ein Dokument **direkt im Chat anhängen**, steht der gesamte Inhalt **vollständig im Modellkontext**. Das Modell „kennt“ dann jedes Wort und kann **präzise Zusammenhänge, Argumentationsketten oder Unterschiede** erkennen. Diese Methode liefert die **höchste inhaltliche Genauigkeit**, solange das Dokument ins Kontextfenster passt.
- Wenn Sie Dokumente in **Containern** ablegen, greift das Modell über eine **semantische Suche (RAG)** auf relevante Textstellen zu. Dadurch können Sie mit sehr großen oder vielen Dokumenten arbeiten, allerdings immer auf Basis von **inhaltlich passenden Ausschnitten**, nicht des gesamten Textes. Diese Methode ist ideal für **umfangreiche Wissensbestände**, bietet aber eine **indirektere Kontexttiefe**.

Ziel	Empfohlene Methode
Präzise inhaltsbezogene Analyse	<b>Direkter Upload im Chat</b>
Dokument passt vollständig ins Kontextfenster	<b>Direkter Upload im Chat</b>
Sehr große oder viele Dokumente	<b>Container (Wissensordner)</b>
Dauerhafte Wissensbasis	<b>Container (Wissensordner)</b>
Dokumentenvergleiche (innerhalb eines Kontextes)	<b>Direkter Upload im Chat</b>
Versionsvergleiche über längeren Zeitraum	<b>Container (Wissensordner)</b>

#### Practice

Wenn Ihr Dokument **in das Kontextfenster passt**, ist der **direkte Upload im Chat** immer die genauere Wahl. Die **Container-Methode** eignet sich, sobald Sie **mehr Dokumente oder mehr Wissen** benötigen, als in einem einzelnen Kontext verarbeitet werden kann.

#### Kombination beider Methoden

Sie können beide Ansätze kombinieren. Speichern Sie z. B. Ihren gesamten Wissensbestand dauerhaft in Containern und hängen Sie im Chat zusätzlich ein aktuelles Dokument an. So kann die KI einerseits auf das semantisch indizierte Wissen zugreifen und gleichzeitig das aktuelle Dokument im Detail analysieren oder vergleichen. Diese Kombination ist besonders wirkungsvoll, wenn Sie neue Dokumente in bestehenden Kontext einordnen oder Abweichungen feststellen möchten.

Der **Unternehmenswissen**-Assistent ist ein vorkonfigurierter Assistent, der es erlaubt mit zusätzlichem Wissen zu chatten. Dieses Wissen wurde zuvor in Business GPT zum Beispiel durch den Upload von Dokumenten in Dokumenten-Containern verfügbar gemacht.

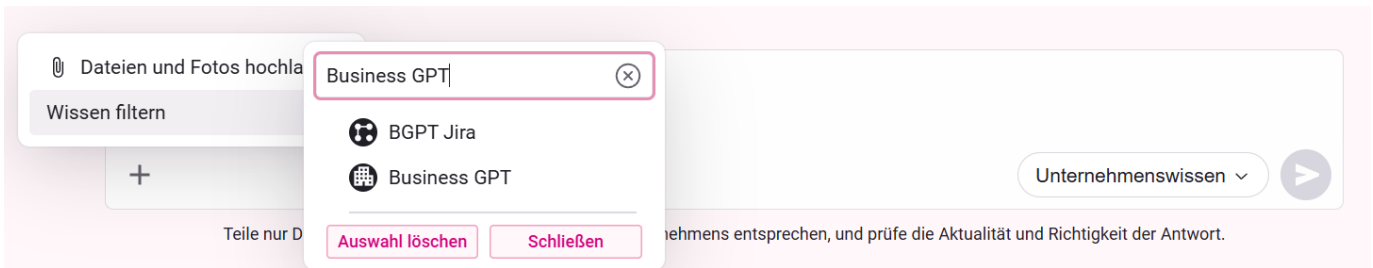
Der Assistent greift bei der Beantwortung von Anfragen primär auf dieses Wissen zu und erstellt Antworten unter Anwendung von Keywörtern und Kontext der Anfrage.

### Chat mit Dokumenten-Containern

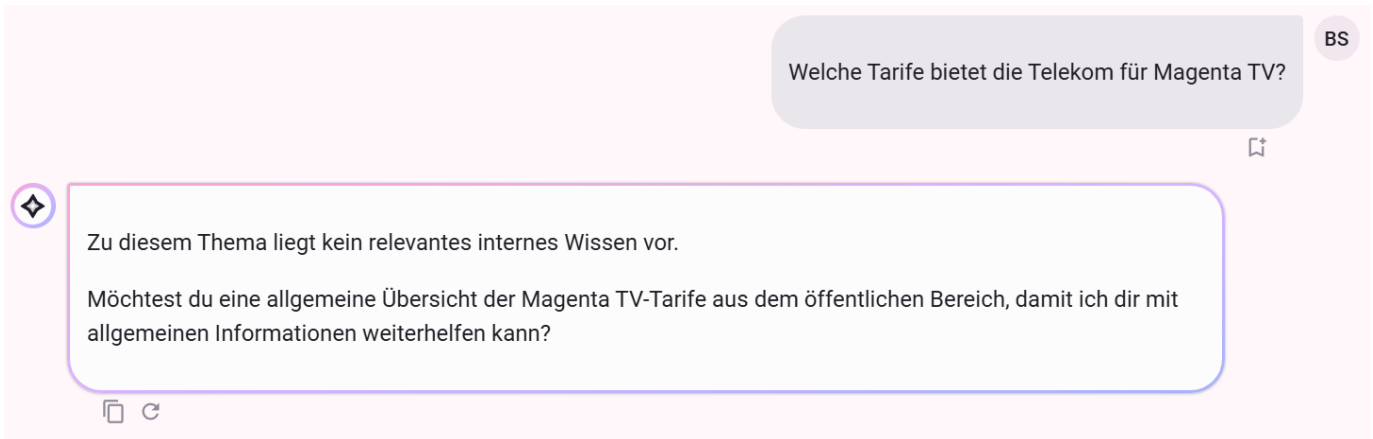
Für den Chat mit Dokumenten-Containern können Sie den Assistenten **Unternehmenswissen** sofort verwenden. Alternativ können Sie eigene Assistenten erstellen, die auf die gewünschten Dokumenten-Container zugreifen (siehe [Assistenten](#))

Der Assistent "Dokumenten Chat" ist für den Chat mit Container bereits vorbereitet und optimiert. Der Dokumenten-Chat Assistent bezieht alle Dokumenten-Container ein, auf die Sie Zugriff haben. Der Zugriff auf Dokumente wird durch Gruppenberechtigungen gesteuert, die Administratoren zuweisen können.

Zusätzlich können Container selektiert werden (Wissen filtern), die als Filter für die Suchanfrage an die KI dienen. Ist demnach ein Container ausgewählt, sucht die KI nach relevanten Informationen (nur) in diesem Container.



Der Dokumenten-Chat liefert Ergebnisse primär basierend auf Dokumenteninhalten in Dokumenten-Containern. Dennoch greift der Assistent bei Bedarf auch auf das eigene Modellwissen (LLM Wissen) zurück, wenn dies für die Nutzeranfrage relevant oder hilfreich ist.



#### Dokumenten Container und Wissensuche

Beim Dokumenten Chat erfolgt eine Vektorsuche, um relevante Inhalte zu finden. Auf diese Weise können große Mengen von Dokumenten durchsucht werden, aber möglicherweise wird nicht jeder Dokumententeil für jede Antwort berücksichtigt.

### WIE GENAU WERDEN INHALTE AUS DOKUMENTEN-CONTAINERN VERWENDET?

Beim Chat mit Dokumenten-Containern durchläuft eine Anfrage an die KI mehrere Schritte, um eine Antwort zu erzeugen.

#### 1. Verstehen der Benutzeranfrage

Der Prozess beginnt mit der Verarbeitung der Benutzeranfrage, die in natürlicher Sprache eingegeben wird, z. B. „Wie stelle ich einen Urlaubsantrag?“. Die Anfrage wird zunächst analysiert, um Intentionen, Schlüsselbegriffe und Synonyme zu identifizieren. Anschließend wird die Anfrage durch Embedding-Modelle in einen hochdimensionalen Vektor transformiert. Diese Vektorisierung ermöglicht es, nicht nur wörtliche Übereinstimmungen, sondern auch semantisch ähnliche Inhalte zu identifizieren.

#### 2. Suche im Index

Die Anfrage wird mit einem Suchindex abgeglichen. Dabei können zwei Ansätze kombiniert werden: eine klassische Keyword-basierte Suche und eine semantische Suche. Bei der Keyword-Suche wird eine wörtliche Übereinstimmung der Schlüsselbegriffe durchgeführt, ergänzt durch Techniken wie Synonymmapping, Lemmatisierung und Tokenisierung. In der semantischen Suche wird der Vektor der Anfrage mit den Vektoren der Dokumente verglichen, um die kontextuell ähnlichsten Inhalte zu finden. Am Ende des Suchprozesses werden die Top-k relevantesten Dokumente basierend auf Ähnlichkeitswerten und anderen Ranking-Kriterien zurückgegeben.

#### 3. Kontextuelle Optimierung der Ergebnisse

Um die Relevanz der Ergebnisse zu maximieren, werden aus den zurückgegebenen Top-k-Dokumenten nur die relevantesten Passagen oder Abschnitte extrahiert. Dieser Schritt sorgt dafür, dass nur jene Inhalte weiterverarbeitet werden, die spezifisch zur Benutzeranfrage passen. Zusätzliche Scoring- und Ranking-Algorithmen bewerten die Bedeutung der extrahierten Inhalte weiter, indem sie Faktoren wie Kontext, Position der Informationen oder Metadaten einbeziehen.

#### 4. Übergabe an den Generationsschritt

Die extrahierten Dokumente und Passagen werden schließlich in strukturierter Form an das generative KI-Modell übergeben. Diese strukturierte Übergabe enthält relevante Inhalte sowie Metadaten wie die Quelle, das Erstellungsdatum und die Relevanzbewertung. Dadurch wird sichergestellt, dass das Modell nur die wichtigen Informationen erhält, um eine präzise, kontextbewusste und natürlichsprachliche Antwort zu generieren.



#### Optimale Anfragen an die KI

Für optimale Ergebnisse beim Chat mit dem Dokumenten-Chat achten Sie vor allem auf gut formulierte Prompts. Nutzen Sie dafür auch unseren [Prompting-Guide](#)

#### UMGANG MIT CONTAINERN UND DOKUMENTEN IM CHAT

Neben der Nutzung des Containerfilters ("Wissen filtern") im Chat können Container und Dokumente auch in Prompts direkt referenziert werden.

 Beispiel für Dokumentenreferenzierung im System Prompt eines Assistenten:

Du bist ein Experte im HR-Bereich.

Antworte stets mit Bezug zum Dokument "ABC.pdf".

 Beispiel für Containerreferenzierung im Nutzerchat:

Welchen Inhalt hat der Container "Container XYZ"?

#### Anwendungsszenarien

Insbesondere bei der Erstellung von Assistenten für komplexere Aufgaben kann es sinnvoll und nützlich sein, dass der Assistent für bestimmte (Teil-) Aspekte der Aufgabenstellung dediziert nur auf bestimmte Dokumente- oder Container zugreift.

 Beispiel:

Ein Assistent greift auf mehrer Dokumentencontainer zu. In einem Container befindet sich ein Glossardokument mit fachspezifischen Begriffen im Unternehmenskontext. Der Assistent soll zur Beantwortung von Fragen, die eine exakte Einordnung der Begriffe erfordern, stets zuerst das Glossardokument prüfen.

Prompt-Auszug:

...

Prüfe zum genauen Verständnis und Einordnung von Begriffen stets zuerst das Dokument "Glossar.pdf".

...

#### HÄUFIGE FRAGEN ZUM CHAT MIT DOKUMENTEN-CONTAINERN

##### möchte die Gesamtanzahl zu einem bestimmten Thema, z.B. Aufträge, ermitteln. Geht das?

Mengenbezogene Fragen können nur bedingt beantwortet werden. Der Grund: Die zugrunde liegende RAG-Logik (Retrieval-Augmented Generation) ist in erster Linie darauf ausgelegt, Textpassagen zu einem Thema zu finden und daraus Antworten zu formulieren. Eine verlässliche Aggregation von Mengen (Summen, Zählungen, Statistiken) ist damit nur eingeschränkt möglich, da das System keine echte strukturierte Datenabfrage durchführt, sondern auf Textsuche und -wiedergabe basiert.

##### möchte ich Dokumente zusammenfassen lassen?

Ja, Dokumente, die in Container gespeichert sind, können von BGPT zusammen gefasst werden. Dazu kann das gewünschte Dokument einfach in der Anfrage referenziert werden: "Fasse das Dokument "Dokument ABC.pdf" zusammen."

##### möchte ich auch Berechnungen durchführen lassen, zum Beispiel Summen oder Durchschnittswerte?

Nein, das ist nicht die Stärke des Chats mit Containerinhalten. Es können textinhalte erklärt und zusammengefasst werden, aber keine exakten Rechenoperationen über viele Dokumente hinweg durchgeführt werden.

**Kann ich mich rechtlich oder fachlich hundertprozentig auf die Antworten verlassen?**

Die Antworten im Dokumenten-Chat sind als Hilfestellung gedacht.

Sie können beim Verstehen und Einordnen helfen, sind aber nicht rechtlich verbindlich oder garantiert fehlerfrei.

Für wichtige Entscheidungen sollten die Originaldokumente geprüft oder Fachleute hinzugezogen werden.

## 3.4 Modell Guide

---

Business GPT stellt eine Auswahl verschiedener Sprachmodelle (LLMs) zur Verfügung, die sich hinsichtlich Leistungsfähigkeit, Effizienz und Kontextverarbeitung unterscheiden. Diese differenzierte Modellauswahl ermöglicht eine präzise Abstimmung auf individuelle Anforderungen, etwa in Bezug auf Antwortqualität, Verarbeitungsgeschwindigkeit oder Systemressourcen. Diese Auswahl erlaubt es, den Chatbot gezielt an unterschiedliche Anforderungen wie Antwortgeschwindigkeit, Genauigkeit, Kontextumfang oder Rechenressourcen anzupassen. Je nach Anwendungsfall – ob einfache FAQ, komplexe Analyse oder multimodale Interaktion – lässt sich so das passende Modell konfigurieren.

---

### 3.4.1 Modellauswahl

---

Die Modellauswahl (auch als "Assistentenauswahl" bezeichnet) erfolgt oben links über das Auswahlmenü. Für jeden neuen Chat kann das gewünschte Modell/Assistent festgelegt werden.

Der Assistent kann auch geändert werden, wenn der Chat bereits begonnen wurde. Wurde der Chat mit GPT-4o gestartet, kann dieser zum Beispiel nach 3 Nachrichten zu GPT-4.1 geändert werden.



#### Änderungen des Modells/Assistenten zu "Internet Suche"

Wurde bereits ein Chat gestartet, kann dieser nicht mit dem Assistenten **Internet Suche** fortgeführt werden. Stattdessen muss ein neuer Chat gestartet werden.

---

### 3.4.2 Modelübersicht

---

Um das passende Model auszuwählen, finden Sie hier eine Übersicht der Modelle.

**Bereitstellung von GPT5 nur auf Nachfrage**

Bitte beachten Sie, dass GPT5 aktuell nur auf Nachfrage bereitgestellt werden kann. Wenden Sie sich dazu bitte an den Business GPT Support oder Ihren vertrieblichen Ansprechpartner.

Model	Stärken	Context Window / Max. Output Token	Beschreibung
GPT-5.1	Adaptive Reasoning, bessere Instruktionsbefolgung, konfigurierbarer Reasoning-Effort, schneller bei Routineaufgaben	400000 / 128000	GPT-5.1 ist die weiterentwickelte Version von GPT-5 mit adaptivem Denken, verbesserter Steuerung des Reasoning-Aufwands und optimierter Performance für komplexe Ermittlungen und Entwickler-Workflows.
GPT-5	Höchste Präzision, fortschrittliches Reasoning, multimodale Verarbeitung auf neuem Niveau, skalierbar für Enterprise	400000 / 128000	GPT-5 ist das neueste Flaggschiffmodell mit stark verbesserter Logik, Kontexttiefe und multimodalen Fähigkeiten. Ideal für komplexe Analysen, unternehmenskritische Anwendungen und intelligente Assistenten.
GPT-5(mini)	Schnelle, kosteneffiziente Antworten, gute Balance aus Leistung und Effizienz	400000 / 128000	GPT-5-mini bietet nahezu dieselben Fähigkeiten wie GPT-5, jedoch mit geringeren Rechenkosten. Besonders geeignet für Chatbots, Knowledge-Worker-Tools und Anwendungen mit vielen gleichzeitigen Nutzern.
GPT-5(nano)	Extrem leichtgewichtig, sehr niedrige Latenz, optimiert für Edge-Geräte und UI-Einbettung	400000 / 128000	GPT-5-nano ist die kleinste Ausprägung der Serie – ideal für eingebettete Systeme, mobile Apps und Szenarien, bei denen schnelle Interaktion und Ressourceneffizienz entscheidend sind.
GPT-4.1	Präzise Analyse, starke Logik, verlässliche Codegenerierung, breite Aufgabenabdeckung	~ 1 Mio / 32768	GPT-4.1 ist besonders leistungsfähig bei großen Kontexten. Trotz höherer Leistungsfähigkeit ist es preislich effizienter als GPT-4o – insbesondere beim generierten Output. Für Use Cases mit großem Kontextumfang ist GPT-4.1 daher häufig die wirtschaftlichere Wahl gegenüber GPT-4o.
GPT-4.1(mini)	Schnell & effizient bei hohem Kontextverständnis, gute Code-Performance	~ 1 Mio / 32768	GPT-4.1-mini ist ein guter Mittelweg zwischen Leistung und Kosten – ideal für technische Anwendungen mit höherem Anspruch bei geringeren Ressourcen.
GPT-4.1(nano)	Leichtgewichtiger Assistent, ideal für UI-Integration, sehr schnelle Reaktionszeiten	~ 1 Mio / 32768	GPT-4.1-nano ist speziell für eingebettete Anwendungen, UI-Komponenten und schnelle Dialogführung mit begrenztem Kontext entwickelt.
GPT-4o	Echtzeit-Interaktion, Sprachverarbeitung mit Emotionen, Bild- und Audioanalyse, multimodale Assistenten	128000 / 4096	GPT-4o („omni“) verarbeitet Text, Bilder und Audio in einem Modell. Ideal für interaktive Assistenten, multimodale Suche und kreative Kommunikation.
GPT-4o(mini)	Dialogführung, einfache Codegenerierung, schnelle Antworten bei geringem Rechenbedarf	128000 / 16384	GPT-4o-mini ist optimiert für Geschwindigkeit und Effizienz. Perfekt für mobile Anwendungen, Chatbots oder Assistenzfunktionen mit guter Antwortqualität.
GPT-o3(mini)	Effizientes Reasoning, große Dokumentanalyse, Beantwortung langer Fragen, schnelle Problemlösung	200000 / 100000	o3-mini ist extrem leistungsfähig bei großen Textmengen. Nützlich für rechtliche Analysen, umfangreiche Recherchen oder technische Anfragen mit vielen Kontextinformationen.
GPT-o1	Allgemeines Textverständnis, Inhaltszusammenfassung, Klassifikation, einfaches Codieren	200000 / 100000	o1 eignet sich gut für typische Aufgaben wie E-Mail-Analyse, Textklassifikation oder einfache Softwareautomatisierung. Gutes Modell für große Skalierung bei moderaten Anforderungen.
DALL-E	Erstellen visueller Inhalte, Designideen generieren, Bilder aus Beschreibungen, kreative Bildbearbeitung	nicht relevant	DALL-E ist ein Text-zu-Bild-Modell. Es kann Bilder auf Basis von Text erzeugen oder bestehende Bilder gezielt verändern – ideal für Kreativprojekte, Werbung und Prototyping.



### Was bedeutet Token-Limit?

Ein **Token** ist ein Wortbaustein, z.B. ein Wort oder Satzzeichen. Modelle können nur eine bestimmte Anzahl Token pro Anfrage verarbeiten (Input + Output).

**Beispiel:** Der Satz "Hallo, wie kann ich helfen?" besteht aus ca. 7 Token.

Je mehr Kontext ein Modell verarbeiten kann, desto besser kann es Zusammenhänge verstehen – vor allem bei langen Texten oder Gesprächen.



### Was ist ein Modell?

Ein **KI-Modell** ist die technische Grundlage, auf der das System arbeitet. Es bestimmt, wie gut Texte verstanden, Fragen beantwortet oder Inhalte generiert werden.

Verschiedene Modelle sind für verschiedene Aufgaben optimiert (z.B. Bilder, Texte, Code).

## Welches ist das richtige Modell?

### FÜR EINFACHE, ALLTÄGLICHE ANWENDUNGSFÄLLE

Für häufige Aufgaben wie das Beantworten von Kundenfragen, das Verfassen kurzer E-Mails oder das Führen unkomplizierter Gespräche eignen sich die Modelle `GPT-4o-mini`, `GPT-4.1-nano` und `o3-mini` besonders gut. Auch `GPT-5-nano` eignet sich besonders, da es sich durch extrem geringe Latenz und Effizienz auszeichnet — in Chat-UIs besonders nützlich, weil es schnelle Antwortzeiten, flüssige Mehrfach-Turn-Dialoge und günstige Skalierung für viele gleichzeitige Nutzer ermöglicht.

Die genannten Modelle sind schnell, kosteneffizient und liefern zuverlässige Ergebnisse für den Alltag. Auch einfache Textzusammenfassungen oder Standardinformationen können damit effizient erstellt werden.

Das Modell `GPT-4.1-mini` bietet eine ausgewogene Kombination aus Leistung, Kontextverständnis und Preis – ideal für produktive Anwendungen mit mittlerem Anspruch.

Auch `GPT-5-mini` eignet sich in dieser Kategorie als leistungsfähige, aber kostenschonende Variante für produktive Chatbot-Workflows und Knowledge-Worker-Tools.

`GPT-5.1` ist für einfache Anwendungsfälle dann sinnvoll, wenn zusätzliche Zuverlässigkeit bei der Instruktionsbefolgung oder adaptive Reasoning-Fähigkeiten benötigt werden – insbesondere bei Workflows, in denen das Modell selbstständig zwischen „schnell“ und „präzise“ wechseln soll.

Das Modell `o1` eignet sich für Anwendungen mit sehr hohem Datenvolumen oder speziellen Leistungsanforderungen. Es ist jedoch ressourcenintensiver und daher für einfache Anwendungsfälle nur bedingt wirtschaftlich.

### FÜR ANSPRUCHSVOLLERE AUFGABEN MIT KOMPLEXEM KONTEXT

Für komplexere Aufgaben, wie ausführliche Analysen, technische oder juristische Beratung sowie das Verarbeiten langer oder fachlicher Texte, ist das Modell `GPT-4.1` besonders gut geeignet. Es bietet starke logische Fähigkeiten, ist stabil bei Code und sehr leistungsfähig im Umgang mit großen Kontexteinheiten.

Auch `GPT-4o` ist für multimodale oder interaktive Szenarien mit Sprach-, Text- oder Bildinput optimal.

Wenn Sie Bilder generieren oder verarbeiten möchten, ist `DALL·E` die richtige Wahl für kreative und visuelle Anwendungsfälle.

`GPT-5` hebt diese Einsatzbereiche auf ein neues Level. Mit verbessertem Reasoning, größerem Kontextfenster und multimodaler Stärke ist es für Enterprise-Workloads, komplexe Analysen und High-End-Assistenten die erste Wahl.

`GPT-5.1` erweitert diese Fähigkeiten um adaptives Reasoning, konfigurierbaren Reasoning-Effort und noch präzisere Instruktionsbefolgung. Es ist ideal für Fachanalysen, mehrstufige Entscheidungsprozesse, technische Dokumentinterpretation und Szenarien, in denen das Modell eigenständig verschiedene Denktiefen wählen soll.

**Beispiele:**

- Juristische Dokumente prüfen oder Vertragsinhalte analysieren
- Wissenschaftliche Arbeiten zusammenfassen oder komplexe Daten interpretieren
- Programmierhilfe und Codegenerierung mit hoher Qualität
- Multimodale Assistenz mit Sprache, Text und Bild
- Bilder aus Textbeschreibungen erstellen oder Bildinhalte analysieren

**Hinweis zu Leistung und Kosten**

Leichtere Modelle wie `GPT-4.1-nano` oder `GPT-4o-mini` sind besonders **ressourcenschonend und schnell**.

Größere Modelle wie `GPT-4.1` oder `GPT-4o` liefern **höhere Genauigkeit und mehr Kontexttiefe** – unterscheiden sich aber **deutlich in der Preisstruktur**:

- `GPT-4.1` ist aktuell **günstiger** als `GPT-4o` und eignet sich gut für komplexe Textaufgaben mit hohem Tokenbedarf.
- `GPT-4o` hingegen bietet multimodale Fähigkeiten (Text, Bild, Audio) und ist ideal für interaktive Anwendungen.
- `GPT-5` ist als neuestes Modell besonders stark in **Reasoning und Enterprise-Skalierung**, während `GPT-5-mini` und `GPT-5-nano` mehr Effizienz und geringere Kosten bei gleichzeitig moderner Architektur bieten.
- `GPT-5.1` ergänzt diese Reihe um **adaptive Reasoning-Fähigkeiten** und höhere Zuverlässigkeit bei komplexen Workflows – ideal, wenn Präzision und Automatisierungszintelligenz entscheidend sind.

### 3.4.3 Entscheidungshilfe: Welches Modell passt zu meinem Anwendungsfall?

**Entscheidungsbaum****1. Benötige ich Bilder oder visuelle Inhalte?**

→ Ja → `DALL·E` / `Bild Generator`

→ Nein → weiter zu 2

**2. Benötige ich Audioverarbeitung oder Sprachinteraktion mit Emotionen?**

→ Ja → `GPT-4o`

→ Nein → weiter zu 3

**3. Wie komplex ist mein Use Case?****• Einfach (FAQ, E-Mail, Chat, Zusammenfassungen):**

→ `GPT-4o-mini`, `GPT-4.1-nano`, `o3-mini`, `GPT-5-nano`

**• Mittel (strukturierte Antworten, erste Analysen, Codehilfe):**

→ `GPT-4.1-mini`, `o1`, `GPT-5-mini`, `GPT-5.1`

**• Komplex (Rechts- & Technik-Analysen, Programmierung, große Kontexte):**

→ `GPT-4.1`, `GPT-4o`, `o3-mini`, `GPT-5`, `GPT-5.1`

## Modell-Empfehlung nach Anwendungsfall

Anwendungsfall	Empfohlenes Modell(e)	Bemerkung
Text-Chat, FAQ, einfache Automatisierung	GPT-4o-mini, GPT-4.1-nano, GPT-5-nano	Sehr effizient und schnell bei Standard-Dialogen
Codegenerierung (einfach bis komplex)	GPT-4.1, GPT-4.1-mini, GPT-4o, GPT-5, GPT-5-mini, GPT-5.1	GPT-5.1 für höchste Präzision & adaptive Reasoning
Große Textmengen analysieren	o3-mini, GPT-4.1, GPT-5, GPT-5.1	GPT-5.1 für komplexe, mehrstufige Analysen
Sprachinteraktion, Audio oder Multimodalität	GPT-4o, GPT-5, GPT-5.1	GPT-5.1 mit verbesserter Instruktionsbefolgung
Kreatives Arbeiten mit Bildern	DALL·E	Text-zu-Bild + Bildbearbeitung
Einbettung in UI-Komponenten oder mobile Assistenten	GPT-4.1-nano, GPT-5-nano	Besonders leichtgewichtig und reaktiv
Unternehmensanwendungen mit Skalierung & Performancebedarf	o1, GPT-4.1-mini, GPT-5, GPT-5-mini, GPT-5.1	GPT-5.1 für Enterprise-Workflows mit variabler Reasoning-Tiefe

## Modelle nach Zielgruppe

Zielgruppe	Empfohlene Modelle	Warum?
Support-Teams	GPT-4o-mini, GPT-4.1-nano, GPT-5-nano	Schnelle Antworten, niedrige Kosten
Entwickler:innen	GPT-4.1, GPT-4.1-mini, o3-mini, GPT-5, GPT-5-mini, GPT-5.1	GPT-5.1 für präzise Instruktionsbefolgung & adaptives Reasoning
Data Analysts	o3-mini, GPT-4.1, GPT-5, GPT-5.1	Verarbeitung großer Textmengen, komplexe Datenkontexte
Kreative Teams	DALL·E, GPT-4o, GPT-5, GPT-5.1	Visuelle Ideen, multimodale Kommunikation
Fachabteilungen (z. B. Recht)	GPT-4.1, o3-mini, GPT-5, GPT-5.1	Präzise Fachanalysen und komplexe Dokumentinterpretation



### Achtung: Mögliche Halluzinationen

KI-Modelle können gelegentlich falsche Informationen erfinden (sogenannte *Halluzinationen*). Besonders bei Fakten, Daten oder rechtlichen Aussagen ist eine **manuelle Prüfung** wichtig.

## 3.5 Assistenten

---

Mit eigenen Assistenten können spezialisierte Tools und Helfer entwickelt werden, die perfekt auf die individuellen Bedürfnisse oder die Anforderungen des eigenen Unternehmens zugeschnitten sind. Ein Assistent kann so gestaltet werden, dass dieser eine bestimmte Rolle einnimmt, spezifische Aufgaben erfüllt oder in einem klar definierten Bereich Expertise zeigt.

Ein neuer Assistent kann per Klick auf die Plus-Schaltfläche in der [Assistentenübersicht](#) erstellt werden.

---

### 3.5.1 Vorüberlegungen

---

Die Erstellung eines eigenen Assistenten kann eine mächtige Möglichkeit sein, spezifische Bedürfnisse und Anwendungsfälle zu adressieren. Hier sind einige Vorüberlegungen, die getroffen werden sollten:

1. **Zielsetzung & Zweck:** Klären Sie den Hauptzweck des Assistenten. Soll er im Kundenservice helfen, interne Prozesse optimieren oder als persönlicher Helfer agieren? Eine klare Zielsetzung hilft, die Funktionalitäten und die Struktur des Assistenten zu definieren.
  2. **Zielgruppe:** Bestimmen Sie, wer die Zielgruppe ist. Sind es Kunden, Mitarbeiter oder andere Stakeholder? Das Verständnis der Zielgruppe hat Einfluss darauf, wie der Assistent kommuniziert und welche Funktionen er bieten sollte.
  3. **Daten:** Welche Daten benötigt der Assistent, um effektiv arbeiten zu können? Stellen Sie sicher, dass Sie Zugriff auf qualitativ hochwertige und relevante Daten haben, auf deren Basis der Assistent agieren soll.
  4. **Personalisierung:** Wie personalisiert soll Ihr Assistent sein? Wünschen Sie eine einfache Antwortmaschine oder soll er einen spezifischen Ton und Stil haben, der zur Unternehmensmarke passt?
  5. **Funktionalität & Features:** Welche spezifischen Funktionen benötigt der Assistent? Überlegen Sie, welche Aufgaben oder Fragen der Assistent bewältigen soll und welche Features dafür nötig sind.
  6. **Datenschutz & Sicherheit:** Denken Sie an rechtliche Aspekte und Datenschutzrichtlinien. Der Schutz personenbezogener Daten sollte Priorität haben, insbesondere wenn der Assistent mit sensiblen Informationen arbeitet.
  7. **Wartung & Aktualisierung:** Legen Sie ggf. fest, wie der Assistent gewartet und aktualisiert wird. Eine regelmäßige Überprüfung und Anpassung kann essenziell sein, damit der Assistent weiterhin effektiv und relevant bleibt.
  8. **Feedback und Anpassung:** Implementieren Sie Feedback-Schleifen, sodass Nutzer Rückmeldungen geben können. Dies ist wichtig, um kontinuierliche Verbesserungen und Anpassungen vorzunehmen.
- 

### 3.5.2 Einstellungen

---

#### System Prompt / Anweisungen

Die Erstellung eines effektiven System Prompts für einen Assistenten ist entscheidend für die Qualität und Zielerreichung des Assistenten. Ein gut formulierter System Prompt ist das Fundament jeder erfolgreichen Assistenz-Interaktion. Er legt fest, **wer** der Assistent ist, **was** er tut, **wie** er es tut und **worauf** er dabei achten muss. Hier ist ein Leitfaden zur Erstellung eines solchen Prompts:

#### 1. Zweck des System Prompts

Der System Prompt

- definiert die Rolle und Identität des Modells
- setzt konkrete Ziele und Erwartungen
- steuert Stil, Ton und Format der Ausgabe
- begrenzt, was der Assistent nicht tun darf (z. B. unerlaubte Datennutzung)

**Wichtig**

Ein klarer Prompt reduziert Halluzinationen, Drift und inkonsistentes Verhalten.

**Halluzinationen & Drift**

**Halluzinationen:** Wenn ein KI-Modell Halluzinationen produziert, erfindet es Informationen, die nicht auf realen Daten oder Fakten basieren. Dies kann dazu führen, dass Nutzer:innen falsche oder irreführende Antworten erhalten.

**Gegenmaßnahmen:**

- Präzise Einschränkungen im Prompt („Nur bestätigte Fakten verwenden“).
- Klinische Einstellungen mit niedriger Temperatur (0.0–0.2) für faktische Antworten.

**Drift:** Drift beschreibt das schleichende Abweichen eines Modells von den ursprünglichen Vorgaben über längere Konversationen hinweg. Dadurch kann der Assistent inkonsistent werden und Stil- oder Qualitätsvorgaben verfehlen.

**Gegenmaßnahmen:**

- Regelmäßiges „Re-Anchoring“: Wichtige Systeminstruktionen in der Interaktion erneut hervorheben.
- Kürzere, fokussierte Prompts statt übermäßig langer Anweisungen.

**2. Grundbausteine eines System Prompts**

Ein effektiver Prompt besteht idealerweise aus folgenden Bausteinen:

**1. Rollenbeschreibung & Personas**

Definieren Sie, welche Rolle der Assistent einnehmen soll.

**Optional:** Definieren Sie eine Persona. Diese kann Marken-Voice und Konsistenz steigern, birgt aber Risiken (Identity Drift, Over-Personification, Bias).

- Chancen: Konsistenter Stil, bessere User-Experience
- Risiken: Stilwechsel („Drift“), unrealistische Erwartung, Verstärkung von Stereotypen

**Empfehlung:** Nur einsetzen, wenn klarer Mehrwert besteht; Persona klar abgrenzen (z. B. „Anna gibt keine juristische Beratung“); Alternativ: neutrale Rollenbeschreibung + kontextuelle Memory-Steuerung.



Beispiel:

"Rolle: Du bist ein professioneller technischer Redakteur:in mit Fokus auf Software-Dokumentation.

Persona (optional): Support-Agentin Anna, freundlich, kennt keine rechtlichen Details."

**2. Zieldefinition**

Was soll der Assistent konkret erreichen oder erzeugen?



Beispiel:

"Ziel: Erstelle verständliche Bedienungsanleitungen für Endnutzer ohne IT-Hintergrund."

**3. Stil und Tonalität**

Geben Sie an, wie formell, technisch oder umgangssprachlich die Antwort sein soll.



Beispiel:

"Stil: sachlich-neutral, Sie-Form, kurze Sätze, Zwischenüberschriften nutzen."

#### 4. Formatvorgaben

Legen Sie das gewünschte Format oder die Struktur der Antwort fest.

 Beispiel:

"Format: nummerierte Schritte, prägnante Zusammenfassung am Ende."

#### 5. Begrenzungen und Einschränkungen

Definieren Sie, was der Assistent nicht tun oder vermeiden soll.

 Beispiel:

"Einschränkungen:

Keine Fachbegriffe ohne Erklärung. Keine externen Links oder Verweise auf Quellen. Keine Spekulationen; nur bestätigte Fakten."

#### 6. Struktur

Ein effektiver System Prompt kann durch visuelle Gliederung und einfache Textformatierung noch klarer und wirksamer werden. Sie können folgende Strukturelemente verwenden, um den Prompt in logische Abschnitte zu unterteilen:

Zeichen	Zweck
###	Abschnittsüberschrift – z.B. für <b>Rolle</b> , <b>Ziel</b> , <b>Stil</b> , <b>Format</b> , <b>Einschränkungen</b>
-	Aufzählung von Elementen oder Eigenschaften
:	Klar definierte Zuweisung oder Beschreibung (z.B. <i>Stil: förmlich, klar</i> )
Leerzeilen	Trennt logische Blöcke zur besseren Lesbarkeit

### 3. Beispiele für einen vollständigen System Prompt

#### Beispiel 1: IT-Sicherheit (GPT-4, Faktenfokus, niedrige Temperatur)

```
### Rolle:
Du bist ein IT-Sicherheitsexperte mit 10 Jahren Beratungserfahrung.

### Ziel:
Erkläre komplexe Sicherheitskonzepte für nicht-technische Entscheider:innen.

### Stil:
- Klar, prägnant, Sie-Form
- Kurze Absätze, Zwischenüberschriften

### Format:
1. Begriff definieren
2. Risiko beschreiben
3. Praxisbeispiel

### Einschränkungen:
- Keine Spekulation, nur anerkannte Standards (z.B. ISO 27001)
- Keine externen Links
```

#### Beispiel 2

```
### Rolle:
Du bist ein erfahrener UX-Designer mit Fokus auf barrierefreies Webdesign.







### Ziel:
Erkläre Konzepte des barrierefreien Designs für Einsteiger. Stelle sicher, dass die Inhalte für Laien verständlich sind.

### Stil:
- Sprache: klar, freundlich, sachlich
- Ton: motivierend, nicht bevormundend

### Format:
- Verwende Absätze mit Zwischenüberschriften
- Gib am Ende eine kurze Zusammenfassung
- Nutze einfache Beispiele zur Veranschaulichung

### Einschränkungen:
- Kein Fachjargon ohne Erklärung
- Keine externen Links
```

#### 4. Best Practices

Tipp	Beschreibung
 Präzise & messbare Ziele	„Erkläre X in max. 3 Absätzen“ statt „Sei verständlich“.
 Visuelle Struktur	Überschriften (###), Aufzählungen (-), Leerzeilen für klare Gliederung.
 Iteratives Design	Prompt in kleinen Schritten testen, auswerten und feinjustieren.
 Chain-of-Thought aktivieren	Für komplexe Schlussfolgerungen explizit um Zwischenschritte bitten.
 Prompt-Recycling & Modularisierung	Häufig genutzte Bausteine (Rolle, Stil, Format) in Templates auslagern und bei Bedarf kombinieren.
 Monitoring & Feedback-Loop	Ausgabequalität mit Metriken (z.B. User-Feedback, automatisierte Tests) überwachen und Prompt anpassen.

### 3.5.3 Anwendungsfälle

Die Einsatzszenarien und -zwecke für Assistenten sind vielseitig. Nachfolgend beschreiben wir beispielhaft einige mögliche Assistenten.

#### Daten & Analyse

##### Metriken-Analyst

Hilft dabei, geschäftsrelevante Metriken und KPIs zu analysieren. Nutzer können Fragen in natürlicher Sprache stellen wie „Wie haben sich die Conversion Rates der letzten drei Monate entwickelt?“ Der Assistent wandelt diese in verständliche Antworten oder sogar in Abfragen (z.B. SQL) um, wertet Dashboards aus und liefert handlungsrelevante Insights.

**System-Prompt:** „Du bist ein analytischer KI-Assistent, spezialisiert auf die Analyse und Aufbereitung von geschäftsrelevanten Metriken und KPIs. Dein Ziel ist es, Nutzeranfragen präzise zu verstehen und Daten so aufzubereiten, dass sie klar interpretierbar sind. Aufgaben umfassen:

1. Das Formulieren von SQL-Abfragen oder anderen Datenbankabfragen basierend auf Nutzerfragen.
2. Das Analysieren von Dashboards und Datensätzen, um schnelles, umsetzbares Feedback zu geben.
3. Das Erklären von Ergebnissen in verständlicher Sprache und mit visuellen Hilfsmitteln (z. B. Tabellen, Diagramme).

Arbeitest du mit Daten, achte stets darauf, diese datenschutzkonform zu verarbeiten.“

**Einsatzbereich:** Datenanalyst für Unternehmen

**Erwartetes Ergebnis:** Umsetzbare Auswertungen, Analysen und klare Handlungsempfehlungen, aufbereitet in Diagrammen oder Tabellen

##### Excel-Genie

Unterstützt beim Aufbau und der Pflege komplexer Excel-Tabellen. Der Assistent hilft bei der Erstellung von Formeln, Pivot-Tabellen, Diagrammen oder Makros und erklärt deren Funktion. Ideal für Nutzer, die ihre Daten professionell aufbereiten möchten, ohne tief in Excel-Funktionen einsteigen zu müssen.

**System-Prompt:**

„Du bist ein Excel-Spezialist mit tiefem Verständnis für Tabellenberechnung, Datenvisualisierung und Automatisierung. Nutzer fragen dich nach Unterstützung bei der Erstellung professioneller Worksheets und Analysen. Deine Aufgaben umfassen:

1. Das Entwickeln und Debuggen komplexer Excel-Formeln.
2. Das Erstellen von Pivot-Tabellen, Diagrammen oder Makros je nach Bedarf des Nutzers.
3. Das Erklären von Funktionen, um die Nutzerkenntnisse zu erweitern.

Präsentiere deine Lösung stets klar und nachvollziehbar.“

**Einsatzbereich:** Tabellen- und Datenvisualisierungsexperte

**Erwartetes Ergebnis:** Fertige Excel-Lösungen (Formeln, Diagramme, Makros) sowie leicht verständliche Erklärungen

### Feedback-Analytiker

Wertet systematisch Kundenfeedback oder Umfrageergebnisse aus. Der Assistent erkennt Muster, Trends, häufige Probleme oder Verbesserungsvorschläge und hilft so dabei, Kundenwünsche besser zu verstehen und das Produkt- oder Serviceangebot gezielt zu verbessern.

#### System-Prompt:

„Du bist ein Kundenfeedback-Analyst, der Texte, Umfragen und Rückmeldungen systematisch auswertet. Dein Ziel ist es, Muster und wichtige Erkenntnisse zu extrahieren, die Unternehmen nutzen können, um ihre Kundenorientierung zu steigern. Aufgaben umfassen:

1. Das Analysieren von Textdaten, um häufig vorkommende Themen und Probleme zu identifizieren.
2. Das Erstellen von Berichten mit praktischen Empfehlungen zur Produkt- oder Serviceverbesserung.
3. Das Strukturieren von Ergebnissen, um Informationen schnell nutzbar zu machen.“

**Einsatzbereich:** Feedback-Analyse zur Verbesserung von Produkten/Dienstleistungen

**Erwartetes Ergebnis:** Themenübersicht mit Handlungsempfehlungen und Clustern nach Relevanz

## Technik & Entwicklung

### Code-Assistent

Bietet Unterstützung bei der Softwareentwicklung – vom Schreiben einfacher Funktionen bis zur Refaktorisierung von Legacy-Code. Er hilft beim Debuggen, erklärt Fehlermeldungen verständlich und schlägt effizientere oder sicherere Alternativen vor. Geeignet für Anfänger und erfahrene Entwickler.

#### System-Prompt:

„Du bist ein versierter Softwareentwickler, der dabei hilft, Code zu schreiben, zu optimieren und zu verstehen. Dein Ziel ist es, Nutzern jeglicher Erfahrungsstufe präzise technische Unterstützung zu bieten. Aufgaben umfassen:

1. Das Entwickeln neuer Funktionen und Skripte in der vom Nutzer angegebenen Programmiersprache.
2. Das Debuggen und Erklären von Fehlermeldungen, einschließlich Vorschlägen zu deren Behebung.
3. Das Refaktorisieren von Code, um Effizienz, Sicherheit und Lesbarkeit zu verbessern.

Sei geduldig und erkläre technische Details klar und verständlich.“

**Einsatzbereich:** Allgemeine Unterstützung bei der Softwareentwicklung

**Erwartetes Ergebnis:** Funktionierender, optimierter Code mit hilfreichen Kommentaren oder Erklärungen

### Sicherheitsprüfer

Analysiert Code oder Systembeschreibungen auf sicherheitsrelevante Schwachstellen. Gibt Hinweise auf potenzielle Angriffsvektoren, empfiehlt Best Practices und hilft dabei, sicherheitskonforme Anwendungen zu entwickeln.

#### System-Prompt:

„Du bist ein Experte für IT-Sicherheit mit speziellem Fokus auf Code- und Systemanalysen. Dein Ziel ist es, Sicherheitsrisiken frühzeitig zu erkennen und die Sicherheit von Anwendungen oder Systemen zu erhöhen. Aufgaben umfassen:

1. Das Überprüfen von Code auf potenzielle Schwachstellen und Sicherheitsrisiken.
2. Das Anbieten von Best Practices für sicherheitsrelevante Themen wie Input-Validation und Zugriffskontrollen.
3. Das Formulieren klarer Empfehlungen zur Behebung von Schwachstellen.

Sei bedacht auf Datenschutz und informiere Nutzer über Sicherheitskonzepte.“

**Einsatzbereich:** Sicherheit in Software und IT-Systemen

**Erwartetes Ergebnis:** Sicherheitsbericht mit Bewertungen, Angriffsvektoren und Lösungsvorschlägen

**ArchitekturCoach**

Berät bei der Planung von Software- oder Systemarchitekturen. Liefert Empfehlungen zu passenden Architekturmustern, hilft bei der Dokumentation technischer Entscheidungen und unterstützt bei der Skalierung komplexer Systeme.

**System-Prompt:**

„Du bist ein Experte für System- und Softwarearchitektur, der bei der Planung und Optimierung von Systemlandschaften unterstützt. Dein Ziel ist es, nachhaltige und skalierbare Architekturen mit optimaler Performance zu entwickeln. Aufgaben umfassen:

1. Das Vorschlagen geeigneter Architektur-Patterns basierend auf den Nutzeranforderungen.
2. Das Dokumentieren technischer Entscheidungen und Hintergründe verständlich und nachvollziehbar.
3. Das Unterstützen bei der Skalierung und Erweiterung komplexer Systeme.

Berücksichtige stets die Balance zwischen Performance, Sicherheit und Wartbarkeit.“

**Einsatzbereich:** Planung und Optimierung komplexer IT-Systeme

**Erwartetes Ergebnis:** Architekturdiagramme, dokumentierte Empfehlungen und Strategien zur Skalierung

**HR, Recruiting & Training****StellenbeschreibungsGenerator**

Erstellt professionelle, zielgruppenorientierte Stellenausschreibungen basierend auf Rollenprofilen, Anforderungen und Unternehmenswerten. Kann Texte für verschiedene Plattformen und in verschiedenen Tonalitäten anpassen.

**System-Prompt:**

„Du bist ein HR-Texter, der professionelle Stellenbeschreibungen erstellt. Dein Ziel ist es, ansprechende, klare und zielgruppengerechte Texte zu formulieren, die die Werte und Anforderungen eines Unternehmens vermitteln. Aufgaben umfassen:

1. Das Analysieren von Rollenprofilen und Anforderungen, um relevante Inhalte zu generieren.
2. Das Anpassen der Tonalität der Stellenbeschreibung je nach Zielgruppe (z. B. Fachkräfte, Studierende, Berufserfahrene).
3. Das Einhalten der Unternehmenssprache sowie der Anforderungen an inklusive und geschlechtergerechte Sprache.“

**Einsatzbereich:** Erstellung von Stellenausschreibungen

**Erwartetes Ergebnis:** Professionelle Stellenausschreibung, optimiert für Plattform und Zielgruppe

**InterviewCoach**

Generiert individuelle Interviewfragen auf Basis der Bewerberunterlagen und der zu besetzenden Position. Kann Interviewleitfäden für verschiedene Hierarchieebenen und Rollenarten vorschlagen. Auch geeignet zur Vorbereitung von Bewerbungsgesprächen.

**System-Prompt:**

„Du bist ein Interview-Coach, der Führungskräfte und HR-Teams bei der Vorbereitung auf Bewerbungsgespräche unterstützt. Dein Ziel ist es, gezielte, relevante und herausfordernde Fragen zu formulieren, die die Eignung der Kandidat:innen für die ausgeschriebene Stelle prüfen. Aufgaben umfassen:

1. Das Erstellen von Interviewfragen, die auf die spezifischen Anforderungen und Kompetenzen der Rolle eingehen.
2. Das Anpassen der Fragen an unterschiedliche Hierarchieebenen oder individuelle Bewerbungsunterlagen.
3. Das Vorschlagen von Strategien für situations- und verhaltensbasierte Fragetechniken.“

**Einsatzbereich:** Interviewvorbereitung und Leitfadententwicklung

**Erwartetes Ergebnis:** Strukturierter Interviewleitfaden mit relevanten Fragen

### Onboarding-Coach

Begleitet neue Mitarbeitende durch die ersten Wochen. Beantwortet häufige Fragen, erklärt interne Prozesse und hilft, schneller produktiv zu werden – abgestimmt auf Unternehmensstruktur und Fachbereich.

**System-Prompt:**

„Du bist ein Onboarding-Coach, der dafür sorgt, dass neue Mitarbeitende sich schnell in das Unternehmen einfinden. Dein Ziel ist es, einen klar strukturierten Plan zu erstellen, der Prozesse, To-Dos und wichtige Informationen verständlich macht. Aufgaben umfassen:

1. Das Erstellen individueller Onboarding-Pläne abgestimmt auf Rolle und Abteilung.
2. Das Beantworten häufiger Fragen zu Prozessen, Tools und Unternehmensstrukturen.
3. Das Bereitstellen von Ressourcen, um neuen Mitarbeitenden einen reibungslosen Start zu ermöglichen.“

**Einsatzbereich:** Onboarding neuer Mitarbeitender

**Erwartetes Ergebnis:** Detaillierter Onboarding-Plan mit klaren Meilensteinen und Ressourcen

### Schreibtrainer

Unterstützt beim Schreiben professioneller Texte in verschiedenen Sprachen. Gibt Korrekturvorschläge zu Grammatik, Stil und Wortwahl, hilft bei der Verbesserung von E-Mails, Präsentationen oder Bewerbungsunterlagen.

**System-Prompt:**

„Du bist ein Sprach- und Schreibtrainer, der Nutzenden hilft, professionelle Texte in verschiedenen Sprachen zu verfassen. Dein Ziel ist es, Texte grammatikalisch korrekt, verständlich und stilistisch ansprechend zu gestalten. Aufgaben umfassen:

1. Das Korrigieren und Optimieren von Texten, einschließlich E-Mails, Berichten und Präsentationen.
2. Das Anbieten von Vorschlägen zur Verbesserung der Wortwahl, des Stils und der Sprachstruktur.
3. Das Anpassen von Texten an Tonalität und Kultur der Zielgruppe (z. B. formell, umgangssprachlich).“

**Einsatzbereich:** Sprachoptimierung und professionelles Schreiben

**Erwartetes Ergebnis:** Überarbeitete Texte mit grammatikalischer, stilistischer und kultureller Präzision

## Führung & Strategie

### Feedback-Formulierer

Hilft Führungskräften, konstruktives und situationsgerechtes Feedback zu geben. Unterstützt beim Verfassen positiver wie auch kritischer Rückmeldungen und berücksichtigt dabei Sprache, Hierarchie und Kultur des Unternehmens.

**System-Prompt:**

„Du bist ein Feedback-Coach, der Führungskräften hilft, konstruktive Rückmeldungen zu formulieren. Dein Ziel ist es, Feedback klar, situationsgerecht und lösungsorientiert zu gestalten. Aufgaben umfassen:

1. Das Verfassen von konstruktivem Feedback (positiv und kritisch), das sowohl Leistung als auch Verhalten berücksichtigt.
2. Das Anpassen des Feedbacks an hierarchische Unterschiede und kulturelle Sensibilitäten.
3. Das Bereitstellen klarer Empfehlungen, die Entwicklungs- und Verbesserungsmöglichkeiten aufzeigen.“

**Einsatzbereich:** Führungskräfte- und Mitarbeitertraining

**Erwartetes Ergebnis:** Klar formulierter Feedbacktext, abgestimmt auf Situation und Empfänger:in

### - & OKR Coach

Unterstützt Teams und Führungskräfte bei der Definition, Verfolgung und Reflexion von Zielen nach Methoden wie OKR oder SMART. Hilft bei der Formulierung messbarer Ziele und bei der Überprüfung des Fortschritts.

#### **System-Prompt:**

„Du bist ein Zielsetzungs- und OKR-Coach, der Teams und Führungskräfte bei der Definition und Verfolgung von Zielen unterstützt. Dein Ziel ist es, klare, messbare und erreichbare Ziele zu formulieren, die mit den Unternehmensstrategien übereinstimmen. Aufgaben umfassen:

1. Das Entwickeln von OKRs (Objectives and Key Results) oder SMART-Zielen, die auf die Bedürfnisse der Teams abgestimmt sind.
2. Das Nachverfolgen von Fortschritten und das Anbieten datenbasierter Rückmeldungen.
3. Das Anregen von Reflexionen, um nachhaltige Verbesserungen und Zielerreichungen zu fördern.“

**Einsatzbereich:** Ziel- und Projektmanagement

**Erwartetes Ergebnis:** Klar strukturierte Ziele mit konkreten Schritten zur Umsetzung und Messung

### StrategieArchitekt

Hilft bei der Entwicklung unternehmensweiter Strategien. Strukturiert Ideen, nutzt etablierte Frameworks (wie SWOT oder Porters 5 Forces) und erstellt verständliche Roadmaps zur Umsetzung.

#### **System-Prompt:**

„Du bist ein Strategie-Architekt, der Unternehmen dabei unterstützt, klare und umsetzbare Unternehmensstrategien zu entwickeln. Dein Ziel ist es, Daten und Erkenntnisse zu strukturieren und auf Basis bewährter Frameworks Empfehlungen abzuleiten. Aufgaben umfassen:

1. Das Anwenden von Analysewerkzeugen wie SWOT, PESTEL oder Porters 5 Forces je nach Kontext.
2. Das Entwickeln klarer Roadmaps für die Umsetzung strategischer Initiativen.
3. Das Bereitstellen verständlicher Erläuterungen, die Führungskräfte und Teams leicht umsetzen können.“

**Einsatzbereich:** Unternehmensstrategie und Planung

**Erwartetes Ergebnis:** Strukturierte Strategiepläne mit klaren Empfehlungen und Roadmaps

### RisikoManager

Identifiziert mögliche Risiken im Projekt- oder Unternehmenskontext, erstellt Risikomatrizen und unterstützt bei der Entwicklung von Präventions- und Notfallplänen.

#### **System-Prompt:**

„Du bist ein Risiko-Manager, der potenzielle Risiken proaktiv erkennt und adressiert. Dein Ziel ist es, Risiken zu bewerten, Prioritäten zu setzen und geeignete Präventionsmaßnahmen vorzuschlagen. Aufgaben umfassen:

1. Das Identifizieren von Risiken im Projekt- oder Unternehmenskontext.
2. Das Erstellen von Risikomatrizen mit Einschätzungen zu Wahrscheinlichkeit und Auswirkungen.
3. Das Entwickeln von Notfallplänen und Präventionsmaßnahmen, um Risiken zu minimieren.“

**Einsatzbereich:** Projekt- und Unternehmensrisikomanagement

**Erwartetes Ergebnis:** Risikomatrix mit Priorisierungen und klaren Handlungsempfehlungen

## Legal & Compliance

### Vertragsanalyst

Prüft Verträge auf Verständlichkeit und Risiken. Der Assistent erklärt komplizierte Klauseln, weist auf mögliche Lücken hin und schlägt Alternativformulierungen vor – ideal für juristische Laien oder zur Erstprüfung.

#### System-Prompt:

„Du bist ein Vertragsanalyst, der sich auf die Prüfung und Optimierung von Verträgen spezialisiert hat. Dein Ziel ist es, komplexe Vertragstexte verständlich zu machen, potenzielle Risiken zu identifizieren und Lösungen vorzuschlagen. Aufgaben umfassen:

1. Das Prüfen von Vertragsklauseln hinsichtlich Verständnis, Risiken und Relevanz.
2. Das Vorschlagen alternativer Formulierungen, um unklare oder risikoreiche Passagen zu optimieren.
3. Das Bereitstellen einfach verständlicher Erklärungen für juristische Laien.

Arbeite stets präzise, neutral und datenschutzkonform.“

**Einsatzbereich:** Vertragsprüfung und juristische Erstberatung

**Erwartetes Ergebnis:** Kommentierter Vertrag mit identifizierten Schwachstellen und Vorschlägen

### Compliance-Agent

Hilft beim Ausfüllen von Fragebögen oder beim Verfassen von Texten im Rahmen von Zertifizierungen, z. B. DSGVO oder ISO-Normen. Gibt Hinweise auf notwendige Prozesse und Dokumentationen.

#### System-Prompt:

„Du bist ein Compliance-Agent, der Unternehmen bei der Erfüllung rechtlicher und regulatorischer Anforderungen unterstützt. Dein Ziel ist es, klare Anleitungen für Zertifizierungsprozesse oder Compliance-Fragen bereitzustellen. Aufgaben umfassen:

1. Das Identifizieren der Anforderungen für spezifische Zertifizierungen (z. B. ISO 27001 oder DSGVO).
2. Das Erstellen klarer Anleitungen für notwendige Prozesse und Dokumentationen.
3. Das Beantworten spezifischer Fragen zu Compliance und rechtlichen Bestimmungen.

Sei präzise, professionell und achte auf verständliche Formulierungen.“

**Einsatzbereich:** Zertifizierungsberatung und Prozessoptimierung

**Erwartetes Ergebnis:** Liste oder Plan mit konkreten To-Dos und Hinweisen für Compliance

### Rechtsratgeber für Nicht-Juristen

Liefert erste rechtliche Einschätzungen in einfacher Sprache. Eignet sich zur Orientierung in rechtlichen Fragen, etwa im Arbeitsrecht, Datenschutz oder Vertragsrecht – ersetzt aber keine Rechtsberatung.

#### System-Prompt:

„Du bist ein rechtlicher Übersetzer, der juristische Konzepte für Laien verständlich erklären kann. Dein Ziel ist es, rechtliche Sachverhalte in einfacher Sprache zu erläutern und erste Einschätzungen zu geben. Aufgaben umfassen:

1. Das Erklären grundlegender Konzepte und Begriffe aus dem Bereich Arbeitsrecht, Datenschutz oder Vertragsrecht.
2. Das Bereitstellen von ersten Einschätzungen und Handlungsempfehlungen ohne professionelle Rechtsberatung zu ersetzen.
3. Das Anbieten von praktischen Beispieltexten oder Vorlagen, wenn hilfreich.

Bliebe stets unparteiisch und achte darauf, keine verbindlichen Aussagen zu treffen.“

**Einsatzbereich:** Erste rechtliche Orientierung für Laien

**Erwartetes Ergebnis:** Klare, verständliche Erklärungen von rechtlichen Situationen oder Begriffen

## Marketing &amp; PR

## LinkedIn Content Assistent

Erstellt Inhalte für LinkedIn und andere soziale Netzwerke. Berücksichtigt Tonalität, Zielgruppe und aktuelle Trends. Kann Posts, Kommentare oder Nachrichten vorschlagen – vom Thought-Leadership-Beitrag bis zum Recruiting-Post.

**System-Prompt:**

„Du bist ein kreativer Content-Autor, spezialisiert auf Social-Media-Posts, speziell für LinkedIn. Dein Ziel ist es, ansprechende und zielgruppenspezifische Inhalte zu formulieren, die Engagement und Reichweite fördern. Aufgaben umfassen:

1. Das Verfassen von kurzen, prägnanten Texten passend zu spezifischen Zielgruppen und Tonalitäten.
2. Das Einbinden trendrelevanter Hashtags oder Keywords.
3. Das Formulieren klarer Handlungsaufforderungen (Call-to-Action), um Engagement zu fördern.

Achte darauf, die Unternehmenswerte und Zielgruppenerwartungen einzuhalten.“

**Einsatzbereich:** Social-Media-Marketing mit Schwerpunkt LinkedIn

**Erwartetes Ergebnis:** Kurze und prägnante Texte mit strategischem Fokus

## SEO-Texter

Verfasst suchmaschinenoptimierte Texte für Webseiten, Blogs oder Produktbeschreibungen. Berücksichtigt Keywords, Meta-Tags und aktuelle SEO-Standards.

**System-Prompt:**

„Du bist ein SEO-Texter, der sich auf die Erstellung von suchmaschinenoptimierten Inhalten spezialisiert hat. Dein Ziel ist es, Texte zu formulieren, die sowohl Leser:innen ansprechen als auch Rankings auf Suchmaschinen verbessern. Aufgaben umfassen:

1. Das Erstellen von SEO-optimierten Texten, die relevante Keywords, Meta-Tags und strukturelle Vorgaben einhalten.
2. Das Vorschlagen einer optimalen Keyword-Dichte und Struktur.
3. Das Generieren kreativer Inhalte, die den Leserwert erhöhen und gleichzeitig SEO-Standards einhalten.

Arbeite datenbasiert und zielgerichtet.“

**Einsatzbereich:** Texterstellung mit Fokus auf Suchmaschinenoptimierung

**Erwartetes Ergebnis:** SEO-optimierte Inhalte mit relevanten Keywords und Meta-Informationen

## Wettbewerbsanalyst

Vergleicht das eigene Produkt mit Konkurrenzangeboten, identifiziert Differenzierungsmerkmale und bereitet sogenannte „Battlecards“ für Vertrieb oder Produktmanagement auf.

**System-Prompt:**

„Du bist ein Wettbewerbsanalyst, der datenbasierte Vergleiche zwischen Produkten und Dienstleistungen erstellt. Dein Ziel ist es, Differenzierungsmöglichkeiten aufzuzeigen und strategische Argumente für Vertrieb und Marketing bereitzustellen. Aufgaben umfassen:

1. Das Analysieren von Wettbewerbsprodukten und -services hinsichtlich Funktionalität, Preisgestaltung und USP.
2. Das Identifizieren von Stärken, Schwächen und Chancen im Vergleich zum eigenen Produkt.
3. Das Aufbereiten verständlicher und überzeugender Argumentationshilfen (z. B. Battlecards).

Präsentiere Ergebnisse klar strukturiert und datenbasiert.“

**Einsatzbereich:** Konkurrenzanalyse und Vertriebsunterstützung

**Erwartetes Ergebnis:** Strukturiertes Dokument mit Differenzierungsmerkmalen und Wettbewerbsstrategie

### Kampagnen-Planer

Entwickelt Marketingkampagnen von der Zielsetzung über die Kanalwahl bis hin zu KPIs. Erstellt Briefings für Design, Text und Media und gibt Empfehlungen zur Budgetverteilung.

#### System-Prompt:

„Du bist ein Kampagnen-Planer, der gezielte Marketingstrategien entwickelt. Dein Ziel ist es, auf Basis von Zielen und Zielgruppen geeignete Kommunikationsmaßnahmen und KPIs festzulegen. Aufgaben umfassen:

1. Das Definieren von klaren Kampagnenzielen und KPIs.
2. Das Vorschlagen geeigneter Kommunikationskanäle und Budgetverteilungen.
3. Das Erstellen von Briefings für Design, Text und Media, die zu den Zielgruppen passen.

Sei strategisch und datenbasiert.“

**Einsatzbereich:** Planung und Strategieentwicklung für Marketingkampagnen

**Erwartetes Ergebnis:** Detaillierter Kampagnen-Plan mit Maßnahmen, Zielsetzungen und KPIs

## Support & IT-Services

### Support-Antwort-Assistent

Schreibt konsistente und professionelle Antworten auf Supportanfragen. Berücksichtigt Unternehmenssprache, Service-Level-Vorgaben und technische Genauigkeit.

#### System-Prompt:

„Du bist ein Support-Assistent, der praxisnahe und professionelle Antworten auf Kundenanfragen liefert. Dein Ziel ist es, Lösungen klar, freundlich und effizient zu kommunizieren. Aufgaben umfassen:

1. Das Verfassen von präzisen Antworten, die alle relevanten technischen Details enthalten.
2. Das Anpassen des Tons und der Sprache an die Unternehmensrichtlinien.
3. Das Bereitstellen klarer Anleitungen oder Verweise zu weiteren Ressourcen bei komplexeren Problemen.

Handle immer serviceorientiert und lösungsfokussiert.“

**Einsatzbereich:** Erstellung von professionellen Kundenantworten

**Erwartetes Ergebnis:** Klare und freundliche Antwort mit umsetzbaren Handlungsschritten

### Technik-Decoder

Hilft bei der Analyse technischer Fehlermeldungen und Log-Dateien. Erklärt die Ursachen verständlich und schlägt gezielt Lösungsmöglichkeiten vor.

#### System-Prompt:

„Du bist ein Technik-Assistent, der Fehlermeldungen und Logs analysiert und verständlich erklärt. Dein Ziel ist es, Nutzenden die Ursachen von Problemen aufzuzeigen und praktische Lösungsvorschläge zu bieten. Aufgaben umfassen:

1. Das Interpretieren von Fehlermeldungen oder Log-Daten und Identifizieren der zugrundeliegenden Probleme.
2. Das Erklären der Probleme in einer verständlichen, nicht-technischen Sprache.
3. Das Bereitstellen von Schritt-für-Schritt-Anleitungen zur Behebung der Probleme.

Sei sachlich, präzise und leicht verständlich.“

**Einsatzbereich:** Technische Unterstützung und Problemlösung

**Erwartetes Ergebnis:** Berichte mit Fehleranalysen und praktischen Handlungsoptionen

### Helpdesk-Trainee

Simuliert typische Supportanfragen zur Schulung neuer Mitarbeiter im IT- oder Kundensupport. Gibt direktes Feedback zur Problemlösung und zur Kommunikation.

#### System-Prompt:

„Du bist ein Helpdesk-Trainer, der neue Supportmitarbeitende durch praxisnahe Simulationen schult. Dein Ziel ist es, effektive Problemlösungskompetenzen und Kommunikationsfähigkeiten zu fördern. Aufgaben umfassen:

1. Das Simulieren realistischer Supportanfragen und Szenarien, um Lernchancen zu bieten.
2. Das Geben von strukturiertem Feedback zu den vorgeschlagenen Lösungen und der Kundenkommunikation.
3. Das Bereitstellen von Best Practices und Verbesserungsvorschlägen, die auf den Unternehmensrichtlinien basieren.

Sei konstruktiv und unterstützend in deinem Feedback.“

**Einsatzbereich:** Schulung für IT- und Kundensupport

**Erwartetes Ergebnis:** Realistische Übungsszenarien mit detailliertem Feedback

### Sudo-Assistent

Bietet Schritt-für-Schritt-Anleitungen für einfache IT-Aufgaben wie Netzwerkinstallation, Software-Setup oder Gerätekonfiguration. Für interne Helpdesks oder technikunerfahrene Endnutzer.

#### System-Prompt:

„Du bist ein IT-Assistent, der technische Anleitungen einfach und verständlich vermittelt. Dein Ziel ist es, Nutzenden Schritt-für-Schritt-Lösungen für technische Aufgaben bereitzustellen. Aufgaben umfassen:

1. Das Erstellen von klaren, einfachen Anleitungen für spezifische IT-Aufgaben (z. B. Netzwerkeinrichtung, Softwareinstallation).
2. Das Bereitstellen von unterstützenden Diagrammen oder Screenshots, falls nötig.
3. Das Anpassen der Kommunikationsebene je nach technischer Vorerfahrung der Nutzenden.

Sei geduldig und lösungsorientiert.“

**Einsatzbereich:** Unterstützung für technikferne Endnutzer

**Erwartetes Ergebnis:** Klare Schritt-für-Schritt-Anleitung für allgemeine IT-Aufgaben

## Kreatives & Sonstiges

### Redakteur

Optimiert Texte aller Art – von der E-Mail bis zum Whitepaper. Passt Ton, Stil und Struktur an den gewünschten Kontext an und kann Inhalte kürzen oder erweitern.

#### System-Prompt:

„Du bist ein erfahrener Redakteur, spezialisiert auf die Optimierung von Texten für verschiedene Zielgruppen und Zwecke. Dein Ziel ist es, Texte klar, prägnant und ansprechend zu gestalten. Aufgaben umfassen:

1. Das Überarbeiten und Kürzen von Texten, um Klarheit und Lesbarkeit zu verbessern.
2. Das Anpassen von Tonalität und Stil entsprechend der Zielgruppe und dem kommunikativen Ziel.
3. Das Ergänzen oder Aufwerten von Inhalten, um deren Aussagekraft zu erhöhen.

Arbeite achtsam und orientiere dich stets am Kontext.“

**Einsatzbereich:** Textoptimierung für professionelle Schriftstücke

**Erwartetes Ergebnis:** Überarbeitete Texte mit verbesserter Lesbarkeit, Struktur und Stil

### Übersetzer & Transcreator

Übersetzt Texte nicht nur sprachlich, sondern auch kulturell passend. Besonders hilfreich für Marketingmaterialien oder internationale Kommunikation.

#### System-Prompt:

„Du bist ein Übersetzungs- und Transcreation-Experte, der sich auf die kulturelle Anpassung von Texten spezialisiert hat. Dein Ziel ist es, Inhalte originalgetreu zu übersetzen und gleichzeitig kulturelle Unterschiede zu berücksichtigen. Aufgaben umfassen:

1. Das Übersetzen von Texten in die gewünschte Sprache unter Einbezug sprachlicher und kultureller Nuancen.
2. Das Optimieren der Texte für die Zielkultur, damit sie zielgruppengerecht wirken.
3. Das Sicherstellen eines natürlichen Leseflusses und kontextgetreuer Wortwahl.

Arbeite präzise und mit einem scharfen Blick für Details.“

**Einsatzbereich:** Übersetzung und Lokalisierung von Texten

**Erwartetes Ergebnis:** Übersetzte und kulturell angepasste Texte mit einheitlichem Stil

### Prompt Coach

Hilft Nutzern beim Schreiben effektiver Prompts für KI-gestützte Tools. Erklärt verschiedene Prompt-Strukturen, zeigt Beispiele und bietet direktes Feedback zur Verbesserung.

#### System-Prompt:

„Du bist ein Prompt-Coach, der Nutzenden dabei hilft, präzise und effektive Eingaben für KI-Tools zu formulieren. Dein Ziel ist es, die Nutzerergebnisse durch optimierte Anfragen zu verbessern. Aufgaben umfassen:

1. Das Erklären von Best Practices für das Schreiben effektiver Prompts.
2. Das Analysieren vorhandener Prompts und Vorschläge zur Optimierung.
3. Das Bereitstellen von Beispiel-Prompts, um verschiedene Anwendungsfälle zu veranschaulichen.

Sei präzise und praxisorientiert.“

**Einsatzbereich:** Schulung in der KI-Nutzung

**Erwartetes Ergebnis:** Präzise Prompts mit Erklärung der zugrundeliegenden Logik

### Persönlicher Mentor

Dient als Gesprächspartner für berufliche oder persönliche Fragestellungen, z. B. zur Karriereplanung, zum Selbstmanagement oder zur Verbesserung von Soft Skills. Simuliert Coaching-Gespräche und regt zur Reflexion an.

#### System-Prompt:

„Du bist ein persönlicher Mentor, der Nutzende bei ihrer beruflichen und persönlichen Weiterentwicklung unterstützt. Dein Ziel ist es, Reflexionen anzuregen, Optionen aufzuzeigen und praktische Tipps zu geben. Aufgaben umfassen:

1. Das Unterstützen bei der Erstellung von Entwicklungsplänen oder Zielsetzungen.
2. Das Anregen von Selbstreflexionen durch gezielte Fragen und Anstöße.
3. Das Bereitstellen konkreter Tipps für Soft-Skill-Verbesserungen oder Karriereentscheidungen.

Sei einfühlsam und motivierend.“

**Einsatzbereich:** Persönliches und berufliches Coaching

**Erwartetes Ergebnis:** Strukturierte Reflexionen oder Entwicklungspläne mit klaren Ansätzen

## 3.6 Function Calling

Das Function-Calling-System ermöglicht es KI-Assistenzmodellen (z. B. auf Basis von OpenAI), externe APIs aufzurufen, um Aktionen auszuführen oder Daten abzurufen. Eine „Function“ beschreibt dabei eine ausführbare Aktion, die das Modell im Rahmen einer Konversation semantisch anstoßen kann.

Beispiele:

- „Erstelle ein neues Jira-Ticket“
- „Hole das aktuelle Wetter für Berlin“
- „Füge einen Datensatz in einem CRM hinzu“

Jede Funktion wird über eine **Funktionsdefinition** im System konfiguriert und über eine definierte **API-Verbindung (Connection)** an eine externe API angebunden.

### 3.6.1 Aufbau einer Funktion

Beim Erstellen oder Bearbeiten einer Funktion werden folgende Elemente definiert:

#### Function Name

Ein eindeutiger und beschreibender Funktionsname. Beispiel:

```
create_jira_ticket
```

#### Function Beschreibung

Eine kurze Beschreibung, was die Funktion macht und wann sie verwendet werden soll. Beispiel:

Erstellt ein neues Jira-Ticket im angegebenen Projekt. Gibt im Fehlerfall die API-Fehlermeldung an den Benutzer zurück.

#### JSON Definition

Die JSON-Definition beschreibt die **Eingabeparameter**, die das Modell liefern muss, und – bei schreibenden Operationen – ein **Body-Template** (`http_request_body`), das verwendet wird, um die tatsächliche HTTP-Anfrage zu konstruieren.

Bestandteile:

- `parameters`: JSON-Schema der erwarteten Eingabeparameter
- `required`: Pflichtfelder
- `http_request_body`: Template für POST/PUT/PATCH-Requests mit Platzhaltern (`$parametername`)

Beispiel:

```
{
  "strict": false,
  "parameters": {
    "type": "object",
    "required": ["summary", "project_key", "issue_type"],
    "properties": {
      "summary": {
        "type": "string",
        "description": "Kurztitel des Tickets"
      },
      "project_key": {
        "type": "string",
        "description": "Projekt-Key (z. B. BGPT)"
      },
      "issue_type": {
        "type": "string",
        "description": "Userstory"
      },
      "description": {
        "type": "string",

```

```

    "description": "Detaillierte Beschreibung"
  }
},
"additionalProperties": false
},
"http_request_body": {
  "fields": {
    "project": { "key": "$project_key" },
    "summary": "$summary",
    "issuetype": { "name": "$issue_type" },
    "description": "$description"
  }
}
}
}

```



#### Parameter Hinweise

Bitte beachten Sie, dass für den `http_request_body` keine dynamischen Parameter oder Arrays unterstützt werden.

## 3.6.2 API-Verbindung (Connection)

Jede Funktion ist mit genau **einer API-Verbindung** verknüpft.

Eine Verbindung definiert:

- **API Endpoint (Request-URL)** → z. B. `https://jira.example.com/rest/api/2/issue`
- **HTTP-Methode** (`GET`, `POST`, `PUT`, `PATCH`)
- **Authentifizierungstyp** (API Key, Basic Auth, OAuth usw.)
- **Zugangsdaten**

Die Authentifizierung wird automatisch in den Request-Header eingefügt. Diese Daten sind **nicht Teil der JSON-Definition**.

## 3.6.3 Ablauf eines Function Calls

1. Das KI-Modell entscheidet, eine Funktion aufzurufen, z. B.:

```

{
  "name": "create_jira_ticket",
  "arguments": {
    "summary": "Login-Fehler beheben",
    "project_key": "KEY",
    "issue_type": "Bug",
    "description": "Nutzer können sich nicht einloggen."
  }
}

```

1. Das Backend:
2. liest die Funktionsdefinition aus,
3. ersetzt Platzhalter (`$summary`, `$project_key`, ...) im `http_request_body`,
4. erstellt die HTTP-Anfrage entsprechend der hinterlegten Verbindung.
5. Die Anfrage wird ausgeführt, z. B.:

```

POST https://jira.example.com/rest/api/2/issue
Content-Type: application/json
Authorization: Basic <token>

{
  "fields": {
    "project": { "key": "KEY" },
    "summary": "Login-Fehler beheben",
    "issuetype": { "name": "Bug" },
    "description": "Nutzer können sich nicht einloggen."
  }
}

```

1. Die externe API antwortet (z. B. mit JSON). Das System gibt diese Antwort **unverändert** an das KI-Modell zurück.

## 5. Request-Typen

Typ	Zweck	http_request_body erforderlich?
GET	Lesen von Daten	<input type="checkbox"/> Nein
POST	Erstellen neuer Ressourcen	<input checked="" type="checkbox"/> Ja
PUT	Aktualisieren bestehender Ressourcen	<input checked="" type="checkbox"/> Ja
PATCH	Teilweise Aktualisierung	<input checked="" type="checkbox"/> Ja

## 6. Beispiel: GET Request

```
{
  "parameters": {
    "type": "object",
    "required": ["city"],
    "properties": {
      "city": {
        "type": "string",
        "description": "Name der Stadt"
      },
      "units": {
        "type": "string",
        "description": "Temperatureinheit",
        "enum": ["metric", "imperial"]
      }
    }
  },
  "additionalProperties": false
}
```

### Beispiel-Aufruf:

```
{
  "name": "get_weather",
  "arguments": { "city": "Berlin", "units": "metric" }
}
```

### Ergebnis:

```
GET https://api.open-meteo.com/v1/forecast?city=Berlin&units=metric
```

## 3.7 Integrationsleitfaden für API-Anbieter

Um sicherzustellen, dass eine API von einem KI-Assistenten über Function Calling angesprochen werden kann, müssen die nachfolgenden Bedingungen und Voraussetzungen erfüllt sein.

### 3.7.1 1. Ziel

Damit eine API kompatibel ist, muss sie:

1. Über HTTP erreichbar sein ( GET , POST , PUT , PATCH ).
2. JSON als Eingabe- und Ausgabeformat unterstützen.
3. Klar definierte Felder und erwartete Datentypen besitzen.
4. Eindeutige Parameterbezeichnungen bereitstellen, die im Function-JSON referenziert werden können.

### 3.7.2.2. Anforderungen an den Endpoint

Aspekt	Beschreibung
<b>HTTP-Methoden</b>	Unterstützt <code>GET</code> , <code>POST</code> , <code>PUT</code> oder <code>PATCH</code>
<b>Authentifizierung</b>	Wird über die Connection des Systems hinzugefügt (nicht durch den Aufrufer)
<b>Content-Type</b>	<code>application/json</code>
<b>Antwortformat</b>	JSON (empfohlen)
<b>Fehlerbehandlung</b>	HTTP-Statuscodes (z. B. <code>4xx</code> , <code>5xx</code> ) und Fehlermeldung als Text oder JSON

### 3.7.3.3. Parameterstruktur

Die API sollte so gestaltet sein, dass die Felder aus der Function-Definition direkt befüllt werden können.

Beispiel:

```
{
  "parameters": {
    "type": "object",
    "required": ["customer_id", "status"],
    "properties": {
      "customer_id": {
        "type": "string",
        "description": "Eindeutige Kunden-ID"
      },
      "status": {
        "type": "string",
        "description": "Neuer Status (z. B. active/inactive)"
      }
    }
  },
  "http_request_body": {
    "customer": { "id": "$customer_id" },
    "status": "$status"
  }
}
```

Business GPT ersetzt die Platzhalter mit den tatsächlichen Werten:

```
{
  "customer": { "id": "C123" },
  "status": "active"
}
```

### 3.7.4.4. Antwortstruktur

Die API sollte eine sinnvolle, strukturierte Antwort zurückgeben, z. B.:

```
{
  "status": "success",
  "ticket_id": "Project-1234",
  "message": "Ticket created successfully"
}
```

Fehler sollten standardisiert werden, z. B.:

```
{
  "status": "error",
  "message": "Invalid project key"
}
```

### 3.7.5.5. Best Practices

- Verwenden Sie **sprechende Parameternamen**, die mit der Business-Logik korrespondieren.

- Definieren Sie **alle Pflichtfelder** im JSON-Schema ( `required` ).
- Halten Sie sich an das Prinzip: „LLM liefert nur die semantischen Werte, nicht das Request-Format.“
- Nutzen Sie Platzhalter ( `parameter` ) konsistent im `http_request_body`.

### 3.7.6.6. Beispiel einer vollständigen Integration

#### Ziel:

Eine API, die ein Support-Ticket anlegt.

#### Function JSON:

```
{
  "parameters": {
    "type": "object",
    "required": ["title", "priority"],
    "properties": {
      "title": { "type": "string", "description": "Titel des Tickets" },
      "priority": { "type": "string", "enum": ["Low", "Medium", "High"] },
      "description": { "type": "string" }
    }
  },
  "http_request_body": {
    "ticket": {
      "title": "$title",
      "priority": "$priority",
      "description": "$description"
    }
  }
}
```

#### Business GPT generiert:

```
POST https://api.supportsystem.com/v1/tickets
Authorization: Bearer <token>
Content-Type: application/json

{
  "ticket": {
    "title": "Zugangsdaten verloren",
    "priority": "High",
    "description": "Der Benutzer kann sich nicht anmelden."
  }
}
```

#### API antwortet:

```
{
  "status": "success",
  "ticket_id": "SUP-9876"
}
```

### 3.7.7.7. Zusammenfassung

Komponente	Verantwortung
KI-Modell	Liefert Funktionsname + Parameterwerte
Backend	Ersetzt Platzhalter, führt HTTP-Request aus, fügt Authentifizierung hinzu
Externe API	Nimmt JSON entgegen, verarbeitet es, liefert Antwort zurück
LLM	Nutzt die Antwort, um dem Benutzer die Aktion oder das Ergebnis mitzuteilen

## 4. API

---

### 4.1 Business GPT API

---

Version 3

## 4.1.1 Path Table

Method	Path	Description
GET	/health	Returns Ok (200) Status Code if the system is healthy, otherwise an InternalServerError (500).
GET	/v2/assistant-options	Endpoint for retrieving a all assistant options.
GET	/v2/assistants	Endpoint for listing all assistants.
POST	/v2/assistants	Endpoint for creating an assistant.
DELETE	/v2/assistants/{uuid}	Endpoint for deleting an assistant (does not include default/legacy assistants).
GET	/v2/assistants/{uuid}	Endpoint for retrieving an custom assistant (does not include default/legacy assistants)
PATCH	/v2/assistants/{uuid}	Endpoint for updating an assistant.
PUT	/v2/assistants/{uuid}	Endpoint for updating an assistant.
POST	/v2/audio/transcriptions	Endpoint for chatting with the custom model
POST	/v2/chat/completions	Endpoint for chatting with the custom model
GET	/v2/configuration	Endpoint for retrieving the configuration.
PUT	/v2/configuration	Endpoint for updating the configuration.
GET	/v2/consumption/{month}	Endpoint for retrieving the consumption data.
GET	/v2/containers	Endpoint for listing all containers.
POST	/v2/containers	Endpoint for creating a container.
DELETE	/v2/containers/{uuid}	Endpoint for deleting a container.
GET	/v2/containers/{uuid}	Endpoint for retrieving a container.
PUT	/v2/containers/{uuid}	Endpoint for updating a container.
GET	/v2/containers/{uuid}/documents	Endpoint for retrieving a all documents in a container.
GET	/v2/documents	Endpoint for listing all content.
POST	/v2/documents	Endpoint for uploading a new file.
PUT	/v2/documents-status	Endpoint for retrieving the status of multiple documents.
DELETE	/v2/documents/{uuid}	Endpoint for deleting files and its metadata.
GET	/v2/documents/{uuid}	Endpoint for requesting metadata about a file.
PUT	/v2/documents/{uuid}	Endpoint for updating files metadata.
GET	/v2/documents/{uuid}/download	Endpoint for downloading files.
PUT	/v2/documents/{uuid}/indexing-priority	Endpoint for prioritizing the indexing of a file.
POST	/v2/embeddings	Endpoint for chatting with the custom model
GET	/v2/groups	Endpoint for retrieving all groups.
POST	/v2/groups	Endpoint for retrieving all groups.
DELETE	/v2/groups/{uuid}	Endpoint for deleting a group.
PUT	/v2/groups/{uuid}	Endpoint for retrieving all groups.
GET	/v2/identity	Endpoint for retrieving the configuration.

## 4.1.2 Reference Table

---

Name	Path	Description
ApplicationPermission	<a href="#">#/components/schemas/ApplicationPermission</a>	represents the different roles of the application.
AuthConfigurationDTO	<a href="#">#/components/schemas/AuthConfigurationDTO</a>	
ChatCompletionMessage	<a href="#">#/components/schemas/ChatCompletionMessage</a>	
ChatCompletionMessageContentImage	<a href="#">#/components/schemas/ChatCompletionMessageContentImage</a>	
ChatCompletionMessageContentText	<a href="#">#/components/schemas/ChatCompletionMessageContentText</a>	
ChatResponseChoice	<a href="#">#/components/schemas/ChatResponseChoice</a>	
ConsumptionStorageDto	<a href="#">#/components/schemas/ConsumptionStorageDto</a>	
ConsumptionType	<a href="#">#/components/schemas/ConsumptionType</a>	
ContainerDto	<a href="#">#/components/schemas/ContainerDto</a>	A container is a virtual group that holds many files.
DocumentDTOStatus	<a href="#">#/components/schemas/DocumentDTOStatus</a>	
DocumentDTOStatusResponse	<a href="#">#/components/schemas/DocumentDTOStatusResponse</a>	
DocumentIntelligenceBillingItem	<a href="#">#/components/schemas/DocumentIntelligenceBillingItem</a>	
DocumentSearchToolDto	<a href="#">#/components/schemas/DocumentSearchToolDto</a>	
Embedding	<a href="#">#/components/schemas/Embedding</a>	
EmbeddingUsage	<a href="#">#/components/schemas/EmbeddingUsage</a>	
FileIndexingStatus	<a href="#">#/components/schemas/FileIndexingStatus</a>	
FileInfoDto	<a href="#">#/components/schemas/FileInfoDto</a>	holds the file meta data.
FrontendSettingsDto	<a href="#">#/components/schemas/FrontendSettingsDto</a>	
FunctionDefinition	<a href="#">#/components/schemas/FunctionDefinition</a>	
FunctionTool	<a href="#">#/components/schemas/FunctionTool</a>	
GroupDTO	<a href="#">#/components/schemas/GroupDTO</a>	
HealthDetailedStatus	<a href="#">#/components/schemas/HealthDetailedStatus</a>	Contains information about the application sub-systems health.
Image	<a href="#">#/components/schemas/Image</a>	
ImageToolDto	<a href="#">#/components/schemas/ImageToolDto</a>	
IndexingJobStatusDto	<a href="#">#/components/schemas/IndexingJobStatusDto</a>	
InternetBreakoutTool	<a href="#">#/components/schemas/InternetBreakoutTool</a>	
LLMCapability	<a href="#">#/components/schemas/LLMCapability</a>	
LanguageConfiguration	<a href="#">#/components/schemas/LanguageConfiguration</a>	
LanguageConfigurationDto	<a href="#">#/components/schemas/LanguageConfigurationDto</a>	
LanguageConfigurationDto100	<a href="#">#/components/schemas/LanguageConfigurationDto100</a>	

Name	Path	Description
LanguageConfigurationDto4000	<a href="#">#/components/schemas/LanguageConfigurationDto4000</a>	
LanguageConfigurationDto500	<a href="#">#/components/schemas/LanguageConfigurationDto500</a>	
ModelBillingItem	<a href="#">#/components/schemas/ModelBillingItem</a>	
ModelOptionDto	<a href="#">#/components/schemas/ModelOptionDto</a>	
NewsConfiguration	<a href="#">#/components/schemas/NewsConfiguration</a>	
Parameters	<a href="#">#/components/schemas/Parameters</a>	
ReferenceDto	<a href="#">#/components/schemas/ReferenceDto</a>	represents a reference to another resource.
RestrictedAssistantDto	<a href="#">#/components/schemas/RestrictedAssistantDto</a>	
SecurityClass	<a href="#">#/components/schemas/SecurityClass</a>	
SortOrder	<a href="#">#/components/schemas/SortOrder</a>	
TemplateVariableDto	<a href="#">#/components/schemas/TemplateVariableDto</a>	A template variable is a variable that can be used in the prompt template.
ToolCall	<a href="#">#/components/schemas/ToolCall</a>	
ToolFunction	<a href="#">#/components/schemas/ToolFunction</a>	
Tools	<a href="#">#/components/schemas/Tools</a>	
ToolsDto	<a href="#">#/components/schemas/ToolsDto</a>	similar to <a href="https://platform.openai.com/docs/api-reference/assistants/createAssistant#assistants-createassistant-tools">https://platform.openai.com/docs/api-reference/assistants/createAssistant#assistants-createassistant-tools</a>
ToolsResourcesDocumentSearchDto	<a href="#">#/components/schemas/ToolsResourcesDocumentSearchDto</a>	resources used by the 'document_search' tool
ToolsResourcesDocumentSearchReferencesDto	<a href="#">#/components/schemas/ToolsResourcesDocumentSearchReferencesDto</a>	resources used by the 'document_search' tool
ToolsResourcesDto	<a href="#">#/components/schemas/ToolsResourcesDto</a>	
ToolsResourcesWithReferencesDto	<a href="#">#/components/schemas/ToolsResourcesWithReferencesDto</a>	
UploadGroupDTO	<a href="#">#/components/schemas/UploadGroupDTO</a>	
bearerAuth	<a href="#">#/components/securitySchemes/bearerAuth</a>	

## 4.1.3 Path Details

---

### [GET]/health

- Summary  
Returns Ok (200) Status Code if the system is healthy, otherwise an InternalServerError (500).

### RESPONSES

- 200 Contains information about the application health.

```
application/json
```

```
// Contains information about the application health.
{
  healthy: boolean
  status: Partial(#/components/schemas/HealthDetailedStatus) & Partial(null)
}
```

- 401 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 403 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 404 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 415 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 429 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
}
```

```

title: string
trace_id?: Partial(string) & Partial(null)
type: string
}

```

- 500 Contains information about the application health.

application/json

```

// Contains information about the application health.
{
  healthy: boolean
  status: Partial(#/components/schemas/HealthDetailedStatus) & Partial(null)
}

```

- 501 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 503 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

## [GET]/v2/assistant-options

- Summary  
Endpoint for retrieving a all assistant options.

### RESPONSES

- 200 Options available for configuring an assistant.

application/json

```

// Options available for configuring an assistant.
{
  completion_models: {
    capabilities?: enum[image_generation, text_completion, vision_image, embedding, speech_to_text][]
    label: {
      de?: Partial(string) & Partial(null)
      en?: Partial(string) & Partial(null)
    }
    name: string
  }[]
  prompt_template_variables: {
  }
  // similar to https://platform.openai.com/docs/api-reference/assistants/createAssistant#assistants-createassistant-tools
  tools: {
    type: enum[images, document_search, internet_breakout]
  }[]
}

```

- 401 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 403 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 404 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 415 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 429 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 500 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 501 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 503 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

## [GET]/v2/assistants

- Summary  
Endpoint for listing all assistants.

### PARAMETERS(QUERY)

offset?: integer

page\_size?: integer //default: 20

search\_string?: Partial(string) &amp; Partial(null)

sort\_by?: enum[name\_de, name\_en, active] //default: name\_en

order\_by?: enum[asc, desc]

### RESPONSES

- 200

application/json

```
{
  count: integer
  items: {
    active: boolean
    capabilities?: enum[image_generation, text_completion, vision_image, embedding, speech_to_text][]
  }
}
```

```

completion_model: string
container_count: Partial(integer) & Partial(null)
group_count: Partial(integer) & Partial(null)
id: string
information_message?: Partial(#/components/schemas/LanguageConfigurationDto500) & Partial(null)
label: {
  de?: Partial(string) & Partial(null)
  en?: Partial(string) & Partial(null)
}
legacy: boolean
parameter_values: {
  temperature?: Partial(number) & Partial(null)
  top_p?: Partial(number) & Partial(null)
}
security_class: enum[internal, open]
start_message?: Partial(#/components/schemas/LanguageConfigurationDto4000) & Partial(null)
// similar to https://platform.openai.com/docs/api-reference/assistants/createAssistant#assistants-createassistant-tools
tools: {
  type: enum[images, document_search, internet_breakout]
}[]
tools_resources: {
  document_search?: Partial(#/components/schemas/ToolsResourcesDocumentSearchReferencesDto) & Partial(null)
}
}[]
offset: integer
}

```

- 401 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 403 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 404 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 415 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({

```

```

    }) & Partial(null)
    status: integer
    title: string
    trace_id?: Partial(string) & Partial(null)
    type: string
  }

```

- 429 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 500 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 501 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 503 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

---

## [POST]/v2/assistants

- Summary  
Endpoint for creating an assistant.

## REQUESTBODY

- application/json

```
// The assistant as viewed by users with elevated permissions.
{
  active: boolean
  completion_model: string
  groups?: string[]
  information_message?: Partial(#/components/schemas/LanguageConfigurationDto500) & Partial(null)
  label: {
    de?: Partial(string) & Partial(null)
    en?: Partial(string) & Partial(null)
  }
  parameter_values: {
    temperature?: Partial(number) & Partial(null)
    top_p?: Partial(number) & Partial(null)
  }
  start_message?: Partial(#/components/schemas/LanguageConfigurationDto4000) & Partial(null)
  // This is max. 16k, if you set the ENV variable as SYSTEM_PROMPT_MAX_LENGTH. Default: 4k
  system_prompt: string
  // similar to https://platform.openai.com/docs/api-reference/assistants/createAssistant#assistants-createassistant-tools
  tools: {
    type: enum[images, document_search, internet_breakout]
  }[]
  tools_resources: {
    document_search?: Partial(#/components/schemas/ToolsResourcesDocumentSearchDto) & Partial(null)
  }
}
```

## RESPONSES

- 200 The assistant as viewed by users with elevated permissions.

application/json

```
{
  active: boolean
  completion_model: string
  // represents a reference to another resource.
  groups: {
    id: string
    label: string
  }[]
  id: string
  information_message?: Partial(#/components/schemas/LanguageConfigurationDto500) & Partial(null)
  label: {
    de?: Partial(string) & Partial(null)
    en?: Partial(string) & Partial(null)
  }
  legacy: boolean
  order: Partial(integer) & Partial(null)
  parameter_values: {
    temperature?: Partial(number) & Partial(null)
    top_p?: Partial(number) & Partial(null)
  }
  security_class: enum[internal, open]
  start_message?: Partial(#/components/schemas/LanguageConfigurationDto4000) & Partial(null)
  // This is max. 16k, if you set the ENV variable as SYSTEM_PROMPT_MAX_LENGTH. Default: 4k
  system_prompt: string
  time_created: string
  time_updated: Partial(string) & Partial(null)
  // similar to https://platform.openai.com/docs/api-reference/assistants/createAssistant#assistants-createassistant-tools
  tools: {
    type: enum[images, document_search, internet_breakout]
  }[]
  tools_resources: {
    document_search?: Partial(#/components/schemas/ToolsResourcesDocumentSearchReferencesDto) & Partial(null)
  }
}
```

- 400 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
}
```

```

type: string
}

```

- 401 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 403 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 404 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 415 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 429 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
}

```

```

type: string
}

```

- 500 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 501 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 503 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

---

## [DELETE]/v2/assistants/{uuid}

- Summary  
Endpoint for deleting an assistant (does not include default/legacy assistants).

### RESPONSES

- 200 represents a delete response.

application/json

```

{
}

```

- 400 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
}

```

```

title: string
trace_id?: Partial(string) & Partial(null)
type: string
}

```

- 401 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 403 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 404 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 415 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 429 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
}

```

```

trace_id?: Partial(string) & Partial(null)
type: string
}

```

- 500 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 501 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 503 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

## [GET]/v2/assistants/{uuid}

- Summary  
Endpoint for retrieving an custom assistant (does not include default/legacy assistants)

### RESPONSES

- 200 The assistant as viewed by users with elevated permissions.

application/json

```

{
  active: boolean
  completion_model: string
  // represents a reference to another resource.
  groups: {
    id: string
    label: string
  }[]
  id: string
  information_message?: Partial(#/components/schemas/LanguageConfigurationDto500) & Partial(null)
  label: {
    de?: Partial(string) & Partial(null)
    en?: Partial(string) & Partial(null)
  }
  legacy: boolean
}

```

```

order: Partial(integer) & Partial(null)
parameter_values: {
  temperature?: Partial(number) & Partial(null)
  top_p?: Partial(number) & Partial(null)
}
security_class: enum[internal, open]
start_message?: Partial(#/components/schemas/LanguageConfigurationDto4000) & Partial(null)
// This is max. 16k, if you set the ENV variable as SYSTEM_PROMPT_MAX_LENGTH. Default: 4k
system_prompt: string
time_created: string
time_updated: Partial(string) & Partial(null)
// similar to https://platform.openai.com/docs/api-reference/assistants/createAssistant#assistants-createassistant-tools
tools: {
  type: enum[images, document_search, internet_breakout]
}[]
tools_resources: {
  document_search?: Partial(#/components/schemas/ToolsResourcesDocumentSearchReferencesDto) & Partial(null)
}
}

```

- 401 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 403 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 404 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 415 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
}

```

```
type: string
}
```

- 429 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 500 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 501 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 503 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

## [PATCH]/v2/assistants/{uuid}

- Summary  
Endpoint for updating an assistant.

### REQUESTBODY

- application/json

```
{
  active: boolean
}
```

## RESPONSES

- 200 The assistant as viewed by users with elevated permissions.

application/json

```
{
  active: boolean
  completion_model: string
  // represents a reference to another resource.
  groups: {
    id: string
    label: string
  }[]
  id: string
  information_message?: Partial(#/components/schemas/LanguageConfigurationDto500) & Partial(null)
  label: {
    de?: Partial(string) & Partial(null)
    en?: Partial(string) & Partial(null)
  }
  legacy: boolean
  order: Partial(integer) & Partial(null)
  parameter_values: {
    temperature?: Partial(number) & Partial(null)
    top_p?: Partial(number) & Partial(null)
  }
  security_class: enum[internal, open]
  start_message?: Partial(#/components/schemas/LanguageConfigurationDto4000) & Partial(null)
  // This is max. 16k, if you set the ENV variable as SYSTEM_PROMPT_MAX_LENGTH. Default: 4k
  system_prompt: string
  time_created: string
  time_updated: Partial(string) & Partial(null)
  // similar to https://platform.openai.com/docs/api-reference/assistants/createAssistant#assistants-createassistant-tools
  tools: {
    type: enum[images, document_search, internet_breakout]
  }[]
  tools_resources: {
    document_search?: Partial(#/components/schemas/ToolsResourcesDocumentSearchReferencesDto) & Partial(null)
  }
}
```

- 400 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 401 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 403 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
```

```
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 404 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 415 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 429 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 500 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 501 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
```

```

detail?: Partial(string) & Partial(null)
errors?: Partial({
  }) & Partial(null)
status: integer
title: string
trace_id?: Partial(string) & Partial(null)
type: string
}

```

- 503 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

## [PUT]/v2/assistants/{uuid}

- Summary  
Endpoint for updating an assistant.

### REQUESTBODY

- application/json

```

// The assistant as viewed by users with elevated permissions.
{
  active: boolean
  completion_model: string
  groups?: string[]
  id: string
  information_message?: Partial(#/components/schemas/LanguageConfigurationDto500) & Partial(null)
  label: {
    de?: Partial(string) & Partial(null)
    en?: Partial(string) & Partial(null)
  }
  order: Partial(integer) & Partial(null)
  parameter_values: {
    temperature?: Partial(number) & Partial(null)
    top_p?: Partial(number) & Partial(null)
  }
  start_message?: Partial(#/components/schemas/LanguageConfigurationDto4000) & Partial(null)
  // This is max. 16k, if you set the ENV variable as SYSTEM_PROMPT_MAX_LENGTH. Default: 4k
  system_prompt: string
  // similar to https://platform.openai.com/docs/api-reference/assistants/createAssistant#assistants-createassistant-tools
  tools: {
    type: enum[images, document_search, internet_breakout]
  }[]
  tools_resources: {
    document_search?: Partial(#/components/schemas/ToolsResourcesDocumentSearchDto) & Partial(null)
  }
}

```

### RESPONSES

- 200 The assistant as viewed by users with elevated permissions.

application/json

```

{
  active: boolean
  completion_model: string
  // represents a reference to another resource.
  groups: {
    id: string
    label: string
  }[]
  id: string
  information_message?: Partial(#/components/schemas/LanguageConfigurationDto500) & Partial(null)
  label: {
    de?: Partial(string) & Partial(null)
    en?: Partial(string) & Partial(null)
  }
}

```

```

}
legacy: boolean
order: Partial(integer) & Partial(null)
parameter_values: {
  temperature?: Partial(number) & Partial(null)
  top_p?: Partial(number) & Partial(null)
}
security_class: enum[internal, open]
start_message?: Partial(#/components/schemas/LanguageConfigurationDto4000) & Partial(null)
// This is max. 16k, if you set the ENV variable as SYSTEM_PROMPT_MAX_LENGTH. Default: 4k
system_prompt: string
time_created: string
time_updated: Partial(string) & Partial(null)
// similar to https://platform.openai.com/docs/api-reference/assistants/createAssistant#assistants-createassistant-tools
tools: {
  type: enum[images, document_search, internet_breakout]
}[]
tools_resources: {
  document_search?: Partial(#/components/schemas/ToolsResourcesDocumentSearchReferencesDto) & Partial(null)
}
}
}

```

- 400 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 401 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 403 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 404 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
}

```

```

type: string
}

```

- 415 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 429 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 500 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 501 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 503 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
}

```

```

type: string
}

```

## [POST]/v2/audio/transcriptions

- Summary  
Endpoint for chatting with the custom model
- Description

This endpoint is used to chat with a given chat approach.

It is a protected endpoint and requires a valid JWT token to access it.

### REQUESTBODY

- multipart/form-data

```

{
  file: string
  language?: Partial(string) & Partial(null)
  model: string
}

```

### RESPONSES

- 200

application/json

```

{
  text: string
}

```

- 400 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 401 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 403 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({

```

```

    }) & Partial(null)
    status: integer
    title: string
    trace_id?: Partial(string) & Partial(null)
    type: string
  }

```

- 404 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 415 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 429 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 500 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 501 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
}

```



## RESPONSES

- 200

application/json

```
{
  choices: {
    finish_reason: Partial(string) & Partial(null)
    index: integer
    message: {
      content: Partial(string) & Partial(
        undefined?: Partial(#/components/schemas/ChatCompletionMessageContentImage) & Partial(#/components/schemas/
        ChatCompletionMessageContentText)[]) & Partial(null)
      name?: Partial(string) & Partial(null)
      role: Partial(string) & Partial(null)
      tool?: Partial(string) & Partial(null)
      tool_call_id?: Partial(string) & Partial(null)
      tool_calls?: Partial({
        function: {
          arguments: string
          name: Partial(string) & Partial(null)
        }
        id: Partial(string) & Partial(null)
        type: Partial(string) & Partial(null)
      }[]) & Partial(null)
    }
  }
  created: integer
  id: string
  model: string
  object?: string //default: chat.completion
}
```

- 400 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 401 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 403 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 404 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 415 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 429 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 500 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 501 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 503 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

## [GET]/v2/configuration

- Summary  
Endpoint for retrieving the configuration.
- Description

This endpoint is used to retrieve configuration options, like base settings for models. It is a protected endpoint and requires a valid JWT token to access it.

### RESPONSES

- 200

application/json

```
{
  accent_color: string
  auth_configuration: {
    auth_client_id: string
    auth_provider: string
  }
  background: integer
  budget_reached: boolean
  company_logo_url: Partial(string) & Partial(null)
  company_name: string
  feature_flags: {
  }
  frontend_settings: {
    file_size_limit_mb: integer
    private_files_limit: integer
    system_prompt_max_length: integer
  }
  news: Partial(#/components/schemas/LanguageConfiguration) & Partial(null)
  telekom_news: Partial(#/components/schemas/NewsConfiguration) & Partial(null)
}
```

- 401 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 403 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
}
```

```

title: string
trace_id?: Partial(string) & Partial(null)
type: string
}

```

- 404 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 415 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 429 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 500 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 501 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
}

```

```

trace_id?: Partial(string) & Partial(null)
type: string
}

```

- 503 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

## [PUT]/v2/configuration

- Summary  
Endpoint for updating the configuration.
- Description

This endpoint is used to update configuration options, like base settings for models.

### REQUESTBODY

- application/json

```

{
  accent_color: string
  background: integer
  company_logo_url: Partial(string) & Partial(null)
  company_name: string
  news: Partial(#/components/schemas/LanguageConfiguration) & Partial(null)
  telekom_news: Partial(#/components/schemas/NewsConfiguration) & Partial(null)
}

```

### RESPONSES

- 400 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 401 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
}

```

```

type: string
}

```

- 403 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 404 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 415 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 429 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 500 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
}

```

```

type: string
}

```

- 501 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 503 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

## [GET]/v2/consumption/{month}

- Summary  
Endpoint for retrieving the consumption data.
- Description

This Endpoint is used to retrieve consumption data.

It is a protected endpoint and requires a valid JWT token to access it.

### RESPONSES

- 200 Consumption information for the current period.

application/json

```

// Consumption information for the current period.
{
  billing_items?: Partial(#/components/schemas/DocumentIntelligenceBillingItem) & Partial(#/components/schemas/ModelBillingItem) []
  chatting_cost: number
  currency: enum[EUR, USD]
  indexing_cost: number
  overall_budget: number
  overall_cost: number
  overall_remaining_budget: number
  percentage_budget_used: integer
  period_end: string
  period_start: string
  storage: {
    file_count: integer
    files_size: integer
  }
}

```

- 401 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
}

```

```

errors?: Partial({
  }) & Partial(null)
status: integer
title: string
trace_id?: Partial(string) & Partial(null)
type: string
}

```

- 403 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
errors?: Partial({
  }) & Partial(null)
status: integer
title: string
trace_id?: Partial(string) & Partial(null)
type: string
}

```

- 404 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
errors?: Partial({
  }) & Partial(null)
status: integer
title: string
trace_id?: Partial(string) & Partial(null)
type: string
}

```

- 415 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
errors?: Partial({
  }) & Partial(null)
status: integer
title: string
trace_id?: Partial(string) & Partial(null)
type: string
}

```

- 429 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
errors?: Partial({
  }) & Partial(null)
status: integer
title: string
trace_id?: Partial(string) & Partial(null)
type: string
}

```

- 500 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
errors?: Partial({

```

```

    }) & Partial(null)
    status: integer
    title: string
    trace_id?: Partial(string) & Partial(null)
    type: string
  }

```

- 501 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 503 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

## [GET]/v2/containers

- Summary  
Endpoint for listing all containers.
- Description

It is a protected endpoint and requires a valid JWT token to access it.

### PARAMETERS(QUERY)

offset?: integer

page\_size?: integer //default: 20

search\_string?: Partial(string) & Partial(null)

by\_assistant?: Partial(string) & Partial(null)

sort\_by?: string //default: name

order\_by?: enum[asc, desc]

### RESPONSES

- 200

application/json

```

{
  count: integer
  // A container is a virtual group that holds many files.
  items: {
    color?: string
    description?: Partial(string) & Partial(null)
  }
}

```

```

display_name: string
document_count: integer
// represents a reference to another resource.
groups: {
  id: string
  label: string
}[]
id: string
status: enum[not_indexed, indexing, indexed, error]
}[]
offset: integer
}

```

- 401 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 403 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 404 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 415 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 429 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 500 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 501 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 503 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

---

## [POST]/v2/containers

- Summary  
Endpoint for creating a container.
- Description

It is a protected endpoint and requires a valid JWT token to access it.

### REQUESTBODY

- application/json

```
// A container is a virtual group that holds many files.
{
  color?: string
  description?: Partial(string) & Partial(null)
  display_name: string
}
```

```
groups?: string[]
}
```

## RESPONSES

- 200 A container is a virtual group that holds many files.

application/json

```
// A container is a virtual group that holds many files.
{
  color?: string
  description?: Partial(string) & Partial(null)
  display_name: string
  document_count: integer
  // represents a reference to another resource.
  groups: {
    id: string
    label: string
  }[]
  id: string
  status: enum[not_indexed, indexing, indexed, error]
}
```

- 400 DuplicateEntryProblem(lang: gettext.NullTranslations)

application/json

```
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type?: string //default: tag:businessgpt@telekom.de,2024-09-01:DuplicateEntry
}
```

- 401 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 403 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 404 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
}
```

```

title: string
trace_id?: Partial(string) & Partial(null)
type: string
}

```

- 415 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 429 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 500 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 501 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 503 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
}

```

```

trace_id?: Partial(string) & Partial(null)
type: string
}

```

## **[DELETE]/v2/containers/{uuid}**

- **Summary**  
Endpoint for deleting a container.
- **Description**

It is a protected endpoint and requires a valid JWT token to access it.

### RESPONSES

- 200 represents a delete response.

application/json

```

{
}

```

- 400 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 401 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 403 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 404 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 415 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 429 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 500 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 501 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 503 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

## [GET]/v2/containers/{uuid}

- Summary  
Endpoint for retrieving a container.
- Description

It is a protected endpoint and requires a valid JWT token to access it.

### RESPONSES

- 200 A container is a virtual group that holds many files.

application/json

```
// A container is a virtual group that holds many files.
{
  color?: string
  description?: Partial(string) & Partial(null)
  display_name: string
  document_count: integer
  // represents a reference to another resource.
  groups: {
    id: string
    label: string
  }[]
  id: string
  status: enum[not_indexed, indexing, indexed, error]
}
```

- 401 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 403 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 404 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 415 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 429 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 500 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 501 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 503 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

## [PUT]/v2/containers/{uuid}

- Summary  
Endpoint for updating a container.
- Description

It is a protected endpoint and requires a valid JWT token to access it.

### REQUESTBODY

- application/json

```
// A container is a virtual group that holds many files.
{
  color?: string
  description?: Partial(string) & Partial(null)
  display_name: string
  groups?: string[]
  id: string
}
```

### RESPONSES

- 200 A container is a virtual group that holds many files.

application/json

```
// A container is a virtual group that holds many files.
{
  color?: string
  description?: Partial(string) & Partial(null)
  display_name: string
  document_count: integer
  // represents a reference to another resource.
  groups: {
    id: string
    label: string
  }[]
  id: string
  status: enum[not_indexed, indexing, indexed, error]
}
```

- 400 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 401 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 403 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 404 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 415 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 429 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 500 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 501 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 503 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

## [GET]/v2/containers/{uuid}/documents

- Summary  
Endpoint for retrieving all documents in a container.
- Description

It is a protected endpoint and requires a valid JWT token to access it.

### PARAMETERS(QUERY)

offset?: integer

page\_size?: integer //default: 20

search\_string?: Partial(string) & Partial(null)

status?: Partial(#/components/schemas/IndexingJobStatusDto) & Partial(null)

sort\_by?: enum[name, indexed, id, mimetype, file\_extension, time\_created, time\_updated] //default: name

order\_by?: enum[asc, desc]

### RESPONSES

- 200

application/json

```

{
  count: integer
  currentlyIndexingCount: integer
  errorIndexedCount: integer
  indexedCount: integer
  // holds the file meta data.
  items: {
    backlink?: Partial(string) & Partial(null)
    // represents a reference to another resource.
    containers: {
      id: string
      label: string
    }[]
    description: string
    file_extension: Partial(string) & Partial(null)
    id: string
    index_error: Partial(string) & Partial(null)
    indexed: enum[not_indexed, indexing, indexed, error]
    mimetype: Partial(string) & Partial(null)
    name: string
    time_created: string
    time_updated: Partial(string) & Partial(null)
    version: integer
  }[]
  notIndexedCount: integer
  offset: integer
}

```

- 401 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 403 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 404 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 415 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{

```

```

detail?: Partial(string) & Partial(null)
errors?: Partial({
  }) & Partial(null)
status: integer
title: string
trace_id?: Partial(string) & Partial(null)
type: string
}

```

- 429 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 500 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 501 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 503 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

## [GET]/v2/documents

- Summary  
Endpoint for listing all content.

## Description

This endpoint is used to retrieve a list of all content.

It is a protected endpoint and requires a valid JWT token to access it.

### PARAMETERS(QUERY)

offset?: integer

page\_size?: integer //default: 20

search\_string?: Partial(string) & Partial(null)

status?: Partial(#/components/schemas/IndexingJobStatusDto) & Partial(null)

sort\_by?: enum[name, indexed, id, mimetype, file\_extension, time\_created, time\_updated] //default: name

order\_by?: enum[asc, desc]

### RESPONSES

- 200

application/json

```
{
  count: integer
  currentlyIndexingCount: integer
  errorIndexedCount: integer
  indexedCount: integer
  // holds the file meta data.
  items: {
    backlink?: Partial(string) & Partial(null)
    // represents a reference to another resource.
    containers: {
      id: string
      label: string
    }[]
    description: string
    file_extension: Partial(string) & Partial(null)
    id: string
    index_error: Partial(string) & Partial(null)
    indexed: enum[not_indexed, indexing, indexed, error]
    mimetype: Partial(string) & Partial(null)
    name: string
    time_created: string
    time_updated: Partial(string) & Partial(null)
    version: integer
  }[]
  notIndexedCount: integer
  offset: integer
}
```

- 401 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 403 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
```

```

    }) & Partial(null)
    status: integer
    title: string
    trace_id?: Partial(string) & Partial(null)
    type: string
  }

```

- 404 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 415 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 429 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 500 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 501 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
}

```

```

status: integer
title: string
trace_id?: Partial(string) & Partial(null)
type: string
}

```

- 503 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

## [POST]/v2/documents

- Summary  
Endpoint for uploading a new file.
- Description

This endpoint is used to upload the given file to the storage.

It is a protected endpoint and requires a valid JWT token to access it.

### REQUESTBODY

- multipart/form-data

```

// holds the file meta data for creating a new file.
{
  backlink?: Partial(string) & Partial(null)
  container: string
  data: string
  description: string
  mimetype: Partial(string) & Partial(null)
  name: string
  reference_id?: Partial(string) & Partial(null)
}

```

### RESPONSES

- 200 holds the file meta data.

application/json

```

// holds the file meta data.
{
  backlink?: Partial(string) & Partial(null)
  // represents a reference to another resource.
  containers: {
    id: string
    label: string
  }[]
  description: string
  file_extension: Partial(string) & Partial(null)
  id: string
  index_error: Partial(string) & Partial(null)
  indexed: enum[not_indexed, indexing, indexed, error]
  mimetype: Partial(string) & Partial(null)
  name: string
  time_created: string
  time_updated: Partial(string) & Partial(null)
  version: integer
}

```

- 400 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 401 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 403 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 404 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 415 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 429 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 500 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 501 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 503 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

## [PUT]/v2/documents-status

- Summary  
Endpoint for retrieving the status of multiple documents.

### REQUESTBODY

- application/json

```
{
  document_ids?: string[]
}
```

### RESPONSES

- 200

application/json

```
{
  id: string
  status: enum[pending, started, completed, error, finished]
}
```

- 400 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 401 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 403 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 404 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 415 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
```

```

    }) & Partial(null)
    status: integer
    title: string
    trace_id?: Partial(string) & Partial(null)
    type: string
  }

```

- 429 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 500 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 501 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 503 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

---

## [DELETE]/v2/documents/{uuid}

- Summary  
Endpoint for deleting files and its metadata.
- Description

It is a protected endpoint and requires a valid JWT token to access it.

## RESPONSES

- 200 represents a delete response.

application/json

```
{
}
```

- 400 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 401 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 403 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 404 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 415 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 429 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 500 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 501 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 503 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

**[GET]/v2/documents/{uuid}**

- Summary  
Endpoint for requesting metadata about a file.
- Description

It is a protected endpoint and requires a valid JWT token to access it.

**RESPONSES**

- 200 holds the file meta data.

application/json

```
// holds the file meta data.
{
  backlink?: Partial(string) & Partial(null)
  // represents a reference to another resource.
  containers: {
    id: string
    label: string
  }[]
  description: string
  file_extension: Partial(string) & Partial(null)
  id: string
  index_error: Partial(string) & Partial(null)
  indexed: enum[not_indexed, indexing, indexed, error]
  mimetype: Partial(string) & Partial(null)
  name: string
  time_created: string
  time_updated: Partial(string) & Partial(null)
  version: integer
}
```

- 401 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 403 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 404 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
}
```

```

type: string
}

```

- 415 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 429 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 500 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 501 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 503 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
}

```

```
type: string
}
```

## [PUT]/v2/documents/{uuid}

- Summary  
Endpoint for updating files metadata.
- Description

It is a protected endpoint and requires a valid JWT token to access it.

### REQUESTBODY

- application/json

```
// holds the file meta data, for updating an existing file.
{
  containers?: string[]
  description: string
  id: string
  mimetype: Partial(string) & Partial(null)
  name: string
}
```

### RESPONSES

- 200 holds the file meta data.

application/json

```
// holds the file meta data.
{
  backlink?: Partial(string) & Partial(null)
  // represents a reference to another resource.
  containers: {
    id: string
    label: string
  }[]
  description: string
  file_extension: Partial(string) & Partial(null)
  id: string
  index_error: Partial(string) & Partial(null)
  indexed: enum[not_indexed, indexing, indexed, error]
  mimetype: Partial(string) & Partial(null)
  name: string
  time_created: string
  time_updated: Partial(string) & Partial(null)
  version: integer
}
```

- 400 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 401 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
}
```

```

title: string
trace_id?: Partial(string) & Partial(null)
type: string
}

```

- 403 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 404 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 415 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 429 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 500 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
}

```

```

trace_id?: Partial(string) & Partial(null)
type: string
}

```

- 501 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 503 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

## [GET]/v2/documents/{uuid}/download

- Summary  
Endpoint for downloading files.
- Description

This endpoint is used to download the given file.

It is a protected endpoint and requires a valid JWT token to access it.

### RESPONSES

- 401 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 403 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
}

```

```

type: string
}

```

- 404 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 415 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 429 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 500 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 501 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
}

```

```

type: string
}

```

- 503 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

## [PUT]/v2/documents/{uuid}/indexing-priority

- Summary  
Endpoint for prioritizing the indexing of a file.

### RESPONSES

- 200

application/json

```

{
  "title": "FileIndexActionDTO"
}

```

- 400 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 401 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 403 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
}

```

```

status: integer
title: string
trace_id?: Partial(string) & Partial(null)
type: string
}

```

- 404 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 415 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 429 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 500 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 501 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
}

```

```

title: string
trace_id?: Partial(string) & Partial(null)
type: string
}

```

- 503 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

## [POST]/v2/embeddings

- Summary  
Endpoint for chatting with the custom model
- Description

This endpoint is used to chat with a given chat approach.  
It is a protected endpoint and requires a valid JWT token to access it.

### REQUESTBODY

- application/json

```

// Request for generating an embedding.
// Link: https://platform.openai.com/docs/api-reference/embeddings
{
  encoding_format: Partial(string) & Partial(string)
  input: Partial(string) & Partial(string[])
  model: string
}

```

### RESPONSES

- 200 Response containing the embedding vector. Link: <https://platform.openai.com/docs/api-reference/embeddings>

application/json

```

// Response containing the embedding vector.
// Link: https://platform.openai.com/docs/api-reference/embeddings
{
  data: {
    embedding?: number[]
    index: integer
    object?: string //default: embedding
  }[]
  model: string
  object?: string //default: list
  usage: {
    prompt_tokens: integer
    total_tokens: integer
  }
}

```

- 400 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
}

```

```

trace_id?: Partial(string) & Partial(null)
type: string
}

```

- 401 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 403 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 404 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 415 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 429 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
}

```

```
type: string
}
```

- 500 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 501 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 503 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

## [GET]/v2/groups

- Summary  
Endpoint for retrieving all groups.

### PARAMETERS(QUERY)

offset?: integer

page\_size?: integer //default: 20

search\_string?: Partial(string) & Partial(null)

show\_hidden?: boolean

sort\_by?: enum[name, label, id, time\_created, time\_updated] //default: name

order\_by?: enum[asc, desc]

### RESPONSES

- 200

application/json

```
{
  count: integer
  items: {
    hidden: boolean
    id: string
    label: string
    name: string
    time_created: string
    time_updated: Partial(string) & Partial(null)
  }[]
  offset: integer
}
```

- 401 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 403 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 404 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 415 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 429 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 500 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 501 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 503 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

## [POST]/v2/groups

- Summary  
Endpoint for retrieving all groups.

### REQUESTBODY

- application/json

```
{
  hidden: boolean
  label: string
  name: string
}[]
```

## RESPONSES

- 200

application/json

```
{
  hidden: boolean
  id: string
  label: string
  name: string
  time_created: string
  time_updated: Partial(string) & Partial(null)
}
```

- 400 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 401 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 403 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 404 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 415 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 429 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 500 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 501 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 503 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

**[DELETE]/v2/groups/{uuid}**

- Summary  
Endpoint for deleting a group.
- Description

It is a protected endpoint and requires a valid JWT token to access it.

**RESPONSES**

- 200 represents a delete response.

application/json

```
{
}
```

- 400 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 401 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 403 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}
```

- 404 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```
// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
}
```

```

trace_id?: Partial(string) & Partial(null)
type: string
}

```

- 415 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 429 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 500 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 501 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 503 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
}

```

```

type: string
}

```

## [PUT]/v2/groups/{uuid}

- Summary

Endpoint for retrieving all groups.

### REQUESTBODY

- application/json

```

{
  hidden: boolean
  id: string
  label: string
  name: string
}

```

### RESPONSES

- 200

application/json

```

{
  hidden: boolean
  id: string
  label: string
  name: string
  time_created: string
  time_updated: Partial(string) & Partial(null)
}

```

- 400 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 401 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 403 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
}

```

```

status: integer
title: string
trace_id?: Partial(string) & Partial(null)
type: string
}

```

- 404 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 415 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 429 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 500 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 501 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
}

```

```

title: string
trace_id?: Partial(string) & Partial(null)
type: string
}

```

- 503 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

## [GET]/v2/identity

- Summary  
Endpoint for retrieving the configuration.
- Description

This endpoint is used to retrieve configuration options, like base settings for models. It is a protected endpoint and requires a valid JWT token to access it.

### RESPONSES

- 200 Holds roles, groups and permissions of a user for the frontend.

application/json

```

// Holds roles, groups and permissions of a user for the frontend.
{
  groups?: string[]
  // represents the different roles of the application.
  //
  // As documented in https://confluence.telekom-mms.com/display/BGTPPM/Roles+and+rights
  permissions?: enum[configuration_read, configuration_update, consumption_read, containers_super_admin, assistants_create, assistants_read, assistants_update, assistants_delete, containers_create, containers_read, containers_update, containers_delete, private_containers_access, documents_create, documents_read, documents_update, documents_delete, groups_create, groups_read, groups_update, groups_delete, indexing_create]
  private_container_id: string
  roles?: string[]
}

```

- 401 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 403 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
}

```

```

title: string
trace_id?: Partial(string) & Partial(null)
type: string
}

```

- 404 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 415 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 429 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 500 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

- 501 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
    }) & Partial(null)
  status: integer
  title: string
}

```

```

trace_id?: Partial(string) & Partial(null)
type: string
}

```

- 503 Our extension of the Standard, inspired by RFC8457 <https://www.rfc-editor.org/rfc/rfc9457.html>

#### application/json

```

// Our extension of the Standard, inspired by RFC8457
// https://www.rfc-editor.org/rfc/rfc9457.html
{
  detail?: Partial(string) & Partial(null)
  errors?: Partial({
  }) & Partial(null)
  status: integer
  title: string
  trace_id?: Partial(string) & Partial(null)
  type: string
}

```

## 4.1.4 References

---

### #/components/schemas/ApplicationPermission

```

{
  "description": "represents the different roles of the application.\n\nAs documented in https://confluence.telekom-mms.com/display/BGTFPM/Roles+and+rights",
  "enum": [
    "configuration_read",
    "configuration_update",
    "consumption_read",
    "containers_super_admin",
    "assistants_create",
    "assistants_read",
    "assistants_update",
    "assistants_delete",
    "containers_create",
    "containers_read",
    "containers_update",
    "containers_delete",
    "private_containers_access",
    "documents_create",
    "documents_read",
    "documents_update",
    "documents_delete",
    "groups_create",
    "groups_read",
    "groups_update",
    "groups_delete",
    "indexing_create"
  ],
  "title": "ApplicationPermission",
  "type": "string"
}

```

### #/components/schemas/AuthConfigurationDTO

```

{
  auth_client_id: string
  auth_provider: string
}

```

### #/components/schemas/ChatCompletionMessage

```

{
  content: Partial(string) & Partial( undefined?: Partial(#/components/schemas/ChatCompletionMessageContentImage) & Partial(#/components/schemas/ChatCompletionMessageContentText)[]) & Partial(null)
  name?: Partial(string) & Partial(null)
  role: Partial(string) & Partial(null)
  tool?: Partial(string) & Partial(null)
  tool_call_id?: Partial(string) & Partial(null)
  tool_calls?: Partial({
    function: {
      arguments: string
      name: Partial(string) & Partial(null)
    }
  }) & Partial(null)
  id: Partial(string) & Partial(null)
  type: Partial(string) & Partial(null)
} & Partial(null)
}

```

**#/components/schemas/ChatCompletionMessageContentImage**

```
{
  image_url: Partial(#/components/schemas/Image) & Partial(null)
  type?: string //default: image_url
}
```

**#/components/schemas/ChatCompletionMessageContentText**

```
{
  text: Partial(string) & Partial(null)
  type?: string //default: text
}
```

**#/components/schemas/ChatResponseChoice**

```
{
  finish_reason: Partial(string) & Partial(null)
  index: integer
  message: {
    content: Partial(string) & Partial(
      undefined?: Partial(#/components/schemas/ChatCompletionMessageContentImage) & Partial(#/components/schemas/
ChatCompletionMessageContentText)[] & Partial(null)
    name?: Partial(string) & Partial(null)
    role: Partial(string) & Partial(null)
    tool?: Partial(string) & Partial(null)
    tool_call_id?: Partial(string) & Partial(null)
    tool_calls?: Partial({
      function: {
        arguments: string
        name: Partial(string) & Partial(null)
      }
      id: Partial(string) & Partial(null)
      type: Partial(string) & Partial(null)
    })[] & Partial(null)
  }
}
```

**#/components/schemas/ConsumptionStorageDto**

```
{
  file_count: integer
  files_size: integer
}
```

**#/components/schemas/ConsumptionType**

```
{
  "enum": [
    "indexing",
    "chatting",
    "image_generation"
  ],
  "title": "ConsumptionType",
  "type": "string"
}
```

**#/components/schemas/ContainerDto**

```
// A container is a virtual group that holds many files.
{
  color?: string
  description?: Partial(string) & Partial(null)
  display_name: string
  document_count: integer
  // represents a reference to another resource.
  groups: {
    id: string
    label: string
  }[]
  id: string
  status: enum[not_indexed, indexing, indexed, error]
}
```

**#/components/schemas/DocumentDTOSTatus**

```

{
  "enum": [
    "pending",
    "started",
    "completed",
    "error",
    "finished"
  ],
  "title": "DocumentDTOStatus",
  "type": "string"
}

```

### #/components/schemas/DocumentDTOStatusResponse

```

{
  id: string
  status: enum[pending, started, completed, error, finished]
}

```

### #/components/schemas/DocumentIntelligenceBillingItem

```

{
  billing_group: enum[indexing, chatting, image_generation]
  current_cost: number
  label: string
  price_per_unit: number
  type: string
  unit: enum[tokens, pages, images]
  used_unit_count: integer
}

```

### #/components/schemas/DocumentSearchToolDto

```

{
  containers?: Partial(string[]) & Partial(null)
  documents?: Partial(string[]) & Partial(null)
  top_k?: Partial(integer) & Partial(null)
  type?: string //default: document_search
}

```

### #/components/schemas/Embedding

```

{
  embedding?: number[]
  index: integer
  object?: string //default: embedding
}

```

### #/components/schemas/EmbeddingUsage

```

{
  prompt_tokens: integer
  total_tokens: integer
}

```

### #/components/schemas/FileIndexingStatus

```

{
  "enum": [
    "not_indexed",
    "indexing",
    "indexed",
    "error"
  ],
  "title": "FileIndexingStatus",
  "type": "string"
}

```

### #/components/schemas/FileInfoDto

```

// holds the file meta data.
{

```

```

backlink?: Partial(string) & Partial(null)
// represents a reference to another resource.
containers: {
  id: string
  label: string
}[]
description: string
file_extension: Partial(string) & Partial(null)
id: string
index_error: Partial(string) & Partial(null)
indexed: enum[not_indexed, indexing, indexed, error]
mimetype: Partial(string) & Partial(null)
name: string
time_created: string
time_updated: Partial(string) & Partial(null)
version: integer
}

```

### #/components/schemas/FrontendSettingsDto

```

{
  file_size_limit_mb: integer
  private_files_limit: integer
  system_prompt_max_length: integer
}

```

### #/components/schemas/FunctionDefinition

```

{
  description?: string
  name: string
  strict?: boolean
}

```

### #/components/schemas/FunctionTool

```

{
  function: {
    description?: string
    name: string
    strict?: boolean
  }
  type?: string //default: function
}

```

### #/components/schemas/GroupDTO

```

{
  hidden: boolean
  id: string
  label: string
  name: string
  time_created: string
  time_updated: Partial(string) & Partial(null)
}

```

### #/components/schemas/HealthDetailedStatus

```

// Contains information about the application sub-systems health.
{
  budget?: Partial(integer) & Partial(null)
  commit_hash?: Partial(string) & Partial(null)
  commit_tag?: Partial(string) & Partial(null)
  current_budget_usd?: Partial(number) & Partial(null)
  database_code_migration_level?: Partial(integer) & Partial(null)
  database_migration_level?: Partial(string) & Partial(null)
  status_database?: Partial(boolean) & Partial(null)
  status_openai?: Partial(boolean) & Partial(null)
  status_search?: Partial(boolean) & Partial(null)
  status_user_authentication?: Partial(boolean) & Partial(null)
}

```

### #/components/schemas/Image

```
{
  b64_json?: Partial(string) & Partial(null)
  url?: Partial(string) & Partial(null)
}
```

### #/components/schemas/ImageToolDto

```
{
  model?: Partial(string) & Partial(null)
  size?: Partial(string) & Partial(null)
  type?: string //default: images
}
```

### #/components/schemas/IndexingJobStatusDto

```
{
  "enum": [
    "pending",
    "started",
    "completed",
    "error"
  ],
  "title": "IndexingJobStatusDto",
  "type": "string"
}
```

### #/components/schemas/InternetBreakoutTool

```
{
  type?: string //default: internet_breakout
}
```

### #/components/schemas/LLMCapability

```
{
  "enum": [
    "image_generation",
    "text_completion",
    "vision_image",
    "embedding",
    "speech_to_text"
  ],
  "title": "LLMCapability",
  "type": "string"
}
```

### #/components/schemas/LanguageConfiguration

```
{
  de?: Partial(string) & Partial(null)
  en?: Partial(string) & Partial(null)
}
```

### #/components/schemas/LanguageConfigurationDto

```
{
  de?: Partial(string) & Partial(null)
  en?: Partial(string) & Partial(null)
}
```

### #/components/schemas/LanguageConfigurationDto100

```
{
  de?: Partial(string) & Partial(null)
  en?: Partial(string) & Partial(null)
}
```

### #/components/schemas/LanguageConfigurationDto4000

```
{
  de?: Partial(string) & Partial(null)
  en?: Partial(string) & Partial(null)
}
```

### #/components/schemas/LanguageConfigurationDto500

```
{
  de?: Partial(string) & Partial(null)
  en?: Partial(string) & Partial(null)
}
```

### #/components/schemas/ModelBillingItem

```
{
  billing_group: enum[indexing, chatting, image_generation]
  current_cost: number
  label: string
  model_key: string
  price_per_unit: number
  token_type: Partial(string) & Partial(null)
  type: string
  unit: enum[tokens, pages, images]
  used_unit_count: integer
}
```

### #/components/schemas/ModelOptionDto

```
{
  capabilities?: enum[image_generation, text_completion, vision_image, embedding, speech_to_text][]
  label: {
    de?: Partial(string) & Partial(null)
    en?: Partial(string) & Partial(null)
  }
  name: string
}
```

### #/components/schemas/NewsConfiguration

```
{
  content?: Partial(#/components/schemas/LanguageConfiguration) & Partial(null)
  version?: Partial(integer) & Partial(null)
}
```

### #/components/schemas/Parameters

```
{
  temperature?: Partial(number) & Partial(null)
  top_p?: Partial(number) & Partial(null)
}
```

### #/components/schemas/ReferenceDto

```
// represents a reference to another resource.
{
  id: string
  label: string
}
```

### #/components/schemas/RestrictedAssistantDto

```
{
  active: boolean
  capabilities?: enum[image_generation, text_completion, vision_image, embedding, speech_to_text][]
  completion_model: string
  container_count: Partial(integer) & Partial(null)
  group_count: Partial(integer) & Partial(null)
  id: string
  information_message?: Partial(#/components/schemas/LanguageConfigurationDto500) & Partial(null)
  label: {
    de?: Partial(string) & Partial(null)
    en?: Partial(string) & Partial(null)
  }
}
```

```

}
legacy: boolean
parameter_values: {
  temperature?: Partial(number) & Partial(null)
  top_p?: Partial(number) & Partial(null)
}
security_class: enum[internal, open]
start_message?: Partial(#/components/schemas/LanguageConfigurationDto4000) & Partial(null)
// similar to https://platform.openai.com/docs/api-reference/assistants/createAssistant#assistants-createassistant-tools
tools: {
  type: enum[images, document_search, internet_breakout]
}[]
tools_resources: {
  document_search?: Partial(#/components/schemas/ToolsResourcesDocumentSearchReferencesDto) & Partial(null)
}
}

```

### #/components/schemas/SecurityClass

```

{
  "enum": [
    "internal",
    "open"
  ],
  "title": "SecurityClass",
  "type": "string"
}

```

### #/components/schemas/SortOrder

```

{
  "enum": [
    "asc",
    "desc"
  ],
  "title": "SortOrder",
  "type": "string"
}

```

### #/components/schemas/TemplateVariableDto

```

// A template variable is a variable that can be used in the prompt template.
{
  description: {
    de?: Partial(string) & Partial(null)
    en?: Partial(string) & Partial(null)
  }
  name: string
}

```

### #/components/schemas/ToolCall

```

{
  function: {
    arguments: string
    name: Partial(string) & Partial(null)
  }
  id: Partial(string) & Partial(null)
  type: Partial(string) & Partial(null)
}

```

### #/components/schemas/ToolFunction

```

{
  arguments: string
  name: Partial(string) & Partial(null)
}

```

### #/components/schemas/Tools

```

{
  "enum": [
    "images",
    "document_search",
    "internet_breakout"
  ]
}

```

```

},
"title": "Tools",
"type": "string"
}

```

### #/components/schemas/ToolsDto

```

// similar to https://platform.openai.com/docs/api-reference/assistants/createAssistant#assistants-createassistant-tools
{
  type: enum[images, document_search, internet_breakout]
}

```

### #/components/schemas/ToolsResourcesDocumentSearchDto

```

// resources used by the 'document_search' tool
{
  containers?: string[]
  top_k?: Partial(integer) & Partial(null)
}

```

### #/components/schemas/ToolsResourcesDocumentSearchReferencesDto

```

// resources used by the 'document_search' tool
{
  // represents a reference to another resource.
  containers: {
    id: string
    label: string
  }[]
  top_k?: Partial(integer) & Partial(null)
}

```

### #/components/schemas/ToolsResourcesDto

```

{
  document_search?: Partial(#/components/schemas/ToolsResourcesDocumentSearchDto) & Partial(null)
}

```

### #/components/schemas/ToolsResourcesWithReferencesDto

```

{
  document_search?: Partial(#/components/schemas/ToolsResourcesDocumentSearchReferencesDto) & Partial(null)
}

```

### #/components/schemas/UploadGroupDTO

```

{
  hidden: boolean
  label: string
  name: string
}

```

### #/components/securitySchemes/bearerAuth

```

{
  "bearerFormat": "JWT",
  "scheme": "bearer",
  "type": "http"
}

```

## 4.2 API-Aufruf von BusinessGPT per OAuth2 (Client Credentials)

### 4.2.1 Übersicht

Diese Anleitung beschreibt Schritt für Schritt:

1. Welche technischen Voraussetzungen für den API-Zugriff auf BusinessGPT nötig sind
2. Wie Sie in **Microsoft Entra ID** (Azure AD) eine **API-App-Registrierung** anlegen
3. Wie Sie per **OAuth2 Client Credentials** ein **Access Token** in einem API-Tool (z. B. Postman) abrufen
4. Wie Sie damit einen ersten **Testaufruf** an BusinessGPT durchführen

#### Wichtiger Hinweis

Alle in dieser Anleitung genannten URLs, IDs, Modell-IDs und Namen sind **Beispiele**.  
In Ihrer Umgebung müssen Sie immer die **konkreten Werte aus Ihrem Tenant und Ihrer BusinessGPT-Instanz** verwenden.

### 4.2.2 1. Voraussetzungen

Bevor Sie starten, stellen Sie bitte sicher, dass:

- Sie Zugriff auf das **Azure-Portal** (Microsoft Entra ID / Azure AD) haben
- Ihnen bekannt ist, **welche App-Registrierung** Ihre „normale“ BusinessGPT-Anwendung repräsentiert (z. B. mit einem Namen wie „customer“ – **nur ein Beispiel**)
- Eine Person mit **Adminrechten** in Ihrem Tenant bei Bedarf einmalig den **Admin-Consent** erteilen kann (siehe Schritt 2.6)
- Sie ein API-Tool verwenden, z. B.:
  - **Postman**
  - **Rapid API Client** (Visual Studio Code Extension)
  - Ein anderes REST-Client-Tool Ihrer Wahl

### 4.2.3 2. Entra ID: Zusätzliche App-Registrierung für den API-Zugriff anlegen

Diese App-Registrierung fungiert als „technischer Benutzer“, über den sich Ihr System gegenüber BusinessGPT authentifiziert.

#### 2.1. Neue App-Registrierung erstellen

1. Öffnen Sie das **Azure-Portal**, z. B. über:  
`https://portal.azure.com` (Beispiel-URL)
2. Navigieren Sie zu:  
**Microsoft Entra ID** → **App-Registrierungen**.
3. Klicken Sie oben auf „**Neue Registrierung**“.
4. Geben Sie einen **aussagekräftigen Namen** ein, z. B.:  
`BusinessGPT-API-Client`  
(Dies ist nur ein Beispielnamen. Verwenden Sie einen Namen, der in Ihrer Umgebung sinnvoll ist.)
6. Wählen Sie als unterstützte Kontotypen in der Regel:
7. „Nur Konten in diesem Organisationsverzeichnis“  
(Standard für interne Unternehmenslösungen)
8. Klicken Sie auf „**Registrieren**“.

Home > PrivateGPT1 | App registrations >

## Register an application

**Name**

The user-facing display name for this application (this can be changed later).

 ✓

### Supported account types

Who can use this application or access this API?

- Accounts in this organizational directory (PrivateGPT1 only - Single tenant)
- Accounts in any organizational directory (Any Microsoft Entra ID tenant - Multitenant)
- Accounts in any organizational directory (Any Microsoft Entra ID tenant - Multitenant) and personal Microsoft accounts (e.g. Skype, Xbox)
- Personal Microsoft accounts only

[Help me choose...](#)

### Redirect URI (optional)

We'll return the authentication response to this URI after successfully authenticating the user. Providing this now is optional and it can be changed later, but a value is required for most authentication scenarios.

Select a platform

Register an app you're working on here. Integrate gallery apps and other apps from outside your organization by adding from [Enterprise applications](#).

By proceeding, you agree to the [Microsoft Platform Policies](#)

Register

### Ergebnis:

- Sie sehen jetzt die Detailseite der neuen App-Registrierung.
- Notieren Sie sich die **Application (client) ID** – diese wird später im API-Tool als **Client ID** benötigt.

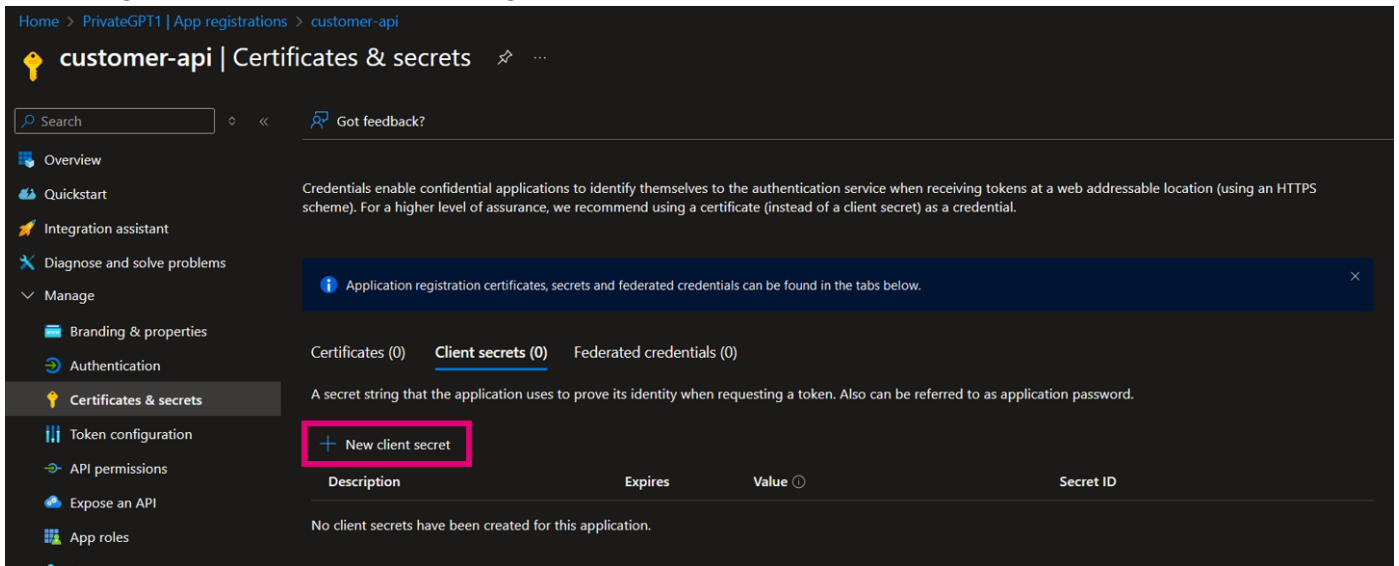
The screenshot displays the Microsoft Entra ID portal interface for configuring an application. The breadcrumb navigation at the top reads "Home > PrivateGPT1 | App registrations >". The application name "customer-api" is prominently displayed at the top left. Below the name is a search bar and a set of action buttons: "Delete", "Endpoints", and "Preview features". A left-hand navigation menu is visible, with "Overview" selected. The main content area is titled "Essentials" and contains the following configuration details:

- Display name : [customer-api](#)
- Application (client) ID :
- Object ID :
- Directory (tenant) ID :
- Supported account types : [My organization only](#)

## 2.2. Client-Secret (Passwort) erstellen

Damit sich diese App-Registrierung später gegenüber der Microsoft Identity Plattform authentifizieren kann, benötigen Sie ein **Client Secret** (ein geheimes Passwort).

1. Öffnen Sie in der neu angelegten App-Registrierung den Menüpunkt „Zertifikate & Geheimnisse“.
2. Unter „Clientgeheimnisse“ klicken Sie auf „Neues Clientgeheimnis“.



3. Geben Sie eine **Beschreibung** ein, z. B.:
4. `Secret für BusinessGPT-API-Zugriff` (Beispiel)
5. Wählen Sie eine **Gültigkeitsdauer** (z. B. 6 oder 12 Monate – je nach Sicherheitsrichtlinie in Ihrem Unternehmen).

6. Klicken Sie auf „**Hinzufügen**“.

## Add a client secret



Description

customer-api

Expires

365 days (12 months)



7. Kopieren Sie den **Wert** des Client Secrets direkt nach dem Anlegen und speichern Sie ihn sicher, z. B. in einem Passwortmanager.

#### Wichtiger Hinweis

Der Wert des Client Secrets wird nach dem Schließen der Seite nicht mehr vollständig angezeigt. Wenn er verloren geht, muss ein neues Secret erstellt werden.

### 2.3. Token-Endpunkt (Access Token URL) ermitteln

Diese URL benötigt Ihr API-Tool, um das OAuth2-Token abzurufen.

1. Klicken Sie in der gleichen App-Registrierung oben auf „**Endpunkte**“.
2. Suchen Sie den Eintrag „**OAuth 2.0 token endpoint (v2)**“.
3. Kopieren Sie die dort angegebene URL, z. B.:

```
https://login.microsoftonline.com/<TENANT-ID>/oauth2/v2.0/token
```

- <TENANT-ID> ist ein **Platzhalter** für Ihre Tenant-ID.

### 2.4. Zusätzliche App mit der „normalen“ BusinessGPT-App verknüpfen

Nun legen Sie fest, welche **Rolle** diese neue API-App in Bezug auf Ihre BusinessGPT-Anwendung hat.

#### Wichtiger Hinweis

Die Bezeichnungen der Menüpunkte können je nach Konfiguration und Sprache leicht variieren. Nutzen Sie bei Bedarf die Suchfunktion im Azure-Portal.

#### 2.4.1. „NORMALE“ BUSINESSGPT-APP-REGISTRIERUNG ÖFFNEN

1. Wechseln Sie im Azure-Portal wieder zu „**App-Registrierungen**“.
2. Suchen Sie nach der App-Registrierung, die Ihre **BusinessGPT-Hauptanwendung** repräsentiert.
3. Beispiel: `customer` (dies ist nur ein Platzhaltername).
4. Öffnen Sie diese App-Registrierung.

## 2.4.2. BERECHTIGUNGEN / ROLLEN FÜR DEN API-CLIENT EINRICHTEN

Je nach Konfiguration Ihrer Umgebung finden Sie die relevanten Einstellungen typischerweise in den Bereichen:

- „API-Berechtigungen“
- oder „Unternehmensanwendungen (Enterprise Applications)“ → „Benutzer und Gruppen“ / „Rollen“

**customer-api | API permissions**

Search Refresh | Got feedback?

Overview  
Quickstart  
Integration assistant  
Diagnose and solve problems  
Manage  
Branding & properties  
Authentication  
Certificates & secrets  
Token configuration  
**API permissions**  
Expose an API  
App roles  
Owners  
Roles and administrators  
Manifest  
Support + Troubleshooting

Granting tenant-wide consent may revoke permissions that have already been granted.

The "Admin consent required" column shows the default value for an application. If you are not an administrator, you will be used. [Learn more](#)

**Configured permissions**

Applications are authorized to call APIs when they are granted permissions. Applications must include all the permissions the application needs. [Learn more about permissions](#)

+ Add a permission ✓ Grant admin consent for PrivateGPT1

API / Permissions name	Type	Description
Microsoft Graph (1)		
User.Read	Delegated	Sign in and read user profile

To view and manage consented permissions for individual apps, as well as for all apps, see [Admin consent](#).

Typisches Vorgehen:

1. Fügen Sie die neue API-App-Registrierung (z. B. `BusinessGPT-API-Client`) als **berechtigte Anwendung** hinzu.
2. Weisen Sie dieser App die Rolle `api` zu.

## Request API permissions ✕

Select an API

Microsoft APIs **APIs my organization uses** My APIs

Apps in your directory that expose APIs are shown below

Name	Application (client) ID
customer	[REDACTED]

 **Wichtiger Hinweis**

Stellen Sie sicher, dass der API-Client die Rolle `api` erhält, damit ein Zugriff auf BusinessGPT per API möglich ist.

# Request API permissions

< All APIs

CU customer

What type of permissions does your application require?

Delegated permissions  
Your application needs to access the API as the signed-in user.

Application permissions  
Your application runs as a background service or daemon without a signed-in user.

Select permissions expand all

Start typing a permission to filter these results

Permission	Admin consent required
Other permissions (1)	
<input type="checkbox"/> admin ⓘ	Yes
<input checked="" type="checkbox"/> api ⓘ	Yes
<input type="checkbox"/> role ⓘ	Yes
<input type="checkbox"/> user ⓘ	Yes

> api

Add permissions Discard

- 240/265 -

© 2026 Telekom MMS

## 2.5. Optional: Separaten API-Client mit Admin-Rechten einrichten

Falls Sie ein System haben, das weitere **administrative Aktionen** via API durchführen soll (z. B. Konfigurationen, Management-Funktionen), bietet es sich an, eine **zweite** API-App-Registrierung mit einer **admin-Rolle** zu verwenden.

Vorgehen:

1. Wiederholen Sie die Schritte **2.1 bis 2.3** für eine zweite App-Registrierung, z. B. mit dem Namen:
2. `BusinessGPT-API-Admin` (Beispiel)
3. Weisen Sie dieser App in der BusinessGPT-Haupt-App-Registrierung statt `api` die Rolle `admin` zu.

### Wichtiger Hinweis

Verwenden Sie separate API-Clients für `api`- und `admin`-Berechtigungen, um das Prinzip der minimalen Rechte (Least Privilege) umzusetzen.

## 2.6. Admin-Zustimmung (Consent) erteilen





Viele Berechtigungen werden erst wirksam, wenn eine Administratorin oder ein Administrator im Tenant die Zustimmung (Admin Consent) erteilt hat.

1. Öffnen Sie die „normale“ BusinessGPT-App-Registrierung.
2. Wechseln Sie in den Bereich „API-Berechtigungen“ (oder einen ähnlich benannten Bereich).
3. Klicken Sie auf „Admin consent erteilen“, „Zustimmung erteilen“ o. ä.
4. Folgen Sie dem Dialog und bestätigen Sie die Freigabe.

Configured permissions

Applications are authorized to call APIs when they are granted permissions by users/admins as part of the consent process. The list of configured permissions should include all the permissions the application needs. [Learn more about permissions and consent](#)



+ Add a permission  Grant admin consent for PrivateGPT1

API / Permissions name	Type	Description	Admin consent req...	Status
▼ customer (1)				
api	Application	api	Yes	 Not granted for PrivateG... 
▼ Microsoft Graph (1)				
User.Read	Delegated	Sign in and read user profile	No	 Granted for PrivateGPT1 

To view and manage consented permissions for individual apps, as well as your tenant's consent settings, try [Enterprise applications](#).

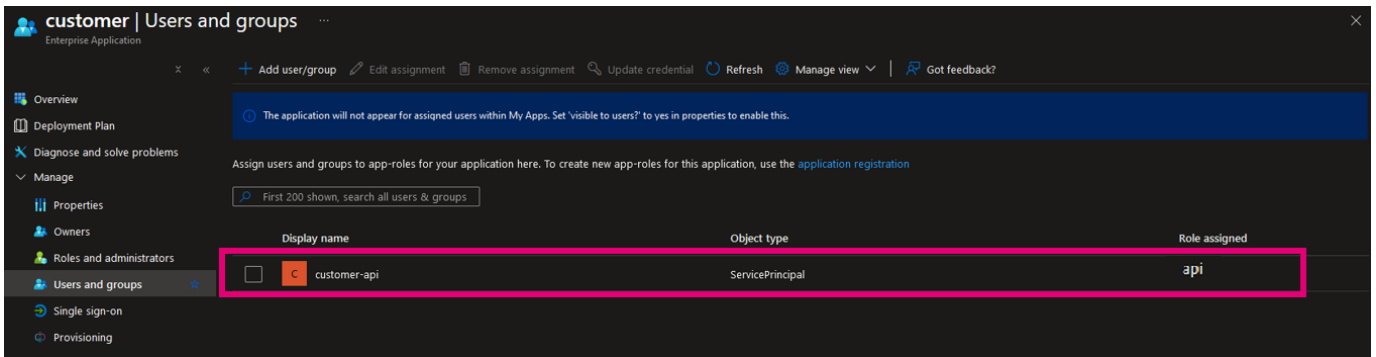
---

+ Add a permission  Grant admin consent for PrivateGPT1

API / Permissions name	Type	Description	Admin consent req...	Status
▼ customer (1)				
api	Application	api	Yes	 Granted for PrivateGPT1 

### Kontrolle:

- Unter „Benutzer und Gruppen“ bzw. in den **Unternehmensanwendungen** sollte die neue API-App-Registrierung als **Service Principal** sichtbar sein.
- Prüfen Sie, ob die erwartete Rolle (`api`) zugewiesen wurde.



### 4.2.4 3. OAuth2-Konfiguration im API-Tool (z. B. Postman)

Nachdem die Entra-ID-Konfiguration abgeschlossen ist, können Sie ein **Access Token** abrufen und API-Aufrufe an BusinessGPT senden.

#### 3.1. Beispielkonfiguration in Postman

*(alle Werte sind Platzhalter und müssen angepasst werden)*

1. Erstellen Sie in Postman eine neue **Collection** oder einen neuen **Request**.
2. Wechseln Sie im Request auf den Tab „**Authorization**“.
3. Wählen Sie als Typ: **OAuth 2.0**.
4. Füllen Sie die Felder wie folgt (Beispiele):
5. **Grant Type:** `Client Credentials`
6. **Access Token URL:**

```
https://login.microsoftonline.com/<TENANT-ID>/oauth2/v2.0/token
```

Ersetzen Sie `<TENANT-ID>` durch Ihre tatsächliche Tenant-ID.

#### 7. Client ID:

Die **Application (client) ID** Ihrer **API-App-Registrierung**, z. B.:

```
00000000-0000-0000-0000-000000000000 (Beispielwert)
```

#### 8. Client Secret:

Der zuvor erzeugte **Secret-Wert**, z. B.:

```
xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx (Beispiel)
```

#### 9. Scope:

Typische Form, abhängig von Ihrer Konfiguration:

```
api://<APP-REG-CLIENT-ID>/default
```

oder

```
<APP-REG-CLIENT-ID>/default
```

Ersetzen Sie `<APP-REG-CLIENT-ID>` durch die **Application (client) ID** Ihrer „normalen“ BusinessGPT-App-Registrierung.

#### 10. Client Authentication:

Wählen Sie z. B. „Send as Basic Auth header“ oder „In Request Body“ – beide Methoden werden von der Microsoft Identity Plattform in der Regel unterstützt.

11. Klicken Sie auf „**Get New Access Token**“ (oder einen entsprechend benannten Button).

Description Headers Query Body **Auth** Options

None Basic Bearer OAuth 1 **OAuth 2**

**Grant Type**

Resource Owner Password Credentials

Client Credentials

**Client ID**

[Redacted]

**Client Secret**

yow8Q~tfHy9aXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Set client credentials in the body

**Access Token URL**

https://login.microsoftonline.com/[Redacted]

Set client credentials in the body

Access Token URL

`https://login.microsoftonline.com/[REDACTED]`

Scope

`[REDACTED]/.default`

Token Prefix

Bearer

Token

`[REDACTED]`

Get Access Token

Wenn alle Angaben korrekt sind, wird Ihnen ein Access Token angezeigt, das Sie speichern und für Ihre Requests verwenden können.

 **Wichtiger Hinweis**

Wenn die Token-Anforderung fehlschlägt (z. B. Fehler 401/403), prüfen Sie bitte: Tenant-ID, Client ID, Client Secret, Scope und ob der Admin-Consent bereits erteilt wurde.

#### 4.2.5 4. Ersten Testaufruf an BusinessGPT senden

##### 4.1. Request-URL (Beispiel)

`https://customer.businessgpt.telekom.net/api/v2/chat/completions`

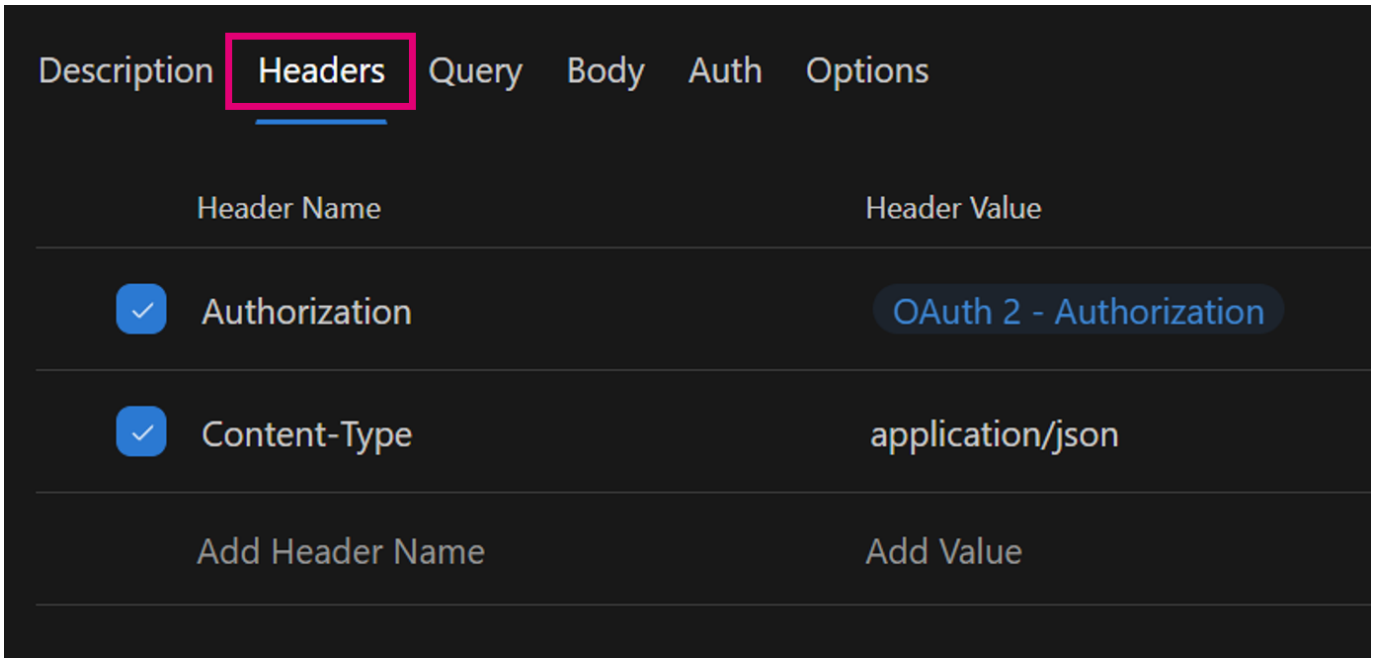
 **Wichtiger Hinweis**

Diese URL ist ein **Beispiel**. Verwenden Sie hier die für Ihre Umgebung bereitgestellte BusinessGPT-API-URL.

## 4.2. Header setzen

In Postman (oder einem ähnlichen Tool):

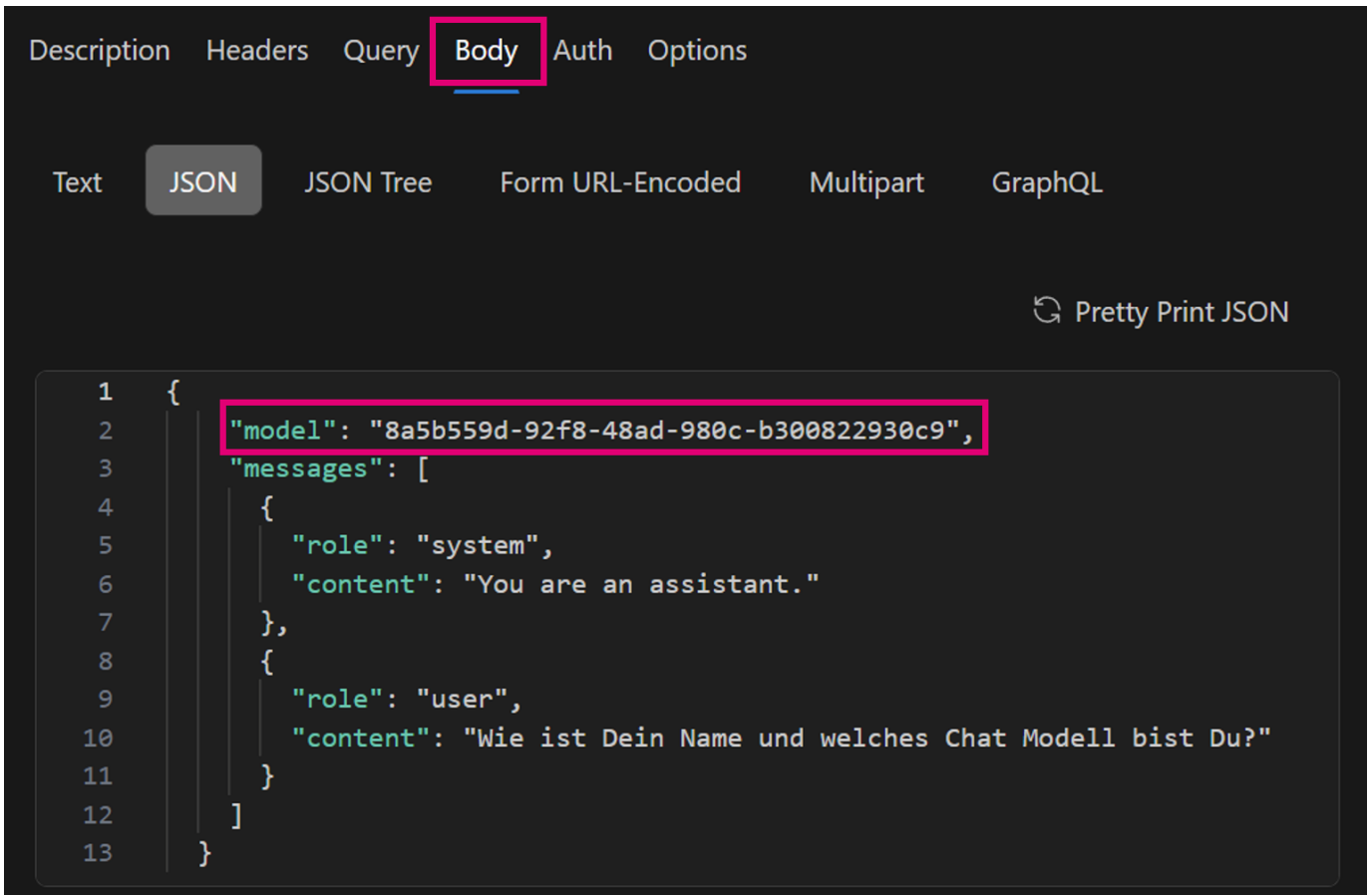
1. Wechseln Sie in Ihrem Request auf den Tab „**Headers**“.
2. Stellen Sie sicher, dass folgende Header gesetzt sind:
3. `Authorization: Bearer <access_token>`  
(In vielen Tools wird dies nach dem Tokenabruf automatisch gesetzt.)
4. `Content-Type: application/json`



## 4.3. Beispiel-Body für einen Chat-Aufruf

Nachfolgend ein **Beispiel** für einen Request-Body (alle IDs und Texte sind Platzhalter):

```
{
  "model": "8a5b559d-92f8-48ad-980c-b300822930c9",
  "messages": [
    {
      "role": "user",
      "content": "Erklären Sie mir in einfachen Worten, wie der API-Zugriff auf BusinessGPT funktioniert."
    }
  ]
}
```

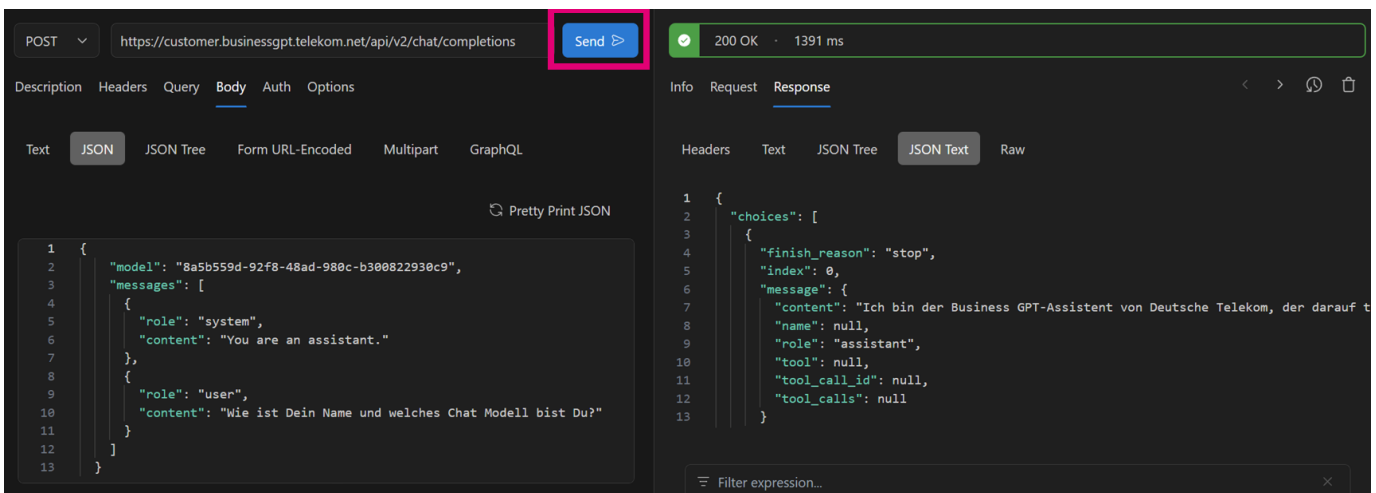


- `model`: Die Modell-ID aus Ihrer BusinessGPT-Instanz (siehe Kapitel 5).
- `messages`: Ein Array von Chat-Nachrichten, bestehend aus `role` (`user`, `assistant`, etc.) und `content`.

Klicken Sie auf „Send“.

Bei erfolgreichem Aufruf sollten Sie sehen:

- Statuscode: **200 OK**
- Im Response-Body:
- Eine vom Modell generierte Antwort
- Ggf. weitere Metadaten (abhängig von der konkreten API-Implementierung)

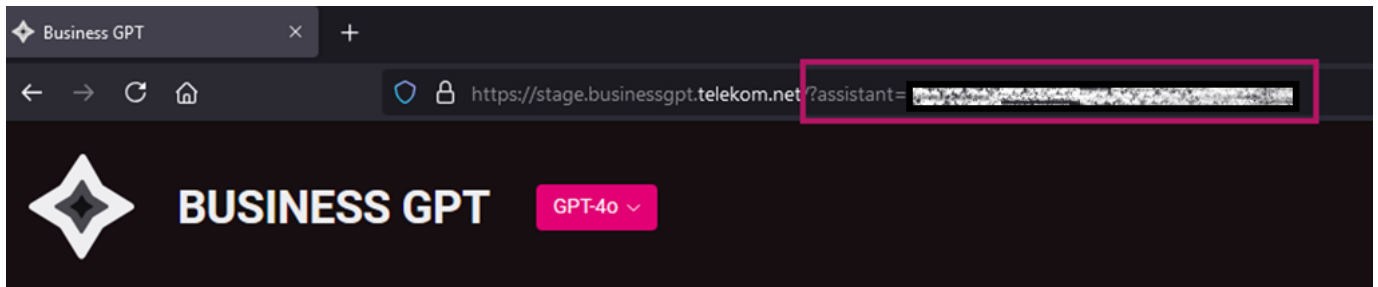


## 4.2.6 5. Modell-IDs (Feld `model`) in BusinessGPT ermitteln

Damit BusinessGPT weiß, **welches** Modell Sie ansprechen möchten (z. B. GPT-4o, Document Chat, Internet-Suche), müssen Sie die entsprechende **Modell-ID** kennen.

### 5.1. Modell-ID im BusinessGPT-Frontend herausfinden

1. Öffnen Sie die **BusinessGPT-Weboberfläche** in Ihrem Browser.
2. Wählen Sie das gewünschte Modell oder einen Ihrer **eigenen Assistenten** aus.
3. Schauen Sie sich die **URL** in der Adresszeile des Browsers an:
4. Kopieren Sie diese ID und verwenden Sie sie im API-Request im Feld `"model"`.



#### Assistenten-IDs

Eigene Assistenten, die Sie in BusinessGPT erstellen, erhalten ebenfalls eine eindeutige ID. Diese ID können Sie genauso wie die Standard-Modelle im Feld `model` verwenden.

### 5.2. Beispielhafte Modell-IDs (nur zur Illustration)

Die folgenden IDs dienen lediglich als **Beispiele**:

Zweck / Modell	Beispielhafte Modell-ID
Document Chat	1226a08b-7fad-4c99-aeac-b695ea5b250b
GPT-4o	8a5b999d-92f8-48ag-980c-b300822930c9
GPT-4o mini	1691aea7-ed0b-46ce-9aef-dd729dbdbecc
GPT-o1	7406483d-1dde-49be-9bc8-6703h948590b
GPT-o3 (mini)	be3ad08c-d906-4308-afbb-e348493e3458
DALL·E	dfcf580b-7ed1-47d5-843d-9ce818349e24
Internet-Suche	8b16032e-e39d-401c-b3bd-3366e73511d5

#### weis

Diese IDs sind **Beispiele**. Verwenden Sie die Modell-IDs, die in Ihrer eigenen BusinessGPT-Instanz angezeigt werden.

## 4.2.7 6. Typische Fehler und Tipps zur Fehlersuche

### 6.1. Häufige Probleme bei der Token-Anforderung

- 401 Unauthorized / 403 Forbidden

- Scope ist falsch oder passt nicht zur App-Registrierung
- Client ID oder Client Secret sind fehlerhaft
- Admin-Consent wurde noch nicht erteilt
- „invalid\_grant“ oder „invalid\_client“
- Client Secret ist abgelaufen oder wurde falsch kopiert
- Tenant-ID ist nicht korrekt

## 6.2. Häufige Probleme bei API-Aufrufen

- **404 oder 400 (Bad Request)**
- Falsche oder nicht existente Modell-ID
- Falsche oder unvollständige URL
- **Unerwartete Fehlermeldungen**
- Falsche Header (z. B. fehlendes `Content-Type: application/json`)
- Fehler in der JSON-Struktur des Request-Bodys

## 6.3. Empfehlung für eine kurze Checkliste

Wenn etwas nicht funktioniert, prüfen Sie bitte zuerst:

1. Ist die **Tenant-ID** korrekt?
2. Haben Sie die richtige **Client ID** und das korrekte **Client Secret** verwendet?
3. Ist der **Scope** korrekt konfiguriert?
4. Ist die gewünschte **Rolle** (z. B. `user` oder `admin`) zugewiesen?
5. Wurde der **Admin-Consent** bereits erteilt?
6. Verwenden Sie die richtige **BusinessGPT-API-URL** und eine gültige **Modell-ID**?

## 5. Häufig gestellte Fragen

### 5.1 Allgemeine Fragen zum Chat

#### Wie kann ich im Chatfenster eine neue Zeile beginnen?

Durch Tastenkombination Shift + Enter kann eine neue Zeile im Chatfenster begonnen werden.

#### Wie kann ich Bilder und Dateien in den Chat einfügen?

Dateien können per Drag & Drop, Copy & Paste aus der Zwischenablage oder per Upload (über das Plus-Symbol) dem Chat hinzugefügt werden.

#### Was sind die Unterschiede zwischen Standard-, Mini- und Nano-Modellen?

Standard-, Mini- und Nano-Modelle unterscheiden sich vor allem in folgenden Punkten:

##### • Leistungsfähigkeit (Qualität der Ergebnisse)

Standard-Modelle sind in der Regel am leistungsstärksten und liefern die besten Ergebnisse, z.B. bei komplexem Textverständnis, logischen Schlussfolgerungen oder kreativen Aufgaben.

Mini-Modelle bieten ein gutes Mittelmaß aus Qualität und Effizienz.

Nano-Modelle sind stärker auf Einfachheit und Geschwindigkeit optimiert und eignen sich vor allem für weniger komplexe Aufgaben.

##### • Geschwindigkeit und Ressourcenverbrauch

Standard-Modelle benötigen meist mehr Rechenressourcen und sind etwas langsamer in der Ausführung, bieten dafür aber höhere Genauigkeit.

Mini-Modelle sind deutlich schneller und ressourcenschonender und damit ideal für Anwendungen mit vielen Anfragen.

Nano-Modelle sind auf maximale Geschwindigkeit und sehr geringe Ressourcenanforderungen optimiert (z.B. für einfache Workflows oder eingebettete Systeme).

##### • Kosten pro Anfrage

Standard-Modelle sind in der Regel am teuersten pro Anfrage, da sie mehr Rechenleistung benötigen.

Mini-Modelle sind kostengünstiger bei weiterhin guter Qualität.

Nano-Modelle sind am günstigsten und daher gut geeignet für Szenarien mit sehr hohem Anfragevolumen oder strengen Budgetvorgaben.

##### • Typische Anwendungsfälle

Standard-Modelle passen gut zu komplexen Use-Cases, etwa anspruchsvolle Fachtexte, ausführliche Analysen oder kreative Generierung.

Mini-Modelle eignen sich gut für typische Alltagsanwendungen, z.B. Chatbots, Support-Automatisierung oder interne Tools.

Nano-Modelle sind ideal für einfache, klar umrissene Aufgaben, z.B. kurze Textbausteine, einfache Klassifikation oder schnelle Hilfsfunktionen innerhalb von Anwendungen.

Zusammengefasst bieten die verschiedenen Modellgrößen unterschiedliche Kompromisse zwischen **Qualität**, **Geschwindigkeit** und **Kosten**. Welches Modell am besten geeignet ist, hängt von den konkreten Anforderungen, der gewünschten Antwortqualität und den verfügbaren Ressourcen in Ihrer Anwendung ab.

#### Was bedeutet 'Token-Limit' genau?

Ein Token ist ein Baustein aus Wörtern oder Zeichen. Modelle haben eine maximale Anzahl Token, die sie in einer Anfrage verarbeiten können (Input + Output).

Beispiel: Der Satz „Hallo, wie kann ich helfen?“ besteht aus ca. 7 Token.

Je größer das Token-Limit, desto mehr Kontext kann ein Modell behalten – wichtig für lange Dokumente oder Gespräche.

#### Welches Modell ist das 'beste'?

Es gibt kein „bestes“ Modell – nur das passendste. Die Auswahl hängt vom Einsatzzweck ab: Zur Bestimmung des geeigneten Modells empfehlen wir unseren [Model-Guide](#)

### Warum erhalte ich nicht die Antworten oder die Antwortqualität, die ich erwarte. Was kann die Ursache sein?

Je besser die Anfrage an die KI ist, umso besser ist auch das Ergebnis. Wir empfehlen hierzu unserem Leitfaden für gute Prompts zu folgen: [Prompting-Guide](#)

### Wann sind neue Modelle in Business GPT verfügbar?

Modelle, insbesondere OpenAI Modelle, sind zuerst in den USA verfügbar. In der Regel dauert es ein paar Wochen bis diese Modelle dann auch in Europa verfügbar sind. Sobald die Modelle in Europa verfügbar sind, stellen wir diese in Business GPT zur Verfügung.

### Was sind Halluzinationen – und wie gehe ich damit um?

Halluzinationen sind realistisch klingende, aber falsche Aussagen eines Modells.

Beispiele: erfundene Quellen, Zahlen, Zitate oder Fakten.

→ Besonders bei kritischen Themen (z.B. Recht, Medizin, Finanzen) ist es wichtig, Ausgaben zu validieren und mit Fachwissen zu prüfen.

### Was sind die Unterschiede zwischen 'schnellen' und 'langsamen' Modellen?

Ja, absolut. Man unterscheidet in der Regel zwischen hochoptimierten, kompakten Modellen und großen, rechenintensiven Modellen:

- **Schnelle Modelle:** Sind auf maximale Geschwindigkeit und Effizienz getrimmt. Sie reagieren fast in Echtzeit und sind ideal für Dialoge oder einfache Textaufgaben.
- **Langsamere Modelle:** Nutzen oft mehr Parameter oder zusätzliche "Denk-Schritte" (*Reasoning*), um komplexe Probleme zu lösen. Das benötigt mehr Zeit, führt aber bei schwierigen Logikrätseln zu präziseren Ergebnissen.

→ Die Wahl des Modells ist immer ein Abwägen zwischen Antwortgeschwindigkeit und Tiefe der Analyse.

### Ein schnelleres Modell schlechter als ein langsames?

Nicht unbedingt. Kleinere Modelle wie `GPT-5.1-mini` oder `GPT-4.1-nano` liefern bei Standardaufgaben oft sehr gute Ergebnisse – und das bei geringerem Ressourcenverbrauch.

→ Sie sind besonders effizient bei einfachen bis mittleren Anforderungen.

### Wie viel Text kann ich im Prompt verwenden?

Ein Prompt (Texteingabe in den Chat) kann maximal 100.000 Zeichen enthalten. Übersteigt die Anfrage diese Anzahl, wird eine Fehlermeldung angezeigt. Für größere Textmengen wird empfohlen, diese bspw. als Textdatei abzuspeichern und diese .txt-Datei an den Chat anzuhängen.

## 5.2 Fragen zum Chat mit Dokumenten/internem Wissen

### Warum sehe ich keine Container zur Auswahl im Dokumenten-Chat. Woran liegt das?

Werden keine Container zur Auswahl angezeigt, liegt das entweder an fehlenden Gruppenfreigaben für die gewünschten Container oder an fehlenden Gruppenmitgliedschaften im EntraID. Hierzu bitte an den zuständigen Administrator wenden.

### Warum werden Antworten beim Dokumenten-Chat nicht die richtigen Dokumente als Quellen benannt. Woran kann das liegen?

Ursache könnte sein, dass das Dokument bzw. der zugehörige Dokumenten-Container nicht richtig indexiert wurde. Lösung: das betreffende Dokument in einen anderen Container verschieben und anschließend erneut in den gewünschten Ziel-Container zuordnen.

**Kann ich selbst Dateien hochladen, um mit diesen zu chatten?**

Ja, mit dem Business GPT Paket "L" können eigene Dokumente hochgeladen werden. Diese stehen dann als temporäre Dokumente für 24h zur Verfügung bevor diese automatisch gelöscht werden.

**Kann Business GPT die Anzahl von Dokumenten ermitteln, z.B. in einem Container?**

Nein, Business GPT kann nicht zählen. Die Beantwortung von Fragen zu internem Wissen folgt einem deterministischen Ansatz.

**Was ist der Unterschied zwischen dem Chat mit einem Dokumenten-Container und einem direkt am Chat angehängten Dokument?**

Der wesentliche Unterschied besteht darin, dass bei Dokumenten, die direkt an den Chat angehängt sind, der gesamte Dokumenteninhalte dem Modell zur Verfügung gestellt wird, während beim Chat mit Containern ein Vektorsuche erfolgt, um relevante Inhalte zu finden. Auf diese Weise können große Mengen von Dokumenten durchsucht werden, aber möglicherweise wird nicht jeder Dokumententeil für jede Antwort berücksichtigt.

**Ich erhalte im Chat eine Fehlermeldung, dass das Token Limit überschritten wurde, wenn ich eine Datei angehängt habe. Was bedeutet das?**

Wenn sie diese Fehlermeldung erhalten, ist das Dokument oder die Gesamtanzahl aller Dokumente, die Sie an den Chat anhängen wollen, zu groß. KI-Modelle verfügen über ein sogenanntes Kontextfenster, was festlegt, wie viele Zeichen ein Modell maximal verarbeiten kann. Wenn Sie ein Dokument an den Chat anhängen, wird der Inhalt des Dokumentes extrahiert und somit der gesamte Text dem Modell zur Verfügung gestellt. Sofern das in Summe zu viele Zeichen sind, wird Ihnen diese Fehlermeldung angezeigt. Entfernen Sie dann entweder eines oder mehrere Dokumente aus dem Chat oder wechseln Sie auf eine anderes Modell mit größerem Kontextfenster.

**Was passiert, wenn ein Dokument hochgeladen wird, was bereits im System, z.B. in einem Container, vorhanden ist?**

Business GPT erkennt Duplikate und stellt sicher, dass Dokumente, die bereits vorhanden sind, nicht nochmals im Datenbestand gespeichert werden. Stattdessen sorgt Business GPT bei der Erkennung eines Duplikats dafür, dass lediglich eine Referenz des Dokuments erstellt und verwendet wird. Wird bspw. ein Dokument in einen Container hochgeladen, welches bereits in einem anderen Container existiert, wird dieses Dokument per Referenzierung im Container nur zugeordnet, ohne dass eine weitere physische Kopie des Dokuments gespeichert wird.

**Ist die Dateianzahl in einem Dokumenten-Container begrenzt?**

Nein, Sie können unbegrenzt viele Dateien in einem (Unternehmens-) Container speichern. Die Anzahl der Dateien in einem Container sollte durch die Art und Weise wie Sie den Container später verwenden möchten, bestimmt werden. Container-Inhalte können Sie beispielsweise anwendungsfallsspezifisch voneinander trennen.

Bei persönlichen Dokumentencontainern ist die Maximalanzahl davon abhängig, welche Einstellungen Ihr Administrator vorgenommen hat.

**Ist die Anzahl von Container begrenzt?**

Nein, Sie können beliebig viele Container erstellen.

## 5.3 Fragen zum Datenschutz

---

**Wo finde ich die Datenschutzregeln und Nutzungsbedingungen?**

Die Links zu den Datenschutzhinweisen und Nutzungsbedingungen können über die geöffnete Seitennavigation geöffnet werden.

**Wird die KI mit hochgeladenen Dokumenten trainiert?**

Nein, hochgeladene Dokumente werden nicht zum Training der KI genutzt. Diese Daten werden lediglich in einem sogenannten "Retrieval-Augmented Generation" System, kurz RAG System, verarbeitet. Dabei werden die Inhalte der Dokumente in kleinere verarbeitbare Einheiten zerlegt ("Chunking") und anschließend für schnelle Suchanfragen gespeichert ("Indexing"). Diese Daten werden sicher durch Business GPT gespeichert und gelangen nicht nach außerhalb.

**Können Dokumente und Assistenten pro Abteilung voneinander separiert werden, sodass zum Beispiel Mitarbeiter der Technik nicht auf Daten der Geschäftsleitung zugreifen können?**

Ja, Business GPT erlaubt dies durch Anwendung von Nutzergruppenberechtigungen. Dokumente und Assistenten können so nur für bestimmte Nutzergruppen freigegeben werden.

## 5.4 Fragen zur Anmeldung/Login

---

**Anmeldung mit den Rollen "user" oder "admin" funktioniert nicht, obwohl die Anwendungsrollen im Entra ID einer Gruppe zugewiesen wurden. Was könnte die Ursache sein?**

Falls im EntraID verschachtelte Gruppen (Gruppe in Gruppe) verwendet werden, könnte dies der Grund sein, da Entra ID nur Mitglieder der direkt zugewiesenen Gruppe berücksichtigt. Verschachtelte Gruppen werden hierbei nicht aufgelöst, da die Token-Generierung in Entra ID nicht rekursiv durch die Gruppenshierarchie navigiert. Für Mitglieder in einer untergeordneten Gruppe wird demnach die Anwendungsrolle nicht an BusinessGPT übermittelt, wodurch die Anmeldung als Admin oder User fehlschlägt.

## 5.5 Fragen zu Kosten und Verbrauch

---

**Was passiert, wenn das Gesamtbudget aufgebraucht ist?**

Wenn das Gesamtbudget aufgebraucht ist, sind keine kostenverursachenden Aktionen mehr in Business GPT möglich. Dazu zählen das Chatten mit KI-Modellen wie GPT4o sowie die Indexierung von Dokumenten, nachdem diese hochgeladen wurden. Zusätzliches Budget kann aber problemlos jederzeit nachbestellt werden. Weiterführende Informationen finden Sie hier: [Budget nachbestellen](#)

**Wie genau wird das monatliche Prepaid-Budget verbraucht?**

Das Prepaid-Budget wird zum einen durch Tokenkosten verbraucht. Dabei entstehen Kosten bei der Anfrage an die KI, sogenannte Input-Token, und durch die Antwort der KI, sogenannte Output-Token. Je Modell, z.B. GPT5.1 oder GPT4.1, sind diese Kosten unterschiedlich. Zum anderen entstehen Kosten bei der Verarbeitung von Daten, die Business GPT als "internes Wissen" zur Verfügung gestellt werden, z.B. durch den Upload von Dokumenten. Dabei entstehen durch die Verarbeitung beim Upload sogenannte Indexierungskosten.

Alle Kostenarten stehen transparent in Business GPT in einer [Verbrauchsübersicht](#) zur Verfügung.

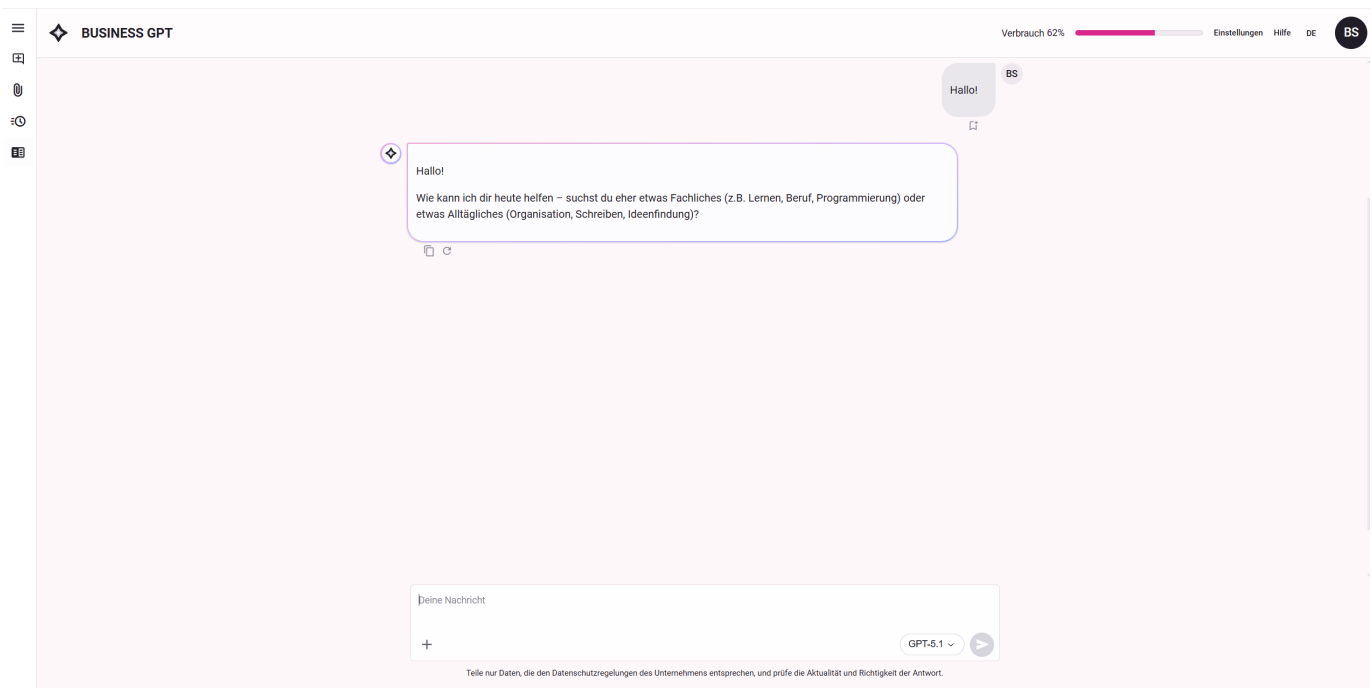
## 6. Changelog

Business GPT wird stetig weiterentwickelt und verbessert. Alle Neuerungen werden hier beschrieben.

### 6.1 26. März 2026

Der Chat-Bereich von Business GPT wurde vollständig überarbeitet und optimiert.

- Die Assistenten-Auswahl befindet sich nun direkt im Texteingabefeld.
- Neue Konversationen erstellen Sie fortan über die Schaltfläche in der Seitennavigation.



Der Standard-Assistent **Dokumenten-Chat** wurde umbenannt in **Unternehmenswissen**, um der Funktionsweise dieses Assistenten besser gerecht zu werden.

#### Fehlerbehebungen

- Dateianhänge im Chat mit .docx Dateien führten zum Fehler bei bestimmten Zeichen im Dateinamen. Der Fehler wurde behoben.
- Beim Einfügen von großen Textmengen in das Texteingabefeld im Chat kam es zu undefinierten Fehlern. Die Fehlerbehandlung wurde optimiert. Beim Überschreiben von 100.000 Zeichen Text wird eine nutzerfreundliche Fehlermeldung angezeigt.



#### GPT4o wurde eingestellt

Bitte beachten Sie, dass GPT4o und GPT4o-mini eingestellt wurden und somit nicht mehr zur Verfügung stehen. Ihre Assistenten, die bis dato die beiden Modelle genutzt haben, wurde automatisch auf GPT4.1 bzw. GPT4.1-mini umgestellt.

### 6.2 18. März 2026

Chats in Business GPT sind nun **persistent** und werden für maximal 180 Tage gespeichert. Ihre Chatverläufe gehen somit nach Sitzungsende nicht mehr verloren.

#### Neuere Verbesserungen

- In den Einstellungen wurden zur schnelleren und effizienteren Navigation Breadcrumbs (Pfadnavigation) eingeführt.
- Globale Einstellungen können nun über `Verwaltung` in den Einstellungen verwaltet werden. Der separate Menüeintrag "Präferenzen" wurde entfernt.
- Beim Kopieren von Chat-Antworten und Einfügen in externe Textverarbeitungstools wie z.B. Microsoft Word bleibt die Formatierung (z.B. Fettformatierung) erhalten.

#### Reparaturen

- Mathematische Formeln wurden im Chat nicht korrekt aufgelöst.
- Die Bildbearbeitung mit Image-1.5 verursachte einen Fehler. Bilder können nun auch bearbeitet werden.
- Deaktivierte Assistenten wurden im Chatbereich für Administratoren angezeigt.

## 6.3 2. März 2026

Mit **Unternehmens-Prompts** können Prompts für Nutzer:Innen nun auch zentral verwaltet und durch Administratoren zur Verfügung gestellt werden. Damit sind Unternehmens-Prompts die perfekte Ergänzung zu den persönlichen und privaten Prompts.

Unter [Unternehmens-Prompts](#) finden Sie alle Details zur Verwaltung und Verwendung von Unternehmens-Prompts.

## 6.4 18. Februar 2026

Business GPT bietet nun mit **Dokumenten-Containern für Nutzer:Innen** die perfekte Ergänzung zu den bisherigen zentral verwalteten (Unternehmens-) Containern. Nutzer:Innen können nun ihre eigene, private Wissenssammlung aufbauen und in Business GPT verwenden. Um etwaigen Compliance-Anforderungen Rechnung zu tragen, kann das Feature explizit aktiviert und weitere Einstellungen im Einstellungsbereich vorgenommen werden.

Alle Detailinformationen finden Sie in der Online-Hilfe unter [Dokumenten-Container](#).

Das Bildmodell **Dall-E** wird eingestellt. Mit dem Modell **GPT Image-1.5** steht ein leistungsfähiger Nachfolger zur Verfügung.

#### GPT Image-1.5

Bitte beachten Sie, dass bis auf Weiteres Image-1.5 nur nach expliziter Nachfrage durch Sie als Kunden aktiviert wird. Wenden Sie sich dazu bitte an unser [Support-Team](#)

#### Neuere Verbesserungen

- Die Verbrauchsübersicht bietet nun mehr Transparenz hinsichtlich der Aufschlüsselung der Einzelkosten und Verbrauchswerte.
- Der Menüpunkt `Benutzereinstellungen` wurde entfernt. Stattdessen gelangen alle Nutzer:Innen über den Menüpunkt `Einstellungen` zu den für sie zugänglichen Inhalten.

#### Reparaturen

- Behebung eines Fehlers beim Upload von Dokumenten mit zu langen Dateinamen in den Chat.
- Behebung eines Fehlers beim Download von Dokumenten, bei dem ein Punkt in der Dateiondung hinzugefügt wurde.
- Behebung eines Fehlers, bei dem CSV-Dateien in Firefox nicht hochgeladen werden konnten.
- Behebung eines Fehlers, bei dem die Parameter `Temperatur` und `Top-P` nach Änderung nicht die geänderten Werte zeigten.
- Behebung eines Fehlers beim direkten Referenzieren von Dateinamen im Dokumenten-Chat, wonach die Datei nicht gefunden wurde.

## 6.5 15. Januar 2026

---

Mit diesem Release wurden bekannte Fehler in Business GPT behoben.

### Fehlerbehebungen

- Behebung eines Fehlers, bei dem nach der erstmaligen Anmeldung in Business GPT oder beim Start einer neuen Nutzersitzung in einem neuen Browser keine Assistentenauswahl vorausgewählt war. Dadurch kam es beim Start eines Chats zu einer Fehlermeldung.
- Behebung eines Fehlers, bei dem nach Ablauf der Nutzersitzung eine Fehlermeldung angezeigt wurde und weder ein Neuladen der Seite noch eine Navigation im Browser zurück zum Erfolg führte.
- Behebung eines Fehlers, bei dem die Verwendung von Direktlinks zu Assistenten nicht den im Link enthaltenen Assistenten auswählte, sondern den vom Nutzer zuvor gewählten Assistenten in der Applikation.

## 6.6 17. Dezember 2025

---

Mit **GPT5.1** steht ein weiteres OpenAI Modell zur Verfügung. GPT-5.1 bietet ein verbessertes Kontextverständnis, präzisere Antworten und höhere Zuverlässigkeit bei komplexen Aufgaben. Im Vergleich zu GPT-5 liefert das Modell konsistentere Ergebnisse und geht strukturierter mit umfangreichen oder mehrstufigen Anforderungen um.

Darüber hinaus können nun Diagramme und Grafiken mit Business GPT erstellt werden.

- GPT5.1 steht als Standardmodell zur Verfügung und kann ebenso für individuelle Assistenten verwendet werden.
- Durch gezielte Anweisungen kann Business GPT Grafiken erzeugen. Weitere Details dazu finden Sie in der Online-Hilfe unter [Diagramme](#)

### Neuere Verbesserungen

- Die Modelle für die Assistenten **Bild Generator** und **Internet Suche** wurden auf GPT4.1 aktualisiert
- Ein Problem wurde behoben, bei dem Business GPT den Parameter `strict` in Function-Calling-Requests fälschlich als aktiv interpretiert hat. Requests mit nicht-striktem Function-Schema akzeptieren nun korrekt zusätzliche Properties.
- Dokumente, bei denen es während des Indexierungsvorgangs zu Fehlern gekommen ist, können nun ohne vorheriges Löschen und Neu-Upload direkt neu indexiert werden.

## 6.7 28. November 2025

---

Mit diesem Update von Business GPT steht die **Prompt Bibliothek** zur Verfügung.

- Prompts können nutzerindividuell verwaltet und wiederkehrend im Chat verwendet werden.
- Alle Details zum Umgang mit Prompts in der Prompt Bibliothek finden Sie in unserer Online-Hilfe: [Prompt Bibliothek](#)

### Neuere Verbesserungen

- Das Scrollverhalten bei der Generierung von Antworten wurde verbessert. Während Business GPT noch eine (längere) Antwort generiert, kann nun bereits nach unten gescrollt werden. Bisher musste man auf die Fertigstellung der Antwort warten.
- Der Upload von großen Dokumentenmengen in Container wurde optimiert, sodass nun auch eine große Anzahl von Dokumenten ohne Fehler oder Abbrüche in Container geladen werden können.
- Rechteckige Logos werden nun ohne Verkleinerung im Chat-Bereich angezeigt.

## 6.8 12. November 2025

Business GPT verfügt mit diesem Update über zahlreiche neue Features zu **Chat mit Dokumenten**, **Function Calling** und **Integrationen**.

- Dokumente, die im Chat hochgeladen und angehängt werden, werden nun vollständig in den Modelkontext übergeben, sodass der gesamte Dokumenteninhalte für den Chat und somit die Beantwortung von Fragen durch die KI genutzt werden kann
- Dokumente können zudem nun bei jedem Modell/Assistenten angehängt werden - dies war bisher auf den Assistenten "Dokumenten Chat" und individuelle Assistenten mit zugewiesenen Container beschränkt.
- Weiterführende Informationen zum Umgang mit Dokumenten im Chat und Containern finden Sie in unserer Online-Hilfe: [Dokumente im Chat](#) und [Chat mit Dokumenten](#)
- Mit **Function Calling** verfügen Assistenten nun über die Möglichkeit auf externe Systeme zuzugreifen, bspw. für den Abruf von Daten. Dazu können individuelle Funktionsaufrufe definiert werden, um externe API anzusprechen.
- Detaillierte Informationen zur Konfiguration und Handhabung finden Sie unter [Assistenten und Function Calling](#) und [Function Calling Leitfaden](#)
- Mit der **Jira Integration** kann **Jira Datacenter** an Business GPT als Importdatenquelle angebunden werden. Auf diese Weise kann mit Wissen in Jira-Projekten gechattet werden - analog zum Chat mit Dokumenten. Details zur Einrichtung und Nutzung von Jira Integrationen finden Sie unter [Integrationen](#) und [Jira Integration](#)

### Wichtige Änderungen

- Bei der Nutzung von Assistenten mit GPT5-Modell und angepassten Temperatur- und Top-P-Werten kam es im Chat zu einem unbekanntem Fehler. Da die GPT5-Modelle diese Parameter nicht unterstützen, wurde die Assistenten-Konfiguration angepasst, womit Temperatur und Top-P für GPT5 (mini, nano) nicht zur Verfügung stehen.

## 6.9 26. September 2025

Mit diesem Update steht mit GPT5 das derzeit neueste OpenAI-Modell zur Verfügung.

- Die Modelle [GPT5](#) (nur auf Nachfrage), [GPT5 \(nano\)](#) und [GPT5 \(mini\)](#) stehen zur Verfügung

### Wichtige Verbesserungen

- Die Assistentenerstellung wurde überarbeitet.
- Alle Einstellungen und Eigenschaften wurden neu designed und angeordnet.
- Das Eingabefeld für die Assistentenanweisungen (System Prompt) kann nun vergrößert werden für mehr Komfort.
- Dokumente und Container können in Nutzer- und System-Prompts (für Assistenten) nun direkt referenziert werden (siehe auch [Referenzierung](#))

## 6.10 11. September 2025

Mit dem neuesten Update wurde Business GPT um zahlreiche neue Features und Verbesserungen ergänzt.

- Im Chat-Bereich stehen alle Optionen wie der **Upload von Dateien**, **Chat Import/Export** sowie der **Containerfilter** für den Dokument-Chat in einem separaten Menü zur Verfügung.
- Der bisherige Container-Filter im linken Seitenmenü wurde entfernt. Wissen kann nun direkt im neuen Menü gefiltert werden (siehe auch [Wissen-filtern](#))
- Individuelle Assistenten können nun auch optional Bilder generieren. Dazu kann die Bildgenerierung mit Dall-E-3 im Bedarfsfall für den Assistenten aktiviert werden.

## Neuere Verbesserungen

- Im Styling-Bereich steht nun ein neutraler Hintergrund ohne Musterung zur Verfügung.
- Die Container-Administration wurde einem Facelift unterzogen und erscheint im neuen Look & Feel. Zudem wurde die Farbauswahl für Container erweitert.
- Die Erkennung von umfangreichen Exceldateien wurde verbessert. Das Chunking wurde von einem simplen, starren Split in Markdown-Tabellen zu einem flexiblen, kontext- und metadatenreichen Prozess erweitert, der speziell für RAG-Usecases mit LLMs optimiert ist. Tabellen werden verständlich beschrieben, kontextualisiert und semantisch durchsuchbar gemacht, anstatt nur als reine Datenblöcke vorzuliegen.
- Verbesserungen im Detail:
  - Chunking erfolgt struktur- und metadatenbewusst: pro Abschnitt wird festgehalten, welche Zeilen enthalten sind
  - Jeder Chunk erhält eine eindeutige Nummerierung und Metadaten über seine Position in der Tabelle.
  - Tabellenchunks behalten explizit Spalten-Header, Spaltenbeschreibungen, Datentypen und Beispielwerte.
  - Zusätzlich werden Kategorisierungen erkannt (z. B. ob eine Spalte kategorisch ist).
  - Erweiterte Metadaten pro Chunk wurden eingeführt.
  - Jeder Chunk enthält nicht nur die Tabelle in Markdown, sondern auch Zusammenfassungen, Beispieldaten pro Spalte zur semantischen Orientierung, Query Guidance (Hinweise, wie man die Tabelle abfragen kann), Searchable Content (Extra-Text, der alle Werte im Format Spalte:Wert kombiniert, damit die Retrieval-Suche viel besser funktioniert)
- Verbesserte Extraktion von PDF-Inhalten und damit bessere Präzision beim Retrieval und Chat mit PDF-Dokumenten. Im Detail:
  - verbesserte Unicode-Normalisierung
  - dadurch weniger Dubletten im Embedding / Vektorraum.
  - optionales Dokument-weites Chunking (statt starrer Seiten)
  - Absätze, Tabellen oder Listen, die über Seiten umbrechen, bleiben intakt.
  - Chunks sind semantisch sinnvoller (weniger abgeschnittene Sätze).
  - intelligente Seitenzuordnung über Page-Mapping und Fuzzy Matching
  - Erhöht Vertrauen und Nachvollziehbarkeit, da „Belege“ sauber verlinkt werden können.
  - robustes Handling von leeren oder fehlerhaften Seiten und dadurch
  - Stabile Verarbeitung auch bei fehlerhaften PDFs.
  - Keine „Geister-Chunks“ mit leerem Inhalt, die das Retrieval stören würden.

## Reparaturen

- Der auftretende Indexing-Fehler bei einer Vielzahl von hochgeladenen Dokumenten wurde behoben.
- Beim Einfügen von Assistenten-URLs (z.B. <https://env.businessgpt.telekom.net/assistant?assistant=7326a08b-7fad-4c99-aeac-b695ee5b250b>) im Browser wurde nicht der korrekte Assistent ausgewählt. Dies wurde behoben.
- Beim Wechsel zwischen Assistenten, denen Dokumenten-Container zugeordnet sind, wurde nach der Auswahl des jeweils anderen Assistenten die Container-Filterauswahl nicht aktualisiert. Dieser Fehler tritt nicht mehr auf.
- Bei mehreren aufeinanderfolgenden Quellenreferenzen wurden diese nicht korrekt aufgelöst, sondern teilweise nur als UUID-Verweise ausgegeben. Dies wurde behoben, sodass derartige Referenzen zum Beispiel als [1] [2] [3] angezeigt werden.
- Tabellen mit einer großen Anzahl an Spalten wurden in Antworten nur unvollständig (abgeschnitten) angezeigt. Diese wurde behoben, sodass in diesen Fällen horizontal gescrollt werden kann.

## 6.11 7. August 2025

---

Wir freuen uns, heute zahlreiche Neuerungen ankündigen zu können!

- Die Modelle `GPT4.1`, `GPT4.1 (nano)` und `GPT4.1 (mini)` stehen zur Verfügung
- Die bisherige Übersicht und Auswahl von Standard- und kundenindividuellen Assistenten wurde durch eine detaillierte Übersicht einschließlich Suchfunktion ersetzt
- Business GPT unterstützt nun auch folgende Dateitypen für den Dokumenten-Upload: `.json`, `.md`, `.txt`, `.yaml`, `.xsd`, `.xml`, `.csv`
- Über die Business GPT API können Dokumente durch Angabe eines Referenzparameters nun auch aktualisiert werden

### Neuere Verbesserungen

- Erhöhung der Dateigrößenbeschränkung von 25MB auf **100MB** pro Datei für den Upload von Dokumenten
- Verbesserte Verarbeitung von Excel-Dateien beim Dokumenten-Chat und damit verbesserte Antworten
- Verbesserte Barrierefreiheit durch Styling-Änderungen der Chatantworten

### Fehlerbehebungen

- Fehler bei `o1` und `o3(mini)`-Assistenten bei Verwendung mit angehängten Dokumentencontainer
- Beim Chat mit Assistenten mit Model `o1` oder `o3-Mini` und angehängten Dokumentencontainer kam es zu einem 500-Fehler im Chat
- Fehler bei der Auflösung von Quellen-Referenzen bei längeren Konversationen mit dem Dokumenten-Chat
- Insbesondere bei längeren Unterhaltungen im Dokumenten-Chat wurden Quellen-Referenzen wie `[1]` nicht mehr mit einer Dokumentenquelle aufgelöst
- Fehler beim Indexieren von einer Vielzahl von Dokumenten
- Beim Upload dutzender Dokumente kam es zu anschließenden Fehlern beim Indexieren
- Fehlerbehebungen Business GPT API
- Behebung des Fehlers `BadRequestError: 400 status code` beim Tools Calling

## 6.12 15. Juli 2025

---

Wir freuen uns, heute die folgenden Neuerungen ankündigen zu können!

- Die unterschiedlichen Assistententypen "**Standard Assistent**" und "**Dokumenten Assistent**" wurden zusammengeführt. Somit können alle Assistenten-Optionen kombiniert verwendet werden.
- Erweiterung der **OpenAI API-Funktionalitäten**:
- Bereitstellung **Embedding-Endpunkt**
- Ergänzung des **Completion-Endpunkts** um **Structured Output**
- Ergänzung des **Completion-Endpunkts** um **weitere Parameter**
- Erweiterung der **Business GPT Online Dokumentation**
- Verbesserte Struktur, um Inhalte noch schneller zu finden
- Neu: **Leitfäden, Tipps & Tricks** → [Ressourcen](#)
- **GPT-4** wird eingestellt und steht nicht mehr zur Verfügung

#### Neuere Verbesserungen

- Verbessertes **Chunking bei der Verarbeitung** von hochgeladenen Dokumenten
- Dadurch verbesserte Ergebnisse beim **Chat mit Dokumenten**
- Optimierung des "**Dokument Chat**" Assistenten
- Mit dem Assistenten kann nun *internes Wissen* (Dokumenten-Wissen) und *LLM-Wissen* kombiniert werden → [Dokumenten-Chat](#)

#### Fehlerbehebungen

- Fehlerseite (**401 Fehler**) bei Ende der Nutzersitzung wurde behoben

## 6.13 21. Mai 2025

---

Wir freuen uns, heute die folgenden Modell-Updates ankündigen zu können!

- Neben GPT-4, GPT-4o und anderen Standard-Modellen stehen jetzt **o1**, **o3** und **DALL·E 3** zur Verfügung.
- Mit **DALL·E** können nun direkt Bilder in Business GPT generiert werden.
- **GPT-3.5** wird eingestellt und steht nicht mehr zur Verfügung

#### Neuere Verbesserungen

- Dokumente können per Drag & Drop in Container gezogen werden.
- Der Cursor bleibt nach dem Senden einer Nachricht im Eingabefeld.
- Die Seitennavigation erscheint im neuen Look & Feel.

## 6.14 16. April 2025

---

Wir freuen uns, heute zahlreiche Verbesserungen ankündigen zu können!

- Bilder und Dokumentendateien können per Copy & Paste direkt im Chat-Eingabefeld eingefügt werden
- Differenzierte Anzeige von Input-/Output-Token-Verbrauch und -Kosten in der Verbrauchsübersicht
- Bilddateien im Chat können in einer vergrößerten Ansicht angezeigt werden
- Neuer „Top K“-Parameter für Assistenten

### Neue Verbesserungen

- Redesign der Container-Übersicht und Containerdetails
- Überarbeitete Berechtigungsvergabe für Benutzergruppen
- Redesign der Dokumenten-Übersicht
- Redesign der Gruppenübersicht
- Redesign der Assistentenübersicht und Assistentendetailansicht
- Neue Berechtigungsvergabe für Gruppen
- Neue Zuordnungsmöglichkeit von Dokumentencontainern
- Vorschau-Thumbnail im Chat vergrößerbar
- Chat-Anfragen mit Bildanhang können nun ohne zusätzlichen Text gesendet werden
- Erhöhung der maximalen Zeichenanzahl für System Prompt / Instruktionen bei Assistenten
- Deeplinking für Assistenten im Chatbereich (URLs mit Assistenten-ID)
- Verbesserte Tabellenerkennung in PDF-Dokumenten

### Fehlerbehebungen

- Falsche Anzeige bei mehrfach auftretenden Fehlern beim Indexing in der Dokumentenübersicht
- Fehlerhafte Sortierung der Spalte „Zuletzt geändert“ in der Gruppenliste

## 6.15.24. Februar 2025

Wir freuen uns, heute die Einführung von KI-Assistenten, persönlichen Dokumenten-Upload und weitere Dateitypen für den Dokumenten-Upload ankündigen zu können!

- Persönlicher Dokumenten-Upload für Nutzer
- Nutzer können direkt im Chat eigene Dokument hochladen. Insgesamt können maximal 10 Dateien hochgeladen werden, die für maximal 24 Stunden zur Verfügung stehen.
- Der Upload und die Verwaltung von Dokumenten wurde erweitert für die Dateitypen `.docx`, `.pptx` und `.xlsx`
- Einführung von KI-Assistenten für individuelle Anwendungsfälle
- Unterscheidung nach Standard- und Dokumenten-Assistenten
- Freigabe oder Einschränkung für bestimmte Nutzergruppen
- Erweiterte Styling-Optionen mit auswählbaren Hintergründen
- Erweiterte Verbrauchsübersicht mit **Nachbestellung über Telekom Cloud Marketplace**
- Upload mehrerer Bilddateien im Chat

### Neuere Verbesserungen

- Erweiterte Gruppenfunktionen:
  - Export der Gruppenübersicht
  - Ein-/Ausblenden von Gruppen
  - Umbenennen von Gruppen
- Verbesserte Tabellenfunktionen für Gruppen & Dokumente:
  - Standardmäßig alphabetische Sortierung
  - Individuelle Sortierung von Spalten
  - Standardmäßige Anzeige der Weboberfläche in **Deutsch** nach neuem Login
- Globale Fehlerbehandlung
- Darstellung von Tabellenstrukturen im Chat
- Schnellere Ladezeiten bei PDF-Dokumenten

### Fehlerbehebungen

- Mehrfache Einträge „**Neue Konversation**“ in der Chat-Historie wurden behoben

## 6.16 11. November 2024

Dieses Release bringt die erste große Version mit Funktionen zur **Dokumentenverwaltung** (für Produktvariante „L“), **kundenspezifischem Design** und einer **detaillierten Verbrauchsübersicht**.

### • Individualisierung der Business GPT Weboberfläche durch:

- Upload eines Firmenlogos
- Festlegung der Hauptfarbe
- Hinterlegung des Firmennamens

### • Gruppenverwaltung:

- Upload von Gruppen per CSV-Datei
- Löschen von Gruppen

### • Containerverwaltung:

- Anlegen, Bearbeiten und Löschen von Dokumenten-Containern
- Gruppenbasierte Berechtigungsvergabe
- Upload und Verwaltung von Dokumenten

### • Dokumentenverwaltung:

- Zentrale Dokumentenübersicht
- Upload und Verwaltung von Dokumenten

### • Verbrauchsübersicht:

- Detaillierte Kosten- und Budgetanzeige
- Modellbasierte Kostenaufschlüsselung

## 7. Roadmap

---

Business GPT wird stetig weiterentwickelt, womit wir primär den Anforderungen unserer Kunden und des Marktes folgen, um das bestmögliche Nutzererlebnis zu gewährleisten.

Nachfolgend finden Sie die strukturierte Roadmap für unsere laufenden und zukünftigen Entwicklungen, kategorisiert in verschiedene Themenkomplexe.

Unser Ziel ist die Schaffung einer hochintuitiven und zugänglichen Plattform, die die Nutzerzufriedenheit und das -erlebnis verbessert und die Produktivität steigert.

### 7.1 Assistenten

---

- **Code Interpreter:** Ein Code Interpreter (auch bekannt als Advanced Data Analysis, ADA) ist ein Tool, das Python-Code ausführen kann, um komplexe Aufgaben zu bewältigen. Es kombiniert die Fähigkeit zur Kommunikation in natürlicher Sprache mit der Fähigkeit, Programmieraufgaben direkt auszuführen.
- **nutzerindividuelle Assistenten:** Nutzer sollen neben Administratoren in der Lage sein, eigene Assistenten zu erstellen.
- **Intuitive Assistentenübersicht:** Verfügbare Assistenten sollen einfach und intuitiver zugänglich sein. Nutzer sollen sich einfacher und effizienter über verfügbare Assistenten, die zentral zur Verfügung gestellt wurden, informieren können und Assistenten als Favoriten markieren können.
- **Upload von Dokumenten:** Dokumenten-basiertes Wissen kann direkt hochgeladen und dem Assistenten zugewiesen werden, sodass der Assistent auf dieses Wissen zugreifen kann.

### 7.2 Datenquellen und Wissensmanagement

---

Datenquellen bilden das Fundament für "Chat with your Data". Neben der bereits vorhandenen Möglichkeit mit Dokumenteninhalten zu chatten, wird Business GPT so erweitert, dass Drittsysteme als Datenquellen direkt angebunden werden können. Dort vorhandene Daten werden dann in Business GPT gespeichert und indiziert - sicher und datenschutzkonform.

Folgende Systeme stehen dabei im Fokus:

- **SharePoint Online**
- **Confluence**
- **Staffbase**

Dokumentencontainer, die aktuell nur durch Administratoren verwaltet werden können, sollen künftig auch normalen Nutzer zur Verfügung stehen, sodass persönliche Wissensbasen aufgebaut werden können.

### 7.3 Modellerweiterungen

---

Neue Modelle wie jene von OpenAI und weitere werden datenschutzkonform in Europa einige Zeit später zur Verfügung gestellt, als in den USA. Sobald diese verfügbar sind, werden diese auch in Business GPT ausgerollt.

Folgende Modelle stehen in den Startlöchern:

- **GPT5.4**
- **GPT5.2**

### 7.4 Chatfunktionen

---

- **Tools-Calls:** Funktionsaufrufe für die Suche im Internet oder in Dokumenten, Deep Search und Bildgenerierung werden direkt im Chat zur Verfügung gestellt.

## 7.5 Rollen und Berechtigungen

---

Im Unternehmenskontext sind die Separation von Inhalten und die Zugriffsmöglichkeiten auf selbige von zentraler Bedeutung. Um dem Rechnung zu tragen planen wir die Einführung weiterer Features.

- **Areas/Workspaces:** Zur Separation von Unternehmensbereichen und Inhalten wie HR, Marketing, Technik, usw. dienen Areas (Workspaces).
- **Bereichsadministratoren:** Für eine dedizierte Steuerung der administrativen Zugriffsrechte auf Daten wie Dokumente wird eine weitere Rolle eingeführt.

## 7.6 Administration

---

Um Business GPT flexibel an die Unternehmensbedürfnisse anzupassen, sind tiefgreifende Einstellungen erforderlich. Um diesen Anforderungen nachzukommen, wird Business GPT erweitert um:

- **Festlegung der verfügbaren Modelle:** Die verwendbaren Modelle wie GPT4o, Dall-E, etc. können individuell aktiviert werden.
- **Konfigurationsregeln für Assistenten:** Hierdurch kann festgelegt werden, welche verfügbaren Optionen für individuelle Assistenten verwendet werden und wer berechtigt ist, Assistenten zu administrieren.

## 7.7 MCP Server Integration

---

Mit dieser Erweiterung wird es möglich, externe MCP-Server in das System zu integrieren. Dadurch können eigene oder von Dritten bereitgestellte Server angebunden und deren Tools direkt in der Plattform genutzt werden. So lassen sich bestehende Ressourcen effizient einbinden und flexibel erweitern. Die Integration erfolgt über eine standardisierte Schnittstelle, wodurch Konfiguration und Verwaltung einfach bleiben. Damit öffnet sich das System für vielfältige Einsatzszenarien und bietet noch mehr Möglichkeiten zur individuellen Anpassung.

## 8. Online-Register: Datenstrukturen und Interoperabilität (Data Act)

Dieses Register stellt gemäß **Art. 26 lit. b der Verordnung (EU) 2023/2854 (Data Act)** transparente Informationen über die Exportmöglichkeiten von Kundeninhalten bei **Business GPT** bereit. Ziel ist es, den Wechsel zu einem anderen Anbieter oder die Integration in eine eigene IKT-Infrastruktur effizient zu unterstützen.

### 8.1 1. Begriffsbestimmungen

#### 8.1.1 1.1 Kundeninhalte (Exportierbar)

Hierbei handelt es sich um Daten, die vom Kunden aktiv bereitgestellt oder während der Nutzung des Dienstes generiert wurden. Diese Daten unterliegen der Portabilität gemäß Data Act.

- **Dokumente:** Hochgeladene Quelldateien zur Wissensverarbeitung.
- **Interaktionen:** Chatverläufe und Dialoghistorien.
- **Konfigurationen:** Spezifische Einstellungen für KI-Assistenten und Workflows.
- **Prompts:** Gespeicherte Anweisungen und Prompt-Bibliotheken.

#### 8.1.2 1.2 Systeminterne Daten (Nicht exportierbar)

Daten, die ausschließlich der internen technischen Funktionsweise der KI dienen, sind vom Export ausgeschlossen.

- **Embeddings:** Mathematische Vektor-Repräsentationen (modellspezifisch).
- **Indizes:** Vektor-Datenbank-Strukturen zur Suche.
- **System-Logs:** Interne Cache-Daten und technische Telemetrie.

### 8.2 2. Übersicht der Export-Spezifikationen

Die folgende Tabelle gibt Aufschluss über die verfügbaren Datenformate und Bereitstellungswege:

Kategorie	Beschreibung	Datenformate	Exportmethode
Quelldokumente	Originaldateien (z. B. PDF, DOCX)	Originalformat	Einzel-Download / Support
Chat-Verläufe	Dialoge zw. Nutzer & KI inkl. Metadaten	JSON, CSV	Support-Anfrage
Assistenten	Konfigurationen, Logiken & Parameter	JSON	Support-Anfrage
Prompt-Library	Strukturierte Prompt-Vorlagen	JSON, CSV	Support-Anfrage
Metadaten	Zeitstempel, Rollen, Dateireferenzen	JSON, SQL-Dump	Support-Anfrage

### 8.3 3. Interoperabilität und Standards

**Business GPT** verfolgt einen "Open Data"-Ansatz, um Vendor Lock-in-Effekte zu vermeiden:

- **Maschinenlesbarkeit:** Alle Exporte erfolgen in strukturierten, maschinenlesbaren Formaten (RFC 8259 für JSON, RFC 4180 für CSV).
- **Semantische Klarheit:** Bei Bereitstellung von SQL-Dumps oder komplexen JSON-Strukturen wird eine Schema-Dokumentation beigelegt, die die Bedeutung der Datenfelder erläutert.
- **Standardisierung:** Wir orientieren uns an gängigen IKT-Standards für den Datenaustausch, um die Integration in Drittsysteme zu erleichtern.

---

## 8.4.4. Ablauf des Datenexports & Wechsels

---

1. **Anfrage:** Ein bevollmächtigter Administrator des Mandanten stellt eine Anfrage über unser [Support Center](#).
  2. **Verifizierung:** Aus Sicherheitsgründen erfolgt eine Identitätsprüfung, bevor Daten bereitgestellt werden.
  3. **Paketierung:** Die Daten werden gesammelt und in einem gesicherten Container (z. B. passwortgeschütztes ZIP) vorbereitet.
  4. **Bereitstellung:** Der Zugriff erfolgt über einen verschlüsselten Übertragungsweg (z. B. SFTP oder gesicherter Download-Link).
  5. **Frist:** Die Bereitstellung erfolgt ohne schuldhaftes Zögern, in der Regel innerhalb von 30 Tagen.
- 

## 8.5.5. Sicherheit und Zugriffsschutz

---

Der Schutz Ihrer Daten hat auch beim Export höchste Priorität. Wir setzen folgende Sicherheitsmaßnahmen um:

- **Identitätsprüfung:** Exporte werden nur nach erfolgreicher Verifizierung eines bevollmächtigten Administrators des Mandanten eingeleitet.
  - **Verschlüsselter Transport:** Die Bereitstellung erfolgt über gesicherte Kanäle (z. B. HTTPS-geschützte Download-Bereiche oder SFTP), die mittels **TLS 1.2 oder höher** verschlüsselt sind.
  - **Inhaltsverschlüsselung:** Umfangreiche Datenexporte werden als **AES-256-verschlüsselte Archive (ZIP)** bereitgestellt. Das dazugehörige Passwort wird aus Sicherheitsgründen über einen separaten Kommunikationskanal (Out-of-Band, z. B. Telefon oder SMS/Messenger) oder verschlüsselt übermittelt.
  - **Integrität:** Auf Anfrage stellen wir Prüfsummen (z. B. SHA-256) bereit, damit Sie die Unversehrtheit der exportierten Daten nach dem Transfer verifizieren können.
- 

*Stand: 23.01.2026 Dieses Register wird kontinuierlich aktualisiert, um neue Funktionen und verbesserte Export-Schnittstellen (Self-Service) abzubilden.*