



Техническа спецификация

Софтуерна платформа за интеграция на три-компонентен изкуствен интелект с механизъм за машинно обучение

I. ПРЕДМЕТ НА ПОРЪЧКАТА

Предмет на публичната покана е „Придобиване на софтуерна платформа за интеграция на три-компонентен изкуствен интелект с механизъм за машинно обучение“.

II. ОБХВАТ И ОПИСАНИЕ НА СОФТУЕРНАТА ПЛАТФОРМА

Поръчката включва доставка, монтаж, инсталация, тестове и пускане в експлоатация на софтуерна платформа за интеграция на три-компонентен изкуствен интелект с механизъм за машинно обучение.

1. Предназначение

Платформата трябва да позволи подаването на потоци от документи (новини и публикации в социални медии) на различни езици към набор от разнообразни, вече разработени семантични софтуери, с цел обогатяване на документи със семантична мета-информация.

Основната цел на придобивания софтуер е да позволи използването в реална производствена среда на софтуерни технологии (три-компонентен изкуствен интелект), разработени от А Дейта Про ООД (АДП). За целта на изпълнителя ще бъде предоставена цялата документация и програмния код на приложенията, разработени от АДП с оглед интегрирането му като основа в софтуерната платформа.

Предвижда се софтуерната платформа да изпълнява четири ключови групи дейности:

1. Софтуерно интегриране на трите компонента, разработени от АДП, така че да работят едновременно/паралелно върху един и същи набор от данни, като обичайно се предвижда използването на повече от един набор от данни едновременно
2. Софтуерно интегриране на трите компонента (разпознаване на именувани обекти, класификация и клъстеризиране на материали) на Serendipity и автоматизация на съответните процеси, така че да е възможно тяхното използване през наличната в дружеството система за управление на съдържание (CMS).



3. Софтуерно интегриране на трите компонента по начин, който да позволява системи за медиен мониторинг на трети страни да комуникират с платформата и да извличат обработени документи, например чрез директно получаване на данни от клиенти или други доставчици. Платформата трябва да притежава приложнопрограмен интерфейс (Application Program Interface), който да позволи пълноценно използване на компонентите.
4. Осигуряване на функционалност за подаване на директна обратна връзка за осъществяване на машинно обучение на интелекта Serendipity от страна на медия анализатори в компанията (работещи през CMS) или от страна на междинни или крайни клиенти на услугата.

2. Изисквания към функционалностите

Минимални изисквания:

Платформата трябва да разполага със следните функционалности и инструменти:

1. Дефиниране на множество потоци от документи.

Платформата трябва да позволява дефинирането на поток от информация и подаването му към релевантни семантични услуги;

Разяснения:

За постигане на необходимата функционалност - дефиниране на множество потоци от документи - платформата трябва да разполага с механизъм за извличане на данни от различни източници на информация в структуриран вид. След извличане на данните, механизмът следва да ги реструктурира и да ги подава към предварително дефиниран семантичен пайплайн (опашка от задачи) в предварително определени момент, формат и структура и по начин, съгласуван с протокол за комуникация със семантичния пайплайн. Чрез този механизъм, платформата трябва да може да поддържа неограничен брой потоци от документи. Потоците от документи ще бъдат постове в твитър, фейсбук постове, статии в традиционни медии, постове в блогове и форуми, таблици и други набори от данни.

Тези потоци от информация ще бъдат подавани към системата в хетерогенен вид. За да бъдат обработвани, следва да бъдат трансформирани до JSON HTTP заявки, съобразени със специфичен JSON формат ползван от трите компонента на изкуствения интелект.

2. Интеграция с услуга за разпознаване на именувани обекти.

Платформата трябва да интегрира вече разработен от АДП софтуер за Разпознаване на именувани обекти (Named Entity Recognition).

Разяснения: Услугата за разпознаване на именувани обекти работи с входящи документи в JSON формат. След обработването им те се подават на JSON парсър, който връща обработения обект. Резултатите следва да се трансформират в RDF и да бъдат заредени в



ЕВРОПЕЙСКИ СЪЮЗ
ЕВРОПЕЙСКИ ФОНД ЗА
РЕГИОНАЛНО РАЗВИТИЕ



„Внедряване на изкуствен интелект "Serendipity" за фундаментална промяна в процеса по предоставяне на услуги за медиа мониторинг и анализи на А Дейта Про" – Договор BG16RFOP002-1.001-0522-C01

А DATA PRO

RDF база данни. Разработваното веб приложение, следователно, следва да трансформира JSON в RDF и обратно.

3. Интеграция с услуга за класифициране на документи.

Платформата трябва да интегрира вече разработен от АДП софтуер за Класифициране на информационни обекти по предварително зададена методология за класифициране, например таксономия или комбинация от различни таксономии;

Разяснения: Услугата за класифициране на документи, подобно на горната (т.2) работи с входящи документи в JSON формат. След обработването им те се подават на JSON парсър, който връща обработения обект. Резултатите следва да се трансформират в RDF и да бъдат заредени в RDF база данни. Разработваното веб приложение, следователно, следва да трансформира JSON в RDF и обратно.

4. Интеграция с клъстеризатор на документи.

Платформата трябва да интегрира вече разработен от АДП софтуер за автоматично клъстеризиране по сходство на набор от документи.

Разяснения: Клъстеризаторът на документи, подобно на т. 2 и 3 по-горе, работи с входящи документи в JSON формат. След обработването им те се подават на JSON парсър, който връща обработения обект. Резултатите следва да се трансформират в RDF и да бъдат заредени в RDF база данни. Разработваното веб приложение, следователно, следва да трансформира JSON в RDF и обратно.

5. Поддръжка на машинно обучение.

Платформата трябва да позволява семантичните софтуери да бъдат обучавани от потребителите на системата. Обучението се извършва, чрез подаване на обработени от хора документи (наричани "златен стандарт");

Разяснения:

Документите, с които семантичните софтуери се захранват, се подават чрез вече съществуващ модул в собствената CMS на възложителя (модул за генериране и обработване на хетерогенни данни).

Системата трябва да разполага с механизъм за управление на процесите на обучение на семантичните услуги. По-конкретно системата трябва да може да извлича специфични извадки от базата данни посредством филтри, определени ръчно през веб форма. Извлечените данни се превръщат във входящи данни за всеки тренировъчен цикъл. Веб приложението, разработвано за изпълнение на този процес, следователно трябва да разполага с механизъм, за подаване на заявки към CMS-а, в който се съхраняват данните от една страна и към семантичните услуги за обучаване на модели, от друга. Веб приложението следва да поддържа бек-ъп и версии на обработваните масиви от



ЕВРОПЕЙСКИ СЪЮЗ
ЕВРОПЕЙСКИ ФОНД ЗА
РЕГИОНАЛНО РАЗВИТИЕ

„Внедряване на изкуствен интелект "Serendipity" за фундаментална промяна в процеса по предоставяне на услуги за медия мониторинг и анализи на А Дейта Про" – Договор BG16RFOP002-1.001-0522-C01



ОПЕРАТИВНА ПРОГРАМА
ИНОВАЦИИ И
КОНКУРЕНТОСПОСОБНОСТ
A DATA PRO

документи, да поддържа специфични мета данни за всеки конкретен тренировъчен модел (като верси, дата, брой документи, език, тип медия и др.).

Това е необходимо с цел анализ и контрол на качеството на обучени, автоматична обработка и качество и производителност при ръчна обработка от анотатори.

6. Специфично машинно обучение за всеки отделен поток от информация.

Платформата трябва да позволява различно обучение на семантичните софтуери за всеки отделен поток от информация;

Разяснения: Поток от информация е масив от данни с информация за специфичен домейн, индустрия, тема, език. Целта е системата да позволява стартиране на индивидуално обучение на семантичните услуги (machine learning), в зависимост от спецификите на всеки поток от информация до получаване на обучени машинни модели. Така обучаваните или обучени машинни модели следва да могат да бъдат извиквани от анотаторите при следващи процеси или цикли на аотиране и обработване на нови документи и масиви от документи от същия домейн. Уеб приложението следва да може да подава заявки към специфична „услуга за обучение“ на семантичните услуги, разработена от възложителя и да извлича и подава данни към общата база данни.

7. Съхранение на обработените потоци от документи.

Платформата трябва да позволява съхранението на вече обработените от семантични софтуери документи в база от данни и във формат, който да позволи визуализацията на документите за нуждите на медийния мониторинг и анализ;

Разяснения: Резултатите от процесите по обработване на данните ще се съхраняват в RDF база данни. Платформата следва да трансформира входящите документи и резултати от JSON формат до RDF формат. С цел коректно форматиране и структуриране на базата тя следва да ползва предварително зададени онтологии, разработени от възложителя.

8. Комуникация със системи за медиен мониторинг и анализ

Платформата трябва да притежава приложно-програмен интерфейс (Application Program Interface), който да позволи системи за медиен мониторинг на трети страни да комуникират с платформата и да извличат обработени документи.

Платформата трябва да има потребителски интерфейс оборудван с контролен панел, който да позволява на клиенти и служители на компанията да анализират обработените документи.

9. Внасяне на данни, обработени от трикомпонентния изкуствен интелект

Платформата трябва да позволи внасянето на данни, посредством приложно програмния интерфейс (API) на платформата.

----- www.eufunds.bg -----

„Внедряване на изкуствен интелект "Serendipity" за фундаментална промяна в процеса по предоставяне на услуги за медия мониторинг и анализи на А Дейта Про" – договор BG16RFOP002-1.001-0522-C01, финансиран от Оперативна програма „Иновации и конкурентоспособност“, съфинансирана от Европейския съюз чрез Европейския фонд за регионално развитие.



ЕВРОПЕЙСКИ СЪЮЗ
ЕВРОПЕЙСКИ ФОНД ЗА
РЕГИОНАЛНО РАЗВИТИЕ

„Внедряване на изкуствен интелект "Serendipity" за фундаментална промяна в процеса по предоставяне на услуги за медия мониторинг и анализи на А Дейта Про" – Договор BG16RFOP002-1.001-0522-C01



ОПЕРАТИВНА ПРОГРАМА
ИНОВАЦИИ И
КОНКУРЕНТОСПОСОБНОСТ
A DATA PRO

Разяснения: Обработените данни ще се съхраняват в RDF база данни. Платформата следва да трансформира изходящите документи, документи и резултати от формат JSON до RDF формат. С цел коректно форматиране и структуриране на базата тя следва да ползва предварително зададени онтологии, разработени от възложителя.

За да е възможно внасяне на вече обработени данни от трикомпонентния изкуствен интелект, системата следва да разполага с уеб форма, която да може да специфицира и филтрира типовете информация, която следва да бъде извлечена. Тъй като всеки компонент работи на собствен сървър, уеб приложението следва да се обръща към семантичните услуги с JSON заявки и да може да представя резултатите от тях чрез модул в съществуващата CMS система на възложителя.

10. Индекс на обработените документи.

Платформата трябва да разполага с индексиращ софтуер, който да позволи различни сечения на обработените документи;

Разяснения: Системата трябва да разполага с механизъм за индексиране на материалите, така че да позволява бързо търсене, вкл. чрез низове за търсене на различни езици. Платформата трябва да поддържа собствен индекс, който да гарантира бързо и ефективно търсене в масиви данни и милиони записи, вкл. чрез използване на логически оператори.

11. Търсене в обработените документи.

Платформата трябва да позволи търсене в индексирания документи по ключова дума, логически оператори и период;

Разяснения: Индексирания документи могат да бъдат на всички официални езици. Броят записи се очаква да нараства значително ежедневно. Търсенето трябва да може да се извършва по ключова дума, набор от ключови думи, нозиве за търсене с приложение на Boolean и други логически оператори. Платформата трябва да връща резултат в рамките на 2 секунди.

12. Визуализация на обработените документи в Графичен Потребителски Интерфейс.

Индекса на Обработените документи трябва да може да бъде визуализиран в графичен потребителски интерфейс, който да позволява анализ на данните. Анализът се осъществява въз основа на различен набор от графики и списък с релевантни документи.

Разяснения: Данните следва да се визуализират в графичен интерфейс, който визуализация на статистически данни за обработван/обработен масив от данни, достъп до мазива от данни. Визуализацията на аналитичните данни се изпълнява чрез представяне на графики и таблици по показатели, критерии и метрики, посочени от



ЕВРОПЕЙСКИ СЪЮЗ
ЕВРОПЕЙСКИ ФОНД ЗА
РЕГИОНАЛНО РАЗВИТИЕ

„Внедряване на изкуствен интелект "Serendipity" за фундаментална промяна в процеса по предоставяне на услуги за медия мониторинг и анализи на А Дейта Про" – Договор BG16RFOP002-1.001-0522-C01



ОПЕРАТИВНА ПРОГРАМА
ИНОВАЦИИ И
КОНКУРЕНТОСПОСОБНОСТ
A DATA PRO

вътрешни за възложителя (анализатори и анотатори) или външни за възложителя (клиенти) потребители.

13. Поддръжка на множество потоци от обогатени документи.

Платформата трябва да позволява визуализацията на различни потоци от обработени документи.

Разяснения: Платформата ще бъде използвана за обработване и анализ на хетерогенни данни по отношение на формат, произход и език. Платформата трябва да позволява извличане на документи с различен произход, език и формат, създаване на потоци - масиви от данни от хетерогенни данни (постове в твитър, фейсбук и други социални медии, статии от преса и традиционни медии онлайн, както и транскрибирани материали от електронни медии (радио и телевизия) в един домейн. Потоците от документи трябва да се представят в потребителски интерфейс (в един формат), който позволява ръчна анотация или представя анотации, изпълнени от машинно обучените модели.

14. Обучение въз основа на обратна връзка от клиенти

Платформата трябва да позволява клиентите да дават обратна връзка, дали документите са обработени правилно. Обратната връзка трябва да може да бъде ползвана за обучения на семантичните услуги описани в 2, 3 и 4.

Разяснения: Платформата следва да позволява подаване на обратна връзка за качеството на изпълнените анализи през потребителски интерфейс. Това следва да става, като вече приложените кодове се променят ръчно от външни или вътрешни за възложителя потребители. Така подадената обратна връзка следва да се подава то разработените изкуствени интелекти за допълнително машинно обучение.

15. Контрол на достъп

Платформата трябва да притежава гъвкав контрол на достъп до интерфейса за анализ на всеки отделен поток от обработени документи;

Разяснения: За постигане на гъвкавост в контрол на достъпа, софтуерната платформа трябва да е интегрирана със системата за управление на съдържание на А Дейта Про и механизмите за определяне на роли с различни нива на достъп и права на служители на компанията. Системата трябва да позволява назначаване на различни роли на едно и също лице при изпълнението на различни проекти или задачи. За тази цел платформата следва да е интегрирана с модул за автентификация, част от CMS системата на възложителя.

16. Пощенски бюлетини

Платформата трябва да поддържа изпращане на пощенски бюлетини (newsletters) до набори от потребители. Пощенските бюлетини трябва да се изготвят автоматично въз

----- www.eufunds.bg -----

„Внедряване на изкуствен интелект "Serendipity" за фундаментална промяна в процеса по предоставяне на услуги за медия мониторинг и анализи на А Дейта Про" – договор BG16RFOP002-1.001-0522-C01, финансиран от Оперативна програма „Иновации и конкурентоспособност“, съфинансирана от Европейския съюз чрез Европейския фонд за регионално развитие.



ЕВРОПЕЙСКИ СЪЮЗ
ЕВРОПЕЙСКИ ФОНД ЗА
РЕГИОНАЛНО РАЗВИТИЕ

„Внедряване на изкуствен интелект "Serendipity" за фундаментална промяна в процеса по предоставяне на услуги за медия мониторинг и анализи на А Дейта Про" – Договор BG16RFOP002-1.001-0522-C01



ОПЕРАТИВНА ПРОГРАМА
ИНОВАЦИИ И
КОНКУРЕНТОСПОСОБНОСТ
A DATA PRO

основа на вече обработената информация, да може да им се задава честота, както и конкретен час, в който да бъдат изпращани. Пощенските бюлетини трябва да бъдат достъпни през основните имейл клиенти.

3. Изисквания към интелектуалната собственост

Авторските права върху системата, в едно с целия изходен код (source code), следва да бъдат изцяло прехвърлени на „А Дейта Про“ ООД.

III. ГАРАНЦИОННИ УСЛОВИЯ

Всички компоненти и системи следва да са обект на гаранционно обслужване, придружени със съответните удостоверяващи документи.

Задължително е гаранционното обслужване да включва обслужване на място и да е с гаранционен срок не по-кратък от 36 месеца считано от датата на подписване на приемо-предавателния протокол след доставка, изпитване и въвеждане в експлоатация.

Гаранционната поддръжка задължително включва осигуряване на безпроблемна работа с вътрешните системи на „А Дейта Про“ ООД, с които платформата комуникира, поддръжка на неограничен брой потребители на платформата и неограничен брой процеси, както и най-малко 20 часа месечно безплатен съпорт от страна на кандидата.

Доставчикът трябва да посочи време (в цели часове) за реакция и диагностика на възникнали проблеми и/или дефекти след подаден сигнал от представител на А Дейта Про ООД в рамките на гаранционния срок.

IV. УСЛОВИЯ ЗА ДОСТАВКА, ИНСТАЛАЦИЯ И ПЪРВОНАЧАЛНО ИЗПИТВАНЕ

Доставката и инсталацията на платформата трябва да се извършат в рамките на 8 месеца след сключване на договор.

Доставчикът следва да инсталира и свърже компонентите на отделните системи в предоставено от фирмата-възложител помещение в гр. Пловдив.

Доставчикът следва да проведе първоначално изпитване в продължение на 7 работни дни и да предаде протокол с получените резултати.

Доставчикът трябва да предостави потребителско ръководство и техническа спецификация на софтуерната платформа на електронен или хартиен носител, на български или на английски език.

----- www.eufunds.bg -----

„Внедряване на изкуствен интелект "Serendipity" за фундаментална промяна в процеса по предоставяне на услуги за медия мониторинг и анализи на А Дейта Про" – договор BG16RFOP002-1.001-0522-C01, финансиран от Оперативна програма „Иновации и конкурентоспособност“, съфинансирана от Европейския съюз чрез Европейския фонд за регионално развитие.



ЕВРОПЕЙСКИ СЪЮЗ
ЕВРОПЕЙСКИ ФОНД ЗА
РЕГИОНАЛНО РАЗВИТИЕ

„Внедряване на изкуствен интелект "Serendipity" за фундаментална промяна в процеса по предоставяне на услуги за медия мониторинг и анализи на А Дейта Про" – Договор BG16RFOP002-1.001-0522-C01



ОПЕРАТИВНА ПРОГРАМА
ИНОВАЦИИ И
КОНКУРЕНТОСПОСОБНОСТ
A DATA PRO

Доставчикът следва да проведе най-малко еднодневно обучение за работа с платформата на експерти, определени от възложителя.

V. ЗАДЪЛЖИТЕЛНО ИЗИСКУЕМИ ДОКУМЕНТИ И ИНФОРМАЦИЯ

В техническото си предложение кандидатът задължително следва да предостави най-малко следната информация:

1. Срок за изпълнение.
2. Описание на изпълнението на всяко от минималните изисквания за платформата, посочени в т. II. ОБХВАТ И ОПИСАНИЕ НА ПЛАТФОРМАТА.
3. Да потвърди прехвърляне на правата върху интелектуална собственост, съгласно изискванията на т. III Гаранционни условия.
4. Срок за гаранционна поддръжка. Минимален срок на гаранционна поддръжка – 36 месеца. Срокът за гаранционна поддръжка задължително се посочва в цели месеци.
5. Екип за изпълнение на дейностите – брой и профил на експертите, които ще бъдат включени в екипа за изпълнение на договора в съответствие с изискванията на т. III.2.3. Технически възможности и квалификация от Публичната покана.
6. План график за изпълнение на доставката с детайлно разписани дейности, срокове и предвидени човекочасове за изпълнение на всяка задача за всеки експерт или функция (ръководител проект, старши програмист, програмист, експерт информационна сигурност), включен в екипа за изпълнение на доставката, съгласно изискванията на т. III.2.3 Технически възможности и квалификация от Публичната покана и общ брой човекочасове – включително за всеки от минимално изискуемите елементи и функционалности (механизми, системи, интерфейси).
7. Предложеният от него брой човекочасове за извършване на допълнителни актуализации на разработената платформа в рамките на предложението от кандидата гаранционен срок, включени в офериранията цена, извън основната разработка на платформата, съгласно изискванията посочени в този документ.
8. Доставчикът трябва да посочи време (в цели часове) за реакция и диагностика на възникнали проблеми и/или дефекти след подаден сигнал от представител на А Дейта Про ООД в рамките на гаранционния срок.

VI. ОЧАКВАНИ РЕЗУЛТАТИ

1. Изработена 1 бр. платформа и всички прилежащи слоеве, механизми, системи и интерфейси;

----- www.eufunds.bg -----

„Внедряване на изкуствен интелект "Serendipity" за фундаментална промяна в процеса по предоставяне на услуги за медия мониторинг и анализи на А Дейта Про" – договор BG16RFOP002-1.001-0522-C01, финансиран от Оперативна програма „Иновации и конкурентоспособност“, съфинансирана от Европейския съюз чрез Европейския фонд за регионално развитие.



ЕВРОПЕЙСКИ СЪЮЗ
ЕВРОПЕЙСКИ ФОНД ЗА
РЕГИОНАЛНО РАЗВИТИЕ



„Внедряване на изкуствен интелект "Serendipity" за фундаментална промяна в процеса по предоставяне на услуги за медиа мониторинг и анализи на А Дейта Про" – Договор BG16RFOP002-1.001-0522-C01

А DATA PRO

2. Проведени 7 (седем) дневни тестове и изпитания и приета с финален премно-предавателен протокол платформа.
3. Разработено и предоставено потребителско ръководство за експлоатация на платформата.
4. Проведено еднодневно обучение от изпълнителя на екип специалисти на възложителя за работа с платформата.

----- www.eufunds.bg -----

„Внедряване на изкуствен интелект "Serendipity" за фундаментална промяна в процеса по предоставяне на услуги за медиа мониторинг и анализи на А Дейта Про" – договор BG16RFOP002-1.001-0522-C01, финансиран от Оперативна програма „Иновации и конкурентоспособност“, съфинансирана от Европейския съюз чрез Европейския фонд за регионално развитие.