



Informatica® Cloud Application Integration
September 2024

Azure OpenAI Chat with File using Guide

© Copyright Informatica LLC 2024

This software and documentation contain proprietary information of Informatica LLC and are provided under a license agreement containing restrictions on use and disclosure and are also protected by copyright law. Reverse engineering of the software is prohibited. No part of this document may be reproduced or transmitted in any form, by any means (electronic, photocopying, recording or otherwise) without prior consent of Informatica LLC. This Software may be protected by U.S. and/or international Patents and other Patents Pending.

Use, duplication, or disclosure of the Software by the U.S. Government is subject to the restrictions set forth in the applicable software license agreement and as provided in DFARS 227.7202-1(a) and 227.7702-3(a) (1995), DFARS 252.227-7013(1)(ii) (OCT 1988), FAR 12.212(a) (1995), FAR 52.227-19, or FAR 52.227-14 (ALT III), as applicable.

The information in this product or documentation is subject to change without notice. If you find any problems in this product or documentation, please report them to us in writing.

Informatica, Informatica Platform, Informatica Data Services, PowerCenter, PowerCenterRT, PowerCenter Connect, PowerCenter Data Analyzer, PowerExchange, PowerMart, Metadata Manager, Informatica Data Quality, Informatica Data Explorer, Informatica B2B Data Transformation, Informatica B2B Data Exchange Informatica On Demand, Informatica Identity Resolution, Informatica Application Information Lifecycle Management, Informatica Complex Event Processing, Ultra Messaging, Informatica Master Data Management, and Live Data Map are trademarks or registered trademarks of Informatica LLC in the United States and in jurisdictions throughout the world. All other company and product names may be trade names or trademarks of their respective owners.

Portions of this software and/or documentation are subject to copyright held by third parties, including without limitation: Copyright DataDirect Technologies. All rights reserved. Copyright © Sun Microsystems. All rights reserved. Copyright © RSA Security Inc. All Rights Reserved. Copyright © Ordinal Technology Corp. All rights reserved. Copyright © Aandacht c.v. All rights reserved. Copyright Genivia, Inc. All rights reserved. Copyright Isomorphic Software. All rights reserved. Copyright © Meta Integration Technology, Inc. All rights reserved. Copyright © Intalio. All rights reserved. Copyright © Oracle. All rights reserved. Copyright © Adobe Systems Incorporated. All rights reserved. Copyright © DataArt, Inc. All rights reserved. Copyright © ComponentSource. All rights reserved. Copyright © Microsoft Corporation. All rights reserved. Copyright © Rogue Wave Software, Inc. All rights reserved. Copyright © Teradata Corporation. All rights reserved. Copyright © Yahoo! Inc. All rights reserved. Copyright © Glyph & Cog, LLC. All rights reserved. Copyright © Thinkmap, Inc. All rights reserved. Copyright © Clearpace Software Limited. All rights reserved. Copyright © Information Builders, Inc. All rights reserved. Copyright © OSS Nokalva, Inc. All rights reserved. Copyright Edifecs, Inc. All rights reserved. Copyright Cleo Communications, Inc. All rights reserved. Copyright © International Organization for Standardization 1986. All rights reserved. Copyright © ej-technologies GmbH. All rights reserved. Copyright © Jaspersoft Corporation. All rights reserved. Copyright © International Business Machines Corporation. All rights reserved. Copyright © yWorks GmbH. All rights reserved. Copyright © Lucent Technologies. All rights reserved. Copyright © University of Toronto. All rights reserved. Copyright © Daniel Veillard. All rights reserved. Copyright © Unicode, Inc. Copyright IBM Corp. All rights reserved. Copyright © MicroQuill Software Publishing, Inc. All rights reserved. Copyright © PassMark Software Pty Ltd. All rights reserved. Copyright © LogiXML, Inc. All rights reserved. Copyright © 2003-2010 Lorenzi Davide, All rights reserved. Copyright © Red Hat, Inc. All rights reserved. Copyright © The Board of Trustees of the Leland Stanford Junior University. All rights reserved. Copyright © EMC Corporation. All rights reserved. Copyright © Flexera Software. All rights reserved. Copyright © Jinfonet Software. All rights reserved. Copyright © Apple Inc. All rights reserved. Copyright © Teleric Inc. All rights reserved. Copyright © BEA Systems. All rights reserved. Copyright © PDFlib GmbH. All rights reserved. Copyright © Orientation in Objects GmbH. All rights reserved. Copyright © Tanuki Software, Ltd. All rights reserved. Copyright © Ricebridge. All rights reserved. Copyright © Sencha, Inc. All rights reserved. Copyright © Scalable Systems, Inc. All rights reserved. Copyright © jQWidgets. All rights reserved. Copyright © Tableau Software, Inc. All rights reserved. Copyright © MaxMind, Inc. All Rights Reserved. Copyright © TMate Software s.r.o. All rights reserved. Copyright © MapR Technologies Inc. All rights reserved. Copyright © Amazon Corporate LLC. All rights reserved. Copyright © Highsoft. All rights reserved. Copyright © Python Software Foundation. All rights reserved. Copyright © BeOpen.com. All rights reserved. Copyright © CNRI. All rights reserved.

This product includes software developed by the Apache Software Foundation (<http://www.apache.org/>), and/or other software which is licensed under various versions of the Apache License (the "License"). You may obtain a copy of these Licenses at <http://www.apache.org/licenses/>. Unless required by applicable law or agreed to in writing, software distributed under these Licenses is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the Licenses for the specific language governing permissions and limitations under the Licenses.

This product includes software which was developed by Mozilla (<http://www.mozilla.org/>), software copyright The JBoss Group, LLC, all rights reserved; software copyright © 1999-2006 by Bruno Lowagie and Paulo Soares and other software which is licensed under various versions of the GNU Lesser General Public License Agreement, which may be found at <http://www.gnu.org/licenses/lgpl.html>. The materials are provided free of charge by Informatica, "as-is", without warranty of any kind, either express or implied, including but not limited to the implied warranties of merchantability and fitness for a particular purpose.

The product includes ACE(TM) and TAO(TM) software copyrighted by Douglas C. Schmidt and his research group at Washington University, University of California, Irvine, and Vanderbilt University, Copyright (©) 1993-2006, all rights reserved.

This product includes software developed by the OpenSSL Project for use in the OpenSSL Toolkit (copyright The OpenSSL Project. All Rights Reserved) and redistribution of this software is subject to terms available at <http://www.openssl.org> and <http://www.openssl.org/source/license.html>.

This product includes Curl software which is Copyright 1996-2013, Daniel Stenberg, <daniel@haxx.se>. All Rights Reserved. Permissions and limitations regarding this software are subject to terms available at <http://curl.haxx.se/docs/copyright.html>. Permission to use, copy, modify, and distribute this software for any purpose with or without fee is hereby granted, provided that the above copyright notice and this permission notice appear in all copies.

The product includes software copyright 2001-2005 (©) MetaStuff, Ltd. All Rights Reserved. Permissions and limitations regarding this software are subject to terms available at <http://www.dom4j.org/license.html>.

The product includes software copyright © 2004-2007, The Dojo Foundation. All Rights Reserved. Permissions and limitations regarding this software are subject to terms available at <http://dojotoolkit.org/license>.

This product includes ICU software which is copyright International Business Machines Corporation and others. All rights reserved. Permissions and limitations regarding this software are subject to terms available at <http://source.icu-project.org/repos/icu/icu/trunk/license.html>.

This product includes software copyright © 1996-2006 Per Bothner. All rights reserved. Your right to use such materials is set forth in the license which may be found at <http://www.gnu.org/software/kawa/Software-License.html>.

This product includes OSSP UUID software which is Copyright © 2002 Ralf S. Engelschall, Copyright © 2002 The OSSP Project Copyright © 2002 Cable & Wireless Deutschland. Permissions and limitations regarding this software are subject to terms available at <http://www.opensource.org/licenses/mit-license.php>.

This product includes software developed by Boost (<http://www.boost.org/>) or under the Boost software license. Permissions and limitations regarding this software are subject to terms available at http://www.boost.org/LICENSE_1_0.txt.

This product includes software copyright © 1997-2007 University of Cambridge. Permissions and limitations regarding this software are subject to terms available at <http://www.pcre.org/license.txt>.

This product includes software copyright © 2007 The Eclipse Foundation. All Rights Reserved. Permissions and limitations regarding this software are subject to terms available at <http://www.eclipse.org/org/documents/epl-v10.php> and at <http://www.eclipse.org/org/documents/edl-v10.php>.

This product includes software licensed under the terms at <http://www.tcl.tk/software/tcltk/license.html>, <http://www.bosrup.com/web/overlib/?License>, <http://www.stlport.org/doc/license.html>, <http://asm.ow2.org/license.html>, <http://www.cryptix.org/LICENSE.TXT>, <http://hsqldb.org/web/hsqldbLicense.html>, <http://httpunit.sourceforge.net/doc/license.html>, <http://jung.sourceforge.net/license.txt>, http://www.gzip.org/zlib/zlib_license.html, <http://www.openldap.org/software/release/license.html>, <http://www.libssh2.org>, <http://slf4j.org/license.html>, <http://www.sente.ch/software/OpenSourceLicense.html>, <http://fusesource.com/downloads/license-agreements/fuse-message-broker-v-5-3-license-agreement>; <http://antlr.org/license.html>; <http://aopalliance.sourceforge.net/>; <http://www.bouncycastle.org/licence.html>; <http://www.jgraph.com/jgraphdownload.html>; <http://www.jcraft.com/jsch/LICENSE.txt>; http://jotm.objectweb.org/bsd_license.html; <http://www.w3.org/Consortium/Legal/2002/copyright-software-20021231>; <http://www.slf4j.org/license.html>; <http://nanoxml.sourceforge.net/orig/copyright.html>; <http://www.json.org/license.html>; <http://forge.ow2.org/projects/javaservice/>; <http://www.postgresql.org/about/license.html>, <http://www.sqlite.org/copyright.html>, <http://www.tcl.tk/software/tcltk/license.html>, <http://www.jaxen.org/faq.html>, <http://www.jdom.org/docs/faq.html>, <http://www.slf4j.org/license.html>; <http://www.iodbc.org/dataspace/iodbc/wiki/IODBC/License>; <http://www.keplerproject.org/md5/license.html>; <http://www.toedter.com/en/jcalendar/license.html>; <http://www.edankert.com/bounce/index.html>; <http://www.net-snmp.org/about/license.html>; <http://www.openmdx.org/#FAQ>; http://www.php.net/license/3_01.txt; <http://srp.stanford.edu/license.txt>; <http://www.schneier.com/blowfish.html>; <http://www.jmock.org/license.html>; <http://xsom.java.net>; <http://benalman.com/about/license/>; <https://github.com/CreateJS/EaselJS/blob/master/src/easeljs/display/Bitmap.js>; <http://www.h2database.com/html/license.html#summary>; <http://jsoncpp.sourceforge.net/LICENSE>; <http://jdbc.postgresql.org/license.html>; <http://protobuf.googlecode.com/svn/trunk/src/google/protobuf/descriptor.proto>; <https://github.com/rantav/hector/blob/master/LICENSE>; <http://web.mit.edu/Kerberos/krb5-current/doc/mitK5license.html>; <http://jibx.sourceforge.net/jibx-license.html>; <https://github.com/lyokato/libgeohash/blob/master/LICENSE>; <https://github.com/hjiang/jsonxx/blob/master/LICENSE>; <https://code.google.com/p/lz4/>; <https://github.com/jedisct1/libsodium/blob/master/LICENSE>; <http://one-jar.sourceforge.net/index.php?page=documents&file=license>; <https://github.com/EsotericSoftware/kryo/blob/master/license.txt>; <http://www.scala-lang.org/license.html>; <https://github.com/tinkerpop/blueprints/blob/master/LICENSE.txt>; <http://gee.cs.oswego.edu/dl/classes/EDU/oswego/cs/dl/util/concurrent/intro.html>; <https://aws.amazon.com/asl/>; <https://github.com/twbs/bootstrap/blob/master/LICENSE>; <https://sourceforge.net/p/xmlunit/code/HEAD/tree/trunk/LICENSE.txt>; <https://github.com/documentcloud/underscore-contrib/blob/master/LICENSE>, and <https://github.com/apache/hbase/blob/master/LICENSE.txt>.

This product includes software licensed under the Academic Free License (<http://www.opensource.org/licenses/afl-3.0.php>), the Common Development and Distribution License (<http://www.opensource.org/licenses/cddl1.php>), the Common Public License (<http://www.opensource.org/licenses/cpl1.0.php>), the Sun Binary Code License Agreement Supplemental License Terms, the BSD License (<http://www.opensource.org/licenses/bsd-license.php>), the new BSD License (<http://opensource.org/licenses/BSD-3-Clause>), the MIT License (<http://www.opensource.org/licenses/mit-license.php>), the Artistic License (<http://www.opensource.org/licenses/artistic-license-1.0>) and the Initial Developer's Public License Version 1.0 (<http://www.firebirdsql.org/en/initial-developer-s-public-license-version-1-0/>).

This product includes software copyright © 2003-2006 Joe Walnes, 2006-2007 XStream Committers. All rights reserved. Permissions and limitations regarding this software are subject to terms available at <http://xstream.codehaus.org/license.html>. This product includes software developed by the Indiana University Extreme! Lab. For further information please visit <http://www.extreme.indiana.edu/>.

This product includes software Copyright (c) 2013 Frank Balluffi and Markus Moeller. All rights reserved. Permissions and limitations regarding this software are subject to terms of the MIT license.

See patents at <https://www.informatica.com/legal/patents.html>.

DISCLAIMER: Informatica LLC provides this documentation "as is" without warranty of any kind, either express or implied, including, but not limited to, the implied warranties of noninfringement, merchantability, or use for a particular purpose. Informatica LLC does not warrant that this software or documentation is error free. The information provided in this software or documentation may include technical inaccuracies or typographical errors. The information in this software and documentation is subject to change at any time without notice.

NOTICES

This Informatica product (the "Software") includes certain drivers (the "DataDirect Drivers") from DataDirect Technologies, an operating company of Progress Software Corporation ("DataDirect") which are subject to the following terms and conditions:

1. THE DATADIRECT DRIVERS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NON-INFRINGEMENT.
2. IN NO EVENT WILL DATADIRECT OR ITS THIRD PARTY SUPPLIERS BE LIABLE TO THE END-USER CUSTOMER FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, CONSEQUENTIAL OR OTHER DAMAGES ARISING OUT OF THE USE OF THE ODBC DRIVERS, WHETHER OR NOT INFORMED OF THE POSSIBILITIES OF DAMAGES IN ADVANCE. THESE LIMITATIONS APPLY TO ALL CAUSES OF ACTION, INCLUDING, WITHOUT LIMITATION, BREACH OF CONTRACT, BREACH OF WARRANTY, NEGLIGENCE, STRICT LIABILITY, MISREPRESENTATION AND OTHER TORTS.

Publication Date: 2024-12-06

Table of Contents

- Preface 5**

- Chapter 1: Introduction to Azure OpenAI Chat with File using Guide recipe.... 6**
 - Azure OpenAI Chat with File using Guide recipe contents. 6
 - Azure OpenAI Chat with File using Guide recipe assets. 7

- Chapter 2: Using the Azure OpenAI Chat with File using Guide recipe..... 8**
 - Step 1: Copy and access the recipe 8
 - Step 2: Configure and publish the Azure OpenAI connection. 9
 - Step 3: Configure and publish the processes. 9
 - Step 4: Publish and run the guide. 10

Preface

Use *Azure OpenAI Chat with File using Guide* to learn how to ask questions to the Azure OpenAI Large Language Model (LLM) based on an uploaded file's contents using a guide. This guide assumes that you have an understanding of the Azure OpenAI Connector concepts.

CHAPTER 1

Introduction to Azure OpenAI Chat with File using Guide recipe

The Azure OpenAI Chat with File using Guide recipe is based on REST and SOAP APIs. You can ask questions to the Azure OpenAI Large Language Model (LLM) based on an uploaded file's contents.

The recipe allows you to upload a file with context, provide a user prompt, read the text from the file, and answer the user's questions based on the contents of the file using the Azure OpenAI Large Language Model (LLM).

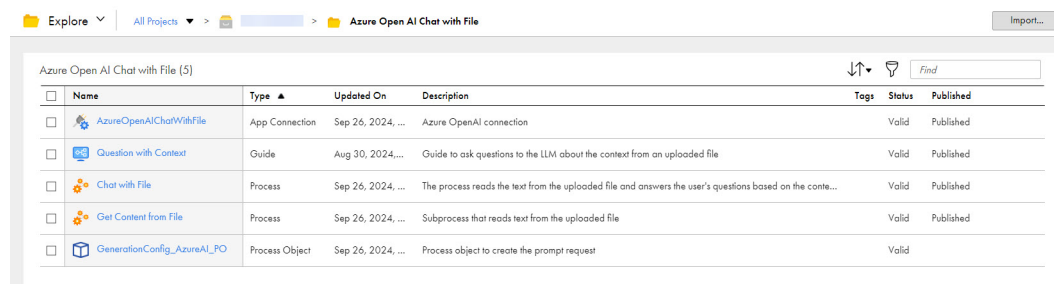
You can upload files of the following formats:

.json, .js, .txt

Azure OpenAI Chat with File using Guide recipe contents

The Azure OpenAI Chat with File using Guide recipe contains a process object, an app connection, a guide, and two processes.

The following image shows the assets that the Azure OpenAI Chat with File using Guide recipe package contains:



The screenshot shows the 'Azure OpenAI Chat with File' recipe package contents in the Azure portal. The interface includes a breadcrumb trail: 'Explore > All Projects > Azure OpenAI Chat with File'. Below the breadcrumb is a table listing the assets:

Name	Type	Updated On	Description	Tags	Status	Published
AzureOpenAIChatWithFile	App Connection	Sep 26, 2024, ...	Azure OpenAI connection		Valid	Published
Question with Context	Guide	Aug 30, 2024, ...	Guide to ask questions to the LLM about the context from an uploaded file		Valid	Published
Chat with File	Process	Sep 26, 2024, ...	The process reads the text from the uploaded file and answers the user's questions based on the conte...		Valid	Published
Get Content from File	Process	Sep 26, 2024, ...	Subprocess that reads text from the uploaded file		Valid	Published
GenerationConfig_AzureAI_PO	Process Object	Sep 26, 2024, ...	Process object to create the prompt request		Valid	

Azure OpenAI Chat with File using Guide recipe assets

The following table lists the assets that the Azure OpenAI Chat with File using Guide recipe package contains:

Asset Name	Asset Type	Description
GenerationConfig_AzureAI_PO	Process object	Creates the prompt request.
AzureOpenAIChatWithFile	App connection	Azure OpenAI connection.
Get Content from File	Process	Subprocess that reads the text from the uploaded file.
Chat with File	Process	Process that reads the text from the uploaded file and answers the user's questions based on the contents of the file using the Azure OpenAI Large Language Model (LLM).
Question with Context	Guide	Guide that allows users to ask questions to the LLM about the context from an uploaded file.

CHAPTER 2

Using the Azure OpenAI Chat with File using Guide recipe

To use the Azure OpenAI Chat with File using Guide recipe, you must perform the following steps manually:

Step 1: Copy and access the recipe

Step 2: Configure and publish the Azure OpenAI connection

Step 3: Configure and publish the processes

Step 4: Publish and run the guide

Step 1: Copy and access the recipe

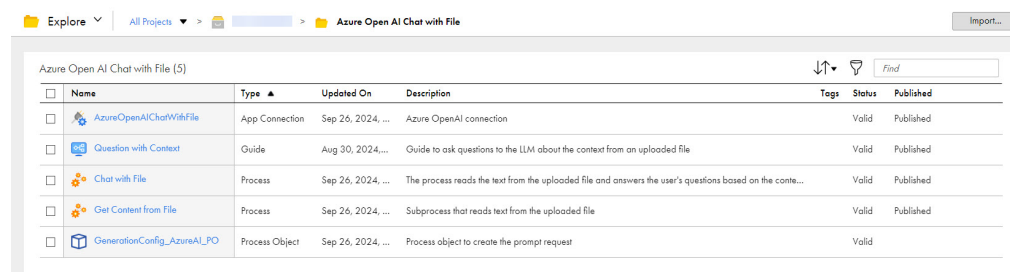
Copy the pre-configured assets in the recipe to a separate project or folder.

1. Open the **Azure OpenAI Chat with File using Guide** recipe and click **Use**.
2. Select the location where you want to copy the recipe, and then click **Continue**.
3. In the **Copying the recipe** dialog box, click **OK**.

It might take some time for the recipe to get copied. You will receive a notification when the recipe is ready for use.

4. After the recipe is copied, click **Explore** to access the recipe content.
5. Navigate to the project or folder where you copied the recipe or enter the recipe name in the **Find** box.

All the assets in the recipe are displayed as shown in the following image:



The screenshot shows a table of assets for the 'Azure OpenAI Chat with File' recipe. The table has columns for Name, Type, Updated On, Description, Tags, Status, and Published. The assets listed are:

Name	Type	Updated On	Description	Tags	Status	Published
AzureOpenAIChatWithFile	App Connection	Sep 26, 2024, ...	Azure OpenAI connection		Valid	Published
Question with Context	Guide	Aug 30, 2024, ...	Guide to ask questions to the LLM about the context from an uploaded file		Valid	Published
Chat with File	Process	Sep 26, 2024, ...	The process reads the text from the uploaded file and answers the user's questions based on the conte...		Valid	Published
Get Content from File	Process	Sep 26, 2024, ...	Subprocess that reads text from the uploaded file		Valid	Published
GenerationConfig_AzureAI_PO	Process Object	Sep 26, 2024, ...	Process object to create the prompt request		Valid	

Step 2: Configure and publish the Azure OpenAI connection

Configure the endpoint URL and API key in the AzureOpenAIChatWithFile connection, and then publish the connection.

1. Open the **AzureOpenAIChatWithFile** connection.
2. In the **Connection Properties** section, enter values for the following properties:

Property	Description
Endpoint_URL	The REST API endpoint for Azure OpenAI. You can find this value in the Keys & Endpoint section when examining your resource from the Azure portal. Alternatively, you can find the value in Azure OpenAI Studio > Playground > Code View .
API_Key	The API key to authenticate Azure OpenAI connection requests. You can find this value in the Keys & Endpoint section when examining your resource from the Azure portal. Alternatively, you can find the value in Azure OpenAI Studio > Playground > Code View .

3. Save and publish the connection.

Step 3: Configure and publish the processes

Configure the deployment details of the LLM model and publish the processes.

1. To publish the **Get Content from File** process, click **Actions** in the row that contains the process and select **Publish**.
2. Open the **Chat with File** process.
3. On the **Temp Fields** tab of the Start step, enter values for the following fields:
 - In the **deployment_id** field, enter the deployment ID of the LLM model.
 - In the **api_version** field, enter the API version of the LLM model.
4. Optionally, in the **Configure Request Parameters** step, enter the prompt instructions in the **Assignments** field by updating the **Prompt_Configuration** field using the Expression Editor, as shown in the following sample code:

```
For Prompt_Configuration:  
<GenerationConfig_AzureAI>  
  <temperature>0.5</temperature>  
  <topP>1</topP>  
  <max_tokens>1000</max_tokens>  
</GenerationConfig_AzureAI>
```

For the **Prompt_Configuration** field, enter values for the following properties:

Property	Description
temperature	Controls the randomness of the model's output. A lower value close to 0 makes the output more deterministic, while a higher value close to 1 increases randomness and creativity. For example, if temperature is set to 0.5, the model balances between deterministic and creative outputs.
topP	Determines the cumulative probability threshold for token selection. The model considers the smallest set of tokens whose cumulative probability meets or exceeds topP. For example, if topP is set to 0.1, the model considers only the top 10% most probable tokens at each step.
max_tokens	Defines the token count of your prompt. The max_tokens can't exceed the model's context length. Most of the models have a context length of 2048 tokens.

- Optionally, in the **Prepare Request** step, enter the prompt instructions in the **Assignments** field by updating the **Request** field using the Expression Editor, as shown in the following sample code:

```
For Request:
<CreateChatCompletionRequest>
  <temperature>{$temp.Prompt_Configuration[1]/temperature }</temperature>
  <top_p>{$temp.Prompt_Configuration[1]/top_p }</top_p>
  <max_tokens>{$temp.Prompt_Configuration[1]/max_tokens }</max_tokens>
<messages>
  <role>user</role>
  <content>You are a helpful assistant.{ $temp.Content_From_File}</content>
</messages>
<messages>
  <role>user</role>
  <content>{$input.User_Prompt}</content>
</messages>
</CreateChatCompletionRequest>
```

- Save and publish the process.

Step 4: Publish and run the guide

You can ask questions to the LLM about the context from an uploaded file. You can use files of the following formats:

.json, .js, .txt

To publish and run the guide, perform the following steps:

- Open the **Question with Context** guide.
- On the **Start** tab of the Start step, ensure that the **Run As** field is set to **Current User**.
- Save and publish the guide.
- On the **Actions** menu, click **Run**. Alternatively, you can copy the execution URL from the **Properties Details** dialog box to run the guide.
- On the **Instructions** page, select the context file that you want to upload from your machine.
- In the **Enter your query below** field, enter a question based on the file contents.
- Click **Continue**.
You will receive a response from the LLM.
- To ask the next question, click **Next Question**.

9. Enter new questions with the same context and click **Continue**.

Note: If the uploaded file is empty or the question is not related to the file context, an error occurs.

You can also use the embed code to embed the guide into an HTML document of a third-party application.