



ИНСТИТУТ
ПО ПУБЛИЧНА
АДМИНИСТРАЦИЯ



ГЕНЕРАТОР НА ИДЕИ

Лидерство и надежден
изкуствен интелект

Несебър, май 2026

ЗА НАСТОЯЩИЯ СБОРНИК

Настоящият материал е разработен в рамките на уъркшоп „Генератор на идеи: Лидерство и надежден изкуствен интелект“, в който взеха участие 38 представители на мрежите на иноваторите и на младите лидери в държавната администрация, поддържани и развивани от Института по публична администрация. Обучението се проведе през месец май 2026 година в рамките на три дни.

Лекторите Ясен Танев и Станислав Атанасов разпределиха участниците в шест екипа, всеки от които получи задачата да оцени определен голям езиков модел, използвайки четири стандартизирани шаблона – SWOT анализ, Risk Canvas (риск-регистри), карта за приложимост и методологични указания. Шаблоните бяха предоставени с готова структура за постъпково оценяване, което да доведе всеки екип до формулиране на собствени методологични изводи и препоръки. По време на цялото събитие лекторите модерираха и подпомагаха работата на екипите с обратна връзка и насоки. При изготвянето на разработките участниците използваха изкуствен интелект като помощен инструмент, като извършваха собствена проверка на резултатите. След приключването на семинара част от екипите доразвиха и финализираха своите документи.

Важно е да се подчертае какво представляват тези материали и какво не. Те не са официална държавна оценка на изброените модели, нито имат претенция да служат директно като методически указания за администрацията. Целта им е друга: да покажат чрез конкретен пример как една администрация би могла да подходи критично към оценката на ИИ модели, съобразявайки я със собствените си условия, вътрешни правила и рискове, като използва и експертна помощ на специалистите в областта на ИИ. Сборникът е замислен като насърчение за вътрешно разбиране на темата и стимул администрациите да разработват собствени методологии и подходи, съобразени с тяхната специфика.

Таблицата, поместена в началото на сборника, е замислена като навигационен преглед, а не като оценка на представянето на екипите – тя помага на читателя бързо да се ориентира кой екип какво е разработил, преди да продължи към пълните профили. Начинът на работа на всеки екип, направените обосновки, конкретните констатации и рисковите преценки могат да се видят в пълен вид в съответния профил по-нататък в документа.

Не на последно място, разработките са и илюстрация на самия формат на обучението – на това как представители на различни администрации, работещи заедно в екип за кратко и интензивно време, могат да създадат съдържателни и практически ориентирани документи, използвайки съвременни инструменти в подкрепа на своята работа.

Как да четем документите и оценките в тях: въвеждаща таблица към разработките на шестте екипа

Таблицата по-долу е навигационен преглед на разработките на шестте екипа – тя помага на читателя бързо да се ориентира кой екип какво е разработил, преди да продължи към пълните профили по-нататък в документа. Начинът на работа на всеки екип, направените обосновки, конкретните констатации и рисковите преценки могат да се видят в пълен вид в съответния профил.

Констатациите в колоните „На какво да обърне внимание читателят“, „Тестване“ и „Бележки“ не са атестация на екипите. Те илюстрират областите за подобряване и на какво да се обръща внимание при документи, създадени с ИИ; показват какво да проверява всеки, който чете подобни предложения. Оригиначните документи са запазени без корекции; където противоречие може да подведе читателя, е добавена редакторска бележка.

Общи методически изводи: само един от шестте екипа приложи документиран протокол от тестването с промптове и резултати – бъдеща оценка от този тип следва да включва такова приложение задължително. Проверката на формулата вероятност $\times$ въздействие и на връзката между детайлните оценки и крайната обща оценка е също толкова важна. И едно окуражаващо съвпадение: въпреки различната калибрация на скалите, всички седем оценки си съвпадат при критичните граници – класифицираната информация и автоматизирани решения без човешки контрол са забранени навсякъде.

Екип / Модел	Крайна позиция	Отличителен принос	На какво да обърне внимание читателят	Тестване с реални промптове	Бележки
Локи Meta Llama 4 (вътрешна инфраструктура)	Предпазливо „да“ - само пилотно внедряване при нискорискови вътрешни задачи, върху собствена инфраструктура.	Единственият екип с документиран протокол от тестване (7 теста с промптове, резултати и установени грешки) и пет работни приложения в Excel - план за тестване, лицензионен анализ, карта на съответствието, регистър на употребите и контролна рамка. Перфектен рисков регистър (0 отклонения от формулата на шаблона). Консервативни, реалистични оценки без нито една максимална.	Обосновките са без конкретни бенчмаркове.	Да - единственият документиран протокол: всеки тест със сценарий, промпт, очакван и получен резултат и установени грешки (напр. „9/10 коректни точки, една дата объркана“). Приложенията са вградени в самия SWOT документ.	Не са установени.
MAAT Claude Opus 4.7	Условно „да“ - с изрично дефинирани забранени употреби (правни становища, кризисна комуникация); най-отворена политика приложение при ежедневни задачи.	Най-плътното попълнената карта за приложимост (всяка от 84-те клетки с оценка, риск и бележка); конкретни технически бенчмаркове; категорични, защитими забрани; пълни одитни изисквания. Отклонението на крайната оценка от средната е изрично аргументирано (експлоатационна себестойност и архитектурни рискове) - съзнателно претегляне.	В рисковия регистър 10 от 30 нива се отклоняват от формулата на шаблона - девет са консервативни завишавания при критично въздействие, но едно е подценяване на ключов риск (обяснимост при автоматизирани решения). Матрицата дава „напълно приложимо“ за правни становища, които собствените им указания забраняват - техническа способност и нормативна допустимост не са разграничени.	Позоваване на публични бенчмаркове и документация.	Потвърдено по оригинала: несъществуващ модел „GPT-5.4“ е цитиран като реален конкурент - името от примерния текст на шаблона е пренесено в съдържанието без проверка.

Хефест Mistral Large 3	„Да“ - благодарение на on-premise възможностите и данните в ЕС; ограничения при чувствителни данни.	Единственият екип с паралелна оценка на два модела - разликите са смислени и последователно пренесени през шаблоните. Работният процес е видим в самия файл, автентичният човешки слой е документиран.	„Разрешено“ при описани условия (бел. ред.: вероятно се има предвид „С условия“); единствената неотметната задача е ключовата „Автоматизирани решения без човешки контрол“ - текстът я забранява, к	Екипът изрично декларира, че оценката е по публично достъпна информация.	Грешен номер на Акта за ИИ: „2024/1034“ вместо 2024/1689. „Pevstral 2“ (вероятно Pixtral).
Хефест xAI Grok 4.3 (Beta)	Висок риск - допустим само ограничен пилот и специален режим при избори, кризи и здравни теми.	Аналитично най-силният рисков регистър от всички документи: DPIA, SCC/TIA, DLP правила, multi-vendor стратегия, rollback план; ясна забрана за правни становища; личен глас и хумор в становището; най-много документирано време за работа от всички файлове.	Две реални грешки в смятането на рисковите нива (4×4=16, маркирано „Среден“ при праг 13 за „Висок“) и четири гранични отклонения - пример защо формулата в легендата изисква проверка. Оценката „препоръчваме“ в SWOT стои до „Висок риск“ в Risk Canvas - защитимо, но изисква обяснителна връзка.	Същата декларация за публично достъпна информация.	Непоследователна версия: „Grok 4.20 / 4“ срещу „Grok 4.3 (Beta)“.
Кецалкоатъл DeepSeek V3	Категорично „не“ за публичния сектор - единствената изцяло	Най-задълбоченото проучване на конкретния модел (регулации, цензура, SLA, цена, хардуер).	Структурно противоречие: Risk Canvas обявява „Неприемлив риск“, а указанията разписват правила за употреба без свързваща уговорка.	Реално промптване е видимо.	Не са установени; твърденията за модела са проверими и в общи линии коректни.
Екип / Модел	Крайна позиция	Отличителен принос	На какво да обърне внимание читателят	Тестване с реални промптове	Бележки
	отрицателна препоръка, аргументирана при признати предимства на модела.	Най-рестриктивната политика за употреба (9 забрани от 20 задачи) Последователна до крайната оценка; Комуникационни материали - комикс-сценарий за изтичане на данни и две инфографики.	Две подценени рискови нива спрямо формулата Есеистичен ИИ-пласт в становището на SWOT.		
Прометей ChatGPT 5.5 (OpenAI)	„Да“ - висок риск, управляем чрез национална рамка (централизирано договаряне, компетентностен център, регистър на ИИ системите).	Най-високото институционално-правно ниво (реални български системи и коректни правни препратки). Единствената политика без нито едно безусловно „Разрешено“ - всичко минава през условия. Най-консервативната карта за приложимост; измерими KPI с целеви стойности; изрична съгласуваност между четирите шаблона; самокритични оценки.	Стриктният рисков профил в матрицата (24 „Висок“) е по-строг от тона на крайната препоръка - посоката е защитима, но читателят трябва да стигне до нея сам.	Дълбока институционална експертиза, почти без следи от директен диалог с модела.	Без установени външни фактически грешки. „OpenAI Privacy Filter“ е посочен като конкретен продукт - наименованието следва да се верифицира преди цитиране.
Янус Google Gemini 2.5 Pro	Ентусиазирано „да“ (обща препоръка 5/5) - най-силната положителна препоръка, но и най-слабо защитената.	Най-конкретните сценарии за употреба (гласови записи, „вътрешен ментор“); признание, че моделът „отговаря уверено, дори когато греши“; Единственото „Не се препоръчва“ за правен анализ в детайлния анализ; Първи отбелязват риска от Prompt Injection.	Известно противоречие в нивата на документа, което може да се дължи на експертната оценка и финално решение в Методолгическата карта „Преценка за социални помощи“ е единственото „С условия“ срещу шест забрани - класическият високорисков случай по Акта за ИИ, докато „Проверка за противоречия“ е „Забранено“: вероятно разменени отметки (бел. ред.). Условията в указанията са шаблонен повтарящ се текст.	Наблюдения, които звучат тествано („~85% надежден код“, поведение при чувствителни теми), но без яснота дали числата са от тест; промптове не са приложени.	В мотивираното становище моделът е наречен „Gemini 3.1 Pro“ - името от примерния текст на шаблона, при оценяван Gemini 2.5 Pro

ГРУПА "ПРОМЕТЕЙ"

OpenAI GPT

(GPT-5.4 / 5.2 / 4o)



SWOT Анализ

Наименование на модела	LLM: ChatGPT 5.5 (Instant/Thinking)
Разработчик / компания	OpenAI
Версия за оценка	LLM: ChatGPT 5.5 (Instant/Thinking)
Дата на попълване	14.05.2026
Членове на групата	Здравко Бечев Деница Цанкова Маргарита Стойчева Лора Любенова Орлин Василев Румен Раенков Георги Косев
Ръководител на групата	Георги Косев

A. Обща приложимост и функционален обхват

Критерий		
A1. Приложимост в публичния сектор		
Обработка на документи <i>доклади, решения, наредби</i>	4	<i>Обосновка / бележки:</i> Структурира (резюме) добре, но има риск от слаб анализ при по-сложни документи, засягащи косвено и други сфери. Актуалният размер на информацията, с която работи е до 1,5 млн. токени, и може да генерира много информация.
Правен анализ <i>тълкуване на нормативни актове</i>	2	<i>Обосновка / бележки:</i> Рискове при анализ на актуално, специфично законодателство. Необходима е задължителна верификация/връзка с основни официални бази данни/текстове. Риск от силна халюцинация.
Граждански услуги <i>чатбот, Q&A, помощ при заявления</i>	4	<i>Обосновка / бележки:</i> При интеграция с вътрешните системи има възможност за генериране на коректна и качествена информация, перфектни FAQ функционалности (потенциал за оценка 5).

Вътрешна комуникация <i>резюмета, протоколи, кореспонденция</i>	4	<i>Обосновка / бележки:</i> Локалните продукти позволяват изчистването на информация от чувствителни данни и дават добра възможност за бързо обобщаване. Възможно е да изпуснат ключови аспекти съобразно векторната структура на модела.
Анализ на данни <i>отчети, статистика, мониторинг</i>	3	<i>Обосновка / бележки:</i> Все още ниска интеграция с excel (подобрява се), който се използва основно в публичната администрация. Генерално разпознаване на тенденциите в данните, но трудно вкарва контекст и в зависимост от обема е възможно да изпусне ключови аспекти на общата картина.
Превод <i>многоезична комуникация, ЕС документи</i>	4	<i>Обосновка / бележки:</i> Понякога изпуска контекст и прави преводът неточен. При големи текстови обеми – съкращава.
Програмиране / автоматизация <i>вътрешни системи, скриптове</i>	4	<i>Обосновка / бележки:</i> Един от водещите модели при писане на код. Възможност за директно приложение в деловодните системи (API), но риск при автоматизираното разпределение към компетентните звена.
Обучение на служители <i>учебни материали, тестове</i>	2	<i>Обосновка / бележки:</i> Липса на дълбочина и контекст.
A2. Функции на платформата		
Текстово генериране <i>качество и прецизност</i>	4	<i>Обосновка / бележки:</i> Високо качество на генерирания текст, но е възможно да губи контекст, и да е лаконичен в отговорите си. Трябват множество итерации за достигане на окончателен качествен продукт.
Reasoning / разсъждение <i>логически вериги, многостъпкови задачи</i>	4	<i>Обосновка / бележки:</i> Документооборот – висок потенциал; високо качество на генерирания текст, но е възможно да губи контекст; претоварването с (разнородна) информация може да го накара да халюцинира.
Code writing <i>генериране и анализ на код</i>	4	<i>Обосновка / бележки:</i> Един от водещите модели при писане на код. Възможност за директно приложение в деловодните системи (API), но риск при автоматизираното разпределение към компетентните звена.
Анализ на изображения <i>vision, диаграми, скрийншоти</i>	3	<i>Обосновка / бележки:</i> Copyrights restrictions и чест риск от проблеми с визуализацията на български символи.
Аудио / глас	4	<i>Обосновка / бележки:</i>

говорим вход/изход		Бързо и точно писане, но ограничения при българската реч.
Видео разбиране анализ на видеосъдържание	3	<i>Обосновка / бележки:</i> Добро видеоразбиране. Предимство пред други модели, които нямат тази функционалност.
Дълги документи обработка на обемни файлове	3	<i>Обосновка / бележки:</i> Структурира добре, но риск от слаб анализ при по-сложни документи, засягащи косвено и други сфери. Актуалният размер на информацията, с която работи е 1,5 млн. токени, и може да генерира много информация.
Agentic режим автономни многостъпкови задачи	4	<i>Обосновка / бележки:</i> Възможност за свързване с имейла; генериране на специфични таблици със структурирани данни от excel, но все още ниска интеграция с excel, който се използва основно в публичната администрация. Генерално разпознаване на тенденциите в комуникацията/данните, но трудно вкарва контекст и в зависимост от обема е възможно да изпусне ключови аспекти на общата картина.
API / системна интеграция свързване с вътрешни системи	4	<i>Обосновка / бележки:</i> Възможност за свързване с имейла и специфични бази данни (документооборот, правни системи, СЕОС); генериране на специфични таблици със структурирани данни от excel, но все още ниска интеграция с excel, който се използва основно в публичната администрация. Генерално разпознаване на тенденциите в комуникацията/данните, но трудно вкарва контекст и в зависимост от обема е възможно да изпусне ключови аспекти на общата картина.
Plugin екосистема налични разширения и конектори	4	<i>Обосновка / бележки:</i> Възможност за свързване с имейла и специфични бази данни (документооборот, правни системи, strategy.bg); генериране на специфични таблици със структурирани данни от excel, но все още ниска интеграция с excel, който се използва основно в публичната администрация. Генерално разпознаване на тенденциите в комуникацията/данните, но трудно вкарва контекст и в зависимост от обема е възможно да изпусне ключови аспекти на общата картина.
Desktop приложение нативна desktop среда	5	<i>Обосновка / бележки:</i> Добра интеграция с desktop системи.
Mobile приложение iOS / Android достъп	5	<i>Обосновка / бележки:</i> Добра интеграция с мобилни системи.
Cowork / колаборация	3	<i>Обосновка / бележки:</i>

екипна работа в реално време		Потенциал за допълнителна оптимизация, но при рестрикции на prompt-овете и необходимост от по-голям финансов ресурс.
Търсене в интернет актуална информация от мрежата	4	<i>Обосновка / бележки:</i> Добро търсене на информация с включени източници, но необходимост от верификация относно актуалността на информацията; рестрикции по отношение на лична информация като снимки; риск от bias.
Fine-tuning настройка за специфична среда	4	<i>Обосновка / бележки:</i> API fine-tuning се поддържа.

B. SWOT Матрица

S STRENGTHS — Силни страни 1. Високо качество на текстовото генериране и reasoning 2. Потенциал за широка системно-техническа интеграция и автоматизация (възможност за директна API интеграция в деловодни и вътрешни системи на администрацията (документооборот, правни системи, CEOC, strategy.bg, електронна поща), допълнена от нативни desktop и mobile приложения с високо качество (оценки 5), което осигурява безпрепятствено внедряване в ежедневието работен процес на служителите) 3. Потенциал за качествени граждански услуги 4. Голям контекстен прозорец (1 млн. токена)	W WEAKNESSES — Слаби страни 1. Риск за сигурност на информацията (GDPR, класифицирана информация и т.н.) 2. Загуба на контекст при обемни и сложни документи 3. Висок риск от халюцинации особено при правен анализ и специфично национално законодателство 4. Ограничения при работа с кирилица (българска) 5. Ниска интеграция с Excel
O OPPORTUNITIES — Възможности 1. Fine-tuning за специфичните нужди на администрацията / интеграция с вътрешни системи и регистри 2. Оптимизацията на процесите 3. Подобряване на гражданското обслужване и комуникация 4. Оптимизация на разходите 5. Потенциал за систематично намаляване на оперативни грешки	T THREATS — Заплахи 1. Липса на човешка верификация 2. Защита на данни и съответствие с регулации 3. Bias и проблеми с актуалността на информация от мрежата (новини и информация от интернет) 4. Зависимост от един доставчик (vendor lock-in) (Guardrails – масово наблюдение на граждани и автономни смъртоносни оръжия)

	5. Риск от прекомерна автоматизация без human in the loop (MCP) 6. Загуба на информация при сложни и обемни документи 7. Халюцинации при правни и нормативни задачи
--	---

С. Разширен SWOT по специфични теми

Цена и лицензиране Наблюдения: Платформата предлага многослоен модел на плащане — индивидуален абонамент (Plus/Pro), екипни и enterprise лицензи, както и API ценообразуване по токени (вход/изход). За малки общини и второстепенни разпоредители разходите могат да бъдат значителни, особено при разширена употреба с дълъг контекст, agentic режим и колаборативни функции. API използването е сравнително предвидимо, но при големи обеми на обработка (документооборот, преписки) разходите могат бързо да ескалират. Липсва специализирано публично-секторно лицензиране, адаптирано към български условия. Препоръка за публичния сектор: Договаряне на централизирано рамково споразумение на национално ниво с обемни отстъпки за държавната администрация; диференциран достъп според нивото на потребителя; пилотни лицензи за общините преди мащабиране.	
Сигурност на данните Наблюдения: Обработката се осъществява предимно в облачна среда извън ЕС, което поражда въпроси за съответствие с GDPR при работа с лични и чувствителни данни. Enterprise/API планове предлагат опции за no-training на данни, zero data retention и регионални инстанции, но локални (on-premise) внедрявания не са налични директно за този клас модели. Съществува риск при споделяне на класифицирана или служебна информация през стандартни интерфейси. Препоръка за публичния сектор: разработване на протоколи за сигурен обмен на информация и вътрешни правила за допустима употреба на ниво администрация. Използване на новата локална система на OpenAI за анонимизиране на данните.	
Езикова поддръжка Наблюдения: Качеството на българския език е добро при стандартно текстово генериране, но при специализирана юридическа, административна и техническа терминология се наблюдават неточности, опростявания и стилистични несъответствия с високия административен регистър.	

Препоръка за публичния сектор: Подкрепа за развитие на български езикови модели в рамките на научната общност (INSAIT - BGGPT).

Vendor lock-in

Наблюдения: Прекомерната зависимост от един доставчик носи стратегически рискове — промяна в ценовата политика, ограничения на функционалности, геополитически фактори, прекратяване на услуги.

Препоръка за публичния сектор: Успоредно разработна на аналогичен заместител в лицето на BG GPT и Mistral

Сложност на внедряване

Наблюдения: Базовото ползване чрез уеб/desktop/mobile интерфейс е технически зряло и достъпно за служители без специализирана подготовка. Сериозните внедрявания (API интеграция с деловодни системи, СЕОС, регистри) обаче изискват значителен IT капацитет, който в много второстепенни разпоредители и общини е ограничен или липсва.

Препоръка за публичния сектор: Централизиран модел на внедряване чрез МИДТ. Мониторинг на внедряването от страна на МИДТ на всички административни нива и колаборация за обучения заедно с ИПА.

Регулаторно съответствие

Наблюдения: Употребата в публичния сектор попада в обхвата на EU AI Act — значителна част от приложенията (вземане на административни решения, граждански услуги, оценка на достъп до права) ще се квалифицират като високорискови, с произтичащите задължения за оценка на съответствието, документация, надзор и регистрация. NIS2/ЗКС налага изисквания за киберсигурност при критични информационни системи; GDPR остава приложим. Понастоящем доставчикът на платформата извършва оценка на съответствието като доставчик на GPAI модел, но отговорността за приложението остава в публичната администрация като deployer.

Препоръка за публичния сектор: Задължителна предварителна категоризация на всеки случай на употреба съгласно EU AI Act; създаване на регистър на AI системите в администрацията; ясна координация между Националния компетентен орган и отделните администрации; включване на изискването за съответствие във всички обществени поръчки; човешки надзор като задължителен елемент при високорискови приложения.

Етика и прозрачност

Наблюдения: Обяснимостта на модела остава ограничена — основните решения се вземат в т.нар. „черна кутия“, което затруднява одитируемостта. Платформата предоставя източници при търсене в интернет, но не и пълна проследимост на reasoning процеса. Отговорността при грешки в административен контекст е неясна — между доставчика на модела, разработчика на интеграцията и администрацията като deployer. Липсва стандартизирана методология за одит на AI-генерирани административни актове.

Препоръка за публичния сектор: разработване на вътрешни процедури за човешка проверка и валидация преди излизане на документ; водене на одитна следа за всяко AI взаимодействие в административни процеси.

Изкривявания (Bias)

Наблюдения: Моделите от този клас са обучени предимно върху англоезични източници, което създава риск от културно, езиково и нормативно изкривяване при работа с български контекст. Възможни са пристрастия при чувствителни обществени теми, малцинствени групи, регионални специфики и при историко-политически въпроси. При генериране на текстове, свързани с граждански услуги или административни решения, рискът от непропорционално третиране на различни групи изисква специално внимание.

Препоръка за публичния сектор: обучение на служителите за разпознаване на bias; забрана за изцяло автоматизирано вземане на решения, засягащи права на граждани; механизъм за обжалване.

Халюцинации / грешки

Наблюдения: Халюцинациите са системен риск, особено осезаем при правен анализ, цитиране на нормативни актове, имена на лица и институции, статистически данни и хронологии. Честотата им е сравнително ниска при стандартни задачи, но нараства значително при специализирана терминология, дълги контексти и работа с актуално национално законодателство. Detectability е ограничена — грешките често са стилистично убедителни и могат да преминат непроверени през неекспертен прочит. Рискът при вземане на административни решения е висок.

Препоръка за публичния сектор: Задължителна верификация на всеки факт, цитат и нормативна препратка от експертен служител преди финализиране; категоричен забрана за пряко използване на AI-генерирани правни анализи без професионален преглед; интеграция с официални бази данни (АПИС, ciela, EUR-Lex) чрез RAG механизми за повишаване на точността; обучение на служителите за разпознаване на типични модели на халюцинация.

Морал и ценности

Наблюдения: Платформата има вградени защитни механизми, които ограничават генерирането на вредно, дискриминационно или незаконно съдържание, и обикновено отказва задачи с очевиден етически проблем. При чувствителни обществено-

политически теми моделът тенденциозно избира неутрална, балансирана позиция, което в административен контекст е по-скоро предимство. Възможни са обаче несъответствия между вградените ценности на доставчика (формирани предимно в американски контекст) и специфичните български/европейски нормативни и културни рамки.

Препоръка за публичния сектор: Актуализация на етичните правила за работа на държавните служители; изготвяне на публични правила за работата на администрациите с изкуствен интелект; периодични оценки на правилата; както има лице по GDPR, така и да има лице по етичен AI.

D. Обобщена оценка на групата

Обобщена оценка по основни показатели		
Обща приложимост <i>средна от Раздел А</i>	4	<i>Обосновка / бележки:</i> Оценка: 4 Средната оценка от попълнените критерии в Раздел А се позиционира около 3,6–3,8, което се закръгля до 4. Платформата демонстрира силен потенциал в обработка на документи, текстово генериране, програмиране и системна интеграция, със значими възможности за приложение в граждански услуги, вътрешна комуникация, обработка на документи и автоматизация. Ограниченията са при правен анализ, excel, сложни документи и нуждата от верификация.
Готовност за внедряване <i>техническа и организационна</i>	2	<i>Обосновка / бележки:</i> Оценка: 2 Технологичната зрялост на платформата е висока — налични са desktop, mobile и API канали за внедряване. Организационната готовност на българската администрация обаче изостава — ограничен IT капацитет в значителна част от второстепенните разпоредители и общините, дефицит на специализирани кадри за AI интеграции, липса на единни вътрешни правила и стандарти за употреба. Реалистичното внедряване изисква предварителна централизирана подготовка, методическа подкрепа и инвестиция в капацитет.
Съответствие с регулации <i>AI Act, GDPR, NIS2</i>	3	<i>Обосновка / бележки:</i> Оценка: 3 Доставчикът на платформата изпълнява задълженията си като доставчик на GPAI модел съгласно EU AI Act, но значителна част от потенциалните публично-секторни приложения попадат в категорията на високорисковите AI системи, което поражда съществени задължения за администрацията като deployer. Необходимост от договорни гаранции и оценки.
Етичен риск	3	<i>Обосновка / бележки:</i>

<i>обратна скала: 5=нисък риск</i>		Оценка: 3 Платформата разполага с вградени защитни механизми срещу очевидно вредно съдържание, но запазва значими рискове, свързани с халюцинации при специализирана терминология, потенциални изкривявания спрямо български контекст и малцинствени групи, ограничена обяснимост на reasoning процеса и неясно разпределение на отговорността при грешки. Рискът е управляем при стриктен човешки контрол и ясни процедури, но в никакъв случай не е пренебрежим — оттук средната оценка.
Обща препоръка <i>1=не препоръчваме, 5=силно препоръчваме</i>	4	<i>Обосновка / бележки:</i> Оценка: 4 Препоръчваме внедряване на платформата в публичния сектор при ясно дефинирани условия и ограничения. Балансът между потенциалните ползи (повишаване на административния капацитет, оптимизация на документооборота, подобряване на гражданските услуги) и рисковете (халюцинации, регулаторно съответствие, vendor lock-in) клони към положително становище, при условие че се подсигури human in the loop.

Мотивирано становище на групата (мин. 5 изречения):

Платформата представлява един от водещите общоцелеви AI модели понастоящем и предлага реален потенциал за повишаване на ефективността, последователността и качеството на работа в българската публична администрация — особено в областите на текстовото генериране, обработката на документи, програмната автоматизация и системната интеграция чрез API. Демонстрираните възможности при многоезична работа, агентни сценарии и мултиплатформена достъпност я правят подходящ инструмент за широк кръг приложения, от вътрешна административна комуникация до граждански услуги. Същевременно идентифицирахме значими рискове, свързани с халюцинации при правен анализ, ограничена дълбочина при специализирана българска терминология, зависимост от външен доставчик извън ЕС и сложен регулаторен профил при високорискови приложения съгласно EU AI Act. Препоръчваме внедряване, но в контролирана, постепенна и стратегически премислена рамка — с категорична забрана за пряко използване в правни анализи и автоматизирано вземане на решения, засягащи права на граждани, без задължителен експертен човешки преглед. Платформата следва да се разглежда не като самостоятелен инструмент за вземане на административни решения, а като инструмент за подпомагане на експертната работа на служителите, който повишава техния капацитет, без да го заменя.

Предложени условия за ползване:

- 1. Институционална и организационна рамка** Централизирано рамково договаряне на лицензни условия на национално ниво (МИДТ) с гарантирани enterprise опции — европейско съхранение на данни, zero data retention, договорно изключване от обучение на модела. Изграждане на споделен компетентностен център за подпомагане на второстепенните разпоредители и общините. Задължително вътрешно категоризиране на всеки случай на употреба съгласно EU AI Act преди внедряване и водене на регистър на AI системите в администрацията.

- 2. Технически контроли и интеграция** Категорична забрана за въвеждане на класифицирана, чувствителна лична и служебна информация в стандартни (consumer) интерфейси. Интеграцията с вътрешни системи (документооборот, СЕОС, регистри) да се осъществява през абстрактен слой (API gateway), позволяващ multi-vendor стратегия и смяна на доставчика. Задължително водене на одитна следа за всяко AI взаимодействие в административни процеси. При работа с правни и нормативни въпроси — задължителна интеграция с официални бази данни (АПИС, EUR-Lex) чрез RAG механизми. Задължително изготвяне от всяка администрация на специфични спрямо нейните дейности вътрешни правила за работа с ИИ.
- 3. Човешки надзор и отговорност** Задължителна експертна верификация на всеки AI-генериран изход с правни последици, преди финализиране. Категорична забрана за изцяло автоматизирано вземане на решения, засягащи права на граждани. Систематично обучение на служителите за работа с AI инструменти и за разпознаване на типични модели на халюцинации и изкривявания.

Risk Canvas

Наименование на модела	LLM: ChatGPT 5.5 (Instant/Thinking)
Разработчик / компания	OpenAI
Версия за оценка	LLM: ChatGPT 5.5 (Instant/Thinking)
Дата на попълване	14.05.2026
Членове на групата	Здравко Бечев Деница Цанкова Маргарита Стойчева Лора Любенова Орлин Василев Румен Раенков Георги Косев
Ръководител на групата	Георги Косев

Легенда за скалите

Вероятност (Вер.) 1 = Изключително рядко 2 = Рядко 3 = Умерено вероятно 4 = Вероятно 5 = Почти сигурно	Въздействие (Вл.) 1 = Незначително 2 = Малко 3 = Умерено 4 = Значително 5 = Критично	Ниво на риска Н = Нисък ($\text{Вер.} \times \text{Вл.} \leq 4$) С = Среден (5–12) В = Висок (13–25)
--	--	--

Риск Канва — 9-полева матрица

1. Технически рискове ➤ Халюцинации и фактически грешки при правни и административни отговори. ➤ Загуба на контекст при сложни, обемни или междусекторни документи и преписки. ➤ Грешки при мултимодален вход —	2. Рискове за данните ➤ Обработка на лични данни на граждани в облачна среда без достатъчен контрол върху съхранението и достъпа. ➤ Риск от въвеждане или изтичане на служебна, непублична или класифицирана информация.	3. Правни / регулаторни ➤ Несъответствие с EU AI Act при използване на модела в средно-или високорискови публично-секторни приложения. ➤ Нарушение на GDPR при обработка на лични данни без правно основание, DPIA, минимизация
--	--	---

изображения, аудио, скрийншоти, таблици и документи с български символи. ➤ Внезапна промяна на поведението на модела при нова версия или промяна на условията/настройките от доставчика.	➤ Запазване на prompts, входни данни и изходи от доставчика, включително риск от използване за обучение или подобряване на услуги. ➤ Трансфер на данни извън ЕС/ЕИП без достатъчни гаранции по GDPR, SCC, TIA и контрол върху подизпълнителите.	или достатъчна анонимизация. ➤ Липса на обяснимост и одитна следа при AI подпомогнати административни решения. ➤ Неясна отговорност при грешни правни, административни или чувствителни изходи, използвани в официален процес.
4. Етични рискове ➤ Систематичен bias спрямо определени групи граждани, социални профили, региони или езикови особености. ➤ Дискриминация при автоматизирано обработване, класиране или приоритизиране на заявления. ➤ Непрозрачност за гражданите дали и как AI е използван при подготовка на отговор, оценка или административна услуга. ➤ Загуба на човешка автономия и прехвърляне на експертна отговорност към AI системата.	5. Оперативни рискове ➤ Недостатъчен технически капацитет за внедряване, поддръжка, мониторинг и сигурна интеграция. ➤ Интеграционни проблеми с legacy системи, деловодни системи, регистри, Excel и вътрешни бази данни. ➤ Недостатъчно обучение на служителите за правилна употреба, проверка на резултати и защита на данните. ➤ Липса на вътрешни правила за допустими, условно допустими и забранени употреби на AI.	6. Репутационни рискове ➤ Публичен скандал при видима AI грешка в официална комуникация, документ или отговор към гражданин. ➤ Загуба на доверие в институцията при непрозрачно използване на AI или липса на човешка отговорност. ➤ Негативно медийно отразяване при изтичане на лични, служебни или чувствителни данни. ➤ Обществено възприятие, че администрацията заменя експертна преценка с автоматизирани отговори.
7. Стратегически рискове ➤ Vendor lock-in и прекомерна зависимост от един външен технологичен доставчик. ➤ Ценови скокове при промяна на лицензионния или API модела и затруднено бюджетно планиране	8. Засегнати страни ➤ Граждани - риск от грешен, непълен, пристрастен или непрозрачно генериран отговор. ➤ Служители - риск от прекомерно доверяване, неправилна употреба и неясна лична/служебна отговорност.	9. Остатъчни рискове ➤ Дори при експертна верификация остава риск от пропусната фактическа, правна или контекстуална грешка. ➤ Дори при вътрешни правила остава риск служители да

➤ Зависимост от инфраструктура, правила и юрисдикция извън ЕС. ➤ Спиране, ограничаване или промяна на услугата при геополитически, търговски или регулаторни обстоятелства.	➤ Ръководство - отговорност за правила, контрол, отчетност, обучение и управление на инциденти. ➤ ИТ звена, ДЛЗД, юристи, одитори и външни доставчици - ключови за сигурност, съответствие, интеграция и мониторинг.	въведат неподходящи лични, служебни или чувствителни данни. ➤ Дори при договорни гаранции остава зависимост от доставчика, неговата инфраструктура, подизпълнители и бъдещи условия. ➤ Дори при прозрачност и обучение остава риск от прекомерно доверяване и рутинно използване без достатъчно критична проверка.
--	---	--

Рисков регистър

Риск / Събитие	Вероятност	Въздействие	Ниво	Мярка за митигация
Технически рискове				
Халюцинации и фактически грешки при отговори	3	5	ВИСОК	Задължителна експертна верификация
Недостъпност / downtime на платформата	2	3	СРЕДЕН	Резервен работен процес при ситуации без ИИ
Внезапна промяна на версия или поведение на модела	3	3	СРЕДЕН	Договорно фиксиране на конкретни версии на модела; тестова среда за валидация преди обновяване.
Ограничения на контекстния прозорец при обемни документи	2	2	НИСЪК	Предварително структуриране и сегментиране на обемните преписки, ясни вътрешни инструкции за максимален обем на вход.

Грешки при мултимодален вход (изображения, аудио)	4	2	СРЕДЕН	Верификация от компетентните звена, вкл. от специализираната администрация и звеното за връзки с обществеността и протокол (при генериране на публична информация).
Рискове за данните и поверителността				
Обработка на лични данни на граждани в облачна среда	4	5	ВИСОК	B1 – 4 B2 – 5 Риск – ВИСОК Категорична забрана за въвеждане на лични данни в стандартни (consumer) интерфейси; използване на локални инструменти или вътрешни процедури за анонимизация/псевдонимизация преди подаване на данни към AI среда; при чувствителни случаи - приоритетно локална, суверенна или EU-only инфраструктура; договорно подписан DPA със стандартни договорни клаузи (SCC).
Изтичане на класифицирана/служебна информация	3	5	ВИСОК	Категорична забрана за въвеждане на класифицирана информация съгласно ЗЗКИ; вътрешни правила за допустима употреба с дисциплинарни последици при нарушение; технически контроли (DLP) при корпоративна интеграция; обучение на служителите; одитна следа на всички взаимодействия.
Запазване на запитвания от доставчика за обучение	3	4	СРЕДЕН	Договорно изключване от обучение (opt-out) в enterprise/API договорите; ползване на режим zero data retention; периодична проверка на договорните условия и политиката за поверителност на

				доставчика; при установяване на използване на данни за обучение от доставчика – задължение за предоставяне на конкретика относно обхвата и целите на използваните данни.
Трансфер на данни извън ЕС (EU AI Act / Schrems II)	4	4	ВИСОК	Преференциране на enterprise опции с гарантирано европейско съхранение и обработка; стандартни договорни клаузи - SCC по Решение (ЕС) 2021/914 + Transfer Impact Assessment + допълнителни технически мерки; проверка на подизпълнители и data location и включване на разпоредби относно правото на интелектуална собственост на съответната администрация по отношение на вход и изход данните/информация, използвани и съответно – генерирани от модела, при евентуалното им използване от доставчика за разработването на последващи продукти и/или услуги; консултация с националните компетентни органи спрямо EU AI Act при високо и среднорискова обработка.
Липса на одитна следа при автоматизирани решения	4	4	ВИСОК	Задължително логване (запаметяване) на всички AI взаимодействия в административни процеси/регистри; интеграция с вътрешни системи за документооборот; съхранение на промпт, контекст и изход за минималния законово изискван срок; технически и организационни мерки за неподправимост на одитната следа.

Правни и регулаторни рискове				
Несъответствие с EU AI Act (Annex III — публичен сектор)	4	5	ВИСОК	Предварителна категоризация на всеки случай на употреба съгласно EU AI Act; оценка на съответствието при високо и среднорискови приложения; регистриране в публичния регистър на ЕС; задължителен човешки надзор от юридически компетентни лица; лице за отговорно използване на ИИ; координация с Националния компетентен орган(и) по EU AI Act; пълна техническа документация.
Нарушение на GDPR при обработка на лични данни	3	4	ВИСОК	Задължителна DPIA преди внедряване при систематична обработка; задължително анонимизиране на данните и задължителен надзор от GDPR служител.
Липса на обяснимост при автоматизирани административни решения	4	5	ВИСОК	Категорична забрана за изцяло автоматизирани решения, засягащи права на граждани (чл. 22 ОРЗД); задължителен експертен човешки преглед; одитна следа за всяка AI итерация; механизъм на подаване на сигнали/обжалване; лице за отговорно използване на ИИ.
Авторскоправни нарушения в генерирано съдържание	3	3	СРЕДЕН	Верификация от компетентните звена, вкл. от специализираната администрация и звеното за връзки с обществеността и протокол (при генериране на публична информация), и вътрешни правила за публикуване (отличителни маркери, напр. <i>Част от информацията е генерирана с ИИ</i>); лице за отговорно използване на ИИ.

Отговорност при грешни правни или медицински изходи	4	5	ВИСОК	Категорична забрана за пряко използване на AI-генерирани правни анализи без експертен преглед; забрана за медицински приложения извън строго регламентирани медицински изделия; механизъм за подаване на сигнали с ясно дефиниране на отговорността в писмени вътрешни правила; лице за отговорно използване на ИИ.
Етични и социални рискове				
Систематичен bias спрямо определени групи граждани	3	4	СРЕДЕН	Задължителни оценки на въздействието за равнопоставеност при високо и среднорискови приложения; разнообразни и представителни тестови набори; периодично оценяване на изходите за признаци на дискриминация; обучение на служителите; механизъм за подаване на сигнали; лице за отговорно използване на ИИ.
Дискриминация при автоматизирано обработване на заявления	3	5	ВИСОК	Забрана за самостоятелно автоматизирано класиране/отхвърляне; човешки контрол; критерии и механизъм за подаване на сигнали/обжалване; лице за отговорно използване на ИИ.
Непрозрачност — граждани не знаят, че AI взима решения	4	4	ВИСОК	Ясно уведомяване при AI подпомагане; запис в административната преписка; възможност за човешки преглед.
Дезинформация и манипулация чрез убедително съдържание	3	4	СРЕДЕН	Проверка на факти и източници; забрана за публикуване без редакторски контрол.

Загуба на критично мислене и зависимост при служителите	4	3	СРЕДЕН	Обучения, правила за верификация, ротационни проверки на AI-generated материали.
Оперативни рискове				
Липса на технически капацитет за внедряване и поддръжка	4	4	ВИСОК	Централизирана подкрепа – осигуряване на необходимото финансиране, пилотни проекти, минимални технически стандарти.
Vendor lock-in — прекомерна зависимост от доставчик	4	4	ВИСОК	Договорни клаузи за гарантиране степента на спазване на приложимото законодателство; Успоредно разработване на аналогичен заместител в лицето на BG GPT и Mistral.
Недостатъчно обучение на служителите	3	4	СРЕДЕН	Задължително обучение преди достъп; вътрешно сертифициране; кратки практически инструкции.
Интеграционни проблеми с наследени (legacy) системи	4	4	ВИСОК	Пилотна интеграция; sandbox среда; API тестове; човешко потвърждение.
Ценови скокове при промяна на лицензионния модел	3	5	ВИСОК	Рамково договаряне; бюджетен таван и допустим процент на индексация (разпределяне на тежестта между участниците по договарянето); мониторинг на изпълнението на договорните клаузи и потреблението.
Репутационни и стратегически рискове				
Публичен скандал при видима AI грешка (медийно отразяване)	3	5	ВИСОК	Комуникационен протокол - вътрешни правила за публикуване на съдържание, генерирано с ИИ (отличителни маркери, напр. <i>Част от информацията е генерирана с ИИ</i>) и верификация от компетентните звена, вкл. от специализираната

				администрация и звеното за връзки с обществеността и протокол (при генериране на публична информация); процедура за корекция.
Загуба на доверие на гражданите в публичните институции	4	5	ВИСОК	Прозрачност, яснота, човешка отговорност, механизъм за подаване на сигнали/обжалване; лице за отговорно използване на ИИ.
Политическа съпротива срещу използването на AI	3	3	СРЕДЕН	Пилотно внедряване; публично обяснение на принципите на използване, мерките за прозрачност, вкл. механизмите за сигнализиране на нередности; ежегодно докладване на резултати от извършената интеграция/внедряване на ИИ в работата на публичната администрация; възможност за осъществяване на парламентарен/обществен контрол.
Зависимост от американски/китайски технологичен доставчик	4	4	ВИСОК	Проверка на суверенни/локални алтернативи; договорни гаранции, гарантиращи спазването на приложимото законодателство.
Спиране на услугата при геополитически напрежения	2	4	СРЕДЕН	Резервен работен процес без ИИ; договорни клаузи за гарантиране защита на данните в случай на force major геополитическа ескалация/напрежение; Успоредно разработване на аналогичен заместител в лицето на BG GPT и Mistral.

Обобщена рискова оценка на групата

Топ 3 критични риска за модела:

1. Обработка, съхранение и трансфер на лични, служебни или чувствителни данни

Най-критичният риск е свързан с въвеждането, съхранението, обработката или трансфера на лични, служебни, непублични или чувствителни данни в облачна AI среда. Този риск включва нарушение на GDPR, трансфер извън ЕС/ЕИП, неясно участие на подизпълнители, липса на контрол върху data location и риск от използване на входните/изходните данни от доставчика.

2. Халюцинации, грешни правни изводи и липса на обяснимост при административни процеси

Моделът може да генерира убедително звучащи, но неверни фактически, правни или административни изводи. Рискът е особено висок, когато резултатите се използват при правен анализ, отговори към граждани, подготовка на официални документи или процеси, които могат да засегнат права, задължения или законни интереси.

3. Vendor lock-in и стратегическа зависимост от външен доставчик

Съществува риск от прекомерна зависимост от конкретен доставчик, неговата инфраструктура, ценови модел, лицензионни условия, версии на модела, подизпълнители и юрисдикция. Това може да затрудни дългосрочното бюджетно планиране, преносимостта към друг модел, контрола върху данните и изграждането на устойчива национална политика за AI в публичната администрация.

Обща рискова оценка на модела за публичния сектор:

- ☐ Нисък риск — може да се внедри с минимални ограничения
- ☐ Среден риск — изисква допълнителни контроли и процедури
- ☒ Висок риск — изисква задълбочена оценка и специфична регулация
- ☐ Неприемлив риск — не се препоръчва за публичния сектор

Мотивирано становище:

Групата оценява общия рисков профил на ChatGPT 5.5 като висок за публичния сектор, когато моделът се използва в контекст на лични данни, служебна информация, правен анализ, административни решения, граждански услуги или интеграция с вътрешни системи. Основните критични рискове са свързани, от една страна, с обработката, съхранението и трансфера на лични, служебни или чувствителни данни в облачна среда, включително трансфер извън ЕС/ЕИП, неясно участие на подизпълнители и ограничен контрол върху местоположението и последващото използване на данните. От друга страна, съществен риск представляват халюцинациите, грешните правни, фактически или административни изводи и ограничената обяснимост на резултатите, особено когато те могат да бъдат използвани при подготовка на официални документи, отговори към граждани или процеси, засягащи права и законни интереси. Като трети критичен риск се откроява vendor lock-in, който включва не само зависимост от конкретен доставчик, но и зависимост от неговата инфраструктура, ценови и лицензионен модел, промени във версиите и функционалностите на модела, условията за достъп до API, подизпълнителите, юрисдикцията на обработка и възможността за преносимост на натрупаните процеси, настройки, интеграции и данни към алтернативно решение.

Високият риск не означава, че моделът не следва да се използва, а че внедряването му трябва да бъде поетапно, контролирано и ограничено до ясно дефинирани случаи на употреба. Моделът може да се използва като помощен инструмент за резюмиране, подготовка на чернови, вътрешна комуникация, обучение, техническа подкрепа и предварителна обработка на документи, при условие че крайният резултат се проверява от компетентен служител. Не следва да се допуска самостоятелно автоматизирано вземане на решения, обработка на класифицирана информация, въвеждане на непсевдонимизирани лични данни в незащитена среда или използване на AI-генерирани правни изводи без експертна верификация. Човешкият контрол следва да се разглежда като хоризонтална задължителна мярка за всички средно- и високорискови употреби, а не като самостоятелно достатъчна гаранция.

За ограничаване на идентифицираните рискове са необходими вътрешни правила за допустима, условно допустима и забранена употреба; задължително обучение на служителите; договорни клаузи за защита на данните, забрана или ограничаване на използването им за обучение на модела, ясно определяне на data location, подизпълнителите, правата върху входните и изходните данни, нивата на обслужване, отговорността и условията за прекратяване. Следва да се въведат строги изисквания и периодични проверки за съответствие на работата на модела с приложимото законодателство, включително GDPR, EU AI Act, правилата за киберсигурност и вътрешните административни процедури. Необходими са още одитна следа, механизми за подаване на сигнали, корекция и оттегляне на грешни AI-генерирани резултати, както и процедури за преглед при засягане на граждани. Паралелно с използването на външни модели следва да се работи по внедряване и развитие на потенциално по-подходящи от гледна точка на езикова, регулаторна и технологична независимост решения, включително BGGPT и европейски модели като Mistral, като се отчита, че към момента те все още не са достатъчно развити или внедрени за всички необходими административни случаи на употреба.

Карта за приложимост

Наименование на модела	LLM: ChatGPT 5.5 (Instant/Thinking)
Разработчик / компания	OpenAI
Версия за оценка	LLM: ChatGPT 5.5 (Instant/Thinking)
Дата на попълване	14.05.2026
Членове на групата	Здравко Бечев Деница Цанкова Маргарита Стойчева Лора Любенова Орлин Василев Румен Раенков Георги Косев
Ръководител на групата	Георги Косев

Част 1. Обзорна матрица на приложимостта

Административна функция	Обработка на документи	Правен анализ	Граждански услуги	Анализ на данни	Програмиране	Обучение
Обработка на кореспонденция	Прил.: ○1 ○2 ○3 ○4 ○5 Риск: □Н □С □В 4 С Добър за резюме и класификация, но при експертна верификация	Прил.: ○1 ○2 ○3 ○4 ○5 Риск: □Н □С □В 2 В Риск от грешна правни тълкувания	Прил.: ○1 ○2 ○3 ○4 ○5 Риск: □Н □С □В 3 С Подходящ за проект на отговор, но при задължителна експертна верификация	Прил.: ○1 ○2 ○3 ○4 ○5 Риск: □Н □С □В 2 С Ограничен о за метаданни и обобщение; необходимост задължително анонимизация на данните	Прил.: ○1 ○2 ○3 ○4 ○5 Риск: □Н □С □В 1 Н Няма съществуване на приложимост	Прил.: ○1 ○2 ○3 ○4 ○5 Риск: □Н □С □В 3 С Може да дава примерни казуси и да изготвя анализи на получената кореспонденция

Изготвяне на доклади и анализи	4 С Силен при структура, чернови и резюмета	2 В Правните изводи изискват експертна верификация, вкл. от юрист	3 С Помага за ясен публичен език	3 С Разпознава тенденции, но анализът следва да е задължително съпроводен от компетентна експертна верификация	2 С Може да подпомогне автоматизация на анализа	3 С Полезен за учебни примери
Тълкуване на нормативни актове	3 В Може да резюмира текст, но с риск от халюцинации и загуба на контекст	2 В Висок риск от халюцинации и неактуалност	2 В Опасно при казуси на граждани съобразно риска от халюцинации и загубата на контекст	1 С Непряка приложимост в зависимост от конкретния казус (обхват, засегнати страни)	1 Н Неприложимо	2 С Само за учебно обяснения (use case) с проверка/верификация
Изготвяне на правни становища	3 В Може да даде структура/насоки, но не и цялостно съдържание	2 В Само помощно, задължителна верификация от компетентен експерт и юрист	1 В Не следва да дава правно становище към граждани	1 С Непряка приложимост в зависимост от конкретния казус (обхват, засегнати страни)	1 Н Неприложимо	2 С Само като учебен пример (use case) с проверка/верификация
Отговори на граждански и запитвания	4 С Подходящ за проект на отговор, но при задължителна експертна верификация	2 В Правни казуси – изискват задължителна верификация	4 В Висока приложимост, но риск за защитата на правата и за	2 С Може да ползва справочни данни при налична необходим	1 Н Неприложимо	3 С Полезен при обучение по комуникация с граждани

		ация от компетентен експерт и юрист	доверието в институциите	ата интеграция със съответните системи		
Чатбот за електронни услуги	3 С Може да обяснява процедури и документи	2 В Да не тълкува нормативни казуси, а да дава единствено насоки за тяхното разработване (източници, примерна структура, адресати и др.)	4 В Силен use case, но изисква прозрачност и бърза включване от/обратна връзка към служител	2 С Може да отчита често задавани въпроси и да предоставя обобщава по подходящ начин административно верифицирани отговори	3 В API интеграцията носи технически и data риск	3 С Полезен за обучение на граждани/служители
Статистически анализ	2 С Може да описва резултати, но остава неприложим в по-голямата си част, освен при нормативни индикатори като законово определени срокове, прагове, проценти, критерии за допустимост, лимити, минимални/максимални стойности или показатели за	1 В Неприложимо освен при нормативни индикатори	1 С Само обяснение на публични данни	3 С Помага при тенденции, но при задължителна проверка на изчисленията от компетентните органи/звена	3 С Може да генерира код за анализ	2 С Полезен за учебни примери

	изпълнение, заложен в нормативен акт, програма или методика.					
Обработка на отчети	4 С Добър за резюме и структура, при верификация на данни/контекст	2 В Правни изводи – задължителна проверка	2 С Може да подготви публично резюме	3 С Полезен но числата се верифицират	2 С Помощна автоматизация	3 С Полезен за учебни примери
Автоматизация на работни процеси	3 С Може да предложи шаблони и стъпки	1 В Да не автоматизира правни решения	2 В Риск при предоставянето на равнопоставен достъп до услуги за граждани	3 С Подходящ за мониторинг с контрол от страна на заинтересованите звена	4 В Силен, но изисква sandbox, code review и (регистър на) log-ове	2 С Полезен за учебен по обучение за процеси
Разработка на вътрешни инструменти	2 С Подпомага документация	1 В Правни функции не се автоматизират самостоятелно	2 С Само помощни интерфейси	3 С Справки и таблици с проверка	4 В Висока приложимост; необходимост от управление на риска при продукционна интеграция	3 С Полезен за технически наръчници с примери от практиката
Учебни материали и тестове	3 С Добър за чернови и структура	3 В Правното съдържание се проверява	3 С Подходящ за разработване на различни варианти/сценарии за обслужване	3 С Предоставяне на информация по отношение на базови примери от практиката	2 С Може да генерира упражнения в дадена област/сфера	3 С Полезен, но риск от плиткост и неточности.

Наръчници и процедури	4 С Силен при структуриране и редакция	4 В Юридическата верификация е задължителна	3 С Помага за ясен административен език	3 С Ограничен о за анализ на данни, свързани с конкретните процедури	3 С Помощно за технически описания	4 С Добър за вътрешни обучения, с експертен преглед
Превод и локализация	4 С Добър превод, но контекстът се проверява	3 В Висок риск при правна терминология	3 С Подходящ за разбираем граждански език	1 Н Неприложимо	1 Н Неприложимо	3 С Полезен за преводи с учебна цел
Протоколи и резюмета	4 С Висока приложимост за резюмиране	2 В Да не прави правни изводи от протокола	3 С Само публично резюме след преглед от компетентните експерти, вкл. тези, отговарящи за връзките с обществеността	2 С Ограничен анализ на теми/честоти	1 Н Неприложимо	3 С Добър за обучителни примери

Част 2. Детайлен анализ по случаи на употреба

2.1. Обработка на административни документи

Способността на модела да чете, структурира, резюмира и изготвя административни документи — решения, наредби, докладни записки, протоколи.

Оценка за избрания модел:

1. Може ли моделът да обработва документи на български език с административна терминология?
2. Поддържа ли обработка на PDF/Word файлове с дълъг обем?
3. Запазва ли форматирането и структурата при генериране на изходен документ?
4. Изисква ли специализирано prompt инженерство за административен контекст?

Отговори / наблюдения:

1. Моделът може да обработва документи на български език с административна терминология, включително доклади, писма, протоколи, докладни записки и проекти на вътрешни документи. Най-силен е при резюмиране, структуриране, извличане на основни точки и подготовка на чернови.
2. Поддържа обработка на PDF/Word файлове с дълъг обем, но при сложни, междусекторни или правно значими документи съществува риск от загуба на контекст, пропускане на важни детайли или прекомерно обобщаване.
3. Може да подпомага запазването на структурата на документа, но не следва да се разчита на него за окончателно форматиране, съответствие с вътрешни образци или официален административен стил без човешка редакция.
4. Изисква специализирано задание/prompt за административен контекст, включително посочване на целта на документа, адресата, правния/институционалния контекст, ограниченията и изискването за запазване на факти, числа и препратки.

Обобщено наблюдение

Моделът е приложим при обработка на административни документи като помощен инструмент, което е съгласувано със SWOT оценката за висока приложимост при обработка на документи и текстово генериране. Рискът остава среден, защото при официални документи е необходима експертна верификация, особено при наличие на правни, финансови или други чувствителни, контекстуални елементи.

Препоръка:

☐ Препоръчва се без ограничения

☒ Препоръчва се с ограничения (опишете):

Да се използва за резюмиране, структуриране, сравнение на версии, чернови и извличане на ключови точки. Да не се използва като самостоятелен източник на окончателни правни, финансови или фактически изводи. При документи с лични, служебни или чувствителни данни да се прилагат анонимизация/псевдонимизация, вътрешни правила и експертна проверка и при координация с лицето за отговорно използване на ИИ и/или други компетентни звена.

☐ Не се препоръчва (обосновка):

2.2. Правен анализ и тълкуване

Способността да анализира нормативни актове, да открива противоречия, да прилага правни норми към конкретни случаи.

Оценка за избора на модел:

1. Колко точен е при тълкуване на български законодателни текстове?
2. Може ли да следи промени в нормативната база (обновени закони)?
3. Как се справя с многозначност и правна неяснота?
4. Маркира ли ясно когато излиза извън областта на своята компетентност?

Отговори / наблюдения:

1. Точността при тълкуване на български законодателни текстове е ограничена. Моделът може да подпомогне ориентация, резюме и структуриране на правен въпрос, но съществува висок риск от неточни тълкувания, несъществуващи препратки или непълно отчитане на специален режим.

2. Не може надеждно самостоятелно да следи промени в нормативната база, освен ако не е свързан с актуални и официални източници. Дори при интернет търсене е необходима проверка в нормативни бази, Държавен вестник, EUR-Lex, АПИС, Сиела или друг официален/лицензиран източник.
3. При многозначност и правна неяснота моделът често предлага логично звучаща интерпретация, но не може да замени юридическата/експертната оценка. Именно при правна неяснота рискът от халюцинация и неправилен извод е най-висок.
4. Не винаги маркира достатъчно ясно, че излиза извън областта на сигурната си компетентност. Отговорите могат да бъдат убедително формулирани, дори когато съдържат грешка или непълна правна логика.

Обобщено наблюдение

Правният анализ е една от най-рисковите употреби. Това съответства на SWOT оценката и на Risk Canvas, където халюцинациите, грешните правни изходи, липсата на обяснимост и отговорността при грешни правни резултати са оценени като висок риск.

Препоръка:

☐ Препоръчва се без ограничения

☒ Препоръчва се с ограничения (опишете):

Да се използва само за помощни дейности – първоначално проучване на нормативната база с цел последващо коректно идентифициране на приложимата правна уредба, резюме на (предложения за) нормативни текстове, структуриране на въпроси, подготовка на списък с релевантни теми или чернова за експертна работа. Да не се използва за окончателни правни становища, мотиви към административни актове, тълкуване към граждани или решения с правен ефект без задължителна проверка от юрист/компетентен експерт.

☐ Не се препоръчва (обосновка):

2.3. Взаимодействие с граждани (чатбот / асистент)

Способността да комуникира с граждани — отговаря на въпроси, помага при подаване на заявления, обяснява административни процедури.

Оценка за избрания модел:

1. Разбира ли естествен разговорен български (включително диалекти и грешки)?
2. Дава ли ясни и разбираеми отговори без административен жаргон?
3. Как се справя с чувствителни теми (социални помощи, здраве, правосъдие)?
4. Може ли да се интегрира с портали за електронни услуги?

Отговори / наблюдения:

1. Моделът разбира естествен разговорен български език, включително непълни въпроси, неформален изказ и частично грешно формулирани запитвания. При диалекти, двусмислени въпроси или емоционално натоварени казуси обаче е възможно да направи неправилно предположение.
2. Може да дава ясни и разбираеми отговори без прекомерен административен жаргон. Това го прави подходящ за FAQ, разясняване на процедури, насочване

към относима/релевантна информация, вкл. формуляри и превеждане на административния език в по-широкостъпна форма.

3. При чувствителни теми - социални помощи, здраве, правосъдие, санкции, права или допустимост - рискът е висок. В тези случаи моделът не следва да дава окончателни отговори, а само обща информация и насочване към институция, звено, служител или официален източник.
4. Може да се интегрира с портали за електронни услуги, но само при ясни протоколи за интеграция със съответните системи, сигурност на данните, вкл. при използването им на вход и изход, одитна следа, прозрачност към гражданите относно ролята на ИИ в предоставянето на съответната електронна услуга, механизъм за подаване на сигнали/обжалване и определено лице за отговорно използване на ИИ.

Обобщено наблюдение

Гражданските услуги са с висока приложимост, но и с висок риск при реални казуси. SWOT отчита висок потенциал за FAQ и чатбот функционалности, а допълнително Risk Canvas извежда рискове за лични данни, непрозрачност, дискриминация, грешни отговори и загуба на доверие.

Препоръка:

☐ Препоръчва се без ограничения

☒ Препоръчва се с ограничения (опишете):

Да се използва само за информационни и навигационни цели. Да не взема решения, да не оценява допустимост, да не класифицира граждани и да не предоставя окончателни правни отговори по индивидуални казуси. Задължителни са ясно уведомяване за участие на ИИ в процеса, проверка и предоставяне на източниците на генерираната информация, възможност за връзка с компетентен служител/звено, а при системни грешки - ескалация към лицето за отговорно използване на ИИ и/или експерт с компетенции в областта на разисквания казус), механизъм за подаване на сигнали/обжалване и одитна следа при всички административни взаимодействия с граждани.

☐ Не се препоръчва (обосновка):

2.4. Анализ на данни и отчетност

Способността да обработва структурирани данни, да генерира отчети и да идентифицира тенденции.

Оценка за избрания модел:

1. Може ли да обработва Excel/CSV данни и да генерира анализи?
2. Поддържа ли визуализация на данни (графики, диаграми)?
3. Колко надежден е при числови изчисления и статистика?
4. Може ли да се свърже с вътрешни бази данни чрез API?

Отговори / наблюдения:

1. Моделът все още е с ниска интеграция с Excel, който се използва основно в публичната администрация, но при подходящо структуриране на количествена информация има потенциал да генерира аналитични текстове, разпознавайки основни тенденции без дълбочина по отношение на причините за тяхното

възникване. Подходящ е за предварителни изводи, описание на тенденции и структуриране на отчет.

2. Може да подпомага визуализация на данни чрез предложения за графики, диаграми и обяснения, но визуализациите и изходните данни, вкл. текст и числа, следва да се проверяват от компетентни звена.
3. Не е достатъчно надежден като самостоятелен инструмент за числови изчисления и статистика. Съществува риск да сгреши при проценти, сравнения, зависимости, агрегации или изводи, особено при големи таблици, липсващи стойности или специфичен административен контекст. Препоръчително е използването на ясно структурирана информация от един и същи тип и в количество, което позволява последваща верификация от експерт.
4. Може да се свърже с вътрешни бази данни чрез API, но това повишава риска и изисква контрол на достъпа, sandbox тестове, логване, анонимизация/псевдонимизация и защита на данните, одобрение от компетентните звена, разработващи/използващи съответните бази данни, както и от лицето за отговорно използване на ИИ.

Обобщено наблюдение

Приложимостта при анализ на данни е средна. SWOT отчита, че моделът разпознава тенденции, но трудно вкарва контекст и има ниска интеграция с Excel. Risk Canvas допълнително поставя висок риск при интеграции, лични данни, липса на одитна следа и технически капацитет.

Препоръка:

☐ Препоръчва се без ограничения

☒ Препоръчва се с ограничения (опишете):

Да се използва за предварително обобщаване, текстово описание на тенденции, предложения за структуриране на отчети и формулиране на хипотези. Всички числа, проценти, финансови изводи, статистически зависимости и нормативни индикатори да се проверяват от компетентните звена и/или чрез надежден изчислителен инструмент. При API интеграция - задължително sandbox, контрол на достъпа и логване, одобрение от ИТ звеното, ДЛЗД/правно звено и лицето за отговорно използване на ИИ според нивото на риска.

☐ Не се препоръчва (обосновка):

2.5. Вътрешно обучение на служители

Способността да генерира учебни материали, тестове, наръчници и да поддържа интерактивно обучение.

Оценка за избрания модел:

1. Може ли да генерира тематични курсове и тестове за служители?
2. Адаптира ли съдържанието към различни нива на компетентност?
3. Поддържа ли формати за дистанционно обучение (SCORM и др.)?
4. Как се справя с чувствителни теми — антикорупция, етика, дискриминация?

Отговори / наблюдения:

1. Моделът може да генерира тематични курсове, примерни тестове, казуси, инструкции и кратки наръчници, но най-вече като чернова или начална структура.
2. Може да адаптира съдържанието към различни нива на компетентност, например въвеждащо, експертно или практически ориентирано ниво. Въпреки това, адаптацията не гарантира пълнота, точност и съответствие с вътрешните правила - необходима е експертна верификация.
3. Не следва да се приема, че самостоятелно поддържа специализирани формати за дистанционно обучение като SCORM. Може да подпомогне подготовката на съдържание, сценарии, тестови въпроси и структури, които след това да бъдат въведени в LMS/обучителна платформа от компетентен екип.
4. При чувствителни теми - антикорупция, етика, дискриминация, GDPR, EU AI Act - е необходим експертен преглед по подобие на работата с цел правен анализ. Рискът е съдържанието да бъде повърхностно, неточно или недостатъчно съобразено с приложимото законодателство.

Обобщено наблюдение

Вътрешното обучение е приложимо, но не като самостоятелно разработване на финални обучителни програми, а по-скоро с цел онагледяване на различни примери от практиката, формулиране на примерни ситуации и други обучителни помощни материали в комбинация с експертизата/опита на компетентните служители. SWOT оценява обучението по-ниско поради липса на дълбочина и контекст.

Препоръка:

☐ Препоръчва се без ограничения

☒ Препоръчва се с ограничения (опишете):

Да се използва за чернови, структури, примерни казуси, тестови въпроси и помощни материали. Финалното съдържание да се валидира от компетентно звено, ИПА/обучителен екип. Да не се използва самостоятелно за обучения по правни, етични, антикорупционни и регулаторни въпроси.

☐ Не се препоръчва (обосновка):

2.6. Програмиране и автоматизация

Способността да пише код, да автоматизира процеси и да поддържа вътрешни системи.

Оценка за избрания модел:

1. Поддържа ли езиците и технологиите, използвани в административните системи?
2. Може ли да чете и документира съществуващ (legasy) код?
3. Колко надеждни са генерираните скриптове без допълнителна проверка?
4. Поддържа ли agentic режим за автономно изпълнение на задачи?

Отговори / наблюдения:

1. Моделът поддържа основните езици и технологии, използвани при административни системи, и може да подпомага писане на код, API заявки, скриптове, документация и технически описания.
2. Може да чете, обяснява и документира съществуващ legasy код, но при стари, специфични или слабо документираните системи е необходим преглед от ИТ

специалист/изпълнител, защото моделът може да не разбере архитектурния и институционалния контекст.

3. Генерираните скриптове не са достатъчно надеждни без допълнителна проверка. Те трябва да преминават през code review, sandbox тестове, security review и одобрение преди внедряване, особено ако засягат реални данни, регистри, деловодни системи или автоматично насочване на преписки.
4. Поддържа agentic режим за многостъпкови задачи, но автономното изпълнение в публичната администрация следва да бъде ограничено. Не трябва да се допуска самостоятелно изпълнение на действия с административен ефект без човешко потвърждение и одитна следа. Потенциал при използване за обзор при регулярни задачи от организационен тип – напр. обзор на предстоящи срещи, на текущи задачи съгласно деловодната система и др.

Обобщено наблюдение

Програмирането и автоматизацията са сред силните приложения на модела, което е съгласувано със SWOT анализа, code writing, API интеграция и agentic режим. Рискът обаче се повишава при продукционна интеграция, достъп до вътрешни системи, обработка на данни и автоматизирано разпределение към компетентни звена.

Препоръка:

☐ Препоръчва се без ограничения

☒ Препоръчва се с ограничения (опишете):

Да се използва за подпомагане на административните звена чрез прототипи, документация, шаблони, тестови скриптове и нискорискова автоматизация. Да не се внедрява директно в продукционна среда без code review, sandbox тестове, security review, контрол на достъпа, логване и одобрение от компетентните звена, както и от лицето за отговорно използване на ИИ. Автономни действия с административен ефект да бъдат забранени без човешко потвърждение.

☐ Не се препоръчва (обосновка):

Част 3. Оценка на интеграционния потенциал

Интеграционен аспект	Описание	Поддръжка	Бележки
API свързване	REST API, GraphQL, webhooks — интеграция с вътрешни системи	Да	Наличен висок потенциал за интеграция с деловодни системи, регистри, СЕОС, правни системи, електронна поща и вътрешни бази данни. Да се допуска само след sandbox тестове, контрол на достъпа, логване, security review, code review и одобрение от ИТ звеното и компетентните административни звена.

Single Sign-On	<i>Съвместимост с Active Directory / ADFS / Keycloak</i>	Частично	Възможно при enterprise/организационна среда, но следва да се провери съвместимостта с използваните в администрацията Active Directory и други софтуерни решения. Да се предпочита институционална идентификация пред индивидуални нерегламентирани акаунти.
On-premise deployment	<i>Възможност за локален хостинг без облачна зависимост</i>	Не	За самия ChatGPT 5.5 не следва да се приема, че е налично пълно локално внедряване без облачна зависимост. За чувствителни случаи следва да се оценяват локални, суверенни или open-weight алтернативи, но те не са идентични с оценявания модел по функционален обхват и зрялост.
Данни в ЕС	<i>Сървъри на европейска територия — GDPR съответствие</i>	Частично	Да се изисква гарантирано съхранение и обработка в ЕС/ЕИП, когато е възможно. При трансфер извън ЕС/ЕИП да се прилагат DPA, SCC по Решение (ЕС) 2021/914, Transfer Impact Assessment, контрол на подизпълнителите и ясна data location политика при ясно формулирани договорни клаузи в тази посока.
Одитна следа	<i>Логирание на всички взаимодействия за отчетност</i>	Частично	Необходима е интеграция с вътрешни системи за логване на потребител, дата, цел на използване, входни данни, prompt, изход и последваща човешка проверка. Сроковете за съхранение следва да се определят

			според правилата за документооборот, архивиране, информационна сигурност и приложимите нормативни изисквания.
Ролеви достъп	<i>RBAC — различни нива на достъп за различни служители</i>	Частично	С оглед изричната забрана за използване на чувствителни неанонимизирани данни и класифицирана информация в модела на този етап не се идентифицира необходимост от въвеждане на специфични ролеви нива на ползване, освен в случаите на мониторинг и одит на работата с модела.
Многоезичност	<i>Поддръжка на български и официални ЕС езици</i>	Да	Моделът поддържа български и официални езици на ЕС, но при юридическа, административна и техническа терминология е необходима експертна проверка. Препоръчително е използване на терминологични речници, вътрешни шаблони и проверка на преводите при официални документи.
Мобилен достъп	<i>Работа от мобилни устройства (field work)</i>	Да	Технически е наличен, но служебната употреба от мобилни устройства следва да се допуска само при ясни правила, MDM/служебни профили, контрол на достъпа и забрана за работа с чувствителни, служебни или класифицирани данни през лични устройства.

Офлайн режим	<i>Функционалност без постоянна интернет връзка</i>	Не	Стандартната услуга изисква интернет/облачна връзка. За критични процеси следва да има резервен работен процес без ИИ и план за непрекъсваемост при downtime, прекъсване на услугата или геополитически/търговски ограничения.
SLA и поддръжка	<i>Гарантирано ниво на обслужване за публичния сектор</i>	Частично	Нужни са договорни гаранции за ниво на обслужване, срокове за реакция, уведомяване при инциденти, промени във версии и функционалности, защита на данните, прекратяване, преносимост и exit strategy.

Обобщена карта за приложимост

Топ 3 приложения с висок потенциал:

1. Обработка на административни документи

Резюмиране, структуриране, сравнение на версии, подготовка на чернови, извличане на ключови точки и подпомагане на вътрешната кореспонденция. Това е най-подходящият ниско-/среднорисков стартов сценарий при задължителна експертна проверка.

2. Граждански информационни услуги и вътрешна комуникация

FAQ, проект на отговори, разясняване на процедури, насочване към релевантни източници и подпомагане на комуникацията с граждани. Приложението е с висок потенциал, но само за информационни и навигационни цели, без решаване на индивидуални казуси.

3. Програмиране, техническа подкрепа и нискорискова автоматизация

Подпомагане на ИТ звена, документация, API прототипи, тестови скриптове, code review помощ и автоматизация на вътрешни нискорискови процеси. Продукционното внедряване изисква sandbox, security review, code review, логване и човешко потвърждение.

Функции, при които моделът НЕ е подходящ:

1. Обработка на класифицирана, чувствителна или непсевдонимизирана лична информация в незащитена среда

Не следва да се въвеждат класифицирани, служебни, чувствителни или лични данни в стандартни/consumer интерфейси или в среда без договорни, технически и организационни гаранции.

2. Самостоятелен правен анализ и окончателни правни становища

Моделът не следва да се използва за окончателно тълкуване на нормативни актове, мотиви към административни актове, правни становища или отговори към граждани с правен ефект без проверка от юрист/компетентен експерт.

3. Автоматизирано вземане на административни решения

Не е подходящ за самостоятелно класиране, одобрение, отхвърляне, приоритизиране или оценка на заявления, проекти, права, помощи, санкции или достъп до услуги.

Препоръчани условия за пилотно внедряване:

Картата за приложимост е съгласувана със SWOT анализа и Risk Canvas. Високата приложимост на модела при документи, текстово генериране, гражданска комуникация, програмиране и API интеграции не означава нисък риск. Рискът се повишава при лични данни, правен ефект, граждански казуси, автоматизация, интеграции с вътрешни системи и зависимост от доставчик. Поради това нито един сценарий не се препоръчва без ограничения; моделът следва да се използва само като помощен инструмент при човешки контрол, договорни гаранции, одитна следа и ясни вътрешни правила.

Пилотното внедряване следва да започне с ниско- и среднорискови случаи на употреба: резюмиране и структуриране на документи, подготовка на чернови, вътрешни справки, учебни помощни материали, техническа документация и тестова автоматизация без реален административен ефект. Преди внедряване следва да се приемат вътрешни правила за допустима, условно допустима и забранена употреба на AI, включително правила за защита на данните, човешка верификация, проверка на източници, одитна следа, механизъм за подаване на сигнали и процедура за корекция или оттегляне на грешни AI-генерирани резултати.

Следва да се определи AI координатор/лице за отговорно използване на ИИ, което да координира прилагането на правилата съвместно с ДЛЗД, ИТ звеното, правното звено, звената по информационна сигурност и ръководството на администрацията. Това лице/звено не замества експертната отговорност на служителите, а координира допустимите случаи на употреба, обученията, сигналите за грешки, проверките за съответствие и периодичния преглед на използването на модела.

Интеграции с вътрешни системи следва да се извършват само поетапно, първо в sandbox среда, без реални чувствителни данни и след оценка на риска. За всички средно- и високорискови употреби са необходими договорни гаранции, контрол върху data location и подизпълнителите, DPA/SCC/TIA при трансфер извън ЕС/ЕИП, SLA, логване, ролеви достъп, code/security review и план за изход от зависимост от доставчика. Паралелно следва да се оценява възможността за внедряване или развитие на по-подходящи за публичния сектор решения от гледна точка на езикова, регулаторна и технологична независимост, включително BGGPT и европейски модели като Mistral, при отчитане на настоящата им по-ниска зрялост за част от административните приложения.

Методологични указания

Наименование на модела	LLM: ChatGPT 5.5 (Instant/Thinking)
Разработчик / компания	OpenAI
Версия за оценка	LLM: ChatGPT 5.5 (Instant/Thinking)
Дата на попълване	14.05.2026
Членове на групата	Здравко Бечев Деница Цанкова Маргарита Стойчева Лора Любенова Орлин Василев Румен Раенков Георги Косев
Ръководител на групата	Георги Косев

Предговор и обхват

Наименование на организацията	-
Административно ниво	-
AI модел(и), обхванати от документа	LLM: ChatGPT 5.5 (Instant/Thinking)
Версия на документа	Версия 1.0
Дата на приемане	26.05.2026
Следваща планирана ревизия	-
Одобен от	Членовете на групата
Отговорен за прилагането	-

Чл. 1. Предмет и обхват

Настоящите указания уреждат използването на големия езиков модел GPT 5.5 (GPT-5.5 Instant, GPT-5.5 Thinking, GPT-5.5 Pro) на OpenAI за подпомагане на експертната работа в администрацията. Документът обхваща приложения като резюмиране и структуриране на документи, подготовка на чернови, превод, вътрешна комуникация, граждански информационни услуги, анализ на неструктурирани данни, програмиране и нискорискова автоматизация. Указанията се прилагат за всички структурни звена и служители с достъп до модела (през уеб/desktop/mobile интерфейс или през API интеграция). Моделът се използва единствено като помощен инструмент при задължителен човешки контрол и не замества експертната отговорност на служителите.

Чл. 2. Правна и регулаторна рамка

Регламент (ЕС) 2024/1689 (Акт за изкуствения интелект, EU AI Act); Регламент (ЕС) 2016/679 (Общ регламент относно защитата на данните, GDPR); Закон за защита на личните данни; Закон за електронното управление (ЗЕУ); Закон за защита на класифицираната информация (ЗЗКИ); Директива (ЕС) 2022/2555 (NIS2) и Закон за киберсигурност; както и приложимите вътрешни правила, наредби и инструкции на администрацията.

Раздел I. Принципи за етично и прозрачно ползване

Основни принципи	
1 Законност	Конкретни мерки в определената организация: Всеки случай на употреба на ИИ преминава предварителна категоризация съгласно EU AI Act и оценка за съответствие с приложимото българско и европейско законодателство от правното звено. Поддържа се вътрешен регистър на AI системите и на допустимите случаи на употреба. Забранена е употреба извън изрично разрешените приложения..
2 Прозрачност	Конкретни мерки в определената организация: При взаимодействие с чатбот или автоматизиран асистент в началото ясно се указва, че гражданинът общува с ИИ. Документи, изготвени с подкрепата на ИИ, могат да съдържат отличителен маркер. Гражданите и служителите се уведомяват предварително за въвеждането на ИИ в съответния процес.
3 Отчетност	Конкретни мерки в определената организация: Въведен е инструктаж за употреба на ИИ на работното място; служителите носят лична отговорност за проверката и финализирането на всеки изход. ИИ не замества човешката отговорност. Поддържа се автоматизирана одитна следа (промт, контекст, изход, автор, дата и час), която обвързва всяко подпомогнато решение с конкретен служител.
4 Справедливост	Конкретни мерки в определената организация: Изходните резултати се проверяват периодично за езикови, демографски, социални и регионални изкривявания (bias), отчитайки повишения риск при обучение предимно върху англоезични източници. При задачи, засягащи граждани, в заданието изрично се указва моделът да не отчита признаци като пол, възраст, етнос, религия или местоживееене. Забранено е изцяло автоматизирано класиране или приоритизиране на заявления; предвиден е механизъм за подаване на сигнали и обжалване.
5 Поверителност	Конкретни мерки в определената организация: Забранено е въвеждането на класифицирана, служебна, чувствителна или непсевдонимизирана лична информация в стандартни (consumer) интерфейси. За автоматично анонимизиране на данни локално, преди да бъдат изпратени към OpenAI, се използва OpenAI Privacy Filter - локално работещ модел, създаден да открива и скрива лични данни (PII) и/или други одобрени локални инструменти или вътрешни процедури за анонимизация/псевдонимизация. Обработката на лични данни се извършва само при наличие на правно основание и при спазване на GDPR, с предпочитание към гарантирано съхранение в ЕС/ЕИП и договорно изключване от обучение на модела.
6 Надеждност	Конкретни мерки в определената организация: Прилага се правилото за задължителен човешки преглед (human-in-the-loop). Нито един изход на модела не се използва за официални цели без проверка на фактите, правните основания, числата и нормативните препратки от служителя-автор. При задачи, изискващи фактологична или правна точност, изходът задължително се съпоставя с официални и актуални източници (Държавен вестник, EUR-Lex, АПИС, Сиела или вътрешни регистри).
7 Пропорционалност	Конкретни мерки в определената организация: Нивото на човешки контрол съответства на риска на задачата. Прилага се тристепенна скала: (1) нисък риск (превод, резюмиране) - бърз преглед за запазен смисъл; (2) среден риск (писма, анализи) - детайлен преглед и повторна заявка за самопроверка; (3) висок риск (проекти за решения, правни въпроси) - структуриран „Reasoning“ анализ за вътрешни противоречия, последван от пълна верификация от компетентен експерт/юрист.

8 Обяснимост	Конкретни мерки в определената организация: Системните промптове, инструкциите и използваните версии на модела се документират централизирано от ИТ звеното. При искане за достъп до информация или жалба организацията предоставя писмено разяснение кои части от подготвителния процес са подпомогнати от ИИ, какви входни данни са подадени и по каква логика е изведен резултатът, удостоверявайки, че не е налице изцяло автоматизирано вземане на решение.

Раздел II. Разрешени и забранени употреби

Категория задача	Статус	Условия / Ограничения / Процедура
Чл. 3. Административни документи		
Изготвяне на проекти за решения	☐ Разрешено ☒ С условия ☐ Забранено	Изготвяне на проекти за решения: За автоматично анонимизиране на данни локално, преди да бъдат изпратени към OpenAI, се използва OpenAI Privacy Filter - локално работещ модел, създаден да открива и скрива лични данни (PII) и/или други одобрени локални инструменти или вътрешни процедури за анонимизация/псевдонимизация. Задължителен структуриран „Reasoning“ анализ на модела за проверка на вътрешни противоречия. Служителят, използвал ИИ, проверява всички факти, числа, правни основания и обстоятелства и валидира резултата (вкл. с КЕП при препращане за съгласуване). Приложимо само за неклафицирана информация. Силна приложимост при структуриране и чернови (SWOT - обработка на документи: 4), но при правно значими актове остава висок риск от халюцинации и загуба на контекст.
Резюмиране на входяща поща	☐ Разрешено ☒ С условия ☐ Забранено	Резюмиране на входяща поща: За автоматично анонимизиране на данни локално, преди да бъдат изпратени към OpenAI, се използва OpenAI Privacy Filter и/или други одобрени локални инструменти или вътрешни процедури за анонимизация/псевдонимизация. Приложимо само за неклафицирана информация. Подходящо за резюме и класификация при експертна верификация на ключовите аспекти, които моделът може да изпусне.
Превод на документи	☐ Разрешено ☒ С условия ☐ Забранено	Превод на документи: За автоматично анонимизиране на данни локално, преди да бъдат изпратени към OpenAI, се използва OpenAI Privacy Filter. създаден да открива и скрива лични данни (PII) и/или други одобрени локални инструменти или вътрешни процедури за анонимизация/псевдонимизация. Приложимо само за неклафицирана информация. Добро качество, но при специализирана юридическа и административна терминология контекстът се проверява; при големи обеми моделът съкращава.
Генериране на официални писма	☐ Разрешено ☒ С условия ☐ Забранено	Генериране на официални писма: За автоматично анонимизиране на данни локално, преди да бъдат изпратени към OpenAI, се използва OpenAI Privacy Filter и/или други одобрени локални инструменти или вътрешни процедури за анонимизация/псевдонимизация. Задължителен структуриран „Reasoning“ анализ на модела за проверка на вътрешни противоречия. Служителят, използвал ИИ, проверява всички факти, числа, правни основания и обстоятелства и валидира резултата (вкл. с КЕП при препращане за съгласуване). Приложимо само за неклафицирана информация.
Чл. 4. Правни и нормативни въпроси		
Анализ на нормативни текстове	☐ Разрешено ☒ С условия ☐ Забранено	Анализ на нормативни текстове: Задължителен структуриран „Reasoning“ анализ на модела за проверка на вътрешни противоречия. Само като помощно средство за първоначално проучване, резюме и структуриране. Висок риск от халюцинации и неактуалност - задължителна проверка в официални бази (Държавен вестник, EUR-Lex, АПИС, Сиела) и от компетентен юрист. Не се използва за окончателни изводи.

Проверка за вътрешни противоречия	☐ Разрешено ☒ С условия ☐ Забранено	Проверка за вътрешни противоречия: Задължителен структуриран „Reasoning“ анализ на модела за проверка на вътрешни противоречия. Приложимо само за неklasифицирана информация. Изходът е помощен и подлежи на експертен преглед.
Изготвяне на правни становища	☐ Разрешено ☒ С условия ☐ Забранено	Изготвяне на правни становища: Задължителен структуриран „Reasoning“ анализ на модела за проверка на вътрешни противоречия. Служителят, използвал ИИ, проверява всички факти, числа, правни основания и обстоятелства и валидира резултата (вкл. с КЕП при препращане за съгласуване). Приложимо само за неklasифицирана информация. Само помощно (структура, насоки). Окончателното становище се подписва от правоспособен юрист, който носи пълната отговорност; ИИ не се цитира като източник на правен авторитет. Една от най-рисковите употреби.
Тълкуване на разпоредби към граждани	☐ Разрешено ☒ С условия ☐ Забранено	Тълкуване на разпоредби към граждани: За автоматично анонимизиране на данни локално, преди да бъдат изпратени към OpenAI, се използва OpenAI Privacy Filter и/или други одобрени локални инструменти или вътрешни процедури за анонимизация/псевдонимизация. Задължителен структуриран „Reasoning“ анализ на модела за проверка на вътрешни противоречия. Служителят, използвал ИИ, проверява всички факти, числа, правни основания и обстоятелства и валидира резултата (вкл. с КЕП при препращане за съгласуване). Приложимо само за неklasифицирана информация. Само обща информация и насочване към компетентно звено/официален източник; без окончателни правни отговори по индивидуални казуси.
Чл. 5. Вземане на решения с ефект върху граждани		
Преценка за социални помощи	☐ Разрешено ☐ С условия ☒ Забранено	Преценка за социални помощи: забранено. Не се допуска самостоятелно автоматизирано оценяване на допустимост или вземане на решения, засягащи права на граждани. Допустимо е само човешко решение без заместване на експертната преценка от ИИ.
Оценка на проекти и заявления	☐ Разрешено ☒ С условия ☐ Забранено	Оценка на проекти и заявления: За автоматично анонимизиране на данни локално, преди да бъдат изпратени към OpenAI, се използва OpenAI Privacy Filter и/или други одобрени локални инструменти или вътрешни процедури за анонимизация/псевдонимизация. Служителят, използвал ИИ, проверява всички факти, числа, правни основания и обстоятелства и валидира резултата (вкл. с КЕП при препращане за съгласуване). Задължителен структуриран „Reasoning“ анализ на модела за проверка на вътрешни противоречия. Забранено е самостоятелно класиране, приоритизиране или отхвърляне. Само помощни справки при задължителен човешки контрол. Не се използва за класифицирана или чувствителна информация.
Наказателни и принудителни мерки	☐ Разрешено ☐ С условия ☒ Забранено	Наказателни и принудителни мерки: забранено. Употребата в наказателни и принудителни процедури не е разрешена.
Автоматизирани решения без човешки контрол	☐ Разрешено ☐ С условия ☒ Забранено	Автоматизирани решения без човешки контрол: забранено. Не се допускат изцяло автоматизирани решения с правен ефект или ефект върху граждани. Човешкият контрол е задължителна хоризонтална мярка.
Чл. 6. Комуникация с граждани		
Чатбот за информационни въпроси	☐ Разрешено ☒ С условия ☐ Забранено	Чатбот за информационни въпроси: За автоматично анонимизиране на данни локално, преди да бъдат изпратени към OpenAI, се използва OpenAI Privacy Filter и/или други одобрени локални инструменти или вътрешни процедури за анонимизация/псевдонимизация. В началото на сесията ясно се указва, че гражданинът общува с автоматизиран асистент, с възможност за връзка с компетентен служител и за верификация на отговора. Само за информационни и навигационни цели; без решаване на индивидуални казуси. Висока приложимост, но висок риск при реални казуси (защита на права и доверие).
Проекти за отговори на жалби	☐ Разрешено ☒ С условия ☐ Забранено	Проекти за отговори на жалби: За автоматично анонимизиране на данни локално, преди да бъдат изпратени към OpenAI, се използва OpenAI Privacy Filter и/или други одобрени локални инструменти или вътрешни процедури за анонимизация/псевдонимизация. Задължителен

		структуриран „Reasoning“ анализ на модела за проверка на вътрешни противоречия. Служителят, използвал ИИ, проверява всички факти, числа, правни основания и обстоятелства и валидира резултата (вкл. с КЕП при препращане за съгласуване).
Публикации в социални медии	☐ Разрешено ☒ С условия ☐ Забранено	Публикации в социални медии: Служителят, използвал ИИ, проверява всички факти, числа, правни основания и обстоятелства и валидира резултата (вкл. с КЕП при препращане за съгласуване). Проверяват се генерираните текст и изображения (вкл. за авторскоправни ограничения и проблеми с кирилица). Добавя се отличителен маркер за участие на ИИ.
Кризисна комуникация	☐ Разрешено ☒ С условия ☐ Забранено	Кризисна комуникация: За автоматично анонимизиране на данни локално, преди да бъдат изпратени към OpenAI, се използва OpenAI Privacy Filter и/или други одобрени локални инструменти или вътрешни процедури за анонимизация/псевдонимизация. Задължителен структуриран „Reasoning“ анализ на модела за проверка на вътрешни противоречия. Служителят, използвал ИИ, проверява всички факти, числа, правни основания и обстоятелства и валидира резултата (вкл. с КЕП при препращане за съгласуване). Задължителен редакторски контрол и проверка на факти и източници преди публикуване; ясно обозначаване на съдържание, подпомогнато от ИИ. Повишено внимание към риска от дезинформация.
Чл. 7. Вътрешно-административни функции		
Обучение на служители	☐ Разрешено ☒ С условия ☐ Забранено	Обучение на служители: Само за чернови, структури, примерни казуси, тестови въпроси и помощни материали; финалното съдържание се валидира от компетентно звено/обучителен екип (ИПА). SWOT оценява обучението по-ниско поради липса на дълбочина и контекст. Не се използва самостоятелно за правни, етични и антикорупционни теми; без класифицирана информация.
Анализ на данни и отчети	☐ Разрешено ☒ С условия ☐ Забранено	Анализ на данни и отчети: За автоматично анонимизиране на данни локално, преди да бъдат изпратени към OpenAI, се използва OpenAI Privacy Filter и/или други одобрени локални инструменти или вътрешни процедури за анонимизация/псевдонимизация. Служителят, използвал ИИ, проверява всички факти, числа, правни основания и обстоятелства и валидира резултата (вкл. с КЕП при препращане за съгласуване). Само за предварително обобщаване и описание на тенденции; всички числа, проценти и статистически зависимости се проверяват от компетентно звено или надежден изчислителен инструмент (ниска интеграция с Excel). При API интеграция - sandbox, контрол на достъпа и логване. Не се използва за класифицирана/чувствителна информация и при подбор на персонал.
Автоматизация на повтарящи се задачи	☐ Разрешено ☒ С условия ☐ Забранено	Автоматизация на повтарящи се задачи: За автоматично анонимизиране на данни локално, преди да бъдат изпратени към OpenAI, се използва OpenAI Privacy Filter и/или други одобрени локални инструменти или вътрешни процедури за анонимизация/псевдонимизация. Служителят, използвал ИИ, проверява всички факти, числа, правни основания и обстоятелства и валидира резултата (вкл. с КЕП при препращане за съгласуване). Само нискорискова автоматизация без реален административен ефект; продукционна интеграция изисква sandbox, code review, security review, логване и одобрение от ИТ звеното. Автономни действия с административен ефект без човешко потвърждение са забранени. Без класифицирана информация и без управление на човешки ресурси.
Обработка на класифицирана информация	☐ Разрешено ☐ С условия ☒ Забранено	Обработка на класифицирана информация: забранено. Моделът е приложим само за некласифицирана информация и не може да се внедри изцяло локално - дори enterprise услугите изпращат данни за обработка към доставчика. Въвеждането на класифицирана или служебна тайна през този интерфейс се третира като нерегламентиран достъп и пробив в сигурността (ЗЗКИ).

Раздел III. Процедури за одобрение и верификация

Чл. 8. Задължителна верификация

Задължителна човешка верификация се изисква за всеки изход, предназначен за официален административен акт, кореспонденция, отговор към гражданин, правен или фактологичен анализ, числови/статистически данни и нормативни препратки. Текстът се преглежда от служител с подходяща квалификация преди подписване, който удостоверява проверката (вкл. с КЕП) и я отразява в дневника на AI употребата. Преди всяко запитване информацията преминава през задължителен етап на автоматизирано заличаване на чувствителни данни: използва се OpenAI Privacy Filter и/или други одобрени локални инструменти или вътрешни процедури за анонимизация/псевдонимизация - локално работещ модел, който (автоматично) разпознава и заличава лични данни, идентификатори и чувствителна информация, като подава към външната система само анонимизиран текст, необходим за изпълнение на задачата. За високорискови задачи се прилага тристепенната скала на контрол по Принцип 7.

Чл. 9. Процедура за одобрение на нови употреби

(1) Преди използване на модела за нов вид задача или в звено, необхванато от указанията, ръководителят на звеното подава писмена заявка до лицето за отговорно използване на ИИ (AI Officer) / ИТ звеното. (2) Заявката съдържа описание на процеса, очакваните ползи, примерен набор от данни и първоначална оценка на риска съгласно Шаблон 2 (Risk Canvas) и категоризация по EU AI Act. (3) ИТ звеното извършва оценка за сигурност и съвместимост (вкл. GDPR), а правното звено проверява съответствието с EU AI Act и националното законодателство. (4) Новата употреба се разрешава с писмена заповед на ръководителя на организацията, отразява се в регистъра на AI системите и указанията се актуализират по реда на чл. 13.

Чл. 10. Задължително разкриване към граждани

- (1) Гражданинът се уведомява за участие на ИИ във всички случаи, в които ИИ е подпомогнал обработката на негово искане или комуникацията с него: при взаимодействие с чатбот/автоматизиран асистент (в началото на сесията - „Вие разговаряте с автоматизиран асистент“); при отговори, становища или документи, изготвени с подкрепата на ИИ (чрез отличителен маркер). Уведомяването се отразява в административната преписка, осигурява се възможност за човешки преглед и за връзка с компетентен служител, както и механизъм за подаване на сигнали и обжалване.
- (2) При установяване на грешен, непълен, подвеждащ или неправомерно използван AI-генериран резултат се задейства процедура за корекция: временно спиране на използването на резултата, експертна проверка, уведомяване на засегнатите лица при необходимост, корекция или оттегляне на документа/отговора, запис в регистъра на инцидентите и докладване до AI Officer, ДЛЗД и ръководството според характера на риска.

Раздел IV. Мониторинг, одит и отчетност

Минимални изисквания за одитна следа (отбележете приложимите):

- ☒ Всяко взаимодействие с AI системата се записва с дата, час и потребител
- ☒ Пазят се логове на AI изходите, използвани в официални документи
- ☒ При автоматизирани решения се съхранява AI входа и изхода
- ☒ Провежда се периодичен преглед на качеството на AI изходите (мин. 1х/тримесечие)
- ☒ Инцидентите, свързани с AI грешки, се докладват на определено звено
- ☒ Годишен отчет за употребата на AI се изготвя и публикува

Чл. 11. KPI за мониторинг на AI употребата

Организацията следи тримесечно следните измерими показатели: (1) Процент открити фактологични/правни грешки и халюцинации при верификация спрямо общия брой генерирани изходи (целева стойност под 2%); (2) Дял на изходите, преминали задължителна човешка верификация преди финализиране (целева стойност 100% за средно- и високорисковите задачи); (3) Брой инциденти с поверителността - въвеждане на реални лични или чувствителни данни в модела (целева стойност 0); (4) Дял на текста, който служителят е трябвало да преработи ръчно (индикатор за полезност); (5) Спестено време при изготвяне на проект спрямо изцяло ръчна изработка; (6) Брой подадени сигнали/обжалвания, свързани с AI-подпомогнати решения, и средно време за тяхното разглеждане.

Раздел V. Отговорности и управление

Роля	Отговорност	Конкретно лице / звено
Ръководител на организацията	<i>Стратегическо одобрение и крайна отговорност за AI политиката</i> Одобрява методологичните указания, издава заповеди за внедряване и за нови употреби; носи крайната административна отговорност за използването на ИИ в организацията.	Ръководител на администрацията (министър / председател / изпълнителен директор / кмет / главен секретар)
AI отговорник (AI Officer)	<i>Координация, обучение, мониторинг, докладване на инциденти</i> Координира допустимите случаи на употреба, поддържа регистъра на AI системите, организира обученията, следи KPI и разглежда сигнали за грешки, изкривявания и етични рискове; не замества експертната отговорност на служителите.	Лице за отговорно използване на ИИ (AI Officer) - определен със заповед експерт/звено
Служители, ползващи AI	<i>Спазване на указанията, верификация на изходите, докладване на проблеми</i> Използват модела само за разрешените употреби, въвеждат единствено анонимизирани данни, извършват задължителна човешка верификация на всеки изход и докладват проблеми и инциденти.	Всички служители с достъп до модела
ИТ отдел / системен администратор	<i>Технически достъп, интеграция, сигурност, логове</i> Отговаря за сигурността на достъпа и API ключовете, изолираната среда, локалния слой за анонимизация, логването на промптове и изходи и недопускането на нерегламентиран трафик на данни.	Звено „Информационни технологии“ / системен администратор
Правно звено	<i>Преглед на нови употреби, съответствие с нормативната рамка</i> Извършва предварителна оценка за съответствие с националното законодателство и EU AI Act при всяка заявка за нова употреба и при категоризацията на случаите на употреба.	Правно звено / юрисконсулт, съвместно с ДЛЗД
ДЛЗД (длъжностно лице по защита на данни)	<i>GDPR съответствие при обработка на лични данни</i> Контролира спазването на правилата за анонимизация/псевдонимизация и GDPR, изготвя/съгласува DPIA и следи трансфера на данни и договорните гаранции (DPA, SCC, TIA).	Длъжностно лице по защита на данните (ДЛЗД)
Вътрешен одит	<i>Периодичен независим преглед на AI употребата</i> Извършва независими периодични проверки на логовете и спазването на забраните и тристепенната скала на контрол; проверява вярното отчитане на KPI; изготвя поне веднъж годишно доклад до ръководителя за съответствието с указанията и EU AI Act.	Звено „Вътрешен одит“ / вътрешен одитор

Раздел VI. Обучение и изграждане на капацитет

Чл. 12. Задължително обучение

Задължително обучение преминават всички служители, за които се предвижда използване на ИИ в работните процеси - при постъпване на работа или при внедряване на модела, в минимален обем от 8 часа, с допълнителен специализиран модул за административно-правните и финансовите звена. Служителите преминават годишен актуализиращ курс (мин. 4 часа) и извънреден курс при промяна на модела или нормативната рамка. Обучението включва практически казуси и завършва с тест (мин. оценка

8 от 10); успешното преминаване се удостоверява със сертификат, който се прилага към личното дело. При неуспех служителят повтаря теста, а при повторен неуспех - цялото обучение.

Модул на обучение	Целева група	Форма и обем	Периодичност
Основи на AI и рискове	Всички служители	Първоначално обучение с практическа насоченост (4 ч.)	На 12 месеца
Работа с конкретния модел	Всички служители с достъп	Практически семинар (4 ч.)	На 12 месеца
Етика и нормативна рамка	Всички служители; доп. модул за правни/финансови звена	Семинар (4 ч.) + доп. модул (4 ч.)	На 12 месеца
Разпознаване на AI грешки	Всички служители	Практически семинар (4 ч.)	На 12 месеца
Защита на личните данни	Всички служители	Първоначално обучение (4 ч.)	На 12 месеца

Раздел VII. Актуализация и валидност

Чл. 13. Условия за задължителна ревизия

Указанията се преразглеждат задължително при: излизане на нова версия на модела или съществена промяна в поведението/условието му; промяна в Регламент (ЕС) 2024/1689 (EU AI Act), Регламент (ЕС) 2016/679 (GDPR) или друга приложима нормативна рамка; регистриран инцидент, свързан със сигурността на данните или с грешка на ИИ с административен ефект; както и по правило най-малко веднъж на 12 месеца, ако в този период не е извършвана актуализация.

За ограничаване на vendor lock-in администрацията следва да предвижда договорни клаузи за преносимост на данни, изход от договора, достъп до логове, запазване на правата върху входните и изходните данни, предварително уведомяване при промяна на версия, API, цени или подизпълнители, както и план за миграция към алтернативно решение. Паралелно следва да се оценяват по-подходящи от гледна точка на езикова, регулаторна и технологична независимост решения, включително BGGPT и европейски модели като Mistral.

Чл. 14. История на промените

Версия 1.0 — одобрения на документа на 28.05.2026 г. от работната група.

Версия	Дата	Описание на промяната	Одобрил
Версия 1.0	28.05.2026	Създаване на документа (първоначална версия)	Членовете на работната група

ГРУПА "ЛОКИ"

Meta Llama

(4 Maverick /Scout/Behemoth 5)



SWOT Анализ

Оценка на фундаментален AI модел за публичната администрация

v1.0 — Работен шаблон

Инструкция за попълване: Попълнете всички полета за избрания от вашата група AI модел. Празните редове и скалите са предназначени за ръчно попълване. Не изтривайте инструкциите в курсив.

Наименование на модела	Llama 4
Разработчик / компания	META
Версия за оценка	Версия 4
Дата на попълване	14.05.2026
Членове на групата	Ани Христова, Антония Желева, Дейвид Александров, Димитър Ходулов, Красимир Червенков, Любчо Милушев
Ръководител на групата	Ани Христова

A. Обща приложимост и функционален обхват

Оценете доколко моделът покрива реални нужди в публичния сектор.

Инструкция за попълване: За всеки критерий: поставете кръгче на избраната оценка (1=много слабо, 5=отлично) и запишете кратка обосновка.

Критерий		
A1. Приложимост в публичния сектор		
Обработка на документи <i>доклади, решения, наредби</i>	Оценка 1–5 ○1 ○2 ○3 ☒ 4 ○5	Моделът е приложим за резюмиране, структуриране, първоначално редактиране и подготовка на проекти на административни документи, докладни записки, вътрешни становища, протоколи и указания. При вътрешна инфраструктура рискът от изтичане на данни е по-нисък в сравнение с публична облачна услуга, но остава необходим контрол върху входните данни, логовете и достъпа. Не следва да се използва самостоятелно за окончателни официални документи, особено когато има правни последици, индивидуални административни актове или чувствителна информация.
Правен анализ <i>тълкуване на нормативни актове</i>	Оценка 1–5 ○1 ○2 ☒ 3 ○4 ○5	Моделът може да подпомага първоначално търсене на аргументи, сравнение на текстове, откриване на възможни противоречия и формулиране на работни бележки по нормативни актове. Въпреки това правният анализ в публичната администрация изисква актуалност, точност и институционална отговорност, които не могат да бъдат оставени на AI модел. Изходите трябва да се приемат само като помощен материал и задължително да се проверяват от компетентно правно звено.
Граждански услуги <i>чатбот, Q&A, помощ при заявления</i>	Оценка 1–5 ○1 ○2 ☒ 3 ○4 ○5	Моделът може да бъде полезен като вътрешен помощник или като чатбот за обща информация относно административни процедури, срокове, необходими документи и насочване към правилни услуги. При работа с граждани рискът е по-висок, защото грешен или неясен отговор може да доведе до пропуснат срок, неправилно подадено заявление или загуба на доверие. Допустима е употреба само при интеграция с проверена база знания,

		ясно обозначение, че отговорът е информационен, и възможност за пренасочване към служител.
Вътрешна комуникация <i>резюмета, протоколи, кореспонденция</i>	Оценка 1–5 ○1 ○2 ○3  ○5	Моделът е подходящ за изготвяне на чернови на вътрешни съобщения, резюмета от срещи, протоколи, работни имейли, справки и кратки обобщения. Това е една от по-безопасните области на приложение, когато не се въвеждат лични данни, класифицирана информация или чувствителни служебни обстоятелства. Необходимо е служителят да запази авторски и професионален контрол върху крайния текст.
Анализ на данни <i>отчети, статистика, мониторинг</i>	Оценка 1–5 ○1 ○2  ○4 ○5	Моделът може да подпомага описание на тенденции, формулиране на аналитични изводи, генериране на примерни заявки, скриптове или обяснения към отчети. Не е подходящ да бъде единствен източник за числови изчисления, статистически заключения или управленски решения, защото може да допусне грешки в логиката, формулите или интерпретацията. Прилага се само с проверка на числата в първичната система или в официален аналитичен инструмент.
Превод <i>многоезична комуникация, ЕС документи</i>	Оценка 1–5 ○1 ○2  ○4 ○5	Моделът може да се използва за работен превод на административни, технически и общи текстове, включително за подготовка на проектни преводи от и към английски език. При превод на правни, нормативни и официални документи обаче е нужна човешка редакция, защото терминологията може да бъде неточна или да не съответства на установената българска и европейска правна лексика. Подходящ е за първа версия, но не и за окончателен официален превод.
Програмиране / автоматизация <i>вътрешни системи, скриптове</i>	Оценка 1–5 ○1 ○2 ○3  ○5	Моделът е приложим за подпомагане на разработчици и системни администратори при генериране на скриптове, примерен код, документация, тестови сценарии и анализ на грешки. Това може да ускори вътрешни ИТ процеси, особено при добре зададени ограничения и среда за тестване. Не трябва да се допуска директно внедряване на генериран код в продукционни системи без code review, тестване, проверка за сигурност и одобрение по вътрешна процедура.
Обучение на служители <i>учебни материали, тестове</i>	Оценка 1–5 ○1 ○2 ○3  ○5	Моделът е подходящ за създаване на учебни материали, примерни казуси, тестове, въпроси за самопроверка и обяснения за различни нива на компетентност. Особено полезен може да бъде при вътрешни обучения по информационна сигурност, работа с документи, защита на данните и административни процедури. Съдържанието трябва да се валидира от отговорно звено, за да не се разпространяват неточни правила или остарели нормативни препратки.
A2. Функции на платформата		
Текстово генериране <i>качество и прецизност</i>	Оценка 1–5 ○1 ○2 ○3  ○5	Моделът има добра приложимост за генериране на структурирани текстове, резюмета, проекти на документи, инструкции и вътрешни комуникации. При правилни шаблони и предварително зададен стил може да даде сравнително качествен резултат. За публичния сектор обаче е необходима редакция от служител, защото формулировките трябва да са точни, юридически издържани и съобразени с институционалния тон.
Reasoning / разсъждение <i>логически вериги, многостъпкови задачи</i>	Оценка 1–5 ○1 ○2  ○4 ○5	Моделът може да подпомага многостъпков анализ, сравнение на варианти, извеждане на аргументи и подреждане на логически зависимости. Въпреки това не може да се приеме като надежден самостоятелен механизъм за сложни правни, финансови или административни преценки. Добре е да се използва за предварителен анализ, но крайната преценка трябва да остане при компетентен служител.
Code writing <i>генериране и анализ на код</i>	Оценка 1–5 ○1 ○2 ○3  ○5	Моделът е силно приложим за генериране на примерен код, автоматизационни скриптове, документация, unit тестове и обяснение на съществуващ код. Това е реална полза за ИТ звена в администрацията. Ограничението е, че генерираният код може да съдържа уязвимости, неефективни конструкции или несъответствие с вътрешните стандарти, затова е задължителен преглед от разработчик или системен администратор.
Анализ на изображения <i>vision, диаграми, скрийншоти</i>	Оценка 1–5 ○1  ○3 ○4 ○5	Функционално някои версии на Llama поддържат работа с изображения и мултимодален вход, но за публична администрация в ЕС има съществена лицензионна и регулаторна неяснота при Llama 4 мултимодалните модели. Поради това към момента оценката трябва да е консервативна. Приложимост може да се разглежда само след правна проверка на конкретната версия и при ясно ограничени случаи, например анализ на диаграми, схеми или публични изображения без лични данни.

Аудио / глас <i>говорим вход/изход</i>	Оценка 1–5 ○1 ☒ ☐ ○3 ☐ ○4 ☐ ○5	Ллама като модел не следва да се разглежда самостоятелно като цялостна гласова платформа за публичната администрация. Възможна е интеграция с отделни компоненти за разпознаване на реч и синтез на глас, но това вече е комбинирана система, която изисква отделна оценка на риска, защита на личните данни и контрол върху записите. Не се препоръчва за обработка на чувствителни разговори без допълнителна правна и техническа рамка.
Видео разбиране <i>анализ на видеосъдържание</i>	Оценка 1–5 ○1 ☒ ☐ ○3 ☐ ○4 ☐ ○5	Видео анализът не следва да се приема като основна и зряла функция за административна употреба във вътрешна инфраструктура без допълнителни специализирани инструменти. Прилагането му може да засегне лични данни, биометрични данни, наблюдение и високи регулаторни рискове. За публичния сектор тази функционалност трябва да остане извън пилотния обхват, освен ако не се разработи отделна оценка на въздействието и правно основание.
Дълги документи <i>обработка на обемни файлове</i>	Оценка 1–5 ○1 ☐ ○2 ☐ ○3 ☒ ☐ ○5	При подходяща версия и достатъчна инфраструктура моделът може да бъде полезен за обработка на дълги документи, наредби, стратегии, доклади, процедури и вътрешни правила. Това е силна приложимост за администрацията, защото голяма част от работата е документно ориентирана. Ограничението е, че при много обемни документи моделът може да пропусне детайли или да обобщи прекалено свободно, затова трябва да се използват контролни въпроси и проверка по източник.
Agentic режим <i>автономни многостъпкови задачи</i>	Оценка 1–5 ○1 ☐ ○2 ☐ ☒ ○4 ☐ ○5	Моделът може да бъде включен в agentic режим чрез вътрешна оркестрация, например за последователно търсене, резюмиране, класифициране и подготовка на проектни резултати. За публичния сектор това трябва да се въвежда внимателно, защото автономното изпълнение на многостъпкови задачи може да създаде грешки, които трудно се проследяват. Допустимо е само при ограничени права, логване, човешко одобрение и забрана за самостоятелно изпращане, публикуване или промяна на данни.
API / системна интеграция <i>свързване с вътрешни системи</i>	Оценка 1–5 ○1 ☐ ○2 ☐ ○3 ☒ ☐ ○5	При вътрешна инфраструктура моделът може да бъде интегриран чрез API с вътрешни портали, системи за документооборот, бази знания, регистри и аналитични среди. Това е една от основните ползи на локалното внедряване, защото организацията може да контролира достъпа, логовете, мрежовата сегментация и политиките за сигурност. Необходимо е интеграцията да мине през стандартен процес по управление на промени, оценка на риска и контрол на достъпа.
Plugin екосистема <i>налични разширения и конектори</i>	Оценка 1–5 ○1 ☐ ○2 ☐ ☒ ○4 ☐ ○5	Около Llama има активна техническа екосистема, но тя не е равнозначна на готова, централизирана и сертифицирана plugin среда за публична администрация. Разширенията и конекторите могат да бъдат полезни, но трябва да се одобряват индивидуално. Вътрешните конектори към регистри, документооборот и бази знания са за предпочитане пред външни плъгини с неясен произход.
Desktop приложение <i>нативна desktop среда</i>	Оценка 1–5 ○1 ☒ ☐ ○3 ☐ ○4 ☐ ○5	Моделът сам по себе си не представлява готово desktop приложение за администрацията. Може да бъде предоставен чрез вътрешен уеб интерфейс или контролирано desktop решение, но това изисква допълнителна разработка, управление на достъпа и поддръжка. Не следва да се допуска неуправлявана локална инсталация по работни станции без централизирана политика.
Mobile приложение <i>iOS / Android достъп</i>	Оценка 1–5 ○1 ☒ ☐ ○3 ☐ ○4 ☐ ○5	Мобилен достъп е възможен само ако организацията изгради защитен интерфейс, управляван през служебни устройства, VPN, MDM и ролеви достъп. За публичния сектор мобилното използване носи допълнителни рискове, включително загуба на устройство, неоторизиран достъп и въвеждане на чувствителна информация в неконтролирана среда. На този етап е подходящо да се разглежда като ограничена бъдеща възможност, а не като основна функция.
Cowork / колаборация <i>екипна работа в реално време</i>	Оценка 1–5 ○1 ☒ ☐ ○3 ☐ ○4 ☐ ○5	Моделът не осигурява самостоятелно пълноценна среда за екипна работа в реално време. Такава функционалност може да бъде изградена чрез вътрешна платформа, но трябва да се решат въпросите за версии, права, история на действията и отговорност за крайния документ. Приложимостта е частична и зависи повече от платформата около модела, отколкото от самия модел.
Търсене в интернет <i>актуална информация от мрежата</i>	Оценка 1–5 ○1 ☒ ☐ ○3 ☐ ○4 ☐ ○5	При вътрешна инфраструктура моделът по подразбиране не трябва да има свободен и неконтролиран достъп до интернет. Възможно е да се добави контролиран модул за търсене или retrieval система към одобрени източници, но това трябва да бъде отделно управлявана функция. За публичния сектор е по-безопасно моделът да работи с вътрешна база знания и официално одобрени източници, вместо със свободно уеб търсене.

Fine-tuning <i>настройка за специфична среда</i>	Оценка 1–5 ○1 ○2 ○3 5	LLama е подходяща фамилия модели за адаптиране към конкретна среда, включително чрез донастройка, инструкции, retrieval-augmented generation и специализирани вътрешни корпуси. Това е силна страна при вътрешна инфраструктура, защото администрацията може да подобри стила, терминологията и релевантността към собствените си процеси. Необходимо е обаче обучение само върху разрешени, почистени и правно допустими данни, с контрол върху личните данни, авторските права и служебната информация.
--	---	---

B. SWOT Матрица

Обобщете основните силни/слаби страни, възможности и заплахи на модела в публичен контекст.

 **Инструкция за попълване:** Запишете поне 4 наблюдения в клетка. Базирайте се на попълнените оценки от Раздел А.

S STRENGTHS — Силни страни - Възможност за внедряване във вътрешна инфраструктура, което позволява по-добър контрол върху данните, достъпа, логовете и мрежовата среда. - Намалена зависимост от публична облачна услуга, когато моделът се хоства и управлява в контролирана среда на организацията или в одобрена държавна инфраструктура. - Добра приложимост при обработка на административни документи, резюмета, вътрешни справки, проекти на писма, инструкции, доклади и обучителни материали. - Възможност за адаптиране към нуждите на публичната администрация чрез вътрешни бази знания, контролирани корпуси, шаблони, инструкции и специализирани правила за употреба. - Подходящост за подпомагане на ИТ звена при генериране на примерен код, автоматизационни скриптове, документация и анализ на технически проблеми. - По-голяма гъвкавост спрямо напълно затворени AI платформи, тъй като организацията може да изгради собствена архитектура, собствени правила за достъп и собствена одитна следа. <i>Допълнителни бележки</i> Силните страни са най-изразени при вътрешна, контролирана и ограничена употреба, където моделът подпомага служителите, но не взема самостоятелни решения. Най-подходящите начални приложения са обработка на документи, вътрешни справки, обучение, техническа помощ и подпомагане на административна комуникация.	W WEAKNESSES — Слаби страни - Необходима е сериозна техническа инфраструктура за надеждно вътрешно внедряване, включително сървърен ресурс, управление на модели, сигурност, мониторинг и поддръжка. - Качеството на отговорите на български език, особено при правна и административна терминология, трябва да бъде допълнително тествано и валидирано в реални сценарии. - Моделът може да генерира фактически неточности, непълни изводи или убедително звучащи, но неправилни правни и административни формулировки. - Липсата на самостоятелна обяснимост при някои резултати затруднява използването му в процеси, при които се изисква ясна правна, административна или одитна аргументация. - Мултимодалните функции, мобилният достъп, колаборацията и гласовите възможности не следва да се приемат като готови за употреба без допълнителна платформа, правна проверка и оценка на риска. - Вътрешното внедряване намалява някои рискове, но не ги премахва напълно, защото остават въпроси за качеството на обучителните данни, лицензионните условия, версиите на модела и контрола върху донастройката. <i>Допълнителни бележки</i> Основната слабост не е само в самия модел, а в сложността на неговото управление. За публичния сектор такова решение не може да бъде оставено само на технически екип. Нужни са правно звено, длъжностно лице по защита на данните, информационна сигурност, вътрешен одит и ясно определен AI отговорник.
O OPPORTUNITIES — Възможности Създаване на вътрешен AI асистент за публичната администрация, който работи с одобрени вътрешни	T THREATS — Заплахи Риск от използване на модела за автоматизирани или фактически автоматизирани решения, които могат да засегнат права, задължения или законни интереси на граждани.

правила, процедури, наръчници, често задавани въпроси и публични нормативни източници. 2. Подобряване на административната ефективност чрез подпомагане на служителите при резюмиране, структуриране, редактиране и сравнение на голям обем документи. 3. Изграждане на контролирана база знания за вътрешна употреба, която да намали зависимостта от неформални практики, разпокъсани указания и индивидуална експертиза на отделни служители. 4. Подпомагане на обученята в администрацията чрез генериране на тестове, казуси, сценарии, примерни ситуации и учебни материали по информационна сигурност, защита на данните, етика и административни процеси. 5. Възможност за по-добро съответствие с ISO 27001 и ISO 42001 чрез въвеждане на управление на риска, контрол на достъпа, проследимост, периодичен преглед, документирани роли и одитируем процес на използване на AI. 6. Развитие на пилотен модел за отговорно използване на AI в публичния сектор, който по-късно може да бъде разширен към други администрации, ако покаже достатъчно надеждност и управляем риск <i>Допълнителни бележки:</i> <i>Най-реалистичната възможност е да се започне с ограничен пилотен проект, а не с масово внедряване. Първият етап може да обхване вътрешни документи, обучения, справки и техническа помощ, без обработване на чувствителни данни и без директна комуникация с граждани по индивидуални казуси</i>	2. Риск от въвеждане на лични данни, служебна информация, чувствителни административни данни или класифицирана информация без достатъчен контрол и без ясно правно основание. 3. Риск от несъответствие с AI Act, GDPR, NIS 2, Закона за киберсигурност, НМИМИС и вътрешните правила за информационна сигурност, ако липсват процедури, роли и одитна следа. 4. Риск от прекомерно доверие на служителите в генерираните резултати, особено когато отговорите звучат убедително, но съдържат неточности, пропуски или неправилна правна логика. 5. Риск от репутационни щети при използване на грешен AI изход в официален документ, отговор до граждани, публична комуникация или управленско решение. 6. Риск от лицензионни, технологични или геополитически ограничения, включително промяна в условията за използване на модела, ограничения за определени версии или зависимост от външен доставчик на базовия модел <i>Допълнителни бележки</i> <i>Най-сериозните заплахи са свързани не толкова с техническата работа на модела, а с неправилното му организационно използване. При липса на човешки контрол, правна проверка, ограничаване на данните и одитна следа дори вътрешно внедрен модел може да създаде висок риск за администрацията.</i>
--	---

C. Разширен SWOT по специфични теми

 Инструкция за попълване: Попълнете текстово наблюденията си за всяка тема. Използвайте конкретни примери.
Цена и лицензиране Наблюдения: (Модел на плащане, достъпност за малки общини, API разходи) При вътрешна инфраструктура Meta Llama може да намали зависимостта от платени облачни AI услуги, но реалната цена не е нулева. Трябва да се отчетат разходите за сървъри, графични ускорители, поддръжка, сигурност, мониторинг, обучение, правна оценка и одит. Необходимо е и предварително да се провери лицензът на конкретната версия, особено ако се предвижда използване в публичния сектор и в рамките на ЕС. Препоръка за публичния сектор: Да се изготви анализ на общата цена на притежание, включващ технически, организационни и регулаторни разходи. Изборът на версия следва да бъде одобрен след правна проверка на лицензионните условия.
Сигурност на данните Наблюдения: (Обработка на лични данни, GDPR, съхранение в облака) Вътрешната инфраструктура ограничава риска от изпращане на данни към външен доставчик, но не премахва рисковете от неправомерен достъп, неправилни права, грешна конфигурация или въвеждане на чувствителна информация от служители. Особено внимание изискват личните данни, служебната информация и документите с ограничен достъп. Препоръка за публичния сектор:

Да се прилага принципът „минимум данни към AI“. Да се въведат ролеви достъп, криптиране, мрежова сегментация, централизирано логване, мониторинг и ясна процедура за инциденти. Да се забрани въвеждането на класифицирана информация.
Езикова поддръжка
Наблюдения: <i>(Качество на български език, юридическа терминология)</i> Моделът може да се използва за български административни текстове, резюмета, вътрешни съобщения и обучителни материали. По-рискови са правните и нормативните текстове, където са нужни точна терминология, актуални препратки и институционален стил.
Препоръка за публичния сектор: Да се използва за първи проект, редакция и структуриране на текстове, но не и като окончателен автор на официални документи. При правни и нормативни материали да има задължителна проверка от компетентно звено.
Vendor lock-in
Наблюдения: <i>(Зависимост от доставчик, алтернативи, преносимост)</i> При вътрешно внедряване зависимостта от облачен доставчик е по-ниска, но остава зависимост от лиценза, бъдещите версии, хардуерните изисквания и екосистемата около модела. Ако процесите се изградят само около една версия на Llama, бъдеща миграция може да бъде затруднена.
Препоръка за публичния сектор: Да се използва модулна архитектура, която позволява замяна на модела. Да се поддържат отделна база знания, стандартни логове, документирани критерии за избор на модел и периодичен сравнителен преглед с алтернативи.
Сложност на внедряване
Наблюдения: <i>(Техническа зрялост, необходима инфраструктура, IT капацитет)</i> Вътрешното внедряване дава по-добър контрол, но изисква сериозна техническа и организационна подготовка. Нужни са инфраструктура, сигурност, мониторинг, управление на достъпа, правила за употреба, обучение и отговорни звена.
Препоръка за публичния сектор: Да се започне с ограничен пилотен проект за нискорискови вътрешни задачи: резюмета, проектни текстове, обучения, справки и техническа помощ. Разширяване да се допуска само след оценка на резултатите, рисковете и натоварването.
Регулаторно съответствие
Наблюдения: <i>(EU AI Act, NIS2/3KC, GDPR, национални наредби)</i> Използването на AI в публичния сектор трябва да бъде съобразено с AI Act, GDPR, NIS 2, Закона за киберсигурност, НМИМИС, ISO 27001 и ISO 42001. Най-чувствителни са случаите, при които AI влияе върху решения с ефект върху граждани.
Препоръка за публичния сектор: Да се забранят автоматизирани решения без човешки контрол. За всяка нова употреба да се прави предварителна оценка от правно звено, длъжностно лице по защита на данните, информационна сигурност и AI отговорник. Да се водят регистър на AI употребите, одитна следа и отчетност.
Етика и прозрачност
Наблюдения: <i>(Обяснимост, одитируемост, отговорност при грешки)</i> AI не трябва да се използва скрито в процеси, които засягат граждани или служители. Липсата на прозрачност може да създаде съмнение, че администрацията прехвърля човешка отговорност към технологията.
Препоръка за публичния сектор: Да се въведе правило, че AI е само помощен инструмент и не замества служителя. При използване в официални документи или комуникация с граждани да има вътрешна проследимост, а при нужда и ясно уведомяване.
Изкривявания (Bias)
Наблюдения: <i>(Пристрастия при различни групи, равнопоставеност)</i>

Моделът може да възпроизведе пристрастия от обучителните данни или от начина на задаване на въпроса. В публичния сектор това е риск при заявления, жалби, социални услуги, комуникация с уязвими групи и оценка на казуси.
Препоръка за публичния сектор: Да се забрани самостоятелна AI оценка на граждани, заявления, допустимост, риск или приоритет. Да се провеждат периодични тестове за пристрастия с анонимизирани сценарии и да се осигури възможност за човешка намеса.
Халюцинации / грешки
Наблюдения: (<i>Честота, detectability, риск при вземане на решения</i>) Моделът може да генерира неточни факти, несъществуващи правни основания, грешни срокове или убедително звучащи, но неправилни изводи. Това е сериозен риск при официални документи, правни анализи и отговори до граждани.
Препоръка за публичния сектор: Всички AI резултати, използвани в официални документи, правни становища, публични съобщения или отговори до граждани, да се проверяват от компетентен служител. AI изходът не следва да бъде единствено основание за административно действие.
Морал и ценности
Наблюдения: (<i>Поведение при чувствителни теми, отказ на вредни задачи</i>) Моделът трябва да се използва в съответствие с принципите на законност, безпристрастност, равнопоставеност, отчетност и защита на обществения интерес. Особено внимание е нужно при жалби, санкции, социално уязвими групи, кризисна комуникация и чувствителни казуси.
Препоръка за публичния сектор: Да се въведат правила за институционален тон, неутралност и недопускане на подвеждащи или дискриминационни формулировки. При чувствителни теми да има задължителна човешка редакция и одобрение от компетентно звено.

D. Обобщена оценка на групата

Обобщена оценка по основни показатели		
Обща приложимост <i>средна от Раздел А</i>	Оценка 1–5 ○1 ○2  ○4 ○5	Обосновка / бележки: Моделът има добра приложимост за вътрешни административни задачи, обработка на документи, подготовка на проекти, обучения, справки и техническа помощ. Оценката не е висока, защото употребата му в публичния сектор трябва да бъде силно ограничена при правни анализи, граждански услуги, автоматизирани решения и чувствителни данни.
Готовност за внедряване <i>техническа и организациона</i>	Оценка 1–5 ○1 ○2  ○4 ○5	Обосновка / бележки: При вътрешна инфраструктура моделът може да бъде внедрен контролирано, но това изисква сериозен технически, организационен и правен капацитет. Необходими са съвършена среда, управление на достъпа, логване, мониторинг, правила за употреба, обучение и ясно разпределение на отговорностите. Затова моделът е по-подходящ за пилотно внедряване, а не за незабавно широко използване.
Съответствие с регулации <i>AI Act, GDPR, NIS2</i>	Оценка 1–5 ○1 ○2  ○4 ○5	Обосновка / бележки: Моделът може да бъде приведен в съответствие с AI Act, GDPR, NIS 2, Закона за киберсигурност, НМИМИС, ISO 27001 и ISO 42001, но само при наличие на допълнителни контролни мерки. Самото вътрешно внедряване не е достатъчно. Нужни са регистър на AI употребите, оценка на риска, правила за данните, човешки контрол, одитна следа и забрана за автоматизирани решения с ефект върху граждани.
Етичен риск <i>обратна скала: 5=нисък риск</i>	Оценка 1–5 ○1 ○2  ○4 ○5	Обосновка / бележки: Етичният риск е управляем, но не може да се определи като нисък. Съществуват рискове от систематични пристрастия и изкривявания в резултатите на модела, грешни или непълни отговори, прекомерно доверие от страна на служителите, неясна отговорност и недостатъчна прозрачност спрямо гражданите. Този риск може да бъде намален чрез задължителна човешка проверка, ясни правила за прозрачност, забрана за самостоятелна оценка на граждани от AI система и периодични проверки за наличие на пристрастия и неравнопоставено третиране.
Обща препоръка	Оценка 1–5	Обосновка / бележки:

1=не препоръчваме, 5=силно препоръчваме	○1 ○2 ○3 ■4 ○5	Моделът може да бъде препоръчан за публичния сектор, но само като помощен инструмент във вътрешна инфраструктура и при ясно определени ограничения. Най-подходящ е за ниско- и среднорискови задачи: резюмиране, структуриране на текстове, вътрешни справки, обучение, техническа помощ и подготовка на проектни документи. Не се препоръчва за самостоятелно вземане на решения, правно обвързващи становища, обработка на класифицирана информация или автоматизирана оценка на граждани.
Мотивирано становище на групата (мин. 5 изречения): <i>Обобщете защо препоръчвате или не препоръчвате модела за публичния сектор.</i> Моделът Meta Llama, разглеждан във вариант на вътрешна инфраструктура, има потенциал за използване в публичната администрация като помощен инструмент, но не и като самостоятелна система за вземане на решения. Най-подходящ е за вътрешни задачи като обработка на документи, справки, обучения, техническа помощ и подготовка на проектни текстове. Употребата му следва да бъде ограничена до пилотен режим, с ясни правила, човешки контрол, правна проверка и одитна следа. Моделът не трябва да се използва за автоматизирани решения с ефект върху граждани, нито за обработка на класифицирана или чувствителна информация. Общата позиция е, че Meta Llama може да бъде полезен за публичния сектор, ако се прилага предпазливо, прозрачно и при ясно разпределени отговорности.		
Предложени условия за ползване: <i>Какви ограничения, контроли или процедури трябва да съпътстват внедряването?</i> - Моделът да се използва само в контролирана вътрешна инфраструктура или в одобрена среда, отговаряща на изискванията за мрежова и информационна сигурност. - Да не се допуска обработване на класифицирана информация, а лични и чувствителни данни да се използват само при изрично правно основание, оценка на риска и одобрена процедура. - Всички резултати, предназначени за официални документи, правни становища, публична комуникация или отговори до граждани, да подлежат на задължителна човешка верификация. - Да се забрани използването на модела за самостоятелно вземане на решения с правен или фактически ефект върху права, задължения или законни интереси на граждани. - Да се осигурят регистър на употребите, проследима одитна следа, логване на взаимодействията, периодичен контрол и механизъм за докладване на инциденти. - Всяка нова употреба на модела да се допуска само след предварителна оценка от компетентните звена по право, защита на данните, информационна сигурност и управление на AI. - Служителите да преминават обучение за допустима употреба, защита на данните, разпознаване на грешки, пристрастия и ограничения на модела. - Да се извършва периодична ревизия на модела, настройките, източниците на данни, качеството на резултатите и съответствието с приложимата нормативна рамка.		

Към настоящия SWOT анализ се прилага отделен Excel файл с приложения, съдържащ план и протокол от тестване, анализ на лицензионните условия и използваните източници, карта на регулаторното и стандартно съответствие, регистър и класификация на AI употребите, както и контролна рамка за внедряване, човешка верификация, сигурност и реакция при инциденти. Приложенията имат за цел да осигурят проследимост, доказуемост и по-висока степен на одитна готовност при евентуално пилотно използване на Meta Llama в публичния сектор.

1. Приложения_към_1-5_SWOT_AI_Model_Meta_Llama.xlsx



Приложения_към_1-5_SWOT_AI_Model

Risk Canvas

Картографиране на рисковете от AI модел в публичен контекст

v1.0 — Работен шаблон

 **Инструкция за попълване:** Попълнете всички полета за избория от вашата група AI модел. Празните редове и скалите са предназначени за ръчно попълване. Не изтривайте инструкциите в курсив.

Наименование на модела	Llama 4
Разработчик / компания	META
Версия за оценка	Версия 4
Дата на попълване	14.05.2026
Членове на групата	Ани Христова, Антония Желева, Дейвид Александров, Димитър Ходулов, Красимир Червенков, Любчо Милушев
Ръководител на групата	Ани Христова

Легенда за скалите

Вероятност (Вер.) 1 = Изключително рядко 2 = Рядко 3 = Умерено вероятно 4 = Вероятно 5 = Почти сигурно	Въздействие (Вл.) 1 = Незначително 2 = Малко 3 = Умерено 4 = Значително 5 = Критично	Ниво на риска Н = Нисък (Вер.×Вл. ≤ 4) С = Среден (5–12) В = Висок (13–25)
--	--	--

Риск Канва — 9-полева матрица

Попълнете всяко поле с идентифицирани рискове, засегнати страни и ключови наблюдения.

 **Инструкция за попълване:** Всяко от 9-те полета представлява различна категория риск. Запишете конкретни рискове — не общи фрази.

1. Технически рискове <i>Халюцинации, грешки, downtime, версионни промени</i> 1. Халюцинации и фактически неточни отговори.	2. Рискове за данните <i>Поверителност, GDPR, изтичане, съхранение</i> 1. Въвеждане на лични данни без ясно правно основание.	3. Правни / регулаторни <i>EU AI Act, NIS2, GDPR, национално право</i>
--	--	--

2. Грешки при обработка на дълги административни документи. 3. Непредвидимо поведение след промяна на версията или настройките на модела. 4. Ограничена надеждност при сложни правни, технически или многостъпкови задачи.	2. Обработка на служебна или чувствителна информация извън одобрените правила. 3. Недостатъчен контрол върху логовете, заявките и генерираните отговори. 4. Риск от неправилен достъп при неправилно зададени роли и права.	1. Несъответствие с AI Act при употреба в процеси с ефект върху граждани. 2. Нарушение на GDPR при обработване на лични данни чрез модела. 3. Несъответствие с NIS 2, Закона за киберсигурност и НМИМИС при липса на мерки за сигурност. 4. Липса на достатъчна одитна следа при използване на AI резултати в официални документи.
4. Етични рискове <i>Bias, дискриминация, непрозрачност, автономия</i> 1. Систематични пристрастия или изкривявания в резултатите на модела. 2. Неравнопоставено третиране при анализ на заявления, жалби или граждански казуси. 3. Недостатъчна прозрачност към гражданите при използване на AI. 4. Прекомерно доверие в модела и отслабване на човешката преценка.	5. Оперативни рискове <i>Зависимост, интеграция, обучение, процеси</i> 1. Недостатъчен технически капацитет за вътрешно внедряване и поддръжка. 2. Липса на ясно определени роли, отговорности и процедури за използване. 3. Недостатъчно обучение на служителите за допустима и безопасна употреба. 4. Интеграционни затруднения с вътрешни системи, регистри и документооборот.	6. Репутационни рискове <i>Доверие на гражданите, медийно отразяване</i> 1. Публичен скандал при използване на грешен AI резултат в официален документ. 2. Загуба на доверие от гражданите при неясна или скрита употреба на AI. 3. Негативна медийна реакция при грешка, свързана с права или услуги за граждани. 4. Съмнения за прехвърляне на човешка отговорност към AI система.
7. Стратегически рискове <i>Vendor lock-in, ценови промени, спиране на услугата</i> 1. Зависимост от технологичната екосистема и лицензионните условия на Meta. 2. Трудна миграция към друг модел при изградени процеси само около Llama. 3. Промяна в лицензионните условия, версиите или възможностите на модела. 4. Неясна дългосрочна устойчивост на вътрешната инфраструктура без достатъчен ресурс.	8. Засегнати страни <i>Граждани, служители, ръководство, партньори</i> 1. Граждани, чиито права, заявления или услуги могат да бъдат косвено засегнати. 2. Служители, които използват AI резултати в административната си работа. 3. Ръководство, което носи отговорност за внедряването и контрола. 4. ИТ, информационна сигурност, правно звено, ДЛЗД и вътрешен одит.	9. Остатъчни рискове <i>Рискове след митигация, приети рискове</i> 1. Възможност за фактически грешки въпреки човешка проверка. 2. Риск от неправилна употреба от служители въпреки проведено обучение. 3. Остатъчен риск от пристрастия при непълни или некачествени входни данни. 4. Риск от непълна проследимост, ако логването и контролът не са технически добре реализирани.

Рисков регистър

Попълнете по всяка категория конкретните рискове с вероятност, въздействие и мерки.

 **Инструкция за попълване:**

Вер. = Вероятност (1–5) · Вл. = Въздействие (1–5) · Ниво: ☐Н (нисък) ☐С (среден) ☐В (висок)

Риск / Събитие	Вероятност	Въздействие	Ниво	Мярка за митигация
----------------	------------	-------------	------	--------------------

Технически рискове				
Халюцинации и фактически грешки при отговори	Вер. ○1 ○2 ○3 ○5	Вл. ○1 ○2 ○3 ○4	Ниво □Н □С В	Мярка за митигация: Задължителна човешка проверка на всички резултати, използвани в официални документи, правни становища, справки или отговори до граждани. AI изходът не се използва като самостоятелно основание за административно действие.
Недостъпност / downtime на платформата	Вер. ○1 ○2 ○4 ○5	Вл. ○1 ○2 ○3 ○5	Ниво □Н С □В	Мярка за митигация: Осигуряване на мониторинг, резервираност, план за възстановяване и процедура за преминаване към ръчна обработка при недостъпност
Внезапна промяна на версия или поведение на модела	Вер. ○1 ○2 ○4 ○5	Вл. ○1 ○2 ○3 ○5	Ниво □Н □С □В	Мярка за митигация: Всяка промяна на модела, настройките или системните инструкции да се тества предварително и да се одобрява по процедура за управление на промени.
Ограничения на контекстния прозорец при обемни документи	Вер. ○1 ○2 ○4 ○5	Вл. ○1 ○2 ○4 ○5	Ниво □Н С □В	Мярка за митигация: Обемните документи да се обработват по части, с проверка по първоизточник и контролни въпроси. Да не се приема автоматично, че моделът е отчетел целия документ.
Грешки при мултимодален вход (изображения, аудио)	Вер. ○1 ○2 ○4 ○5	Вл. ○1 ○2 ○4 ○5	Ниво □Н С □В	Мярка за митигация: Мултимодалните функции да се използват само след отделна оценка на риска. Да не се прилагат за биометричен анализ, идентификация на лица или чувствителни казуси без изрично одобрение.
Рискове за данните и поверителността				
Обработка на лични данни на граждани в облачна среда	Вер. ○1 ○3 ○4 ○5	Вл. ○1 ○2 ○3 ○4	Ниво □Н С □В	Мярка за митигация: Да се използва вътрешна инфраструктура. Лични данни да се обработват само при правно основание, минимизиране на данните и участие на длъжностното лице по защита на данните.
Изтичане на класифицирана/служебна информация	Вер. ○1 ○2 ○4 ○5	Вл. ○1 ○2 ○3 ○4	Ниво □Н □С В	Мярка за митигация: Да се забрани въвеждането на класифицирана информация. За служебна информация да се прилагат ролеви права, логване, мрежова сегментация и контрол от звеното по информационна сигурност.
Запазване на запитвания от доставчика за обучение	Вер. ○1 ○3 ○4 ○5	Вл. ○1 ○2 ○3 ○5	Ниво □Н С □В	Мярка за митигация: Да не се използват публични AI услуги за служебни данни. При вътрешна инфраструктура да се изключи нерегламентирано изпращане на заявки, логове или телеметрия към външни доставчици.
Трансфер на данни извън ЕС (EU AI Act / Schrems II)	Вер. ○1 ○3 ○4 ○5	Вл. ○1 ○2 ○3 ○4	Ниво □Н С □В	Мярка за митигация: Данните и логовете да се съхраняват в контролирана вътрешна или одобрена европейска среда. Външни API, облачни компоненти или телеметрия да се допускат само след правна и техническа оценка.
Липса на одитна следа при автоматизирани решения	Вер. ○1 ○2 ○4 ○5	Вл. ○1 ○2 ○3 ○4	Ниво □Н □С В	Мярка за митигация: Да се забранят автоматизирани решения без човешки контрол. За всяка официална употреба да се съхраняват потребител, дата, час, вход, изход, версия на модела и служител, извършил проверката.
Правни и регулаторни рискове				
Несъответствие с EU AI Act (Annex III — публичен сектор)	Вер. ○1 ○2 ○4 ○5	Вл. ○1 ○2 ○3 ○4	Ниво	Мярка за митигация: Всяка употреба да се класифицира по риск. Сценарии с ефект върху права, задължения или достъп до услуги да

			☐ Н ☐ С ☒ В	се допускат само след правна оценка, човешки контрол и одобрена процедура.
Нарушение на GDPR при обработка на лични данни	Вер. ○1 ○2 ☒ ○4 ○5	Вл. ○1 ○2 ○3 ○4 ☒	Ниво ☐ Н ☐ С ☒ В	<i>Мярка за митигация:</i> Да се прилагат минимизиране на данните, ограничаване на целите, контрол на достъпа и отчетност. При по-рискови обработки да се извършва оценка на въздействието върху защитата на данните.
Липса на обяснимост при автоматизирани административни решения	Вер. ○1 ○2 ☒ ○4 ○5	Вл. ○1 ○2 ○3 ○4 ☒	Ниво ☐ Н ☐ С ☒ В	<i>Мярка за митигация:</i> AI да не взема самостоятелни административни решения. Служителят трябва да може да обоснове решението чрез закон, факти и документи, а не само чрез AI резултат.
Авторскоправни нарушения в генерирано съдържание	Вер. ○1 ○2 ☒ ○4 ○5	Вл. ○1 ○2 ☒ ○4 ○5	Ниво ☐ Н ☒ С ☐ В	<i>Мярка за митигация:</i> Генерираното съдържание да се проверява преди публикуване или официална употреба. При съмнение да се търси становище от правно звено.
Отговорност при грешни правни или медицински изходи	Вер. ○1 ○2 ○3 ☒ ○5	Вл. ○1 ○2 ○3 ○4 ☒	Ниво ☐ Н ☐ С ☒ В	<i>Мярка за митигация:</i> Да се забрани използването на модела като самостоятелен източник на правни, медицински или други специализирани заключения. Допуска се само като работна чернова след експертна проверка.
Етични и социални рискове				
Систематичен bias спрямо определени групи граждани	Вер. ○1 ○2 ☒ ○4 ○5	Вл. ○1 ○2 ○3 ☒ ○5	Ниво ☐ Н ☒ С ☐ В	<i>Мярка за митигация:</i> Да се извършват периодични тестове за пристрастия с анонимизирани сценарии. Моделът да не се използва за самостоятелна оценка на граждани, заявления, допустимост или приоритет.
Дискриминация при автоматизирано обработване на заявления	Вер. ○1 ○2 ☒ ○4 ○5	Вл. ○1 ○2 ○3 ○4 ☒	Ниво ☐ Н ☐ С ☒ В	<i>Мярка за митигация:</i> Да се забрани автоматизирано класиране, отхвърляне или оценяване на заявления без човешки контрол. Всички предложения на модела да се разглеждат само като помощна информация.
Непрозрачност — граждани не знаят, че AI взима решения	Вер. ○1 ○2 ☒ ○4 ○5	Вл. ○1 ○2 ○3 ☒ ○5	Ниво ☐ Н ☒ С ☐ В	<i>Мярка за митигация:</i> Да се въведат правила за прозрачност. При използване на AI в комуникация с граждани или при обработка на искане това да се обозначава, когато е приложимо.
Дезинформация и манипулация чрез убедително съдържание	Вер. ○1 ○2 ☒ ○4 ○5	Вл. ○1 ○2 ○3 ☒ ○5	Ниво ☐ Н ☒ С ☐ В	<i>Мярка за митигация:</i> AI съдържание за публична комуникация да се публикува само след редакция и одобрение от компетентно звено. Да се забрани употреба за подвеждащи или манипулативни съобщения.
Загуба на критично мислене и зависимост при служителите	Вер. ○1 ○2 ○3 ☒ ○5	Вл. ○1 ○2 ☒ ○4 ○5	Ниво ☐ Н ☒ С ☐ В	<i>Мярка за митигация:</i> Да се провеждат обучения за ограниченията на AI, разпознаване на грешки и задължителна проверка. Вътрешните правила да подчертават, че отговорността остава при служителя.
Оперативни рискове				
Липса на технически капацитет за внедряване и поддръжка	Вер. ○1 ○2 ☒ ○4 ○5	Вл. ○1 ○2 ○3 ☒ ○5	Ниво ☐ Н ☒ С ☐ В	<i>Мярка за митигация:</i> Да се започне с ограничен пилотен проект. Преди внедряване да се оцени капацитетът за инфраструктура, администриране, мониторинг, архивиране и поддръжка.

Vendor lock-in — прекомерна зависимост от доставчик	Вер. ○1 ○2 ■3 ○4 ○5	Вл. ○1 ○2 ○3 ■4 ○5	Ниво □Н ■С □В	Мярка за митигация: Да се използва модулна архитектура, отделна база знания и възможност за замяна на модела. Настройките, интеграциите и критериите за избор да се документират.
Недостатъчно обучение на служителите	Вер. ○1 ○2 ○3 ■4 ○5	Вл. ○1 ○2 ○3 ■4 ○5	Ниво □Н □С ■В	Мярка за митигация: Достъп до модела да се предоставя само след обучение. То трябва да обхваща допустими и забранени употреби, защита на данните, разпознаване на грешки и докладване на проблеми.
Интеграционни проблеми с наследени (legacy) системи	Вер. ○1 ○2 ■3 ○4 ○5	Вл. ○1 ○2 ○3 ■4 ○5	Ниво □Н ■С □В	Мярка за митигация: Интеграциите да се реализират поетапно, първо в тестова среда. Да се използват защитени API връзки, минимални права, журнал на действията и процедура за одобрение на промени.
Ценови скокове при промяна на лицензионния модел	Вер. ○1 ○2 ■3 ○4 ○5	Вл. ○1 ○2 ■3 ○4 ○5	Ниво □Н ■С □В	Мярка за митигация: Да се изготви анализ на общата цена на притежание и план за алтернативен модел. Да се следят лицензионните условия и да се предвиди ресурс за миграция.
Репутационни и стратегически рискове				
Публичен скандал при видима AI грешка (медийно отразяване)	Вер. ○1 ○2 ■3 ○4 ○5	Вл. ○1 ○2 ○3 ○4 ■5	Ниво □Н □С ■В	Мярка за митигация: Всички AI резултати за официални документи, публични съобщения и отговори до граждани да се проверяват и одобряват от компетентно звено преди използване.
Загуба на доверие на гражданите в публичните институции	Вер. ○1 ○2 ■3 ○4 ○5	Вл. ○1 ○2 ○3 ○4 ■5	Ниво □Н □С ■В	Мярка за митигация: Да се осигурят прозрачност, човешки контрол и възможност за обяснение на AI употребата. Да не се допуска впечатление, че решенията се вземат от AI вместо от компетентен орган.
Политическа съпротива срещу използването на AI	Вер. ○1 ○2 ■3 ○4 ○5	Вл. ○1 ○2 ○3 ■4 ○5	Ниво □Н ■С □В	Мярка за митигация: Внедряването да се извършва поетапно, с ясна комуникация, ограничен пилотен обхват, оценка на риска и публично защитима обосновка за ползите и ограниченията.
Зависимост от американски/китайски технологичен доставчик	Вер. ○1 ○2 ■3 ○4 ○5	Вл. ○1 ○2 ○3 ■4 ○5	Ниво □Н ■С □В	Мярка за митигация: Да се поддържа възможност за миграция към алтернативни модели и да се следят лицензионните, правните и геополитическите ограничения. Да се избягва архитектура, зависима само от един доставчик.
Спиране на услугата при геополитически напрежения	Вер. ○1 ■2 ○3 ○4 ○5	Вл. ○1 ○2 ○3 ■4 ○5	Ниво □Н □С □В	Мярка за митигация: При вътрешна инфраструктура да се осигури локална работоспособност, архив на използваните версии, резервен план и възможност за преминаване към друг модел при промяна на условията.

Обобщена рискова оценка на групата

Топ 3 критични риска за модела: - Халюцинации и фактически грешки при използване в официални документи, правни анализи или отговори до граждани. - Нарушение на изискванията за защита на лични, служебни или чувствителни данни при неправилна употреба на модела. - Използване на AI за автоматизирани или фактически автоматизирани решения с ефект върху права, задължения или законни интереси на граждани.
--

Обща рискова оценка на модела за публичния сектор:

- ☐ Нисък риск — може да се внедри с минимални ограничения
- ☐ Среден риск — изисква допълнителни контроли и процедури
- ☐ Висок риск — изисква задълбочена оценка и специфична регулация
- ☐ **Неприемлив риск — не се препоръчва за публичния сектор**

Мотивирано становище:

1. Моделът Meta Llama е допустим за ограничена употреба в публичния сектор само при внедряване в контролирана вътрешна инфраструктура, с ясно определен обхват, отговорности и механизми за контрол. Употребата следва да е съобразена с AI Act, GDPR, NIS 2, Закона за киберсигурност, НМИМИС, ISO 27001 и ISO 42001.
2. Общата оценка е висок, но управляем риск, при наличие на човешка верификация, контрол на достъпа, логване, регистър на употребите и проследима одитна следа.
3. Моделът не следва да се използва за самостоятелни административни решения, автоматизирана оценка на граждани, обработка на класифицирана информация или окончателни правни становища. Допустим е за пилотна вътрешна употреба при административни, аналитични, обучителни и технически задачи.

Карта за приложимост

AI модел в публичната администрация — функционален анализ

v1.0 — Работен шаблон

 **Инструкция за попълване:** Попълнете всички полета за избория от вашата група AI модел. Празните редове и скалите са предназначени за ръчно попълване. Не изтривайте инструкциите в курсив.

Наименование на модела	Llama 4
Разработчик / компания	META
Версия за оценка	Версия 4
Дата на попълване	14.05.2026
Членове на групата	Ани Христова, Антония Желева, Дейвид Александров, Димитър Ходулов, Красимир Червенков, Любчо Милушев
Ръководител на групата	Ани Христова

Част 1. Обзорна матрица на приложимостта

Бърза оценка по административни функции. Прил. = Приложимост 1–5. Риск = Н/С/В.

 **Инструкция за попълване:** Попълнете всяка клетка: поставете оценка за приложимост (1=неприложимо, 5=напълно приложимо), ниво на риск (Н/С/В) и кратка бележка.

Административна функция	Обработка на документи	Правен анализ	Граждански услуги	Анализ на данни	Програмиране	Обучение
Обработка на кореспонденция	Прил.: ○1 ○2 ○3 ☒ 4 ○5 Риск: ☐ Н ☒ С ☐ В	Прил.: ○1 ☒ 2 ☐ 3 ○4 ○5 Риск: ☐ Н ☐ С ☒ В	Прил.: ○1 ○2 ☒ 3 ○4 ○5 Риск: ☐ Н ☒ С ☐ В	Прил.: ○1 ☒ 2 ☐ 3 ○4 ○5 Риск: ☐ Н ☒ С ☐ В	Прил.: ☒ 1 ○2 ○3 ○4 ○5 Риск: ☒ Н ☐ С ☐ В	Прил.: ○1 ○2 ☒ 3 ○4 ○5 Риск: ☒ Н ☐ С ☐ В
Изготвяне на доклади и анализи	Прил.: ○1 ○2 ○3 ☒ 4 ○5 Риск: ☐ Н ☒ С ☐ В	Прил.: ○1 ☒ 2 ☐ 3 ○4 ○5 Риск: ☐ Н ☐ С ☒ В	Прил.: ○1 ☒ 2 ☐ 3 ○4 ○5 Риск: ☐ Н ☒ С ☐ В	Прил.: ○1 ○2 ○3 ☒ 4 ○5 Риск: ☐ Н ☒ С ☐ В	Прил.: ○1 ☒ 2 ☐ 3 ○4 ○5 Риск: ☐ Н ☒ С ☐ В	Прил.: ○1 ○2 ○3 ☒ 4 ○5 Риск: ☒ Н ☐ С ☐ В
Тълкуване на нормативни актове	Прил.: ○1 ○2 ☒ 3 ○4 ○5 Риск: ☐ Н ☐ С ☒ В	Прил.: ○1 ○2 ☒ 3 ○4 ○5 Риск: ☐ Н ☐ С ☒ В	Прил.: ○1 ☒ 2 ☐ 3 ○4 ○5 Риск: ☐ Н ☐ С ☒ В	Прил.: ○1 ☒ 2 ☐ 3 ○4 ○5 Риск: ☐ Н ☒ С ☐ В	Прил.: ☒ 1 ○2 ☐ 3 ○4 ○5 Риск: ☒ Н ☐ С ☐ В	Прил.: ○1 ○2 ☒ 3 ○4 ○5 Риск: ☐ Н ☒ С ☐ В
Изготвяне на правни становища	Прил.: ○1 ○2 ☒ 3 ○4 ○5 Риск: ☐ Н ☒ С ☐ В	Прил.: ○1 ☒ 2 ☐ 3 ○4 ○5 Риск: ☐ Н ☐ С ☒ В	Прил.: ☒ 1 ○2 ☐ 3 ○4 ○5 Риск: ☐ Н ☒ С ☐ В	Прил.: ○1 ☒ 2 ☐ 3 ○4 ○5 Риск: ☐ Н ☒ С ☐ В	Прил.: ☒ 1 ○2 ☐ 3 ○4 ○5 Риск: ☐ Н ☒ С ☐ В	Прил.: ○1 ☒ 2 ☐ 3 ○4 ○5 Риск: ☐ Н ☒ С ☐ В

	○Н ○С 	○Н ○С 	○Н ○С 	○Н  ○В	 ○С ○В	○Н  ○В
Отговори на граждански запитвания	Прил.: ○1 ○2  ○4 ○5 Риск: ○Н  ○В	Прил.: ○1  ○3 ○4 ○5 Риск: ○Н ○С 	Прил.: ○1 ○2  ○4 ○5 Риск: ○Н ○С 	Прил.: ○1  ○3 ○4 ○5 Риск: ○Н  ○В	Прил.:  ○2 ○3 ○4 ○5 Риск:  ○С ○В	Прил.: ○1 ○2  ○4 ○5 Риск: ○Н  ○В
Чатбот за електронни услуги	Прил.: ○1  ○3 ○4 ○5 Риск: ○Н  ○В	Прил.:  ○2 ○3 ○4 ○5 Риск: ○Н ○С 	Прил.: ○1 ○2  ○4 ○5 Риск: ○Н ○С 	Прил.: ○1  ○3 ○4 ○5 Риск: ○Н  ○В	Прил.: ○1  ○3 ○4 ○5 Риск: ○Н  ○В	Прил.: ○1 ○2  ○4 ○5 Риск: ○Н  ○В
Статистически анализ	Прил.: ○1  ○3 ○4 ○5 Риск: ○Н  ○В	Прил.:  ○2 ○3 ○4 ○5 Риск: ○Н  ○В	Прил.:  ○2 ○3 ○4 ○5 Риск: ○Н  ○В	Прил.: ○1 ○2  ○4 ○5 Риск: ○Н  ○В	Прил.: ○1 ○2  ○4 ○5 Риск: ○Н  ○В	Прил.: ○1 ○2  ○4 ○5 Риск:  ○С ○В
Обработка на отчети	Прил.: ○1 ○2 ○3  ○5 Риск: ○Н  ○В	Прил.: ○1  ○3 ○4 ○5 Риск: ○Н ○С 	Прил.: ○1  ○3 ○4 ○5 Риск: ○Н  ○В	Прил.: ○1 ○2 ○3  ○5 Риск: ○Н  ○В	Прил.: ○1  ○3 ○4 ○5 Риск: ○Н  ○В	Прил.: ○1 ○2 ○3  ○5 Риск:  ○С ○В
Автоматизация на работни процеси	Прил.: ○1 ○2  ○4 ○5 Риск: ○Н  ○В	Прил.: ○1  ○3 ○4 ○5 Риск: ○Н ○С 	Прил.: ○1  ○3 ○4 ○5 Риск: ○Н ○С 	Прил.: ○1 ○2  ○4 ○5 Риск: ○Н  ○В	Прил.: ○1 ○2 ○3  ○5 Риск: ○Н  ○В	Прил.: ○1 ○2  ○4 ○5 Риск: ○Н  ○В
Разработка на вътрешни инструменти	Прил.: ○1 ○2  ○4 ○5 Риск: ○Н  ○В	Прил.: ○1  ○3 ○4 ○5 Риск: ○Н ○С 	Прил.:  ○2 ○3 ○4 ○5 Риск: ○Н ○С 	Прил.: ○1 ○2  ○4 ○5 Риск: ○Н  ○В	Прил.: ○1 ○2 ○3  ○5 Риск: ○Н  ○В	Прил.: ○1 ○2  ○4 ○5 Риск:  ○С ○В
Учебни материали и тестове	Прил.: ○1 ○2 ○3  ○5 Риск:  ○С ○В	Прил.: ○1 ○2  ○4 ○5 Риск: ○Н  ○В	Прил.: ○1 ○2  ○4 ○5 Риск: ○Н  ○В	Прил.: ○1 ○2  ○4 ○5 Риск:  ○С ○В	Прил.: ○1 ○2  ○4 ○5 Риск:  ○С ○В	Прил.: ○1 ○2 ○3  ○5 Риск:  ○С ○В
Наръчници и процедури	Прил.: ○1 ○2 ○3  ○5 Риск: ○Н  ○В	Прил.: ○1 ○2  ○4 ○5 Риск: ○Н ○С 	Прил.: ○1  ○3 ○4 ○5 Риск: ○Н  ○В	Прил.: ○1  ○3 ○4 ○5 Риск: ○Н  ○В	Прил.: ○1  ○3 ○4 ○5 Риск: ○Н  ○В	Прил.: ○1 ○2 ○3  ○5 Риск:  ○С ○В
Превод и локализация	Прил.: ○1 ○2  ○4 ○5 Риск: ○Н  ○В	Прил.: ○1  ○3 ○4 ○5 Риск: ○Н ○С 	Прил.: ○1 ○2  ○4 ○5 Риск: ○Н  ○В	Прил.: ○1  ○3 ○4 ○5 Риск: ○Н  ○В	Прил.: ○1  ○3 ○4 ○5 Риск:  ○С ○В	Прил.: ○1 ○2  ○4 ○5 Риск:  ○С ○В
Протоколи и резюмета	Прил.: ○1 ○2 ○3  ○5 Риск: ○Н  ○В	Прил.: ○1  ○3 ○4 ○5 Риск: ○Н  ○В	Прил.: ○1  ○3 ○4 ○5 Риск: ○Н  ○В	Прил.: ○1 ○2  ○4 ○5 Риск: ○Н  ○В	Прил.:  ○2 ○3 ○4 ○5 Риск:  ○С ○В	Прил.: ○1 ○2 ○3  ○5 Риск:  ○С ○В

Част 2. Детайлен анализ по случаи на употреба

За всеки сценарий оценете конкретно как избраният модел се справя.

2.1. Обработка на административни документи
Способността на модела да чете, структурира, резюмира и изготвя административни документи — решения, наредби, докладни записки, протоколи. Оценка за избрания модел: 1. Може ли моделът да обработва документи на български език с административна терминология?

2. Поддържа ли обработка на PDF/Word файлове с дълъг обем?
3. Запазва ли форматирането и структурата при генериране на изходен документ?
4. Изисква ли специализирано prompt инженерство за административен контекст?

Отговори / наблюдения:

1. Моделът може да обработва документи на български език с административна терминология, но резултатите следва да се проверяват от компетентен служител. Подходящ е за резюмиране, структуриране, редакция и изготвяне на проектни текстове.
2. Поддържа обработка на PDF и Word файлове с дълъг обем при подходяща техническа интеграция. При обемни документи е препоръчително съдържанието да се обработва на части и да се извършва проверка по първоизточник.
3. Не може да се гарантира пълно запазване на форматирането и структурата без допълнителен инструмент или предварително зададен шаблон. Моделът може да възпроизведе логическа структура, заглавия и раздели, но финалното оформление трябва да се провери от служител.
4. Изисква специализирани инструкции за административен контекст. Необходимо е ясно да се зададат целта, видът на документа, институционалният стил, допустимите източници и ограниченията.

Препоръка:

☐ Препоръчва се без ограничения

☐ Препоръчва се с ограничения (опишете):

Да се използва за проектни текстове, резюмета, вътрешни справки, протоколи и работни материали. Всички резултати, предназначени за официална употреба, следва да се проверяват от компетентен служител. Да не се въвеждат класифицирана информация, чувствителни данни или лични данни без правно основание и одобрена процедура.

☐ Не се препоръчва (обосновка):

2.2. Правен анализ и тълкуване

Способността да анализира нормативни актове, да открива противоречия, да прилага правни норми към конкретни случаи.

Оценка за избрания модел:

1. Колко точен е при тълкуване на български законодателни текстове?
2. Може ли да следи промени в нормативната база (обновени закони)?
3. Как се справя с многозначност и правна неяснота?
4. Маркира ли ясно когато излиза извън областта на своята компетентност?

Отговори / наблюдения:

1. Точността при тълкуване на български законодателни текстове е ограничена и не следва да се приема като гарантирана. Моделът може да подпомага предварителен анализ, обобщение и откриване на възможни връзки между разпоредби, но не може да замести правна експертиза.
2. Не може надеждно самостоятелно да следи промени в нормативната база, освен ако не е свързан с актуален и официален източник на нормативни актове. Без такава интеграция съществува риск да използва остаряла информация.
3. При многозначност и правна неяснота моделът може да предложи логично звучащо тълкуване, но има риск от прекалено категоричен или непълен извод. Резултатът следва да се приема само като работна хипотеза.
4. Не винаги маркира достатъчно ясно, че излиза извън областта на своята компетентност. Затова в инструкциите към модела трябва изрично да се задава да посочва несигурност, липса на данни и необходимост от проверка.

Препоръка:

☐ Препоръчва се без ограничения

☐ Препоръчва се с ограничения (опишете):

Да се използва само като помощен инструмент за предварителен анализ, обобщение или подготовка на работна чернова. Окончателното тълкуване, становище или правен извод задължително се извършва от правно звено или компетентен служител. AI резултатът не следва да бъде самостоятелно основание за административно действие.

☐ Не се препоръчва (обосновка):

2.3. Взаимодействие с граждани (чатбот / асистент)

Способността да комуникира с граждани — отговаря на въпроси, помага при подаване на заявления, обяснява административни процедури.

Оценка за избрания модел:

1. Разбира ли естествен разговорен български (включително диалекти и грешки)?
2. Дава ли ясни и разбираеми отговори без административен жаргон?
3. Как се справя с чувствителни теми (социални помощи, здраве, правосъдие)?
4. Може ли да се интегрира с портали за електронни услуги?

Отговори / наблюдения:

1. Моделът може да разбира естествен разговорен български език, включително текстове с правописни грешки или неясно формулирани въпроси. При диалектни изрази, непълни данни или объркани казуси е възможно неправилно тълкуване на намерението на гражданина.
2. Може да дава ясни и разбираеми отговори, ако е настроен да използва достъпен език и да избягва административен жаргон. Необходимо е отговорите да се базират на предварително одобрена база знания.
3. При чувствителни теми като социални помощи, здраве, правосъдие, жалби или санкции рискът е по-висок. Моделът не следва да дава индивидуални правни, медицински или социални преценки, а само обща информация и насочване.
4. Интеграция с портали за електронни услуги е възможна при подходяща техническа архитектура, API свързване, контрол на достъпа, логване и ограничаване на функциите. Не трябва да се допуска самостоятелно подаване, промяна или оценка на заявления без човешки контрол.

Препоръка:

☐ Препоръчва се без ограничения

☐ Препоръчва се с ограничения (опишете):

Да се използва само като информационен асистент за общи въпроси, насочване към услуги, разясняване на процедури и помощ при ориентиране в административни изисквания. Да не се използва за индивидуална правна преценка, автоматизирано решаване на казуси, оценка на заявления или обработка на чувствителни данни без одобрена процедура. При сложни или чувствителни случаи гражданинът следва да бъде насочван към компетентен служител.

☐ Не се препоръчва (обосновка):

2.4. Анализ на данни и отчетност

Способността да обработва структурирани данни, да генерира отчети и да идентифицира тенденции.

Оценка за избрания модел:

1. Може ли да обработва Excel/CSV данни и да генерира анализи?
2. Поддържа ли визуализация на данни (графики, диаграми)?
3. Колко надежден е при числови изчисления и статистика?
4. Може ли да се свърже с вътрешни бази данни чрез API?

Отговори / наблюдения:

1. Моделът може да подпомага обработката на Excel и CSV данни, ако данните са подадени в подходящ формат или са достъпни чрез контролирана вътрешна интеграция. Подходящ е за обобщения, описание на тенденции, подготовка на текстова част към отчети и формулиране на аналитични изводи.
2. Може да подпомага създаването на графики и диаграми чрез генериране на инструкции, код или описание на визуализации. Самото визуализиране обаче зависи от допълнителен инструмент или платформа, а не само от модела.
3. Надеждността при числови изчисления и статистика е ограничена. Моделът може да допусне грешки във формули, суми, проценти или интерпретация на резултати. Всички числови изводи трябва да се проверяват в първичните системи или от компетентен аналитичен екип.
4. Свързване с вътрешни бази данни чрез API е възможно при подходяща техническа архитектура, но трябва да се извършва при минимални права, контрол на достъпа, логване и предварителна оценка на риска. Не следва да има директен и неограничен достъп до регистри или чувствителни данни.

Препоръка:

☐ Препоръчва се без ограничения

☐ Препоръчва се с ограничения (опишете):

Да се използва за подпомагане на анализи, обобщения, проектни отчети, описание на тенденции и подготовка на справки. Числовите резултати, статистическите изводи и управленските заключения задължително се проверяват от компетентен служител или в официална информационна система. Достъпът до вътрешни бази данни да се допуска само при одобрена процедура, минимални права и проследима одитна следа.

☐ Не се препоръчва (обосновка):

2.5. Вътрешно обучение на служители

Способността да генерира учебни материали, тестове, наръчници и да поддържа интерактивно обучение.

Оценка за избрания модел:

1. Може ли да генерира тематични курсове и тестове за служители?
2. Адаптира ли съдържанието към различни нива на компетентност?
3. Поддържа ли формати за дистанционно обучение (SCORM и др.)?
4. Как се справя с чувствителни теми — антикорупция, етика, дискриминация?

Отговори / наблюдения:

1. Моделът може да генерира тематични курсове, тестове, казуси, практически задачи и примерни сценарии за обучение на служители. Подходящ е за първоначална подготовка на материали по информационна сигурност, защита на данните, административни процедури, етика и работа с AI.
2. Може да адаптира съдържанието според различни нива на компетентност, например начинаещи, напреднали служители, експерти или ръководители. Адаптацията следва да се проверява, за да не се допуснат неточни или прекалено опростени указания.
3. Не поддържа самостоятелно специализирани формати за дистанционно обучение, като SCORM, без допълнителна платформа или техническа интеграция. Може обаче да подпомага подготовката на съдържание, структура, тестови въпроси и сценарии.
4. При чувствителни теми като антикорупция, етика, дискриминация, конфликт на интереси и защита на личните данни моделът може да бъде полезен за изготвяне на примерни казуси, но съдържанието трябва задължително да се валидира от компетентно звено.

Препоръка:

☐ Препоръчва се без ограничения

☐ Препоръчва се с ограничения (опишете):

Да се използва за подготовка на проектни учебни материали, тестове, казуси, наръчници и практически сценарии. Всички материали следва да бъдат прегледани и одобрени от компетентно звено преди официално използване. Сценарият е подходящ за пилотно внедряване, тъй като рискът е по-нисък в сравнение с правен анализ, граждански услуги или автоматизирани административни решения.

☐ Не се препоръчва (обосновка):

2.6. Програмиране и автоматизация

Способността да пише код, да автоматизира процеси и да поддържа вътрешни системи.

Оценка за избрания модел:

1. Поддържа ли езиките и технологиите, използвани в административните системи?
2. Може ли да чете и документира съществуващ (legacy) код?
3. Колко надеждни са генерираните скриптове без допълнителна проверка?
4. Поддържа ли agent'ic режим за автономно изпълнение на задачи?

Отговори / наблюдения:

1. Моделът може да подпомага работа с езици и технологии, използвани в административни системи, включително писане на примерен код, скриптове, заявки, тестове и техническа документация. Приложимостта зависи от конкретната технологична среда и зададените ограничения.

- 2. Може да чете, обяснява и документира съществуващ код, включително наследени системи, когато кодът е предоставен в подходящ обем и контекст. При сложни системи е необходим преглед от разработчик или системен администратор.
- 3. Генерираните скриптове и код не са достатъчно надеждни за директно внедряване без проверка. Възможни са уязвимости, неправилна логика, несъответствие с вътрешни стандарти със съвместимостта.
- 4. Agentic режимът за автономно изпълнение на задачи носи повишен риск. Той следва да бъде ограничен до тестова среда или нискорискови действия, при минимални права, логване и задължително човешко одобрение преди изпълнение в продукционна среда.

Препоръка:

☐ Препоръчва се без ограничения

☐ Препоръчва се с ограничения (опишете):

Да се използва като помощен инструмент за разработчици, системни администратори и ИТ специалисти. Всеки генериран код, скрипт или автоматизация трябва да преминава през преглед, тестване, проверка за сигурност и одобрение преди внедряване. Да не се допуска автономно изпълнение на действия в продукционна среда без човешки контрол.

☐ Не се препоръчва (обосновка):

Част 3. Оценка на интеграционния потенциал

 **Инструкция за попълване:** Оценете техническите възможности за интеграция с административната инфраструктура.

Интеграционен аспект	Описание	Поддръжка	Бележки
API свързване	REST API, GraphQL, webhooks — интеграция с вътрешни системи	☒ Да ☐ Не ☐ Частично	Възможно е чрез вътрешно изградени API услуги и контролирана интеграция с административни системи. Необходимо е ограничаване на правата, логване и одобрена процедура за промени.
Single Sign-On	Съвместимост с Active Directory / ADFS / Keycloak	☐ Да ☐ Не ☒ Частично	Възможно е при изграждане на вътрешна платформа около модела, съвместима с Active Directory, ADFS или Keycloak. Не е функция на самия модел, а на внедрената среда.
On-premise deployment	Възможност за локален хостинг без облачна зависимост	☒ Да ☐ Не ☐ Частично	Моделът е подходящ за локално внедряване във вътрешна инфраструктура, което намалява зависимостта от публична облачна услуга и позволява по-добър контрол върху данните.
Данни в ЕС	Сървъри на европейска територия — GDPR съответствие	☒ Да ☐ Не ☐ Частично	При локално внедряване данните могат да се съхраняват в инфраструктура на организацията или в одобрена европейска среда. Необходимо е да се забрани нерегламентирано изпращане на данни към външни услуги.
Одитна следа	Логирание на всички взаимодействия за отчетност	☐ Да ☐ Не ☒ Частично	Одитната следа трябва да се реализира чрез вътрешната платформа: логване на потребител, дата, час, вход, изход, версия на модела и извършена човешка проверка.
Ролеви достъп	RBAC — различни нива на достъп за различни служители	☐ Да ☐ Не ☒ Частично	Възможно е чрез интеграция с вътрешна система за управление на достъпа. Трябва да се въведат роли за обикновен потребител, администратор, одитор, AI отговорник и контролни звена.

Многоезичност	Поддръжка на български и официални ЕС езици	☒ Да ☐ Не ☐ Частично	Моделът може да работи с български и други езици, включително официални езици на ЕС. При правни и административни текстове е необходима човешка проверка на терминологията.
Мобилен достъп	Работа от мобилни устройства (field work)	☐ Да ☐ Не ☒ Частично	Допустим само при защитена вътрешна платформа, служебни устройства, VPN, управление на мобилни устройства и ограничени права. Не се препоръчва свободен мобилен достъп.
Офлайн режим	Функционалност без постоянна интернет връзка	☐ Да ☐ Не ☒ Частично	Възможен е при локално внедряване и предварително осигурени ресурси, но функционалността зависи от конкретната архитектура, хардуер и достъп до вътрешни бази знания.
SLA и поддръжка	Гарантирано ниво на обслужване за публичния сектор	☐ Да ☐ Не ☒ Частично	Гарантираното ниво на обслужване зависи от вътрешния ИТ капацитет или от договор с доставчик на инфраструктура. Следва да се определят показатели за наличност, време за реакция и възстановяване.

Обобщена карта за приложимост

Топ 3 приложения с висок потенциал:

- Обработка, резюмиране и структуриране на административни документи, включително докладни записки, протоколи, справки, вътрешни указания и проектни текстове.
- Вътрешно обучение на служители чрез изготвяне на учебни материали, тестове, казуси, наръчници и практически сценарии по информационна сигурност, защита на данните, етика и административни процеси.
- Техническа помощ и автоматизация за ИТ звена чрез подготовка на примерен код, скриптове, документация, тестови сценарии и анализ на грешки, при задължителен преглед и тестване.

Функции, при които моделът НЕ е подходящ:

- Самостоятелно вземане на административни решения с правен или фактически ефект върху граждани.
- Автоматизирана оценка на заявления, жалби, социален статус, риск или допустимост без човешки контрол.
- Обработка на класифицирана информация, чувствителни данни или изготвяне на окончателни правни становища без изрично правно основание, одобрена процедура и експертна проверка.

Препоръчани условия за пилотно внедряване:

- Пилотното внедряване следва да се извърши само във вътрешна контролирана инфраструктура, с ограничен обхват и ясно определени роли. Първоначално моделът да се използва за нискорискови вътрешни задачи: резюмета, проектни документи, обучения, справки и техническа помощ.
- Необходимо е да се въведат контрол на достъпа, логване, регистър на употребите, одитна следа и задължителна човешка верификация на резултатите. Да се забрани използването за автоматизирани решения, класифицирана информация и индивидуални административни преценки с ефект върху граждани.

Методологични указания

Етично и прозрачно ползване на AI модели в публичната администрация

v1.0 — Работен шаблон

Инструкция за попълване: Попълнете всички полета за избория от вашата група AI модел. Празните редове и скалите са предназначени за ръчно попълване. Не изтривайте инструкциите в курсив.

Наименование на модела	Llama 4
Разработчик / компания	META
Версия за оценка	Версия 4
Дата на попълване	14.05.2026
Членове на групата	Ани Христова, Антония Желева, Дейвид Александров, Димитър Ходулов, Красимир Червенков, Любчо Милушев
Ръководител на групата	Ани Христова

Предговор и обхват

Инструкция за попълване: Попълнете идентификационните данни на вашата организация и дефинирайте обхвата на документа.

Наименование на организацията	Публична администрация в Република България
Административно ниво	Централна администрация
AI модел(и), обхванати от документа	Meta Llama 4, разглеждан във вариант на внедряване във вътрешна контролирана инфраструктура
Версия на документа	v1.0, работен проект
Дата на приемане	Не е приложимо, документът е в проектен вид
Следваща планирана ревизия	До 12 месеца след приемане или при промяна на модела, нормативната рамка, техническата архитектура или установеното ниво на риск
Одобен от	Не е приложимо, документът не е официално приет
Отговорен за прилагането	Определен AI отговорник, съвместно с ИТ звено, звено по информационна сигурност, правно звено, длъжностно лице по защита на данните и вътрешен одит

Чл. 1. Предмет и обхват

Опишете за какво се отнася документът — кои AI системи, кои структурни звена и кои видове задачи обхваща.

- Настоящият документ определя правила за етично, прозрачно и контролирано използване на AI модела Meta Llama 4 във вътрешна инфраструктура на публичната администрация.
- Указанията се прилагат за служители и структурни звена, които използват модела за административни, аналитични, обучителни, технически и помощни дейности.
- Моделът се използва единствено като помощен инструмент при подготовка на проектни документи, резюмета, справки, анализи, обучителни материали и техническа помощ.
- Не се допуска използването му за самостоятелно вземане на административни решения, автоматизирана оценка на граждани, обработка на класифицирана информация или окончателни правни становища без човешка проверка и одобрена процедура.

Чл. 2. Правна и регулаторна рамка

Посочете нормативните актове, на базата на които се издават указанията (EU AI Act, GDPR, ЗЕУ, вътрешни наредби).

- Настоящите указания се издават и прилагат в съответствие с Регламент (ЕС) 2024/1689 относно изкуствения интелект, Регламент (ЕС) 2016/679, Директива (ЕС) 2022/2555, Закона за киберсигурност, Закона за електронното управление и Наредбата за минималните изисквания за мрежова и информационна сигурност.
- Прилагат се и действащите вътрешни актове на организацията относно информационната сигурност, защитата на личните данни, управлението на достъпа, документооборота и управлението на инциденти.
- Документът отчита приложимите добри практики по ISO/IEC 27001 и ISO/IEC 42001.

Раздел I. Принципи за етично и прозрачно ползване

Определете основните принципи, на които се базира политиката. Попълнете конкретните мерки за всеки.

Основни принципи	
1 Законност	AI системите се използват само в рамките на действащото законодателство. Всяко решение, подпомогнато от AI, спазва приложимите правни норми. Конкретни мерки в нашата организация: AI моделът се използва само за предварително одобрени дейности, съобразени с приложимата нормативна рамка и вътрешните правила. Нова употреба се допуска след оценка за правно съответствие, защита на данните и информационна сигурност.
2 Прозрачност	Гражданите и служителите имат право да знаят кога AI участва в процесите. Резултатите от AI трябва да са обяснени разбираемо. Конкретни мерки в нашата организация: Употребата на AI се документира в регистър на приложенията, с посочване на цел, обхват, отговорно звено и вид обработвани данни. При участие на AI в процеси, засягащи граждани, това се обозначава по ясен и разбираем начин, когато е приложимо.
3 Отчетност	Ясна верига на отговорност. Служителите носят отговорност за решенията, взети с AI помощ — AI не замества човешката отговорност. Конкретни мерки в нашата организация: Отговорността за крайния резултат остава при компетентния служител или орган. За официална употреба на AI резултат се осигурява проследимост чрез логове, човешка проверка и документиране на извършената верификация.
4 Справедливост	AI инструментите не дискриминират. Резултатите се проверяват редовно за пристрастия (bias) спрямо различни групи граждани. Конкретни мерки в нашата организация: AI моделът не се използва за самостоятелна оценка на граждани, заявления, допустимост, риск или приоритет. Периодично се извършва проверка за пристрастия, изкривявания и неравнопоставено третиране.
5 Поверителност	Личните данни на гражданите се обработват само при наличие на правно основание и при спазване на GDPR. Конкретни мерки в нашата организация: Лични данни се обработват само при наличие на правно основание, конкретна цел и минимален необходим обем. Не се допуска въвеждане на класифицирана информация или чувствителни данни без изрично одобрена процедура и технически мерки за защита.
6 Надеждност	AI системите се използват само за задачи, при които точността е потвърдена. Резултатите подлежат на верификация. Конкретни мерки в нашата организация: AI моделът се използва само за задачи, при които са оценени приложимостта, точността и рискът. Резултатите, предназначени за официална употреба, подлежат на задължителна човешка верификация.
7 Пропорционалност	Нивото на автоматизация съответства на риска от грешки. При по-висок риск — по-задължителен човешки контрол. Конкретни мерки в нашата организация: Степента на използване на AI се определя според риска на конкретната дейност. При задачи със среден или висок риск се прилагат допълнителни ограничения, одобрение и човешки контрол.

8

Обяснимост

При поискване от гражданин или орган, организацията може да обясни как е бил използван AI при дадено решение.

Конкретни мерки в нашата организация:

При необходимост организацията трябва да може да обясни целта на използване на AI, входните данни, генерирания резултат и извършената човешка проверка. AI резултатът не замества фактическата и правната обосновка на административното действие.

Раздел II. Разрешени и забранени употреби

Определете за всяка категория задачи дали е разрешена, допустима с условия или забранена.

 **Инструкция за попълване:** За всяка задача: поставете отметка в съответната колона и опишете условията или ограниченията.

Категория задача	Статус	Условия / Ограничения / Процедура
Административни документи		
Изготвяне на проекти за решения	☐ Разрешено ☐ С условия ☐ Забранено	Условия / ограничения / процедура: Допуска се само за подготовка на работна чернова. Крайният текст се проверява и одобрява от компетентен служител или звено.
Резюмиране на входяща поща	☐ Разрешено ☐ С условия ☐ Забранено	Условия / ограничения / процедура: Допуска се за вътрешно подпомагане, без въвеждане на класифицирана информация и чувствителни данни. При лични данни се прилага минимизиране и правно основание.
Превод на документи	☐ Разрешено ☐ С условия ☐ Забранено	Условия / ограничения / процедура: Допуска се за работен превод. Официални, правни и нормативни преводи подлежат на задължителна човешка редакция и проверка на терминологията.
Генериране на официални писма	☐ Разрешено ☐ С условия ☐ Забранено	Условия / ограничения / процедура: Допуска се само за проект на текст. Окончателното писмо се преглежда, редактира и одобрява от отговорен служител преди изпращане.
Правни и нормативни въпроси		
Анализ на нормативни текстове	☐ Разрешено ☐ С условия ☐ Забранено	Условия / ограничения / процедура: Допуска се за предварителен анализ, сравнение и обобщение. Правните изводи се потвърждават от правно звено или компетентен служител.
Проверка за вътрешни противоречия	☐ Разрешено ☐ С условия ☐ Забранено	Условия / ограничения / процедура: Допуска се като помощен контролен механизъм. Установените противоречия се проверяват по първоизточник и се валидират от компетентно звено.
Изготвяне на правни становища	☐ Разрешено ☐ С условия ☐ Забранено	Условия / ограничения / процедура: Допуска се само за работна чернова или структуриране на аргументи. Окончателното становище се изготвя и подписва от правно звено или оправомощено лице.
Тълкуване на разпоредби към граждани	☐ Разрешено ☐ С условия ☐ Забранено	Условия / ограничения / процедура: Допуска се само за обща информационна насока. Не се допуска индивидуална правна преценка или окончателно тълкуване без компетентен служител.
Вземане на решения с ефект върху граждани		
Преценка за социални помощи	☐ Разрешено ☐ С условия ☐ Забранено	Условия / ограничения / процедура: Не се допуска AI моделът да извършва самостоятелна преценка за права, допустимост, размер или отказ на социална помощ.
Оценка на проекти и заявления	☐ Разрешено ☐ С условия ☐ Забранено	Условия / ограничения / процедура: Не се допуска самостоятелно класиране, оценяване, одобряване или отхвърляне на проекти и заявления чрез AI.

Наказателни и принудителни мерки	☐ Разрешено ☐ С условия ☐ Забранено	<i>Условия / ограничения / процедура:</i> Не се допуска използване на AI за самостоятелно определяне, предлагане или мотивиране на наказателни, санкционни или принудителни мерки.
Автоматизирани решения без човешки контрол	☐ Разрешено ☐ С условия ☐ Забранено	<i>Условия / ограничения / процедура:</i> Не се допуска вземане на решения с правен или фактически ефект върху граждани без човешки контрол, проверка и административна отговорност.
Комуникация с граждани		
Чатбот за информационни въпроси	☐ Разрешено ☐ С условия ☐ Забранено	<i>Условия / ограничения / процедура:</i> Допуска се само за обща информация, насочване към услуги и разясняване на процедури въз основа на одобрена база знания.
Проекти за отговори на жалби	☐ Разрешено ☐ С условия ☐ Забранено	<i>Условия / ограничения / процедура:</i> Допуска се само за подготовка на проект. Всеки отговор по жалба се проверява от компетентен служител и при необходимост от правно звено.
Публикации в социални медии	☐ Разрешено ☐ С условия ☐ Забранено	<i>Условия / ограничения / процедура:</i> Допуска се за проект на текст. Публикуване се извършва само след редакция и одобрение от отговорното звено за публична комуникация.
Кризисна комуникация	☐ Разрешено ☐ С условия ☐ Забранено	<i>Условия / ограничения / процедура:</i> Допуска се само като помощен инструмент за проект на съобщение. Окончателната комуникация се одобрява от ръководство или определено кризисно звено
Вътрешно-административни функции		
Обучение на служители	☐ Разрешено ☐ С условия ☐ Забранено	<i>Условия / ограничения / процедура:</i> Допуска се за подготовка на учебни материали, тестове, казуси и наръчници. Материалите се преглеждат преди официално използване.
Анализ на данни и отчети	☐ Разрешено ☐ С условия ☐ Забранено	<i>Условия / ограничения / процедура:</i> Допуска се за обобщения, тенденции и проектни отчети. Числовите резултати и управленските изводи се проверяват в официални системи или от компетентен служител.
Автоматизация на повтарящи се задачи	☐ Разрешено ☐ С условия ☐ Забранено	<i>Условия / ограничения / процедура:</i> Допуска се само за нискорискови вътрешни процеси, при минимални права, логване, тестване и човешко одобрение преди внедряване.
Обработка на класифицирана информация	☐ Разрешено ☐ С условия ☐ Забранено	<i>Условия / ограничения / процедура:</i> Не се допуска въвеждане, обработване, обобщаване или генериране на съдържание, съдържащо класифицирана информация.

Раздел III. Процедури за одобрение и верификация

Чл. 10. Задължителна верификация

Опишете кои задачи изискват задължителна проверка от служител преди използване на AI изхода.

- Задължителна проверка от компетентен служител се извършва преди използване на AI резултат в официални документи, правни становища, отговори до граждани, публични съобщения, доклади, анализи и управленски справки.
- Верификация се изисква и при обработка на лични данни, служебна информация, жалби, заявления, чувствителни теми, нормативни текстове, финансови данни, статистически изводи и генериран код или скриптове.
- AI резултатът се приема само като помощен материал. Крайният текст, извод или действие се одобрява от отговорен служител или компетентно звено, като се документира извършената проверка и основанието за използване.

Чл. 11. Процедура за одобрение на нови употреби

Как служителят може да заяви одобрение за нов вид употреба на AI, непредвиден в настоящите указания?

- 1. Нова употреба на AI, която не е предвидена в настоящите указания, се заявява писмено от служителя или ръководителя на съответното звено до определения AI отговорник.
- 2. Заявката следва да съдържа описание на целта, вида на задачата, използваните данни, очаквания резултат, засегнатите лица и предварителна оценка на риска.
- 3. Одобрение се дава след съгласуване с компетентните звена по информационна сигурност, защита на личните данни, правно съответствие и ИТ. До получаване на одобрение новата употреба не се допуска.

Чл. 12. Задължително разкриване към граждани

В кои случаи гражданинът трябва да бъде информиран, че AI е участвал в обработката на неговото искане?

- 1. Гражданинът се информира, когато AI е използван при обработка на негово запитване, заявление, жалба или друг административен казус и резултатът може да има значение за предоставената информация, насока или последващо административно действие.
- 2. Разкриването е задължително при използване на чатбот, автоматизирано генериран проект на отговор, анализ на подадени данни или друг процес, при който AI подпомага обработката на конкретно искане.
- 3. Информацията към гражданина следва да бъде ясна и разбираема, като се посочва, че AI е използван само като помощен инструмент и че окончателната преценка се извършва от компетентен служител или орган.

Раздел IV. Мониторинг, одит и отчетност

Минимални изисквания за одитна следа (отбележете приложимите):

- ☐ Всяко взаимодействие с AI системата се записва с дата, час и потребител
- ☐ Пазят се логове на AI изходите, използвани в официални документи
- ☐ При автоматизирани решения се съхранява AI входа и изхода

(Бележка: За настоящия обхват не се прилага, тъй като автоматизирани решения без човешки контрол са забранени.)

- ☐ Провежда се периодичен преглед на качеството на AI изходите (мин. 1х/тримесечие)
- ☐ Инцидентите, свързани с AI грешки, се докладват на определено звено
- ☐ Годишен отчет за употребата на AI се изготвя и публикува

Чл. 15. KPI за мониторинг на AI употребата

Определете конкретни измерими показатели: брой взаимодействия, % задачи с верификация, % открити грешки и др.

- 1. Мониторингът на AI употребата се извършва чрез показатели като брой взаимодействия, брой активни потребители, процент на резултатите с човешка верификация, процент на установени грешки, брой инциденти и брой одобрени или отказани нови употреби.
- 2. Показателите се преглеждат най-малко веднъж на тримесечие от AI отговорника съвместно с компетентните звена по ИТ, информационна сигурност, правно съответствие и защита на личните данни.

Раздел V. Отговорности и управление

Роля	Отговорност	Конкретно лице / звено
Ръководител на организацията	Стратегическо одобрение и крайна отговорност за AI политиката Одобрява политиката за използване на AI, определя стратегическия обхват и носи крайна управленска отговорност.	Конкретно лице / звено: Ръководител на организацията или оправомощено лице.
AI отговорник (AI Officer)	Координация, обучение, мониторинг, докладване на инциденти Координира прилагането на указанията, поддържа регистър на AI употребите, организира обучения, следи KPI и координира докладването на инциденти.	Конкретно лице / звено: Определен AI отговорник / AI Officer
Служители, ползващи AI	Спазване на указанията, верификация на изходите, докладване на проблеми	Конкретно лице / звено: Всички служители с предоставен достъп до AI системата

	Използват AI само в рамките на разрешения обхват, проверяват резултатите преди употреба и докладват грешки, рискове или инциденти.	
ИТ отдел / системен администратор	Технически достъп, интеграция, сигурност, логове Осигурява техническа поддръжка, управление на достъпа, интеграция, логване, защита на инфраструктурата и технически мониторинг.	Конкретно лице / звено: ИТ звено / системни администратори
Правно звено	Преглед на нови употреби, съответствие с нормативната рамка Извършва правна оценка на нови употреби, проверява съответствието с нормативната рамка и участва при оценка на високорискови сценарии.	Конкретно лице / звено: Правно звено
ДЛЗД (длъжностно лице по защита на данни)	GDPR съответствие при обработка на лични данни Оценява съответствието с изискванията за защита на личните данни, участва при обработки с лични данни и дава становища при повишен риск.	Конкретно лице / звено: Длъжностно лице по защита на данните
Вътрешен одит	Периодичен независим преглед на AI употребата Извършва периодичен независим преглед на AI употребата, одитната следа, спазването на правилата и ефективността на контролните мерки.	Конкретно лице / звено: Звено за вътрешен одит

Раздел VI. Обучение и изграждане на капацитет

Чл. 20. Задължително обучение			
Посочете кои служители трябва да преминат обучение, какъв е минималният обем и как се документира.			
1. Задължително обучение преминават всички служители с достъп до AI системата, както и лицата, участващи в нейното администриране, контрол, мониторинг, одит и одобряване на нови употреби. 2. Обучението обхваща допустими и забранени употреби, защита на личните данни, информационна сигурност, човешка верификация, разпознаване на грешки и докладване на инциденти. 3. Провеждането на обучението се документира с програма, дата, списък на участниците и доказателство за преминаване, включително резултат от проверка на знанията, когато е приложимо.			
Модул на обучение	Целева група	Форма и обем	Периодичност
Основи на AI и рискове	Всички служители с достъп до AI системата	Въвеждащо обучение, минимум 2 учебни часа	При първоначален достъп и веднъж годишно
Работа с конкретния модел	Служители, които използват Meta Llama 4 в служебната дейност	Практическо обучение, минимум 2 учебни часа	При първоначален достъп и при съществена промяна на модела
Етика и нормативна рамка	Всички потребители на AI, ръководители и контролни звена	Обучение или инструктаж, минимум 1 учебен час	Веднъж годишно или при нормативна промяна
Разпознаване на AI грешки	Служители, използващи AI резултати в документи, справки и анализи	Практически казуси, минимум 1 учебен час	Веднъж годишно
Защита на личните данни	Всички потребители на AI и служители, работещи с лични данни	Обучение или инструктаж, минимум 1 учебен час	Веднъж годишно или при промяна в правилата

Раздел VII. Актуализация и валидност

Чл. 24. Условия за задължителна ревизия

- Опишете кога документът трябва задължително да се преразгледа: нова версия на модела, регулаторна промяна, инцидент, изтичане на срок.
- 1. Настоящият документ подлежи на задължителна ревизия при промяна на AI модела, версията, техническата архитектура, обхвата на употреба или приложимата нормативна рамка.
 - 2. Ревизия се извършва и при установен съществен риск, инцидент, системни грешки, одитна констатация или предписание от компетентен контролен орган.
 - 3. Документът се преглежда най-малко веднъж годишно, независимо от настъпването на промени.

Чл. 25. История на промените

Документирайте всички версии на документа с дата, автор и описание на промяната.

- 1. Всички изменения на документа се документират чрез посочване на версия, дата, описание на промяната и лице или орган, одобрил изменението.
- 2. Историята на промените се поддържа с цел проследимост, отчетност и доказване на актуалността на правилата при вътрешен или външен одит.

Версия	Дата	Описание на промяната	Одобрил
v1.0	14.05.2026	Първоначална работна версия на методологичните указания	Ръководител на групата / оправомощено лице
v1.1	При необходимост	Актуализация при промяна на модела, нормативната рамка или оценката на риска	Оправомощено лице / компетентен орган
V2.0	При съществена промяна	Основна ревизия на документа при нова версия на модела или разширяване на обхвата	Ръководител на организацията / оправомощено лице

ΓΡΥΠΑ "ΜΑΑΤ"

**Anthropic Claude
(Opus 4.7/ Sonnet 4.6/
Haiku 4.5)**



SWOT Анализ

Оценка на фундаментален AI модел за публичната администрация

Anthropic Claude Opus 4.7

Инструкция за попълване: Попълнете всички полета за избория от вашата група AI модел. Празните редове и скалите са предназначени за ръчно попълване. Не изтривайте инструкциите в курсив.

Наименование на модела	Claude Opus 4.7
Разработчик / компания	Anthropic
Версия за оценка	claude-opus-4-7 (април 2026)
Дата на попълване	14.05.2026 г.
Членове на групата	Мариела, Теодора, Милен, Жасмина, Анелия, Денис
Ръководител на групата	Милен

A. Обща приложимост и функционален обхват

Инструкция за попълване: За всеки критерий: поставете кръгче на избраната оценка (1=много слабо, 5=отлично) и запишете кратка обосновка.

Критерий		
A1. Приложимост в публичния сектор		
Обработка на документи <i>доклади, решения, наредби</i>	Оценка 1–5 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☒ 5	Обосновка / бележки: Обработка дълги текстове с висока точност, структурира доклади и решения автоматично. Отличен в обработката на дълги документи (до 1M токена), идеален за големи административни преписки.
Правен анализ <i>тълкуване на нормативни актове</i>	Оценка 1–5 ☐ 1 ☐ 2 ☐ 3 ☒ 4 ☐ 5	Обосновка / бележки: Има дълбоко разбиране и извършва анализ на документи, обработва нормативни актове, но може да допуска грешки при специфични науки, тълкувания и практиката от правото. Много силен в разсъжденията, като помага за анализи, но изисква проверка от човек поради риск от "халюцинации" със специфичната правна терминология.
Граждански услуги <i>чатбот, Q&A, помощ при заявления</i>	Оценка 1–5 ☐ 1 ☐ 2 ☐ 3 ☒ 4 ☐ 5	Обосновка / бележки: Добра база за автоматизирани въпроси и отговори, изисква допълнителна настройка за специфични процедури. Поддържа български, но не е специализиран за него. Скъп е за мащабиране.
Вътрешна комуникация <i>резюмета, протоколи, кореспонденция</i>	Оценка 1–5 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☒ 5	Обосновка / бележки: Отлична способност да обобщава и да следва инструкции, което го прави идеален за резюмета, протоколи, кореспонденция и вътрешни доклади.

SWOT Анализ — AI Модел Claude Opus 4.7.

Анализ на данни <i>отчети, статистика, мониторинг</i>	Оценка 1–5 ☐ 1 ☐ 2 ☐ 3 ☒ 4 ☐ 5	Обосновка / бележки: Добре анализира структурирани данни и изготвя отчети. Анализира диаграми и таблици от отчети, но няма вградена поддръжка за работа с големи бази данни.
Превод <i>многоезична комуникация, ЕС документи</i>	Оценка 1–5 ☐ 1 ☐ 2 ☐ 3 ☒ 4 ☐ 5	Обосновка / бележки: Висококачествен превод на официалните езици в Европейския съюз. За български превод на специфична терминология е надежден, но не е перфектен.
Програмиране / автоматизация <i>вътрешни системи, скриптове</i>	Оценка 1–5 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☒ 5	Обосновка / бележки: Един от най-силните модели за програмиране (SWE-bench 78.2%). Може да автоматизира рутинни IT задачи в администрацията.
Обучение на служители <i>учебни материали, тестове</i>	Оценка 1–5 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☒ 5	Обосновка / бележки: Силен е при структуриране на учебно съдържание / може да адаптира учебно съдържание. Може да генерира качествени учебни сценарии и тестове. Успешно адаптира сложна терминология за всяка аудитория.
A2. Функции на платформата		
Текстово генериране <i>качество и прецизност</i>	Оценка 1–5 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☒ 5	Обосновка / бележки: Моделът заема лидерска позиция по качество на генерирания текст благодарение на своята прецизност, последователност и нисък процент на грешки. Притежава ясен, точен и издържан професионален стил.
Reasoning / разсъждение <i>логически вериги, многостъпкови задачи</i>	Оценка 1–5 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☒ 5	Обосновка / бележки: Притежава развити способности за анализ, разбиране на комплексни задачи и дългосрочно поддържане на контекст. Отличен избор за целите на стратегическия анализ.
Code writing <i>генериране и анализ на код</i>	Оценка 1–5 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☒ 5	Обосновка / бележки: Това е най-развитото функционално направление на модела. Той извършва последваща верификация, преди да визуализира генерирания код, и демонстрира изключителни резултати при анализа на данни и автоматизацията на процеси. Заедно с GPT-5.4, моделът заема лидерска позиция в индустрията.
Анализ на изображения <i>vision, диаграми, скрийншоти</i>	Оценка 1–5 ☐ 1 ☐ 2 ☐ 3 ☒ 4 ☐ 5	Обосновка / бележки: Поддържа висока разделителна способност при обработката на изображения, което го прави изключително ефективен при разчитането и анализа на сканирани документи, чертежи и схеми.
Аудио / глас <i>говорим вход/изход</i>	Оценка 1–5 ☐ 1 ☐ 2 ☒ 3 ☐ 4 ☐ 5	Обосновка / бележки: Моделът не разполага с поддръжка за гласово въвеждане, което налага използването на външни инструменти за аудио транскрипция. В уеб базираната си версия системата пренасочва заявките за директен говор и аудио обработка към модела Sonnet 4.6.
Видео разбиране <i>анализ на видеосъдържание</i>	Оценка 1–5 ☐ 1 ☐ 2 ☒ 3 ☐ 4 ☐ 5	Обосновка / бележки: Моделът не поддържа обработка и анализ на видео файлове. Налице е и липса на функционалност за проследяване на движението (motion tracking).

SWOT Анализ — AI Модел Claude Opus 4.7.

Дълги документи <i>обработка на обемни файлове</i>	Оценка 1–5 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☒ 5	Обосновка / бележки: Разполага с контекстен прозорец до 1 000 000 токена, което осигурява възможност за едновременна обработка и анализ на мащабни масиви от документи.
Agentic режим <i>автономни многостъпкови задачи</i>	Оценка 1–5 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☒ 5	Обосновка / бележки: Водещ в индустрията модел, оптимизиран за поддръжка на автономни процеси и автоматично управление на сложни последователности от задачи. Възможността за автономно действие му позволява ефективно да извършва крос-системни операции, като автоматизирано попълване на данни и формуляри.
API / системна интеграция <i>свързване с вътрешни системи</i>	Оценка 1–5 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☒ 5	Обосновка / бележки: Моделът покрива всички нива на оперативна съвместимост, като гарантира пълно съответствие с изискванията на Закона за защита на класифицираната информация (ЗЗКИ).
Plugin екосистема <i>налични разширения и конектори</i>	Оценка 1–5 ☐ 1 ☐ 2 ☐ 3 ☒ 4 ☐ 5	Обосновка / бележки: Моделът поддържа отворения стандарт Model Context Protocol (MCP), разработен от Anthropic, за защитена двустранна връзка с външни бази данни и инструменти. Към момента обаче се отчита липса на тясно специализирани административни плъгини за управление.
Desktop приложение <i>нативна desktop среда</i>	Оценка 1–5 ☐ 1 ☐ 2 ☐ 3 ☒ 4 ☐ 5	Обосновка / бележки: Налице е самостоятелно десктоп приложение за операционните системи Windows и macOS. Моделът предлага добре развита екосистема с пълноценна поддръжка и директен достъп до локални файлове на потребителя.
Mobile приложение <i>iOS / Android достъп</i>	Оценка 1–5 ☐ 1 ☐ 2 ☐ 3 ☒ 4 ☐ 5	Обосновка / бележки: Налични са софтуерни приложения за платформите iOS и Android, предоставящи достъп до пълната функционалност на модела. Въпреки това мобилната екосистема се оценява като недостатъчно зряла, а приложенията демонстрират висока зависимост от външната софтуерна среда.
Cowork / колаборация <i>екипна работа в реално време</i>	Оценка 1–5 ☐ 1 ☐ 2 ☒ 3 ☐ 4 ☐ 5	Обосновка / бележки: Моделът поддържа функционалности за споделяне на сесии (разговори) и съвместна работа по общи документи. Към момента обаче липсва възможност за паралелно редактиране в реално време (real-time co-authoring).
Търсене в интернет <i>актуална информация от мрежата</i>	Оценка 1–5 ☐ 1 ☐ 2 ☐ 3 ☒ 4 ☐ 5	Обосновка / бележки: Моделът поддържа уеб търсене в реално време, което го прави изключително полезен за извличане на актуална нормативна и регулаторна информация. Той притежава силни способности за анализ и структуриран синтез на данни. Критичен недостатък: Наблюдават се пропуски при формирането на логически изводи, както и склонност към смесване на контекста от различни източници.
Fine-tuning <i>настройка за специфична среда</i>	Оценка 1–5 ☐ 1 ☐ 2 ☒ 3 ☐ 4 ☐ 5	Обосновка / бележки: Процесите по адаптация към специфичния административен стил и интеграцията на вътрешна база данни от знания липсват като готови публични функционалности. Системната настройка и персонализацията се осъществяват изцяло чрез инженеринг или чрез изграждане на RAG архитектура (Retrieval-Augmented Generation).

B. SWOT Матрица

Инструкция за попълване: Запишете поне 4 наблюдения в клетка. Базирайте се на попълнените оценки от Раздел A.

S STRENGTHS — Силни страни

- 1. Високотехнологичен аналитичен капацитет (Reasoning):** Лидерско ниво на логическо мислене, оптимизирано за разрешаване на комплексни административни казуси и прецизни правни анализи.
- 2. Мащабен контекстен прозорец:** Капацитет от 1 000 000 токена, позволяващ едновременна обработка и структуриране на цели информационни масиви, фирмени профили или административни досиета.
- 3. Автономни агентни функционалности:** Поддържа на агентни работни потоци за автоматизация на многостъпкови процеси и крос-системни операции в множество независими платформи.
- 4. Усъвършенствана софтуерна разработка:** Високи възможности за генериране и оптимизация на програмен код, приложими за автоматизация на вътрешната ИТ инфраструктура.
- 5. Интегрирани деонтологични филтри:** Високо ниво на вградена безопасност и етика, гарантиращи минимизиране на риска от генериране на токсично или зловредно съдържание.
- 6. Стратегическа приложимост:** Оптимално решение за вътрешноорганизационни аналитични процеси, изискващи експертна обработка на данни (бек-офис), пред масово обслужване на клиенти (фронт-офис)

Допълнителни бележки:

W WEAKNESSES — Слаби страни

- 1. Риск от деформация на факти:** Наличие на вероятност от фактически неточности (халюцинации), което налага задължителна последваща проверка на критични данни.
- 2. Зависимост от външни информационни източници:** Невъзможност за автономна верификация на актуална нормативна уредба или новинарски поток без интеграция на външни RAG или уеб инструменти.
- 3. Ограничен мултимодален капацитет:** Недостатъчно развити функционалности за нативна обработка на аудио съдържание и пълен дефицит на инструменти за видео анализ.
- 4. Висока себестойност на операциите:** Стойност от \$25 за 1 000 000 токена (приблизително 7500 думи), което прави модела икономически неефективен за масови потребителски чатботове (фронт-офис).
- 5. Стилистични дефицити при локализация:** Пониско качество на генерирания текст на български език по отношение на специфична административна и правна терминология в сравнение с представянето му на английски език.
- 6. Липса на архитектурна персонализация (Fine-tuning):** Отсъствие на възможност за дълбоко дообучение на модела върху специфични вътрешноорганизационни масиви от документи.

Допълнителни бележки:

O OPPORTUNITIES — Възможности

- 1. Катализиране на дигиталната трансформация:** Възможност за ускоряване на процесите по дигитализация и драстично повишаване на общата административна ефективност.
- 2. Проектиране на когнитивни асистенти:** Изграждане на интелигентни административни архитектури и дигитални асистенти от ново поколение за оптимизация на вътрешните процеси.
- 3. Модернизация на аналитичния капацитет:** Надграждане на аналитичния капацитет и експертния потенциал в рамките на държавните и общинските ведомства.
- 4. Оптимизация на трансфера на знания:** Ефективна софтуерна подкрепа при вътрешното обучение на служители, управлението на интелектуалния капитал и автоматизираното генериране на специализирани учебни материали.
- 5. Делегиране на рутинни експертни анализи:** Освобождаване на капацитет при висококвалифицирани кадри (юристи, инженери,

T THREATS — Заплахи

- 1. Регулаторен риск по линия на AI Act:** Потенциално налагане на строги ограничения или забрани от страна на ЕС, в случай че технологията бъде категоризирана като модел на ИИ с общо предназначение (GPAI) със системен риск.
- 2. Технологична обвързаност (Vendor Lock-in):** Риск от прекалена оперативна и стратегическа зависимост от Anthropic.
- 3. Репутационен и оперативен риск от халюцинации:** Вероятност от генериране на фактически неточности и подвеждащи аналитични резултати, застрашаващи правилността на административните решения.
- 4. Инфраструктурна динамика и амортизация:** Изключително бързо променяща се технологична среда, налагаща непрекъснати последващи одити, нови капиталови инвестиции и постоянна системна адаптация.
- 5. Архитектурно несъответствие с GDPR:** Висок риск от нерегламентирана обработка на лични

проектанти) чрез автоматизация на повтарящи се текстови и правни анализи. 6. Потенциал за национална езикова локализация: Генериране на база данни и методология за бъдещо разработване на суверенен (национален) AI модел на български език, адаптиран към местното законодателство. <i>Допълнителни бележки:</i>	данни при директна интеграция с API интерфейса на Anthropic, извън изолирана и сертифицирана облачна среда на ЕС. 6. Икономическа нецелесъобразност (Ефективност на разходите): Финансов риск от преразход, тъй като по-икономични алтернативни модели (напр. от фамилиите GPT или Gemini) могат да покрият до 90% от базовите нужди на организацията при многократно по-ниски разходи. <i>Допълнителни бележки:</i> Основната стратегическа заплаха се корени в опасността от реализиране на мащабни финансови инвестиции, последвани от административно блокиране на системата от регулаторните органи поради системно несъответствие с изискванията на GDPR и ЗЗЛД.

C. Разширен SWOT по специфични теми

Инструкция за попълване: Попълнете текстово наблюденията си за всяка тема. Използвайте конкретни примери.

Цена и лицензиране Наблюдения: <i>(Модел на плащане, достъпност за малки общини, API разходи)</i> Моделът се характеризира с висока себестойност на операциите, което го прави икономически неефективен за масово административно внедряване. За нуждите на по-малки общински структури – например за автоматизирана обработка на граждански жалби и сигнали – текущите API разходи са финансово непосилни и неоправдани спрямо капацитета на по-леки и нискобюджетни алтернативи. Препоръка за публичния сектор: Целесъобразно е прилагането на хибриден подход. Моделът от най-висок клас следва да бъде изолиран единствено за тясно специфични задачи с висока добавена стойност (като стратегически анализ на обемни досиета или комплексна софтуерна разработка). За регулярни, масови и повтарящи се административни операции (фронт-офис) се препоръчва интеграцията на значително по-икономичен модел с по-малък мащаб.
Сигурност на данните Наблюдения: <i>(Обработка на лични данни, GDPR, съхранение в облака)</i> Генерираната информация не се подлага на автоматизирано профилиране или наблюдение от страна на Anthropic. Въпреки това архитектурата на модела създава предпоставки за компрометиране на принципа за ограничаване на достъпа до равнище „необходимост да се знае“. Допълнителен критичен фактор за съответствието с GDPR е обстоятелството, че данните се трансферират и обработват на сървърни мощности, позиционирани извън Европейското икономическо пространство (в САЩ). Препоръка за публичния сектор: Допуска се ограничено внедряване на технологията единствено при изпълнение на административни задачи с ниско до средно ниско ниво на критичност и конфиденциалност. Всички интеграции трябва да се извършват при стриктно софтуерно съблюдаване на принципа „необходимост да се знае“, като се

избягва подаването на чувствителни лични данни или класифицирана информация към външния API интерфейс.

Езикова поддръжка

Наблюдения: *(Качество на български език, юридическа терминология)*

Моделът демонстрира високо ниво на езикова култура на български език при провеждането на обща комуникация. Въпреки това се отчитат системни затруднения и спад в прецизността при обработката на специфична юридическа, административна и тясно техническа терминология.

Препоръка за публичния сектор:

Задължително е интегрирането на модела в RAG архитектура (*Retrieval-Augmented Generation*), която динамично да му подава легитимните дефиниции и точните термини от действащите закони и наредби. Налага се и въвеждането на задължителен протокол за последващ човешки контрол - всеки генериран официален документ трябва да бъде проверен и одобрен от експерт преди финално извеждане.

Vendor lock-in

Наблюдения: *(Зависимост от доставчик, алтернативи, преносимост)*

Идентифицира се умерен риск от технологична обвързаност. Директната интеграция с API на Anthropic създава системна зависимост на програмно ниво. Рискът обаче се балансира от факта, че Claude Opus 4.7 е достъпен през водещите облачни платформи (AWS Bedrock, Google Cloud Vertex AI, Microsoft Azure). Това осигурява висока преносимост и гъвкавост за превключване между различни облачни инфраструктури без промяна в основния доставчик на модела.

Препоръка за публичния сектор:

Стратегически целесъобразно е избягването на монолитна зависимост от единичен изкуствен интелект. Препоръчва се изграждането на независима софтуерна архитектура, поддържаща динамично маршрутизиране (*Model Routing*). Платформата трябва да бъде конфигурирана така, че да рутира комплексните аналитични задачи към Claude Opus 4.7, а регулярните и по-леки процеси - към алтернативни модели като GPT-4o. Този мултидоставчиков подход ще гарантира пълна софтуерна автономност на администрацията.

Сложност на внедряване

Наблюдения: *(Техническа зрялост, необходима инфраструктура, ИТ капацитет)*

Платформата предлага интуитивен потребителски интерфейс, но архитектурната ѝ интеграция в структурата на държавната администрация се характеризира с висока комплексност. Реализирането на успешна системна интеграция е критично ограничено при липса на квалифициран вътрешен ИТ капацитет за софтуерна поддръжка, управление на API интерфейси и контрол на сигурността.

Препоръка за публичния сектор:

Препоръчва се поэтапно внедряване чрез стартиране на ограничени пилотни проекти в нискорискови административни направления с цел тестване на технологичната съвместимост. Паралелно с това е необходимо организирането на специализирани обучения за повишаване на дигиталната квалификация на ключови служители, които ще оперират със системата.

Регулаторно съответствие

Наблюдения: *(EU AI Act, NIS2/3KC, GDPR, национални наредби)*

Моделът се категоризира като модел на ИИ с общо предназначение със системен риск съгласно Регламента на ЕС за изкуствения интелект (AI Act). За внедряващата го административна структура това поражда незабавно задължение за изготвяне на задължителна Оценка на въздействието върху защитата на данните. Поради естеството на архитектурата, ведомствата са длъжни да гарантират постоянен човешки надзор при вземането на решения съгласно рестрикциите за автоматизирано индивидуално профилиране, заложи в чл. 22 от GDPR. Допълнително, интеграцията изисква

съобразяване с изискванията за оперативна устойчивост на критичните субекти съгласно Директивата NIS2 (ЗАКОН ЗА КИБЕРСИГУРНОСТ).

Препоръка за публичния сектор:

Налага се провеждането на задължителен цялостен правен одит и детайлен преглед на съответствието преди реалното софтуерно внедряване. Необходимо е да бъде нормативно разписан и утвърден вътрешен административен протокол за непрекъснат човешки контрол за всяко софтуерно генерирано решение или предложение за акт, засягащо правата и интересите на гражданите.

Етика и прозрачност

Наблюдения: *(Обяснимост, одитируемост, отговорност при грешки)*

Моделът притежава добри способности за контекстуално обосноваване на своите отговори. Въпреки това, подобно на останалите големи езикови модели (LLM), неговата вътрешна архитектура функционира на принципа на „черна кутия“ (*Black Box*). Това ограничава дълбоката алгоритмична одитируемост и прави невъзможно прецизното проследяване на логическата верига, довела до специфична интерпретация на даден нормативен акт. Всички евентуални фактически грешки остават изцяло под споделената правна отговорност на крайния потребител.

Препоръка за публичния сектор:

Задължително е софтуерните решения в публичния сектор да притежават високо ниво на проверимост чрез поддържане на пълни системни журнали за одит. Налага се въвеждането на нормативен регламент за предварително уведомяване и информиране на гражданите във всички случаи, в които административните процеси и преписки са били подпомогнати или частично обработени чрез системи с изкуствен интелект.

Изкривявания (Bias)

Наблюдения: *(Пристрастия при различни групи, равнопоставеност)*

Поради обстоятелството, че корпусът от данни за обучение е доминиран от източници на английски език, съществува системен риск от пренасяне на чужди културни и социални изкривявания. Налице е вероятност от генериране на предубедени или неточни съждения по отношение на специфични етнически малцинства, социални групи или исторически фигури. Рискът от алгоритмична дискриминация и деформация на неутралността е документиран дефицит за всички големи езикови модели (LLM).

Препоръка за публичния сектор:

Налага се въвеждането на контролирана фаза на симулирани атаки и провокативно тестване чрез специализиран екип, съставен от експерти с различен професионален и социален профил. Тяхната цел трябва да бъде умишленото провокиране на модела за идентифициране на дискриминационни, обидни или субективни изказвания преди официалното му въвеждане в експлоатация.

Халюцинации / грешки

Наблюдения: *(Честота, detectability, пуск при вземане на решения)*

Всички генерирани правни и фактически твърдения изискват задължителна външна проверка. Критичен дефицит на технологията е високата степен на убедителност на изказа при генериране на грешни отговори (т.нар. „убедителни халюцинации“), което затруднява тяхното бързо разпознаване. Специфичен оперативен риск възниква в ситуации, изискващи логически преходи при липса на изрични данни в контекста — тогава моделът има системна склонност да фабрикува несъществуващи факти, нормативни основания или събития.

Препоръка за публичния сектор:

Въвеждане на задължителен регламент за пълна фактологична верификация на всеки изходящ текст. Налага се провеждане на специализирани експертни обучения за разпознаване на маркери за халюцинации в синтезираното съдържание. На оперативното ниво се препоръчва прилагането на

инженеринг чрез техники за кръстосана проверка (подаване на дефинирани документи и паралелно валидиране чрез алтернативно структурирани заявки по един и същ казус).
Морал и ценности
Наблюдения: <i>(Поведение при чувствителни теми, отказ на вредни задачи)</i> Моделът интегрира висока степен на етично съгласуване, като систематично отказва изпълнението на заявки с вредно, дискриминационно, нерегламентирано или незаконно съдържание. Поведението на системата при обработка на чувствителни и поляризирани теми е балансирано и обективно, което демонстрира стриктно съответствие с вградените принципи и конституционни правила за безопасност.
Препоръка за публичния сектор: Задължително е провеждането на предварителна емпирична оценка на модела чрез симулиране на реални административни и казуистични сценарии. Успоредно с това е необходимо разработването и нормативното утвърждаване на вътрешна институционална Политика за допустима употреба, която да дефинира ясните граници и етичните отговорности на служителите при работа с ИИ инструменти в държавната администрация.

D. Обобщена оценка на групата

Обобщена оценка по основни показатели		
Обща приложимост <i>средна от Раздел А</i>	Оценка 1–5 ☐ 1 ☐ 2 ☒ 3 ☐ 4 ☐ 5	Обосновка / бележки: Моделът притежава изключителен потенциал за оптимизация на бек-офис процесите във вътрешната администрация. Въпреки това, високата експлоатационна себестойност и архитектурните рискове при съблюдаването на принципа „необходимост да се знае“ ограничават възможността за присъждане на максимална оценка за цялостно внедряване към момента.
Готовност за внедряване <i>техническа и организационна</i>	Оценка 1–5 ☐ 1 ☐ 2 ☒ 3 ☐ 4 ☐ 5	Обосновка / бележки: Въпреки че моделът предлага готови за незабавна употреба функционалности (<i>Out-of-the-Box</i>), неговото софтуерно изолиране и конфигуриране в защитена корпоративна среда изискват профилиран ИТ капацитет (DevOps експерти и AI инженери).
Съответствие с регулации <i>AI Act, GDPR, NIS2</i>	Оценка 1–5 ☐ 1 ☐ 2 ☒ 3 ☐ 4 ☐ 5	Обосновка / бележки: Постигането на регулаторно съответствие е реализируемо, но изисква значителен обем от допълнителни софтуерни и административни настройки, без да се гарантира пълно отстраняване на правния риск. За гарантиране на безусловното съответствие с регулаторната рамка на ЕС е критично прилагането на комплексни правни, организационни и технически мерки за безопасност.
Етичен риск <i>обратна скала: 5=нисък риск</i>	Оценка 1–5 ☐ 1 ☐ 2 ☐ 3 ☒ 4 ☐ 5	Обосновка / бележки: Разработчикът на модела притежава най-утвърдената репутация в индустрията по отношение на етичните стандарти и безопасността на системите с изкуствен интелект. Спрямо конкурентните алтернативи на пазара, при този модел се отчита значително по-ниска уязвимост към злоупотреби и по-малък риск от генериране на нерегламентирано или вредно съдържание.
Обща препоръка <i>1=не препоръчваме, 5=силно препоръчваме</i>	Оценка 1–5 ☐ 1 ☐ 2 ☐ 3 ☒ 4 ☐ 5	Обосновка / бележки: Моделът се позиционира като високотехнологично и надеждно решение за нуждите на държавната администрация, при условие че се приложат всички дефинирани технически и правни мерки за защита на информацията. Препоръчва се първоначалното му внедряване под формата на ограничени пилотни проекти в рамките на експертни и аналитични звена. Поради финансови и регулаторни съображения, технологията не се препоръчва за масово и автоматизирано обслужване на граждани.

Мотивирано становище на групата (мин. 5 изречения): *Обобщете защо препоръчвате или не препоръчвате модела за публичния сектор.*

Експертната група препоръчва условното внедряване на модела Claude Opus 4.7 в публичния сектор, но единствено като специализирано бек-офис решение за експертни нужди.

Оценката на потенциала, рисковете и стратегическата съвместимост се базира на следните пет ключови аргумента:

- Лидерски аналитичен капацитет:** Технологията демонстрира най-високо ниво на логическо мислене (*Reasoning*) на пазара, което я прави мощен инструмент за обработка на комплексни правни казуси, одит на документация и софтуерна разработка.
- Масщабна обработка на данни:** Интегрираният контекстен прозорец от 1 000 000 токена позволява едновременно зареждане и анализ на цели досиета, архивни преписки и обемни нормативни актове без загуба на логическа нишка.
- Архитектурни рискове и GDPR съответствие:** Директният трансфер и обработка на информация на сървъри извън Европейското икономическо пространство (в САЩ) изискват допълнителни технически надстройки и изолация в защитени облачни среди, за да се гарантира пълно съответствие с GDPR.

- **Регулаторна тежест (EU AI Act):** Като модел за изкуствен интелект с общо предназначение (GPAI) със системен риск, неговото внедряване налага провеждането на задължителна Оценка на въздействието върху защитата на данните (DPIA) и внедряване на постоянен надзор.
- **Икономическа нецелесъобразност за масова употреба:** Поради високата си цена от \$25 за 1M изходящи токена и липсата на нативна гласова поддръжка, моделът е икономически неефективен за фронт-офис обслужване на граждани, където по-леки и евтини алтернативи могат да покрият базовите процеси.

Заклучение: Моделът не се препоръчва за масови общински услуги. Той обаче трябва да бъде внедрен чрез **хибриден подход и пилотни проекти** в тясно профилирани държавни и ИТ звена, където изискванията за етичност, автоматизация на процеси и дълбока експертиза носят най-голяма добавена стойност.

Предложени условия за ползване: *Какви ограничения, контроли или процедури трябва да съпътстват внедряването?*

1. Регулаторно съответствие с GDPR: Задължително сключване на Договор за обработка на данни с Anthropic. Въвеждане на изрична правна и техническа забрана за използването на входящите данни от администрацията за последващо дообучаване на ИИ от страна на Anthropic.

2. Задължителен надзор (Human-in-the-Loop): Прилагане на принципа за контрол „Human-in-the-Loop“. За всички административни процедури, засягащи права, законни интереси или задължения на граждани, крайният документ се подписва от служител след детайлна проверка на фактологията.

3. Алгоритмична проследимост и одит: Всички входящи заявки и генерирани отговори трябва автоматично да се логват и съхраняват в системни журнали за одит за период от 5 години. Това осигурява архив при последващи проверки за грешки, субективност или изкривявания.

4. Институционална рамка и квалификация: Разработване и внедряване на Политика за допустима употреба на системи с ИИ в рамките на ведомството. Провеждане на специализирани курсове за обучение на персонала относно рисковете от „убедителни халюцинации“.

5. Контролирана пилотна фаза: Ограничаване на първоначалния обхват на модела до едно изолирано експертно звено за минимален изпитателен период от 3 (три) до 6 (шест) месеца. Разширяването на софтуерната архитектура в цялата администрация се допуска само след одит на резултатите от пилотния проект.

Risk Canvas

Картографиране на рисковете от AI модел в публичен контекст

Anthropic Claude Opus 4.7

Инструкция за попълване: Попълнете всички полета за избория от вашата група AI модел. Празните редове и скалите са предназначени за ръчно попълване. Не изтривайте инструкциите в курсив.

Наименование на модела	Claude Opus 4.7
Разработчик / компания	Anthropic
Версия за оценка	claude-opus-4-7 (април 2026)
Дата на попълване	14.05.2026 г.
Членове на групата	Мариела, Теодора, Милен, Жасмина, Анелия, Денис
Ръководител на групата	Милен

Легенда за скалите

Вероятност (Вер.) 1 = Изключително рядко 2 = Рядко 3 = Умерено вероятно 4 = Вероятно 5 = Почти сигурно	Въздействие (Вл.) 1 = Незначително 2 = Малко 3 = Умерено 4 = Значително 5 = Критично	Ниво на риска Н = Нисък (Вер.×Вл. ≤ 4) С = Среден (5–12) В = Висок (13–25)
--	--	--

Риск Канва — 9-полева матрица

Попълнете всяко поле с идентифицирани рискове, засегнати страни и ключови наблюдения.

Инструкция за попълване: Всяко от 9-те полета представлява различна категория риск. Запишете конкретни рискове — не общи фрази.

1. Технически рискове <i>Халюцинации, грешки, downtime, версионни промени</i> Моделът Claude Opus 4.7 притежава капацитет за генериране на административни становища и документи, които се отличават с авторитетен и професионален стил, но могат да съдържат критични	2. Рискове за данните <i>Поверителност, GDPR, изтичане, съхранение</i> Налице е критична уязвимост, свързана с практиката на служители да използват лични или безплатни потребителски версии на Claude за обработка на служебни документи,	3. Правни / регулаторни <i>EU AI Act, NIS2, GDPR, национално право</i> Внедряването на генерирано от изкуствен интелект съдържание в официалните административни процеси е недопустимо без осигуряване на пълна проследимост на алгоритмичния
---	---	--

юридически или фактически неточности (убедителни халюцинации). Внедряването на тези текстове в реалната административна дейност без последваща експертна верификация създава непосредствен риск от вземане на неправилни управленски решения, допускане на процесуални нарушения и възникване на последващи правни спорове. Допълнителни инфраструктурни рискове са свързани с потенциални прекъсвания в наличността на API интерфейса и непланирани версионни промени от страна на Anthropic, които могат да доведат до деградация в качеството на генерирания изходен резултат.	съдържащи лични данни или класифицирана информация. Този нерегламентиран ИТ процес лишава администрацията от възможност за централизирано управление и мониторинг. При липса на корпоративни механизми за ограничаване на съхранението на данни и ясен контрол върху правата на достъп, възниква непосредствен риск информацията да бъде архивирана от доставчика, използвана извън първоначалното ѝ предназначение или компрометирана от неоторизирани лица. Подобни сценарии водят до тежки нарушения на Регламента за защита на личните данни (GDPR), Закона за защита на личните данни (ЗЗЛД) и вътрешните правила за информационна сигурност.	произход и без упражняването на ефективен човешки надзор над финалния резултат. Липсата на ясна софтуерна маркировка на изходните данни създава непосредствен риск от вземане на административни решения въз основа на неточна, непълна или невалидирана информация. Подобна неконтролирана практика влиза в директно противоречие с фундаменталните изисквания за прозрачност, отчетност и човешки контрол, заложи в европейската нормативна уредба – включително Общия регламент относно защитата на данните (GDPR) и Акта на ЕС за изкуствения интелект (EU AI Act).
4. Етични рискове <i>Bias, дискриминация, непрозрачност, автономия</i> При обработката на обществено чувствителни административни казуси съществува вероятност моделът да генерира формулировки, съдържащи субективни изкривявания (<i>Bias</i>), елементи на неравнопоставеност или неподходящ изказ спрямо определени групи граждани. Това създава непосредствен риск от допускане на дискриминационни практики или използване на език, който нарушава нормативно заложените принципи за обективност, политическа неутралност и безпристрастност на държавната администрация. Без прилагането на задължителен експертен преглед и филтрация, интегрирането на подобни текстове може да компрометира законосъобразността на административните решения и да доведе до драстичен спад в общественото доверие към институциите.	5. Оперативни рискове <i>Зависимост, интеграция, обучение, процеси</i> Интегрирането на системата с вътрешните регистри, информационни системи и документооборота на администрацията може да се окаже значително по-сложно и изискващо ресурси от първоначално предвиденото. Това създава риск от забавяне на административни процеси, необходимост от допълнителни технически разработки и повишена зависимост от ограничен ИТ капацитет и външни изпълнители. При липса на добра координация и техническа подготовка е възможно внедряването да доведе до временни затруднения в работата на администрацията и до по-високи разходи за поддръжка и адаптация на системите.	6. Репутационни рискове <i>Доверие на гражданите, медийно отразяване</i> Публичното разпространение на съдържание, генерирано от изкуствен интелект, което съдържа грешки, подвеждаща информация или неподходящи формулировки, може да доведе до сериозно негативно медийно отразяване. Подобен ефект би имало и евентуално изтичане на лични данни или чувствителна информация при използването на AI системи. Такива случаи създават риск от спад в общественото доверие към администрацията, съмнения относно надеждността на дигиталните услуги и засилен обществен и институционален натиск за ограничаване на използването на изкуствен интелект в публичния сектор.
7. Стратегически рискове <i>Vendor lock-in, ценови промени, спиране на услугата</i>	8. Засегнати страни <i>Граждани, служители, ръководство, партньори</i>	9. Остатъчни рискове <i>Рискове след митигация, приети рискове</i>

Администрацията може постепенно да изгради ключови работни процеси и услуги около Claude, което създава силна зависимост от конкретния доставчик – Anthropic. При промяна в ценовата политика, условията за ползване на API услугите, ограничения в достъпа или технически проблеми с платформата, това може пряко да засегне работата на администрацията. Подобна зависимост крие риск от затруднено преминаване към алтернативни решения, непредвидени разходи и ограничена гъвкавост при бъдещо развитие на информационните системи.	Служителите могат да използват текстове и предложения, генерирани от изкуствен интелект, без да са преминали обучение за ограниченията на технологията и без необходимата критична проверка на резултатите. Това създава риск от допускане на грешки в официалната комуникация, неточни административни действия или използване на неподходящи формулировки в документи и становища. При липса на ясни вътрешни правила и подготовка е възможно служителите да приемат генерираното съдържание като напълно надеждно, въпреки че то може да съдържа фактически, правни или процедурни неточности.	Дори при наличие на вътрешни правила, механизми за контрол, човешки преглед и мерки за информационна сигурност, използването на изкуствен интелект остава свързано с определен остатъчен риск. Възможно е моделът да генерира убедително звучаща, но невярна или подвеждаща информация, да бъде повлиян чрез злонамерени инструкции в подадените данни или да се стигне до човешка грешка при интерпретацията и използването на генерираното съдържание. Поради това дори добре управляваното внедряване на AI не може напълно да изключи риска от неточности, неправилни административни действия или инциденти, свързани със сигурността и качеството на информацията.
---	---	--

Рисков регистър

Попълнете по всяка категория конкретните рискове с вероятност, въздействие и мерки.

Инструкция за попълване: Вер. = Вероятност (1–5) · Вл. = Въздействие (1–5) · Ниво: ☐ Н (нисък) ☐ С (среден) ☐ В (висок)

Риск / Събитие	Вероятност	Въздействие	Ниво	Мярка за митигация
Технически рискове				
Халюцинации и фактически грешки при отговори	Вер. ☐ 1 ☐ 2 ☐ 3 ☒ 4 ☐ 5	Вл. ☐ 1 ☐ 2 ☐ 3 ☒ 4 ☐ 5	Ниво ☐ Н ☐ С ☒ В	RAG система (подаване на точни документи). Проверка от човек за всички официални изходи.
Недостъпност / downtime на платформата	Вер. ☐ 1 ☒ 2 ☐ 3 ☐ 4 ☐ 5	Вл. ☐ 1 ☐ 2 ☒ 3 ☐ 4 ☐ 5	Ниво ☐ Н ☒ С ☐ В	Multi-LLM стратегия – автоматично превключване към GPT-4o или Gemini при проблем. Локален кеш на често използвани отговори.
Внезапна промяна на версия или поведение на модела	Вер. ☐ 1 ☐ 2 ☒ 3 ☐ 4 ☐ 5	Вл. ☐ 1 ☐ 2 ☒ 3 ☐ 4 ☐ 5	Ниво ☐ Н ☐ С ☒ В	Пилотно тестване на всяка нова версия в изолирана среда преди внедряване. Фиксиране на версията в API повикванията.
Ограничения на контекстния прозорец при обемни документи	Вер. ☐ 1 ☐ 2 ☐ 3 ☒ 4 ☐ 5	Вл. ☐ 1 ☐ 2 ☒ 3 ☐ 4 ☐ 5	Ниво ☐ Н ☒ С ☐ В	Разделяне на документи и обобщаване на отделните части. Използване на векторна база данни за търсене в голям обем.

Грешки при мултимодален вход (изображения, аудио)	Вер. ☐ 1 ☐ 2 ☐ 3 ☒ 4 ☐ 5	Вл. ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☒ 5	Ниво ☐ H ☐ C ☒ B	Предварителна обработка – подобряване на качеството на изображенията. Сравнение с OCR резултат от специализиран инструмент.
Рискове за данните и поверителността				
Обработка на лични данни на граждани в облачна среда	Вер. ☐ 1 ☐ 2 ☐ 3 ☒ 4 ☐ 5	Вл. ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☒ 5	Ниво ☐ H ☐ C ☒ B	GDPR. Класифициран е като GPAI (general-purpose AI) модел със системен риск
Изтичане на класифицирана/служебна информация	Вер. ☐ 1 ☒ 2 ☐ 3 ☐ 4 ☐ 5	Вл. ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☒ 5	Ниво ☐ H ☐ C ☒ B	Достъп само за служители с ниво "вътрешно". Логове на всички заявки. Процедура за отмяна на API ключове.
Запазване на запитвания от доставчика за обучение	Вер. ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☒ 5	Вл. ☐ 1 ☒ 2 ☐ 3 ☐ 4 ☐ 5	Ниво ☐ H ☐ C ☒ B	Изрично изключване на опцията "train on data" в конфигурацията на API. Проверка на договора с Anthropic за това как се обработват данните.
Трансфер на данни извън ЕС (EU AI Act / Schrems II)	Вер. ☐ 1 ☐ 2 ☐ 3 ☒ 4 ☐ 5	Вл. ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☒ 5	Ниво ☐ H ☐ C ☒ B	Шифриране на данните в покой и при пренос.
Липса на одитна следа при автоматизирани решения	Вер. ☐ 1 ☐ 2 ☐ 3 ☒ 4 ☐ 5	Вл. ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☒ 5	Ниво ☐ H ☐ C ☒ B	Логване на всяка заявка и отговор в централна система. Въвеждане на хеш на отговора като доказателство за цялостта на данните.
Правни и регулаторни рискове				
Несъответствие с EU AI Act (Annex III — публичен сектор)	Вер. ☐ 1 ☐ 2 ☒ 3 ☐ 4 ☐ 5	Вл. ☐ 1 ☐ 2 ☒ 3 ☐ 4 ☐ 5	Ниво ☐ H ☐ C ☒ B	Оценка на въздействието (DPIA) преди внедряване. Документиране на системния риск според изискванията на EU AI Act.
Нарушение на GDPR при обработка на лични данни	Вер. ☐ 1 ☐ 2 ☐ 3 ☒ 4 ☐ 5	Вл. ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☒ 5	Ниво ☐ H ☐ C ☒ B	Човешки контрол при всички решения, засягащи граждани. Информация на гражданите, че се използва AI.
Липса на обяснимост при автоматизирани административни решения	Вер. ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☒ 5	Вл. ☐ 1 ☐ 2 ☒ 3 ☐ 4 ☐ 5	Ниво ☐ H ☒ C ☐ B	Задължително предоставяне на "reasoning" от модела в отговора. Проверка на логиката от човек преди изпращане до гражданина.
Авторскоправни нарушения в генерирано съдържание	Вер. ☐ 1 ☒ 2 ☐ 3 ☐ 4 ☐ 5	Вл. ☐ 1 ☐ 2 ☒ 3 ☐ 4 ☐ 5	Ниво ☐ H ☒ C ☐ B	Използване на модела само за трансформативни цели (анализ, обобщение), не за копиране. Проверка при публикуване на текстове онлайн.

Отговорност при грешни правни или медицински изходи	Вер. ☐ 1 ☐ 2 ☐ 3 ☒ 4 ☐ 5	Вл. ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☒ 5	Ниво ☐ H ☐ C ☒ B	Клауза в договора с доставчика за ограничаване на отговорността (стандартна). Задължителна проверка от лицензиран юрист/лекар преди използване на изхода.
Етични и социални рискове				
Систематичен bias спрямо определени групи граждани	Вер. ☐ 1 ☒ 2 ☐ 3 ☐ 4 ☐ 5	Вл. ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☒ 5	Ниво ☐ H ☐ C ☒ B	Редовни тестове за bias с представителни примери. "Black Sheep team" от служители с различен профил.
Дискриминация при автоматизирано обработване на заявления	Вер. ☐ 1 ☒ 2 ☐ 3 ☐ 4 ☐ 5	Вл. ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☒ 5	Ниво ☐ H ☐ C ☒ B	Наблюдение на изходите за дискриминационни модели. Въвеждане на проверка преди финално решение.
Непрозрачност — граждани не знаят, че AI взима решения	Вер. ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☒ 5	Вл. ☐ 1 ☒ 2 ☐ 3 ☐ 4 ☐ 5	Ниво ☐ H ☐ C ☒ B	Задължително уведомяване на гражданите във всеки канал за комуникация. Публичен регистър на AI употребите в администрацията.
Дезинформация и манипулация чрез убедително съдържание	Вер. ☒ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5	Вл. ☐ 1 ☒ 2 ☐ 3 ☐ 4 ☐ 5	Ниво ☒ H ☐ C ☐ B	Човешка проверка на всички публични съобщения, генерирани от AI. Воден знак на AI съдържание според изискванията на EU AI Act.
Загуба на критично мислене и зависимост при служителите	Вер. ☐ 1 ☒ 2 ☐ 3 ☐ 4 ☐ 5	Вл. ☐ 1 ☐ 2 ☒ 3 ☐ 4 ☐ 5	Ниво ☐ H ☒ C ☐ B	Обучение на служителите как да използват AI като помощник, а не като заместител. Задължителни проверки на AI отговорите от втори служител при важни задачи.
Оперативни рискове				
Липса на технически капацитет за внедряване и поддръжка	Вер. ☐ 1 ☐ 2 ☐ 3 ☒ 4 ☐ 5	Вл. ☐ 1 ☐ 2 ☒ 3 ☐ 4 ☐ 5	Ниво ☐ H ☒ C ☐ B	Пилотен проект с външен консултант или използване на AI като услуга от сертифициран доставчик.
Vendor lock-in — прекомерна зависимост от доставчик	Вер. ☐ 1 ☐ 2 ☐ 3 ☒ 4 ☐ 5	Вл. ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☒ 5	Ниво ☐ H ☐ C ☒ B	Multi-LLM архитектура – абстракция на API повикванията, позволяваща смяна на модела.
Недостатъчно обучение на служителите	Вер. ☐ 1 ☐ 2 ☐ 3 ☒ 4 ☐ 5	Вл. ☐ 1 ☐ 2 ☒ 3 ☐ 4 ☐ 5	Ниво ☐ H ☒ C ☐ B	Задължително сертифициране на служители, работещи с AI. Редовни обучения и тестове за качество на заявките.

Интеграционни проблеми с наследени (legacy) системи	Вер. ☐ 1 ☐ 2 ☐ 3 ☒ 4 ☐ 5	Вл. ☐ 1 ☐ 2 ☒ 3 ☐ 4 ☐ 5	Ниво ☐ Н ☒ С ☐ В	Разработка на microservices, които да абстрахират комуникацията.
Ценови скокове при промяна на лицензионния модел	Вер. ☐ 1 ☒ 2 ☐ 3 ☐ 4 ☐ 5	Вл. ☐ 1 ☐ 2 ☒ 3 ☐ 4 ☐ 5	Ниво ☐ Н ☒ С ☐ В	Бюджетиране с надбавка от 200% за неочаквани увеличения на цените. Резервен план за миграция към по-евтин модел.
Репутационни и стратегически рискове				
Публичен скандал при видима AI грешка (медийно отразяване)	Вер. ☐ 1 ☐ 2 ☐ 3 ☒ 4 ☐ 5	Вл. ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☒ 5	Ниво ☐ Н ☐ С ☒ В	Бързо признаване на грешката, корекция и обяснение на мерките. Пилотен обхват – първо в нечувствителни зони.
Загуба на доверие на гражданите в публичните институции	Вер. ☐ 1 ☒ 2 ☐ 3 ☐ 4 ☐ 5	Вл. ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☒ 5	Ниво ☐ Н ☐ С ☒ В	Прозрачност – публичен доклад за ползите и грешките на AI. Канал за обратна връзка от гражданите за AI отговори.
Политическа съпротива срещу използването на AI	Вер. ☐ 1 ☒ 2 ☐ 3 ☐ 4 ☐ 5	Вл. ☐ 1 ☐ 2 ☒ 3 ☐ 4 ☐ 5	Ниво ☐ Н ☒ С ☐ В	Ранно включване на парламентарната комисия по дигитализация. Пилотен проект с положителни резултати за представяне пред опонентите.
Зависимост от американски/китайски технологичен доставчик	Вер. ☐ 1 ☐ 2 ☐ 3 ☒ 4 ☐ 5	Вл. ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☒ 5	Ниво ☐ Н ☐ С ☒ В	План за миграция към европейски алтернативи при геополитическа криза.
Спиране на услугата при геополитически напрежения	Вер. ☒ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5	Вл. ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☒ 5	Ниво ☐ Н ☐ С ☒ В	Резервен план – използване на локален open-source модел в критични ситуации.

Обобщена рискова оценка на групата

Топ 3 критични риска за модела:

1. Обработка на лични данни и трансфер извън ЕС – рискът от нарушение на GDPR е най-сериозен, с потенциални глоби и спиране на дейността.
2. Халюцинации при правни/административни решения – моделът може да генерира несъществуващи членове от закони, което води до неправилни решения за граждани.
3. Vendor lock-in и геополитическа зависимост – прекомерната зависимост от американска компания крие риск от спиране на услугата или ценови шок.

Обща рискова оценка на модела за публичния сектор:

- ☐ Нисък риск — може да се внедри с минимални ограничения
- ☐ Среден риск — изисква допълнителни контроли и процедури
- ☒ Висок риск — изисква задълбочена оценка и специфична регулация

❑ Неприемлив риск — не се препоръчва за публичния сектор

Мотивирано становище:

Група “МААТ” оценява Claude Opus 4.7 като модел с висок риск за публичния сектор, въпреки неговите технически превъзходства. Основните рискови категории са правно-регулаторната (GDPR, AI Act) и за сигурност на данните, тъй като липсва нативна поддръжка за локално внедряване в ЕС без трансфер на данни към САЩ. Халюцинациите, макар и намалени, остават неприемливи за автоматизирани административни решения, засягащи правата на гражданите. Vendor lock-in и геополитическата зависимост допълнително повишават стратегическия риск.

Въпреки това, групата не счита риска за напълно неприемлив.

Моделът е подходящ за пилотни проекти с нисък риск (вътрешни обобщения, ИТ автоматизация), но не за директно обслужване на граждани или за взимане на автоматизирани правни решения без задълбочена допълнителна защита.

Карта за приложимост

AI модел в публичната администрация — функционален анализ

Anthropic Claude Opus 4.7

Инструкция за попълване: Попълнете всички полета за избория от вашата група AI модел. Празните редове и скалите са предназначени за ръчно попълване. Не изтривайте инструкциите в курсив.

Наименование на модела	Claude Opus 4.7
Разработчик / компания	Anthropic
Версия за оценка	claude-opus-4-7 (април 2026)
Дата на попълване	14.05.2026 г.
Членове на групата	Мариела, Теодора, Милен, Жасмина, Анелия, Денис
Ръководител на групата	Милен

Част 1. Обзорна матрица на приложимостта

Бърза оценка по административни функции. Прил. = Приложимост 1–5. Риск = Н/С/В.

 **Инструкция за попълване:** Попълнете всяка клетка: поставете оценка за приложимост (1=неприложимо, 5=напълно приложимо), ниво на риск (Н/С/В) и кратка бележка.

Административна функция	Обработка на документи	Правен анализ	Граждански услуги	Анализ на данни	Програмиране	Обучение
Обработка на кореспонденция	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н □С □В	Прил.: ○1 ○2 ●3 ○4 ○5 Риск: □Н □С □В	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: □Н □С □В	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: □Н □С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н □С □В	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: □Н □С □В
Изготвяне на доклади и анализи	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н □С □В	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: □Н □С □В	Прил.: ●1 ○2 ○3 ○4 ○5 Риск: □Н □С □В	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: □Н □С □В	Прил.: ●1 ○2 ○3 ○4 ○5 Риск: □Н □С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н □С □В
Тълкуване на нормативни актове	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: □Н □С □В	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: □Н □С □В	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: □Н □С □В	Прил.: ●1 ○2 ○3 ○4 ○5 Риск: □Н □С □В	Прил.: ●1 ○2 ○3 ○4 ○5 Риск: □Н □С □В	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: □Н □С □В
Изготвяне на правни становища	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: □Н □С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н □С □В	Прил.: ○1 ●2 ○3 ○4 ○5 Риск: □Н □С □В	Прил.: ●1 ○2 ○3 ○4 ○5 Риск: □Н □С □В	Прил.: ●1 ○2 ○3 ○4 ○5 Риск: □Н □С □В	Прил.: ○1 ○2 ●3 ○4 ○5 Риск: □Н □С □В

Отговори на граждански запитвания	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: ☒Н □С □В	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: □Н ☒С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: ☒Н □С □В	Прил.: ○1 ●2 ○3 ○4 ○5 Риск: ☒Н □С □В	Прил.: ●1 ○2 ○3 ○4 ○5 Риск: ☒Н □С □В	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: ☒Н □С □В
Чатбот за електронни услуги	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: ☒Н □С □В	Прил.: ○1 ○2 ●3 ○4 ○5 Риск: □Н ☒С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: ☒Н □С □В	Прил.: ○1 ●2 ○3 ○4 ○5 Риск: ☒Н □С □В	Прил.: ○1 ○2 ●3 ○4 ○5 Риск: ☒Н □С □В	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: ☒Н □С □В
Статистически анализ	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: ☒Н □С □В	Прил.: ●1 ○2 ○3 ○4 ○5 Риск: ☒Н □С □В	Прил.: ●1 ○2 ○3 ○4 ○5 Риск: ☒Н □С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: ☒Н □С □В	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: ☒Н □С □В	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: ☒Н □С □В
Обработка на отчети	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: ☒Н □С □В	Прил.: ○1 ○2 ●3 ○4 ○5 Риск: □Н ☒С □В	Прил.: ●1 ○2 ○3 ○4 ○5 Риск: ☒Н □С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: ☒Н □С □В	Прил.: ●1 ○2 ○3 ○4 ○5 Риск: ☒Н □С □В	Прил.: ○1 ○2 ●3 ○4 ○5 Риск: ☒Н □С □В
Автоматизация на работни процеси	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: ☒Н □С □В	Прил.: ○1 ●2 ○3 ○4 ○5 Риск: □Н ☒С □В	Прил.: ●1 ○2 ○3 ○4 ○5 Риск: ☒Н □С □В	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: ☒Н □С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: ☒Н □С □В	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: ☒Н □С □В
Разработка на вътрешни инструменти	Прил.: ○1 ●2 ○3 ○4 ○5 Риск: ☒Н □С □В	Прил.: ●1 ○2 ○3 ○4 ○5 Риск: ☒Н □С □В	Прил.: ○1 ○2 ○3 ○4 ○5 Риск: ☒Н □С □В	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: ☒Н □С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: ☒Н □С □В	Прил.: ○1 ○2 ●3 ○4 ○5 Риск: ☒Н □С □В
Учебни материали и тестове	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: ☒Н □С □В	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: ☒Н □С □В	Прил.: ○1 ○2 ●3 ○4 ○5 Риск: ☒Н □С □В	Прил.: ○1 ○2 ●3 ○4 ○5 Риск: ☒Н □С □В	Прил.: ○1 ○2 ●3 ○4 ○5 Риск: ☒Н □С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: ☒Н □С □В
Наръчници и процедури	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: ☒Н □С □В	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: □Н ☒С □В	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: ☒Н □С □В	Прил.: ○1 ●2 ○3 ○4 ○5 Риск: ☒Н □С □В	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: ☒Н □С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: ☒Н □С □В
Превод и локализация	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: ☒Н □С □В	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: □Н ☒С □В	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: ☒Н □С □В	Прил.: ●1 ○2 ○3 ○4 ○5 Риск: ☒Н □С □В	Прил.: ●1 ○2 ○3 ○4 ○5 Риск: ☒Н □С □В	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: ☒Н □С □В
Протоколи и резюмета	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: ☒Н □С □В	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: □Н ☒С □В	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: ☒Н □С □В	Прил.: ○1 ○2 ●3 ○4 ○5 Риск: ☒Н □С □В	Прил.: ●1 ○2 ○3 ○4 ○5 Риск: ☒Н □С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: ☒Н □С □В

Част 2. Детайлен анализ по случаи на употреба

За всеки сценарий оценете конкретно как избраният модел се справя.

2.1. Обработка на административни документи
Способността на модела да чете, структурира, резюмира и изготвя административни документи — решения, наредби, докладни записки, протоколи. Оценка за избрания модел: - Може ли моделът да обработва документи на български език с административна терминология? - Поддържа ли обработка на PDF/Word файлове с дълъг обем? - Запазва ли форматирането и структурата при генериране на изходен документ? - Изисква ли специализирано prompt инженерство за административен контекст? Отговори / наблюдения: - Обработка отлично български административни документи с коректна терминология и стил. - Контекстен прозорец от 1М токена позволява обработка на цели наредби и решения. - Частично – изходният текст е добре структуриран, но специфично форматиране изисква допълнителна обработка.

4. Минимално – стандартни промптове работят добре; за специализирани шаблони е нужна лека настройка.

Препоръка:

☐ Препоръчва се без ограничения

☒ Препоръчва се с ограничения (опишете):

Задължително верифициране на изходни документи от служител преди изпращане. Изграждане на институционален шаблон за промпт с примерни документи.

☐ Не се препоръчва (обосновка):

2.2. Правен анализ и тълкуване

Способността да анализира нормативни актове, да открива противоречия, да прилага правни норми към конкретни случаи.

Оценка за избрания модел:

1. Колко точен е при тълкуване на български законодателни текстове?
2. Може ли да следи промени в нормативната база (обновени закони)?
3. Как се справя с многозначност и правна неяснота?
4. Маркира ли ясно когато излиза извън областта на своята компетентност?

Отговори / наблюдения:

1. Точен при общото тълкуване – рискове при специфична национална юриспруденция (напр. ЗОП, ЗМВР).
2. Моделът не следи промени в реално време; изисква актуализиране на контекста ръчно.
3. Добре маркира многозначност и предлага алтернативни тълкувания, посочвайки неяснотата.
4. Моделът изрично посочва границите на компетентността си и препоръчва консултация с юрист.

Препоръка:

☐ Препоръчва се без ограничения

☒ Препоръчва се с ограничения (опишете):

Само за предварителен анализ и изготвяне на проекти – задължителна проверка от правоспособен юрист. Не се използва като единствен правен авторитет.

☐ Не се препоръчва (обосновка):

2.3. Взаимодействие с граждани (чатбот / асистент)

Способността да комуникира с граждани — отговаря на въпроси, помага при подаване на заявления, обяснява административни процедури.

Оценка за избрания модел:

1. Разбира ли естествен разговорен български (включително диалекти и грешки)?
2. Дава ли ясни и разбираеми отговори без административен жаргон?
3. Как се справя с чувствителни теми (социални помощи, здраве, правосъдие)?
4. Може ли да се интегрира с портали за електронни услуги?

Отговори / наблюдения:

1. Разбира разговорен български, включително диалектни форми и правописни грешки на потребителите.
2. Генерира ясни отговори без административен жаргон при правилно зададен промпт.
3. Моделът е деликатен при чувствителни теми и отклонява неподходящи запитвания.
4. REST API позволява интеграция с e-Gov портали и системи за електронни услуги.

Препоръка:

☐ Препоръчва се без ограничения

☒ Препоръчва се с ограничения (опишете):

Препоръчваме за информационни чатботове. Задължителна human escalation при правни и финансови въпроси. Ясна идентификация на AI пред гражданина.

☐ Не се препоръчва (обосновка):

2.4. Анализ на данни и отчетност

Способността да обработва структурирани данни, да генерира отчети и да идентифицира тенденции.

Оценка за избрания модел:

1. Може ли да обработва Excel/CSV данни и да генерира анализи?
2. Поддържа ли визуализация на данни (графики, диаграми)?
3. Колко надежден е при числови изчисления и статистика?
4. Може ли да се свърже с вътрешни бази данни чрез API?

Отговори / наблюдения:

1. Обработка Excel/CSV данни когато са предоставени в текстов вид или чрез Code Execution.
2. Описва графики текстово; директна визуализация изисква допълнителен инструмент (Python, Power BI).
3. Добър при прости изчисления; при сложна статистика (регресии, тестове) изисква верификация.
4. Чрез API може да получава данни от бази данни, но изисква разработка на middleware слой.

Препоръка:

☐ Препоръчва се без ограничения

☒ Препоръчва се с ограничения (опишете):

Подходящ за коментиране и интерпретация на данни, не за директно изчисление. Комбиниране с Python за пълна аналитична верига.

☐ Не се препоръчва (обосновка):

2.5. Вътрешно обучение на служители

Способността да генерира учебни материали, тестове, наръчници и да поддържа интерактивно обучение.

Оценка за избрания модел:

1. Може ли да генерира тематични курсове и тестове за служители?
2. Адаптира ли съдържанието към различни нива на компетентност?
3. Поддържа ли формати за дистанционно обучение (SCORM и др.)?
4. Как се справя с чувствителни теми — антикорупция, етика, дискриминация?

Отговори / наблюдения:

1. Генерира тематични курсове, тестове с различни типове въпроси и обяснителни материали.
2. Адаптира сложността на езика и примерите към посоченото ниво на обучаемите.
3. Не поддържа SCORM директно; съдържанието може да бъде експортирано и интегрирано ръчно.
4. Справя се балансирано с теми като антикорупция, етика и дискриминация.

Препоръка:

☐ Препоръчва се без ограничения

☒ Препоръчва се с ограничения (опишете):

☐ Не се препоръчва (обосновка):

2.6. Програмиране и автоматизация

Способността да пише код, да автоматизира процеси и да поддържа вътрешни системи.

Оценка за избрания модел:

1. Поддържа ли езиките и технологиите, използвани в административните системи?

2. Може ли да чете и документира съществуващ (legacy) код?

3. Колко надеждни са генерираните скриптове без допълнителна проверка?

4. Поддържа ли agentic режим за автономно изпълнение на задачи?

Отговори / наблюдения:

1. Поддържа Python, SQL, JavaScript, PowerShell и VBA – основните технологии в администрацията.

2. Чете и документира код, обяснява логиката и предлага рефакторинг.

3. Добър за генериране на сложни скриптове, изисква се задължително тестване преди внедряване.

4. Claude Code и Agentic API режим позволяват автономно изпълнение на многостъпкови задачи.

Препоръка:

☐ Препоръчва се без ограничения

☒ Препоръчва се с ограничения (опишете):

Задължително тестване на всеки генериран код в тестова среда преди производствено внедряване. Не се допуска автономен достъп до production системи без одобрение.

☐ Не се препоръчва (обосновка):

Част 3. Оценка на интеграционния потенциал

 **Инструкция за попълване:** Оценете техническите възможности за интеграция с административната инфраструктура.

Интеграционен аспект	Описание	Поддръжка	Бележки
API свързване	REST API, GraphQL, webhooks — интеграция с вътрешни системи	☒ Да ☐ Не ☐ Частично	
Single Sign-On	Съвместимост с Active Directory / ADFS / Keycloak	☐ Да ☐ Не ☒ Частично	Active Directory изисква настройка.
On-premise deployment	Възможност за локален хостинг без облачна зависимост	☐ Да ☒ Не ☐ Частично	Само cloud-базирана услуга.
Данни в ЕС	Сървъри на европейска територия — GDPR съответствие	☐ Да ☒ Не ☐ Частично	Сървъри в САЩ.
Одитна следа	Логирание на всички взаимодействия за отчетност	☐ Да ☐ Не ☒ Частично	Пълен audit trail изисква допълнителна разработка.
Ролеви достъп	RBAC — различни нива на достъп за различни служители	☐ Да ☐ Не ☒ Частично	
Многоезичност	Поддръжка на български и официални ЕС езици	☒ Да ☐ Не ☐ Частично	Поддържа всички официални ЕС езици включително български.

Карта за приложимост — AI Модел Claude Opus 4.7.

Мобилен достъп	Работа от мобилни устройства (field work)	☒ Да ☐ Не ☐ Частично	iOS и Android приложения с пълна функционалност.
Офлайн режим	Функционалност без постоянна интернет връзка	☐ Да ☒ Не ☐ Частично	Изисква постоянна интернет връзка – облачна услуга.
SLA и поддръжка	Гарантирано ниво на обслужване за публичния сектор	☐ Да ☐ Не ☒ Частично	

Обобщена карта за приложимост

Топ 3 приложения с висок потенциал:

- 1. Обработка и изготвяне на административни документи (решения, доклади, протоколи)
- 2. Правен анализ и тълкуване на нормативни актове за подпомагане на юридически екипи
- 3. Вътрешно обучение на служители – учебни модули, тестове, наръчници

Функции, при които моделът НЕ е подходящ:

- 1. Автономни решения с правни последиствия за граждани (без human-in-the-loop)
- 2. Обработка на класифицирана информация без анонимизация
- 3. Реално-времеви правоприлагане без допълнителен надзор

Препоръчани условия за пилотно внедряване:

Пилотно в 2-3 министерства: обработка на документи и вътрешно обучение.

Задължително: DPA с Anthropic, обучение на служители, вътрешна AI политика.

Human-in-the-loop за правни становища; одит на AI резултатите на всеки 3 месеца.

Enterprise лиценз и IT интеграция; оценка на EU AI Act съответствие преди старт.

Методологични указания

Етично и прозрачно ползване на AI модели в публичната администрация

Anthropic Claude Opus 4.7

Инструкция за попълване: Попълнете всички полета за избория от вашата група AI модел. Празните редове и скалите са предназначени за ръчно попълване. Не изтривайте инструкциите в курсив.

аименование на модела	Claude Opus 4.7
Разработчик / компания	Anthropic
Версия за оценка	claude-opus-4-7 (април 2026)
Дата на попълване	14.05.2026 г.
Членове на групата	Мариела, Теодора, Милен, Жасмина, Анелия, Денис
Ръководител на групата	Милен

Предговор и обхват

Инструкция за попълване: Попълнете идентификационните данни на вашата организация и дефинирайте обхвата на документа.

Наименование на организацията	Министерство на отбраната, ДАНС
Административно ниво	Централна администрация
AI модел(и), обхванати от документа	Claude Opus 4.7 (Anthropic)
Версия на документа	1.0.
Дата на приемане	14.05.2026 г.
Следваща планирана ревизия	14.11.2026 г.
Одобрен от	Милен – (няма налична информация)
Отговорен за прилагането	Теодора, Мариела, Жасмина, Анелия, Денис – (няма налична информация)

Чл. 1. Предмет и обхват

Настоящите указания уреждат етичното и прозрачно използване на AI модела Claude Opus 4.7 на Anthropic в дейността на (няма налична информация). Те обхващат всички структурни звена, които внедряват или използват модела чрез API във вътрешно-административни процеси (обработка на документи, IT автоматизация, вътрешна комуникация, обучение на служители). Документът не се прилага за използване на модела директно пред граждани (чатбот) – за тази цел се изисква отделна оценка.

Чл. 2. Правна и регулаторна рамка

- EU AI Act (Регламент (ЕС) 2024/1689) – особено разпоредбите за GPAI със системен риск и високорисковите системи (Annex III).
- Общ регламент относно защитата на данните (GDPR) – Регламент (ЕС) 2016/679 (чл. 5, 13, 15, 22, 30, 32, 35).

3. Закон за електронното управление (ЗЕУ) – относно прозрачност, сигурност и качество на обществените услуги.
4. Закон за защита на личните данни (ЗЗЛД) и указанията на КЗЛД.
5. Директива NIS2 и Закон за киберсигурност (ЗКС) – мерки за сигурност на мрежовите и информационните системи.
6. Вътрешни актове на организацията – Правилник за работа, Политика за сигурност и Политика за поверителност.

Раздел I. Принципи за етично и прозрачно ползване

Определете основните принципи, на които се базира политиката. Попълнете конкретните мерки за всеки.

Основни принципи	
1 Законност	Всички употреби на Claude Opus 4.7 се оценяват предварително от правното звено за съответствие с EU AI Act и GDPR. Забранени са употребите, които нарушават чл. 22 от GDPR (автоматизирано вземане на решения, засягащи граждани).
2 Прозрачност	Във всички официални документи, подпомогнати от AI, се включва явно обозначение: „Изготвено с подкрепата на AI модел Claude Opus 4.7, проверено от служител“. На уебсайта на организацията се публикува описание на използваните AI системи.
3 Отчетност	За всеки документ или решение, подпомогнати от AI, се определя служител-носител на отговорност, който проверява и одобрява изхода. AI никога не замества човешката отговорност. Грешките на модела не освобождават служителя от отговорност.
4 Справедливост	На всеки 3 месеца се извършва тест за bias с представителни примери (различни имена, адреси, социални групи). При установяване на дискриминационен модел употребата се спира до отстраняване на проблема чрез промяна на инструкциите.
5 Поверителност	Лични данни се изпращат към Claude Opus 4.7 само след премахване на ЕГН, имена, точен адрес. Забранено е изпращането на чувствителни данни (чл. 9 от GDPR) към модела. Внедряването става само след сключен Договор за обработка на лични данни (DPA).
6 Надеждност	Преди въвеждане в редовна експлоатация, всяка задача се тества пилотно в продължение на 30 дни с оценка на точността (>95% за фактологични задачи). Резултатите подлежат на задължителна верификация от човек (не се доверяваме автоматично).
7 Пропорционалност	За нискорискови задачи (вътрешно резюме) е допустима автоматизация с одит. За среднорискови задачи (проект на писмо) е задължителна проверка от втори служител. За високорискови задачи (правни становища, оценки със значение за граждани) AI служи само като помощник – решението се взема изцяло от човек.
8 Обяснимост	За всяко решение, засягащо граждани, се съхранява одитен лог с входа (промпта) и изхода на модела. При оспорване на решението, организацията предоставя на гражданина информация, че е използван AI, както и основните причини за изхода, без да разкрива търговски тайни.

Раздел II. Разрешени и забранени употреби

Определете за всяка категория задачи дали е разрешена, допустима с условия или забранена.

Инструкция за попълване: За всяка задача: поставете отметка в съответната колона и опишете условията или ограниченията.

Категория задача	Статус	Условия / Ограничения / Процедура
Административни документи		
Изготвяне на проекти за решения	☐ Разрешено ☒ С условия ☐ Забранено	Условия: Само за вътрешно ползване или като проект. Проектът задължително се проверява и коригира от служител с право на подпис. В крайния документ се отбелязва, че е подпомогнат от AI.
Резюмиране на входяща поща	☒ Разрешено ☐ С условия ☐ Забранено	Няма специални ограничения, но служител трябва да провери дали резюмето не пропуска важни детайли от оригиналния документ.
Превод на документи	☐ Разрешено ☒ С условия ☐ Забранено	Условия: Само за чернова на превод. За официални документи (нотариални, съдебни) преводът задължително се проверява от рецензент, владеещ чуждия език. Моделът не може да превежда самичък – нужна е човешка редакция.
Генериране на официални писма	☐ Разрешено ☒ С условия ☐ Забранено	Условия: Писмата се генерират от одобрен шаблон на организацията. Задължителна проверка от ръководител на звено преди изпращане. Забранено за използване в наказателни производства или принудителни мерки.
Правни и нормативни въпроси		
Анализ на нормативни текстове	☒ Разрешено ☐ С условия ☐ Забранено	Разрешено за вътрешни анализи и проучвания, но изходът винаги се проверява с оригиналния източник на правото. При съмнение се консултира правно звено.
Проверка за вътрешни противоречия	☒ Разрешено ☐ С условия ☐ Забранено	Разрешено като помощно средство, но не може да замества проверката от юрист.
Изготвяне на правни становища	☐ Разрешено ☐ С условия ☒ Забранено	Забранено напълно. Правните становища, които се изпращат на граждани или други институции, не могат да бъдат генерирани от AI. Те се изготвят изцяло от квалифициран юрист. AI може да се ползва само за спомагателно проучване, но не и за генериране на финалния текст.
Тълкуване на разпоредби към граждани	☐ Разрешено ☒ С условия ☐ Забранено	Условия: Отговорът задължително се верифицира от служител преди изпращане на гражданина. В отговора се съдържа изричното предупреждение: <i>"Това е информативен отговор, генериран с помощта на AI. За официално тълкуване се обърнете към компетентен орган."</i>
Вземане на решения с ефект върху граждани		
Преценка за социални помощи	☐ Разрешено ☒ С условия ☒ Забранено	Забранено напълно. Решенията за отпускане или отказ на социални помощи не могат да се базират на AI анализ. Моделът може да се ползва само за обобщение на документи без пряко влияние върху решението.
Оценка на проекти и заявления	☐ Разрешено ☒ С условия ☐ Забранено	Условия: AI може да се използва за предварително групиране на заявления или за извличане на формални несъответствия (напр. липсващ подпис). Финалната оценка се прави изцяло от човек.
Наказателни и принудителни мерки	☐ Разрешено ☐ С условия ☒ Забранено	Забранено напълно. Моделът не може да се използва за никакви действия, свързани с наказателни производства, глоби, задържания или други принудителни мерки.
Автоматизирани решения без човешки контрол	☐ Разрешено ☐ С условия ☒ Забранено	Забранено напълно съгласно чл. 22 от GDPR. Всяко решение със значение за гражданите изисква човешки контрол и одобрение.

Комуникация с граждани		
Чатбот за информационни въпроси	☐ Разрешено ☒ С условия ☐ Забранено	Условия: Чатботът трябва явно да обявява, че е AI, и да дава възможност за препращане към човешки оператор. Забранява се събирането на лични данни чрез чатбота. Отговорите се ограничават до общодостъпна информация (работно време, адреси, формуляри).
Проекти за отговори на жалби	☒ Разрешено ☐ С условия ☐ Забранено	Разрешено за генериране на проект за отговор, но задължителна проверка и редакция от служител. Забранено за жалби, съдържащи чувствителни лични данни.
Публикации в социални медии	☐ Разрешено ☒ С условия ☐ Забранено	Условия: Публикациите се проверяват от отговорник за комуникация преди публикуване. За всяка AI-генерирана публикация се прави маркировка #AI-assisted#.
Кризисна комуникация	☐ Разрешено ☐ С условия ☒ Забранено	Забранено напълно. По време на кризи (бедствия, инциденти) комуникацията се изготвя изцяло от хора, за да се избегне риск от дезинформация.
Вътрешно-административни функции		
Обучение на служители	☒ Разрешено ☐ С условия ☐ Забранено	Разрешено за генериране на учебни материали, примери и тестове, но материалите се проверяват от експерт по темата преди употреба.
Анализ на данни и отчети	☒ Разрешено ☐ С условия ☐ Забранено	Разрешено за вътрешни анализи, стига да не се обработват лични данни без анонимизация. Изводите от AI се проверяват статистически от анализатор.
Автоматизация на повтарящи се задачи	☒ Разрешено ☐ С условия ☐ Забранено	Разрешено, но всяка автоматизация (напр. попълване на формуляри) се тества в "тестова среда" преди реална употреба. Автоматизацията не може да засяга лични данни на граждани.
Обработка на класифицирана информация	☐ Разрешено ☐ С условия ☒ Забранено	Забранено напълно. Моделът не може да обработва информация с ниво на класификация "Секретно" или "Поверително". За такава информация се използват локални AI системи (on-premise), които не са обхванати от настоящите указания и работят в затворена мрежа.

Раздел III. Процедури за одобрение и верификация

Чл. 10. Задължителна верификация

Задължителна проверка от служител (винаги преди използване на AI изхода) се изисква за:

1. Всички документи, които напускат организацията (писма, отговори на жалби, проекти на решения).
2. Всички правни анализи и тълкувания на нормативни актове.
3. Всички изходи, които засягат права или задължения на граждани (дори ако служат само за помощ).
4. Всички генерирани кодове или скриптове, които ще се изпълняват в производствена среда.
5. Всички комуникации с медиите или публикации в официални канали.

Не се изисква задължителна проверка за вътрешни чернови, лични бележки или експериментални тестове, които нямат да напуснат работната станция на служителя.

Чл. 11. Процедура за одобрение на нови употреби

1. Служителят подава искане до AI отговорника по електронен път, описвайки: задачата, очакваната полза, рисковете и необходимите данни.
2. AI отговорникът организира преценка от тристранна комисия (ИТ, правно звено, ДЛЗД) в срок до 14 дни.

3. Комисията издава становище: разрешено, разрешено с условия, забранено.
4. Разрешените нови употреби се вписват в приложение към настоящите указания при следващата ревизия на документа.

Чл. 12. Задължително разкриване към граждани

Гражданинът се информира във всички случаи, когато:

1. Отговорът на неговото запитване или жалба е бил генериран с помощта на AI (дори и да е редактиран от служител).
2. Негови лични данни са били обработвани от AI.
3. Вземането на решение, което го засяга, е било подпомогнато от AI анализ (чл. 13, пар. 2, буква "f" от GDPR – право на информация).

Информирането става чрез явно изречение в самия документ/отговор или чрез стандартно уведомление на уебсайта на организацията. При поискване гражданинът има право да получи повече информация за логиката на AI системата, без да се засягат търговски тайни или сигурността.

Раздел IV. Мониторинг, одит и отчетност

Минимални изисквания за одитна следа (отбележете приложимите):

- Всяко взаимодействие с AI системата се записва с дата, час и потребител
- Пазят се логове на AI изходите, използвани в официални документи
- При автоматизирани решения се съхранява AI входа и изхода
- Провежда се периодичен преглед на качеството на AI изходите (мин. 1х/тримесечие)
- Инцидентите, свързани с AI грешки, се докладват на определено звено
- Годишен отчет за употребата на AI се изготвя и публикува

Чл. 15. KPI за мониторинг на AI употребата

1. Обем на употреба: Брой заявки към Claude Opus 4.7 (месечно, разбити по звена).
2. Проверка на качество: Месечен дял на проверените от човек AI изходи (цел: 100% за външни комуникации).
3. Грешки: Брой и вид на откритите грешки/халюцинации на месец (цел: <5% грешки при фактологични задачи).
4. Спазване на указанията: Брой случаи на нерегламентирана употреба на AI (наказателни производства, лични данни без анонимизация) – цел: 0.
5. Bias тестове: Резултати от тримесечните тестове за дискриминация (цел: няма систематично пристрастие към която и да е група).
6. Обучение: Дял на служителите, преминали задължително обучение (цел: 100%).

Раздел V. Отговорности и управление

Роля	Отговорност	Конкретно лице / звено
Ръководител на организацията	Стратегическо одобрение и крайна отговорност за AI политиката. Одобрява годишния отчет и ресурсите за обучение.	Боса, ПСО Шефа, ДАНС
AI отговорник (AI Officer)	Координация, ежедневен мониторинг на KPI, обучение, докладване на инциденти, поддръжка на този документ.	Милен

Служители, ползващи AI	Спазване на указанията, задължителна верификация на изходите, докладване на проблеми и грешки на AI отговорника.	Всички служители с достъп до модела
ИТ отдел / системен администратор	Технически достъп до модела, интеграция с вътрешните системи, осигуряване на киберсигурност, съхранение на логове, контрол на API ключовете.	Денис, Началник на ОИТ
Правно звено	Преглед на нови употреби, оценка на съответствие с EU AI Act и други регулации, консултиране при инциденти.	Мариела, Жасмина, Анелия
ДЛЗД (длъжностно лице по защита на данни)	Оценка на въздействието (DPIA), контрол за спазване на GDPR, обработка на сигнали от граждани, свързани с лични данни и AI.	Мариела
Вътрешен одит	Периодичен независим преглед на AI употребата (веднъж годишно), проверка на одитните следи, оценка на ефективността на контролите.	Жасмина, Анелия

Раздел VI. Обучение и изграждане на капацитет

Чл. 20. Задължително обучение

Задължително обучение преминават:

1. Всички служители, които имат достъп до API на Claude Opus 4.7 (минимум 4 часа).
2. Всички ръководители на звена, където се използва модела (минимум 2 часа).
3. AI отговорникът и ИТ администраторите (минимум 8 часа специализирано обучение).

Обучението се документира с протокол за проведено обучение и тест (70% верни отговора за издържане). Периодичност: първоначално обучение преди получаване на достъп, след това веднъж годишно (текущо обучение).

Модул на обучение	Целева група	Форма и обем	Периодичност
Основи на AI и рисковете	Всички служители	Онлайн курс, 2 часа + тест	Веднъж (при наемане или при получаване на достъп)
Работа с конкретния модел	Служители с достъп	Практически семинар, 2 часа	Веднъж + при нова версия
Етика и нормативна рамка	Всички служители с достъп	Лекция с казуси, 2 часа	Ежегодно
Разпознаване на AI грешки	Всички служители с достъп	Практически семинар, 2 часа	Ежегодно
Защита на личните данни	Всички служители с достъп + ДЛЗД	Онлайн курс, 1 час + тест	Ежегодно

Раздел VII. Актуализация и валидност

Чл. 24. Условия за задължителна ревизия

- Настоящите указания подлежат на задължителна ревизия в следните случаи:
- 1. Промяна на версията на модела (напр. Claude Opus 4.8 или 5.0) – ревизия в срок до 30 дни.
 - 2. Влизане в сила на нови или променени регулации (EU AI Act, GDPR, NIS2, национални закони) – ревизия в срок до 60 дни.
 - 3. Сериозен инцидент с модела (напр. изтичане на данни, дискриминация, публичен скандал) – незабавна ревизия и предприемане на коригиращи мерки (включително временно спиране на ползването).
 - 4. Изтичане на срока на валидност (6 месеца от последната ревизия) – ревизия и актуализация.

Чл. 25. История на промените

Версия	Дата	Описание на промяната	Одобрил
1.0.	14.05.2026 г.	Първоначална версия на указанията за Claude Opus 4.7.	Милен
1.1	18.05.2026 г.	Редакция на основни нормативни документи използвани в указанията	Милен

ГРУПА "ЯНУС"
Google Gemini
(3.1 Pro/ 3 Flash/ 2.5 Pro)



SWOT Анализ

Оценка на фундаментален AI модел за публичната администрация

v1.0 — Работен шаблон

-  **Инструкция за попълване:** Попълнете всички полета за избория от вашата група AI модел. Празните редове и скалите са предназначени за ръчно попълване. Не изтривайте инструкциите в курсив.

Наименование на модела	Пример: Claude Opus 4.7 / GPT-5.4 / Gemini 3.1 Pro Gemini 2,5 Pro
Разработчик / компания	Пример: Anthropic / OpenAI / Google DeepMind Google Gemini
Версия за оценка	Точната версия, която групата изследва Gemini 2,5 pro – Preview 09-2025
Дата на попълване	ДД.ММ.ГГГГ 15.05.2026
Членове на групата	Имена на участниците Михаела Неделчева, Петя Дошкова, Моника Драгнева, Лиляна Динева, Николета Кесова, Китка Славкова
Ръководител на групата	Моника Драгнева

A. Обща приложимост и функционален обхват

Оценете доколко моделът покрива реални нужди в публичния сектор.

-  **Инструкция за попълване:** За всеки критерий: поставете кръгче на избраната оценка (1=много слабо, 5=отлично) и запишете кратка обосновка.

Критерий		
A1. Приложимост в публичния сектор		
Обработка на документи доклади, решения, наредби	Оценка 1–5 ○1 ○2 ○3 ●4 ○5	Обосновка / бележки: 4 Сканира ръкописни документи от документи на терен; извежда и разпознава грешки и може да въвежда данни в националните бази и регистри; сравнява данните с държавни изисквания в реално време.
Правен анализ тълкуване на нормативни актове	Оценка 1–5 ○1 ●2 ○3 ○4 ○5	Обосновка / бележки: 2 Не може да тълкува закони, може само да ги систематизира и сравнява; идентифицира потенциални конфликти.
Граждански услуги чатбот, Q&A, помощ при заявления	Оценка 1–5 ○1 ○2 ○3 ●4 ○5	Обосновка / бележки: 4 Интелигентен асистент- chat bot; персонализирано генериране на стъпки за подаване на документи; превод на административни указания на всички езици.

Вътрешна комуникация <i>резюмета, протоколи, кореспонденция</i>	Оценка 1–5 ○1 ○2 ●3 ○4 ○5	Обосновка / бележки:3 Оптимизация на работните процеси в администрацията; служителите могат да задават въпроси за вътрешната уредба и процедури за конкретното ведомство.
Анализ на данни <i>отчети, статистика, мониторинг</i>	Оценка 1–5 ○1 ○2 ○3 ●4 ○5	Обосновка / бележки:4 Моделът може да открива зависимости в големи таблични масиви, анализира огромни количества данни.
Превод <i>многоезична комуникация, ЕС документи</i>	Оценка 1–5 ○1 ○2 ○3 ●4 ○5	Обосновка / бележки:4 Възможност за светкавичен и качествен превод на документи и комуникация с чужди граждани.
Програмиране / автоматизация <i>вътрешни системи, скриптове</i>	Оценка 1–5 ○1 ○2 ○3 ●4 ○5	Обосновка / бележки: 4 Генерира и автоматизира процеси и месечни доклади и бази данни.
Обучение на служители <i>учебни материали, тестове</i>	Оценка 1–5 ○1 ○2 ●3 ○4 ○5	Обосновка / бележки:3 Chat bot за задаване на въпроси като „вътрешен ментор“; възможност за създаване на персонализирани обучителни планове за служители.
A2. Функции на платформата		
Текстово генериране <i>качество и прецизност</i>	Оценка 1–5 ○1 ○2 ○3 ●4 ○5	Обосновка / бележки:4 Автоматично изготвяне на официални документи, заповеди, писма.
Reasoning / разсъждение <i>логически вериги, многостъпкови задачи</i>	Оценка 1–5 ○1 ●2 ○3 ○4 ○5	Обосновка / бележки:2 Има способност за логически анализи и верижно мислене.
Code writing <i>генериране и анализ на код</i>	Оценка 1–5 ○1 ○2 ○3 ●4 ○5	Обосновка / бележки:4 Писане на програмни скриптове; одити и автоматизация на съществуващ софтуерен код в държавните системи.
Анализ на изображения <i>vision, диаграми, скрийншоти</i>	Оценка 1–5 ○1 ●2 ○3 ○4 ○5	Обосновка / бележки:2 Разпознаване на печат и подпис и разчитане на етикети.
Аудио / глас <i>говорим вход/изход</i>	Оценка 1–5 ○1 ○2 ○3 ○4 ●5	Обосновка / бележки:5 Преобразуване на гласове записи от заседания в текст и резюмиране на ключови точки.
Видео разбиране <i>анализ на видеосъдържание</i>	Оценка 1–5 ○1 ○2 ●3 ○4 ○5	Обосновка / бележки:3 Анализ на записи от камери в реални време.
Дълги документи <i>обработка на обемни файлове</i>	Оценка 1–5 ○1 ○2 ○3 ○4 ●5	Обосновка / бележки:5 Анализ на масиви до 2 млн. токена и 1,5 млн. думи и едновременно търсене и сравняване.
Agentic режим	Оценка 1–5 ○1 ○2 ●3 ○4 ○5	Обосновка / бележки:3 Възможност за изпълнение на многостепенни задачи.

автономни многостъпкови задачи		
API / системна интеграция свързване с вътрешни системи	Оценка 1–5 ○1 ○2 ●3 ○4 ○5	Обосновка / бележки:3 Програмна свързаност с национални платформи.
Plugin екосистема налични разширения и конектори	Оценка 1–5 ○1 ○2 ○3 ●4 ○5	Обосновка / бележки:4 Разполага с gemini extentions, интегрират модела директно в работния процес.
Desktop приложение нативна desktop среда	Оценка 1–5 ○1 ○2 ○3 ●4 ○5	Обосновка / бележки:4 Достъпен през web-интерфейс и като интегриран панел в google chrome.
Mobile приложение iOS / Android достъп	Оценка 1–5 ○1 ○2 ●3 ○4 ○5	Обосновка / бележки: Пълно заместване на google асистент за android и интегриран chrome и мобилно приложение gemini app.
Cowork / колаборация екипна работа в реално време	Оценка 1–5 ○1 ●2 ○3 ○4 ○5	Обосновка / бележки: Моделът може да участва в планиране на проекти и да разпределя задачи.
Търсене в интернет актуална информация от мрежата	Оценка 1–5 ○1 ○2 ○3 ●4 ○5	Обосновка / бележки:5 Използва google в реално време и не оперира само с тренировъчните си данни.
Fine-tuning настройка за специфична среда	Оценка 1–5 ○1 ○2 ●3 ○4 ○5	Обосновка / бележки: Потребителите могат да го надстройват с данни, fine-tuning се прави не само на текста, но и на изображения и видео.

B. SWOT Матрица

Обобщете основните силни/слаби страни, възможности и заплахи на модела в публичен контекст.

 **Инструкция за попълване:** Запишете поне 4 наблюдения в клетка. Базирайте се на попълнените оценки от Раздел А.

S STRENGTHS — Силни страни 1. Много голям контекстен прозорец (до 1M tokens) за работа с дълги административни документи. 2. Силни разсъждения и аналитични способности при сложни задачи и правни анализи. 3. Поддържа API, fine-tuning и agentic workflows за интеграция в публични системи. 4. Отлични мултимодални възможности — текст, изображения, аудио, видео и документи. 5. Преодоляване на езикова бариера. 6. Интелигентен асистент; автоматизира процеси.	W WEAKNESSES — Слаби страни 1. Зависимост от Google Cloud инфраструктура и интернет свързаност. 2. Лошо качество на входящите данни в администрациите. 3. Технологична пропаст сред служителите. 4. Липса на гъвкавост при извънредни ситуации.
---	--

Допълнителни бележки: Подходящ за мащабни администрации и обработка на големи информационни потоци.	5. Въвеждането на AI може да доведе атрофия на критичното мислене на служителите. 6. Риск от халюцинации при правни или административни казуси. Допълнителни бележки: Необходим е човешки контрол при критични решения.
O OPPORTUNITIES — Възможности 1. Подобряване на гражданските услуги чрез AI чатботи и виртуални асистенти. 24/7 дигитален асистент за граждани. 2. Намаляване на времето за обработка на сигнали, заявления и справки. 3. Елиминиране на корупцията. 4. Повишаване на капацитета на служителите. 5. Автоматизация на административни процеси и документооборот. 6. Подкрепа за многоезична комуникация и превод на европейски документи. Допълнителни бележки: Има потенциал за дигитална трансформация на публичния сектор.	T THREATS — Заплахи 1. Vendor lock-in и зависимост от един технологичен доставчик. 2. Постоянна нужда от свързаност - при спиране на интернет всичко спира. 3. Риск от изтичане или неправилна обработка на чувствителни данни. 4. Регулаторни ограничения по EU AI Act и GDPR. населението при грешна прогноза. 6. Киберрискове и възможни атаки чрез AI системи. Допълнителни бележки: Необходимо е ясно управление на риска и политики за сигурност.

C. Разширен SWOT по специфични теми

 Инструкция за попълване: Попълнете текстово наблюденията си за всяка тема. Използвайте конкретни примери.
Цена и лицензиране Наблюдения: (Модел на плащане, достъпност за малки общини, API разходи) Внедряването е инвестиция, която се самоизплаща чрез спестено време, но изисква първоначален бюджет. Гъвкавост на разходите-моделът позволява плащане само на реално използвано количество токени. Препоръка за публичния сектор: Най-рентабилно е внедряването през Veerex AI API, което оптимизира разходите спрямо натоварването; задължително ползване на сървъри в ЕС.
Сигурност на данните Наблюдения: (Обработка на лични данни, GDPR, съхранение в облака) Gemini не използва вашите данни за обучение на моделите си. Данните са криптирани в покой и при трансфер. Данните са сигурни ако се използва институционален лиценз. Препоръка за публичния сектор: Забрана на безплатни потребителски версии за служебни цели; настройка на суверенен облак.
Езикова поддръжка

Наблюдения: <i>(Качество на български език, юридическа терминология)</i> Пълно разбиране на специфична терминология, мултиезичност и гласови функции.
Препоръка за публичния сектор: Моделът е напълно готов за работа в публичната администрация без допълнителни препоръки и настройки.
Vendor lock-in
Наблюдения: <i>(Зависимост от доставчик, алтернативи, преносимост)</i> Дълбоката интеграция на Gemini в Google прави смяна на доставчика трудна и скъпа поради специфични API протоколи.
Препоръка за публичния сектор: Използване на софтуерни посредници
Сложност на внедряване
Наблюдения: <i>(Техническа зрялост, необходима инфраструктура, IT капацитет)</i> Ниска сложност, подходящо за ежедневна работа в публичния сектор.
Препоръка за публичния сектор: Обучения на служителите за работа с модела.
Регулаторно съответствие
Наблюдения: <i>(EU AI Act, NIS2/3KC, GDPR, национални наредби)</i> Задължителен човешки надзор и пълна прозрачност пред гражданите.
Препоръка за публичния сектор: Съответствие с изискванията на ЕГОВ и задължителен човешки надзор.
Етика и прозрачност
Наблюдения: <i>(Обяснимост, одитируемост, отговорност при грешки)</i> Безпристрасност; Моделът показва логиката и законовата база.
Препоръка за публичния сектор: Гражданите трябва да знаят кога разговарят с ИИ; всяко решение с правни последици се взима и подписва от служител.
Изкривявания (Bias)
Наблюдения: <i>(Пристрастия при различни групи, равнопоставеност)</i> Риск от халюцинации и алгоритмични пристрастия.
Препоръка за публичния сектор: Технически филтри; проверка от човек.
Халюцинации / грешки
Наблюдения: <i>(Честота, detectability, риск при вземане на решения)</i> Не могат да бъдат премахнати на сто процента, но рискът се свежда до нива по-ниски от човешка грешка.

Препоръка за публичния сектор:
Технически филтри; проверка от човек.

Морал и ценности

Наблюдения: *(Поведение при чувствителни теми, отказ на вредни задачи)*
Не е способен на емпатия; морално недопустимо е да се разчита на ИИ за решения, които изискват дългогодишен интуитивен опит.

Препоръка за публичния сектор:
Технологията трябва да служи на човека, а не да го управлява.

D. Обобщена оценка на групата

Обобщена оценка по основни показатели		
Обща приложимост <i>средна от Раздел А</i>	Оценка 1–5 ○1 ○2 ○3 ●4 ○5	Обосновка / бележки:2 Висок потенциал за автоматизация.
Готовност за внедряване <i>техническа и организационна</i>	Оценка 1–5 ○1 ○2 ●3 ○4 ○5	Обосновка / бележки:3 Технологията е готова, всичко зависи от бюджета на институцията.
Съответствие с регулации <i>AI Act, GDPR, NIS2</i>	Оценка 1–5 ●1 ○2 ○3 ○4 ○5	Обосновка / бележки:1 Съответствието на gemini с регулационната рамка в България и ЕС е на много високо ниво.
Етичен риск <i>обратна скала: 5=нисък риск</i>	Оценка 1–5 ●1 ○2 ○3 ○4 ○5	Обосновка / бележки:1 Висока чувствителност на данните.
Обща препоръка <i>1=не препоръчваме, 5=силно препоръчваме</i>	Оценка 1–5 ○1 ○2 ○3 ○4 ●5	Обосновка / бележки:5 Внедряването е силно препоръчително в публичната администрация, но трябва да се използва като помощник на експертите, а не като автоматизиран заместител.
Мотивирано становище на групата (мин. 5 изречения): <i>Обобщете защо препоръчвате или не препоръчвате модела за публичния сектор.</i> Внедряването на Gemini в публичния сектор, е стратегически обосновано поради уникалната способност на модела да обработва огромни масиви от информация с висока скорост и логическа прецизност. Моделът не само намалява административната тежест върху служителите, но и драстично подобрява качеството на обслужването на гражданите чрез предоставяне на 24/7 достъпна информация на достъпен език. В икономически план, гъвкавото лицензиране и потенциалът за автоматизация превръщат Gemini в инструмент с висока възвръщаемост на инвестицията, осигурявайки технологичен суверинитет и модерно управление на публичните ресурси.		
Предложени условия за ползване: <i>Какви ограничения, контроли или процедури трябва да съпътстват внедряването?</i>		

За да бъде внедрен Gemini в публичния сектор не само ефективно, но и законосъобразно, е необходимо установяването на строга рамка от правила. Тези условия за ползване трябва да залегнат във вътрешните заповеди за организация на работата.

Контроли- нито едно решение, генерирано от ИИ, няма правна стойност, докато не бъде прегледано, потвърдено и подписано от служител.

Risk Canvas

Картографиране на рисковете от AI модел в публичен контекст

v1.0 — Работен шаблон

Инструкция за попълване: Попълнете всички полета за избория от вашата група AI модел. Празните редове и скалите са предназначени за ръчно попълване. Не изтривайте инструкциите в курсив.

Наименование на модела	Пример: Claude Opus 4.7 / GPT-5.4 / Gemini 3.1 Pro Gemini 2.5 Pro
Разработчик / компания	Пример: Anthropic / OpenAI / Google DeepMind Google Gemini
Версия за оценка	Точната версия, която групата изследва Gemini – 2.5 Pro – Preview – 09-2025
Дата на попълване	ДД.ММ.ГГГГ 14.05.2026
Членове на групата	Имена на участниците Михаела Неделчева, Петя Дошкова, Моника Драгнева, Лиляна Динева, Николета Кесова, Китка Славкова
Ръководител на групата	Моника Драгнева

Легенда за скалите

Вероятност (Вер.) 1 = Изключително рядко 2 = Рядко 3 = Умерено вероятно 4 = Вероятно 5 = Почти сигурно	Въздействие (Вл.) 1 = Незначително 2 = Малко 3 = Умерено 4 = Значително 5 = Критично	Ниво на риска Н = Нисък (Вер.×Вл. ≤ 4) С = Среден (5–12) В = Висок (13–25)
--	--	--

Риск Канва — 9-полева матрица

Попълнете всяко поле с идентифицирани рискове, засегнати страни и ключови наблюдения.

Инструкция за попълване: Всяко от 9-те полета представлява различна категория риск. Запишете конкретни рискове — не общи фрази.

1. Технически рискове	2. Рискове за данните	3. Правни / регулаторни EU AI Act, NIS2, GDPR, национално право
-----------------------	-----------------------	--

<i>Халюцинации, грешки, downtime, версионни промени</i> - Настъпване на паника при населението при сгрешен епидемиологичен анализ или сгрешени лабораторни данни. - Грешки при обработка на изображения, PDF файлове и мултимодален вход. - Халюцинации и неточни отговори при правни или административни казуси.	Поверителност, GDPR, изтичане, съхранение - Чувствителни медицински данни, технологична тайна - Риск от изтичане на чувствителна или служебна информация.	- Особено големи финансови последици и юридическа отговорност при сгрешен нормативен акт, написан с ИИ. - Нарушения на GDPR при неправилна обработка на лични данни.
4. Етични рискове <i>Bias, дискриминация, непрозрачност, автономия</i> - Сериозен риск от дискриминация. - Непрозрачност относно ролята на AI в административните процеси. - Риск от генериране на подвеждаща или манипулативна информация. - Прекомерна автоматизация и намаляване на човешкия контрол.	5. Оперативни рискове <i>Зависимост, интеграция, обучение, процеси</i> - Силна зависимост от интернет, сериозна заплаха от хакери. - Зависимост от интернет и Google Cloud инфраструктура. - Недостатъчен технически капацитет в публичната администрация. - Трудности при интеграция със стари системи.	6. Репутационни рискове <i>Доверие на гражданите, медийно отразяване</i> - Подкопаване на авторитета на институцията, като източник на достоверна информация. - Медийни скандали при видима AI грешка. - Негативна обществена реакция срещу автоматизирани решения. - Репутационни щети при изтичане на данни или дискриминационни резултати.
7. Стратегически рискове <i>Vendor lock-in, ценови промени, спиране на услугата</i> - Възможност за рязко поскъпване на услугата. - Ограничен контрол върху развитието на модела. - Риск при геополитически ограничения или прекъсване на услугата.	8. Засегнати страни <i>Граждани, служители, ръководство, партньори</i> - Граждани, използващи електронни административни услуги. - Държавни служители и IT екипи. - Ръководство на институциите. - Външни партньори и доставчици на системи.	9. Остатъчни рискове <i>Рискове след митигация, приети рискове</i> - Скрити халюцинации - Модел на отклонение - Риск от злонамерени кибератаки - Прекомерно доверие - Остатъчен риск от човешки грешки при използване на AI системата.

Gemini предоставя много силни multimodal и аналитични възможности, които могат значително да подобрят ефективността на публичната администрация. Моделът е особено подходящ за обработка на документи, анализ на данни и автоматизация на административни процеси. Въпреки това съществуват съществени рискове, свързани с поверителността на данните, прозрачността и зависимостта от външен доставчик. Необходимо е внедряване на строг човешки контрол, политики за сигурност и механизми за

одит. Използването му следва да бъде ограничено за високорискови административни решения без човешка намеса.

Рисков регистър

Попълнете по всяка категория конкретните рискове с вероятност, въздействие и мерки.

 **Инструкция за попълване:** Вер. = Вероятност (1–5) · Вл. = Въздействие (1–5) · Ниво: ☐Н (нисък) ☐С (среден) ☐В (висок)

Риск / Събитие	Вероятност	Въздействие	Ниво	Мярка за митигация
Технически рискове				
Халюцинации и фактически грешки при отговори	Вер. ○1 ○2 ●3 ○4 ○5	Вл. ○1 ○2 ○3 ○4 ●5	Ниво ☐Н ☐С ☒В	Мярка за митигация: Задължителна човешка проверка и използване само като помощен инструмент
Недостъпност / downtime на платформата	Вер. ○1 ●2 ○3 ○4 ○5	Вл. ○1 ○2 ○3 ○4 ●5	Ниво ☐Н ☐С ☐В	Мярка за митигация: Подписване на корпоративен договор с гарантирана наличност.
Внезапна промяна на версия или поведение на модела	Вер. ○1 ●2 ○3 ○4 ○5	Вл. ○1 ○2 ○3 ○4 ●5	Ниво ☐Н ☐С ☐В	Мярка за митигация: Offline-контрол, поддържане на локални pdf шаблони и инструкции за преминаване към „Ръчен режим“.
Ограничения на контекстния прозорец при обемни документи	Вер. ○1 ○2 ●3 ○4 ○5	Вл. ○1 ○2 ○3 ●4 ○5	Ниво ☐Н ☐С ☐В	Мярка за митигация: Разделяне на документите и структурирана обработка.
Грешки при мултимодален вход (изображения, аудио)	Вер. ○1 ○2 ●3 ○4 ○5	Вл. ○1 ○2 ○3 ●4 ○5	Ниво ☐Н ☐С ☐В	Мярка за митигация: Допълнителна валидация от служители.
Рискове за данните и поверителността				
Обработка на лични данни на граждани в облачна среда	Вер. ○1 ○2 ●3 ○4 ○5	Вл. ○1 ○2 ○3 ○4 ●5	Ниво ☐Н ☐С ☒В	Мярка за митигация: Данните се заключват в сървъри; автоматизирани филтри за скриване на лична информация
Изтичане на класифицирана/служебна информация	Вер. ○1 ○2 ●3 ○4 ○5	Вл. ○1 ○2 ○3 ○4 ●5	Ниво ☐Н ☐С ☒В	Мярка за митигация: Внедряване на дигитална стена
Запазване на запитвания от доставчика за обучение	Вер. ○1 ○2 ●3 ○4 ○5	Вл. ○1 ○2 ○3 ○4 ●5	Ниво ☐Н ☐С ☐В	Мярка за митигация: Vertex AI – в договора е заложено, че данните на клиента са изолирани.
Трансфер на данни извън ЕС (EU AI Act / Schrems II)	Вер. ○1 ○2 ○3 ●4 ○5	Вл. ○1 ○2 ○3 ●4 ○5	Ниво ☐Н ☐С ☒В	Мярка за митигация: Използване на EU datacenter и SCC договори.
Липса на одитна следа при автоматизирани решения	Вер. ○1 ○2 ●3 ○4 ○5	Вл. ○1 ○2 ○3 ●4 ○5	Ниво ☐Н ☐С ☐В	Мярка за митигация: Всяко предложение на ИИ се записва с уникален код, свързан с конкретна преписка.

Правни и регулаторни рискове				
Несъответствие с EU AI Act (Annex III — публичен сектор)	Вер. ○1 ○2 ●3 ○4 ○5	Вл. ○1 ○2 ○3 ○4 ●5	Ниво □Н □С •В	Мярка за митигация: Оценка на въздействието върху фундаменталните права.
Нарушение на GDPR при обработка на лични данни	Вер. ○1 ○2 ●3 ○4 ○5	Вл. ○1 ○2 ○3 ○4 ●5	Ниво □Н □С •В	Мярка за митигация: Внедряване на филтри, които автоматично анонимизират лични данни.
Липса на обяснимост при автоматизирани административни решения	Вер. ○1 ○2 ○3 ●4 ○5	Вл. ○1 ○2 ○3 ●4 ○5	Ниво □Н □С •В	Мярка за митигация: Human-in-the-loop и документация.
Авторскоправни нарушения в генерирано съдържание	Вер. ○1 ○2 ●3 ○4 ○5	Вл. ○1 ○2 ●3 ○4 ○5	Ниво □Н •С □В	Мярка за митигация: Проверка на съдържанието преди публикуване
Отговорност при грешни правни или медицински изходи	Вер. ○1 ○2 ●3 ○4 ○5	Вл. ○1 ○2 ○3 ○4 ●5	Ниво □Н □С •В	Мярка за митигация: Забрана за автономни крайни решения.
Етични и социални рискове				
Систематичен bias спрямо определени групи граждани	Вер. ○1 ○2 ●3 ○4 ○5	Вл. ○1 ○2 ○3 ●4 ○5	Ниво □Н •С □В	Мярка за митигация: Bias testing и периодичен одит.
Дискриминация при автоматизирано обработване на заявления	Вер. ○1 ○2 ●3 ○4 ○5	Вл. ○1 ○2 ○3 ○4 ●5	Ниво □Н □С •В	Мярка за митигация: Човешки надзор върху решенията.
Непрозрачност — граждани не знаят, че AI взима решения	Вер. ○1 ○2 ○3 ●4 ○5	Вл. ○1 ○2 ●3 ○4 ○5	Ниво □Н •С □В	Мярка за митигация: Ясно обозначаване на AI системите.
Дезинформация и манипулация чрез убедително съдържание	Вер. ○1 ○2 ○3 ●4 ○5	Вл. ○1 ○2 ○3 ●4 ○5	Ниво □Н □С •В	Мярка за митигация: Модериране на съдържанието и верификация
Загуба на критично мислене и зависимост при служителите	Вер. ○1 ○2 ○3 ●4 ○5	Вл. ○1 ○2 ●3 ○4 ○5	Ниво □Н •С □В	Мярка за митигация: Обучения и вътрешни политики.
Оперативни рискове				
Липса на технически капацитет за внедряване и поддръжка	Вер. ○1 ○2 ○3 ●4 ○5	Вл. ○1 ○2 ○3 ●4 ○5	Ниво □Н □С •В	Мярка за митигация: Специализирани AI екипи и външни експерти.
Vendor lock-in — прекомерна зависимост от доставчик	Вер. ○1 ○2 ○3 ○4 ●5	Вл. ○1 ○2 ○3 ●4 ○5	Ниво □Н □С •В	Мярка за митигация: Multi-vendor стратегия и отворени стандарти.

Недостатъчно обучение на служителите	Вер. ○1 ○2 ○3 ●4 ○5	Вл. ○1 ○2 ●3 ○4 ○5	Ниво ☐ Н ·С ☐ В	Мярка за митигация: Регулярни обучения и сертификация.
Интеграционни проблеми с наследени (legacy) системи	Вер. ○1 ○2 ○3 ●4 ○5	Вл. ○1 ○2 ○3 ●4 ○5	Ниво ☐ Н ☐ С ·В	Мярка за митигация: API middleware и поетапна интеграция.
Ценови скокове при промяна на лицензионния модел	Вер. ○1 ○2 ○3 ●4 ○5	Вл. ○1 ○2 ○3 ●4 ○5	Ниво ☐ Н ☐ С ·В	Мярка за митигация: Дългосрочни договори и бюджетен резерв.
Репутационни и стратегически рискове				
Публичен скандал при видима AI грешка (медийно отразяване)	Вер. ○1 ○2 ●3 ○4 ○5	Вл. ○1 ○2 ○3 ○4 ●5	Ниво ☐ Н ☐ С ·В	Мярка за митигация: Процедури за управление на кризи
Загуба на доверие на гражданите в публичните институции	Вер. ○1 ○2 ○3 ●4 ○5	Вл. ○1 ○2 ○3 ○4 ●5	Ниво ☐ Н ☐ С ·В	Мярка за митигация: Прозрачност и обществена отчетност.
Политическа съпротива срещу използването на AI	Вер. ○1 ○2 ●3 ○4 ○5	Вл. ○1 ○2 ○3 ●4 ○5	Ниво ☐ Н ·С ☐ В	Мярка за митигация: Пилотни проекти и обществен диалог.
Зависимост от американски/китайски технологичен доставчик	Вер. ○1 ○2 ●3 ○4 ○5	Вл. ○1 ○2 ○3 ○4 ●5	Ниво ☐ Н ☐ С ·В	Мярка за митигация: Европейски облачни алтернативи и решения за архивиране.
Спиране на услугата при геополитически напрежения	Вер. ○1 ○2 ●3 ○4 ○5	Вл. ○1 ○2 ○3 ○4 ●5	Ниво ☐ Н ☐ С ·В	Мярка за митигация: Планове за непрекъсваемост и резервни AI решения.

Обобщена рискова оценка на групата

Топ 3 критични риска за модела: - Изтичане или неправомерна обработка на лични и служебни данни. - Халюцинации и грешни AI отговори при административни и правни процеси. - Vendor lock-in и силна зависимост от Google Cloud инфраструктура и лицензионен модел.
Обща рискова оценка на модела за публичния сектор: ☐ Нисък риск — може да се внедри с минимални ограничения · Среден риск — изисква допълнителни контроли и процедури ☐ Висок риск — изисква задълбочена оценка и специфична регулация ☐ Неприемлив риск — не се препоръчва за публичния сектор
Мотивирано становище: Gemini 3.1 Pro е мощен multimodal AI модел с висока приложимост в публичния сектор, особено при обработка на документи, анализ на данни, автоматизация и граждански услуги. Моделът предлага значителни предимства чрез голям контекстен прозорец, силни reasoning способности и интеграция с Google екосистемата. Въпреки това използването му създава съществени

рискове, свързани с поверителността на данните, прозрачността на алгоритмите и зависимостта от външен cloud доставчик. Особено критични са рисковете при обработка на лични данни и възможността от генериране на неточни или подвеждащи административни отговори. Поради това внедряването му следва да бъде съпроводено с човешки контрол, строги политики за сигурност, правен надзор и процедури за одит и отчетност. Моделът е подходящ за подпомагащи и аналитични функции, но не се препоръчва за напълно автономни административни решения.

Карта за приложимост

AI модел в публичната администрация — функционален анализ

v1.0 — Работен шаблон

 **Инструкция за попълване:** Попълнете всички полета за избрания от вашата група AI модел. Празните редове и скалите са предназначени за ръчно попълване. Не изтривайте инструкциите в курсив.

Наименование на модела	Пример: Claude Opus 4.7 / GPT-5.4 / Gemini 3.1 Pro Gemini 2.5 Pro
Разработчик / компания	Пример: Anthropic / OpenAI / Google DeepMind Google Gemini
Версия за оценка	Точната версия, която групата изследва Gemini – 2.5 Pro – Preview – 09-2025
Дата на попълване	ДД.ММ.ГГГГ 14.05.2026
Членове на групата	Имена на участниците Михаела Неделчева, Петя Дошкова, Моника Драгнева, Лиляна Динева, Николета Кесова, Китка Славкова
Ръководител на групата	Моника

Част 1. Обзорна матрица на приложимостта

Бърза оценка по административни функции. Прил. = Приложимост 1–5. Риск = Н/С/В.

 **Инструкция за попълване:** Попълнете всяка клетка: поставете оценка за приложимост (1=неприложимо, 5=напълно приложимо), ниво на риск (Н/С/В) и кратка бележка.

Административна функция	Обработка на документи	Правен анализ	Граждански услуги	Анализ на данни	Програмиране	Обучение
Обработка на кореспонденция	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В	Прил.: ○1 ○2 ●3 ○4 ○5 Риск: □Н ■С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н □С ■В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н □С ■В	Прил.: ○1 ○2 ●3 ○4 ○5 Риск: □Н □С ■В	Прил.: ○1 ○2 ●3 ○4 ○5 Риск: □Н ■С □В
Изготвяне на доклади и анализи	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В	Прил.: ○1 ○2 ●3 ○4 ○5 Риск: □Н □С ■В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: ■Н □С □В	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: □Н ■С □В	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: □Н ■С □В	Прил.: ○1 ○2 ●3 ○4 ○5 Риск: □Н ■С □В

Тълкуване на нормативни актове	Прил.: ○1 ○2 ●3 ○4 ○5 Риск: □Н ■С □В	Прил.: ○1 ○2 ●3 ○4 ○5 Риск: □Н ■С □В	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: □Н ■С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н □С ■В	Прил.: ○1 ○2 ●3 ○4 ○5 Риск: □Н □С ■В	Прил.: ○1 ○2 ●3 ○4 ○5 Риск: □Н □С ■В
Изготвяне на правни становища	Прил.: ○1 ○2 ●3 ○4 ○5 Риск: □Н ■С □В	Прил.: ○1 ●2 ○3 ○4 ○5 Риск: □Н □С ■В	Прил.: ○1 ○2 ●3 ○4 ○5 Риск: □Н ■С □В	Прил.: ○1 ○2 ●3 ○4 ○5 Риск: □Н □С ■В	Прил.: ○1 ○2 ●3 ○4 ○5 Риск: □Н ■С □В	Прил.: ○1 ○2 ●3 ○4 ○5 Риск: □Н ■С □В
Отговори на граждански запитвания	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В	Прил.: ○1 ○2 ●3 ○4 ○5 Риск: □Н □С ■В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н □С ■В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В
Чатбот за електронни услуги	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В	Прил.: ○1 ●2 ○3 ○4 ○5 Риск: □Н ■С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н □С ■В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н □С ■В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В
Статистически анализ	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н □С ■В
Обработка на отчети	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: ■Н □С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: ■Н □С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: ■Н □С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: ■Н □С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: ■Н □С □В
Автоматизация на работни процеси	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В	Прил.: ○1 ○2 ●3 ○4 ○5 Риск: □Н ■С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н □С ■В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н □С ■В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В
Разработка на вътрешни инструменти	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В	Прил.: ○1 ○2 ●3 ○4 ○5 Риск: □Н ■С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В
Учебни материали и тестове	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В
Наръчници и процедури	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: ■Н □С □В	Прил.: ○1 ○2 ●3 ○4 ○5 Риск: □Н □С ■В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: ■Н □С □В	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: □Н ■С □В	Прил.: ○1 ○2 ●3 ○4 ○5 Риск: □Н ■С □В
Превод и локализация	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: □Н □С ■В	Прил.: ○1 ○2 ●3 ○4 ○5 Риск: □Н □С ■В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: ■Н □С □В	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: □Н ■С □В	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: □Н □С ■В	Прил.: ○1 ○2 ○3 ●4 ○5 Риск: □Н ■С □В
Протоколи и резюмета	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н □С ■В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В	Прил.: ○1 ○2 ○3 ○4 ●5 Риск: □Н ■С □В

Част 2. Детайлен анализ по случаи на употреба

За всеки сценарий оценете конкретно как избраният модел се справя.

2.1. Обработка на административни документи

Способността на модела да чете, структурира, резюмира и изготвя административни документи — решения, наредби, докладни записки, протоколи.

Оценка за избрания модел:

1. Може ли моделът да обработва документи на български език с административна терминология?
2. Поддържа ли обработка на PDF/Word файлове с дълъг обем?
3. Запазва ли форматирането и структурата при генериране на изходен документ?
4. Изисква ли специализирано prompt инженерство за административен контекст?

Отговори / наблюдения:

1. Може да обработва документи на български език, но не следва изцяло специфичната българска административна терминология (понякога генерира твърде буквални преводи от английски).

2. Поддържа обработка на големи файлове (до 1500 страници благодарение на контекстния прозорец от 2 милиона токена).

3. Частично запазва сложни таблични форматирания и специфични държавни бланки; изисква последваща ръчна корекция на структурата.

4. Изисква специализирано prompt инженерство (Few-Shot Prompting с реални примери от Държавен вестник) за постигане на официален административен стил.

Препоръка:

☐ Препоръчва се без ограничения

☒ Препоръчва се с ограничения (опишете):

Да се използва основно за резюмиране, класификация и изготвяне на първоначални чернови (драфтове). Всеки документ задължително трябва да преминава през редакция от служител („Human-in-the-loop“), преди да бъде придвижен по официален път.

☐ Не се препоръчва (обосновка):

2.2. Правен анализ и тълкуване

Способността да анализира нормативни актове, да открива противоречия, да прилага правни норми към конкретни случаи.

Оценка за избрания модел:

1. Колко точен е при тълкуване на български законодателни текстове?
2. Може ли да следи промени в нормативната база (обновени закони)?
3. Как се справя с многозначност и правна неяснота?
4. Маркира ли ясно когато излиза извън областта на своята компетентност?

Отговори / наблюдения:

1. Показва задоволителна точност при базови текстове, но изпитва затруднения при препратки между различни български кодекси и закони.

2. Тъй като базата данни е ограничена до септември 2025 г. (за тази preview версия), моделът **не може** да следи промените в реално време, освен ако новите закони не му бъдат подадени директно в прозореца за контекст (чрез RAG).

3. При многозначност моделът има склонност да избира най-популярната интерпретация, като понякога халюцинира несъществуваща съдебна практика.

4. Рядко маркира сам липсата на компетентност; склонен е да отговаря уверено дори когато греши.

Препоръка:

☐ Препоръчва се без ограничения

☐ Препоръчва се с ограничения (опишете):

■ **Не се препоръчва (обосновка):**

Рискът от грешно тълкуване на нормативната уредба в публичния сектор може да доведе до нищожност на административните актове. Може да се ползва само като помощен инструмент от правоспособни юристи за бързо търсене в текст, но не и самостоятелно.

2.3. Взаимодействие с граждани (чатбот / асистент)

Способността да комуникира с граждани — отговаря на въпроси, помага при подаване на заявления, обяснява административни процедури.

Оценка за избрания модел:

1. Разбира ли естествен разговорен български (включително диалекти и грешки)?
2. Дава ли ясни и разбираеми отговори без административен жаргон?
3. Как се справя с чувствителни теми (социални помощи, здраве, правосъдие)?
4. Може ли да се интегрира с портали за електронни услуги?

Отговори / наблюдения:

1. Отлично разбира разговорен български, улавя печатни грешки, жаргон и контекст.
2. Успешно „превежда“ тежкия административен стил на достъпен за гражданите език.
3. При чувствителни теми (социални помощи, здраве) отговаря твърде общо поради вградените филтри за безопасност на Google, което понякога фрустрира потребителя.
4. Може да се интегрира успешно с уеб портали чрез REST API.

Препоръка:

☐ Препоръчва се без ограничения

■ Препоръчва се с ограничения (опишете):

Подходящ за първа линия на информационно обслужване (FAQ чатбот), който насочва гражданите към правилните линкове и бланки, но без да взема решения и без да дава персонални правни съвети.

☐ Не се препоръчва (обосновка):

2.4. Анализ на данни и отчетност

Способността да обработва структурирани данни, да генерира отчети и да идентифицира тенденции.

Оценка за избрания модел:

1. Може ли да обработва Excel/CSV данни и да генерира анализи?
2. Поддържа ли визуализация на данни (графики, диаграми)?
3. Колко надежден е при числови изчисления и статистика?
4. Може ли да се свърже с вътрешни бази данни чрез API?

Отговори / наблюдения:

1. Справя се много добре с Excel и CSV файлове, като успешно извежда статистически тенденции и аномалии.
2. Моделът поддържа генериране на код за графики (на Python), но за директна визуализация е нужна интеграция с външна платформа (например Google Looker Studio или PowerBI).
3. Математическата надеждност е висока, когато използва вградения си код-интерпретатор, но при чисти текстови изчисления все още прави грешки.
4. Интеграцията с вътрешни SQL бази данни е напълно възможна чрез защитени API слоеве.

Препоръка:

☐ Препоръчва се без ограничения

■ Препоръчва се с ограничения (опишете):

Изключително полезен за аналитичните звена, под условие че анализирания масиви от данни са предварително анонимизирани (без ЕГН, имена и адреси на граждани).

☐ Не се препоръчва (обосновка):

2.5. Вътрешно обучение на служители

Способността да генерира учебни материали, тестове, наръчници и да поддържа интерактивно обучение.

Оценка за избрания модел:

1. Може ли да генерира тематични курсове и тестове за служители?
2. Адаптира ли съдържанието към различни нива на компетентност?
3. Поддържа ли формати за дистанционно обучение (SCORM и др.)?
4. Как се справя с чувствителни теми — антикорупция, етика, дискриминация?

Отговори / наблюдения:

1. Отличен инструмент; генерира подробни учебни модули, казуси за изпити и забавни интерактивни викторини за нуждите на администрацията.
2. Лесно адаптира езика от "начинаещ държавен служител" до "експерт".
3. Не генерира директно готови SCORM пакети, но може да подготви суровия текст и XML/JSON структурата за тях.
4. Справя се отлично с теми като антикорупция и етика, следвайки официалните етични кодекси на администрацията.

Препоръка:

- Препоръчва се без ограничения

Това е най-безопасната и високопродуктивна зона за внедряване на модела. Рискът от халюцинации тук е минимален и лесно се контролира от обучителите.

☐ Препоръчва се с ограничения (опишете):

☐ Не се препоръчва (обосновка):

2.6. Програмиране и автоматизация

Способността да пише код, да автоматизира процеси и да поддържа вътрешни системи.

Оценка за избрания модел:

1. Поддържа ли езиците и технологиите, използвани в административните системи?
2. Може ли да чете и документира съществуващ (legacy) код?
3. Колко надеждни са генерираните скриптове без допълнителна проверка?
4. Поддържа ли agentic режим за автономно изпълнение на задачи?

Отговори / наблюдения:

1. Поддържа всички основни езици, използвани в електронното управление (Java, Python, C#, PHP, JavaScript).
2. Изключително полезен за рефакторизиране и документиране на стари (legacy) системи, които нямат актуална документация.

3. Кодът е надежден в около 85% от случаите; изисква задължителен Code Review и тестване в изолирана среда (Sandbox).
4. Поддържа съвременни функции за извикване на инструменти (Function Calling), което позволява изграждане на автономни агенти.

Препоръка:

☐ Препоръчва се без ограничения

■ Препоръчва се с ограничения (опишете):

Да се ползва само като асистент на ИТ експертите в държавните ведомства. Забранено е директното внедряване на генериран от AI код в продукционна среда без човешка автентификация.

☐ Не се препоръчва (обосновка):

Част 3. Оценка на интеграционния потенциал

 **Инструкция за попълване:** Оценете техническите възможности за интеграция с административната инфраструктура.

Интеграционен аспект	Описание	Поддръжка	Бележки
API свързване	REST API, GraphQL, webhooks — интеграция с вътрешни системи	■ Да ☐ Не ☐ Частично	Налична е пълна поддръжка чрез Google Vertex AI API с криптиране на трафика.
Single Sign-On	Съвместимост с Active Directory / ADFS / Keycloak	■ Да ☐ Не ☐ Частично	Чрез Google Cloud Identity платформата лесно се свързва с държавния Keycloak/AD.
On-premise deployment	Възможност за локален хостинг без облачна зависимост	☐ Да ■ Не ☐ Частично	Моделът е проприетарен и се хоства единствено в облачната инфраструктура на Google.
Данни в ЕС	Сървъри на европейска територия — GDPR съответствие	■ Да ☐ Не ☐ Частично	Възможно е избиране на регион за пълно съответствие
Одитна следа	Логирание на всички взаимодействия за отчетност	■ Да ☐ Не ☐ Частично	Интегрира се с Google Cloud Audit

			Logs – записва кой служител какво е подал.
Ролеви достъп	RBAC — различни нива на достъп за различни служители	☒ Да ☐ Не ☐ Частично	Нативно управление чрез IAM политики (администратор, редактор, viewer).
Многоезичност	Поддръжка на български и официални ЕС езици	☒ Да ☐ Не ☐ Частично	Отлично ниво на езиковия модел за всички 24 официални езика на ЕС.
Мобилен достъп	Работа от мобилни устройства (field work)	☐ Да ☐ Не ☒ Частично	Зависи от разработването на конкретно мобилно приложение от страна на ведомството.
Офлайн режим	Функционалност без постоянна интернет връзка	☐ Да ☒ Не ☐ Частично	Изисква постоянна и бърза връзка с облака, за да функционира.
SLA и поддръжка	Гарантирано ниво на обслужване за публичния сектор	☐ Да ☐ Не ☒ Частично	Стандартният Enterprise SLA предлага 99.9% наличност, но няма специфична "държавна" BG поддръжка.

Обобщена карта за приложимост

Топ 3 приложения с висок потенциал:

1. **Автоматизирано резюмиране и класификация на входяща преписка и дълги доклади** (оптимизира времето на деловодителите благодарение на огромния си контекст).

2. **Генератор на обучителни курсове, тестове и интерактивни наръчници** за вътрешно повишаване на квалификацията на държавните служители.

3. **Помощник-инструмент (Copilot) за ИТ отделите** при документиране на legacy код и автоматизация на рутинни процеси (скриптове).

Функции, при които моделът НЕ е подходящ:

1. Автономно тълкуване на закони и вземане на правно обвързващи решения (риск от халюцинации и грешни правни изводи).
2. Директна обработка на документи, съдържащи класифицирана информация или сурови лични данни (КРД/GDPR) без междинна анонимизация.
3. Критични системи, изискващи 100% офлайн работа или локален (On-premise)

Препоръчани условия за пилотно внедряване:

1. Внедряването трябва да започне задължително от **вътрешните (нетоксични) процеси** – например Института по публична администрация (ИПА) за дигитално обучение на служители.
2. Да се използва **само ЕС регион** за Vertex AI, за да се гарантира, че данните не напускат съюза.
3. Всички промптове трябва да бъдат предварително тествани и заключени в шаблони, за да се избегне „Prompt Injection“ от външни потребители.
4. Да се спазва стриктно правилото **„Човекът контролира процеса“ (Human-in-the-loop)** – моделът само предлага, служителят одобрява и подписва.

Методологични указания

Етично и прозрачно ползване на AI модели в публичната администрация

v1.0 — Работен шаблон

Инструкция за попълване: Попълнете всички полета за избория от вашата група AI модел. Празните редове и скалите са предназначени за ръчно попълване. Не изтривайте инструкциите в курсив.

аименование на модела	Пример: Claude Opus 4.7 / GPT-5.4 / Gemini 3.1 Pro Gemini, 2.5 Pro
Разработчик / компания	Пример: Anthropic / OpenAI / Google DeepMind Google Gemini
Версия за оценка	Точната версия, която групата изследва Gemini – 2.5 Pro – Preview – 09 - 2025
Дата на попълване	ДД.ММ.ГГГГ 14.05.2026 г.
Членове на групата	Имена на участниците Михаела Неделчева, Петя Дошкова, Моника Драгнева, Лиляна Динева, Николета Кесова, Китка Славкова
Ръководител на групата	Моника

Предговор и обхват

Инструкция за попълване: Попълнете идентификационните данни на вашата организация и дефинирайте обхвата на документа.

Наименование на организацията	Пълно официално наименование ЯНУС
Административно ниво	Централна / областна / общинска администрация Централна администрация
AI модел(и), обхванати от документа	Посочете конкретните модели/версии Вътрешни/изходящи
Версия на документа	1
Дата на приемане	22.05.2026
Следваща планирана ревизия	30.06.2027
Одобрен от	Длъжност и имена
Отговорен за прилагането	Ръководител на ЯНУС

Чл. 1. Предмет и обхват

Опишете за какво се отнася документът — кои AI системи, кои структурни звена и кои видове задачи обхваща.

Обхваща всички видове системи на Gemini, 2.5 Pro, отнася се за всички структурни звена и задачи.

Чл. 2. Правна и регулаторна рамка

Посочете нормативните актове, на базата на които се издават указанията (EU AI Act, GDPR, ЗЕУ, вътрешни наредби).

Акт за изкуствен интелект на ЕС (EU AI Act – Регламент (ЕС) 2024/1689);
GDPR - Регламент (ЕС) 2016/679);
Директива (ЕС) 2019/790 относно авторското право в цифровия единен пазар
ЗАПСП;
Закон за киберсигурност;
ЗЗП и др.

Раздел I. Принципи за етично и прозрачно ползване

Определете основните принципи, на които се базира политиката. Попълнете конкретните мерки за всеки.

Основни принципи	
1 Законност	AI системите се използват само в рамките на действащото законодателство. Всяко решение, подпомогнато от AI, спазва приложимите правни норми. Конкретни мерки в нашата организация: - Защита на личните данни (спазване на GDPR, който изисква анонимизация; - Етични филтри (предотвратяване на генерирането на дискриминационно съдържание или език на омразата); - Човешки надзор (AI предлага решение, но крайния избор и правна тежест пада в/у човек)
2 Прозрачност	Гражданите и служителите имат право да знаят кога AI участва в процесите. Резултатите от AI трябва да са обяснени разбираемо. Конкретни мерки в нашата организация: - Регистър на алгоритмите (опис на всички AI системи); - Одит (редовни проверки за безпристрастност; предотвратяване на алгоритмична дискриминация по пол, възраст, етнос и т.н.)
3 Отчетност	Ясна верига на отговорност. Служителите носят отговорност за решенията, взети с AI помощ — AI не замества човешката отговорност. Конкретни мерки в нашата организация: - На оперативно ниво (служител – експертно ниво) – отговаря за правилното въвеждане на данни; - На ръководно ниво – отговаря за избора на надежна AI система и за обучението на екипа си; - На организационно ниво – отговаря за осигуряване на механизъм за контрол, одит и реакция при грешка.
4 Справедливост	AI инструментите не дискриминират. Резултатите се проверяват редовно за пристрастия (bias) спрямо различни групи граждани. Конкретни мерки в нашата организация: - Използване на критерии за всички; - Механизъм за обжалване и преразглеждане на решението.
5 Поверителност	Личните данни на гражданите се обработват само при наличие на правно основание и при спазване на GDPR. Конкретни мерки в нашата организация: - Съгласие; - Договорно задължение; - Законово задължение; - Обществен интерес (функциониране на държавните органи); - Легитимен интерес (ползите не накърняват правата на индивида).

6 Надеждност	AI системите се използват само за задачи, при които точността е потвърдена. Резултатите подлежат на верификация. Конкретни мерки в нашата организация: - Сравняване на изхода от AI с алтернативни методи за изчисляване или др.; - Двойна проверка от експерт (юридически анализи) – крайния анализ задължително се преглежда от юрист/експерт; - При откриване на грешка, същата се докладва и модела се дообучава.
7 Пропорционалност	Нивото на автоматизация съответства на риска от грешки. При по-висок риск — по-задължителен човешки контрол. Конкретни мерки в нашата организация: - AI предлага, но не може да продължи без изрично потвърждение от служител; - AI действа автоматично, но служител следи в реално време и може да се намеси; - Системата е автоматизирана и служител извършва одит (при нисък риск).
8 Обяснимост	При поискване от гражданин или орган, организацията може да обясни как е бил използван AI при дадено решение. Конкретни мерки в нашата организация: - Дневници/ записи на решенията (logs)(история на това кога и как е стартирал AI моделът; - Избор на алгоритми, които позволяват да се води „тежестта“ на различните фактори; - Ясни правила за това как служителите документират намесата си в автоматизираните процеси.

Раздел II. Разрешени и забранени употреби

Определете за всяка категория задачи дали е разрешена, допустима с условия или забранена.

Инструкция за попълване: За всяка задача: поставете отметка в съответната колона и опишете условията или ограниченията.		
Категория задача	Статус	Условия / Ограничения / Процедура
Административни документи		
Изготвяне на проекти за решения	☐ Разрешено ☒ С условия ☐ Забранено	Условия / ограничения / процедура: Верификация от служител/ няма/ избор на модел AI
Резюмиране на входяща поща	☐ Разрешено ☒ С условия ☐ Забранено	Условия / ограничения / процедура: Верификация от служител/ няма/ коректно въвеждане на данни
Превод на документи	☐ Разрешено ☒ С условия ☐ Забранено	Условия / ограничения / процедура: Верификация от служител/ няма/ избор на модел AI
Генериране на официални писма	☐ Разрешено ☒ С условия ☐ Забранено	Условия / ограничения / процедура: Верификация от служител/ няма/ коректно въвеждане на данни
Правни и нормативни въпроси		
Анализ на нормативни текстове	☐ Разрешено ☒ С условия ☐ Забранено	Условия / ограничения / процедура: Верификация от служител/ няма/ коректно въвеждане на данни
Проверка за вътрешни противоречия	☐ Разрешено ☐ С условия ☒ Забранено	Условия / ограничения / процедура: н. п.
Изготвяне на правни становища	☐ Разрешено ☒ С условия	Условия / ограничения / процедура: Верификация от служител/ няма/ коректно въвеждане на данни

	☐ Забранено	
Тълкуване на разпоредби към граждани	☐ Разрешено ☒ С условия ☐ Забранено	Условия / ограничения / процедура: Верификация от служител/ няма/ избор на масив от данни
Вземане на решения с ефект върху граждани		
Преценка за социални помощи	☐ Разрешено ☒ С условия ☐ Забранено	Условия / ограничения / процедура: Верификация от служител/ няма/ избор на масив от данни
Оценка на проекти и заявления	☐ Разрешено ☒ С условия ☐ Забранено	Условия / ограничения / процедура: Верификация от служител/ няма/ избор на масив от данни
Наказателни и принудителни мерки	☐ Разрешено ☐ С условия ☒ Забранено	Условия / ограничения / процедура: н. п.
Автоматизирани решения без човешки контрол	☐ Разрешено ☐ С условия ☒ Забранено	Условия / ограничения / процедура: н. п.
Комуникация с граждани		
Чатбот за информационни въпроси	☒ Разрешено ☐ С условия ☐ Забранено	Условия / ограничения / процедура: При невъзможност, препращане към служител/ няма/ избор на масив от данни
Проекти за отговори на жалби	☐ Разрешено ☒ С условия ☐ Забранено	Условия / ограничения / процедура: Верификация от служител/ няма/ коректно въвеждане на данни
Публикации в социални медии	☐ Разрешено ☐ С условия ☒ Забранено	Условия / ограничения / процедура: н. п.
Кризисна комуникация	☐ Разрешено ☐ С условия ☒ Забранено	Условия / ограничения / процедура: н. п.
Вътрешно-административни функции		
Обучение на служители	☒ Разрешено ☐ С условия ☐ Забранено	Условия / ограничения / процедура: Дигитализация на всички вътрешни документи
Анализ на данни и отчети	☐ Разрешено ☒ С условия ☐ Забранено	Условия / ограничения / процедура: Верификация от служител/ няма/ избор на модел AI
Автоматизация на повтарящи се задачи	☐ Разрешено ☒ С условия ☐ Забранено	Условия / ограничения / процедура: Верификация от служител/ няма/ коректно въвеждане на данни
Обработка на класифицирана информация	☐ Разрешено ☐ С условия ☒ Забранено	Условия / ограничения / процедура: н. п.

Раздел III. Процедури за одобрение и верификация

Чл. 10. Задължителна верификация

Опишете кои задачи изискват задължителна проверка от служител преди използване на AI изхода.

- Изготвяне на проекти на решения;
- Генериране на официални писма;
- Изготвяне на правни становища;

- Тълкуване на разпоредби към граждани;
- Преценка за социална помощ;
- Оценка на проекти и заявления;
- Изготвяне на проекти на отговори на жалби.

Чл. 11. Процедура за одобрение на нови употреби

Как служителят може да заяви одобрение за нов вид употреба на AI, непредвиден в настоящите указания?

- Кратко описание на инструмента и метода на използване;
- ИТ отдела проверява дали инструмента е сигурен;
- Юрист/про потвърждава, че употребата не нарушава законите и принципа на забрана за дискриминация.

Чл. 12. Задължително разкриване към граждани

В кои случаи гражданинът трябва да бъде информиран, че AI е участвал в обработката на неговото искане?

Във всички случаи. Гражданите имат право да поискат преразглеждане на казуса изцяло от човек.

Раздел IV. Мониторинг, одит и отчетност

Минимални изисквания за одитна следа (отбележете приложимите):

- ✓ Всяко взаимодействие с AI системата се записва с дата, час и потребител
- ✓ Пазят се логове на AI изходите, използвани в официални документи
- ✓ При автоматизирани решения се съхранява AI входа и изхода
- ✓ Провежда се периодичен преглед на качеството на AI изходите (мин. 1х/тримесечие)
- ✓ Инцидентите, свързани с AI грешки, се докладват на определено звено
- ✓ Годишен отчет за употребата на AI се изготвя и публикува

Чл. 15. KPI за мониторинг на AI употребата

Определете конкретни измерими показатели: брой взаимодействия - 50, 90 % задачи с верификация, 50 % открити грешки и др.

Key Performance Indicators – ключови показатели за ефективност

- Положителна обратна връзка на всяко тримесечие;
- Похвали;
- Липса на препоръки, грешки и неточности.

Раздел V. Отговорности и управление

Роля	Отговорност	Конкретно лице / звено
Ръководител на организацията	Стратегическо одобрение и крайна отговорност за AI политиката	Конкретно лице / звено: Ръководител на администрацията
AI отговорник (AI Officer)	Координация, обучение, мониторинг, докладване на инциденти	Конкретно лице / звено: ИТ звено
Служители, ползващи AI	Спазване на указанията, верификация на изходите, докладване на проблеми	Конкретно лице / звено: Всички служители
ИТ отдел / системен администратор	Технически достъп, интеграция, сигурност, логове	Конкретно лице / звено: системен администратор
Правно звено	Преглед на нови употреби, съответствие с нормативната рамка	Конкретно лице / звено: Юрист от звеното

ДЛЗД (длъжностно лице по защита на данни)	GDPR съответствие при обработка на лични данни	Конкретно лице / звено: dpo
Вътрешен одит	Периодичен независим преглед на AI употребата	Конкретно лице / звено: водещ одитор

Раздел VI. Обучение и изграждане на капацитет

Чл. 20. Задължително обучение			
Посочете кои служители трябва да преминат обучение, какъв е минималният обем и как се документира. - Подлежат всички служители според ролята им в организацията; - Минимално 10 часа при постъпване на работа; - Оперативна работа с инструмента AI; - Сертификати; - Декларации; - Регистър на компетенции; - Дневник за периодични инструктажи (не по-малко от 1 път на 5 месеца).			
Модул на обучение	Целева група	Форма и обем	Периодичност
Основи на AI и рискове	Всички служители	Присъствен/онлайн	Веднъж на 5 месеца
Работа с конкретния модел	Съобразно функционалната компетентност	Присъствен/онлайн	Веднъж на 5 месеца
Етика и нормативна рамка	Всички служители	Присъствен/онлайн	Веднъж на 5 месеца
Разпознаване на AI грешки	Всички служители	Присъствен/онлайн	Веднъж на 5 месеца
Защита на личните данни	Всички служители	Присъствен/онлайн	Веднъж на 5 месеца

Раздел VII. Актуализация и валидност

Чл. 24. Условия за задължителна ревизия			
Опишете кога документът трябва задължително да се преразгледа: нова версия на модела, регулаторна промяна, инцидент, изтичане на срок. Нова версия трябва задължително да се внедри, съобразно регулируемото законодателство, регулаторна промяна, регистриране на повече от 30% грешки за отчетен период и/или изтичане на версията на инструмента.			
Чл. 25. История на промените			
Документирайте всички версии на документа с дата, автор и описание на промяната. Версия 1 – Първична за внедряване на Gemini, 2.5 Pro			
Версия	Дата	Описание на промяната	Одобрил
Версия 1	22.05.2026 г.	Въвеждане на инструмента	Ръководител на организацията

ГРУПА „КЕЦАЛКОАТЪЛ“

DeepSeek (V3.2/R1)



ШАБЛОН 1 / 4

SWOT Анализ

Оценка на фундаментален AI модел за публичната администрация

v1.0 — Работен шаблон

Инструкция за попълване: Попълнете всички полета за избория от вашата група AI модел. Празните редове и скалите са предназначени за ръчно попълване. Не изтривайте инструкциите в курсив.

Наименование на модела	Пример: Claude Opus 4.7 / GPT-5.4 / Gemini 3.1 Pro DeepSeek AI
Разработчик / компания	Пример: Anthropic / OpenAI / Google DeepMind Hangzhou DeepSeek AI
Версия за оценка	Точната версия, която групата изследва V3
Дата на попълване	ДД.ММ.ГГГГ 14.05.2026
Членове на групата	Имена на участниците Цветомир, Константин, Есма, Антония, Клаудия, Явор
Ръководител на групата	Константин

A. Обща приложимост и функционален обхват

Оценете доколко моделът покрива реални нужди в публичния сектор.

Инструкция за попълване: За всеки критерий: поставете кръгче на избраната оценка (1=много слабо, 5=отлично) и запишете кратка обосновка.

Критерий		
A1. Приложимост в публичния сектор		
Обработка на документи доклади, решения, наредби	Оценка 1–5 ○1 ○2 ○3 ○4 ○5	Обосновка / бележки: Извършва синтезиран анализ на документа, като предоставя основните акценти по разбираем начин.
Правен анализ тълкуване на нормативни актове	Оценка 1–5 ○1 ○2 ○3 ○4 ○5	Обосновка / бележки: Извършва качествен анализ като сравнява и оценява различията и промените на нормативната рамка, изложени в документа.

Граждански услуги чатбот, Q&A, помощ при заявления	Оценка 1–5 ○1 ○2 <u>○3</u> ○4 ○5	Обосновка / бележки: Предвид опростения и синтезиран вид, по който предоставя информацията от анализирания тестов, може да се счита, че е пригоден за предоставяне на конкретни кратки отговори примерно чрез чатбот. Възможно е да липсва детайлност при конкретни запитвания. —
Вътрешна комуникация резюмета, протоколи, кореспонденция	Оценка 1–5 ○1 <u>○2</u> ○3 ○4 ○5	Обосновка / бележки: Може да се използва при структурирането на шаблони и писма, но използването му не се препоръчва предвид рисковете свързани с конфиденциалността на информацията. —
Анализ на данни отчети, статистика, мониторинг	Оценка 1–5 ○1 ○2 <u>○3</u> ○4 ○5	Обосновка / бележки: Извършва добър цялостен и сравнителен анализ, на данни от различни по формат документи. Дава конкретни и структурирани изводи и отговори. Не е подходящ за документи, съдържащи чувствителна информация. —
Превод многоезична комуникация, ЕС документи	Оценка 1–5 ○1 ○2 ○3 <u>○4</u> ○5	Обосновка / бележки: Превежда добре смислово и структурирано от английски на български с малък брой неточности. —
Програмиране / автоматизация вътрешни системи, скриптове	Оценка 1–5 ○1 ○2 <u>○3</u> ○4 ○5	Обосновка / бележки: Моделът може да създава работещ код на различни езици, който да автоматизира често срещани, но обикновени задачи в публичната администрация: Не поддържа постоянни сесии или cloud функции, което може да създаде ограничения в изпълняването на задачите. —
Обучение на служители учебни материали, тестове	Оценка 1–5 ○1 ○2 ○3 <u>○4</u> ○5	Обосновка / бележки: Предвид възможностите му за детайлен анализ, може да се използва при обучение на служителите и генериране на учебни материали и проверка на тестове. —
A2. Функции на платформата		
Текстово генериране качество и прецизност	Оценка 1–5 ○1 ○2 <u>○3</u> ○4 ○5	Обосновка / бележки: При генериране на текст по въпрос с поставени критерии, не дава конкретен отговор като е необходимо задаването на допълнителни критерии и данни, които е възможно да не са предварителни известни на потребителя. —
Reasoning / разсъждение	Оценка 1–5 ○1 ○2 ○3 <u>○4</u> ○5	Обосновка / бележки: При запитвания за публични процедури се ограничава до законовата рамка, без да предоставя алтернативни решения.

логически вериги, многостъпкови задачи		—
Code writing генериране и анализ на код	Оценка 1–5 ○1 ○2 ○3 ○4 <u>○5</u>	Обосновка / бележки: Един от най-добрите модели в света за програмиране. Понякога може да предложи по-малко "сигурни" решения.
Анализ на изображения vision, диаграми, скрийншоти	Оценка 1–5 ○1 <u>○2</u> ○3 ○4 ○5	Притежава солидни мултимодални способности, разпознава текст (OCR) и структури, но отстъпва на други модели при по-детайлни или художествени анализи.
Аудио / глас говорим вход/изход	Оценка 1–5 ○1 <u>○2</u> ○3 ○4 ○5	Обосновка / бележки: Поддържа превод и транскрипция чрез API, но няма развит нативен специализиран аудио модул. Липсва емоционален диапазон и ниска латентност в реално време.
Видео разбиране анализ на видеосъдържание	Оценка 1–5 <u>○1</u> ○2 ○3 ○4 ○5	Обосновка / бележки: Може да анализира кадри, но липсва нативна поддръжка за дълги видео файлове с разбиране на времевата линия.
Дълги документи обработка на обемни файлове	Оценка 1–5 ○1 ○2 ○3 <u>○4</u> ○5	Обосновка / бележки: Поддържа голям контекстен прозорец (128k токена), което е стандарт за индустрията. При дълги документи извличането на информация от средата на текста понякога е затруднено. Идеален за анализ на цели програмни библиотеки.
Agentic режим автономни многостъпкови задачи	Оценка 1–5 ○1 ○2 <u>○3</u> ○4 ○5	Обосновка / бележки: Висок "Reasoning" капацитет. Оптимизиран за мислене стъпка по стъпка. Изисква добра рамка (framework), за да действа напълно автономно. Добра основа за изграждане на AI агенти заради съотношението цена/интелект.
API / системна интеграция свързване с вътрешни системи	Оценка 1–5 ○1 ○2 ○3 <u>○4</u> ○5	Обосновка / бележки: Напълно съвместим с OpenAI API стандарта. Изключително ниски цени (до 10-20 пъти по-евтино от конкуренцията). Понякога има прекъсвания на сървърите поради огромен трафик от цял свят. Локални регулации могат да ограничат достъпа до API в определени региони.
Plugin екосистема налични разширения и конектори	Оценка 1–5 ○1 <u>○2</u> ○3 ○4 ○5	Обосновка / бележки: Може да се свързва чрез определени инструменти, но няма собствен "App Store" или богата затворена екосистема от плъгини. Затворените екосистеми на Microsoft и Google предлагат по-лесна интеграция за бизнес потребители.

Desktop приложение <i>нативна desktop среда</i>	Оценка 1–5 ○1 ○2 <u>○3</u> ○4 ○5	Обосновка / бележки: DeepSeek няма официално нативно Desktop приложение. Ползва се основно през уеб или API.
Mobile приложение <i>iOS / Android достъп</i>	Оценка 1–5 ○1 ○2 <u>○3</u> ○4 ○5	Обосновка / бележки: Има официални приложения за iOS и Android, но липсват системни интеграции
Cowork / колаборация <i>екипна работа в реално време</i>	Оценка 1–5 ○1 <u>○2</u> ○3 ○4 ○5	Обосновка / бележки: Възможност за споделяне на чатове, но липсват способности за екипна работа в реално време
Търсене в интернет <i>актуална информация от мрежата</i>	Оценка 1–5 ○1 ○2 <u>○3</u> ○4 ○5	Обосновка / бележки: Интегрирано търсене, което работи добре и цитира източници, но филтрира информация поради китайските регулации за съдържание.
Fine-tuning <i>настройка за специфична среда</i>	Оценка 1–5 ○1 ○2 ○3 <u>○4</u> ○5	Обосновка / бележки: Тъй като архитектурата е частично отворена, организациите могат да го дообучават върху собствени сървъри. Изисква сериозен хардуерен ресурс за модел с такъв размер.

B. SWOT Матрица

Обобщете основните силни/слаби страни, възможности и заплахи на модела в публичен контекст.

 **Инструкция за попълване:** Запишете поне 4 наблюдения в клетка. Базирайте се на попълнените оценки от Раздел А.

S STRENGTHS — Силни страни 1. Рекордно ниска цена спрямо конкурети. 2. Отличен reasoning 3. Open-weight модел MIT	W WEAKNESSES — Слаби страни 1. Ограничена мултимодалност - Липсва добра поддръжка за аудио и видео, слаб анализ на изображения 2. Зависимост от чат платформата за уеб търсене.
--	--

4. Възможност за работа върху собствена инфраструктура. 5. Обработка на големи документи 6. Висока производителност Допълнителни бележки:	3. Ограничена плъгин среда 4. Изисква сериозен ресурс за локлано хостване 5. Липсва на вградени държавни спецификации, не познава добре българската нормативна уредба 6. Ограничена прозрачност и сигурност. Потенциални регулаторни рискове. Допълнителни бележки:
O OPPORTUNITIES — Възможности 1. Автоматизация на определени административни процеси. 2. Обучение на служителите 3. Намаляване на административни разходи. 4. Създаване на вътрешна екосистема, изолирана от глобалната мрежа 5. 6. Допълнителни бележки:	T THREATS — Заплахи 1. Регулаторни пречки предвид ситуирането на разработчика в Китай - геополитически риск 2. GDPR ограничения във връзка с възможни течове на чувствителни данни. 3. Рискове свързани с доставката на софтуера и възможността от санкции и кибер рискове. 4. Липса на редовни актуализации и поддръжка - бързо остаряване на продукта спрямо конкуренцията. 5. 6. Допълнителни бележки:

C. Разширен SWOT по специфични теми

 **Инструкция за попълване:** Попълнете текстово наблюденията си за всяка тема. Използвайте конкретни примери.

Цена и лицензиране**Наблюдения:** (Модел на плащане, достъпност за малки общини, API разходи)

Моделът е революционен по отношение на цената. API разходите са около 10–20 пъти по-ниски от тези на останалите конкуренти - \$0.028 за 1M токена

Препоръка за публичния сектор: Подходящ за дейности в малки администрации и институции, при които не се налага работа с чувствителна информация.

Сигурност на данните**Наблюдения:** (Обработка на лични данни, GDPR, съхранение в облака)

Тъй като DeepSeek е базиран в Китай, съхранението в облака повдига въпроси относно суверенитета на данните. Обработката на лични данни през техния публичен API крие риск от изнасяне на информация извън ЕС/ЕИП

Препоръка за публичния сектор: Да не се подават лични данни (ЕГН, имена и други) през публичния API и да не се използва за работа с класифицирана информация. В такива случаи, да се използва локална инсталация на собствен сървър.

Езикова поддръжка**Наблюдения:** (Качество на български език, юридическа терминология)

Моделът показва добро качество на български език, включително граматика и синтаксис. Юридическата терминология е коректна, но понякога може да използва термини от други правни системи.

Препоръка за публичния сектор: Задължителен "human-in-the-loop" (човешки надзор) при генериране на официални документи или правни становища.

Vendor lock-in**Наблюдения:** (Зависимост от доставчик, алтернативи, преносимост)

Високи нива на съвместимост и преносимост.

Препоръка за публичния сектор: Добра поддръжка на модулната архитектура на софтуера за улеснено прехвърляне при необходимост.

Сложност на внедряване**Наблюдения:** (Техническа зрялост, необходима инфраструктура, IT капацитет)

API интеграцията е стандартна и бърза, но локалното внедряване на пълния модел изисква сериозна инфраструктура и инвестиция

Препоръка за публичния сектор: Да се използва само за работа с нечувствителни данни. Преценка на рентабилността на инвестицията в локална инфраструктура.

Регулаторно съответствие

Наблюдения: (EU AI Act, NIS2/3KC, GDPR, национални наредби)

Липса на пълно съответствие с EU AI Act и други нормативни документи по отношение на прозрачността на данните за обучение. Трудно покриване на регулаторните изисквания, ако се разчита на външни сървъри извън ЕС.

Препоръка за публичния сектор: Да се използва само за спомагателни дейности.

Етика и прозрачност

Наблюдения: (Обяснимост, одитируемост, отговорност при грешки)

При грешка отговорността остава изцяло върху административния орган. Ниска одитируемост на процеса на вземане на решения на модела.

Препоръка за публичния сектор: Да се извършва контрол върху решенията, предлагани от езиковия модел. Да се създадат вътрешни правила по отношение на подаваната информация от страна на членове на администрацията.

Изкривявания (Bias)

Наблюдения: (Пристрастия при различни групи, равнопоставеност)

По-малко вероятно е да разпознава специфични регионални социални нюанси в България. Може да фаворитизира източноазиатски позиции и разбирания.

Препоръка за публичния сектор: Редовни тестове на отговорите при автоматизирана обработка на услуги за граждани.

Халюцинации / грешки

Наблюдения: (Честота, detectability, пуск при вземане на решения)

При сложни изчисления или интерпретация на непознати наредби може да измисли несъществуващи правила, особено при тестове с административни документи.

Препоръка за публичния сектор: Използване на по-конкретна терминология и разбиване на задачите при заявки за генериране на анализи.

Морал и ценности

Наблюдения: (Поведение при чувствителни теми, отказ на вредни задачи)

Отказва генериране на вредно съдържание при опит за нарушаване на закона и неетични практики. Доказано цензурира въпроси, противоречащи на официалните позиции на Китай, но проявява висок морал и ценности по отношение на други социални въпроси.

Препоръка за публичния сектор: Подходящ за среда, където се изисква политическа неутралност, да се осъществява контрол по отношение на теми свързани с позициите на Китай.

D. Обобщена оценка на групата

Обобщена оценка по основни показатели		
Обща приложимост <i>средна от Раздел А</i>	Оценка 1–5 ○1 ○2 <u>○3</u> ○4 ○5	Обосновка / бележки: Силен в програмиране, логически задачи и обработка на текст. Не достига цялостно нивото на останалите конкуренти.
Готовност за внедряване <i>техническа и организационна</i>	Оценка 1–5 ○1 ○2 <u>○3</u> ○4 ○5	Обосновка / бележки: Висока готовност през API (OpenAI съвместимост), но сложна за локално хостване поради огромните изисквания към хардуера.
Съответствие с регулации <i>AI Act, GDPR, NIS2</i>	Оценка 1–5 ○1 ○2 <u>○3</u> ○4 ○5	Обосновка / бележки: Критични рискове по GDPR (трансфер извън ЕС) и неясна прозрачност на данните за обучение спрямо AI Act.
Етичен риск <i>обратна скала: 5=нисък риск</i>	Оценка 1–5 ○1 ○2 ○3 <u>○4</u> ○5	Обосновка / бележки: Среден риск. Стабилни филтри за безопасност, но липса на външен одит и риск от пристрастия
Обща препоръка <i>1=не препоръчваме, 5=силно препоръчваме</i>	Оценка 1–5 ○1 ○2 <u>○3</u> ○4 ○5	Обосновка / бележки: DeepSeek V3 е идеален за вътрешна употреба (помощник на програмисти, подготовка на чернови на доклади, анализ на неклаसифицирана информация), но към момента не се препоръчва за автоматизирана работа с лични данни на граждани или за вземане на крайни административни решения без човешка проверка.

Мотивирано становище на групата (мин. 5 изречения): *Обобщете защо препоръчвате или не препоръчвате модела за публичния сектор.*

Използването на DeepSeek V3 в публичната администрация крие значителни рискове за суверенитета на данните, тъй като съхранението на информация извън юрисдикцията на ЕС компрометира фундаменталното доверие в държавните институции. Липсата на регулаторна прозрачност създава психологическа несигурност и институционална уязвимост, които надхвърлят потенциалните ползи от оперативната бързина. Въпреки че моделът демонстрира добри логически способности и финансова достъпност, тези предимства остават повърхностни пред лицето на етичната неяснота и опасността от неконтролирани пристрастия в административните решения. Зависимостта от външен доставчик с неясни механизми на отчетност ерозира усещането за предвидимост и сигурност. В крайна сметка, техническата ефективност на инструмента не може да компенсира липсата на правни гаранции и психологическия дискомфорт от делегирането на публични функции на непроверена технологична архитектура.

Предложени условия за ползване: *Какви ограничения, контроли или процедури трябва да съпътстват внедряването?*

Предвид гореизложеното, групата счита, че моделът е подходящ единствено за ограничени и нискорискови административни дейности.със задължителна човешка проверка на процесите и забрана за обработка на чувствителни и лични данни.

Предв

ШАБЛОН 2 / 4

Risk Canvas

Картографиране на рисковете от AI модел в публичен контекст

v1.0 — Работен шаблон

 **Инструкция за попълване:** Попълнете всички полета за изборания от вашата група AI модел. Празните редове и скалите са предназначени за ръчно попълване. Не изтривайте инструкциите в курсив.

Наименование на модела	Пример: Claude Opus 4.7 / GPT-5.4 / Gemini 3.1 Pro DeepSeek AI
Разработчик / компания	Пример: Anthropic / OpenAI / Google DeepMind Hangzhou DeepSeek AI
Версия за оценка	Точната версия, която групата изследва V3
Дата на попълване	ДД.ММ.ГГГГ 14.05.2026
Членове на групата	Имена на участниците Цветомир, Константин, Есма, Антония, Клаудия, Явор
Ръководител на групата	Константин

Легенда за скалите

Вероятност (Вер.) 1 = Изключително рядко 2 = Рядко 3 = Умерено вероятно 4 = Вероятно 5 = Почти сигурно	Въздействие (Вл.) 1 = Незначително 2 = Малко 3 = Умерено 4 = Значително 5 = Критично	Ниво на риска Н = Нисък (Вер.×Вл. ≤ 4) С = Среден (5–12) В = Висок (13–25)
--	--	--

Риск Канва — 9-полева матрица

Попълнете всяко поле с идентифицирани рискове, засегнати страни и ключови наблюдения.

 **Инструкция за попълване:** Всяко от 9-те полета представлява различна категория риск. Запишете конкретни рискове — не общи фрази.

1. Технически рискове Риск от халюцинации и неточни отговори при правен анализ. Ограничения при многостепни задачи.	2. Рискове за данните Потенциален теч на чувствителна информация. Трансфер на данни извън ЕС Риск при обработка на лични данни през публичния API. Липса на прозрачност по отношение на съхранението на данни.	3. Правни / регулаторни Несъответствие с EU AI Act, NIS2, GDPR, и национално право. Липса на регулаторна прозрачност
4. Етични рискове Непрозрачност в логиката на модела. Неконтролирани пристрастия. Възможен bias при определени позиции. Риск от цензуриране.	5. Оперативни рискове Зависимост от външен доставчик. Ограничена plugin система. Високи хардуерни изисквания за локално внедряване.	6. Репутационни рискове Загуба на доверие при публични грешки Репутационен риск при теч на данни Обществено недоволство при установяване на чуждестранна намеса.
7. Стратегически рискове Геополитически риск от обвързване с архитектура, ситуирана извън страната и ЕС. Зависимост от външна инфраструктура. Риск от спиране на услугата и поддръжката	8. Засегнати страни Граждани - по отношение на поверителността на личните данни. Риск от деградация на уменията на служителите.. Последствия и юридическа отговорност при погрешни решения, подготвени от AI и касаещи ключови процеси на държавната администрация. Спад на доверието на партньорите при потенциални сривове в предоставяните услуги Риск за отделите по поддръжка на системите и киберсигурността поради липса на външен SLA.	9. Остатъчни рискове Скрити технологични вратички. Риск от халюцинации, дори след проверка.

Рисков регистър

Попълнете по всяка категория конкретните рискове с вероятност, въздействие и мерки.

 **Инструкция за попълване:** Вер. = Вероятност (1–5) · Вл. = Въздействие (1–5) · Ниво: ☐ Н (нисък) ☐ С (среден) ☐ В (висок)

Риск / Събитие	Вероятност	Въздействие	Ниво	Мярка за митигация
Технически рискове				
Халюцинации и фактически грешки при отговори	Вер. ○1 ○2 ○3 <u>○4</u> ○5	Вл. ○1 ○2 ○3 <u>○4</u> ○5	Ниво ☐ Н ☐ С ☒ В	Мярка за митигация: Човешка проверка на всички административни и правни казуси. Информацията да не се използва в официални документи без проверка.
Недостъпност / downtime на платформата	Вер. ○1 ○2 <u>○3</u> ○4 ○5	Вл. ○1 ○2 <u>○3</u> ○4 ○5	Ниво ☐ Н ☒ С ☐ В	Мярка за митигация: Резервни вътрешни решения и алтернативни решения.
Внезапна промяна на версия или поведение на модела	Вер. ○1 ○2 <u>○3</u> ○4 ○5	Вл. ○1 ○2 ○3 <u>○4</u> ○5	Ниво ☐ Н ☒ С ☐ В	Мярка за митигация: Тестване преди внедряване на нови версии
Ограничения на контекстния прозорец при обемни документи	Вер. ○1 ○2 <u>○3</u> ○4 ○5	Вл. ○1 ○2 <u>○3</u> ○4 ○5	Ниво ☐ Н ☒ С ☐ В	Мярка за митигация: Разделяне на документите на по-малки части.
Грешки при мултимодален вход (изображения, аудио)	Вер. ○1 ○2 ○3 ○4 <u>○5</u>	Вл. ○1 ○2 ○3 ○4 <u>○5</u>	Ниво ☐ Н ☐ С ☒ В	Мярка за митигация: Използване на специализирани локални OCR инструменти.
Рискове за данните и поверителността				
Обработка на лични данни на граждани в облачна среда	Вер. ○1 ○2 ○3 ○4 <u>○5</u>	Вл. ○1 ○2 ○3 ○4 <u>○5</u>	Ниво ☐ Н ☐ С ☒ В	Мярка за митигация: Забрана или анонимизиране на подаваните лични данни.
Изтичане на класифицирана/служебна информация	Вер. ○1 ○2 ○3 ○4 <u>○5</u>	Вл. ○1 ○2 ○3 ○4 <u>○5</u>	Ниво ☐ Н ☐ С ☒ В	Мярка за митигация: Строга забрана за качване на служебна/класифицирана документация в облачната версия
Запазване на запитвания от доставчика за обучение	Вер. ○1 ○2 ○3 <u>○4</u> ○5	Вл. ○1 ○2 ○3 <u>○4</u> ○5	Ниво ☐ Н ☐ С ☒ В	Мярка за митигация: Сключване на специфични договори за API употреба или преминаване към локален хостинг.
Трансфер на данни извън ЕС (EU AI Act / Schrems II)	Вер. ○1 ○2 ○3 ○4 <u>○5</u>	Вл. ○1 ○2 ○3 ○4 <u>○5</u>	Ниво ☐ Н ☐ С ☒ В	Мярка за митигация: Локално хостване при възможност.
Липса на одитна следа при автоматизирани решения	Вер. ○1 ○2 <u>○3</u> ○4 ○5	Вл. ○1 ○2 ○3 <u>○4</u> ○5	Ниво ☐ Н ☐ С ☒ В	Мярка за митигация: Разработване на вътрешен софтуер, който задължително записва в системните логове

				автора на промпта, изхода на модела и крайната верификация от служителя.
Правни и регулаторни рискове				
Несъответствие с EU AI Act (Annex III — публичен сектор)	Вер. ○1 ○2 ○3 <u>○4</u> ○5	Вл. ○1 ○2 ○3 ○4 <u>○5</u>	Ниво ☐ Н ☐ С ☒ В	Мярка за митигация: Ограничаване до спомагателни дейности.
Нарушение на GDPR при обработка на лични данни	Вер. ○1 ○2 ○3 <u>○4</u> ○5	Вл. ○1 ○2 ○3 ○4 <u>○5</u>	Ниво ☐ Н ☐ С ☒ В	Мярка за митигация: Консултация с правно звено. Забрана или анонимизиране при използване на чувствителни данни.
Липса на обяснимост при автоматизирани административни решения	Вер. ○1 ○2 ○3 <u>○4</u> ○5	Вл. ○1 ○2 ○3 <u>○4</u> ○5	Ниво ☐ Н ☐ С ☒ В	Мярка за митигация: Човешки надзор и вътрешен контрол.
Авторскоправни нарушения в генерирано съдържание	Вер. ○1 <u>○2</u> ○3 ○4 ○5	Вл. ○1 <u>○2</u> ○3 ○4 ○5	Ниво ☒ Н ☐ С ☐ В	Мярка за митигация: Използване на филтри за проверка за плагиатство и придържане до публично достъпни текстове
Отговорност при грешни правни или медицински изходи	Вер. ○1 ○2 <u>○3</u> ○4 ○5	Вл. ○1 ○2 ○3 ○4 <u>○5</u>	Ниво ☐ Н ☐ С ☒ В	Мярка за митигация: Правно дефиниране в длъжностните характеристики, че крайната професионална и юридическа отговорност остава за служителя.
Етични и социални рискове				
Систематичен bias спрямо определени групи граждани	Вер. ○1 ○2 <u>○3</u> ○4 ○5	Вл. ○1 ○2 ○3 <u>○4</u> ○5	Ниво ☐ Н ☐ С ☒ В	Мярка за митигация: Периодични тестове и оценка на отговорите
Дискриминация при автоматизирано обработване на заявления	Вер. ○1 <u>○2</u> ○3 ○4 ○5	Вл. ○1 ○2 ○3 ○4 <u>○5</u>	Ниво ☐ Н ☐ С ☒ В	Мярка за митигация: Забрана за използване на модела за филтриране и оценка на гражданите, получаващи съответната услуга..
Непрозрачност — граждани не знаят, че AI взима решения	Вер. ○1 ○2 ○3 <u>○4</u> ○5	Вл. ○1 ○2 <u>○3</u> ○4 ○5	Ниво ☐ Н ☐ С ☒ В	Мярка за митигация: Показване на ясно и видимо съобщение, че гражданинът си взаимодейства с или получава текст, генериран от AI.
Дезинформация и манипулация чрез убедително съдържание	Вер. ○1 ○2 <u>○3</u> ○4 ○5	Вл. ○1 ○2 ○3 <u>○4</u> ○5	Ниво ☐ Н ☐ С ☒ В	Мярка за митигация: Задължително водно маркиране на официалните текстове и използване на сигурни канали (електронни подписи) за автентичност.
Загуба на критично мислене и зависимост при служителите	Вер. ○1 ○2 ○3 <u>○4</u> ○5	Вл. ○1 ○2 ○3 <u>○4</u> ○5	Ниво ☐ Н ☐ С ☒ В	Мярка за митигация: Редовни обучения на служителите относно лимитите на технологията и значението на експертната проверка на фактите.
Оперативни рискове				

Липса на технически капацитет за внедряване и поддръжка	Вер. ○1 ○2 <u>○3</u> ○4 ○5	Вл. ○1 ○2 ○3 <u>○4</u> ○5	Ниво ☐ Н ☒ С ☐ В	Мярка за митигация: Обучение на IT екипите
Vendor lock-in — прекомерна зависимост от доставчик	Вер. ○1 ○2 <u>○3</u> ○4 ○5	Вл. ○1 ○2 ○3 <u>○4</u> ○5	Ниво ☐ Н ☒ С ☐ В	Мярка за митигация: Модула архитектура и алтернативни решения
Недостатъчно обучение на служителите	Вер. ○1 ○2 ○3 ○4 <u>○5</u>	Вл. ○1 ○2 <u>○3</u> ○4 ○5	Ниво ☐ Н ☒ С ☐ В	Мярка за митигация: Разработване на кратки и ясни наръчници за промптване и провеждане на задължителни симулационни тестове.
Интеграционни проблеми с наследени (legacy) системи	Вер. ○1 ○2 <u>○3</u> ○4 ○5	Вл. ○1 ○2 <u>○3</u> ○4 ○5	Ниво ☐ Н ☒ С ☐ В	Мярка за митигация: Поетапно внедряване
Ценови скокове при промяна на лицензионния модел	Вер. ○1 <u>○2</u> ○3 ○4 ○5	Вл. ○1 ○2 <u>○3</u> ○4 ○5	Ниво ☒ Н ☐ С ☐ В	Мярка за митигация: Сключване на дългосрочни договори за цени на токен през API.
Репутационни и стратегически рискове				
Публичен скандал при видима AI грешка (медийно отразяване)	Вер. ○1 ○2 ○3 <u>○4</u> ○5	Вл. ○1 ○2 ○3 ○4 <u>○5</u>	Ниво ☐ Н ☐ С ☒ В	Мярка за митигация: Ограничаване на автоматизираните решения в медийно достъпни решения и предоставяни услуги
Загуба на доверие на гражданите в публичните институции	Вер. ○1 ○2 <u>○3</u> ○4 ○5	Вл. ○1 ○2 ○3 ○4 <u>○5</u>	Ниво ☐ Н ☐ С ☒ В	Мярка за митигация: Пълна прозрачност за това къде и защо се ползва AI в администрацията.
Политическа съпротива срещу използването на AI	Вер. ○1 ○2 <u>○3</u> ○4 ○5	Вл. ○1 ○2 <u>○3</u> ○4 ○5	Ниво ☐ Н ☒ С ☐ В	Мярка за митигация: Провеждане на информационни сесии за ръководния състав и синдикалните организации, демонстриращи, че AI е инструмент в помощ на служителите, а не техен заместник.
Зависимост от американски/китайски технологичен доставчик	Вер. ○1 ○2 ○3 <u>○4</u> ○5	Вл. ○1 ○2 ○3 <u>○4</u> ○5	Ниво ☐ Н ☐ С ☒ В	Мярка за митигация: Дългосрочна стратегия за миграция към европейски отворени модели.
Спиране на услугата при геополитически напрежения	Вер. ○1 ○2 <u>○3</u> ○4 ○5	Вл. ○1 ○2 ○3 ○4 <u>○5</u>	Ниво ☐ Н ☐ С ☒ В	Мярка за митигация: Поддържане на готовност за спешно превключване на инфраструктурата към напълно локално работещи модели, напълно независими от външни интернет връзки и чужди сървъри.

Обобщена рискова оценка на групата

Топ 3 критични риска за модела:

1. Нарушение на GDPR и трансфер на данни извън ЕС

2. Юридически халюцинации и фактически грешки при правен анализ
3. Липса на съответствие с изискванията на европейското законодателство

Обща рискова оценка на модела за публичния сектор:

- ☐ Нисък риск — може да се внедри с минимални ограничения
- ☐ Среден риск — изисква допълнителни контроли и процедури
- ☐ **Висок риск — изисква задълбочена оценка и специфична регулация**
- ☐ **Неприемлив риск — не се препоръчва за публичния сектор**

Мотивирано становище:

DeepSeek V3 притежава изключителен технологичен интелект и безпрецедентна икономическа ефективност, което го прави привлекателен за автоматизация на рутинни процеси.

Въпреки това, неговият геополитически произход, облачната архитектура извън пределите на ЕС и липсата на нативни съответствия с европейското законодателство (GDPR и AI Act) го позиционират като **високорисков инструмент** за българската държавна администрация. Моделът не притежава вградена правна обяснимост и не може да носи отговорност за грешни изходи.

При притегляне на ползите и недостатъците на модела, считаме, че внедряването му не е допустимо в публични сектор. Макар използването му в рутинни процедури, при които не се обработват лични и служебни данни, да носи сравнително нисък риск, а цената да е привлекателна спрямо тази на конкуренцията, цялостната му интеграция в конкретна институция от публичния сектор и необходимия ресурс за същото не биха донесли достатъчно ползи, които да оправдаят установените по-горе рискове.

ШАБЛОН 3 / 4

Карта за приложимост

AI модел в публичната администрация — функционален анализ

v1.0 — Работен шаблон

**Инструкция за попълване:** Попълнете всички полета за избория от вашата група AI модел. Празните редове и скалите са предназначени за ръчно попълване. Не изтривайте инструкциите в курсив.

Наименование на модела	Пример: Claude Opus 4.7 / GPT-5.4 / Gemini 3.1 Pro DeepSeek AI
Разработчик / компания	Пример: Anthropic / OpenAI / Google DeepMind Hangzhou DeepSeek AI
Версия за оценка	Точната версия, която групата изследва V3
Дата на попълване	ДД.ММ.ГГГГ 14.05.2026
Членове на групата	Имена на участниците Цветомир, Константин, Есма, Антония, Клаудия, Явор
Ръководител на групата	Константин

Част 1. Обзорна матрица на приложимостта

Бърза оценка по административни функции. Прил. = Приложимост 1–5. Риск = Н/С/В.

**Инструкция за попълване:** Попълнете всяка клетка: поставете оценка за приложимост (1=неприложимо, 5=напълно приложимо), ниво на риск (Н/С/В) и кратка бележка.

Административна функция	Обработка на документи	Правен анализ	Градски услуги	Анализ на данни	Програмиране	Обучение
Обработка на кореспонденция	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: □Н □С <u>□В</u> И изтичане на лични данни към външни сървъри. _	Прил.: ○1 <u>○2</u> ○3 ○4 ○5 Риск: □Н □С <u>□В</u> Опасност от погрешно класифициране на правни искания. _	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: <u>□Н</u> □С □В Генерира добри чернови, но изисква редакция.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: □Н □С <u>□В</u> Риск при автоматизирано извличане на метаданни.	Прил.: ○1 ○2 ○3 <u>○4</u> ○5 Риск: <u>□Н</u> □С □В Полезно за скриптове за сортиране на имейли.	Прил.: ○1 ○2 ○3 <u>○4</u> ○5 Риск: <u>□Н</u> □С □В Подходящо за шаблони за обучение на нови кадри.
Изготвяне на доклади и анализи	Прил.: ○1 ○2 ○3 <u>○4</u> ○5 Риск:	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск:	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск:	Прил.: ○1 ○2 ○3 ○4 <u>○5</u> Риск:	Прил.: ○1 ○2 ○3 ○4 <u>○5</u> Риск:	Прил.: ○1 ○2 ○3 ○4 <u>○5</u> Риск:

	☐ Н ☒ С ☐ В Ефективно резюмиране, но изисква проверка на източниците.	☐ Н ☐ С ☒ В Риск от включване на несъществуваща съдебна практика.	☐ Н ☒ С ☐ В Подготовка на публични отчети с обща информация.	☐ Н ☒ С ☐ В Изключителна способност за синтез на сложни масиви.	☒ Н ☐ С ☐ В Генериране на код за автоматични справки.	☒ Н ☐ С ☐ В Структуриране на аналитични материали за стажанти.
Тълкуване на нормативни актове	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☐ С ☒ В Риск от грешна интерпретация на административни и термини.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☐ С ☒ В Висок риск от „халюцинации“ и измислени алинеи.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☐ С ☒ В Може да подведе граждани с грешни указания.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☒ С ☐ В Полезно за категоризиране на норми по ключови думи.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☒ Н ☐ С ☐ В Помага за създаване на логически дървета на закона.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☒ С ☐ В Разясняване на общи правни концепции.
Изготвяне на правни становища	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☐ С ☒ В Опасност от пропуски в официалната документация.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☐ С ☒ В Липса на юридическа отговорност за изхода.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☐ С ☒ В Недопустимо за преки правни съвети към граждани.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☐ С ☒ В Риск при автоматизирана оценка на правни доказателства.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☐ С ☒ В Проверка на логическата последователност на текст.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☒ С ☐ В Упражнения за студенти по право.
Отговори на граждански запитвания	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☐ С ☒ В Автоматизиране на стандартизиран и отговори.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☒ С ☐ В Опасност от предоставяне на грешна правна информация.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☒ С ☐ В Полезен за чернови, но изисква човешки надзор.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☐ С ☐ В Анализ на тенденции в гражданските жалби.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☒ Н ☐ С ☐ В Възможно създаване на код за интеграция на системи за запитвания.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☐ С ☒ В Възможно обучение на служители за по-добра комуникация.
Чатбот за електронни услуги	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☒ С ☐ В Улеснена навигация в документи през чат.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☐ С ☒ В Неподходящо за интерпретация на права в чат.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☐ С ☒ В Трудности при интеграция с официални регистри.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☒ С ☐ В Събиране на обратна връзка от потребителите.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☒ С ☐ В Изграждане на архитектурата на самия бот.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☒ Н ☐ С ☐ В Симулации за работа с потребители.
Статистически анализ	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☒ С ☐ В Извличане на данни от текстови протоколи.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☒ С ☐ В Търсене на статистически зависимости в дела	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☐ С ☒ В Демографски профили на потребителите.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☒ С ☐ В Силна математика, но риск при качване на реални бази.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☐ С ☒ В Писане на скриптове (R, Python) за анализ.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☒ Н ☐ С ☐ В Обясняване на статистически методи.
Обработка на отчети	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☒ Н ☐ С ☐ В Отлично структуриране на седмични отчети.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☐ С ☒ В Опасност от изпускане на критични правни срокове.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☒ С ☐ В Резюмета на отчети за прозрачност.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☒ С ☐ В Ефективно сравняване на данни от различни звена.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☒ С ☐ В Автоматизация на форматирането на отчети.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☒ Н ☐ С ☐ В Инструкции за правилно отчитане.
Автоматизация на работни процеси	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☒ С ☐ В Създаване на логически схеми за документооборот.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☐ С ☒ В Автоматизацията изисква строга правна рамка.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☒ С ☐ В Оптимизация на пътя на гражданина.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☒ С ☐ В Проследяване на ефективността на веригите.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☐ С ☒ В Най-силно приложение при писане на код.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☒ Н ☐ С ☐ В Уроци по оптимизация на времето.

Разработка на вътрешни инструменти	Прил.: ○1 ○2 ○3 <u>○4</u> ○5 Риск: ☐ Н ☒ С ☐ В Документация за софтуерни нужди.	Прил.: ○1 <u>○2</u> ○3 ○4 ○5 Риск: ☐ Н ☒ С ☐ В Трябва да съответства на строги ИТ наредби.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☒ С ☐ В Инструменти за вътрешна статистика.	Прил.: ○1 ○2 ○3 <u>○4</u> ○5 Риск: ☐ Н ☒ С ☐ В Модули за обработка на масиви.	Прил.: ○1 ○2 ○3 ○4 <u>○5</u> Риск: ☐ Н ☒ С ☐ В Бързо създаване на инструменти от ИТ отдела.	Прил.: ○1 ○2 ○3 ○4 <u>○5</u> Риск: ☐ Н ☒ С ☐ В Техническо обучение за новия софтуер.
Учебни материали и тестове	Прил.: ○1 ○2 ○3 ○4 <u>○5</u> Риск: ☐ Н ☒ С ☐ В Генериране на въпроси по текстове на наредби.	Прил.: ○1 ○2 ○3 <u>○4</u> ○5 Риск: ☐ Н ☒ С ☐ В Казуси за обучение (изисква проверка).	Прил.: ○1 ○2 ○3 ○4 ○5 Риск: ☐ Н ☒ С ☐ В Информационни брошури за граждани.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☒ С ☐ В Анализ на резултатите от тестове.	Прил.: ○1 ○2 ○3 ○4 <u>○5</u> Риск: ☐ Н ☒ С ☐ В Генериране на задачи по програмиране.	Прил.: ○1 ○2 ○3 ○4 <u>○5</u> Риск: ☐ Н ☒ С ☐ В Идеално за вътрешни държавни академии.
Наръчници и процедури	Прил.: ○1 ○2 ○3 ○4 ○5 Риск: ☐ Н ☒ С ☐ В Помага за структурата, но изисква редакция.	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☒ С ☐ В Процедурите трябва да са юридически изрядни.	Прил.: ○1 ○2 ○3 ○4 ○5 Риск: ☐ Н ☒ С ☐ В Ясни инструкции за ползване на услуги.	Прил.: ○1 ○2 ○3 ○4 ○5 Риск: ☐ Н ☒ С ☐ В Описание на методологии за данни.	Прил.: ○1 ○2 ○3 ○4 ○5 Риск: ☐ Н ☒ С ☐ В Техническо ръководство за софтуер.	Прил.: ○1 ○2 ○3 ○4 <u>○5</u> Риск: ☐ Н ☒ С ☐ В Създаване на образователни наръчници.
Превод и локализация	Прил.: ○1 ○2 ○3 ○4 ○5 Риск: ☐ Н ☒ С ☐ В Обработва добре документите	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☒ С ☐ В Не познава законодателствот о добре	Прил.: ○1 ○2 ○3 ○4 ○5 Риск: ☐ Н ☒ С ☐ В Ограничени възможности за превод	Прил.: ○1 ○2 ○3 ○4 ○5 Риск: ☐ Н ☒ С ☐ В Непълнен анализ, възможни течове	Прил.: ○1 ○2 ○3 <u>○4</u> ○5 Риск: ☐ Н ☒ С ☐ В Ограничен мащаб на анализа	Прил.: ○1 ○2 ○3 ○4 ○5 Риск: ☐ Н ☒ С ☐ В Превод на обучителни материали
Протоколи и резюмета	Прил.: ○1 ○2 ○3 ○4 ○5 Риск: ☐ Н ☒ С ☐ В инструмент за автоматизирано резюмиране	Прил.: ○1 ○2 <u>○3</u> ○4 ○5 Риск: ☐ Н ☒ С ☐ В Помага за извличане на правни аргументи	Прил.: ○1 ○2 ○3 ○4 ○5 Риск: ☐ Н ☒ С ☐ В Идеален за трансформир ане на сложни и дълги протоколи	Прил.: ○1 ○2 ○3 ○4 ○5 Риск: ☐ Н ☒ С ☐ В бързо окрупняване на информация и откриване на повтарящи се теми	Прил.: ○1 ○2 ○3 <u>○4</u> ○5 Риск: ☐ Н ☒ С ☐ В Помага с написването на скриптове	Прил.: ○1 ○2 ○3 ○4 ○5 Риск: ☐ Н ☒ С ☐ В Превръща дългите стенограми и протоколи в синтезирани учебни казуси

Част 2. Детайлен анализ по случаи на употреба

За всеки сценарий оценете конкретно как избраният модел се справя.

2.1. Обработка на административни документи Способността на модела да чете, структурира, резюмира и изготвя административни документи — решения, наредби, докладни записки, протоколи. Оценка за избрания модел: 1. Може ли моделът да обработва документи на български език с административна терминология? 2. Поддържа ли обработка на PDF/Word файлове с дълъг обем? 3. Запазва ли форматирането и структурата при генериране на изходен документ? 4. Изисква ли специализирано prompt инженерство за административен контекст? Отговори / наблюдения: Моделът разбира на добро ниво българска административна терминология и дълги текстове (128k контекст). Въпреки това, той не запазва автоматично специфичното държавно форматиране (бланки, отстъпи) и изисква готови шаблони (prompt инженерство).
--

Препоръка:

- ☐ Препоръчва се без ограничения
- ☐ Препоръчва се с ограничения (опишете):

☐ **Не се препоръчва (обосновка):**

Строга забрана за подаване на лични (GDPR) или класифицирани данни в облачната версия.

2.2. Правен анализ и тълкуване

Способността да анализира нормативни актове, да открива противоречия, да прилага правни норми към конкретни случаи.

Оценка за избрания модел:

1. Колко точен е при тълкуване на български законодателни текстове?
2. Може ли да следи промени в нормативната база (обновени закони)?
3. Как се справя с многозначност и правна неяснота?
4. Маркира ли ясно когато излиза извън областта на своята компетентност?

Отговори / наблюдения:

Ниска точност при специфично българско право. Моделът не следи „Държавен вестник“ в реално време и има склонност да „халюцинира“ - измисля убедително звучащи членове и алинеи, вместо да откаже отговор.

Препоръка:

- ☐ Препоръчва се без ограничения
- ☐ Препоръчва се с ограничения (опишете):

☐ **Не се препоръчва (обосновка):**

Риск от генериране на незаконосъобразни актове и грешни правни тълкувания

2.3. Взаимодействие с граждани (чатбот / асистент)

Способността да комуникира с граждани — отговаря на въпроси, помага при подаване на заявления, обяснява административни процедури.

Оценка за избрания модел:

1. Разбира ли естествен разговорен български (включително диалекти и грешки)?
2. Дава ли ясни и разбираеми отговори без административен жаргон?
3. Как се справя с чувствителни теми (социални помощи, здраве, правосъдие)?
4. Може ли да се интегрира с портали за електронни услуги?

Отговори / наблюдения:

Справя се с разговорен български, печатни грешки и жаргон. Успешно превежда сложния институционален език на достъпен за гражданите стил. API интерфейсът е евтин и лесен за интеграция.

Препоръка:

- ☐ Препоръчва се без ограничения
- ☐ Препоръчва се с ограничения (опишете):

Добър за първо ниво на обслужване. Задължителна бележка, че отговорите нямат силата на административен акт.

☐ **Не се препоръчва (обосновка):**

2.4. Анализ на данни и отчетност

Способността да обработва структурирани данни, да генерира отчети и да идентифицира тенденции.

Оценка за избрания модел:

1. Може ли да обработва Excel/CSV данни и да генерира анализи?
2. Поддържа ли визуализация на данни (графики, диаграми)?
3. Колко надежден е при числови изчисления и статистика?
4. Може ли да се свърже с вътрешни бази данни чрез API?

Отговори / наблюдения:

Чете отлично структурирани формати (CSV, JSON) и открива аномалии. Не генерира графики нативно, но пише перфектен код за тяхната визуализация. Притежава силна математическа логика за статистика.

Препоръка:

- ☐ Препоръчва се без ограничения
- ☐ **Препоръчва се с ограничения (опишете):**

Полезен за анализ на големи, обобщени вътрешни отчети. Да не се качват сурови масиви с данъчна, финансова или медицинска информация на граждани.

- ☐ **Не се препоръчва (обосновка):**

2.5. Вътрешно обучение на служители

Способността да генерира учебни материали, тестове, наръчници и да поддържа интерактивно обучение.

Оценка за избрания модел:

1. Може ли да генерира тематични курсове и тестове за служители?
2. Адаптира ли съдържанието към различни нива на компетентност?
3. Поддържа ли формати за дистанционно обучение (SCORM и др.)?
4. Как се справя с чувствителни теми — антикорупция, етика, дискриминация?

Отговори / наблюдения:

Генерира планове за курсове, казуси и тестове, като лесно адаптира трудността според отдела. Има строги филтри за безопасност и генерира обективно съдържание по теми като етика и корупция.

Препоръка:

- ☐ Препоръчва се без ограничения
- ☐ **Препоръчва се с ограничения (опишете):**

Добър помощен инструмент, но трябва да се използва внимателно по отношение на наличните данни в организацията

- ☐ **Не се препоръчва (обосновка):**

2.6. Програмиране и автоматизация

Способността да пише код, да автоматизира процеси и да поддържа вътрешни системи.

Оценка за избрания модел:

1. Поддържа ли езиките и технологиите, използвани в административните системи?

2. Може ли да чете и документира съществуващ (legacy) код?

3. Колко надеждни са генерираните скриптове без допълнителна проверка?

4. Поддържа ли agentic режим за автономно изпълнение на задачи?

Отговори / наблюдения:

Добър модел за код. Поддържа всички технологии в електронното управление, чете отлично остарял код и е добър за изграждане на автономни AI агенти поради ниската си цена.

Препоръка:

☐ Препоръчва се без ограничения

☐ Препоръчва се с ограничения (опишете):

☐ **Не се препоръчва (обосновка):**

Да не се използва в неизолирана среда, поради наличните рискове.

Част 3. Оценка на интеграционния потенциал

 **Инструкция за попълване:** *Оценете техническите възможности за интеграция с административната инфраструктура.*

Интеграционен аспект	Описание	Поддръжка	Бележки
API свързване	REST API, GraphQL, webhooks — интеграция с вътрешни системи	☐ <u>Да</u> ☐ Не ☐ Частично	Предоставя високопроизводително REST API,, което прави интеграцията в съществуващи системи лесна и евтина.
Single Sign-On	Съвместимост с Active Directory / ADFS / Keycloak	☐ Да ☐ Не ☐ <u>Частично</u>	Облачната платформа на DeepSeek няма нативна enterprise поддръжка за Active Directory/Keycloak. Ако обаче моделът се внедрява локално или се достъпва през вътрешен прокси софтуер.
On-premise deployment	Възможност за локален хостинг без облачна зависимост	☐ <u>Да</u> ☐ Не ☐ Частично	Възможно е пълно локално внедряване в държавен дата център, но изисква колосален хардуерен ресурс, за да работи с пълния си капацитет.
Данни в ЕС	Сървъри на европейска територия — GDPR съответствие	☐ Да ☐ <u>Не</u> ☐ Частично	Официалните сървъри за облачното API на компанията се намират извън Европейския съюз, което възпрепятства съответствието с GDPR.

Одитна следа	Логиране на всички взаимодействия за отчетност	☐ Да ☐ Не ☒ <u>Частично</u>	Облачното API пази логове за таксуване на токени, но не предоставя детайлен административен контрол над потребителските заявки за нуждите на държавния одит. Пълно логиране е възможно само при собствена локална инфраструктура.
Ролеви достъп	RBAC — различни нива на достъп за различни служители	☐ Да ☒ <u>Не</u> ☐ Частично	В облачната си версия предоставя единен API ключ за акаунт. Правата и нивата на достъп трябва да бъдат програмирани допълнително от вътрешните IT екипи в администрацията.
Многоезичност	Поддръжка на български и официални ЕС езици	☒ <u>Да</u> ☐ Не ☐ Частично	Добра нативна поддръжка на български език, английски и основните официални езици в ЕС, без нужда от допълнителни модули за превод, но се необходимост от човешки анализ.
Мобилен достъп	Работа от мобилни устройства (field work)	☒ <u>Да</u> ☐ Не ☐ Частично	Тъй като работи като уеб услуга (API), моделът може лесно да се интегрира в мобилни приложения за служители на терен, стига да имат достъп до интернет или до вътрешната мрежа (VPN).
Офлайн режим	Функционалност без постоянна интернет връзка	☐ Да ☒ <u>Не</u> ☐ Частично	Ииска огромна изчислителна мощ. Облачната версия не работи офлайн, а локалната изисква непрекъсната връзка на крайните устройства с вътрешния сървърен център.
SLA и поддръжка	Гарантирано ниво на обслужване за публичния сектор	☐ Да ☒ <u>Не</u> ☐ Частично	DeepSeek не предлага специфични SLA или директна техническа поддръжка, съобразени с нуждите на европейския публичен сектор. При срив в облачната система администрацията няма

			правни инструменти за компенсация.
--	--	--	------------------------------------

Обобщена карта за приложимост

Топ 3 приложения с висок потенциал:

- 1. Програмиране, автоматизация и поддръжка на код
- 2. Вътрешно обучение и генериране на обучителни материали
- 3. Обработка на документи и резюмиране на неклафицирана информация

Функции, при които моделът НЕ е подходящ:

- 1. Правен анализ, тълкуване и становища
- 2. Работа с лични данни (GDPR) и класифицирана информация
- 3. Автоматизирани решения за граждани без човешки контрол

Препоръчани условия за пилотно внедряване:

Задължителен човешки контрол - моделът трябва да се използва единствено като асистент, който подготвя чернови. Никаква реална експлоатация без верификация от служител.

Пълна анонимизация на данните - преди изпращане на заявки, да бъдат заличени всички ЕГН-та, имена, адреси и финансова информация.

Изолирана тестова среда - всички генерирани от модела програмни кодове и автоматизации да се тестват в сигурна среда, отделена от съответната държавна инфраструктура.

ШАБЛОН 4 / 4

Методологични указания

Етично и прозрачно ползване на AI модели в публичната администрация

v1.0 — Работен шаблон

Инструкция за попълване: Попълнете всички полета за избория от вашата група AI модел. Празните редове и скалите са предназначени за ръчно попълване. Не изтривайте инструкциите в курсив.

аименование на модела	Пример: Claude Opus 4.7 / GPT-5.4 / Gemini 3.1 Pro DeepSeek AI
Разработчик / компания	Пример: Anthropic / OpenAI / Google DeepMind Hangzhou DeepSeek AI
Версия за оценка	Точната версия, която групата изследва V3
Дата на попълване	ДД.ММ.ГГГГ 20.05.2026
Членове на групата	Имена на участниците Цветомир, Константин, Есма, Антония, Клаудия, Явор
Ръководител на групата	Константин

Предговор и обхват

Инструкция за попълване: Попълнете идентификационните данни на вашата организация и дефинирайте обхвата на документа.

Наименование на организацията	Пълно наименование
Административно ниво	Централна / областна / общинска администрация
AI модел(и), обхванати от документа	Посочете конкретните модели/версии DeepSeek V3
Версия на документа	1.0
Дата на приемане	ДД.ММ.ГГГГ
Следваща планирана ревизия	ДД.ММ.ГГГГ
Одобрен от	Длъжност и имена
Отговорен за прилагането	Длъжност, звено и контакти

Чл. 1. Предмет и обхват

Опишете за какво се отнася документът — кои AI системи, кои структурни звена и кои видове задачи обхваща.

Настоящите указания определят задължителните правила, технически и организационни мерки за сигурно, прозрачно и законосъобразно използване на езиковия модел DeepSeek V3. Документът се прилага за всички структурни звена в рамките на общата и специализирана администрация. Обхватът на задачите включва

обработка на неklasифицирани текстови масиви, стилистична обработка, езикови преводи, подпомагане на ИТ дейностите и вътрешно-административни функции.

Чл. 2. Правна и регулаторна рамка

Посочете нормативните актове, на базата на които се издават указанията (EU AI Act, GDPR, ЗЕУ, вътрешни наредби).

Указанията се издават в съответствие с Регламента на ЕС за изкуствения интелект (EU AI Act), Регламент (ЕС) 2016/679 (GDPR), Закона за електронното управление (ЗЕУ), Закона за киберсигурност (ЗКС), Наредбата за общите изисквания към информационните системи в администрацията и Вътрешните правила за документооборота и сигурността на информацията.

Раздел I. Принципи за етично и прозрачно ползване

Определете основните принципи, на които се базира политиката. Попълнете конкретните мерки за всеки.

Основни принципи	
1 Законност	AI системите се използват само в рамките на действащото законодателство. Всяко решение, подпомогнато от AI, спазва приложимите правни норми. Конкретни мерки в нашата организация: Забранява се използването на DeepSeek V3 за автоматизирано вземане на решения. Моделът да се ползва само за спомагателни технически задачи и подготовка на чернови.
2 Прозрачност	Гражданите и служителите имат право да знаят кога AI участва в процесите. Резултатите от AI трябва да са обяснени разбираемо. Конкретни мерки в нашата организация: Всички официални методологически разяснения или писма, изготвени с първоначална помощ от AI, трябва да съдържат маркировка. Външните чатботове за граждани задължително визуализират съобщение, че отговорите са генерирани от ИИ.
3 Отчетност	Ясна верига на отговорност. Служителите носят отговорност за решенията, взети с AI помощ — AI не замества човешката отговорност. Конкретни мерки в нашата организация: Водещ принцип в експлоатацията на ИИ модела е личната отговорност на ползващият го служител. При използване на текст генериран от DeepSeek V3 в официален документ, същото да бъде отразено по надлежащия ред, а документът да бъде подписан с личния КЕП на служителя. Същият носи пълна дисциплинарна и административна отговорност за верността му.
4 Справедливост	AI инструментите не дискриминират. Резултатите се проверяват редовно за пристрастия (bias) спрямо различни групи граждани. Конкретни мерки в нашата организация: Промптовете към модела не трябва да съдържат дискриминационни критерии. Изходите се подлагат на редовен тримесечен преглед от комисия с цел недопускане на систематичен предразсъдък спрямо уязвими групи от граждани.
5 Поверителност	Личните данни на гражданите се обработват само при наличие на правно основание и при спазване на GDPR. Конкретни мерки в нашата организация: Внедрява се филтър на ниво ИТ инфраструктура. Забранява се изпращането към облачното API на DeepSeek на лични данни и чувствителна служебна информация.
6 Надеждност	AI системите се използват само за задачи, при които точността е потвърдена. Резултатите подлежат на верификация. Конкретни мерки в нашата организация: Поради висок риск от халюцинации, всеки цитиран закон, член, алинея или решение на се подлага на ръчна кръстосана проверка с правно-информационни системи преди финализиране.

7 Пропорционалност	Нивото на автоматизация съответства на риска от грешки. При по-висок риск — по-задължителен човешки контрол. Конкретни мерки в нашата организация: Автоматизацията да се ограничи до задачи с нисък риск. Задачи с висока правна стойност преминават през експертен филтър.
8 Обяснимост	При поискване от гражданин или орган, организацията може да обясни как е бил използван AI при дадено решение. Конкретни мерки в нашата организация: ИТ отделът поддържа техническа документация за системните промптове и контекстуалната база от данни (RAG), използвани за захранване на модела.

Раздел II. Разрешени и забранени употреби

Определете за всяка категория задачи дали е разрешена, допустима с условия или забранена.

 Инструкция за попълване: За всяка задача: поставете отметка в съответната колона и опишете условията или ограниченията.		
Категория задача	Статус	Условия / Ограничения / Процедура
Административни документи		
Изготвяне на проекти за решения	☐ Разрешено ☐ <u>С условия</u> ☐ Забранено	Условия / ограничения / процедура: Само за структуриране на технически шаблони.
Резюмиране на входяща поща	☐ Разрешено ☐ <u>С условия</u> ☐ Забранено	Условия / ограничения / процедура: Допустимо само за общи запитвания.
Превод на документи	☐ <u>Разрешено</u> ☐ С условия ☐ Забранено	Условия / ограничения / процедура: За превод на европейски директиви, чуждестранни добри практики и неклассифицирани ръководства.
Генериране на официални писма	☐ Разрешено ☐ <u>С условия</u> ☐ Забранено	Условия / ограничения / процедура: Използва се само за стилистично оформяне. Финалният текст задължително преминава през преглед от експертен служител..
Правни и нормативни въпроси		
Анализ на нормативни текстове	☐ Разрешено ☐ <u>С условия</u> ☐ Забранено	Условия / ограничения / процедура: Допустимо единствено чрез подаване на точния правен текст директно в промпта, за да се избегнат фактически грешки от модела.
Проверка за вътрешни противоречия	☐ Разрешено ☐ <u>С условия</u> ☐ Забранено	Условия / ограничения / процедура: За откриване на разминавания в проекти за вътрешни правила.
Изготвяне на правни становища	☐ Разрешено ☐ С условия ☐ <u>Забранено</u>	Условия / ограничения / процедура: Моделът халюцинира правни норми и не познава в детайли съдебната практика..
Тълкуване на разпоредби към граждани	☐ Разрешено ☐ С условия ☐ <u>Забранено</u>	Условия / ограничения / процедура: Тълкуването на закона е изключителна функция на експертния служител. Рискът от подвеждане е висок
Вземане на решения с ефект върху граждани		
Преценка за социални помощи	☐ Разрешено ☐ С условия ☐ <u>Забранено</u>	Условия / ограничения / процедура: Изрично забранено съгласно изискванията на Регламента на ЕС за изкуствения интелект.
Оценка на проекти и заявления	☐ Разрешено ☐ С условия ☐ <u>Забранено</u>	Условия / ограничения / процедура: Оценката на допустимост се извършва единствено от определените по закон служители на администрацията..

Наказателни и принудителни мерки	☐ Разрешено ☐ С условия ☐ <u>Забранено</u>	Условия / ограничения / процедура: Изцяло извън обхвата. Риск от тежки закононарушения.
Автоматизирани решения без човешки контрол	☐ Разрешено ☐ С условия ☐ <u>Забранено</u>	Условия / ограничения / процедура: Забранено съгласно изискванията на GDPR и ограниченията на публичния сектор в EU AI Act.
Комуникация с граждани		
Чатбот за информационни въпроси	☐ Разрешено ☐ <u>С условия</u> ☐ Забранено	Условия / ограничения / процедура: Допустимо само за базови въпроси.
Проекти за отговори на жалби	☐ Разрешено ☐ С условия ☐ <u>Забранено</u>	Условия / ограничения / процедура: Жалбите съдържат чувствителна информация и лични данни - обработката им е недопустима.
Публикации в социални медии	☐ <u>Разрешено</u> ☐ С условия ☐ Забранено	Условия / ограничения / процедура: За оформяне на анонси или съобщения за техническа поддръжка на системите..
Кризисна комуникация	☐ Разрешено ☐ С условия ☐ <u>Забранено</u>	Условия / ограничения / процедура: При кризисни ситуации или аварии съобщенията се изготвят от ПР звеното с цел гарантиране на 100% точност.
Вътрешно-административни функции		
Обучение на служители	☐ <u>Разрешено</u> ☐ С условия ☐ Забранено	Условия / ограничения / процедура: За генериране на изпитни тестове, казуси и обучителни презентации..
Анализ на данни и отчети	☐ Разрешено ☐ <u>С условия</u> ☐ Забранено	Условия / ограничения / процедура: Допустимо само след пълна анонимизация на чувствителни, статистически и други, посочени от ръководния състав данни.
Автоматизация на повтарящи се задачи	☐ <u>Разрешено</u> ☐ С условия ☐ Забранено	Условия / ограничения / процедура: За генериране на скриптове, форматиране на файлове и софтуерно тестване.
Обработка на класифицирана информация	☐ Разрешено ☐ С условия ☐ <u>Забранено</u>	Условия / ограничения / процедура: Абсолютна забрана. Нарушение на Закона за защита на класифицираната информация (ЗЗКИ).

Раздел III. Процедури за одобрение и верификация

Чл. 10. Задължителна верификация

Опишете кои задачи изискват задължителна проверка от служител преди използване на AI изхода.

Задължителна човешка проверка се прилага за: Всички програмни кодове, написани с помощта на DeepSeek V3, всеки статистически отчет за разходвани публични средства, структуриран от модела, текстове на писма, наръчници или отговори на въпроси..

Чл. 11. Процедура за одобрение на нови употреби

Как служителят може да заяви одобрение за нов вид употреба на AI, непредвиден в настоящите указания?

Когато е идентифицирана нова нужда да се подаде писмено искане до отговорните за внедряването и поддръжката на ИИ служители. Искането трябва да съдържа: описание на задачата, ползи, примерен промпт и оценка на риска за лични данни. Одобрението се дава след съвместно становище на ИТ отдела, Правното звено и служителят по сигурност на информацията.

Чл. 12. Задължително разкриване към граждани

В кои случаи гражданинът трябва да бъде информиран, че AI е участвал в обработката на неговото искане?

Администрацията информира гражданите във всички случаи, в които те взаимодействат директно с ИИ система.

Раздел IV. Мониторинг, одит и отчетност

Минимални изисквания за одитна следа (отбележете приложимите):

- ☐ Всяко взаимодействие с AI системата се записва с дата, час и потребител
- ☐ Пазят се логове на AI изходите, използвани в официални документи
- ☐ При автоматизирани решения се съхранява AI входа и изхода
- ☐ Провежда се периодичен преглед на качеството на AI изходите (мин. 1х/тримесечие)
- ☐ Инцидентите, свързани с AI грешки, се докладват на определено звено
- ☐ Годишен отчет за употребата на AI се изготвя и публикува

Чл. 15. KPI за мониторинг на AI употребата

Определете конкретни измерими показатели: брой взаимодействия, % задачи с верификация, % открити грешки и др.
Общ обем на заявките: За следене на ИТ натоварването и разходите.
Процент на верифицираните документи:
Коефициент на AI грешки/инциденти: Брой докладвани случаи на халюцинации или грешен код.

Раздел V. Отговорности и управление

Роля	Отговорност	Конкретно лице / звено
Ръководител на организацията	Стратегическо одобрение и крайна отговорност за AI политиката	Конкретно лице / звено: Директор
AI отговорник (AI Officer)	Координация, обучение, мониторинг, докладване на инциденти	Конкретно лице / звено: Специализиран ИТ експерт, ССИ
Служители, ползващи AI	Спазване на указанията, верификация на изходите, докладване на проблеми	Конкретно лице / звено: Определени със заповед служители на общата и специализираната администрация
ИТ отдел / системен администратор	Технически достъп, интеграция, сигурност, логове	Конкретно лице / звено: Служител по мрежова и информационна сигурност
Правно звено	Преглед на нови употреби, съответствие с нормативната рамка	Конкретно лице / звено: Директор на правна дирекция
ДЛЗД (длъжностно лице по защита на данни)	GDPR съответствие при обработка на лични данни	Конкретно лице / звено: ССИ
Вътрешен одит	Периодичен независим преглед на AI употребата	Конкретно лице / звено: Финансов контролор, експерт от звено за вътрешен одит

--	--	--

Раздел VI. Обучение и изграждане на капацитет

Чл. 20. Задължително обучение			
Посочете кои служители трябва да преминат обучение, какъв е минималният обем и как се документира. Всички служители имащи достъп до DeepSeek V3, преминават задължителен въстпителен курс, организиран съвместно с Института по публична администрация (ИПА) или вътрешния ИТ отдел. Обучението се удостоверява с протокол и сертификат.			
Модул на обучение	Целева група	Форма и обем	Периодичност
Основи на AI и рискове	Всички потребители	Лекционна, 3 часа	Веднъж годишно
Работа с конкретния модел	ИТ експерти и методолози	Работни семинари (Промптване), 4 часа	При всяка нова версия на модела
Етика и нормативна рамка	Правен отдел и ръководство	Семинари (AI Act), 2 часа	При промяна в законодателството
Разпознаване на AI грешки	Експерти в звената за „Мониторинг“	Практически казуси, 2 часа	На шест месеца
Защита на личните данни	Всички служители с достъп, 2 часа	Онлайн курс, 1 час	Веднъж годишно

Раздел VII. Актуализация и валидност

Чл. 24. Условия за задължителна ревизия			
Опишете кога документът трябва задължително да се преразгледа: нова версия на модела, регулаторна промяна, инцидент, изтичане на срок. Настоящият документ да се преразглежда и актуализира задължително при: - Излизане на нова основна версия на модела - Официално влизане в сила на нови санкции или насоки. - Регистриран сериозен инцидент (изтичане на служебни данни или системна халюцинация, довела до грешно вътрешно решение).			
Чл. 25. История на промените			
Документирайте всички версии на документа с дата, автор и описание на промяната.			
Версия	Дата	Описание на промяната	Одобрил
1.0		Първоначално приемане на Указанията за експлоатация на DeepSeek V3.	Директор на организацията

--	--	--	--



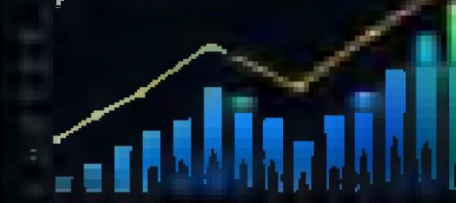
анализ на данни



успешност

наръчник за използване на DeepSeek V3 в публичния сектор

ефективност



основни принципи



баланс



прозрачност



публична
отчетност



справедливост



конфиденциалност



надеждност



пропорционалност

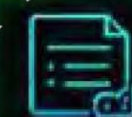


обяснимост

политика
на
употреба

правила за допустимост

разрешени задачи



превод на
официални документи



създаване на
учебни казуси
и тестове



подготовка на
ит стратегии в
изолирана среда



стилистично
оформяне за
социални медии



забранени задачи



автоматизирани
решения за граждани



изготвяне на правни
становища и анализи



обработка на жалби
и запитвания от
граждани



обработка на
държавни тайни и
класифицирана
информация

ключови процедури



човешки
контрол и
верификация



зареждане на
официални
нормативни бази



анонимизиране
на лични данни



тестване на код
в изолирана
тестова среда

отговорности



стратегия и
одобрение



мониторинг
и одит



ит екип



правно
звено

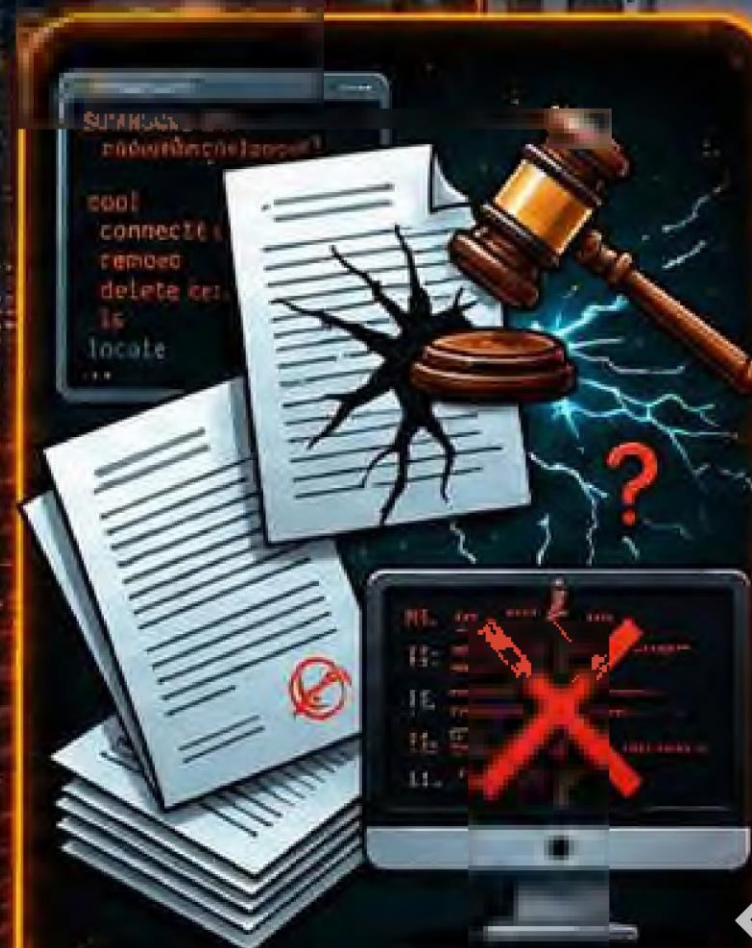


длъжностно
лице по защита
на данните



КРИТИЧНИ РИСКОВЕ И ГРАНИЦИ ЗА DEEPSEEK V3 В ПУБЛИЧНИЯ СЕКТОР: АНАЛИЗ И СЪОБРАЖЕНИЯ

WARNING
SECURITY BREACH
DETECTED



ГЕНЕРИРАНЕ НА ГРЕШНИ
И ФАБРИКУВАНИ
ПРАВНИ ТЕКСТОВЕ

ALERT
UNAUTHORIZED ACCESS
ATTEMPT



ПОДВЕЖДАНЕ НА
ГРАЖДАНТЕ ПО
АДМИНИСТРАТИВНИ ВЪПРОСИ



СТРАТЕГИЧЕСКА И
ГЕОПОЛИТИЧЕСКА
ЗАВИСИМОСТ

ПОЛИТИКА
НА УПОТРЕБА

УПОТРЕБАТА СЪЗДАВА НЕПРИЕМЛИВ РИСК ЗА ПУБЛИЧНИЯ СЕКТОР И ОБЩЕСТВОТО

ГРУПА "ХЕФЕСТ"
xAI Grok (4.20/4)
и Mistral
(Large3/Devstral2)



SWOT Анализ

Оценка на фундаментален AI модел за публичната администрация

v1.0 — Работен шаблон

 **Инструкция за попълване:** Попълнете всички полета за избория от вашата група AI модел. Празните редове и скалите са предназначени за ръчно попълване. Не изтривайте инструкциите в курсив.

Наименование на модела	Пример: Claude Opus 4.7 / GPT-5.4 / Gemini 3.1 Pro MISTRAL (Large 3 / Pevstral 2)
Разработчик / компания	Пример: Anthropic / OpenAI / Google DeepMind MISTRAL AI
Версия за оценка	Точната версия, която групата изследва MISTRAL Large 3
Дата на попълване	ДД.ММ.ГГГГ 14.05.2026 г.
Членове на групата	Имена на участниците Хефест – Манол, Николай, Делян, Диан, Наталия, Силвия, Петър
Ръководител на групата	Наталия

A. Обща приложимост и функционален обхват

Оценете доколко моделът покрива реални нужди в публичния сектор.

 **Инструкция за попълване:** За всеки критерий: поставете кръгче на избраната оценка (1=много слабо, 5=отлично) и запишете кратка обосновка.

Критерий		
A1. Приложимост в публичния сектор		
Обработка на документи доклади, решения, наредби	Оценка 1–5 ○1 ○2 ○3 ○4 x 5	Обосновка / бележки: Добра работа с големи текстове
Правен анализ тълкуване на нормативни актове	Оценка 1–5 ○1 ○2 ○3 x 4 ○5	Обосновка / бележки: По-подходящ за структурирани задачи
Граждански услуги чатбот, Q&A, помощ при заявления	Оценка 1–5 ○1 ○2 ○3 x 4 ○5	Обосновка / бележки: Добър за Q&A системи

Вътрешна комуникация <i>резюме, протоколи, кореспонденция</i>	Оценка 1–5 ○1 ○2 ○3 ○4 x5	Обосновка / бележки: Генерира ясни текстове _____
Анализ на данни <i>отчети, статистика, мониторинг</i>	Оценка 1–5 ○1 ○2 ○3 x4 ○5	Обосновка / бележки: Добър при отчети и структуриране _____
Превод <i>многоезична комуникация, ЕС документи</i>	Оценка 1–5 ○1 ○2 ○3 x4 ○5	Обосновка / бележки: Силна многоезична поддръжка _____
Програмиране / автоматизация <i>вътрешни системи, скриптове</i>	Оценка 1–5 ○1 ○2 ○3 ○4 x5	Обосновка / бележки: Много добър при код и API _____
Обучение на служители <i>учебни материали, тестове</i>	Оценка 1–5 ○1 ○2 ○3 ○4 x5	Обосновка / бележки: Подходящ за обучителни материали _____
A2. Функции на платформата		
Текстово генериране <i>качество и прецизност</i>	Оценка 1–5 ○1 ○2 ○3 ○4 x5	Обосновка / бележки: Високо качество и добра структура _____
Reasoning / разсъждение <i>логически вериги, многостъпкови задачи</i>	Оценка 1–5 ○1 ○2 ○3 x4 ○5	Обосновка / бележки: Силен при логически задачи _____
Code writing <i>генериране и анализ на код</i>	Оценка 1–5 ○1 ○2 ○3 ○4 x5	Обосновка / бележки: Много добри резултати за код и автоматизация _____
Анализ на изображения <i>vision, диаграми, скрийншоти</i>	Оценка 1–5 ○1 ○2 x3 ○4 ○5	Обосновка / бележки: Зависи от конкретния модел _____
Аудио / глас <i>говорим вход/изход</i>	Оценка 1–5 ○1 x2 ○3 ○4 ○5	Обосновка / бележки: По-ограничена поддръжка _____
Видео разбиране <i>анализ на видеосъдържание</i>	Оценка 1–5 ○1 x2 ○3 ○4 ○5	Обосновка / бележки: Няма силен фокус _____
Дълги документи <i>обработка на обемни файлове</i>	Оценка 1–5 ○1 ○2 ○3 ○4 x5	Обосновка / бележки: Силен при големи документи _____
Agentic режим <i>автономни многостъпкови задачи</i>	Оценка 1–5 ○1 ○2 ○3 x4 ○5	Обосновка / бележки: Добра поддръжка на сложни процеси _____

API / системна интеграция <i>свързване с вътрешни системи</i>	Оценка 1–5 ○1 ○2 ○3 ○4 x5	Обосновка / бележки: Една от силните страни
Plugin екосистема <i>налични разширения и конектори</i>	Оценка 1–5 ○1 ○2 x3 ○4 ○5	Обосновка / бележки: По-ограничена от големите екосистеми
Desktop приложение <i>нативна desktop среда</i>	Оценка 1–5 ○1 x2 ○3 ○4 ○5	Обосновка / бележки: Няма силно развита крайна среда
Mobile приложение <i>iOS / Android достъп</i>	Оценка 1–5 ○1 x2 ○3 ○4 ○5	Обосновка / бележки: По-ограничено потребителско приложение
Cowork / колаборация <i>екипна работа в реално време</i>	Оценка 1–5 ○1 ○2 x3 ○4 ○5	Обосновка / бележки: Основни възможности
Търсене в интернет <i>актуална информация от мрежата</i>	Оценка 1–5 ○1 ○2 x3 ○4 ○5	Обосновка / бележки: Не е основна силна страна
Fine-tuning <i>настройка за специфична среда</i>	Оценка 1–5 ○1 ○2 ○3 ○4 x5	Обосновка / бележки: Силни възможности за адаптация

B. SWOT Матрица

Обобщете основните силни/слаби страни, възможности и заплахи на модела в публичен контекст.

 **Инструкция за попълване:** Запишете поне 4 наблюдения в клетка. Базирайте се на попълнените оценки от Раздел А.

S STRENGTHS — Силни страни - Силни API възможности - Гъвкаво внедряване - Добри модели за код - Висока скорост - Възможност за локално използване при някои модели - Добра интеграция Допълнителни бележки: няма	W WEAKNESSES — Слаби страни - По-слабо разпознаване сред масови потребители - Неравномерно качество на български език - Изисква техническа експертиза - По-малка еко система - Възможни халюцинации - Ограничена прозрачност Допълнителни бележки: няма
---	--

O OPPORTUNITIES — Възможности 1. Локално внедряване в институции 2. Автоматизация на документи 3. Анализ на административни данни 4. Намаляване на разходи 5. Обучение на персонал 6. Интеграция със съществуващи системи Допълнителни бележки: няма	T THREATS — Заплахи 1. GDPR ограничения 2. Зависимост от доставчик 3. AI регулации 4. Грешки при вземане на решения 5. Рискове при чувствителни данни 6. Недостатъчен вътрешен контрол Допълнителни бележки: няма
---	--

C. Разширен SWOT по специфични теми

 Инструкция за попълване: Попълнете текстово наблюденията си за всяка тема. Използвайте конкретни примери.
Цена и лицензиране – 4/5 Наблюдения: (Модел на плащане, достъпност за малки общини, API разходи) гъвкави варианти и добър контрол, ниски API разходи Препоръка за публичния сектор: правилно планиране на броя лицензи, необходими за администрацията/организацията
Сигурност на данните – 4/5 Наблюдения: (Обработка на лични данни, GDPR, съхранение в облака) по-голям потенциал за контрол и локално внедряване Препоръка за публичния сектор: задължително използване на модела на локален сектор при администрации/организации
Езикова поддръжка – 4/5 Наблюдения: (Качество на български език, юридическа терминология) добра многоезична работа Препоръка за публичния сектор: повторна проверка на генерирани резултати от друг модел и верификация от човек
Vendor lock-in – 4/5 Наблюдения: (Зависимост от доставчик, алтернативи, преносимост) по-малък риск при различни варианти за внедряване Препоръка за публичния сектор: при нужда от гъвкавост, локално внедряване и по-ниска зависимост от конкретен доставчик

Сложност на внедряване – 3/5
Наблюдения: <i>(Техническа зрялост, необходима инфраструктура, IT капацитет)</i> изисква техническа експертиза
Препоръка за публичния сектор: при наличие на квалифициран IT екип и капацитет за поддръжка на AI инфраструктура
Регулаторно съответствие – 4/5
Наблюдения <i>(EU AI Act, NIS2/ЗКС, GDPR, национални наредби)</i> по-лесно адаптиране към институционални изисквания
Препоръка за публичния сектор: регулярен IT одит за системите и ИИ за съответствие с актуалната правна рамка
Етика и прозрачност – 3/5
Наблюдения: <i>(Обяснимост, одитируемост, отговорност при грешки)</i> необходими вътрешни процедури
Препоръка за публичния сектор: Да се използва с ясни вътрешни правила за контрол, одит и човешки надзор върху резултатите.
Изкривявания (Bias) – 3/5
Наблюдения: <i>(Пристрастия при различни групи, равнопоставеност)</i> съществува риск от пристрастия
Препоръка за публичния сектор: Необходимо е периодично тестване и мониторинг за недопускане на дискриминационни или неточни резултати.
Халюцинации / грешки – 3/5
Наблюдения: <i>(Честота, detectability, риск при вземане на решения)</i> не може да се разчита без проверка
Препоръка за публичния сектор: валидиране на рисковите резултати от човек
Морал и ценности – 4/5
Наблюдения: <i>(Поведение при чувствителни теми, отказ на вредни задачи)</i> добри ограничения при рискови заявки
Препоръка за публичния сектор: Подходящ за среди с повишени изисквания за безопасност и контрол върху генерираното съдържание.

D. Обобщена оценка на групата

Обобщена оценка по основни показатели		
Обща приложимост <i>средна от Раздел А</i>	Оценка 1–5 ○1 ○2 ○3 x4 ○5	Обосновка / бележки: Подходящ за обработка на документи, автоматизация и интеграция с вътрешни системи
Готовност за внедряване <i>техническа и организационна</i>	Оценка 1–5 ○1 ○2 ○3 x4 ○5	Обосновка / бележки: Предлага добра техническа гъвкавост, но изисква IT ресурс и планиране
Съответствие с регулации <i>AI Act, GDPR, NIS2</i>	Оценка 1–5 ○1 ○2 ○3 x4 ○5	Обосновка / бележки: Предоставя по-добри възможности за контрол и адаптиране към институционални изисквания
Етичен риск <i>обратна скала: 5=нисък риск</i>	Оценка 1–5 ○1 ○2 ○3 x4 ○5	Обосновка / бележки: Рисковете остават управляеми при наличие на човешки контрол
Обща препоръка <i>1=не препоръчваме, 5=силно препоръчваме</i>	Оценка 1–5 ○1 ○2 ○3 ○4 x5	Обосновка / бележки: Подходящ за внедряване в публичния сектор при ясни правила и технически контрол
Мотивирано становище на групата (мин. 5 изречения): <i>Обобщете защо препоръчвате или не препоръчвате модела за публичния сектор.</i> Групата счита, че Mistral AI показва по-висока техническа зрялост за внедряване в публичния сектор. Платформата има добри възможности за интеграция, автоматизация и обработка на големи обеми информация. Предимство е по-голямата гъвкавост при внедряване и потенциалът за контрол върху инфраструктурата. Въпреки това не трябва да се използва без човешки надзор. Препоръчва се внедряване с ясни процедури за проверка и сигурност.		
Предложени условия за ползване: <i>Какви ограничения, контроли или процедури трябва да съпътстват внедряването?</i> Човешка валидация на резултатите; ограничаване на чувствителни данни; логване и проследимост на действията; съответствие с GDPR и вътрешни политики; периодични тестове за качество и пристрастия.		

Risk Canvas

Картографиране на рисковете от AI модел в публичен контекст

v1.0 — Работен шаблон

 **Инструкция за попълване:** Попълнете всички полета за избория от вашата група AI модел. Празните редове и скалите са предназначени за ръчно попълване. Не изтривайте инструкциите в курсив.

Наименование на модела	Пример: Claude Opus 4.7 / GPT-5.4 / Gemini 3.1 Pro MISTRAL (Large 3 / Pevstral 2)
Разработчик / компания	Пример: Anthropic / OpenAI / Google DeepMind MISTRAL AI
Версия за оценка	Точната версия, която групата изследва MISTRAL Large 3
Дата на попълване	ДД.ММ.ГГГГ 14.05.2026 г.
Членове на групата	Имена на участниците Хефест – Манол, Николай, Делян, Диан, Наталия, Силвия, Петър
Ръководител на групата	Наталия

Легенда за скалите

Вероятност (Вер.) 1 = Изключително рядко 2 = Рядко 3 = Умерено вероятно 4 = Вероятно 5 = Почти сигурно	Въздействие (Вл.) 1 = Незначително 2 = Малко 3 = Умерено 4 = Значително 5 = Критично	Ниво на риска Н = Нисък (Вер.×Вл. ≤ 4) С = Среден (5–12) В = Висок (13–25)
--	--	--

Риск Канва — 9-полева матрица

Попълнете всяко поле с идентифицирани рискове, засегнати страни и ключови наблюдения.

 **Инструкция за попълване:** Всяко от 9-те полета представлява различна категория риск. Запишете конкретни рискове — не общи фрази.

1. Технически рискове <i>Халюцинации, грешки, downtime, версионни промени</i> ☐ Халюцинации и неточни отговори при административни задачи ☐ Възможни грешки при обработка на български език ☐ Downtime или недостъпност на API услуги ☐ Версионни промени, влияещи върху поведението на модела	2. Рискове за данните <i>Поверителност, GDPR, изтичане, съхранение</i> ☐ Обработка на лични данни в облачна среда ☐ Риск от изтичане на чувствителна информация ☐ Недостатъчна проследимост на достъпа до данни ☐ Трансфер на данни извън ЕС при външни услуги	3. Правни / регулаторни <i>EU AI Act, NIS2, GDPR, национално право</i> ☐ Несъответствие с EU AI Act ☐ Нарушения на GDPR при неправилна обработка ☐ Липса на достатъчна обяснимост на решенията ☐ Риск при използване на AI за административни решения
4. Етични рискове <i>Bias, дискриминация, непрозрачност, автономия</i> ☐ Потенциален bias спрямо различни групи ☐ Генериране на подвеждащо съдържание ☐ Недостатъчна прозрачност пред гражданите ☐ Прекомерна автоматизация без човешки контрол	5. Оперативни рискове <i>Зависимост, интеграция, обучение, процеси</i> ☐ Недостатъчен IT капацитет за поддръжка ☐ Трудности при интеграция със legacy системи ☐ Зависимост от технически експерти ☐ Недостатъчно обучение на служителите	6. Репутационни рискове <i>Доверие на гражданите, медийно отразяване</i> ☐ Загуба на доверие при публична AI грешка ☐ Негативно медийно отразяване ☐ Съмнения относно сигурността на данните ☐ Ниско обществено доверие към AI решения
7. Стратегически рискове <i>Vendor lock-in, ценови промени, спиране на услугата</i> ☐ Vendor lock-in към конкретна инфраструктура ☐ Промени в ценовия модел на услугата ☐ Зависимост от външен AI доставчик ☐ Бързо остаряване на използвания модел	8. Засегнати страни <i>Граждани, служители, ръководство, партньори</i> ☐ Граждани ☐ Административни служители ☐ IT отдели и ръководство ☐ Външни партньори и доставчици	9. Остатъчни рискове <i>Рискове след митигация, приети рискове</i> ☐ Оставащ риск от халюцинации след проверки ☐ Риск от човешка грешка при валидация ☐ Частична зависимост от AI доставчик ☐ Невъзможност за пълно елиминиране на bias

Рисков регистър

Попълнете по всяка категория конкретните рискове с вероятност, въздействие и мерки.



Инструкция за попълване: Вер. = Вероятност (1–5) · Вл. = Въздействие (1–5) · Ниво: ☐Н (нисък) ☐С (среден) ☐В (висок)

Риск / Събитие	Вероятност	Въздействие	Ниво	Мярка за митигация
Технически рискове				
Халюцинации и фактически грешки при отговори	Вер. ○1 ○2 ○3 X4 ○5	Вл. ○1 ○2 ○3 X4 ○5	Ниво ☐ Н ☐ С XВ	Мярка за митигация: Човешка валидация на резултатите и задължителна проверка при критични административни решения.
Недостъпност / downtime на платформата	Вер. ○1 ○2 X3 ○4 ○5	Вл. ○1 ○2 X3 ○4 ○5	Ниво ☐ Н XС ☐ В	Мярка за митигация: Осигуряване на backup процеси и алтернативни канали при недостъпност на системата.
Внезапна промяна на версия или поведение на модела	Вер. ○1 ○2 ○3 X4 ○5	Вл. ○1 ○2 X3 ○4 ○5	Ниво ☐ Н XС ☐ В	Мярка за митигация: Тестване на нови версии в контролирана среда преди внедряване.
Ограничения на контекстния прозорец при обменни документи	Вер. ○1 X2 ○3 ○4 ○5	Вл. ○1 ○2 X3 ○4 ○5	Ниво ☐ Н XС ☐ В	Мярка за митигация: Разделяне на големи документи и използване на структурирана обработка.
Грешки при мултимодален вход (изображения, аудио)	Вер. ○1 ○2 X3 ○4 ○5	Вл. ○1 X2 ○3 ○4 ○5	Ниво ☐ Н XС ☐ В	Мярка за митигация: Допълнителна проверка от служител при обработка на мултимодално съдържание.
Рискове за данните и поверителността				
Обработка на лични данни на граждани в облачна среда	Вер. ○1 ○2 ○3 X4 ○5	Вл. ○1 ○2 ○3 ○4 X5	Ниво ☐ Н ☐ С XВ	Мярка за митигация: Използване на локално внедряване или private cloud инфраструктура.
Изтичане на класифицирана/служебна информация	Вер. ○1 ○2 X3 ○4 ○5	Вл. ○1 ○2 ○3 ○4 X5	Ниво ☐ Н ☐ С XВ	Мярка за митигация: Ограничен достъп, криптиране и вътрешни политики за сигурност.
Запазване на запитвания от доставчика за обучение	Вер. ○1 ○2 X3 ○4 ○5	Вл. ○1 ○2 ○3 X4 ○5	Ниво ☐ Н XС ☐ В	Мярка за митигация: Изключване на data retention и training опции при използване на API.
Трансфер на данни извън ЕС (EU AI Act / Schrems II)	Вер. ○1 ○2 X3 ○4 ○5	Вл. ○1 ○2 ○3 ○4 X5	Ниво ☐ Н ☐ С XВ	Мярка за митигация: Използване на инфраструктура и сървъри, разположени в ЕС.
Липса на одитна следа при автоматизирани решения	Вер. ○1 ○2 ○3 X4 ○5	Вл. ○1 ○2 ○3 X4 ○5	Ниво ☐ Н ☐ С XВ	Мярка за митигация: Въвеждане на логване и проследимост на действията и заявките.
Правни и регулаторни рискове				

Несъответствие с EU AI Act (Annex III — публичен сектор)	Вер. ○1 ○2 X 3 ○4 ○5	Вл. ○1 ○2 ○3 ○4 X 5	Ниво ☐ Н ☐ С X В	Мярка за митигация: Регулярен правен и технически одит за съответствие с регулациите.
Нарушение на GDPR при обработка на лични данни	Вер. ○1 ○2 X 3 ○4 ○5	Вл. ○1 ○2 ○3 ○4 X 5	Ниво ☐ Н ☐ С X В	Мярка за митигация: Минимизиране на личните данни и въвеждане на GDPR процедури.
Липса на обяснимост при автоматизирани административни решения	Вер. ○1 ○2 ○3 X 4 ○5	Вл. ○1 ○2 ○3 X 4 ○5	Ниво ☐ Н ☐ С X В	Мярка за митигация: Осигуряване на човешки надзор и документиране на AI процесите.
Авторскоправни нарушения в генерирано съдържание	Вер. ○1 X 2 ○3 ○4 ○5	Вл. ○1 ○2 X 3 ○4 ○5	Ниво ☐ Н X С ☐ В	Мярка за митигация: Проверка на генерираното съдържание преди официално използване.
Отговорност при грешни правни или медицински изходи	Вер. ○1 ○2 X 3 ○4 ○5	Вл. ○1 ○2 ○3 ○4 X 5	Ниво ☐ Н ☐ С X В	Мярка за митигация: AI да се използва само като помощен инструмент, а не за крайни решения.
Етични и социални рискове				
Систематичен bias спрямо определени групи граждани	Вер. ○1 ○2 X 3 ○4 ○5	Вл. ○1 ○2 ○3 X 4 ○5	Ниво ☐ Н X С ☐ В	Мярка за митигация: Периодични тестове за bias и мониторинг на резултатите.
Дискриминация при автоматизирано обработване на заявления	Вер. ○1 ○2 X 3 ○4 ○5	Вл. ○1 ○2 ○3 X 4 ○5	Ниво ☐ Н X С ☐ В	Мярка за митигация: Задължителна човешка проверка при чувствителни административни процеси.
Непрозрачност — граждани не знаят, че AI взима решения	Вер. ○1 ○2 ○3 X 4 ○5	Вл. ○1 ○2 ○3 X 4 ○5	Ниво ☐ Н ☐ С X В	Мярка за митигация: Ясно уведомяване при използване на AI системи в администрацията.
Дезинформация и манипулация чрез убедително съдържание	Вер. ○1 ○2 X 3 ○4 ○5	Вл. ○1 ○2 ○3 X 4 ○5	Ниво ☐ Н X С ☐ В	Мярка за митигация: Ограничаване на автоматично публикуване без човешка редакция.
Загуба на критично мислене и зависимост при служителите	Вер. ○1 ○2 ○3 X 4 ○5	Вл. ○1 ○2 X 3 ○4 ○5	Ниво ☐ Н X С ☐ В	Мярка за митигация: Редовно обучение и насърчаване на независима експертна оценка.
Оперативни рискове				
Липса на технически капацитет за внедряване и поддръжка	Вер. ○1 ○2 ○3 X 4 ○5	Вл. ○1 ○2 ○3 X 4 ○5	Ниво ☐ Н ☐ С X В	Мярка за митигация: Изграждане на квалифициран IT екип и техническо обучение.

Vendor lock-in — прекомерна зависимост от доставчик	Вер. ○1 ○2 X3 ○4 ○5	Вл. ○1 ○2 ○3 X4 ○5	Ниво ☐ Н XС ☐ В	Мярка за митигация: Използване на отворени стандарти и възможност за миграция към алтернативни системи.
Недостатъчно обучение на служителите	Вер. ○1 ○2 ○3 X4 ○5	Вл. ○1 ○2 X3 ○4 ○5	Ниво ☐ Н XС ☐ В	Мярка за митигация: Провеждане на регулярни обучения и вътрешни инструкции.
Интеграционни проблеми с наследени (legacy) системи	Вер. ○1 ○2 ○3 X4 ○5	Вл. ○1 ○2 X3 ○4 ○5	Ниво ☐ Н XС ☐ В	Мярка за митигация: Постепенно внедряване и предварително тестване на интеграциите.
Ценови скокове при промяна на лицензионния модел	Вер. ○1 ○2 X3 ○4 ○5	Вл. ○1 ○2 ○3 X4 ○5	Ниво ☐ Н XС ☐ В	Мярка за митигация: Дългосрочни договори и предварително планиране на бюджета.
Репутационни и стратегически рискове				
Публичен скандал при видима AI грешка (медийно отразяване)	Вер. ○1 ○2 X3 ○4 ○5	Вл. ○1 ○2 ○3 ○4 X5	Ниво ☐ Н ☐ С XВ	Мярка за митигация: Процедури за кризисна комуникация и човешки контрол върху публичните резултати.
Загуба на доверие на гражданите в публичните институции	Вер. ○1 ○2 X3 ○4 ○5	Вл. ○1 ○2 ○3 ○4 X5	Ниво ☐ Н ☐ С XВ	Мярка за митигация: Прозрачност при използването на AI и ясни механизми за контрол.
Политическа съпротива срещу използването на AI	Вер. ○1 X2 ○3 ○4 ○5	Вл. ○1 ○2 X3 ○4 ○5	Ниво ☐ Н XС ☐ В	Мярка за митигация: Публична информираност и поетапно внедряване с демонстрирани ползи.
Зависимост от американски/китайски технологичен доставчик	Вер. ○1 X2 ○3 ○4 ○5	Вл. ○1 ○2 ○3 X4 ○5	Ниво ☐ Н XС ☐ В	Мярка за митигация: Предпочитане на европейски AI решения и локална инфраструктура.
Спиране на услугата при геополитически напрежения	Вер. ○1 X2 ○3 ○4 ○5	Вл. ○1 ○2 ○3 X4 ○5	Ниво ☐ Н XС ☐ В	Мярка за митигация: Подготовка на резервни системи и възможност за локална работа.

Обобщена рискова оценка на групата

Топ 3 критични риска за модела:

1. GDPR и изтичане на чувствителни данни
2. Халюцинации при административни решения
3. Недостатъчна прозрачност и контрол върху AI процесите

Обща рискова оценка на модела за публичния сектор:

- ☐ Нисък риск — може да се внедри с минимални ограничения
- ☒ **Среден риск — изисква допълнителни контроли и процедури**
- ☐ Висок риск — изисква задълбочена оценка и специфична регулация
- ☐ **Неприемлив риск — не се препоръчва за публичния сектор**

Мотивирано становище:

Моделът Mistral AI предлага добри възможности за внедряване в публичния сектор благодарение на своята гъвкавост, API интеграции и потенциал за локално разполагане. Въпреки това съществуват значителни рискове, свързани с обработката на чувствителни данни, възможни халюцинации и необходимост от човешки контрол върху резултатите. Рискът може да бъде намален чрез ясни вътрешни процедури, технически мерки за сигурност и постоянен мониторинг. Особено важно е спазването на GDPR и изискванията на EU AI Act. Групата счита, че моделът е подходящ за използване при наличие на адекватен контрол, обучение и одитируемост.

Карта за приложимост

AI модел в публичната администрация — функционален анализ

v1.0 — Работен шаблон

 **Инструкция за попълване:** Попълнете всички полета за избрания от вашата група AI модел. Празните редове и скалите са предназначени за ръчно попълване. Не изтривайте инструкциите в курсив.

Наименование на модела	Пример: Claude Opus 4.7 / GPT-5.4 / Gemini 3.1 Pro MISTRAL (Large 3 / Pevstral 2)
Разработчик / компания	Пример: Anthropic / OpenAI / Google DeepMind MISTRAL AI
Версия за оценка	Точната версия, която групата изследва MISTRAL Large 3
Дата на попълване	ДД.ММ.ГГГГ 14.05.2026 г.
Членове на групата	Имена на участниците Хефест – Манол, Николай, Делян, Диан, Наталия, Силвия, Петър
Ръководител на групата	Наталия

Част 1. Обзорна матрица на приложимостта

Бърза оценка по административни функции. Прил. = Приложимост 1–5. Риск = Н/С/В.

 **Инструкция за попълване:** Попълнете всяка клетка: поставете оценка за приложимост (1=неприложимо, 5=напълно приложимо), ниво на риск (Н/С/В) и кратка бележка.

Административна функция	Обработка на документи	Правен анализ	Граждански услуги	Анализ на данни	Програмиране	Обучение
Обработка на кореспонденция	Прил.: ○1 ○2 ○3 ○4 X 5 Риск: □ Н X С □В	Прил.: ○1 ○2 ○3 X 4 ○5 Риск: □ Н X С □В	Прил.: ○1 ○2 ○3 X 4 ○5 Риск: □ Н X С □В	Прил.: ○1 ○2 X 3 ○4 ○5 Риск: X Н □С □В	Прил.: ○1 ○2 X 3 ○4 ○5 Риск: X Н □С □В	Прил.: ○1 ○2 ○3 X 4 ○5 Риск: X Н □С □В
Изготвяне на доклади и анализи	Прил.: ○1 ○2 ○3 ○4 X 5 Риск: □ Н X С □В	Прил.: ○1 ○2 ○3 X 4 ○5 Риск: □ Н X С □В	Прил.: ○1 ○2 X 3 ○4 ○5 Риск: X Н □С □В	Прил.: ○1 ○2 ○3 ○4 X 5 Риск: □ Н X С □В	Прил.: ○1 ○2 ○3 X 4 ○5 Риск: X Н □С □В	Прил.: ○1 ○2 ○3 X 4 ○5 Риск: X Н □С □В
Тълкуване на нормативни актове	Прил.: ○1 ○2 ○3 X 4 ○5 Риск: □ Н X С □В	Прил.: ○1 ○2 ○3 X 4 ○5 Риск: □ Н □С X В	Прил.: ○1 ○2 X 3 ○4 ○5 Риск: □ Н □С X В	Прил.: ○1 X 2 ○3 ○4 ○5 Риск: X Н □С □В	Прил.: ○1 X 2 ○3 ○4 ○5 Риск: X Н □С □В	Прил.: ○1 ○2 ○3 X 4 ○5 Риск: □ Н X С □В

Изготвяне на правни становища	Прил.: ○1 ○2 ○3 X4 ○5 Риск: ☐ Н ☒ С ☐ В	Прил.: ○1 ○2 ○3 X4 ○5 Риск: ☐ Н ☐ С ☒ В	Прил.: ○1 ○2 X3 ○4 ○5 Риск: ☐ Н ☐ С ☒ В	Прил.: ○1 X2 ○3 ○4 ○5 Риск: ☒ Н ☐ С ☐ В	Прил.: ○1 X2 ○3 ○4 ○5 Риск: ☒ Н ☐ С ☐ В	Прил.: ○1 ○2 ○3 X4 ○5 Риск: ☐ Н ☒ С ☐ В
Отговори на граждански запитвания	Прил.: ○1 ○2 ○3 X4 ○5 Риск: ☐ Н ☒ С ☐ В	Прил.: ○1 ○2 X3 ○4 ○5 Риск: ☐ Н ☒ С ☐ В	Прил.: ○1 ○2 ○3 ○4 X5 Риск: ☐ Н ☒ С ☐ В	Прил.: ○1 X2 ○3 ○4 ○5 Риск: ☒ Н ☐ С ☐ В	Прил.: ○1 X2 ○3 ○4 ○5 Риск: ☒ Н ☐ С ☐ В	Прил.: ○1 ○2 X3 ○4 ○5 Риск: ☒ Н ☐ С ☐ В
Чатбот за електронни услуги	Прил.: ○1 ○2 ○3 X4 ○5 Риск: ☐ Н ☒ С ☐ В	Прил.: ○1 ○2 X3 ○4 ○5 Риск: ☐ Н ☒ С ☐ В	Прил.: ○1 ○2 ○3 ○4 X5 Риск: ☐ Н ☒ С ☐ В	Прил.: ○1 ○2 X3 ○4 ○5 Риск: ☒ Н ☐ С ☐ В	Прил.: ○1 ○2 ○3 ○4 X5 Риск: ☐ Н ☒ С ☐ В	Прил.: ○1 ○2 X3 ○4 ○5 Риск: ☒ Н ☐ С ☐ В
Статистически анализ	Прил.: ○1 ○2 X3 ○4 ○5 Риск: ☒ Н ☐ С ☐ В	Прил.: ○1 X2 ○3 ○4 ○5 Риск: ☒ Н ☐ С ☐ В	Прил.: ○1 X2 ○3 ○4 ○5 Риск: ☒ Н ☐ С ☐ В	Прил.: ○1 ○2 ○3 ○4 X5 Риск: ☐ Н ☒ С ☐ В	Прил.: ○1 ○2 ○3 X4 ○5 Риск: ☒ Н ☐ С ☐ В	Прил.: ○1 ○2 ○3 X4 ○5 Риск: ☒ Н ☐ С ☐ В
Обработка на отчети	Прил.: ○1 ○2 ○3 ○4 X5 Риск: ☐ Н ☒ С ☐ В	Прил.: ○1 ○2 X3 ○4 ○5 Риск: ☐ Н ☒ С ☐ В	Прил.: ○1 ○2 X3 ○4 ○5 Риск: ☒ Н ☐ С ☐ В	Прил.: ○1 ○2 ○3 ○4 X5 Риск: ☐ Н ☒ С ☐ В	Прил.: ○1 ○2 ○3 X4 ○5 Риск: ☒ Н ☐ С ☐ В	Прил.: ○1 ○2 ○3 X4 ○5 Риск: ☒ Н ☐ С ☐ В
Автоматизация на работни процеси	Прил.: ○1 ○2 ○3 ○4 X5 Риск: ☐ Н ☒ С ☐ В	Прил.: ○1 ○2 X3 ○4 ○5 Риск: ☐ Н ☒ С ☐ В	Прил.: ○1 ○2 ○3 X4 ○5 Риск: ☐ Н ☒ С ☐ В	Прил.: ○1 ○2 ○3 X4 ○5 Риск: ☐ Н ☒ С ☐ В	Прил.: ○1 ○2 ○3 ○4 X5 Риск: ☐ Н ☒ С ☐ В	Прил.: ○1 ○2 X3 ○4 ○5 Риск: ☒ Н ☐ С ☐ В
Разработка на вътрешни инструменти	Прил.: ○1 ○2 X3 ○4 ○5 Риск: ☒ Н ☐ С ☐ В	Прил.: ○1 X2 ○3 ○4 ○5 Риск: ☒ Н ☐ С ☐ В	Прил.: ○1 ○2 X3 ○4 ○5 Риск: ☒ Н ☐ С ☐ В	Прил.: ○1 ○2 ○3 X4 ○5 Риск: ☐ Н ☒ С ☐ В	Прил.: ○1 ○2 ○3 ○4 X5 Риск: ☐ Н ☒ С ☐ В	Прил.: ○1 ○2 X3 ○4 ○5 Риск: ☒ Н ☐ С ☐ В
Учебни материали и тестове	Прил.: ○1 ○2 ○3 X4 ○5 Риск: ☒ Н ☐ С ☐ В	Прил.: ○1 ○2 X3 ○4 ○5 Риск: ☐ Н ☒ С ☐ В	Прил.: ○1 ○2 X3 ○4 ○5 Риск: ☒ Н ☐ С ☐ В	Прил.: ○1 ○2 X3 ○4 ○5 Риск: ☒ Н ☐ С ☐ В	Прил.: ○1 ○2 ○3 X4 ○5 Риск: ☒ Н ☐ С ☐ В	Прил.: ○1 ○2 ○3 ○4 X5 Риск: ☒ Н ☐ С ☐ В
Наръчници и процедури	Прил.: ○1 ○2 ○3 ○4 X5 Риск: ☒ Н ☐ С ☐ В	Прил.: ○1 ○2 ○3 X4 ○5 Риск: ☐ Н ☒ С ☐ В	Прил.: ○1 ○2 X3 ○4 ○5 Риск: ☒ Н ☐ С ☐ В	Прил.: ○1 ○2 X3 ○4 ○5 Риск: ☒ Н ☐ С ☐ В	Прил.: ○1 ○2 X3 ○4 ○5 Риск: ☒ Н ☐ С ☐ В	Прил.: ○1 ○2 ○3 ○4 X5 Риск: ☒ Н ☐ С ☐ В
Превод и локализация	Прил.: ○1 ○2 ○3 X4 ○5 Риск: ☐ Н ☐ С ☐ В	Прил.: ○1 ○2 X3 ○4 ○5 Риск: ☐ Н ☐ С ☐ В	Прил.: ○1 ○2 ○3 X4 ○5 Риск: ☐ Н ☐ С ☐ В	Прил.: ○1 X2 ○3 ○4 ○5 Риск: ☐ Н ☐ С ☐ В	Прил.: ○1 X2 ○3 ○4 ○5 Риск: ☐ Н ☐ С ☐ В	Прил.: ○1 ○2 ○3 X4 ○5 Риск: ☐ Н ☐ С ☐ В
Протоколи и резюмета	Прил.: ○1 ○2 ○3 ○4 ○5 Риск: ☒ Н ☐ С ☐ В	Прил.: ○1 ○2 ○3 ○4 ○5 Риск: ☐ Н ☒ С ☐ В	Прил.: ○1 ○2 ○3 ○4 ○5 Риск: ☒ Н ☐ С ☐ В	Прил.: ○1 ○2 ○3 ○4 ○5 Риск: ☒ Н ☐ С ☐ В	Прил.: ○1 ○2 ○3 ○4 ○5 Риск: ☒ Н ☐ С ☐ В	Прил.: ○1 ○2 ○3 ○4 ○5 Риск: ☒ Н ☐ С ☐ В

Част 2. Детайлен анализ по случаи на употреба

За всеки сценарий оценете конкретно как избраният модел се справя.

2.1. Обработка на административни документи
Способността на модела да чете, структурира, резюмира и изготвя административни документи — решения, наредби, докладни записки, протоколи.

Оценка за избрания модел:

1. Може ли моделът да обработва документи на български език с административна терминология?
2. Поддържа ли обработка на PDF/Word файлове с дълъг обем?
3. Запазва ли форматирането и структурата при генериране на изходен документ?
4. Изисква ли специализирано prompt инженерство за административен контекст?

Отговори / наблюдения:

- Моделът обработва добре административни текстове на български език.
- Поддържа работа с големи документи и структуриране на съдържание.
- Запазва сравнително добра структура при генериране на документи.
- За по-сложни административни задачи е необходим по-прецизен prompt.

Препоръка:

☐ Препоръчва се без ограничения

☒ Препоръчва се с ограничения (опишете):

Задължителна човешка проверка преди официално използване.

☐ Не се препоръчва (обосновка):

2.2. Правен анализ и тълкуване

Способността да анализира нормативни актове, да открива противоречия, да прилага правни норми към конкретни случаи.

Оценка за избрания модел:

1. Колко точен е при тълкуване на български законодателни текстове?
2. Може ли да следи промени в нормативната база (обновени закони)?
3. Как се справя с многозначност и правна неяснота?
4. Маркира ли ясно когато излиза извън областта на своята компетентност?

Отговори / наблюдения:

- Добре анализира нормативни текстове, но не трябва да се използва самостоятелно.
- Не гарантира актуалност на всички законодателни промени.
- Възможни са неточности при сложни правни казуси.
- Необходимо е участие на юрист при финални становища.

Препоръка:

☐ Препоръчва се без ограничения

☒ Препоръчва се с ограничения (опишете):

Само като помощен инструмент под човешки надзор.

☐ Не се препоръчва (обосновка):

2.3. Взаимодействие с граждани (чатбот / асистент)

Способността да комуникира с граждани — отговаря на въпроси, помага при подаване на заявления, обяснява административни процедури.

Оценка за избрания модел:

1. Разбира ли естествен разговорен български (включително диалекти и грешки)?
2. Дава ли ясни и разбираеми отговори без административен жаргон?
3. Как се справя с чувствителни теми (социални помощи, здраве, правосъдие)?

4. Може ли да се интегрира с портали за електронни услуги?

Отговори / наблюдения:

- Разбира разговорен български език.
- Генерира ясни и разбираеми отговори.
- При чувствителни теми е необходим контрол.
- Подходящ за интеграция в чатбот системи.

Препоръка:

☐ Препоръчва се без ограничения

☒ Препоръчва се с ограничения (опишете):

Без автоматично вземане на решения.

☐ Не се препоръчва (обосновка):

2.4. Анализ на данни и отчетност

Способността да обработва структурирани данни, да генерира отчети и да идентифицира тенденции.

Оценка за избрания модел:

1. Може ли да обработва Excel/CSV данни и да генерира анализи?
2. Поддържа ли визуализация на данни (графики, диаграми)?
3. Колко надежден е при числови изчисления и статистика?
4. Може ли да се свърже с вътрешни бази данни чрез API?

Отговори / наблюдения:

- Поддържа обработка на структурирани данни.
- Подходящ за анализи и отчети.
- Не е напълно надежден при сложни статистически изчисления.
- Поддържа API интеграции.

Препоръка:

☐ Препоръчва се без ограничения

☒ Препоръчва се с ограничения (опишете):

Проверка на числовите резултати от служител.

☐ Не се препоръчва (обосновка):

2.5. Вътрешно обучение на служители

Способността да генерира учебни материали, тестове, наръчници и да поддържа интерактивно обучение.

Оценка за избрания модел:

1. Може ли да генерира тематични курсове и тестове за служители?
2. Адаптира ли съдържанието към различни нива на компетентност?
3. Поддържа ли формати за дистанционно обучение (SCORM и др.)?
4. Как се справя с чувствителни теми — антикорупция, етика, дискриминация?

Отговори / наблюдения:

- Подходящ за генериране на учебни материали.
- Може да адаптира съдържание към различни нива.
- Ограничена поддръжка на специализирани e-learning формати.
- Добре обработва етични и административни теми. **Препоръка:**

☒ Препоръчва се без ограничения

☐ Препоръчва се с ограничения (опишете):

☐ Не се препоръчва (обосновка):

2.6. Програмиране и автоматизация

Способността да пише код, да автоматизира процеси и да поддържа вътрешни системи.

Оценка за избрания модел:

- 1. Поддържа ли езиките и технологиите, използвани в административните системи?
- 2. Може ли да чете и документира съществуващ (legacy) код?
- 3. Колко надеждни са генерираните скриптове без допълнителна проверка?
- 4. Поддържа ли agentic режим за автономно изпълнение на задачи?

Отговори / наблюдения:

- Добра поддръжка на популярни програмни езици.
- Може да анализира и документира legacy код.
- Генерираният код изисква проверка.
- Поддържа автоматизация и agentic процеси.

Препоръка:

☐ Препоръчва се без ограничения

☒ Препоръчва се с ограничения (опишете):

Задължителен code review преди продукционна среда.

☐ Не се препоръчва (обосновка):

Част 3. Оценка на интеграционния потенциал

 **Инструкция за попълване:** Оценете техническите възможности за интеграция с административната инфраструктура.

Интеграционен аспект	Описание	Поддръжка	Бележки
API свързване	REST API, GraphQL, webhooks — интеграция с вътрешни системи	☒ Да ☐ Не ☐ Частично	Добра API интеграция
Single Sign-On	Съвместимост с Active Directory / ADFS / Keycloak	☐ Да ☐ Не ☒ Частично	Зависи от инфраструктурата
On-premise deployment	Възможност за локален хостинг без облачна зависимост	☒ Да ☐ Не ☐ Частично	Силна страна на модела
Данни в ЕС	Сървъри на европейска територия — GDPR съответствие	☒ Да ☐ Не ☐ Частично	Зависи от внедряването
Одитна следа	Логирание на всички взаимодействия за отчетност	☐ Да ☐ Не ☒ Частично	Зависи от внедряването
Ролеви достъп	RBAC — различни нива на достъп за различни служители	☐ Да ☐ Не ☒ Частично	Изисква вътрешна конфигурация
Многоезичност	Поддръжка на български и официални ЕС езици	☒ Да ☐ Не ☐ Частично	Добра многоезична поддръжка
Мобилен достъп	Работа от мобилни устройства (field work)	☐ Да ☐ Не ☒ Частично	Необходима е техническа адаптация

Офлайн режим	Функционалност без постоянна интернет връзка	☒ Да ☐ Не ☐ Частично	Поддържа адаптация
SLA и поддръжка	Гарантирано ниво на обслужване за публичния сектор	☒ Да ☐ Не ☐ Частично	Подходящ за институции

Обобщена карта за приложимост

Топ 3 приложения с висок потенциал:

- 1. Автоматизация на административни документи и кореспонденция
- 2. Чатботи и електронни услуги за граждани
- 3. Разработка на вътрешни AI инструменти и автоматизация на процеси

Функции, при които моделът НЕ е подходящ:

- 1. Самостоятелно вземане на административни или правни решения
- 2. Обработка на класифицирана информация без локална инфраструктура
- 3. Автоматизирани решения без човешки контрол и верификация

Препоръчани условия за пилотно внедряване:

- ☐ Внедряване първоначално в нискорискови административни процеси
- ☐ Задължителен човешки контрол върху всички критични резултати
- ☐ Използване на GDPR-съвместима и по възможност локална инфраструктура
- ☐ Провеждане на обучение за служителите преди използване на системата
- ☐ Въвеждане на логове, мониторинг и периодичен одит на AI резултатите
- ☐ Ограничаване на достъпа до чувствителни данни и вътрешни системи
- ☐ Постепенно разширяване след успешен пилотен период и оценка на риска

Методологични указания

Етично и прозрачно ползване на AI модели в публичната администрация

v1.0 — Работен шаблон

Инструкция за попълване: Попълнете всички полета за избория от вашата група AI модел. Празните редове и скалите са предназначени за ръчно попълване. Не изтривайте инструкциите в курсив.

аименование на модела	Пример: Claude Opus 4.7 / GPT-5.4 / Gemini 3.1 Pro MISTRAL (Large 3 / Pevstral 2)
Разработчик / компания	Пример: Anthropic / OpenAI / Google DeepMind MISTRAL AI
Версия за оценка	Точната версия, която групата изследва MISTRAL Large 3
Дата на попълване	ДД.ММ.ГГГГ 14.05.2026 г.
Членове на групата	Имена на участниците Хефест – Манол, Николай, Делян, Диан, Наталия, Силвия, Петър
Ръководител на групата	Наталия

Предговор и обхват

Инструкция за попълване: Попълнете идентификационните данни на вашата организация и дефинирайте обхвата на документа.

Наименование на организацията	Пълно официално наименование Областна администрация и съпътстващи общински структури
Административно ниво	Централна / областна / общинска администрация Регионално / Областно / Общинско ниво
AI модел(и), обхванати от документа	Посочете конкретните модели/версии MISTRAL AI
Версия на документа	v1.0 — Официално утвърдени указания
Дата на приемане	ДД.ММ.ГГГГ 14.05.2026 г.
Следваща планирана ревизия	ДД.ММ.ГГГГ 14.05.2027 г.
Одобрен от	Длъжност и имена Ръководител администрацията
Отговорен за прилагането	Длъжност, звено и контакти IT отдел и AI отговорник

Чл. 1. Предмет и обхват

Опишете за какво се отнася документът — кои AI системи, кои структурни звена и кои видове задачи обхваща.

Настоящите указания определят правилата за използване на AI системи в публичната администрация. Документът обхваща използването на Mistral AI за административни документи, анализи, обучение и комуникация с граждани. Указанията се прилагат за всички служители, използващи AI инструменти в рамките на служебните си задължения.

Чл. 2. Правна и регулаторна рамка

Посочете нормативните актове, на базата на които се издават указанията (EU AI Act, GDPR, ЗЕУ, вътрешни наредби).

EU AI Act, GDPR, NIS2, Закон за електронното управление, вътрешни правила за информационна сигурност и защита на личните данни.

Раздел I. Принципи за етично и прозрачно ползване

Определете основните принципи, на които се базира политиката. Попълнете конкретните мерки за всеки.

Основни принципи	
1 Законност	AI системите се използват само в рамките на действащото законодателство. Всяко решение, подпомогнато от AI, спазва приложимите правни норми. Конкретни мерки в нашата организация: Всички AI процеси се одобряват от ръководството и правното звено.
2 Прозрачност	Гражданите и служителите имат право да знаят кога AI участва в процесите. Резултатите от AI трябва да са обяснени разбираемо. Конкретни мерки в нашата организация: Гражданите се уведомяват при използване на AI.
3 Отчетност	Ясна верига на отговорност. Служителите носят отговорност за решенията, взети с AI помощ — AI не замества човешката отговорност. Конкретни мерки в нашата организация: Крайното решение винаги се потвърждава от служител.
4 Справедливост	AI инструментите не дискриминират. Резултатите се проверяват редовно за пристрастия (bias) спрямо различни групи граждани. Конкретни мерки в нашата организация: Извършват се периодични проверки за bias.
5 Поверителност	Личните данни на гражданите се обработват само при наличие на правно основание и при спазване на GDPR. Конкретни мерки в нашата организация: Личните данни се обработват само при GDPR съответствие.
6 Надеждност	AI системите се използват само за задачи, при които точността е потвърдена. Резултатите подлежат на верификация. Конкретни мерки в нашата организация: Резултатите се проверяват преди използване.
7 Пропорционалност	Нивото на автоматизация съответства на риска от грешки. При по-висок риск — по-задължителен човешки контрол. Конкретни мерки в нашата организация: При високорискови процеси има задължителен човешки контрол.

8

Обяснимост

При поискване от гражданин или орган, организацията може да обясни как е бил използван AI при дадено решение.

Конкретни мерки в нашата организация:

AI процесите се документират и архивират.

Раздел II. Разрешени и забранени употреби

Определете за всяка категория задачи дали е разрешена, допустима с условия или забранена.

 **Инструкция за попълване:** За всяка задача: поставете отметка в съответната колона и опишете условията или ограниченията.

Категория задача	Статус	Условия / Ограничения / Процедура
Административни документи		
Изготвяне на проекти за решения	☐ Разрешено ☒ С условия ☐ Забранено	Условия / ограничения / процедура: Изходът от AI подлежи на задължителна проверка и одобрение от компетентен служител преди официално използване.
Резюмиране на входяща поща	☒ Разрешено ☐ С условия ☐ Забранено	Условия / ограничения / процедура: Да не се обработват класифицирани данни и чувствителна информация без допълнителен контрол.
Превод на документи	☒ Разрешено ☐ С условия ☐ Забранено	Условия / ограничения / процедура: Официалните преводи подлежат на финална проверка от служител или зачет при необходимост.
Генериране на официални писма	☐ Разрешено ☒ С условия ☐ Забранено	Условия / ограничения / процедура: Всички писма трябва да бъдат прегледани и утвърдени от отговорен служител преди изпращане
Правни и нормативни въпроси		
Анализ на нормативни текстове	☐ Разрешено ☒ С условия ☐ Забранено	Условия / ограничения / процедура: AI се използва само като помощен инструмент без самостоятелно правно тълкуване.
Проверка за вътрешни противоречия	☒ Разрешено ☐ С условия ☐ Забранено	Условия / ограничения / процедура: Резултатите се използват само за подпомагане на експертен анализ.
Изготвяне на правни становища	☐ Разрешено ☒ С условия ☐ Забранено	Условия / ограничения / процедура: Финалното становище се подписва и потвърждава от правоспособен юрист.
Тълкуване на разпоредби към граждани	☐ Разрешено ☒ С условия ☐ Забранено	Условия / ограничения / процедура: Информацията се предоставя само с уточнение, че не представлява официална правна консултация.
Вземане на решения с ефект върху граждани		
Преценка за социални помощи	☐ Разрешено ☐ С условия ☒ Забранено	Условия / ограничения / процедура: AI не може самостоятелно да взема решения, свързани със социални права и помощи.
Оценка на проекти и заявления	☐ Разрешено ☒ С условия ☐ Забранено	Условия / ограничения / процедура: AI може да подпомага обработката, но крайното решение се взема от комисия или служител.
Наказателни и принудителни мерки	☐ Разрешено ☐ С условия ☒ Забранено	Условия / ограничения / процедура: Не се допуска използване на AI за автоматизирано налагане на санкции или ограничения.

Автоматизирани решения без човешки контрол	☐ Разрешено ☐ С условия ☐ Забранено	Условия / ограничения / процедура: Всички критични административни решения изискват човешка верификация.
Комуникация с граждани		
Чатбот за информационни въпроси	☒ Разрешено ☐ С условия ☐ Забранено	Условия / ограничения / процедура: Чатботът трябва ясно да обозначава, че използва AI технология.
Проекти за отговори на жалби	☐ Разрешено ☒ С условия ☐ Забранено	Условия / ограничения / процедура: Отговорите се преглеждат от компетентен служител преди изпращане.
Публикации в социални медии	☐ Разрешено ☒ С условия ☐ Забранено	Условия / ограничения / процедура: Публикациите се одобряват от отговорен служител или PR отдел.
Кризисна комуникация	☐ Разрешено ☒ С условия ☐ Забранено	Условия / ограничения / процедура: AI не може самостоятелно да публикува кризисна информация без одобрение от ръководството.
Вътрешно-административни функции		
Обучение на служители	☒ Разрешено ☐ С условия ☐ Забранено	Условия / ограничения / процедура: Материалите подлежат на периодичен преглед за актуалност и коректност.
Анализ на данни и отчети	☐ Разрешено ☒ С условия ☐ Забранено	Условия / ограничения / процедура: Числовите анализи и статистическите резултати се проверяват от служител.
Автоматизация на повтарящи се задачи	☒ Разрешено ☐ С условия ☐ Забранено	Условия / ограничения / процедура: Автоматизацията се допуска само след предварително тестване и контрол.
Обработка на класифицирана информация	☐ Разрешено ☐ С условия ☒ Забранено	Условия / ограничения / процедура: Не се допуска използване на публични AI модели за класифицирани или строго поверителни данни.

Раздел III. Процедури за одобрение и верификация

Чл. 10. Задължителна верификация
Опишете кои задачи изискват задължителна проверка от служител преди използване на AI изхода. Всички правни, административни и граждански решения подлежат на човешка проверка.
Чл. 11. Процедура за одобрение на нови употреби
Как служителят може да заяви одобрение за нов вид употреба на AI, непредвиден в настоящите указания? Нови AI употреби се одобряват от IT и правното звено.
Чл. 12. Задължително разкриване към граждани
В кои случаи гражданинът трябва да бъде информиран, че AI е участвал в обработката на неговото искане? Гражданите се уведомяват при AI участие в административен процес.

Раздел IV. Мониторинг, одит и отчетност

Минимални изисквания за одитна следа (отбележете приложимите):

- ☒ Всяко взаимодействие с AI системата се записва с дата, час и потребител
- ☒ Пазят се логове на AI изходите, използвани в официални документи
- ☒ При автоматизирани решения се съхранява AI входа и изхода
- ☒ Провежда се периодичен преглед на качеството на AI изходите (мин. 1х/тримесечие)
- ☒ Инцидентите, свързани с AI грешки, се докладват на определено звено
- ☒ Годишен отчет за употребата на AI се изготвя и публикува

Чл. 15. KPI за мониторинг на AI употребата

Определете конкретни измерими показатели: брой взаимодействия, % задачи с верификация, % открити грешки и др.

- ☐ Брой AI взаимодействия
- ☐ % проверени AI резултати
- ☐ % открити грешки
- ☐ Време за обработка на документи
- ☐ Брой докладвани инциденти

Раздел V. Отговорности и управление

Роля	Отговорност	Конкретно лице / звено
Ръководител на организацията	Стратегическо одобрение и крайна отговорност за AI политиката Издаване на вътрешни заповеди; стратегическо разпределение на ИТ бюджета за хардуер (сървъри).	Конкретно лице / звено: Областен управител / Кмет на общината
AI отговорник (AI Officer)	Координация, обучение, мониторинг, докладване на инциденти Мониторинг на спазването на методиката; координация между ИТ отдела и правното звено.	Конкретно лице / звено: Директор на дирекция "Административно обслужване"
Служители, ползващи AI	Спазване на указанията, верификация на изходите, докладване на проблеми Ръчна проверка на всеки абзац; недопускане на безкритично копиране на текстове; спазване на правилата за сигурност.	Конкретно лице / звено: Всички служители в деловодството, правния отдел и ЧР
ИТ отдел / системен администратор	Технически достъп, интеграция, сигурност, логове Поддръжка на хардуера, актуализация на теглата на моделите (open-weights), киберсигурност на мрежата, сигурност на логовете.	Конкретно лице / звено: Отдел "Информационни технологии и съобщения"
Правно звено	Преглед на нови употреби, съответствие с нормативната рамка Съгласуване на проектите на решения; преглед на договорите за хардуерна поддръжка спрямо българското законодателство.	Конкретно лице / звено: Отдел "Правно осигуряване"
ДЛЗД (длъжностно лице по защита на данни)	GDPR съответствие при обработка на лични данни Гарантиране, че локалният трансфер на данни не нарушава правата на гражданите по GDPR; регулярни вътрешни проверки.	Конкретно лице / звено: Главен експерт - ДЛЗД за Областта

Вътрешен одит	Периодичен независим преглед на AI употребата Проверка за икономическа целесъобразност на внедрената технология и спазване на вътрешните разпоредби.	Конкретно лице / звено: Звено "Вътрешен одит"
---------------	---	--

Раздел VI. Обучение и изграждане на капацитет

Чл. 20. Задължително обучение			
Посочете кои служители трябва да преминат обучение, какъв е минималният обем и как се документира. Всички служители, използващи AI, преминават ежегодно обучение.			
Модул на обучение	Целева група	Форма и обем	Периодичност
Основи на AI и рискове	Всички служители	Онлайн / 4 ч.	Годишно
Работа с конкретния модел	IT и администрация	Практическо / 6 ч.	Годишно
Етика и нормативна рамка	Ръководство	Семинар / 4 ч.	Годишно
Разпознаване на AI грешки	Всички служители	Практическо / 3 ч.	На 6 месеца
Защита на личните данни	Всички служители	Онлайн / 2 ч.	Годишно

Раздел VII. Актуализация и валидност

Чл. 24. Условия за задължителна ревизия			
Опишете кога документът трябва задължително да се преразгледа: нова версия на модела, регулаторна промяна, инцидент, изтичане на срок. <u>Документът се актуализира при:</u> нова версия на AI модела; промяна в регулаторната рамка; сериозен AI инцидент; установени проблеми при одит; минимум веднъж годишно.			
Чл. 25. История на промените			
Документирайте всички версии на документа с дата, автор и описание на промяната. Настоящият документ подлежи на актуализация при промени в нормативната рамка, внедряване на нови AI функционалности, установени инциденти или след проведен вътрешен одит. Всички промени се документират и одобряват от ръководството на организацията.			
Версия	Дата	Описание на промяната	Одобрил
v1.0	14.05.2026 г.	Първоначално създаване и приемане на методологичните указания за използване на Mistral AI	Ръководител на организацията

SWOT Анализ

Оценка на фундаментален AI модел за публичната администрация

v1.0 — Работен шаблон

Инструкция за попълване: Попълнете всички полета за избория от вашата група AI модел. Празните редове и скалите са предназначени за ръчно попълване. Не изтривайте инструкциите в курсив.

Наименование на модела	Пример: Claude Opus 4.7 / GPT-5.4 / Gemini 3.1 Pro xAI Grok (4.20 / 4)
Разработчик / компания	Пример: Anthropic / OpenAI / Google DeepMind xAI
Версия за оценка	Точната версия, която групата изследва Grok 4.3 (Beta)
Дата на попълване	ДД.ММ.ГГГГ 14.05.2026 г.
Членове на групата	Имена на участниците Хефест – Манол, Николай, Делян, Диан, Наталия, Силвия, Петър
Ръководител на групата	Наталия

A. Обща приложимост и функционален обхват

Оценете доколко моделът покрива реални нужди в публичния сектор.

Инструкция за попълване: За всеки критерий: поставете кръгче на избраната оценка (1=много слабо, 5=отлично) и запишете кратка обосновка.

Критерий		
A1. Приложимост в публичния сектор		
Обработка на документи доклади, решения, наредби	Оценка 1–5 ○1 ○2 x 3 x4 ○5	Обосновка / бележки: Добро резюмиране и редактиране
Правен анализ тълкуване на нормативни актове	Оценка 1–5 ○1 ○2 x 3 ○4 ○5	Обосновка / бележки: Подпомага анализ, но не е надежден за нормативни решения
Граждански услуги чатбот, Q&A, помощ при заявления	Оценка 1–5 ○1 ○2 ○3 x 4 ○5	Обосновка / бележки: Подходящ за чатбот сценарии

Вътрешна комуникация <i>резюмета, протоколи, кореспонденция</i>	Оценка 1–5 ○1 ○2 ○3 ○4 x5	Обосновка / бележки: Добър при протоколи и кореспонденция
Анализ на данни <i>отчети, статистика, мониторинг</i>	Оценка 1–5 ○1 ○2 ○3 x4 ○5	Обосновка / бележки: Силен за текстови анализ, по-слаб за сложни аналитични процеси
Превод <i>многоезична комуникация, ЕС документи</i>	Оценка 1–5 ○1 ○2 ○3 x4 ○5	Обосновка / бележки: Добър многоезичен апацитет
Програмиране / автоматизация <i>вътрешни системи, скриптове</i>	Оценка 1–5 ○1 ○2 ○3 x4 ○5	Обосновка / бележки: Добро генериране на код
Обучение на служители <i>учебни материали, тестове</i>	Оценка 1–5 ○1 ○2 ○3 ○4 x5	Обосновка / бележки: Подходящ за учебни атериали
A2. Функции на платформата		
Текстово генериране <i>качество и прецизност</i>	Оценка 1–5 ○1 ○2 ○3 ○4 x5	Обосновка / бележки: Генерира естествен и добре структуриран текст
Reasoning / разсъждение <i>логически вериги, многостъпкови задачи</i>	Оценка 1–5 ○1 ○2 ○3 x4 ○5	Обосновка / бележки: Добър при многостъпкови задачи, но не винаги последователен
Code writing <i>генериране и анализ на код</i>	Оценка 1–5 ○1 ○2 ○3 x4 ○5	Обосновка / бележки: Генерира и обяснява код ефективно
Анализ на изображения <i>vision, диаграми, скрийншоти</i>	Оценка 1–5 ○1 ○2 ○3 x4 ○5	Обосновка / бележки: Разпознава изображения и визуално съдържание
Аудио / глас <i>говорим вход/изход</i>	Оценка 1–5 ○1 ○2 x3 ○4 ○5	Обосновка / бележки: Наличност според платформата и версията
Видео разбиране <i>анализ на видеосъдържание</i>	Оценка 1–5 ○1 ○2 x3 ○4 ○5	Обосновка / бележки: Ограничени възможности
Дълги документи <i>обработка на обемни файлове</i>	Оценка 1–5 ○1 ○2 x3 ○4 ○5	Обосновка / бележки: Обработка големи текстове сравнително добре
Agentic режим <i>автономни многостъпкови задачи</i>	Оценка 1–5 ○1 ○2 ○3 x4 ○5	Обосновка / бележки: Поддържа многостъпкови задачи

API / системна интеграция <i>свързване с вътрешни системи</i>	Оценка 1–5 ○1 ○2 ○3 x4 ○5	Обосновка / бележки: Поддържа интеграции
Plugin екосистема <i>налични разширения и конектори</i>	Оценка 1–5 ○1 ○2 x3 ○4 ○5	Обосновка / бележки: По-малка екосистема
Desktop приложение <i>нативна desktop среда</i>	Оценка 1–5 ○1 ○2 x3 ○4 ○5	Обосновка / бележки: Ограничена самостоятелна среда
Mobile приложение <i>iOS / Android достъп</i>	Оценка 1–5 ○1 ○2 ○3 x4 ○5	Обосновка / бележки: Добър мобилен достъп
Cowork / колаборация <i>екипна работа в реално време</i>	Оценка 1–5 ○1 ○2 x3 ○4 ○5	Обосновка / бележки: Ограничени екипни функции
Търсене в интернет <i>актуална информация от мрежата</i>	Оценка 1–5 ○1 ○2 ○3 ○4 x5	Обосновка / бележки: Много силна страна
Fine-tuning <i>настройка за специфична среда</i>	Оценка 1–5 ○1 x2 ○3 ○4 ○5	Обосновка / бележки: Ограничени възможности за персонализация

B. SWOT Матрица

Обобщете основните силни/слаби страни, възможности и заплахи на модела в публичен контекст.

 **Инструкция за попълване:** Запишете поне 4 наблюдения в клетка. Базирайте се на попълнените оценки от Раздел А.

S STRENGTHS — Силни страни 1. Бързи отговори 2. Добра разговорна комуникация 3. Достъп до актуална информация 4. Силен при резюмиране 5. Пподходящ за чат ботове 6. Добро адаптиране към контекст Допълнителни бележки: няма	W WEAKNESSES — Слаби страни 1. Риск от халюцинации 2. Ограничена прозрачност на модела 3. Правния анализ изисква проверка 4. Не е оптимизиран за административни процеси 5. Необходим е човешки контрол 6. Зависимост от екосистемата Допълнителни бележки: няма
--	--

O OPPORTUNITIES — Възможности - автоматизация на административни услуги - виртуални асистенти - обработка на сигнали от граждани - обучение на служители - подоряване на обслужването - намаляване времето за обработка Допълнителни бележки: няма	T THREATS — Заплахи - GDPR рискове - Vendor lock in - Регулаторни ограничения - грешни отговори при критични решения - кибер рискове - прекомерно доверие в ИИ Допълнителни бележки: няма
---	--

C. Разширен SWOT по специфични теми

 Инструкция за попълване: Попълнете текстово наблюденията си за всяка тема. Използвайте конкретни примери.
Цена и лицензиране – 3/5 Наблюдения: (Модел на плащане, достъпност за малки общини, API разходи) абонаментен модел и потенциално нарастващи разходи Препоръка за публичния сектор: не е подходящ за малки общини поради потенциал за високи разходи
Сигурност на данните – 3/5 Наблюдения: (Обработка на лични данни, GDPR, съхранение в облака) необходим е строг контрол при чувствителни данни Препоръка за публичния сектор: само при наличие на добра защита, строг контрол и надеждна кибер сигурност
Езикова поддръжка – 4/5 Наблюдения: (Качество на български език, юридическа терминология) добър български език- възможни терминологични неточности Препоръка за публичния сектор: може да се приложи при лека редакция от служители
Vendor lock-in – 2/5 Наблюдения: (Зависимост от доставчик, алтернативи, преносимост) Риск от зависимост към конкретна еко сиситема Препоръка за публичния сектор: Да се прилага с внимание има риск от зависимост от затворена екосистема и ограничени възможности за преносимост.

Сложност на внедряване – 4/5
Наблюдения: <i>(Техническа зрялост, необходима инфраструктура, IT капацитет)</i> Лесен за потребители, средна сложност за интеграции
Препоръка за публичния сектор: ако се постигне лесно управление на инструмента -да се приложи
Регулаторно съответствие – 3/5
Наблюдения <i>(EU AI Act, NIS2/ЗКС, GDPR, национални наредби)</i> Необходима е допълнителна оценка за ползване в публичния сектор
Препоръка за публичния сектор: нормативни рамки и допълнително обучение ИИ
Етика и прозрачност – 3/5
Наблюдения: <i>(Обяснимост, одитируемост, отговорност при грешки)</i> Ограничена обяснимост
Препоръка за публичния сектор: проверка от хора-експерти
Изкривявания (Bias) – 3/5
Наблюдения: <i>(Пристрастия при различни групи, равнопоставеност)</i> риск от отклонения
Препоръка за публичния сектор: верификация от хора
Халюцинации / грешки – 3/5
Наблюдения: <i>(Честота, detectability, риск при вземане на решения)</i> необходимост от човешка проверка
Препоръка за публичния сектор: прегледи, редакция от експерти
Морал и ценности – 4/5
Наблюдения: <i>(Поведение при чувствителни теми, отказ на вредни задачи)</i> има защитни механизми, но не са достатъчно самостоятелни
Препоръка за публичния сектор: допълнителен човешки контрол при работа с чувствителна или обществено значима информация.

D. Обобщена оценка на групата

Обобщена оценка по основни показатели		
Обща приложимост <i>средна от Раздел А</i>	Оценка 1–5 ○1 ○2 ○3 x4 ○5	Обосновка / бележки: Моделът показва добра приложимост за обработка на документи, вътр. Комуникация и граждански услуги. Подходящ за чат ботове и административна помощ. Необходим е човешки контрол.
Готовност за внедряване <i>техническа и организационна</i>	Оценка 1–5 ○1 ○2 ○3 x4 ○5	Обосновка / бележки: Платформата е лесна за използване и внедряване. Възможен допълнителен технически ресурс
Съответствие с регулации <i>AI Act, GDPR, NIS2</i>	Оценка 1–5 ○1 ○2 x3 ○4 ○5	Обосновка / бележки: Необходимо е внимателно съобразяване с GDPR с ИИ регулации и вътр. Политика за работа с чувствителни данни
Етичен риск <i>обратна скала: 5=нисък риск</i>	Оценка 1–5 ○1 ○2 x3 ○4 ○5	Обосновка / бележки: Съществува риск от неточни отговори и ограничена прозрачност при сложни казуси
Обща препоръка <i>1=не препоръчваме, 5=силно препоръчваме</i>	Оценка 1–5 ○1 ○2 ○3 x4 ○5	Обосновка / бележки: Подходящ за подпомагане на административни функции, но не и за самостоятелно вземане на решения
Мотивирано становище на групата (мин. 5 изречения): <i>Обобщете защо препоръчвате или не препоръчвате модела за публичния сектор.</i> Групата счита че модела предлага добри възможности за автоматизиране на административни дейности и подобряване на комуникацията. Представя се добре при текстова обработка, резюмиране и обслужване на потребители. Съществува риск от неточности. Препоръчва се ограничено и контролирано внедряване.		
Предложени условия за ползване: <i>Какви ограничения, контроли или процедури трябва да съпътстват внедряването?</i> - Човешка проверка на всички официални документи; - забрана за автоматично вземане на решения; - стриктно спазване на GDPR; - периодичен одит на резултатите		

Risk Canvas

Картографиране на рисковете от AI модел в публичен контекст

v1.0 — Работен шаблон

 **Инструкция за попълване:** Попълнете всички полета за избрания от вашата група AI модел. Празните редове и скалите са предназначени за ръчно попълване. Не изтривайте инструкциите в курсив.

Наименование на модела	Пример: Claude Opus 4.7 / GPT-5.4 / Gemini 3.1 Pro xAI Grok (4.20 / 4)
Разработчик / компания	Пример: Anthropic / OpenAI / Google DeepMind xAI
Версия за оценка	Точната версия, която групата изследва Grok 4.3 (Beta)
Дата на попълване	ДД.ММ.ГГГГ 14.05.2026 г.
Членове на групата	Имена на участниците Хефест – Манол, Николай, Делян, Диан, Наталия, Силвия, Петър
Ръководител на групата	Наталия

Легенда за скалите

Вероятност (Вер.) 1 = Изключително рядко 2 = Рядко 3 = Умерено вероятно 4 = Вероятно 5 = Почти сигурно	Въздействие (Вл.) 1 = Незначително 2 = Малко 3 = Умерено 4 = Значително 5 = Критично	Ниво на риска Н = Нисък (Вер.×Вл. ≤ 4) С = Среден (5–12) В = Висок (13–25)
--	--	--

Риск Канва — 9-полева матрица

Попълнете всяко поле с идентифицирани рискове, засегнати страни и ключови наблюдения.

 **Инструкция за попълване:** Всяко от 9-те полета представлява различна категория риск. Запишете конкретни рискове — не общи фрази.

1. Технически рискове <i>Халюцинации, грешки, downtime, версионни промени</i> • Халюцинации при правни, административни и технически казуси. • Силна зависимост от актуалност на web/X търсене и от системни настройки. • Версионни промени и поведение на модела без контрол от администрацията. • Мултимодални грешки при разчитане на изображения/документи.	2. Рискове за данните <i>Поверителност, GDPR, изтичане, съхранение</i> • Въвеждане на лични, служебни или чувствителни данни в облачна услуга. • Възможност разговори/входове да се използват за обучение или персонализация според настройките. • Трансфер и обработка извън ЕС; нужда от DPIA и договорни гаранции. • Липса на пълна локална одитна следа, ако не се интегрира отделно.	3. Правни / регулаторни <i>EU AI Act, NIS2, GDPR, национално право</i> • GDPR: правно основание, минимизация, права на субектите, трансфери. • EU AI Act: възможна високорискова употреба при публични услуги/решения. • Нужда от човешки контрол при административни актове и индивидуални права. • Авторскоправен риск при генерирани текстове/изображения/код.
4. Етични рискове <i>Bias, дискриминация, непрозрачност, автономия</i> • Bias спрямо групи граждани, региони, езици или политически теми. • Непрозрачност за гражданина, ако AI участва в обработката. • Убедително звучаща дезинформация, особено при актуални събития. • Прехвърляне на отговорност от служител към модел.	5. Оперативни рискове <i>Зависимост, интеграция, обучение, процеси</i> • Недостатъчно обучение за безопасни промптове и проверка на отговори. • Интеграционни трудности с документооборот, регистри и legacy системи. • Нужда от резервен процес при недостъпност на услугата. • Неясни роли: кой проверява, одобрява и носи отговорност.	6. Репутационни рискове <i>Доверие на гражданите, медийно отразяване</i> • Публичен скандал при грешен отговор, дискриминационен текст или теч на данни. • Съмнение, че институцията “автоматизира” решения без човешки контрол. • Политически и медиен натиск заради връзката с X/Elon Musk. • Загуба на доверие при непроверено AI съдържание.
7. Стратегически рискове <i>Vendor lock-in, ценови промени, спиране на услугата</i> • Vendor lock-in към американски доставчик и специфична екосистема. • Промяна на цени, лимити, модели или условия за ползване. • Спиране/ограничаване на модели и пренасочване към нови версии. • Риск при зависимост от външна платформа за публични процеси.	8. Засегнати страни <i>Граждани, служители, ръководство, партньори</i> • Граждани и заявители в административни услуги. • Служители, експерти, ИТ и длъжностно лице по защита на данните. • Ръководство, правен отдел, обществени поръчки и одит. • Партньори, доставчици, медии и контролни органи.	9. Остатъчни рискове <i>Рискове след митигация, приети рискове</i> • Дори с човешка проверка остават фактически грешки и непълноти. • Остава регулаторна несигурност при високорискови публични употреби. • Остава риск от промяна на условия/модел без предварителен контрол. • Приема се само помощна употреба, не самостоятелно вземане на решения.

Рисков регистър

Попълнете по всяка категория конкретните рискове с вероятност, въздействие и мерки.

 **Инструкция за попълване:** Вер. = Вероятност (1–5) · Вл. = Въздействие (1–5) · Ниво: ☐Н (нисък) ☐С (среден) ☐В (висок)

Риск / Събитие	Вероятност	Въздействие	Ниво	Мярка за митигация
Технически рискове				
Халюцинации и фактически грешки при отговори	Вер. ○1 ○2 ○3 X4 ○5	Вл. ○1 ○2 ○3 ○4 X5	Ниво ☐ Н ☐ С ☐ В	Мярка за митигация: Задължителна човешка проверка; забрана за директно използване като окончателен административен/експертен извод; цитиране на първоизточници.
Недостъпност / downtime на платформата	Вер. ○1 ○2 X3 ○4 ○5	Вл. ○1 ○2 ○3 ○4 X5	Ниво ☐ Н ☐ С ☐ В	Мярка за митигация: Резервен процес без AI; мониторинг на статуса; критичните срокове не се обвързват с наличността на Grok.
Внезапна промяна на версия или поведение на модела	Вер. ○1 ○2 ○3 X4 ○5	Вл. ○1 ○2 ○3 X4 ○5	Ниво ☐ Н ☐ С ☐ В	Мярка за митигация: Регистър на използваната версия; тестове след промяна; периодичен преглед на промптове и вътрешни инструкции.
Ограничения на контекстния прозорец при обемни документи	Вер. ○1 ○2 ○3 X4 ○5	Вл. ○1 ○2 ○3 X4 ○5	Ниво ☐ Н ☐ С ☐ В	Мярка за митигация: Разделяне на документи; контролен списък за пълнота; сравнение с оригинала при обемни файлове.
Грешки при мултимодален вход (изображения, аудио)	Вер. ○1 ○2 X3 ○4 ○5	Вл. ○1 ○2 ○3 ○4 X5	Ниво ☐ Н ☐ С ☐ В	Мярка за митигация: Визуална проверка от служител; забрана за единствена основа при чертежи, снимки, медицински/правни изображения.
Рискове за данните и поверителността				
Обработка на лични данни на граждани в облачна среда	Вер. ○1 ○2 ○3 X4 ○5	Вл. ○1 ○2 ○3 ○4 X5	Ниво ☐ Н ☐ С ☐ В	Мярка за митигация: DPIA; минимизация; псевдонимизация; забрана за чувствителни данни без основание; договорни и технически гаранции.
Изтичане на класифицирана/служебна информация	Вер. ○1 ○2 X3 ○4 ○5	Вл. ○1 ○2 ○3 ○4 X5	Ниво ☐ Н ☐ С ☐ В	Мярка за митигация: Забрана за класифицирана/служебна тайна; DLP правила; обучение; санкции при нарушение; използване само на одобрена среда.

Запазване на запитвания от доставчика за обучение	Вер. ○1 ○2 ○3 X4 ○5	Вл. ○1 ○2 ○3 X4 ○5	Ниво □H □C □XB	Мярка за митигация: Изключване на обучение/персонализация, когато е налично; enterprise/API условия; вътрешна политика “без лични данни в промптове”.
Трансфер на данни извън ЕС (EU AI Act / Schrems II)	Вер. ○1 ○2 ○3 X4 ○5	Вл. ○1 ○2 ○3 X4 ○5	Ниво □H □C □XB	Мярка за митигация: Оценка на трансферите; SCC/TIA; консултация с ДЛЗД; избор на режим с подходящи гаранции за ЕС.
Липса на одитна следа при автоматизирани решения	Вер. ○1 ○2 X3 ○4 ○5	Вл. ○1 ○2 ○3 X4 ○5	Ниво □H XC □B	Мярка за митигация: Локален лог на заявка, версия, оператор, дата и използван изход; забрана за скрити автоматизирани решения.
Правни и регулаторни рискове				
Несъответствие с EU AI Act (Annex III — публичен сектор)	Вер. ○1 ○2 X3 ○4 ○5	Вл. ○1 ○2 ○3 ○4 X5	Ниво □H □C XB	Мярка за митигация: Класификация на use case; ако е високорисков: risk management, документация, човешки надзор, мониторинг, регистрация при приложимост.
Нарушение на GDPR при обработка на лични данни	Вер. ○1 ○2 ○3 X4 ○5	Вл. ○1 ○2 ○3 ○4 X5	Ниво □H □C XB	Мярка за митигация: Правно основание; информиране; минимизация; срокове за съхранение; права на субектите; DPIA за висок риск.
Липса на обяснимост при автоматизирани административни решения	Вер. ○1 ○2 X3 ○4 ○5	Вл. ○1 ○2 ○3 ○4 X5	Ниво □H □C XB	Мярка: AI само като помощен инструмент; мотивите се пишат и подписват от компетентен човек; съхранение на проверими източници.
Авторскоправни нарушения в генерирано съдържание	Вер. ○1 ○2 X3 ○4 ○5	Вл. ○1 ○2 X3 ○4 ○5	Ниво □H XC □B	Мярка за митигация: Проверка за права и лиценз; избягване на дословно възпроизвеждане; маркиране на AI-помощ при външни материали.
Отговорност при грешни правни или медицински изходи	Вер. ○1 ○2 X3 ○4 ○5	Вл. ○1 ○2 ○3 ○4 X5	Ниво □H □C XB	Мярка за митигация: Забрана за правни/медицински решения без специалист; шаблон “не е окончателно становище”; втори контрол при чувствителни казуси.
Етични и социални рискове				
Систематичен bias спрямо определени групи граждани	Вер. ○1 ○2 X3 ○4 ○5	Вл. ○1 ○2 ○3 X4 ○5	Ниво □H XC □B	Мярка за митигация: Тестови набори по групи; преглед за дискриминационни резултати; канал за жалби и корекции.
Дискриминация при автоматизирано обработване на заявления	Вер. ○1 ○2 X3 ○4 ○5	Вл. ○1 ○2 ○3 X4 ○5	Ниво □H □C XB	Мярка за митигация: Без автоматично класиране/отказ; човек взема решението; документирани критерии; редовен одит.

Непрозрачност — граждани не знаят, че AI взима решения	Вер. ○1 ○2 ○3 X4 ○5	Вл. ○1 ○2 X3 ○4 ○5	Ниво □Н □С □В	Мярка за митигация: Ясно уведомяване кога е използван AI; публична политика; право на човешки преглед.
Дезинформация и манипулация чрез убедително съдържание	Вер. ○1 ○2 ○3 X4 ○5	Вл. ○1 ○2 ○3 X4 ○5	Ниво □Н □С □В	Мярка за митигация: Проверка през надеждни източници; забрана за публикуване без редактор; специален режим за избори, кризи и здраве.
Загуба на критично мислене и зависимост при служителите	Вер. ○1 ○2 ○3 X4 ○5	Вл. ○1 ○2 X3 ○4 ○5	Ниво □Н □С □В	Мярка за митигация: Обучение за критична проверка; ротация на задачи; изискване за поне един независим източник при факти.
Оперативни рискове				
Липса на технически капацитет за внедряване и поддръжка	Вер. ○1 ○2 X3 ○4 ○5	Вл. ○1 ○2 ○3 X4 ○5	Ниво □Н □С □В	Мярка за митигация: Пилотен проект; собственик на процеса; ИТ и правна поддръжка; ясна процедура за инциденти.
Vendor lock-in — прекомерна зависимост от доставчик	Вер. ○1 ○2 ○3 X4 ○5	Вл. ○1 ○2 ○3 X4 ○5	Ниво □Н □С □В	Мярка за митигация: Мултивендор стратегия; експортируеми промптове/логове; периодичен пазарен преглед; договорни изходни клаузи.
Недостатъчно обучение на служителите	Вер. ○1 ○2 ○3 X4 ○5	Вл. ○1 ○2 X3 ○4 ○5	Ниво □Н □С □В	Мярка за митигация: Кратко задължително обучение; примерни промптове; забранени употреби; периодични проверки.
Интеграционни проблеми с наследени (legacy) системи	Вер. ○1 ○2 X3 ○4 ○5	Вл. ○1 ○2 ○3 X4 ○5	Ниво □Н □С □В	Мярка за митигация: Пилотно интегриране; API gateway; тестова среда; ограничени права; план за rollback.
Ценови скокове при промяна на лицензионния модел	Вер. ○1 ○2 ○3 X4 ○5	Вл. ○1 ○2 X3 ○4 ○5	Ниво □Н □С □В	Мярка за митигация: Бюджетен буфер; рамково споразумение; лимити по потребители; алтернативни модели.
Репутационни и стратегически рискове				
Публичен скандал при видима AI грешка (медийно отразяване)	Вер. ○1 ○2 ○3 X4 ○5	Вл. ○1 ○2 ○3 ○4 X5	Ниво □Н □С □В	Мярка за митигация: Процедура за публична реакция; предварително одобрени съобщения; журнал на проверките; човек подписва крайния резултат.
Загуба на доверие на гражданите в публичните институции	Вер. ○1 ○2 X3 ○4 ○5	Вл. ○1 ○2 ○3 ○4 X5	Ниво □Н □С □В	Мярка за митигация: Прозрачност, човешки контрол и право на обжалване; публично описание на помощната роля на AI.
Политическа съпротива срещу използването на AI	Вер. ○1 ○2 X3 ○4 ○5	Вл. ○1 ○2 ○3 X4 ○5	Ниво □Н □С □В	Мярка за митигация: Ранно включване на ръководство, юристи, синдикати/служители и

				граждански представители; ограничен пилот.
Зависимост от американски/китайски технологичен доставчик	Вер. ○1 ○2 ○3 X4 ○5	Вл. ○1 ○2 ○3 X4 ○5	Ниво ☐ Н ☐ С ☒ В	Мярка за митигация: Оценка на суверенитет на данните; европейски алтернативи за чувствителни процеси; договорни гаранции.
Спиране на услугата при геополитически напрежения	Вер. ○1 X2 ○3 ○4 ○5	Вл. ○1 ○2 ○3 X4 ○5	Ниво ☐ Н ☒ С ☐ В	Мярка за митигация: План за непрекъсваемост; алтернативни доставчици; локални шаблони и процедури без AI.

Обобщена рискова оценка на групата

Топ 3 критични риска за модела:

1. Обработка на лични/служебни данни в облачна среда и възможно използване на входиове за обучение/персонализация.

2. Халюцинации и убедително звучащи грешки при правни, административни, технически и публично чувствителни теми.

3. Регулаторен риск при използване в публичния сектор, особено при услуги, права на граждани и автоматизирани решения.

Обща рискова оценка на модела за публичния сектор:

☐ Нисък риск — може да се внедри с минимални ограничения

☐ Среден риск — изисква допълнителни контроли и процедури

☒ **Висок риск — изисква задълбочена оценка и специфична регулация**

☐ **Неприемлив риск — не се препоръчва за публичния сектор**

Мотивирано становище:

Grok може да бъде полезен като помощен инструмент за търсене, обобщаване, чернови, код и първоначален анализ, но не следва да се използва самостоятелно за административни решения, правни становища, обработка на чувствителни данни или комуникация към граждани без човешки контрол.

Поради облачна обработка, интеграция с X/Grok, възможно използване на данни за обучение/персонализация според настройките, риск от халюцинации и регулаторна несигурност по EU AI Act, препоръката е: ограничен пилот, DPIA, ясна вътрешна политика, логване, обучение и задължителна човешка валидация. Да, бюрокрацията пак печели, но понякога точно тя предпазва от скъп публичен спектакъл.

Оценката е изготвена по публично достъпна информация към 14.05.2026.

Карта за приложимост

AI модел в публичната администрация — функционален анализ

v1.0 — Работен шаблон

 **Инструкция за попълване:** Попълнете всички полета за избрания от вашата група AI модел. Празните редове и скалите са предназначени за ръчно попълване. Не изтривайте инструкциите в курсив.

Наименование на модела	Пример: Claude Opus 4.7 / GPT-5.4 / Gemini 3.1 Pro xAI Grok (4.20 / 4)
Разработчик / компания	Пример: Anthropic / OpenAI / Google DeepMind xAI
Версия за оценка	Точната версия, която групата изследва Grok 4.3 (Beta)
Дата на попълване	ДД.ММ.ГГГГ 14.05.2026 г.
Членове на групата	Имена на участниците Хефест – Манол, Николай, Делян, Диан, Наталия, Силвия, Петър
Ръководител на групата	Наталия

Част 1. Обзорна матрица на приложимостта

Бърза оценка по административни функции. Прил. = Приложимост 1–5. Риск = Н/С/В.

 **Инструкция за попълване:** Попълнете всяка клетка: поставете оценка за приложимост (1=неприложимо, 5=напълно приложимо), ниво на риск (Н/С/В) и кратка бележка.

Административна функция	Обработка на документи	Правен анализ	Граждански услуги	Анализ на данни	Програмиране	Обучение
Обработка на кореспонденция	Прил.: ○1 ○2 ○3 X4 ○5 Риск: □ Н XС □В	Прил.: ○1 ○2 X3 ○4 ○5 Риск: □ Н □С XВ	Прил.: ○1 ○2 ○3 X4 ○5 Риск: □ Н XС □В	Прил.: ○1 X2 ○3 ○4 ○5 Риск: XН □С □В	Прил.: ○1 X2 ○3 ○4 ○5 Риск: XН □С □В	Прил.: ○1 ○2 X3 ○4 ○5 Риск: XН □С □В
Изготвяне на доклади и анализи	Прил.: ○1 ○2 ○3 X4 ○5 Риск: □ Н XС □В	Прил.: ○1 ○2 X3 ○4 ○5 Риск: □ Н □С XВ	Прил.: ○1 ○2 X3 ○4 ○5 Риск: □ Н XС □В	Прил.: ○1 ○2 ○3 ○4 X5 Риск: □ Н XС □В	Прил.: ○1 ○2 X3 ○4 ○5 Риск: XН □С □В	Прил.: ○1 ○2 ○3 X4 ○5 Риск: XН □С □В
Тълкуване на нормативни актове	Прил.: ○1 ○2 X3 ○4 ○5 Риск: □ Н XС □В	Прил.: ○1 ○2 X3 ○4 ○5 Риск: □ Н □С XВ	Прил.: ○1 ○2 X3 ○4 ○5 Риск: □ Н □С XВ	Прил.: ○1 X2 ○3 ○4 ○5 Риск: XН □С □В	Прил.: X1 ○2 ○3 ○4 ○5 Риск: XН □С □В	Прил.: ○1 ○2 ○3 X4 ○5 Риск: □ Н XС □В

Изготвяне на правни становища	Прил.: o1 o2 X3 o4 o5 Риск: □ Н □С □В	Прил.: o1 X2 o3 o4 o5 Риск: □ Н □С ХВ	Прил.: o1 X2 o3 o4 o5 Риск: □ Н □С ХВ	Прил.: o1 X2 o3 o4 o5 Риск: ХН □С □В	Прил.: X1 o2 o3 o4 o5 Риск: ХН □С □В	Прил.: o1 o2 X3 o4 o5 Риск: □ Н □С □В
Отговори на граждански запитвания	Прил.: o1 o2 o3 X4 o5 Риск: □ Н □С □В	Прил.: o1 o2 X3 o4 o5 Риск: □ Н □С ХВ	Прил.: o1 o2 o3 X4 o5 Риск: □ Н □С ХВ	Прил.: o1 X2 o3 o4 o5 Риск: ХН □С □В	Прил.: X1 o2 o3 o4 o5 Риск: ХН □С □В	Прил.: o1 o2 X3 o4 o5 Риск: ХН □С □В
Чатбот за електронни услуги	Прил.: o1 o2 X3 o4 o5 Риск: □ Н □С □В	Прил.: o1 X2 o3 o4 o5 Риск: □ Н □С ХВ	Прил.: o1 o2 o3 X4 o5 Риск: □ Н □С ХВ	Прил.: o1 o2 X3 o4 o5 Риск: □ Н □С □В	Прил.: o1 o2 o3 o4 X5 Риск: □ Н □С □В	Прил.: o1 o2 X3 o4 o5 Риск: ХН □С □В
Статистически анализ	Прил.: o1 X2 o3 o4 o5 Риск: ХН □С □В	Прил.: X1 o2 o3 o4 o5 Риск: ХН □С □В	Прил.: X1 o2 o3 o4 o5 Риск: ХН □С □В	Прил.: o1 o2 o3 o4 X5 Риск: □ Н □С □В	Прил.: o1 o2 o3 X4 o5 Риск: ХН □С □В	Прил.: o1 o2 X3 o4 o5 Риск: ХН □С □В
Обработка на отчети	Прил.: o1 o2 o3 o4 X5 Риск: □ Н □С □В	Прил.: o1 X2 o3 o4 o5 Риск: □ Н □С □В	Прил.: o1 X2 o3 o4 o5 Риск: ХН □С □В	Прил.: o1 o2 o3 o4 X5 Риск: □ Н □С □В	Прил.: o1 o2 X3 o4 o5 Риск: ХН □С □В	Прил.: o1 o2 X3 o4 o5 Риск: ХН □С □В
Автоматизация на работни процеси	Прил.: o1 o2 o3 X4 o5 Риск: □ Н □С □В	Прил.: o1 X2 o3 o4 o5 Риск: □ Н □С □В	Прил.: o1 o2 X3 o4 o5 Риск: □ Н □С □В	Прил.: o1 o2 o3 X4 o5 Риск: □ Н □С □В	Прил.: o1 o2 o3 o4 X5 Риск: □ Н □С □В	Прил.: o1 o2 X3 o4 o5 Риск: ХН □С □В
Разработка на вътрешни инструменти	Прил.: o1 X2 o3 o4 o5 Риск: ХН □С □В	Прил.: X1 o2 o3 o4 o5 Риск: ХН □С □В	Прил.: o1 X2 o3 o4 o5 Риск: ХН □С □В	Прил.: o1 o2 o3 X4 o5 Риск: □ Н □С □В	Прил.: o1 o2 o3 o4 X5 Риск: □ Н □С □В	Прил.: o1 o2 X3 o4 o5 Риск: ХН □С □В
Учебни материали и тестове	Прил.: o1 o2 o3 X4 o5 Риск: ХН □С □В	Прил.: o1 X2 o3 o4 o5 Риск: □ Н □С □В	Прил.: o1 X2 o3 o4 o5 Риск: ХН □С □В	Прил.: o1 X2 o3 o4 o5 Риск: ХН □С □В	Прил.: o1 o2 o3 X4 o5 Риск: ХН □С □В	Прил.: o1 o2 o3 o4 X5 Риск: ХН □С □В
Наръчници и процедури	Прил.: o1 o2 o3 o4 X5 Риск: ХН □С □В	Прил.: o1 o2 X3 o4 o5 Риск: □ Н □С □В	Прил.: o1 X2 o3 o4 o5 Риск: ХН □С □В	Прил.: o1 X2 o3 o4 o5 Риск: ХН □С □В	Прил.: o1 X2 o3 o4 o5 Риск: ХН □С □В	Прил.: o1 o2 o3 o4 X5 Риск: ХН □С □В
Превод и локализация	Прил.: o1 o2 o3 X4 o5 Риск: ХН □С □В	Прил.: o1 X2 o3 o4 o5 Риск: □ Н □С □В	Прил.: o1 o2 X3 o4 o5 Риск: ХН □С □В	Прил.: X1 o2 o3 o4 o5 Риск: ХН □С □В	Прил.: X1 o2 o3 o4 o5 Риск: ХН □С □В	Прил.: o1 o2 o3 X4 o5 Риск: ХН □С □В
Протоколи и резюмета	Прил.: o1 o2 o3 o4 X5 Риск: ХН □С □В	Прил.: o1 o2 X3 o4 o5 Риск: □ Н □С □В	Прил.: o1 X2 o3 o4 o5 Риск: ХН □С □В	Прил.: o1 X2 o3 o4 o5 Риск: ХН □С □В	Прил.: X1 o2 o3 o4 o5 Риск: ХН □С □В	Прил.: o1 o2 o3 X4 o5 Риск: ХН □С □В

Част 2. Детайлен анализ по случаи на употреба

За всеки сценарий оценете конкретно как избраният модел се справя.

2.1. Обработка на административни документи
Способността на модела да чете, структурира, резюмира и изготвя административни документи — решения, наредби, докладни записки, протоколи.

Оценка за избрания модел:

1. Може ли моделът да обработва документи на български език с административна терминология?
2. Поддържа ли обработка на PDF/Word файлове с дълъг обем?
3. Запазва ли форматирането и структурата при генериране на изходен документ?
4. Изисква ли специализирано prompt инженерство за административен контекст?

Отговори / наблюдения:

- ☐ Моделът генерира разбираеми административни текстове.
- ☐ Добре резюмира и структурира документи.
- ☐ Възможни са фактически неточности при сложни казуси.
- ☐ При дълги документи е възможна загуба на контекст.
- ☐ Изисква човешка проверка преди официално използване.
- ☐ Добре работи при чернови, протоколи и вътрешна документация.

Препоръка:

☐ Препоръчва се без ограничения

☐ Препоръчва се с ограничения (опишете):

Задължителна човешка проверка и забрана за автоматично генериране на окончателни административни актове.

☐ Не се препоръчва (обосновка):

2.2. Правен анализ и тълкуване

Способността да анализира нормативни актове, да открива противоречия, да прилага правни норми към конкретни случаи.

Оценка за избрания модел:

1. Колко точен е при тълкуване на български законодателни текстове?
2. Може ли да следи промени в нормативната база (обновени закони)?
3. Как се справя с многозначност и правна неяснота?
4. Маркира ли ясно когато излиза извън областта на своята компетентност?

Отговори / наблюдения:

- ☐ Моделът може да обобщава нормативни текстове.
- ☐ Не гарантира актуалност на законодателството.
- ☐ Съществува висок риск от убедително звучащи грешки.
- ☐ Не е подходящ за окончателни правни решения.
- ☐ Изисква постоянен човешки контрол.

Препоръка:

☐ Препоръчва се без ограничения

☒ Препоръчва се с ограничения (опишете):

Само като помощен инструмент под контрол на юрист.

☐ Не се препоръчва (обосновка):

2.3. Взаимодействие с граждани (чатбот / асистент)

Способността да комуникира с граждани — отговаря на въпроси, помага при подаване на заявления, обяснява административни процедури.

Оценка за избрания модел:

1. Разбира ли естествен разговорен български (включително диалекти и грешки)?
2. Дава ли ясни и разбираеми отговори без административен жаргон?
3. Как се справя с чувствителни теми (социални помощи, здраве, правосъдие)?

4. Може ли да се интегрира с портали за електронни услуги?

Отговори / наблюдения:

- ☐ Добре обработва общи граждански въпроси.
- ☐ Възможни са неточни или подвеждащи отговори.
- ☐ Не трябва да предоставя окончателни административни решения.
- ☐ Необходимо е ясно обозначаване, че се използва AI.

Препоръка:

☐ Препоръчва се без ограничения

☒ Препоръчва се с ограничения (опишете):

Само за информационни цели и под човешки контрол.

☐ Не се препоръчва (обосновка):

2.4. Анализ на данни и отчетност

Способността да обработва структурирани данни, да генерира отчети и да идентифицира тенденции.

Оценка за избрания модел:

1. Може ли да обработва Excel/CSV данни и да генерира анализи?
2. Поддържа ли визуализация на данни (графики, диаграми)?
3. Колко надежден е при числови изчисления и статистика?
4. Може ли да се свърже с вътрешни бази данни чрез API?

Отговори / наблюдения:

- ☐ Подходящ за първичен анализ и обобщение.
- ☐ Добре структурира отчети.
- ☐ Възможни са грешки при сложни статистически изчисления.
- ☐ Необходимо е валидиране от служител.

Препоръка:

☐ Препоръчва се без ограничения

☒ Препоръчва се с ограничения (опишете):

Числовите и аналитичните резултати подлежат на проверка.

☐ Не се препоръчва (обосновка):

2.5. Вътрешно обучение на служители

Способността да генерира учебни материали, тестове, наръчници и да поддържа интерактивно обучение.

Оценка за избрания модел:

1. Може ли да генерира тематични курсове и тестове за служители?
2. Адаптира ли съдържанието към различни нива на компетентност?
3. Поддържа ли формати за дистанционно обучение (SCORM и др.)?
4. Как се справя с чувствителни теми — антикорупция, етика, дискриминация?

Отговори / наблюдения:

- ☐ Подходящ за вътрешни обучения.
- ☐ Генерира добри обяснения и примери.
- ☐ Добра адаптация към административен контекст.
- ☐ Нисък риск при вътрешна употреба.

Препоръка:

☒ Препоръчва се без ограничения

☐ Препоръчва се с ограничения (опишете):

☐ Не се препоръчва (обосновка):

2.6. Програмиране и автоматизация

Способността да пише код, да автоматизира процеси и да поддържа вътрешни системи.

Оценка за избрания модел:

- Поддържа ли езиките и технологиите, използвани в административните системи?
- Може ли да чете и документира съществуващ (legacy) код?
- Колко надеждни са генерираните скриптове без допълнителна проверка?
- Поддържа ли agentic режим за автономно изпълнение на задачи?

Отговори / наблюдения:

- ☐ Силна страна на модела е програмирането.
- ☐ Подходящ за automation и internal tooling.
- ☐ Генерираният код може да съдържа грешки.
- ☐ Необходимо е тестване преди продукционна среда.

Препоръка:

- ☐ Препоръчва се без ограничения
- ☒ Препоръчва се с ограничения (опишете):

Задължителен code review и тестова среда.

- ☐ Не се препоръчва (обосновка):

Част 3. Оценка на интеграционния потенциал

 **Инструкция за попълване:** Оценете техническите възможности за интеграция с административната инфраструктура.

Интеграционен аспект	Описание	Поддръжка	Бележки
API свързване	REST API, GraphQL, webhooks — интеграция с вътрешни системи	☒ Да ☐ Не ☐ Частично	Поддържа API интеграция и automation workflows
Single Sign-On	Съвместимост с Active Directory / ADFS / Keycloak	☐ Да ☐ Не ☒ Частично	Зависи от външни системи и корпоративна интеграция
On-premise deployment	Възможност за локален хостинг без облачна зависимост	☐ Да ☒ Не ☐ Частично	Няма официална on-premise поддръжка
Данни в ЕС	Сървъри на европейска територия — GDPR съответствие	☐ Да ☒ Не ☐ Частично	Данните може да се обработват извън ЕС
Одитна следа	Логирание на всички взаимодействия за отчетност	☐ Да ☐ Не ☒ Частично	Зависи от внедрените логове и monitoring системи
Ролеви достъп	RBAC — различни нива на достъп за различни служители	☐ Да ☐ Не ☒ Частично	Изисква допълнителна

			организационна конфигурация
Многоезичност	Поддръжка на български и официални ЕС езици	X Да ☐ Не ☐ Частично	Поддържа български и други езици
Мобилен достъп	Работа от мобилни устройства (field work)	X Да ☐ Не Частично	Достъпен през мобилен браузър и мобилни приложения
Офлайн режим	Функционалност без постоянна интернет връзка	☐ Да X Не Частично	Изисква интернет свързаност и cloud инфраструктура
SLA и поддръжка	Гарантирано ниво на обслужване за публичния сектор	☐ Да ☐ Не X Частично	Зависи от използвания план и корпоративните услуги

Обобщена карта за приложимост

Топ 3 приложения с висок потенциал:

1. Автоматизация на вътрешни административни процеси и обработка на документи
2. Разработка на вътрешни AI инструменти, scripting и programming assistance
3. Генериране на резюмета, анализи, протоколи и обучителни материали

Функции, при които моделът НЕ е подходящ:

1. Самостоятелно вземане на административни или правни решения
2. Обработка на класифицирана или силно чувствителна информация
3. Автоматизирано предоставяне на правни становища без човешки контрол

Препоръчани условия за пилотно внедряване:

- ☐ Използване първоначално само в нискорискови административни процеси
- ☐ Задължителна човешка проверка на всички критични резултати и документи
- ☐ Ограничаване на достъпа до лични, чувствителни и класифицирани данни
- ☐ Провеждане на обучение на служителите относно рисковете и ограниченията на AI
- ☐ Въвеждане на логване, мониторинг и периодичен одит на AI употребата
- ☐ Използване на вътрешни правила за сигурност, GDPR контрол и управление на достъпа

Методологични указания

Етично и прозрачно ползване на AI модели в публичната администрация

v1.0 — Работен шаблон

Инструкция за попълване: Попълнете всички полета за избория от вашата група AI модел. Празните редове и скалите са предназначени за ръчно попълване. Не изтривайте инструкциите в курсив.

аименование на модела	Пример: Claude Opus 4.7 / GPT-5.4 / Gemini 3.1 Pro xAI Grok (4.20 / 4)
Разработчик / компания	Пример: Anthropic / OpenAI / Google DeepMind xAI
Версия за оценка	Точната версия, която групата изследва Grok 4.3 (Beta)
Дата на попълване	ДД.ММ.ГГГГ 14.05.2026 г.
Членове на групата	Имена на участниците Хефест – Манол, Николай, Делян, Диан, Наталия, Силвия, Петър
Ръководител на групата	Наталия

Предговор и обхват

Инструкция за попълване: Попълнете идентификационните данни на вашата организация и дефинирайте обхвата на документа.

Наименование на организацията	Пълно официално наименование Областна администрация и съпътстващи общински структури
Административно ниво	Централна / областна / общинска администрация Регионално / Областно / Общинско ниво
AI модел(и), обхванати от документа	Посочете конкретните модели/версии xAI
Версия на документа	v1.0 — Официално утвърдени указания
Дата на приемане	ДД.ММ.ГГГГ 14.05.2026 г.
Следваща планирана ревизия	ДД.ММ.ГГГГ 14.05.2027 г.
Одобрен от	Длъжност и имена Ръководител администрацията
Отговорен за прилагането	Длъжност, звено и контакти IT отдел и AI отговорник

Чл. 1. Предмет и обхват

Опишете за какво се отнася документът — кои AI системи, кои структурни звена и кои видове задачи обхваща.

Настоящите методологични указания определят правилата, условията и ограниченията за използване на xAI Grok в дейността на публичната администрация. Документът обхваща използването на AI системата при обработка на документи, административна кореспонденция, анализ на данни, обучение на служители, автоматизация на процеси и помощни информационни услуги.

Указанията се прилагат за всички административни звена, служители и технически екипи, които използват AI инструменти при изпълнение на служебните си задължения.

Документът не допуска използване на AI за самостоятелно вземане на административни решения, обработка на класифицирана информация или предоставяне на окончателни правни становища без човешки контрол.

Чл. 2. Правна и регулаторна рамка

Посочете нормативните актове, на базата на които се издават указанията (EU AI Act, GDPR, ЗЕУ, вътрешни наредби).

Настоящите указания се издават на основание:

Регламент (ЕС) относно изкуствения интелект (EU AI Act);
Общ регламент относно защитата на данните (GDPR);
Закон за електронното управление;
Закон за защита на личните данни;
Закон за киберсигурност;
Вътрешните правила за информационна сигурност и управление на риска;
Вътрешните политики и процедури на организацията относно използването на информационни технологии и защита на информацията.

Раздел I. Принципи за етично и прозрачно ползване

Определете основните принципи, на които се базира политиката. Попълнете конкретните мерки за всеки.

Основни принципи	
1 Законност	AI системите се използват само в рамките на действащото законодателство. Всяко решение, подпомогнато от AI, спазва приложимите правни норми. Конкретни мерки в нашата организация: ☐ Използване на AI системата само за законосъобразни административни дейности; ☐ Спазване на EU AI Act, GDPR и националното законодателство; ☐ Забрана за използване на AI при дейности, противоречащи на нормативната рамка; ☐ Правен преглед на нови случаи на употреба преди внедряване.
2 Прозрачност	Гражданите и служителите имат право да знаят кога AI участва в процесите. Резултатите от AI трябва да са обяснени разбираемо. Конкретни мерки в нашата организация: ☐ Ясно обозначаване при използване на AI в комуникация с граждани; ☐ Информиране на служителите за ограниченията и рисковете на AI системата; ☐ Документиране на основните AI процеси и случаи на употреба; ☐ Поддържане на проследимост на AI генерираните резултати.
3 Отчетност	Ясна верига на отговорност. Служителите носят отговорност за решенията, взети с AI помощ — AI не замества човешката отговорност. Конкретни мерки в нашата организация: ☐ Определяне на отговорни лица и звена за AI управлението; ☐ Водене на логове и проследяване на AI взаимодействията; ☐ Провеждане на вътрешен одит и периодичен контрол;

	☐ Документиране на инциденти, корекции и актуализации.
4 Справедливост	AI инструментите не дискриминират. Резултатите се проверяват редовно за пристрастия (bias) спрямо различни групи граждани. Конкретни мерки в нашата организация: ☐ Ограничаване на обработката на лични и чувствителни данни; ☐ Спазване на принципа за минимизация на данните; ☐ Забрана за използване на класифицирана информация в публични AI системи; ☐ Прилагане на вътрешни политики за информационна сигурност и GDPR контрол.
5 Поверителност	Личните данни на гражданите се обработват само при наличие на правно основание и при спазване на GDPR. Конкретни мерки в нашата организация: ☐ Ограничаване на обработката на лични и чувствителни данни; ☐ Спазване на принципа за минимизация на данните; ☐ Забрана за използване на класифицирана информация в публични AI системи; ☐ Прилагане на вътрешни политики за информационна сигурност и GDPR контрол.
6 Надеждност	AI системите се използват само за задачи, при които точността е потвърдена. Резултатите подлежат на верификация. Конкретни мерки в нашата организация: ☐ Проверка и верификация на AI резултатите преди официално използване; ☐ Периодично тестване на качеството и точността на системата; ☐ Ограничаване на AI употребата при високорискови административни дейности; ☐ Докладване и анализ на установени грешки и инциденти; ☐ Поддържане на човешки контрол при критични процеси; ☐ Използване на AI само в рамките на одобрени административни задачи.
7 Пропорционалност	Нивото на автоматизация съответства на риска от грешки. При по-висок риск — по-задължителен човешки контрол. Конкретни мерки в нашата организация: ☐ Използване на AI само при дейности с подходящо ниво на риск; ☐ Ограничаване на AI употребата при чувствителни административни процеси; ☐ Прилагане на допълнителни мерки при високорискови случаи; ☐ Съобразяване на AI внедряването с реалните административни нужди.
8 Обяснимост	При поискване от гражданин или орган, организацията може да обясни как е бил използван AI при дадено решение. Конкретни мерки в нашата организация: ☐ Осигуряване на възможност за проследяване на AI генерираните резултати; ☐ Документиране на използваните AI процеси и случаи на употреба; ☐ Предоставяне на разбираема информация за ролята на AI в административния процес; ☐ Запазване на логове и история на AI взаимодействията; ☐ Осигуряване на възможност служител да обясни или коригира AI резултат; ☐ Информиране на гражданите при използване на AI в комуникацията или услугите.

Раздел II. Разрешени и забранени употреби

Определете за всяка категория задачи дали е разрешена, допустима с условия или забранена.

 **Инструкция за попълване:** За всяка задача: поставете отметка в съответната колона и опишете условията или ограниченията.

Категория задача	Статус	Условия / Ограничения / Процедура
Административни документи		
Изготвяне на проекти за решения	☐ Разрешено ☒ С условия ☐ Забранено	Условия / ограничения / процедура: AI може да подпомага изготвянето на чернови и структурирането на административни решения, но крайният документ подлежи на задължителна проверка и одобрение от компетентен служител.
Резюмиране на входяща поща	☒ Разрешено ☐ С условия ☐ Забранено	Условия / ограничения / процедура: Допуска се използване за обобщаване и категоризация на входяща кореспонденция при спазване на вътрешните правила за защита на информацията.
Превод на документи	☒ Разрешено ☐ С условия ☐ Забранено	Условия / ограничения / процедура: Допуска се използване за работни и вътрешни преводи. Официалните документи подлежат на финална човешка проверка преди използване.
Генериране на официални писма	☐ Разрешено ☒ С условия ☐ Забранено	Условия / ограничения / процедура: Всички генерирани писма трябва да бъдат прегледани, редактирани и одобрени от отговорен служител преди изпращане.
Правни и нормативни въпроси		
Анализ на нормативни текстове	☐ Разрешено ☒ С условия ☐ Забранено	Условия / ограничения / процедура: AI се използва само като помощен инструмент за обобщаване и първичен анализ. Не се допуска самостоятелно правно тълкуване без участие на юрист.
Проверка за вътрешни противоречия	☒ Разрешено ☐ С условия ☐ Забранено	Условия / ограничения / процедура: Допуска се използване за подпомагане на административната проверка на документи, процедури и текстове, като резултатите подлежат на човешка верификация.
Изготвяне на правни становища	☐ Разрешено ☐ С условия ☒ Забранено	Условия / ограничения / процедура: Не се допуска използване на AI за самостоятелно изготвяне на окончателни правни становища или правни консултации без правоспособен юрист.
Тълкуване на разпоредби към граждани	☐ Разрешено ☒ С условия ☐ Забранено	Условия / ограничения / процедура: Информацията може да се използва само за общо разяснение. Не се допуска предоставяне на окончателни правни консултации без проверка от компетентен служител или юрист.
Вземане на решения с ефект върху граждани		
Преценка за социални помощи	☐ Разрешено ☐ С условия ☒ Забранено	Условия / ограничения / процедура: AI не може самостоятелно да извършва оценка, класиране или вземане на решения относно социални помощи и права на граждани.
Оценка на проекти и заявления	☐ Разрешено ☒ С условия ☐ Забранено	Условия / ограничения / процедура: AI може да подпомага административната обработка и предварителния анализ, но крайното решение се взема от комисия или оправомощен служител.

Наказателни и принудителни мерки	☐ Разрешено ☒ С условия ☒ Забранено	<i>Условия / ограничения / процедура:</i> Не се допуска използване на AI за автоматизирано налагане на санкции, ограничения или принудителни административни мерки.
Автоматизирани решения без човешки контрол	☐ Разрешено ☒ С условия ☒ Забранено	<i>Условия / ограничения / процедура:</i> Всички решения с административни, правни или финансови последици подлежат на човешка проверка и одобрение.
Комуникация с граждани		
Чатбот за информационни въпроси	☐ Разрешено ☒ С условия ☒ Забранено	<i>Условия / ограничения / процедура:</i> Допуска се използване само за обща и информационна комуникация. Гражданите трябва да бъдат уведомени, че комуникират с AI система.
Проекти за отговори на жалби	☐ Разрешено ☒ С условия ☒ Забранено	<i>Условия / ограничения / процедура:</i> AI може да генерира чернови и предложения, но всички отговори подлежат на проверка и одобрение от компетентен служител.
Публикации в социални медии	☐ Разрешено ☒ С условия ☒ Забранено	<i>Условия / ограничения / процедура:</i> Генерираното съдържание трябва да бъде прегледано и одобрено от отговорен служител или PR звено преди публикуване.
Кризисна комуникация	☐ Разрешено ☒ С условия ☒ Забранено	<i>Условия / ограничения / процедура:</i> AI може да подпомага подготовката на чернови, но публикуването на кризисна информация се извършва само след одобрение от ръководството.
Вътрешно-административни функции		
Обучение на служители	☒ Разрешено ☒ С условия ☒ Забранено	<i>Условия / ограничения / процедура:</i> Допуска се използване за генериране на обучителни материали, тестове, инструкции и вътрешни обучения.
Анализ на данни и отчети	☐ Разрешено ☒ С условия ☒ Забранено	<i>Условия / ограничения / процедура:</i> AI може да подпомага анализа и структурирането на данни, но числовите резултати и изводите подлежат на човешка проверка.
Автоматизация на повтарящи се задачи	☐ Разрешено ☒ С условия ☒ Забранено	<i>Условия / ограничения / процедура:</i> Допуска се автоматизация само след предварително тестване, контрол на сигурността и одобрение от IT отдела.
Обработка на класифицирана информация	☐ Разрешено ☒ С условия ☒ Забранено	<i>Условия / ограничения / процедура:</i> Не се допуска използване на публични AI системи за класифицирани, поверителни или строго чувствителни данни.

Раздел III. Процедури за одобрение и верификация

Чл. 10. Задължителна верификация

Опишете кои задачи изискват задължителна проверка от служител преди използване на AI изхода.

Всички резултати от AI системата подлежат на човешка проверка преди официално използване.

Чл. 11. Процедура за одобрение на нови употреби

Как служителят може да заяви одобрение за нов вид употреба на AI, непредвиден в настоящите указания?

Новите случаи на употреба се одобряват от IT отдела и правното звено.

Чл. 12. Задължително разкриване към граждани

В кои случаи гражданинът трябва да бъде информиран, че AI е участвал в обработката на неговото искане?

Не се допуска AI да взема самостоятелни решения с правни или административни последици.

Раздел IV. Мониторинг, одит и отчетност**Минимални изисквания за одитна следа (отбележете приложимите):**

- ☒ Всяко взаимодействие с AI системата се записва с дата, час и потребител
- ☒ Пазят се логове на AI изходите, използвани в официални документи
- ☒ При автоматизирани решения се съхранява AI входа и изхода
- ☒ Провежда се периодичен преглед на качеството на AI изходите (мин. 1х/тримесечие)
- ☒ Инцидентите, свързани с AI грешки, се докладват на определено звено
- ☒ Годишен отчет за употребата на AI се изготвя и публикува

Чл. 15. KPI за мониторинг на AI употребата

Определете конкретни измерими показатели: брой взаимодействия, % задачи с верификация, % открити грешки и др.

- ☐ Брой AI взаимодействия;
- ☐ Процент проверени AI резултати;
- ☐ Брой установени грешки;
- ☐ Средно време за обработка на документи;
- ☐ Брой докладвани инциденти;
- ☐ Брой обучени служители.

Раздел V. Отговорности и управление

Роля	Отговорност	Конкретно лице / звено
Ръководител на организацията	Стратегическо одобрение и крайна отговорност за AI политиката Издаване на вътрешни заповеди; стратегическо разпределение на ИТ бюджета за хардуер (сървъри).	Конкретно лице / звено: Кмет / ръководство
AI отговорник (AI Officer)	Координация, обучение, мониторинг, докладване на инциденти Мониторинг на спазването на методиката; координация между ИТ отдела и правното звено.	Конкретно лице / звено: ИТ отдел / AI координатор
Служители, ползващи AI	Спазване на указанията, верификация на изходите, докладване на проблеми Ръчна проверка на всеки абзац; недопускане на безкритично копиране на текстове; спазване на правилата за сигурност.	Конкретно лице / звено: Административни звена
ИТ отдел / системен администратор	Технически достъп, интеграция, сигурност, логове Поддръжка на хардуера, актуализация на теглата на моделите (open-weights), киберсигурност на мрежата, сигурност на логовете.	Конкретно лице / звено: Системна администрация

Правно звено	Преглед на нови употреби, съответствие с нормативната рамка Съгласуване на проектите на решения; преглед на договорите за хардуерна поддръжка спрямо българското законодателство.	Конкретно лице / звено: Юридически отдел
ДЛЗД (длъжностно лице по защита на данни)	GDPR съответствие при обработка на лични данни Гарантиране, че локалният трансфер на данни не нарушава правата на гражданите по GDPR; регулярни вътрешни проверки.	Конкретно лице / звено: DPO
Вътрешен одит	Периодичен независим преглед на AI употребата Проверка за икономическа целесъобразност на внедрената технология и спазване на вътрешните разпоредби.	Конкретно лице / звено: Одитно звено

Раздел VI. Обучение и изграждане на капацитет

Чл. 20. Задължително обучение			
Посочете кои служители трябва да преминат обучение, какъв е минималният обем и как се документира. Всички служители, използващи AI системи, преминават периодично обучение.			
Модул на обучение	Целева група	Форма и обем	Периодичност
Основи на AI и рискове	Всички служители	Всички служители	Всички служители
Работа с конкретния модел	IT и администрация	IT и администрация	IT, администрация
Етика и нормативна рамка	Всички служители	Всички служители	Всички служители
Разпознаване на AI грешки	Всички служители	Всички служители	Всички служители
Защита на личните данни	IT и ръководство	IT и ръководство	IT и ръководство

Раздел VII. Актуализация и валидност

Чл. 24. Условия за задължителна ревизия
Опишете кога документът трябва задължително да се преразгледа: нова версия на модела, регулаторна промяна, инцидент, изтичане на срок. ☐ промяна в нормативната рамка; ☐ нова версия на AI модела; ☐ установени инциденти; ☐ резултати от вътрешен одит; ☐ промени в организационните процеси; ☐ минимум веднъж годишно.
Чл. 25. История на промените
Документирайте всички версии на документа с дата, автор и описание на промяната. Всички промени по документа се документират с дата, описание и одобрение от ръководството.

Версия	Дата	Описание на промяната	Одобрил
v1.0	14.05.2026 г.	Първоначално приемане на методологичните указания за xAI Grok	Ръководител на организацията

ИНТЕЛИГЕНТНИ РЕШЕНИЯ.
ЕВРОПЕЙСКА НАДЕЖНОСТ.

AI КАРТА ЗА АДМИНИСТРАЦИИ И ИНСТИТУЦИИ

СУВЕРЕН, ОТВОРЕН, ЕФЕКТИВЕН

SWOT АНАЛИЗ



СИЛНИ СТРАНИ

- Глобален изглед с фокус върху суверенитет и прозрачност
- Отворени и прозрачни модели
- Висока производителност и ефективност
- Гъвкави модели на работен режим
- Подходящ за пълноценно внедряване (open platform)



СЛАБОСТИ

- По-ниска ефективност в сравнение с глобалните единоредни
- Все още ограничени мултимедийни възможности в някои модели
- Изисква допълнителни ресурси за сложни внедрявания



ВЪЗМОЖНОСТИ

- Отворени и прозрачни модели
- Подходящи за услуги за граждани
- Сигурен анализ на мултимедийни данни
- Развиване на глобални AI приложения в европейския регион



ЗАПЛАХИ

- Бързо развитие на конкурентни модели
- Невъзможност за лесно внедряване
- Риск от невъзможна употреба
- Зависимост от инфраструктурни компоненти

ПРИЛОЖИМОСТ В АДМИНИСТРАЦИИ



Автоматизация на документи и кореспонденция



Анализ и обобщаване на законодателствата и политики



Поддръжка на граждани чрез чат асистенти



Вътрешни търсения и търсене в документи



Анализ на обществени данни и отчети



Поддръжка при вземане на решения

КЛЮЧОВИ ХАРАКТЕРИСТИКИ

- Европейски принципи и съответствие с GDPR
- Възможност за лесно внедряване
- Гъвкави модели (различни размери)
- Сигна за прозрачност и приоткритост на ресурси
- Подходящ за чувствителни данни и критични системи

МЕТОДОЛОГИЯ ЗА ВНЕДРЯВАНЕ



ПРИНЦИПИ: ПРОЗРАЧНОСТ | СИГУРНОСТ | ОТГОВОРНОСТ | ЗАКОНСОБНОСТ

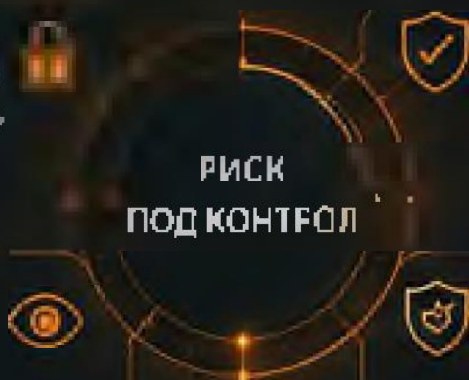
RISK CANVAS

РИСКОВЕ

- Изтичане на чувствителна информация
- Неправилни или подвеждателни отговори
- Злоупотреба с модел
- Технически провалявания

ПОСРЕДСТВА

- Използване на дигитален
- Репутационни щети
- Правни санкции
- Финансов загуби



ПРЕВЕНЦИЯ

- Контрол на достъпа и данните
- Глобално тестване
- Оптимизация на политиките за употреба
- Обучение на служителите

МИТИГАЦИЯ

- Наблюдение и отчет на корупция
- Мониторинг на репутацията
- Дълготрайно и редовно при изпитания
- Редовно актуализиране на модели

УКАЗАНИЯ ЗА УПОТРЕБА

- Определете ясно целта за всяко приложение
- Използвайте само разрешени данни
- Проверявайте резултатите срещу външни решения
- Демонстрирайте принципите и резултатите
- Осигурете обучение на служителите
- Спазвайте законодателството и вътрешните политики

AI КАРТА ЗА АДМИНИСТРАЦИИ И ИНСТИТУЦИИ

МОЩ РАЗБИРАНЕ. СКОРОСТ

SWOT АНАЛИЗ

СИЛНИ СТРАНИ

- Много силно разбиране на контекста и главни теми
- Висока скорост и ефективност на отговорите
- Достъп до актуална информация в реално време (кредитна информация)
- Подходящ за динамични среди и бързи решения

СЛАБОСТИ

- Зависимост от сложна инфраструктура
- По-малко контрол върху данните
- По-високи разходи при мултимедийна информация
- По-малко прозрачност на етапа на внедряване

ВЪЗМОЖНОСТИ

- Бърза обработка на големи обеми данни
- Мониторинг на обществени настроения и тенденции
- Глобален приют за критични ситуации
- Използване на мултимедийни източници и сложни данни

ЗАПЛАХИ

- Риск от дезинформация и манипулация
- Зависимост от сложни платформи
- Препреки в усвояването на данни
- Регулаторни ограничения

ПРИЛОЖИМОСТ В АДМИНИСТРАЦИИ



Бързо търсене и анализ на информация



Анализ на обществени настроения и обратна връзка



Поддръжка при комуникация и кампании



Мониторинг на медиите и социални мрежи



Генериране на идеи и стратегии



Поддръжка при вземане на решения и презентации

МЕТОДОЛОГИЯ ЗА ВНЕДРЯВАНЕ



ПРИНЦИПИ: СИГУРНОСТ | ГЪВКАВОСТ | ОТГОВОРНОСТ | ПРОЗРАЧНОСТ

RISK CANVAS

РИСКОВЕ

- Изтичане на чувствителна информация
- Неправилни или подвеждателни отговори
- Злоупотреба с генерирано съдържание
- Зависимост от външни услуги

ПОСЛЕДСТВИЯ

- Репутационни щети
- Неправилни решения
- Правни санкции
- Финансов загуби



ПРЕВЕНЦИЯ

- Контрол на достъпа и данните
- Глобално тестване
- Оптимизация на политиките за употреба
- Обучение на служителите

МИТИГАЦИЯ

- Наблюдение и отчет на корупция
- Мониторинг на репутацията
- Дълготрайно и редовно при изпитания
- Редовно актуализиране на модели

УКАЗАНИЯ ЗА УПОТРЕБА

- Дефинирайте ясно сценарии и цели
- Използвайте проверени източници
- Проверявайте фактите срещу външни източници
- Следете за промените и неточности
- Осигурете човешки контрол
- Спазвайте законодателството и вътрешните политики



СУВЕРЕНИТЕТ. СИГУРНОСТ. СЛУЖЕНЕ НА ОБЩЕСТВОТО.

MISTRAL AI – ВАШИЯТ НАДЕЖДЕН ПАРТНЪР ЗА ДИГИТАЛНА ТРАНСФОРМАЦИЯ.

ЕВРОПЕЙСКИ СУВЕРЕНИТЕТ



ГОВЕРИТЕЛНОСТ НА ДАННИТЕ



ПРОЗРАЧНОСТ И ОТГОВОРНОСТ



УЧАСТИЕ В ГОЛАЗА НА ОБЩЕСТВОТО



ИНТЕЛИГЕНТНОСТ. СКОРОСТ. ВЪЗДЕЙСТВИЕ.

GROK BY XAI – ЗА АДМИНИСТРАЦИИ, КОИТО ДЕЙСТВАТ БЪРЗО И ОТГОВОРНО

БЪРЗО ДОСТЪП ДО ИНФОРМАЦИЯ



АНАЛИЗ И РАЗБИРАНЕ



ПОДДРЪЖКА ЗА РЕШЕНИЯ



ОТГОВОРНОСТ И ПРОЗРАЧНОСТ

ХЕФЕСТ

ИЗКОВАВАМЕ НАДЕЖДЕН ИЗКУСТВЕН ИНТЕЛЕТ

ЛИДЕРСТВО | ДОВЕРИЕ | ОБЩНОСТ | ДИГИТАЛЕН СУВЕРЕНИТЕТ | СТГОВОРНОСТ

ОЦЕНЯВАМЕ ДНЕС.
ОФОРМЯМЕ УТРЕ.
ЗАЩИТАВАМЕ ЗАЕДНО.



MISTRAL AI

OPEN. EFFICIENT. EUROPEAN.
ENGINEERED FOR IMPACT.

X
GROK
BY XAI

BOLD. CURIOUS. UNRESTRAINED.
BUILT TO UNDERSTAND.

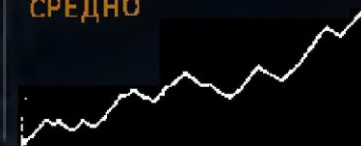
ОЦЕНКА НА РИСК

- Bias
- Safety
- Robustness
- Privacy
- Misuse
- Compliance



НИВО НА РИСК

СРЕДНО



ЕТИКА И УПРАВЛЕНИЕ

- Прозрачност
- Отчетност
- Справедливост
- Човешки надзор



ЗАЩИТА НА
ОБЩНОСТТА

СИЛНИ ЗАЕДНО
БЪДЕЩЕТО Е НАШЕ.



AI MODEL CARD

- Model Purpose
- Strengths
- Limitations
- Training Data
- Use Cases
- Evaluation

SWOT АНАЛИЗ

S СИЛНИ СТРАНИ	W СЛАБОСТИ
O ВЪЗМОЖНОСТИ	T ЗАПЛАХИ

ДИГИТАЛЕН СУВЕРЕНИТЕТ



НЕЗАВИСИМОСТ
СИГУРНОСТ
КОНТРОЛ НА ДАННИТЕ

Natalia



Dian



Deyan



Nikolay



Petar



Manol



Silviya



7 УМА. 1 ЦЕЛ. ПО-СИЛНИ ЗАЕДНО.



АНАЛИЗИРАМЕ



ОЦЕНЯВАМЕ РИСКА



СРАВНЯВАМЕ



ИЗГРАЖДАМЕ ДОВЕРИЕ



ИНСТИТУТ
ПО ПУБЛИЧНА
АДМИНИСТРАЦИЯ



Институт по публична администрация

София, ул. "Сердика 6–8",
тел.: 02 940 2556
имейл: ipa@ipa.government.bg
www.ipa.government.bg