

ГЕНЕРАТИВЕН ИЗКУСТВЕН ИНТЕЛЕКТ В ПОМОЩ НА ЗЕМЕДЕЛСКИ СТОПАНИ

Станимир Стоянов, Венета Табакова-Комсалова,
Любка Дуковска, Милена Кюмюрджиева

Резюме. Тази статия представя разработването на съветваща система, реализирана като услуга и използваща възможностите на генеративния изкуствен интелект и по-специално на големия езиков модел (LLM). Услугата ще прави оценка на рисковете за земеделските производители при изготвяне на проекти по различни програми, съобразявайки се с българското законодателство, свързано със земеделието, както и изискванията на съответната програма. Резултатите от анализа на осъществимостта са обобщени в статията. Освен това се обсъждат възможни архитектури на услугата. Услугата ще бъде предназначена основно за подпомагане на малки и средни земеделски стопани.

Ключови думи: генеративен изкуствен интелект, голям езиков модел, интелигентно земеделие, съветстващи системи.

Въведение

Напоследък, като основна движеща сила на дигиталната трансформация възниква генеративният изкуствен интелект (GAI). Централна концепция за GAI са големите езикови модели (LLM) [1], които направиха революция в областта на обработката на естествени езици. Тези модели, снабдени с милиарди параметри и обучени с огромни публични набори от данни, представляват значителен скок в стремежа ни да накараме машините да разбират и генерират човешки език. LLM се интегрират все повече в различни приложения, подобрявайки тяхната интелигентност. Интелигентното земеделие също може да бъде интересна и много полезна област на приложение, която може да се възползва от LLM.

Платформата ЗЕМЕЛА [2] е предназначена да подпомага предимно малки и средни фермери, работещи в условията на интелигентно земеделие. Платформата има регионален характер, т.е. стреми се да отчита в максимална степен характерните особености на конкретния земеделски район. Тази статия представя идеята за разработване на консултантска

услуга, използваща възможностите на генеративния изкуствен интелект и по-специално на LLM. Услугата ще прави оценка на рисковете за земеделските стопани при изготвяне на проекти по различни програми, съобразявайки се с българското законодателство, свързано със земеделието, както и изискванията на съответната програма. Освен това възнамеряваме да интегрираме услугата в ЗЕМЕЛА. По този начин се надяваме да усилим и усъвършенстваме оперативните възможности на платформата. Подобна интеграция е сериозно предизвикателство, за справяне с което е необходим задълбочен анализ, както на възможностите на LLM и текущото състояние на платформата ЗЕМЕЛА, така и на българското законодателство и нормативна уредба, касаеща земеделието. Статията обобщава първите резултати от този анализ. Освен това е представена общата архитектура на консултантската услуга.

Какво представляват големите езикови модели

LLM, по силата на своя дизайн и размер, притежават необичайна способност да схващат нюансите, идиомите и сложността на човешкия език. Архитектурата им, базирана основно на модела Transformer [3], им позволява да управляват различни зависимости в текстови данни. Така LLM могат да разберат контекста върху по-дълги части от текст, което прави техните резултати последователни и контекстуално подходящи. Силата на LLMs не е съобразена само с разбирането. Те са генеративни по природа, способни да произвеждат съдържание, което варира от отговори на сложни въпроси до изработване на есета, създаване на поезия и дори генериране на код. Тази гъвкавост се дължи на обучението им с огромни и разнообразни данни, включително книги, статии, уебсайтове и др. Освен това те са способни да се настройват фино за конкретни задачи, което ги прави безценни активи в области като обслужване на клиенти, изследвания, развлечения и образование, поддържайки целостта на спецификациите. Въпреки това, използването на LLMs не е без предизвикателства. Техният огромен размер изисква значителни изчислителни ресурси, което поражда опасения относно достъпността и въздействието върху околната среда. Освен това LLM понякога могат да произвеждат резултати, които са неправилни или контекстуално неподходящи. Това, че са обучени с огромни и нефилтрирани набори от данни, означава, че те могат по невнимание да показват пристрастия (да халюцинират) или да разпространяват дезинформация.

Фината настройка е подобна на специализирано обучение след широкообхватно образование. Точно както завършил медицина, след придобиване на фундаментални познания по медицина, може да специализира в кардиология или неврология, LLM моделът, постулирайки своето основно обучение, може да бъде фино настроен за конкретни задачи или набори от данни за да постигне експертно представяне в определени области. Това включва вземане на предварително обучен модел и продължаване на процеса на обучение с по-тесен набор от данни, подходящ за целевите задачи. Много приложения на изкуствения интелект изискват разбиране на специфичен жаргон, терминология или контекст. Например, моделът за анализ на правни документи трябва да схване правната терминология, която може да не присъства достатъчно в основните данни. Фината настройка на правните документи гарантира, че моделът е опитен в тази конкретна задача. Въпреки това, фината настройка не е без предизвикателства. Човек трябва да гарантира, че моделът не „забравя“ своите фундаментални знания, феномен, наречен „катастрофално забравяне“. Съществува и риск от пренастройване, ако наборът от данни за фина настройка е твърде малък, което кара моделът да работи лошо върху данни от реалния свят.

В исторически план фината настройка е била основният метод за специализирани модели. Чрез продължаване на обучението на предварително обучен LLM върху конкретни данни, моделът се адаптира, прецизирайки своите тегла и отклонения, за да стане по-компетентен в определени области или задачи. Фината настройка обаче изисква допълнителни набори от данни, изчислителни ресурси и опит. Подсказващото инженерство (prompt engineering) е нюансирано изкуство за взаимодействие с LLMs. Вместо да се задълбочава в трудната задача на преквалификацията, подсказващото инженерство се фокусира върху изработването и прецизиране на входните подкани, за да предизвика желаните резултати. Става дума за разбиране и използване на съществуващите знания и поведение на модела.

Интеграция на LLM в ЗЕМЕЛА

В много публикации (напр. [4]), се поставя въпросът за търсене на възможности собствени хранилища данни да бъдат взети предвид от LLM реализации. Платформата ЗЕМЕЛА представя цялостно киберфизическо-социално пространство, основано на агентно-ориентиран ди-

зайн, онтологии и специализирани подпространства. Смятаме, че съществува значителен потенциал за нейното подобрене на всеки отделен неин компонент чрез интеграция с LLM. Резултатите от досегашния анализ могат да бъдат обобщени както следва:

- Подобро управление на знанията в платформата – фокусът на платформата да съхранява и обработва огромни количества специализирани селскостопански данни и знания я прави идеален кандидат за интегриране на LLM. LLM може да подпомогне семантичното организиране на неструктурирани селскостопански данни, като ги прави по-достъпни и разбираеми, както и подпомагане разработването на онтологии чрез предлагане на възможни взаимоотношения и йерархии въз основа на огромно предварително съществуващо знание в модела.
- Поддръжка за разширено моделиране за целите на откриване и идентифициране на аномалии – LLM може да окаже съдействие при формулиране и интерпретация на възможни отклонения, осигурявайки обяснения на естествен език на сложни процеси за тяхното възникване.
- Симулация и анализ на специфични сценарии – LLM може да се използва за генериране на сценарии, симулиращи различни селскостопански ситуации въз основа на исторически и прогнозиранни данни.
- Усъвършенстване на персоналните асистенти – LLMs могат да играят ключова роля в усъвършенстването на взаимодействията между земеделските оператори и техните съответни асистенти, посредством въвеждане на интерфейс на естествен език, позволяващ по-интуитивно взаимодействие за потребителите.
- Увеличаване на подпространството на гардовете – ролята на гардовете, поддържащи взаимодействието между виртуалния и физическия светове може да бъде подсилена с LLM чрез обработка на естествен език на сензорни данни, преобразуване на редови данни в разбираеми прозрения.

Интегрирането на LLMs в платформата ЗЕМЕЛА не само ще разшири нейните съществуващи възможности, но също така ще въведе множество нови функционалности, особено в областта на интерпретацията на данни, взаимодействието с потребителите и подкрепата за вземане на

решения.

Използване на собствени знания и онтологии с LLM

ЗЕМЕЛА разполага с различни онтологии [5], разработени през последните няколко години. Използването на знания, представени в онтологии, позволява по-добро привеждане в съответствие между структурираните знания и огромното неструктурирано знание, което притежават LLM. Използването на собствените онтологии с LLM може да бъде осъществено в следните насоки:

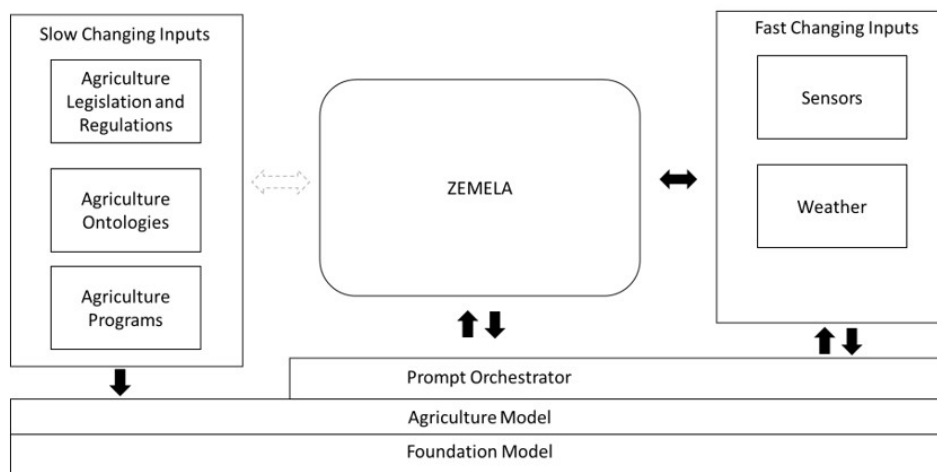
- Представяне на знания в онтологии – допълнително към стандартните формати може да се избира подходящата детайлност за необходимото представяне на знания.
- Извличане на знания за LLM обучение – онтологичните данни могат да бъдат преобразувани на естествен език или да бъдат комбинирани с данни от различни източници на данни.
- Фина настройка на LLM – проектиране на подкани, които са в съответствие със структурата и съдържанието на използваната онтология, което ще помогне при насочването на LLM да произвежда желаните резултати.
- Запитване и взаимодействия – след обучението може да използваме подсказващо инженерство за взаимодействие с LLMs по начин, който извлича или използва специфичните знания от интересувашата ни онтология.
- Непрекъснато обучение – итеративно усъвършенстване с добавяне на повече данни към онтологията или с развитието на знанията за домейна, с итеративно прецизирана и обучена LLM.

Чрез интегриране на онтологии с LLMs, ние по същество свързваме структурирани и неструктурирани знания. Като член на Европейския съюз, България следва национални и европейски разпоредби, свързани със селското стопанство. За да интегрираме правната рамка в ЗЕМЕЛА ще трябва да вземем предвид също национални закони и актове за земеделието, регламенти на Европейския съюз за общата селскостопанска политика, регламенти на държавни институции, международни споразумения.

Архитектура на услугата

Въз основа на резултатите от анализа, представени по-горе, предлагаме два възможни подхода за създаване на архитектура на услугата. Първият подход е да се разширява ЗЕМЕЛА към екосистема, включваща следните компоненти (Фигура 1):

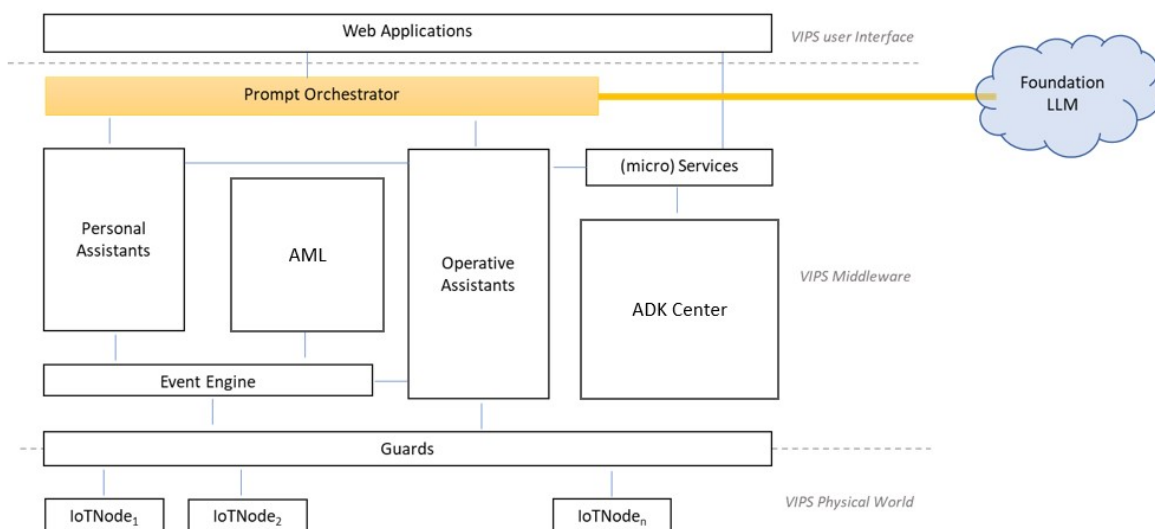
- Основен модел – този компонент ще бъде внедрен като предварително обучен общ LLM.
- Модел на селското стопанство – този компонент ще бъде добре обучен основен модел от специализирани текстове за селското стопанство.
- Prompt Orchestrator – поддържа подготовката на подходящи подкани за правилно предаване на потребителски заявки.



Фигура 1. Екосистема ЗЕМЕЛА

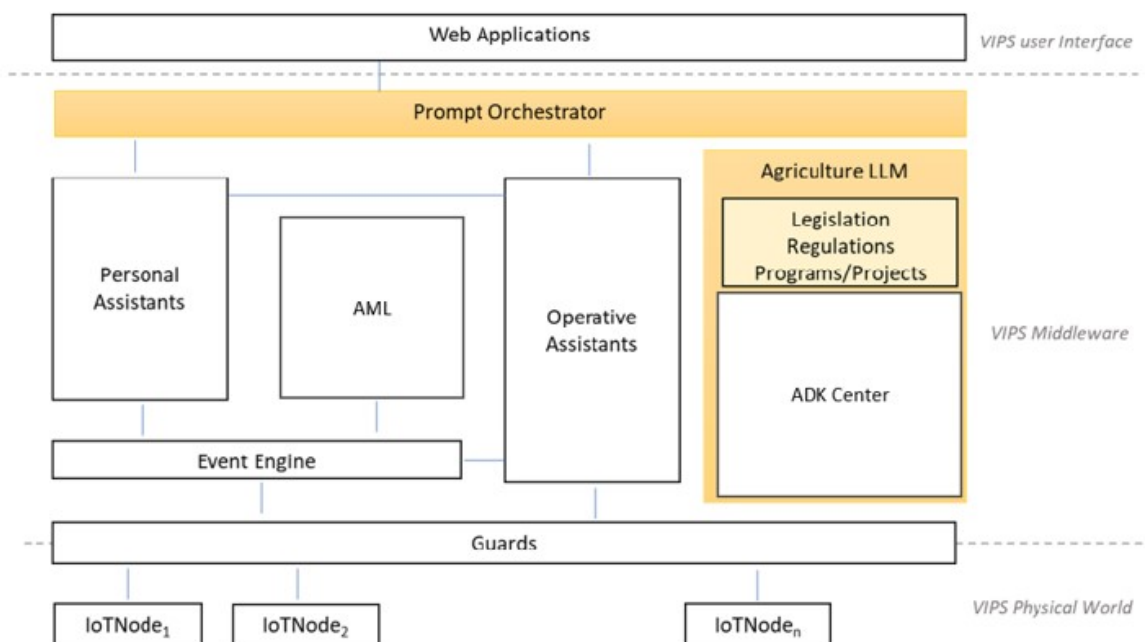
За да се подготви за работа, средата използва два основни входни потока:

- Бавно променящи се входящи ресурси – тези входни данни могат да бъдат земеделско законодателство и нормативни документи и установени земеделски програми, хранилищата на специализирани земеделски знания, земеделски онтологии. Общото между тези документи е, че не подлежат на чести промени.
- Бързо променящи се входове – тези входове се характеризират с честа промяна във времето. Те обикновено се получават от сензорни мрежи или от различни бюлетини (например за времето).



Фигура 2. ЗЕМЕЛА с нова услуга – първи етап

Вторият подход е разширяване на самата платформа с нова услуга. Предлагаме това да стане на два етапа. По време на първия етап Prompt Orchestrator ще бъде поставен между уеб приложенията и асистентите на платформата (Фигура 2). Във втората фаза (Фигура 3), LLM по селско стопанство ще бъде създаден чрез фина настройка на основен модел. В тази версия Prompt orchestrator служи като абстракционен слой между потребителския интерфейс и асистентите на ZEMELA.



Фигура 3. ЗЕМЕЛА с нова услуга – втори етап

Заклучение

Настоящата разработка представя идеята и архитектурата на услуга, която консултира малки и средни фермери при подготовката и изпълнението на проекти по различни програми. Планираме да интегрираме услугата в платформата ЗЕМЕЛА. Следващите стъпки са да се проведат експерименти и с двата подхода, за да се реши коя е най-ефективната архитектура за разработване на прототип на услуга.

Благодарности

Това изследване е подкрепено от Националната програма за научни изследвания „Интелигентно растениевъдство“, Договор за безвъзмездна финансова помощ № D01-65/19.03.2021 г., одобрен с Решение на Министерския съвет № 866/26.11.2020 г., както и от проект BG05M2OP001-1.001-0003: „Център за върхови постижения по Информатика и информационни и комуникационни технологии“, финансиран по Оперативна програма „Наука и образование за интелигентен растеж“.

Литература

- [1] M. Chen, J. Tworek, H. Jun, Q. Yuan, H. Pinto, J. Kaplan, H. Edwards, Y. Burda, N. Joseph, G. Brockman, et al., Evaluating large language models trained on code, 2021, arXiv preprint arXiv:2107.03374;
- [2] S. Stoyanov, J. Todorov, I. Stoyanov, V. Tabakova-Komsalova, L. Dukovska, ZEMELA – An Intelligent Agriculture Platform, *Big Data, Knowledge and Control Systems Engineering – BdKCSE'2021*, 28–29 October 2021, Sofia, Bulgaria.
- [3] D. Rothman, *Transformers for Natural Language Processing*, Packt Publishing Ltd., 2022.
- [4] J. Nekrasov, How to use ChatGPT with your own (proprietary) data, <https://www.twoday.lt/blog/how-to-use-chat-gpt-with-your-own-proprietary-data>
- [5] A. Stoyanova-Doycheva, E. Doychev, V. Ivanova, V. Valkanov, V. Tabakova-Komsalova, Event Ontology about Wheat Cultivation, *AgriControl 2022 The 7th IFAC Conference on Sensing, Control and Automation Technologies for Agriculture*, 14–16 September 2022, Munich, Germany.

Станимир Стоянов¹, Венета Табакова-Комсалова²,

Любка Дуковска³, Милена Кюмюрджиева⁴

^{1,2} Пловдивски университет „Паисий Хилендарски“

Факултет по математика и информатика,

бул. „България“ № 236, Пловдив, България

^{3,4} Институт по информационни и комуникационни технологии,

Българска академия на науките, София

Автор за кореспонденция: stani@uni-plovdiv.bg

GENERATIVE ARTIFICIAL INTELLIGENCE SUPPORTING FARMERS

Stanimir Stoyanov, Veneta Tabakova-Komsalova,
Lyubka Dukovska, Milena Kyumurdzhieva

Abstract. *This paper presents the developing an advisory system implemented as a service that uses the capabilities of generative artificial intelligence and in particular the large language model (LLM). The service is able to assess the risks for farmers when preparing projects under various programs, complying with the Bulgarian legislation related to agriculture, as well as the requirements of the relevant program. The results of the feasibility analysis are summarized in this article. Furthermore, possible service architectures are discussed. The service will be mainly intended to support small and medium-sized farmers.*

Key words: Generative artificial intelligence, Large language model, Smart agriculture, Advisory systems.