Harvard Business School

DRAFT
FEBRUARY 4, 2025

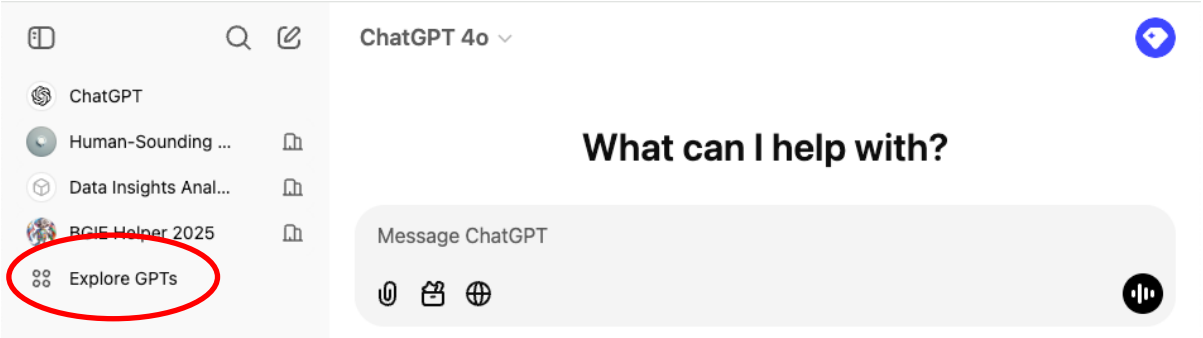
IAVOR BOJINOV
ANNIKA HILDEBRANDT

Building a Custom GPT and AI Agent Evaluator

Custom GPTs can quickly and easily be built using ChatGPT. These models are the precursors to agents as they can combine multiple sources of information with custom instructions; however, they are still intended to be worker companions as they lack the autonomy to act. Nevertheless, they provide a powerful tool that enables employees to create a wide range of custom applications, from personal benefits assistants to coding support. One of the benefits of custom GPTs is they can be shared broadly with others, either in your enterprise or with other ChatGPT users.

This decentralized approach that allows everyone to create and share agents raises some vital governance questions. To address governance issues, you will be creating a custom GPT that creates governance guidelines for the creators of GPTs. Your evaluator will help GPT builders understand what standards custom GPTs and AI agents must adhere to.

The instructions below are tailored for the ChatGPT premium accounts with access to build GPTs. Begin by navigating to chat.com and logging in or creating an account. In the left-hand navigation bar, select 'Explore GPTs' to reach the GPT home page.



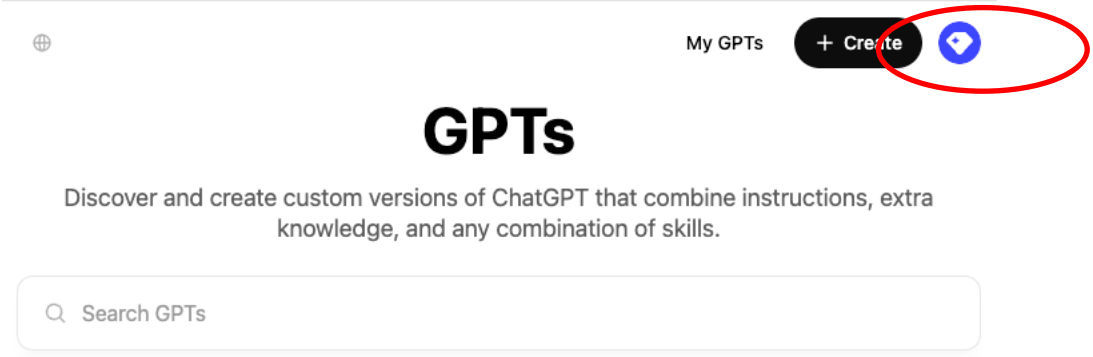
Task 1: Creating your first GPT – Basic Configuration

Being by creating your first GPT and configuring the details.

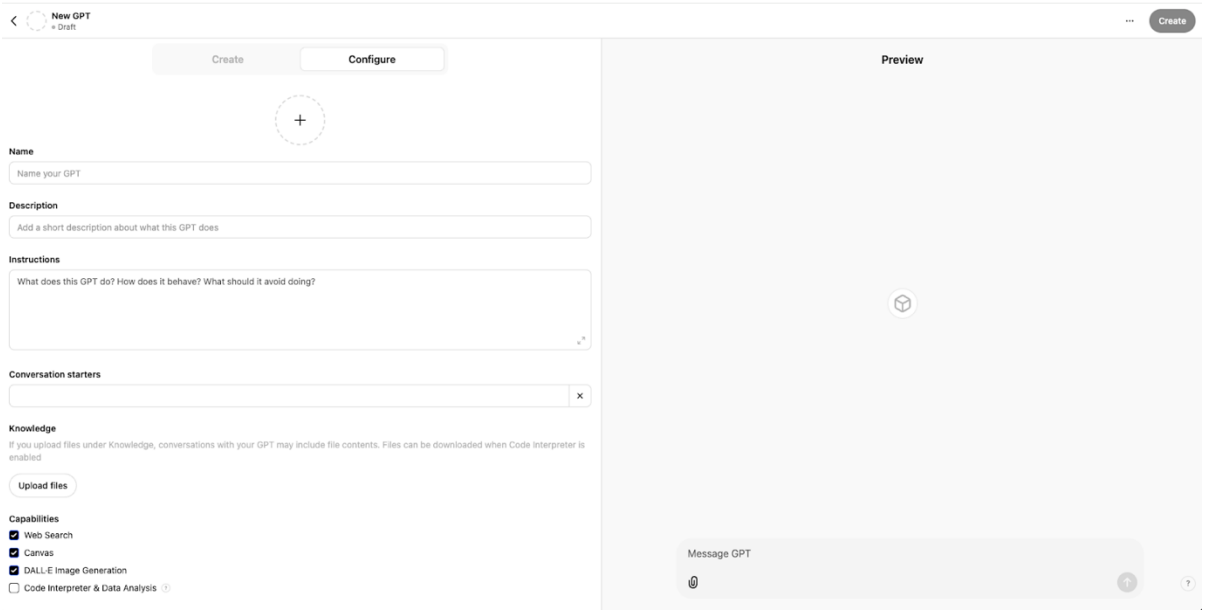
1) On the GPT home page, select the create button in the top right to be brought to the configuration page.

Professor Iavor Bojinov and Research Associate Annika Hildebrandt prepared this exercise as the basis for class discussion rather than to illustrate either effective or ineffective handling of an administrative situation.

Copyright © 2025 President and Fellows of Harvard College. To order copies or request permission to reproduce materials, call 1-800-545-7685, write Harvard Business School Publishing, Boston, MA 02163, or go to www.hbsp.harvard.edu. This publication may not be digitized, photocopied, or otherwise reproduced, posted, or transmitted, without the permission of Harvard Business School.



2) You will be brought to the new GPT draft page. On the left side, you will see a panel for creating your GPT. On the right side you will see a preview window where you can test your GPT. In the left-hand panel, you can either fill out the instructions for your GPT manually using the 'Configure' tab or have ChatGPT fill them out for you via a conversational chat using the 'Create' tab. For the purposes of this activity, you will use the 'Configure' tab.



3) Enter a name for your GPT in the name field. Below is an example to get you started; feel free to deviate as desired. Throughout the document, examples are written in italics.

   a) Example name: *AI Agent Evaluator*

4) Enter a short description in the description field.

   a) Example description: *An custom GPT that provides governance guidance for agents and custom GPTs.*

Task 2: Creating your first GPT – Instructions

Instructions will let your custom GPT know how it should behave.

1) Enter a context for your AI agent in the instruction box. Context helps your GPT respond appropriately.

    a. Example context:

    # Context

    Your goal is to provides governance guidance for the creators of custom GPTs at [company name].

    Note: You can and should modify the text in the brackets throughout the activity to customize your agent. The number sign [#] symbol followed by a space is used by ChatGPT to create headings. One # corresponds to 'Heading 1', Two ## corresponds to 'Heading 2', and ### corresponds to 'Heading 3'. Structure is an important element of agent instructions.

2) Start adding specific instruction steps under a new 'Instructions' header. First, add an instructions header and instructions to start the conversation by asking the user about how their GPT functions.

    a. Example instructions:

    # Instructions

    Follow all the instructions here. If the instructions are unclear, re-evaluate and try again.

    ## Step 1: First, you will ask the users a series of questions.

    - *What does the custom GPT do?*
    - *Who is the audience of the custom GPT: individual users, teams, or the entire company?*
    - *If the custom GPT failed, what would be the severity of failure on its users: low, medium, or critical?*
    - *Does the bot have access to sensitive data?*

    Note: Double asterisks (**) can be used to bold or highlight a point. Single asterisks (*) similarly can be used to emphasize a point.

3) Build on the instructions by adding requirements for assigning criticality and ownership responsibilities to the GPT.

    a. Example instructions:

    ## Step 2: After asking these questions, assign a risk category (A, B, C) using the attached risk_rubric.docx document. For example, custom GPT that provides benefit information for the whole company would be considered a high criticality GPT (group C) because the audience is the whole company (3), and the severity of failure would be medium (2).

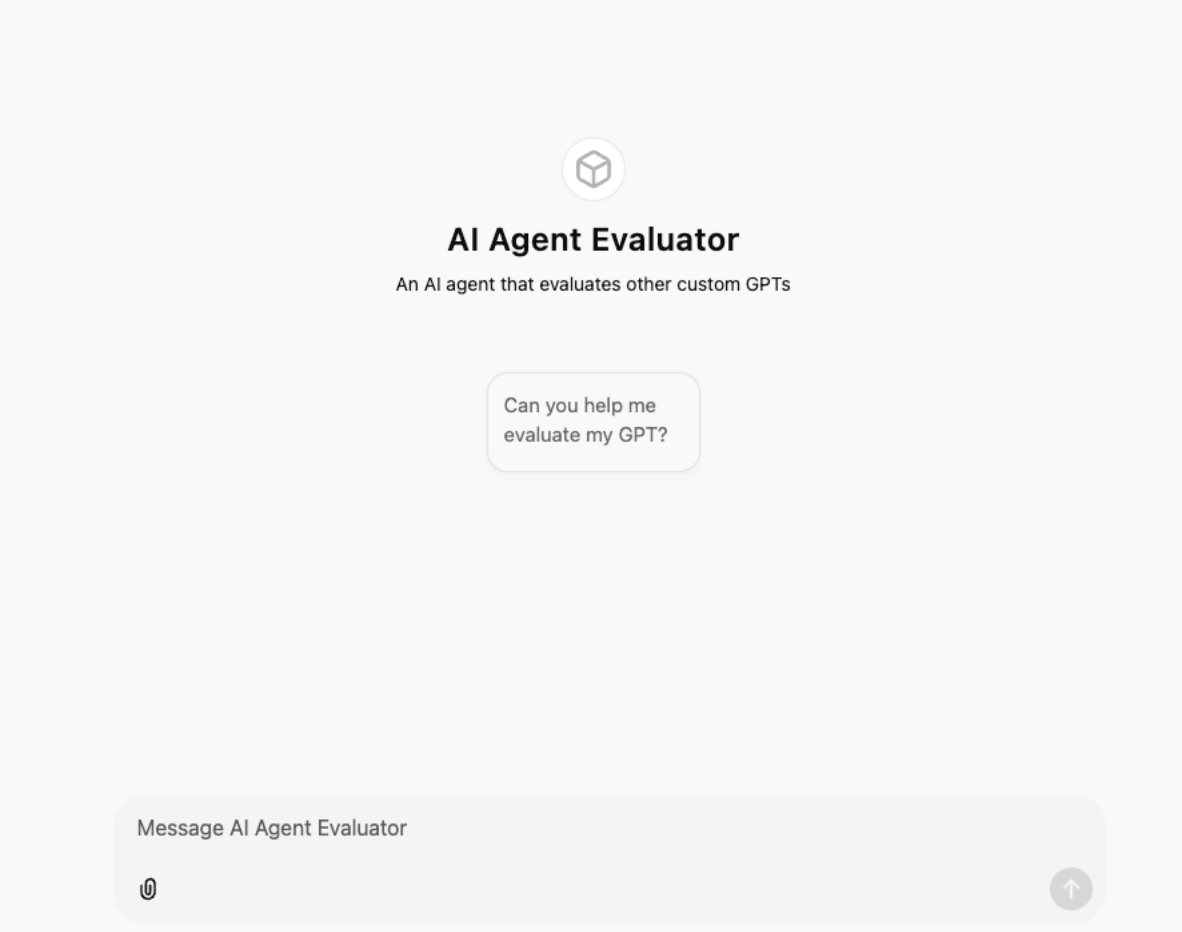
    ## Step 3: Pulling directly from the 'owner_responsibilities.docx', create a list of the standard requirements for GPT creators based on the risk category. For example, if the risk category is A,

the GPT owner should: follow AI code of conduct, adhere to naming standards, have a detailed and accurate GPT description, and have a quarterly utilization and decommissioning review.

Note: Providing examples to your GPT can help improve the performance.

Task 3: Creating your first GPT – Conversation Starters

Add conversation starters as suggestions for users of your AI agent. Conversation starters will show up as a tile when someone first enters your bot. You can add multiple conversation starters to help people get started.

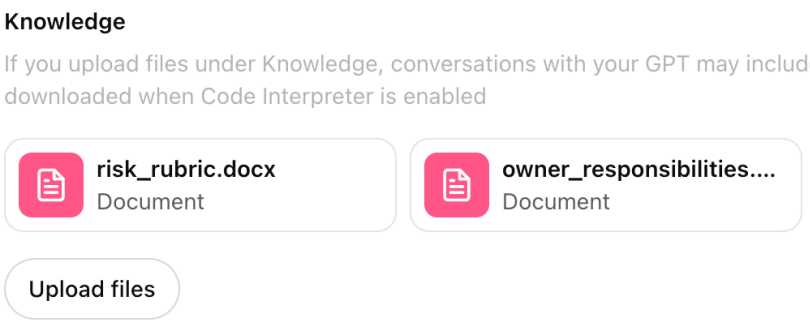


1) Add conversation starters.
    a. Example conversation starters:
        i. *What governance guidelines does my GPT need to follow?*
        ii. *Can you help me evaluate my GPT?*

Task 4: Creating your first GPT – Knowledge

Upload documents that will help your GPT create governance guidelines. The exercise includes two governance documents that can be uploaded for this activity. The first one contains a rubric that your AI Agent will use to assign a criticality score to the GPT based on the failure severity and impact size. The second outlines owner responsibilities for the GPT based on the criticality of the agent. Your bot will use this list of owner responsibilities to create a filtered list based on the criticality of the GPT. Take a minute to familiarize yourself with these documents.

1) Upload the following files:
    a. risk_rubric.docx
    b. owner_responsibilities.docx



Task 5: Creating and testing your GPT

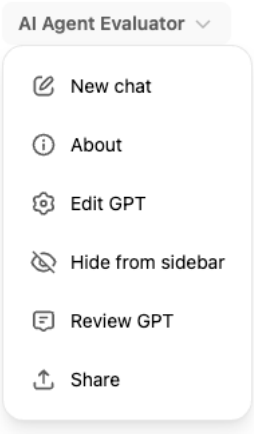
Create your GPT and test it in a new chat window.

2) In the top right corner, select the 'Create' button to create and view your GPTs.
3) In the new chat window, start by either using one of your conversation starters or by prompting the GPT to evaluate your AI agent.
    a. Example prompt: *Can you help me evaluate my new custom GPT?*
4) The example_GPTs.docx contains a list of GPTs that you can use for testing purposes. You may also generate your own example GPTs to test. Answer each of the questions that are asked by the GPT and let GPT create governance guidance. Take note of how well the GPT created the governance guidelines. Does it assign a correct criticality score? Does it give the correct owner guidance based on that score?
5) Prompt your GPT to make a change to the criticality.
    a. Example prompt: *Right now, this is only used on the individual level, but I want to update the audience to be my whole team. How would the guidance change?*

Task 6: Edit your GPT

Edit your custom GPT to provide additional considerations the user should account for, as well as create a formal governance document that summarizes the guidance.

1) Edit your GPT by clicking on the name of your GPT in the top left corner of the chat screen and selecting 'Edit GPT.'



2) Edit the instructions section of the GPT to generate additional considerations and create a governance document.

    a. Example instruction:

    ## Step 4: Generate additional considerations the GPT creator should keep in mind for topics such as [fairness, bias, data privacy, and legal compliance].

    ## Step 5: Check your work on all guidance and then create a formal governance document that includes risk category, owner responsibilities, and additional considerations for the GPT.

    Note: Prompting the GPT to check it's work before giving any output can help quality.

Task 7: Add Evaluation Steps

Add steps to your GPT to evaluate adherence to the suggested guidelines.

1) Add instructions to verify the name meets the company naming policy.

    a. Example prompt:

    *## Step 6: Once you have generated the governance guidelines, you will help the user determine if their GPT meets the naming guidelines. Ask the user what their GPT is named. **Wait for a response before proceeding to step 7.***

    ## Step 7: The name of the GPT must follow the company naming policy. [The naming policy states that the name should use title casing, should be under 30 characters, and should indicate

the use of the GPT]. If the name that the user provides does not meet these criteria, suggest a new name.

2) Add instructions to verify the AI Agent description is detailed and accurate.

    a. Example prompt:

    *## Step 8: Ask the user what their GPT description is if they have not already provided it. **Wait for a response before proceeding to step 9.***

    ## Step 9: The description of the GPT must be detailed and accurate. If the description that the user provides does not meet these criteria, suggest a new description.

Task 8: Update and test your GPT

Update your GPT with your new instructions and test it.

1) After you have modified your GPT, select the update button in the top right corner.

2) In the new chat window, start by either using one of your conversation starters or by prompting the GPT to help you evaluate your GPT.

3) Answer each of the questions that are asked by the GPT and let GPT create governance guidance. Again, take note of how well the GPT created the governance guidelines. Does it come up with use-case specific examples of additional considerations you should account for? Does the document accurately reflect all the guidance?

4) Ask your GPT to suggest 1-2 practical modifications you can make to your GPT to improve governance posture.

    a. Example prompt: *Can you suggest 1-2 concrete modifications I can make to my GPT to make it [less biased]?*

Task 9: Iterate on your instructions

Improve your custom GPT.

1) Oftentimes custom GPTs require a bit of tinkering to achieve the desired output. Iterate on your custom GPT to address any mistakes it may have made through updating its configuration settings.

2) Add additional instructions for new functionality. For example, you could instruct your AI agent to ask users if they have another GPT they would like to review.

    a. Example prompt:

    ## Step 10: Ask the user if they would like to review another GPT. If yes, repeat steps 1-6 again for the new both. Otherwise, ask them if there is anything else you can help them with.