



Финансирано от
Европейския съюз

Следващо Поколение ЕС



План за възстановяване и
устойчивост



Република България

Доклад за направено онлайн проучване за IfGPT – голям набор от езикови данни за български език



Брой участници: 21
28 февруари 2026 г.
Найден Найденов

Резюме

Настоящият доклад представя резултатите от анкетно проучване, проведено сред 21 специалисти в областта на информационните технологии, изкуствения интелект и академичната общност. Целта на проучването е да се оцени нивото на приложението, функционалността, техническото изпълнение и общата зрялост на големия набор от данни IfGPT.

Резултатите показват, че анкетираните дават висока оценка на големия набор данни IfGPT – средна оценка 4.33/5 за качество на езиковите данни и 4.33/5 за документацията върху набора данни и функциите за достъп до него.

Над 80% от участниците смятат, че наборът е подходящ за научни изследвания и образователни цели. Технологичната зрялост се оценява предимно на ниво TRL 6, а мнозинството участници потвърждават готовността на платформата за валидиране в релевантна среда за постигане на ниво TRL 7.

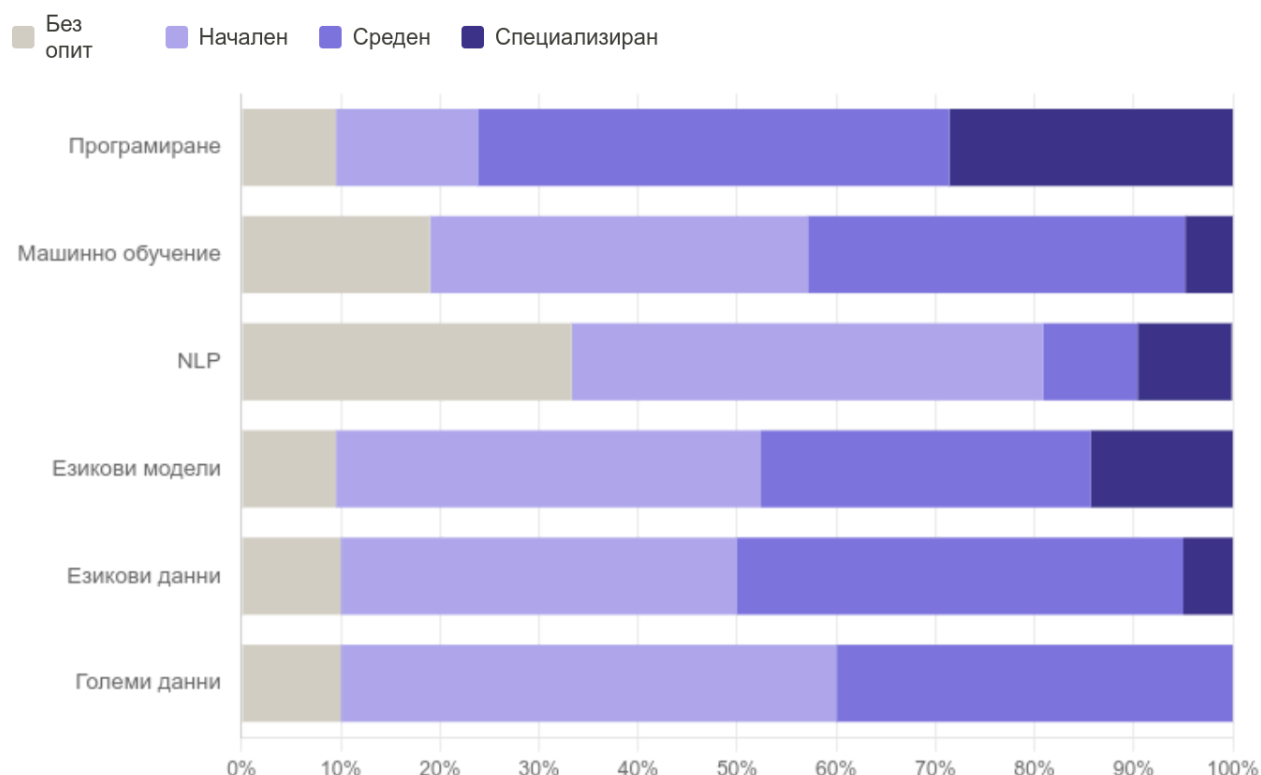
1. Профил на участниците

Проучването е попълнено от 21 участници с разнообразен професионален профил, представляващи основните целеви групи за използване на набора от данни IfGPT.

Професионален профил

Категория	Брой	Дял
Студенти по ИТ / информатика	10	50,0%
Изследователи / преподаватели	8	40,0%
Специалисти в сферата на ИТ	7	35,0%
Специалисти в областта на ИИ и компютърните науки	2	10,0%
Финансист	1	5,0%
Други	1	5,0%

Опит в ключови технологични области



Анализът на нивата на компетентност по основните специализирани области показва, че преобладаващата част от участниците имат среден до специализиран опит поне в някоя от изброените области, най-вече в програмирането, по-малко в езиковите модели, използването на езикови и големи данни и машинното обучение.

2. Познаване на набора от данни IfGPT

Резултатите показват, че над 66% от участниците поне частично познават IfGPT – 38,1% са работили директно с него, а 14,3% са преглеждали документацията. Едва 33,3% не са запознати, което свидетелства за добра видимост на платформата сред целевата аудитория.

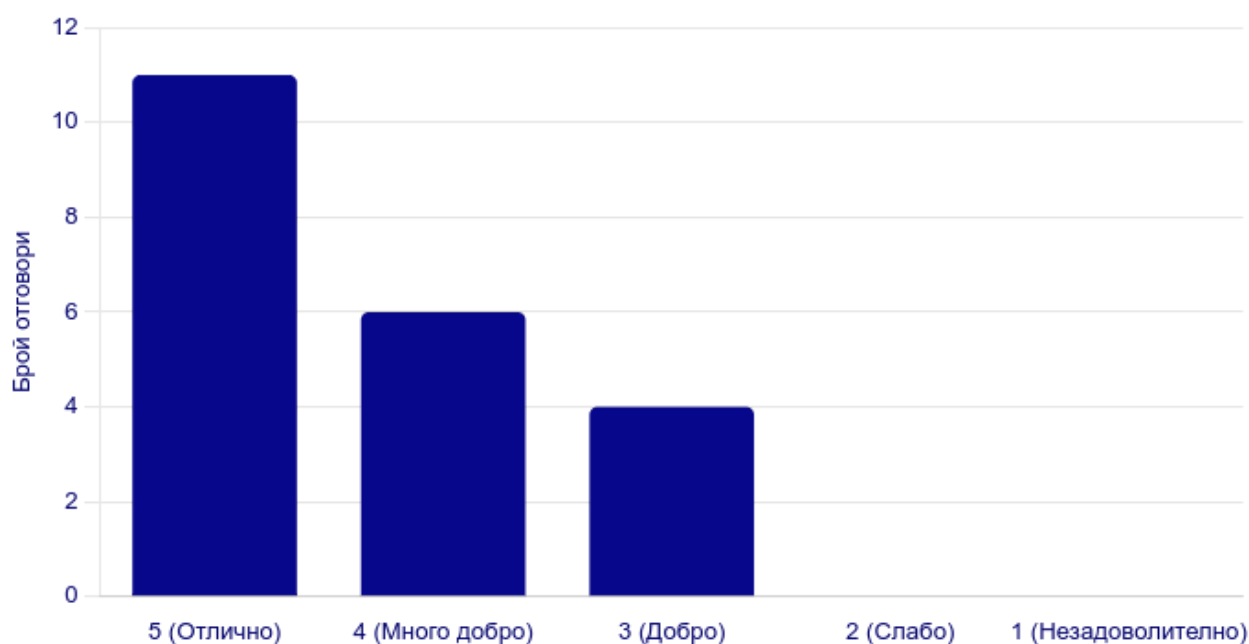
Механизъм на запознаване с набора данни IfGPT

Канал	Брой	Дял
Официалният уебсайт	14	66,7%
Документация	2	9,5%
Научна публикация	2	9,5%
Демонстрация / тест	1	4,8%
Друго	1	4,8%

Официалният уебсайт е доминиращият канал за запознаване с IfGPT – 66,7% от участниците са научили за платформата именно оттам, което подчертава необходимостта от поддържане и актуализиране на страницата. Като метод на запознаване следват документацията (налична на сайта, както и в платформите GitHub и HuggingFace) и научните публикации имат по-ограничен принос (по 9,5%).

3. Оценка на функционалността и качеството

Качество на езиковите данни



Качеството на езиковите данни получава висока оценка от участниците с обща средна оценка 4.33 (от 5). Над 80% са дали положителна оценка 4 или 5, което показва

значителна удовлетвореност. Нито един участник в проучването не е дал оценка под 3, което е особено показателно за надеждността на съдържанието.

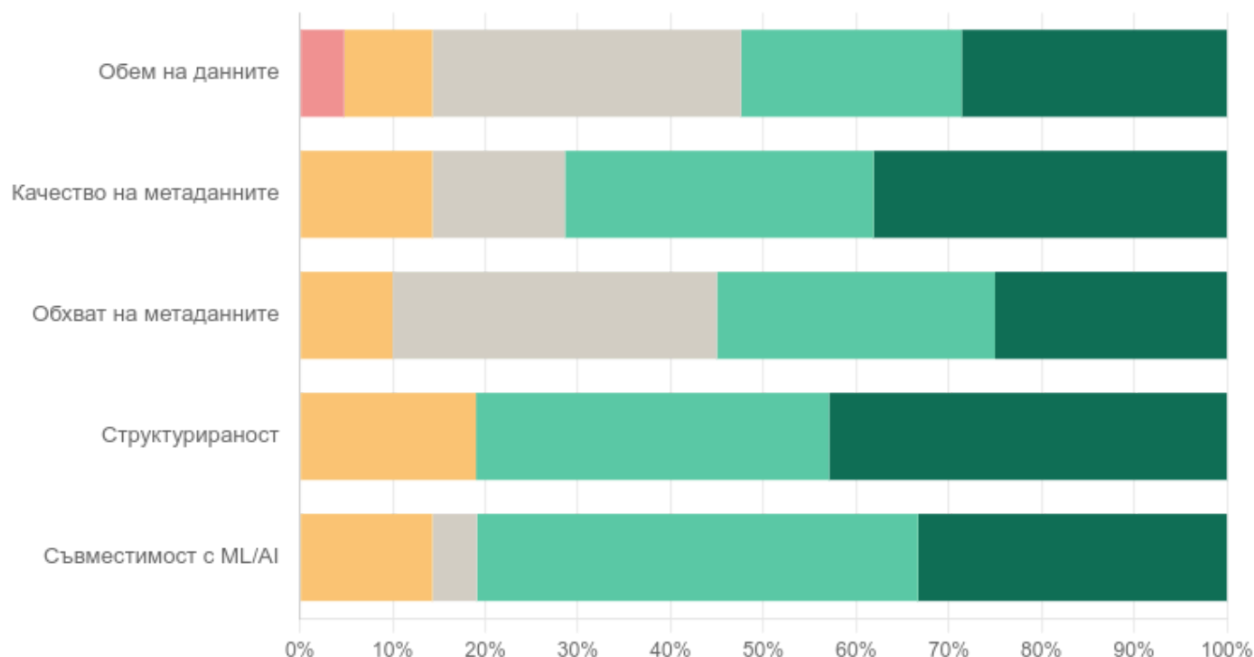
Приложимост на големия набор данни IfGPT

Приложение	Брой	Дял
Образователни цели	19	90,5%
Научни изследвания	18	85,7%
Обучение на езикови модели	10	47,6%
Прототипи и експериментални приложения	8	38,1%
Не са подходящи за научни разработки	1	4,8%

Големият набор от данни IfGPT и предоставеният достъп до него се възприемат като изключително подходящи за образователни цели (90,5%) и научни изследвания (85,7%). Приложимостта за обучение на езикови модели, както и за изработване на прототипи и експериментални приложения, се оценява по-умерено. Само един участник е преценил, че данните не са подходящи за научни разработки.

Оценка на характеристиките

1 (Незадоволително) 2 (Слабо) 3 (Добро) 4 (Много добро) 5 (Отлично)



Ключовите характеристики – обем на данните, качество на метаданните, обхват на метаданните, структурираност, съвместимост с различни приложения, са оценени по петстепенна скала.

- **Обем на данните** – средно 3,62 от 5 (с най-широко разпределение и единствената незадоволителна оценка);
- **Качество на метаданните** – средно 3,95 от 5 (над 70% дават положителна оценка);

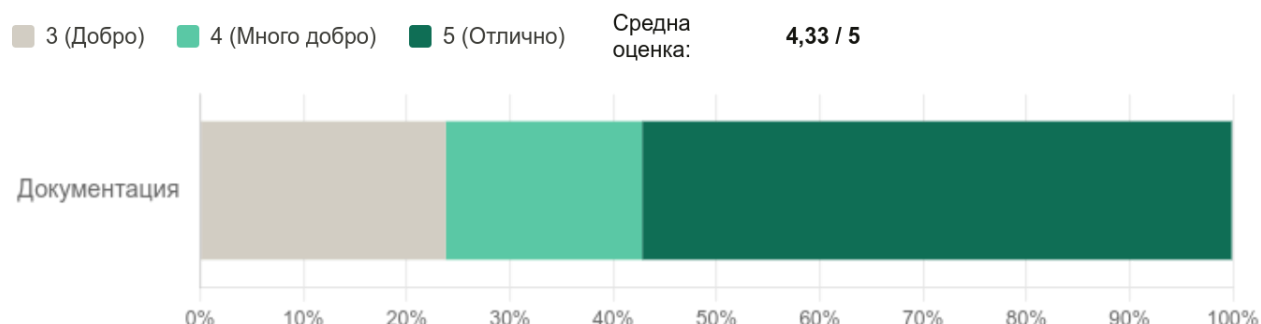
- **Обхват на метаданните** – средно 3,70 от 5 (около половината дават положителна оценка, следвана от голям дял добра оценка);
- **Структурираност** — средно 4,05 от 5 (високо оценена с 80% положителни оценки);
- **Съвместимост с ML/AI** — средно 4,00 от 5 (80% положителни оценки).

Ефективно използване в реална среда

Становище	Брой	Дял
Напълно съгласен	11	52,4%
По-скоро съгласен	6	28,6%
Не мога да преценя	5	23,8%
По-скоро не съм съгласен	0	0,0%
Не съм съгласен	0	0,0%

Над 80% от анкетираните са съгласни, че IfGPT може да се използва ефективно в реална работна среда, а нито един участник не е изразил несъгласие – резултат, потвърждаващ потенциалната практическа приложимост на платформата.

Оценка за качеството на документацията



Близо 60% от участниците дават отлична оценка за документацията, а близо 20% – много добра оценка.

Интеграционен потенциал

Среда за интеграция	Брой	Дял
Системи за обработка на естествен език	15	71,4%
Академични проекти	15	71,4%
Приложения за машинно обучение	10	47,6%
Прототипни приложения	7	33,3%
Трудни за интеграция	2	9,5%
ML pipelines	1	4,8%

IfGPT се смята за лесно интегрируем в системи за обработка на естествен език, както и в академични проекти и приложения, използващи машинно обучение. Близко 10% от участниците обаче не виждат конкретни приложения за интеграция.

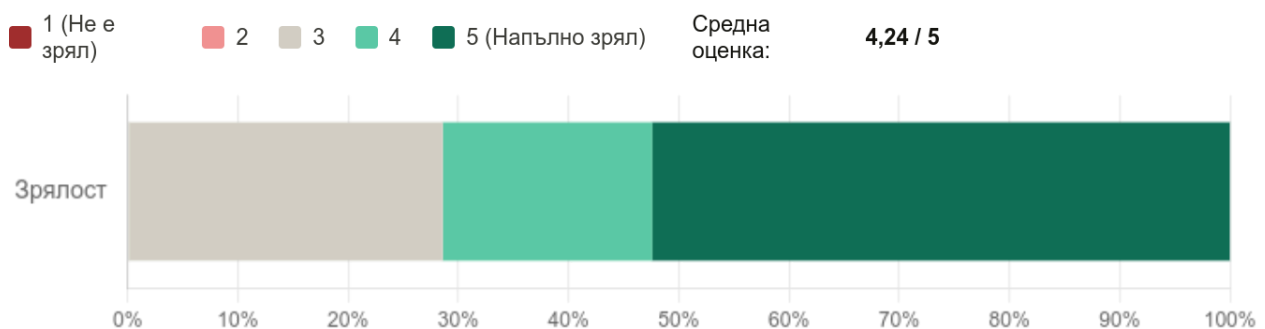
Ниво на технологична готовност (TRL)

Разпределение на отговорите за нивото на технологична готовност (n = 21)



Мнозинството от участниците оценяват, че големият набор от данни IfGPT и достъпът до него съответстват на ниво на технологична готовност TRL 6 – демонстрация в релевантна среда, което потвърждава нейната техническа зрялост и готовност за по-широко внедряване.

Зрялост за валидиране



Повече от половината участници смятат, че IfGPT е достатъчно зрял за валидиране в релевантна среда, а нито един не е изразил категоричен скептицизъм. Комбинацията от TRL оценка и оценката за зрялост подчертава добрата позиция на набора от данни за преминаване към по-практически приложения и TRL 7.

Основни предимства

Предимство	Брой	Дял
------------	------	-----

Достъп до големи данни на български език	17	81,0%
Подбор на данни на български език	16	76,2%
Отворена AI инфраструктура	12	57,1%
Позволява фина настройка на модели	7	33,3%
Изгражда прототипна технология	4	19,0%
Включва оценка и сравнение	3	14,3%
IfGPT спестява време	1	4,8%

Водещото предимство на IfGPT е достъпът до големи текстови данни на български език, следвано от подбора на данните и отворената AI инфраструктура.

Препоръки за подобрение

Предимство	Брой	Дял
Повече данни на разговорен български език	10	47,6%
Повече данни на официален/формален български	9	42,9%
Възможност за търсене по подтеми	8	38,1%
Автоматизирана фина настройка	6	28,6%
Подробна документация и обучителни материали	5	23,8%
По-лесен потребителски интерфейс	4	19,0%
Интеграция с реални системи	3	14,3%
Повече оценъчни тестове	2	9,5%
Поддръжка на повече езици	1	4,8%

Коментари:

- *Проектът изглежда перспективен и полезен за по-добра подкрепа на българския език в AI технологии.*
- *Иновативен и различен продукт.*
- *Полезно.*
- *Като потребителка оценявам продукта като много обещаващ, изглежда стабилен за работа в тестова среда, с ясен потенциал за практическо внедряване.*
- *Платформата е полезна и лесна за използване.*