



ИНСТИТУТ
ПО ПУБЛИЧНА
АДМИНИСТРАЦИЯ

ДЕМИСТИФИЦИРАНЕ НА ПРОЦЕСА ПО СЪЗДАВАНЕ НА ДИГИТАЛНИ АСИСТЕНТИ:

ОПИТЪТ НА ПЪРВАТА
ДЪРЖАВНА ИНСТИТУЦИЯ
С ИИ В БЪЛГАРИЯ

ФЕВРУАРИ 2025

Дигиталните асистенти не са нещо мистично или сложно за разбиране. Те са инструменти със съвсем конкретни възможности и ограничения. За успешното им внедряване е важно да разберем, че те са помощници, създадени за решаване на конкретни задачи, не универсални решения. Те изискват внимателно планиране и постоянна поддръжка, а успехът им зависи от качеството на информацията, с която разполагат.

Как работи дигиталният асистент?

Представете си служител в информационен център – нека това бъде Мария, която познава отлично цялата документация на институцията, знае наизуст всички най-често задавани въпроси и отговори и когато при нея дойде гражданин с въпрос, тя може много бързо да съдейства без да губи време. Дигиталният асистент работи по подобен начин: той има "база знание" (това формално и неформално знание на Мария), която е създадена от експертите в институцията. Когато някой зададе въпрос, то асистентът:

- Първо търси в тази база подходящата информация.
- Ако е сигурен в отговора, го предоставя веднага.
- Ако не е напълно сигурен, то насочва към човек от екипа.
- Може да отговаря на множество хора едновременно, по всяко време - 24/7.

За разлика от Мария обаче, дигиталният асистент не може да взема самостоятелни решения или да прави изключения от правилата. Затова при по-сложни случаи винаги насочва към експерт. Също като Мария, дигиталният асистент се учи от опита чрез експертите от институцията, които помагат като наблюдават неговите отговори и постоянно подобряват базата му от знания, за да става все по-полезен. Както професионалните спортисти тренират почти всеки ден, така и експертите тренират асистента, изисква се воля и целеустременост.

Опитът с ЧИПА и техническите аспекти

Дигиталният асистент на Института по публична администрация, чат-ботът **ЧИПА**, в тази негова версия, е разработен с две основни цели:

- 1) Бързо предоставяне на релевантна информация.
- 2) Съдействие по въпроси за обученията на ИПА.

Нашият дигитален асистент е разработен с помощта на GPT-4o-mini, като е интегриран в система, която комбинира векторна база данни и алгоритми за извличане на релевантна информация. ИТ експертите, с които работихме и които доброволно отделиха от времето си, създадоха векторна база данни, в която бяха съхранени предварително обработени от нас документи.

Важен момент беше подборът на информация – решихме асистентът да работи с предварително подбрани и проверени данни, вместо да разчита на информация от интернет и сайта на Института. Събрахме както релевантната публична информация за ИПА, така и често задавани въпроси по телефон и имейл, информация за различни инициативи и професионални мрежи, поддържани от ИПА.

За да направим документацията разбираема за езиковия модел, тя беше трансформирана чрез векторизация – процес, при който текстът се представя като математически вектори в многомерно пространство. Това позволява търсенето на информация не чрез точни съвпадения на думи, а въз основа на смисловата близост.

За да насочим основния модел в желаната от нас посока, използвахме специфична формулировка на първоначалния prompt/запитване, който задава целия тон на комуникацията на асистента. Там изрично сме заложили ограниченията, които ръководят кореспонденцията и споделяната информация.

Дигиталният асистент използва MMR (Maximal Marginal Relevance) алгоритъм за оптимизиране на информационното извличане. MMR алгоритъмът помага за балансиране между:

- Релевантност – доколко даден документ е свързан с потребителския въпрос.
- Разнообразие – избягване на твърде сходни резултати, за да се представи по-широк контекст.

Този баланс се управлява чрез параметър, който определя относителната тежест на двете характеристики. В нашия случай акцентът е поставен върху релевантността, за да се гарантира, че отговорите са максимално точни и близки до потребителското запитване.

За да гарантираме безопасността и коректността на взаимодействията, приложихме следните мерки:

- Защита срещу злонамерени атаки – филтриране на запитвания, които могат да манипулират модела или да го изкарат извън контекста.
- Проследяване чрез логове – анализ на потребителските взаимодействия за подобряване на точността и откриване на потенциални проблеми.
- Ограничение на сесиите – всяка потребителска сесия е ограничена до 50 заявки, след което се рестартира автоматично. Това предотвратява прекомерното натрупване на контекст и потенциални опити за експлоатация.

Важен детайл е, че така подредената база данни и модел на ИПА може да бъде „закачен“ към алтернативни на Open AI системи за ИИ като BGGpt или други.

Предизвикателства и решения

Когато създавате дигитални асистенти, ще срещнете различни предизвикателства – от технически проблеми до случаи на "халюцинации" на асистента. Това се случи и с ЧИПА. Тук от особена важност беше, че екипът по внедряването спазваше няколко важни правила:

- Постоянно наблюдение и корекции, основно извън стандартното работно време за администрацията;
- Първоначално включване в теста на 5 човека и лично изпълнителният директор, поетапно включване на други експерти от ИПА, привличане на доброволци за тестване от различни институции и общности в системата на държавната администрация;
- Изключително строг контрол върху правилните отговори;
- Редовно обогатяване на базата знания;
- Спазване на принципите за етичност и добросъвестност и на добри практики за киберсигурност,

в резултат на които изготвихме и условията за ползване на ЧИПА, които да отразяват тези наши принципи <https://www.ipa.government.bg/bg/usloviya-za-polzvane-na-chipa-chatbot-na-ipa>.

Включването на външни експерти и колеги в процеса на тестване се оказва изключително ценно не само за валидацията на модела, но и като катализатор за неговото усъвършенстване. Техният поглед "отвън" съдейства за обогатяване на контекстуалната база от знания чрез различните професионални перспективи и идентифициране на допълнителни сценарии за приложение.

Развитие

Моделът е проектиран като "жив" организъм - той продължава да се обучава и усъвършенства въз основа на взаимодействията с потребителите на ИПА. За да се гарантира качеството на отговорите се използва контролиран подход към тренирането, като всяка нова информация преминава през процес на валидация, преди да бъде интегрирана в знанията на модела.

Създаването на дигитален асистент е и възможност за развитие на компетентностите на екипите, защото всички ангажирани с процеса се учехме как да бъдат преодолените различни предизвикателства - учене през целия живот в действие ☺. Екипна работа.

ИПА има по-големи амбиции и ще тества лимитите на ИИ като основната цел ще е, наред с дейността по отговор на въпроси, да се персонализира съдържание и обучение, да се пилотира дигитален помощник в процеса на въвеждане на КАФ, и всички възможни дейности за подобряване на обучителното преживяване на държавните служители.

А защо да го направим и ние, ще се запитате?

Дигиталният асистент е този ваш колега, който:

- Реагира на момента по въпроси на потребители.
- Отговаря търпеливо на повтарящи се въпроси.
- Помага на много хора едновременно.
- Осигурява равен достъп до информация, през цялото денонощие.
- Помага ви да работите по-добре, да се концентрирате изцяло на същинската ви работа, когато вие имате срокове и спешни задачи, които изискват цялото Ви внимание.

За администрацията това означава по-ефективна работа, а за гражданите – по-достъпни услуги по всяко време.

Дигиталният асистент не замества служителите, а им помага, като поема бюрократичната работа, работата по повтарящите се въпроси. Това ни оставя повече време за по-важните задачи, които изискват човешка преценка и внимание. В крайна сметка, смисълът е работата на администрацията да бъде по-ефективна чрез този нов, дигитален помощник в екипа, който да оставя време за същинската аналитична работа.

Автори: Таня Иванова-Чикова, Павел Иванов, Ясен Танев, Станислав Атанасов.

София, февруари 2025 г.

С подкрепата на Европейски цифров иновационен хъб "Тракия".

Благодарности за съдействието на Цветелина Калчева, Петя Димитрова, Тодор Николов.