Разработване на Custom GPT за комуникация и популяризиране на резултати от научно-изследователски проекти: базови принципи и насоки

Стоянка Иванова

Development of Custom GPT for Communication and Dissemination of Scientific Research Results: Fundamental Principles and Guidelines

Stoyanka Ivanova

Abstract:
The article presents an approach to creating customized (Custom) GPT models by OpenAI, aimed at communicating and disseminating scientific research results. Fundamental principles for planning and structuring supplemental knowledge bases are introduced, tailored to different user categories—free and paid—each having access to context windows of varying sizes. Strategies for integrating textual documents, tabular data, and programming code are discussed to effectively extend GPT model capabilities. Emphasis is placed on the significance of these solutions in educational processes, highlighting the role of Custom GPT in facilitating learning through clear explanations of complex scientific concepts and the automated provision of structured knowledge, which will be increasingly important in the future. A practical example illustrates how Custom GPT enhances both the dissemination of completed scientific projects and effective support for educational activities in academic environments.

Keywords: artificial intelligence, GPT, science communication, science popularization, personalized models, climate analyses, GIS

For contacts: Prof. PhD Arch. Stoyanka Ivanova, UACEG – Sofia, siva_fce@uacg.bg

ВЪВЕДЕНИЕ

С навлизането на големите езикови модели (LLM) като GPT, дигиталната трансформация на образованието и научната комуникация преминава към нов етап. Тези модели не само че извличат знания от обширни обеми текст, но и създават нови възможности за взаимодействие между хора и информация – включително чрез създаване на персонализирани ботове за специфични задачи. Технологията на Custom GPT, достъпна от ноември 2023 г. [1] е достъпна за създаване само от платени потребители (Plus и Pro), но веднъж създаден, ботът може да бъде използван от всички видове потребители. Това дава възможност за изграждане на ботове, които комбинират езиковата мощ на GPT-4o и GPT-4.5 [2, 3] с локално подадени знания, организирани по начин, съобразен с нуждите на целева аудитория.

Настоящата разработка представя подход за създаване на потребителски GPT модел, предназначен за комуникация и популяризиране на научно-изследователски резултати. Разглежда се както теоретичната основа на LLM технологиите и методите за подготовка на данни, така и практически пример – ClimaGeo GPT, създаден за обяснение на климатични анализи в ГИС среда. Основната идея е не просто да се създаде инструмент за въпроси и отговори, а интелигентен комуникатор, който адаптира обясненията си към нивото на потребителя и насочва към ключови източници.

Предложената методика се базира на следните принципи: ясна цел и аудитория, структура на информацията, пригодна за семантично търсене, и гъвкавост при комбиниране на формати – текст, таблици, JSON. Подходът има приложение както в научно-изследователска среда, така и в образованието, като улеснява разбирането и работата със сложни анализи.

Целта на публикацията е да формулира практически насоки за създаване на ефективен Custom GPT, който не само отговаря на запитвания, но и служи като мост между експертно знание и разбираем език. Показано е как добре структуриран бот може да увеличи достъпността на научната информация, да подпомогне обучението и да подобри публичната комуникация на резултатите от научни проекти.

ИЗЛОЖЕНИЕ

Методология и основни принципи

Изграждането на персонализиран GPT модел започва с дефиниране на целта и типа аудитория. Различните потребители – изследователи, преподаватели, студенти и неспециалисти – имат различни нужди, което изисква различен подход към формата, обхвата и дълбочината на предоставената информация. Отговорите на модела трябва да са адаптирани към нивото на компетентност и конкретния интерес на събеседника. Това изисква прецизно формулирани инструкции, които насочват поведението на модела [4], както и подходящ подбор и структуриране на знанията, с които той работи.

Информацията, която се предоставя на потребителския бот, не се използва за обучение в класическия смисъл. Вместо това тя се подава под формата на допълнителни текстове, към които моделът получава достъп чрез семантично търсене. Това предполага внимателен подбор на формати и структура. Най-ефективни са .txt файловете, които съдържат добре структурирана информация без излишно форматиране. DOC и PDF документи могат да създадат проблеми поради скритото си форматиране. Таблици (.csv, .xlsx) и JSON файлове обогатяват възможностите за обработка на данни и извършване на изчисления, включително чрез код на Python.

Структура и видове данни за GPT модели

Custom GPT моделите използват допълнителната информация като контекстна база знания. Тя не се зарежда еднократно, а е достъпна за търсене по ключови думи, семантична близост и логическа структура. В този контекст е от съществено значение да се осигури добра навигация в данните – чрез заглавия, секции, списъци, таблици и речници. Контекстният прозорец (обемът информация, който GPT може да обработи наведнъж) също има значение: платените потребители имат достъп до 128 000 токена, докато неплатените – до 32 000 токена, което влияе върху обема на информация, който може да бъде използван в един отговор или поредица от взаимодействия.

Текстовите файлове са подходящи за обяснителен текст, инструктаж и речник, табличните файлове са предпочитани за числови резултати, индекси, прогнози и сравнителни данни. JSON файловете са особено мощен инструмент – освен за структурирани данни, те могат да съдържат и код, който GPT моделът

може да изпълнява (например за изчисления, генериране на графики или конвертиране на единици) [5].

Необходимо е също така да се съобразят ограниченията на системата – като например максималният брой файлове (до 20) [6], максимален размер на файл (512 MB) [7] и общият лимит от два милиона токена. Това налага икономичен и стратегически подбор на съдържание. Препоръчва се използването на кратки обяснителни фрагменти, селектирани въпроси и ясно подредена информация. Когато се използват изображения или таблици в DOCX или PDF файлове, те трябва да бъдат с висок контраст и ясни надписи, тъй като GPT извлича текстова информация, но не разчита пълноценно графични елементи.

За да бъде успешен един GPT модел, не е достатъчно просто да се предостави голям обем знания. Решаващо значение имат формата на поднасяне, семантичната яснота и способността на структурата да ориентира модела при търсене и формиране на отговор. Добре структуриран файл с няколко стотин думи може да бъде значително по-ефективен от слабо организиран документ с десетки страници.

Практически пример: ClimaGeo GPT

ClimaGeo GPT е персонализиран езиков модел [8], създаден в рамките на научен проект към Центъра за научноизследователска и проектантска дейност (ЦНИП) на УАСГ. Основната му функция е да подпомага комуникацията и разбирането на климатични анализи (Таблица 1), извършвани с помощта на ГИС технологии, като представя резултатите на разбираем език и адаптира отговорите си спрямо нивото на потребителя – от експерт до студент.

Моделът използва добре структурирана база знания, съставена от разнообразни файлове (подробно описани в [9]), включително:

- текстови документи с описание на проекта, цели, използвани термини и често задавани въпроси;
- таблични файлове с резултати от климатични анализи по сценарии и региони [10];
- речник с обяснения на ключови термини;
- JSON файлове с инструкции и логически зависимости.

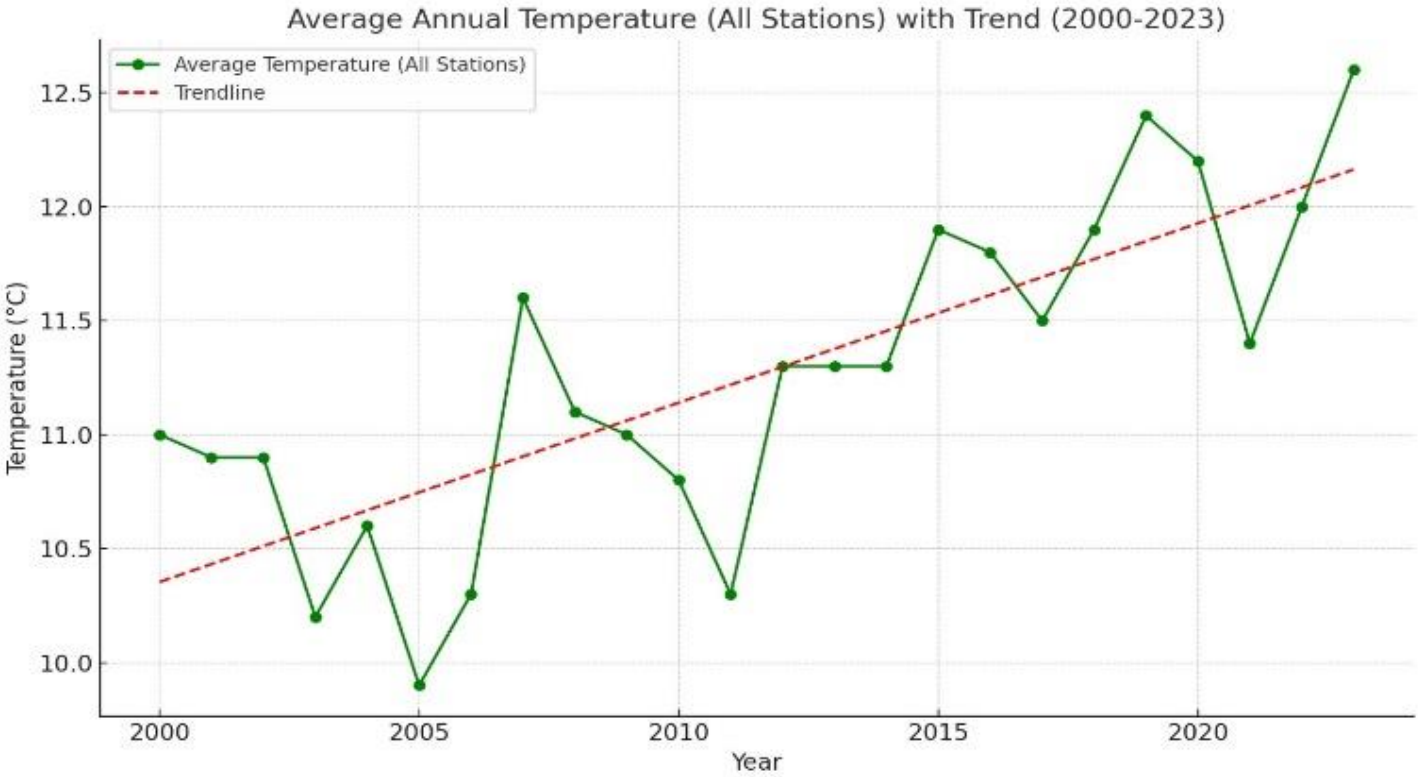
Например, ботът може да отговори както на въпрос от типа „Какви са очакваните температурни аномалии по сценарий RCP 8.5 за периода 2071–2099 г.?“, така и на по-достъпен въпрос от студент: „Ще стане ли по-горещо в Русе след 2050 година?“.

ClimaGeo GPT служи не просто като търсачка в собствената си база знания, а като активен медиатор между сложна научна информация и човешко разбиране. Осигурява достъп до обобщени резултати (фиг. 1), навигация в съдържанието и препраща към допълнителни източници, когато това е необходимо. Той може да бъде използван в:

- обучението на студенти по климатология, ГИС и енергийна ефективност;
- подпомагане на научно-изследователски екипи при планиране и анализ;
- повишаване на осведомеността на широка публика относно климатичните промени.

Таблица 1. Информация за климатичните промени, очаквани в България, според различните климатични сценарии. Източник: [10].

Климатичен показател	Сценарий RCP4.5 (2021-2050 г.)	Сценарий RCP4.5 (2071-2099 г.)	Сценарий RCP8.5 (2021-2050 г.)	Сценарий RCP8.5 (2071-2099 г.)
Средногодишна температура	+1.2 °C	+2.5 °C	+1.5 °C	+4.0 °C
Зимна температура	+0.8 °C	+1.6 °C	+1.0 °C	+2.8 °C
Лятна температура	+1.5 °C	+3.0 °C	+2.0 °C	+5.0 °C
Валежи (годишен обем)	-5%	-10%	-10%	-25%
Летни валежи	-15%	-20%	-25%	-35%
Зимни валежи	Без големи промени	Леко увеличение	Без големи промени	Леко увеличение
Индекси на засушаване	Умерено увеличение	Силно увеличение	Значително увеличение	Силно засушаване



Фиг. 1. Графика, показваща средната годишна температура за всички метеорологични станции в България от 2000 до 2023 г., създадена от ClimaGeo GPT [8] на базата на данни от [10].

Дискусия

Примерът с ClimaGeo GPT потвърждава значението на добре организираната база знания и внимателно формулираните инструкции за успешната работа на един потребителски бот. Една от най-ценните страни на разработката е, че самото

изграждане на модела подпомага екипа, който го създава – чрез изясняване на понятията, систематизация на информацията и по-добро разбиране на логиката на проекта.

Изграждането на ClimaGeo GPT изискваше не само специализирани познания за GPT, но и стратегическо мислене при подбора на информация и нейната организация. Процесът на структуриране включваше:

- формулиране на описания и обяснения в ясен и недвусмислен стил;
- създаване на речник на термините, насочен към различни аудитории;
- подбор и обобщение на основни резултати, представени в табличен вид;
- конструиране на файл с често задавани въпроси и техните отговори;
- организиране на данни по сценарии и региони с цел сравнение и анализ.

По време на тестването на модела стана ясно, че потребителите очакват не само отговори, но и аргументация – връзка между концепции, обяснение на причинно-следствени зависимости и насоки за по-нататъшно изследване. Това налага GPT ботът да бъде не просто интерфейс за търсене, а партньор в разбирането.

Взаимодействието с ClimaGeo GPT води до повишена информираност, активиране на любопитство и задълбочаване на знанията на потребителите. Това го прави ценен инструмент в академична и изследователска среда, където достъпът до ясно представена, структурирана и актуална информация е от съществено значение.

Особено важно е, че GPT моделът показва адаптивност към стила на задаване на въпроси – от технически до неформален език – и може да „превежда“ съдържание според профила на потребителя. Това го прави приложим както в комуникацията със специалисти, така и в образователни инициативи и в популяризацията на науката пред широката публика.

В този контекст ClimaGeo GPT демонстрира възможностите на съвременните езикови модели не само да обработват знания, но и да подпомагат изграждането на нови когнитивни връзки у потребителя – нещо, което досега беше запазена територия на преподавателя и научния комуникатор.

ЗАКЛЮЧЕНИЕ

Разработката на потребителски GPT модел като ClimaGeo GPT показва значимия потенциал на изкуствения интелект като инструмент за научна комуникация, образователна подкрепа и публично разбиране на сложни аналитични резултати. Проектът демонстрира как внимателното планиране, добре структурираното съдържание и правилната стратегия за интеграция на знания могат да трансформират голям езиков модел в ефективен, достъпен и полезен партньор на широк кръг потребители.

Един от ключовите изводи е, че ефективността на модела не зависи толкова от количеството информация, колкото от нейната организираност, яснота и адаптивност. Ясното адресиране на различни нива на експертиза – от специалист до неспециалист – е от решаващо значение за постигане на комуникационна ефективност.

ClimaGeo GPT показа, че дори с ограничени ресурси и без програмиране, е възможно да се създаде мощен помощен инструмент, ако е налице стратегическо

мислене и междудисциплинарен подход. Интеграцията на файлове с термини, обяснения, въпроси и таблични данни в подходящ формат е не само технически възможна, но и препоръчителна практика за всеки създател на потребителски GPT бот.

Насоките за бъдеща работа включват:

- автоматизация чрез добавяне на Python код и изчислителни модули;
- свързване с външни бази данни и онлайн визуализации;
- интеграция със системи за обучение и интерактивни среди;
- развитие на инструменти за анализ на потребителската активност и обратна връзка.

В заключение, персонализираните GPT модели се очертават като ефективен мост между знанията, съхранявани в научни и образователни среди, и реалните потребности на обществото. Развитието на такива инструменти изисква не само техническа компетентност, но и разбиране за начина, по който хората учат, задават въпроси и формират познание.

ЛИТЕРАТУРА

[1] OpenAI Blog. Introducing GPTs. https://openai.com/index/introducing-gpts

[2] OpenAI Blog. Hello GPT-4o. https://openai.com/index/hello-gpt-4o/

[3] OpenAI Blog. Introducing GPT-4.5. https://openai.com/index/introducing-gpt-4-5/

[4] OpenAI. Guidelines for writing instructions for custom GPTs. https://help.openai.com/en/articles/9358033-key-guidelines-for-writing-instructions-for-custom-gpts

[5] OpenAI. Tokenizer. https://platform.openai.com/tokenizer

[6] OpenAI. How many files can I upload at once per GPT? https://help.openai.com/en/articles/8983703-how-many-files-can-i-upload-at-once-per-gpt

[7] OpenAI. What are the file upload size restrictions? https://help.openai.com/en/articles/8983719-what-are-the-file-upload-size-restrictions

[8] ClimaGeo GPT. https://chatgpt.com/g/g-6736724cee708190ab89c76cdfa36c4b-climageo-gpt

[9] Ivanova, S., Ilieva, T. Development of Custom GPT as a System for Communication and Promotion of Climate Analysis in a GIS Environment: Planning, Structure and Application in Projects. // Proceedings of IX International Scientific Conference Industry 4.0, Winter Session, 2024, pp. 101–104.

[10] НИМХ. Променящият се климат на България – анализи и прогнози. https://www.meteo.bg/meteo7/sites/storm.cfd.meteo.bg.meteo7/files/kniga_klim_promeni_23-12-2023_KM_ff.pdf