Commonwealth of Massachusetts AI Assistant & Implementation Partner Response to questions

OpenAI (via Carahsoft)
10.30.25

Proprietary & Confidential — Not For Public Distribution
This document contains confidential and proprietary information of the Offeror. Unauthorized disclosure, reproduction, or distribution is prohibited.

Questions in blue, Responses in black

Accessibility Conformance Report (ACR):

Please submit a current ACR based on the latest version of the Voluntary Product Accessibility Template (VPAT). The report should evaluate conformance with WCAG 2.1 or WCAG 2.2 Level A and AA success criteria.

Relevant documents are attached: OpenAI-Accessibility Rider.pdf, OpenAI-VPAT.pdf

Best and Final Offer (BAFO):
Please provide a comprehensive BAFO that includes implementation, training, and ongoing support costs for the following license tiers for Enterprise Core:

10,000 licenses (annual fee)
20,000 licenses (annual fee)
30,000 licenses (annual fee)
40,000 licenses (annual fee)

Core Platform Annual Fee (seat price)

10k = $1,560,000 ($13/seat)

20k = $2,640,00 ($11/seat)

30k = $3,240,000 ($9/seat)

40k = $3,360,000 ($7/seat)

90 days of Enablement and customer support for the duration of the signed contract is included in the price. A dedicated enablement partner will be available for the duration of the first year in the 25,000+ seats tier. Here is an overview of our

Enablement options by tier:

Enablement Element / Tier	10k-29,999k seats	30k+ seats
Approach	Launch strategy tailored to your organization's goals. OpenAI's experts collaborate directly with your stakeholders to plan the rollout, deliver executive and leadership training, and set the foundation for long-term adoption across teams.	Strategic partnership to guide your AI adoption journey. OpenAI works side-by-side with your team to design a deployment plan, build internal champions, and ensure measurable long-term results. The engagement includes practical, role-based training—helping employees apply AI tools directly in their daily work to drive real productivity gains and organizational change.
Personnel	Shared Expert - Executes a successful launch, enables your team to build and replicate the launch and enablement motion, and provides QA and escalation around issues.	Dedicated Strategic Expert - Consistent Partner and POC for the year. Your Executive teammate to drive true transformation across the organization, departments, etc. Deep feature dives, assists with new feature roll-out adoption and advocates for your organization within OpenAI. Up to date best practices from OpenAI's commercial and government adoption practices.
Executive Workshops	Yes- Limited	Yes- In person training
Executive Light-House Use Cases	2-3* *depends on complexity of use case*	3-6* *Depends on complexity of use cases*
Train the Trainers	Yes-Limited	Yes- In person training
Active Champion Network	Initiated in Launch	Initiated in Launch + supported throughout the year
Custom Live Trainings	For Launch	Launch + throughout the year
On-sites, Hackathons, other	Yes, for kick off	Yes- Kick-Off, plus, multiple
Webinars/ OpenAI Academy	Yes	Yes
Dedicated Enterprise Resource Hub	Yes- Standard Enterprise	Yes, with some potential customization of resources
Value- Survey and Benchmarks	standard	custom
Communication Cadence	Launch Cadence, then Hand-off cadence, then support cadence	Regularized and consistent cadence, initiative and light-house sprints.
Kick-Off and Planning	Standard	In-person, virtual, and customized
Co-Lead Champion Sessions	1-2	2+ (Depending on right fit for organizational structure and culture)
Full- Leadership Engagement Pack/ Communications plan	General Roll-out plan	Detailed Suggested Roll-out script, including best practice communications for leaders creating AI-enabled organizations.
Role- Based Trainings	Virtual, Recorded	More customized, GPT and automation focus
Playbook	Standard	Customized
Launch Sessions	101, 102, Feature Deep Dives- mainly virtual	101, 102, Feature Deep Dives, New Feature Roll-outs, virtual, custom recorded and in person
Prompt Pack	Pre-Generated	Custom Live Generation Tool

1. Please describe how credits are consumed and tracked, including concrete examples by function or use case.
2. Please provide an estimated annual credit usage for a typical policy analyst.
3. Do you offer an alternative licensing model for advanced features based on a fixed monthly fee? If yes, please describe.

a. Credits consumption and tracking (with examples)

ChatGPT Enterprise uses Advanced Model Units (AMUs) to measure and manage consumption of advanced model capabilities (e.g., GPT-5 Thinking and Pro models, Image Generation, ChatGPT Agent, Deep Research, Multimodal Chat).

- Each AMU represents a fixed quantity of compute for advanced model interactions.
- Credits are consumed per message or task depending on model complexity, length of input/output, and whether advanced features (such as image generation or code execution) are used. Please refer to the below table for details on AMUs consumed by common advanced features or models. An up-to-date rate card is maintained at https://help.openai.com/en/articles/11481834-chatgpt-rate-card

Advanced Model / Feature	Unit	Credits
GPT-5 Thinking	1 message	10 credits
GPT-5 Pro	1 message	50 credits
Agent	1 message	30 credits
Deep Research	1 task	50 credits
ImageGen	1 message	5 credits
Voice	1 minute	5 credits

- All usage of AMUs is centrally tracked through the Admin Console, allowing administrators to monitor consumption across the entire workspace or by individual users. Admins can additionally disable advanced features like Agent or Deep Research for custom roles in order to limit AMU consumption by team if needed. If AMUs are depleted, all users still have access to a fully functioning Core model.
- Admins can set Overage Limits and Credit Usage Alerts from the Workspace Settings for even greater control and visibility.
- We are including 60 days of unlimited AMUs for the Commonwealth from the start date of the contract for the Commonwealth to evaluate the efficacy of the AMUs and pricing model variables.

Examples by role:

- Technical Staff (Data/IT): analyzing datasets or writing scripts → ~30–50 AMUs/day.
- Policy Analyst: drafting legislation summaries, comparing multi-page bills → ~15–25 AMUs/day.
- Grants Administrator: rewriting compliance documentation with attachments → ~10–15 AMUs/day.

b. Estimated annual usage for a typical policy analyst

Based on pilot programs and enterprise benchmarks, a typical policy analyst uses **~3,500–5,000 AMUs annually**, depending on the volume of longform or analytical tasks. This corresponds roughly to 10–20 AMUs per workday (assuming 240 active

workdays per year). For context, this usage pattern aligns with other analytical functions in state and local government where ChatGPT Enterprise supports drafting, summarization, and policy comparison.

c. Alternative licensing model for advanced features

Although we do not offer an alternative pricing model, we can implement granular usage thresholds (at the user, group, workspace level based on your requirements) to ensure alignment with monthly budget maximums.

4.   Concurrent Users

Are you able to offer a concurrent user licensing option?

Yes, our proposal uses an enterprise concurrent model (not named users) so agencies can reallocate seats without needing to purchase additional licenses. A concurrent user licensing option lets an organization purchase a pool of licenses that may be reallocated among the organization over time. In this structure, only a fixed number of users may be provisioned at the same time, but if one user is deprovisioned, that opens a seat to provision a different user. In other words, this option places a cap on simultaneously provisioned users rather than distinct named or unique users.