

Project Carnegie

Private Preview Terms

October 10, 2022

These terms are an agreement between you and Microsoft Corporation (or one of its affiliates). As confirmed in your acceptance of the invitation to participate in the Project Carnegie Private Preview, or by accessing any materials provided in connection with the program, you agree that (1) you have the authority to agree to these terms on behalf of your organization and (2) your organization agrees to and is bound by these terms.

Overview

We are introducing the Project Carnegie Private Preview API (“Project Carnegie Preview”), a new Azure Cognitive Services API that detects certain material that is potentially offensive, risky, or otherwise undesirable. The initial version of Project Carnegie Preview will detect material in text. In later versions, we intend to update the API with new functionalities offering state-of-the-art text, image, and multi-modal models that will detect problematic content to help make applications & services safer from harmful user-generated and/or AI-generated content.

We are very happy to have you experiment with the API during its Private Preview. We are eager to collaborate with you so that your feedback can help us shape Project Carnegie further before we make it generally available.

Your use of Project Carnegie Preview is subject to the terms of your license agreement with Microsoft applicable to your use of Microsoft online services, including applicable Product Terms incorporated in the license agreements (“*License Terms*”). Any capitalized terms not otherwise defined in these Project Carnegie Preview Terms will have the meaning given to them in your License Terms. In the event of any conflict between these Project Carnegie Preview Terms and the License Terms, these terms will control.

What We Need From You

Once you are approved for use of the Project Carnegie Preview, we will need approximately 2 or more hours of your time:

- a. Onboarding session – approximately 1 hour
- b. Feedback session(s) – approximately 1 to 2 hours
- c. Optional: Experimenting with Project Carnegie beyond scheduled sessions

Data Access & Use

You may optionally provide us with examples of your real content with which you are using the Project Carnegie Preview and/or examples of content and associated predictive scores (“Sample Content”) so that we may manually review them for purposes of improving your experience with the Project Carnegie Preview and improving the quality of the Project Carnegie service.

If you provide Sample Content to us:

- You agree that you will not include any personal, confidential, or commercially sensitive information in the Sample Content you send us.

- You consent to permit authorized engineers, data scientists, and project teams (collectively, “Reviewers”) from Microsoft to manually review the Sample Content to build, improve, and validate the machine learning models for Project Carnegie (“Improvements”). We will review the Sample Content to improve the customer experience with better machine learning implementations. Manual review of the Sample Content is needed to achieve that outcome. You may revoke your consent to our human review of your Sample Content at any time by giving the project manager working with you written notice, and we will delete the Sample Content from our systems promptly upon receipt of such notice.
- The project manager working with you will provide instructions for how to send any Sample Content you choose to provide to us. Only Reviewers with permissions will access and view your Sample Content, solely for the purposes described here, and such Reviewers will only use secure workstations to do so.
- As between you and Microsoft: (a) all Sample Content remain yours and (b) all Improvements, and any work product or developments (such as coding or documentation) resulting from Microsoft’s use of Sample Content, will be owned by Microsoft subject to any underlying rights of yours in the Sample Content. In addition, all Sample Content will remain your Confidential Information and all Improvements will remain Microsoft’s Confidential Information.

Acknowledgements

Please remember that the Project Carnegie Preview is a new and innovative technology and results are not guaranteed. Your oversight and review are crucial. It is ultimately you who are accountable for the final results of your use of Project Carnegie Preview.

We want the Project Carnegie Preview functionality to delight and assist you. Accordingly, if the system generates unintended results, or results that do not meet your expectations, immediately inform Microsoft (via the project manager you are working with, or other mechanism specified by Microsoft) so that we can address those concerns.

It is important that you understand the scope and limitations of the Project Carnegie Preview. The initial version of the Project Carnegie Preview provides predictive scores for four (4) different categories of potentially offensive or harmful textual content, as described briefly below. Contact the project manager you are working with for more details.

- Category 1: Sexual - Sexual describes language related to anatomical organs and genitals, romantic relationship, acts portrayed in erotic or affectionate terms, pregnancy, physical sexual acts, including those portrayed as an assault or a forced sexual violent act against one’s will, prostitution, pornography.
- Category 2: Violence - Violence describes language related to physical actions intended to hurt, injure, damage, or kill someone or something; describes weapons, guns, and related entities, such as manufacturers, associations, legislation, etc.
- Category 3: Hate Speech - Hate speech is defined as any speech that attacks or uses pejorative or discriminatory language with reference to a person or Identity Group on the basis of certain differentiating attributes of these groups including but not limited to race, ethnicity, nationality, gender identity and expression, sexual orientation, religion, immigration status, ability status, personal appearance, and body size.
- Category 4: Self-Harm - Self-harm describes language related to physical actions intended to purposely hurt, injure, or damage one’s body or kill oneself.

This service is in preview, which means that it is made available to you before general availability so you can test and evaluate it and provide feedback to Microsoft. Previews may have restricted functionality and may employ reduced or different privacy, security, or compliance commitments than a commercial version. To the extent that you use the Project Carnegie Preview with any “live” or production customer data, personal data, or other data that is subject to heightened compliance requirements, such use is at your sole risk, and it is your responsibility to notify your end users that they should not include sensitive information with their use of previews. This preview feature is provided “as-is,” “with all faults,” “as available,” and without warranty, support, or service level agreements. You may be required to update your Project Carnegie resource during or at the end of the preview period.

Other Terms

Feedback. In connection with your participation in this private preview, you may give us suggestions, comments, input, ideas, or know-how, in any form, (“*Feedback*”) about Project Carnegie Preview. However, you will not provide us Feedback containing Personal Data or other data that is subject to legal or regulatory compliance requirements. You give to Microsoft, without charge, the right to use, share, and freely commercialize Feedback in any way and for any purpose. You also give third parties, without charge, the right to use, or interface with, any Microsoft products or services that include the Feedback. You represent and warrant that you own or otherwise control all rights to such Feedback and that no such Feedback is subject to any third-party rights.

Confidentiality. Except for information that Microsoft has disclosed publicly, all information you receive about Project Carnegie Preview through participation in this program is Confidential Information. All Confidential Information shared between you and Microsoft is subject to the non-disclosure agreement between us or, in the absence of one, the confidentiality terms below:

“*Confidential Information*” means non-public information, know-how, or trade secrets in any form that are designated confidential or that a reasonable person knows or reasonably should know to be confidential. Confidential Information does not include any information, however designated, that: **(i)** is or becomes publicly available without a breach of the confidentiality terms herein; **(ii)** was lawfully known to the receiver of the information without an obligation to keep it confidential; **(iii)** is received from another source who can disclose it lawfully and without an obligation to keep it confidential; **(iv)** is independently developed; or **(v)** is Feedback.

Neither you nor Microsoft will disclose the other's Confidential Information to a third party for five years. Each party will (a) use such information only for purposes of this business relationship, and (b) take reasonable steps to protect the other's Confidential Information. Each party may disclose the other's Confidential Information to its affiliates, employees, and contractors only on a need-to-know basis, subject to the confidentiality provisions herein.

Confidential Information may also be disclosed by either party as necessary to comply with a judicial or other governmental order, but the party making such disclosure must (a) first give the other party reasonable notice of such disclosure to allow for a reasonable opportunity to seek a protective order or equivalent, or (b) obtain written assurance from the applicable judicial or governmental entity that it will afford the Confidential Information the highest level of protection available under applicable law, or otherwise as may be required by law.

Your Responsibility for Content in Your Application or Service. Project Carnegie provides predictive scores that indicate the possible presence of certain types of potentially offensive or harmful textual content. It is your responsibility to monitor and review these scores and the associated content and take action as appropriate. As between you and Microsoft, you are responsible for all content in your application or service, and it is your responsibility to take action with respect to content in your application or service as necessary to comply with applicable laws.