ДӘРІС КОНСПЕКТІСІ

Интеллектуалды ақпараттық жүйелерге кіріспе

Дәрістің мақсаты

Студенттерге интеллектуалды ақпараттық жүйелердің мәнін, құрылымын, негізгі технологияларын және жауапты қолдану қағидаларын тұтас жүйе ретінде түсіндіру; нақты міндет үшін орынды интеллектуалды әдісті таңдауға негіз қалыптастыру.

Күтілетін оқу нәтижелері

- интеллектуалды ақпараттық жүйе, жасанды интеллект және дәстүрлі ақпараттық жүйе ұғымдарын ажыратады;
- ИАЖ-дың негізгі қабаттары мен компоненттерін түсіндіреді;
- ережелік жүйе, машиналық оқыту, терең оқыту, анық емес логика, NLP және RAG тәсілдерінің қолданылу жағдайларын салыстырады;
- қарапайым пәндік кейске ИАЖ архитектурасын ұсынады;
- дәлдікпен қатар түсіндірмелілік, әділдік, құпиялық және адам бақылауын бағалайды.

Проблемалық кіріспе

Бүгінгі тақырыпты қарапайым салыстырудан бастайық. Университеттің ақпараттық жүйесі студенттің аты-жөнін, пәндерін, бағаларын және сабаққа қатысуын сақтайды. Ол сұрау бойынша транскрипт жасайды немесе есеп береді. Ал егер сол жүйе студенттің оқу траекториясын талдап, үлгермеу қаупін алдын ала болжаса, тәуекелдің себептерін көрсетсе және кураторға қандай қолдау шарасы тиімді болуы мүмкін екенін ұсынса, біз интеллектуалды функция туралы айта бастаймыз.

Демек, мәселе тек компьютердің көп есептеуінде емес. Интеллектуалдылық деректен мағына шығару, белгісіздік жағдайында қорытынды жасау, тәжірибеден үйрену, контексті ескеру және шешімді адамға түсінікті түрде жеткізу қабілеттерімен байланысты. Осы дәрісте үш негізгі сұраққа жауап іздейміз: жүйе нені біледі, ол қалай ой қорытады және нәтижесіне неге сенуге болады?

Аудиторияға сұрақ: Калькулятор, электрондық журнал және ChatGPT тәрізді көмекшінің қайсысын интеллектуалды жүйе деп санайсыз? Жауабыңызды бір белгімен негіздеңіз.

Күтілетін жауап: калькулятор белгіленген алгоритмді орындайды; электрондық журналдың өзі негізінен деректерді сақтау мен есепке алуға арналған; ал тілдік көмекші контекстке қарай жаңа жауап құра алады. Бірақ генерация жасай алу автоматты түрде дұрыс, қауіпсіз немесе сенімді шешім дегенді білдірмейді. Сондықтан интеллектуалдылық пен сапаны бөлек бағалау керек.

Негізгі ұғымдар және айырмашылықтар

Ақпараттық жүйе, жасанды интеллект және ИАЖ

Ақпараттық жүйе (АЖ) - ақпаратты жинау, сақтау, өңдеу, іздеу және пайдаланушыға ұсыну үшін өзара байланысқан адамдар, процестер, деректер және бағдарламалық-техникалық құралдар жиынтығы. Оның басты құндылығы - ұйымдағы операцияларды тәртіпке келтіру және басқарушылық шешімге қажетті ақпаратты уақытында беру.

Жасанды интеллект (ЖИ) - қабылдау, үйрену, пайымдау, жоспарлау, тілмен жұмыс істеу және әрекет ету сияқты интеллектуалды міндеттерді компьютерлік әдістермен шешуді зерттейтін ғылыми және инженерлік бағыт. ЖИ - кең сала; ол алгоритмдерді, модельдерді және теорияларды қамтиды.

Интеллектуалды ақпараттық жүйе (ИАЖ) - ақпараттық жүйенің деректерді жай өңдеумен шектелмей, білімді немесе үйретілген модельді пайдаланып талдау, болжау, ұсыныс беру, түсіндіру не шешім қабылдауды қолдау қабілеті бар түрі. Яғни ИАЖ ұйымдық ақпараттық орта мен ЖИ әдістерінің түйіскен жерінде орналасады.

> **Есте сақтайтын формула:** *ИАЖ = ақпараттық жүйенің инфрақұрылымы + деректер/білім + интеллектуалды әдіс + пайдаланушыға берілетін негізделген нәтиже + кері байланыс.*

Өлшем	Дәстүрлі ақпараттық жүйе	Интеллектуалды ақпараттық жүйе
Негізгі мақсат	Деректі тіркеу және операцияны автоматтандыру	Белгісіздік жағдайында талдау, болжау немесе ұсыным беру
Логика	Алдын ала анықталған бизнес-процесс	Ереже, ықтималдық модель, үйрену немесе бірнеше әдістің үйлесімі
Нәтиже	Есеп, жазба, транзакция	Класс, ықтималдық, диагноз нұсқасы, ұсыным, табиғи тілдегі жауап
Өзгеріске бейімделу	Көбіне бағдарламаны қайта баптау қажет	Модельді қайта оқыту, білім базасын жаңарту немесе кері байланыспен бейімдеу мүмкін
Негізгі тәуекел	Деректің қате енгізілуі, қолжетімділік	Bias, model drift, қате сенімділік, түсіндірілмеу, автоматтандыруға артық сенім

Шекара әрқашан қатаң емес. Мысалы, ережелік сараптамалық жүйе тәжірибеден өздігінен үйренбеуі мүмкін, бірақ пәндік білімді формалды ережелермен ұсынып, қорытынды шығаратындықтан ИАЖ қатарына жатады. Керісінше, күрделі интерфейсі бар жүйе тек тіркелген сценарийлерді орындаса, оның сыртқы әсерлілігі интеллектуалдылықтың дәлелі болмайды.

Қысқаша даму логикасы

Кезең	Басым тәсіл	ИАЖ үшін маңызы
1950-1960 жж.	Символдық ЖИ, логикалық есептер	Интеллектіні формалды символдар мен ережелер арқылы модельдеу
1970-1980 жж.	Сараптамалық жүйелер	Пән маманының білімін ережелер базасына енгізу
1990-2000 жж.	Машиналық оқыту және деректерді өндіру	Ережені қолмен жазудан деректен заңдылық үйренуге көшу
2010 жж.	Терең оқыту	Кескін, сөйлеу және мәтіндегі күрделі белгілерді автоматты үйрену
2020 жж.	Генеративті ЖИ, үлкен тілдік модельдер, RAG	Мазмұн генерациясы мен ұйым деректеріне негізделген диалогтық интерфейстер

Бұл кезеңдер бірін-бірі толық жойған жоқ. Қазіргі сенімді жүйелердің көбі гибридті: мысалы, тілдік модель пайдаланушы сұрағын түсінеді, іздеу модулі бекітілген құжатты табады, ережелер қауіпсіздік шектеулерін тексереді, ал адам соңғы шешімді қабылдайды.

ИАЖ архитектурасы

ИАЖ-ды тек бір модель деп қарастыру - жиі кездесетін қате. Өндірістік жүйе деректер көзінен бастап мониторингке дейінгі тұтас архитектурадан тұрады. Модель жоғары дәлдік көрсеткенімен, деректер кешіксе, интерфейс түсініксіз болса немесе кері байланыс жиналмаса, жүйе нақты ортада құндылық бере алмайды.

Логикалық ағын: *1) дерек келеді; 2) оның сапасы тексеріліп, белгіге айналады; 3) білім базасы немесе модель қорытынды жасайды; 4) нәтиже түсіндіріледі; 5) пайдаланушы әрекет етеді; 6) нәтиже мен кері байланыс жүйені бақылауға қайтады.*

Қабат	Мысал компоненттер	Негізгі сұрақ
Деректер қабаты	Дерекқор, сенсор, LMS/ERP/CRM, құжат, веб-қызмет	Толықтық, дәлдік, өзектілік, рұқсат
Алдын ала өңдеу	Тазалау, біріктіру, кодтау, белгілер құру, мәтінді бөлу	Қате және leakage болдырмау
Білім/модель қабаты	Ережелер базасы, онтология, білім графы, ML/DL моделі, векторлық индекс	Пәндік білім мен заңдылықты ұсыну
Инференс және шешім	Логикалық шығару, жіктеу, ранжирлеу, жоспарлау, генерация	Кірістен қорытындыға өту
Түсіндіру	Себеп, сенімділік, пайдаланылған дереккөз, қарсы мысал	Сенім және тексеру мүмкіндігі
Интерфейс/интеграция	Дашборд, мобильді қосымша, чат, API, хабарлама	Нәтижені жұмыс процесіне енгізу
Кері байланыс/мониторинг	Лог, пайдаланушы бағасы, қате, drift, қайта оқыту	Сапаны уақыт бойынша сақтау

Барлық қабатты кесіп өтетін талаптар бар: ақпараттық қауіпсіздік, қолжетімділікті басқару, аудит ізі, құпиялық, модель нұсқаларын басқару және жауапты тұлғаларды белгілеу. Бұлар соңында қосылатын қосымша емес, архитектураның бастапқы шарты.

Білімді ұсыну және инференс

Білімді ұсыну - жүйенің пәндік әлем туралы мәліметті қандай формада сақтайтынын анықтайды. Ережелер «ЕГЕР шарт, ОНДА қорытынды» түрінде беріледі. Онтология ұғымдарды, олардың қасиеттері мен байланыстарын анықтайды. Білім графы нақты объектілер мен қатынастарды түйіндер және қырлар ретінде көрсетеді. Машиналық оқыту моделі білімді параметрлерде жасырын түрде сақтайды, ал векторлық ұсыну мағынасы ұқсас мәтіндерді математикалық кеңістікте жақын орналастырады.

Инференс - бар дерек пен ұсынылған білім негізінде жаңа қорытынды алу процесі. Ережелік жүйеде ол тікелей немесе кері тізбектеу арқылы жүреді; ықтималдық модельде - класс ықтималдығын есептеу арқылы; ұсыным жүйесінде - нысандарды релеванттылық бойынша ранжирлеу арқылы; тілдік модельде - келесі токендердің ықтималдық таралуына сүйеніп жауап құрастыру арқылы орындалады.

Интеллектуалды әдістердің негізгі топтары

Әдісті технология сәніне қарап емес, міндеттің табиғатына, дерекке, қателік құнына және түсіндіру талабына қарап таңдаймыз. Бір әдіс барлық мәселені шешпейді. Төмендегі салыстыру бастапқы бағдар береді.

Әдіс	Қашан орынды	Күшті жағы	Шектеуі
Ережелік сараптамалық жүйе	Білім анық, ережелер тұрақты, аудит маңызды	Түсінікті қорытынды; аз дерекпен жұмыс істейді	Ережелер көбейгенде күрделенеді; жаңа жағдайға әлсіз
Машиналық оқыту	Тарихи мысалдардан болжау/жіктеу керек	Жасырын заңдылықтарды табады; өлшенетін сапа	Сапалы дерек қажет; bias және drift болуы мүмкін
Терең оқыту	Кескін, сөйлеу, мәтін сияқты күрделі дерек	Белгілерді автоматты үйренеді; жоғары өнімділік	Көп дерек/есептеу; түсіндіру қиынырақ
Анық емес логика	«төмен», «орташа», «жоғары» тәрізді бұлыңғыр ұғым	Адамдық пайымдауға жақын; басқаруға ыңғайлы	Мүшелік функциялары субъективті бапталуы мүмкін
Прецеденттік пайымдау	Ұқсас бұрынғы жағдайдан шешім бейімдеу	Тәжірибені тікелей қайта қолданады	Ұқсастық өлшемі мен кейс сапасына тәуелді
NLP және тілдік модель	Мәтінді түсіну, іздеу, қысқарту, диалог, генерация	Табиғи интерфейс; көп міндетке бейімделу	Галлюцинация, контекст шектеуі, құпия дерек тәуекелі
RAG	Жауапты ұйым құжаттарына негіздеу қажет	Өзекті дереккөзді жауапқа қосады; тексеру жеңілдейді	Іздеу қатесі генерацияға өтеді; дереккөз сапасы шешуші
Гибридті жүйе	Бір міндетте ереже, модель, іздеу және адам бақылауы қажет	Әдістердің күшті жақтарын біріктіреді	Архитектура, тестілеу және қызмет көрсету күрделірек

Ережелік сараптамалық жүйе

Сараптамалық жүйе пән маманының білімін білім базасында сақтап, инференс механизмі арқылы қорытынды шығарады. Оның классикалық компоненттері: білім базасы, жұмыс жады, шығару механизмі, білімді толықтыру модулі және түсіндіру модулі. Мысал ереже: «ЕГЕР құрылғы температурасы жоғары ЖӘНЕ діріл деңгейі өссе, ОНДА подшипник ақауының ықтималдығы жоғары». Мұндай жүйе шешім жолын ережелер тізбегімен көрсете алады.

Артықшылығы - ашық логика және пәндік шектеулерді дәл орындау. Кемшілігі - білім алу тарлығы: сарапшының үнсіз тәжірибесін толық формалдау қиын. Ереже саны өскенде қайшылық пен сүйемелдеу қиындығы пайда болады.

Машиналық оқыту және терең оқыту

Машиналық оқытуда жүйе кіріс белгілері мен мақсатты нәтиже арасындағы заңдылықты мысалдардан үйренеді. Бақыланатын оқытуда дұрыс жауап белгіленген: мысалы, бұрынғы студенттердің дерегі бойынша оқу қаупін болжау. Бақыланбайтын оқытуда дайын белгі жоқ: алгоритм ұқсас объектілерді кластерге бөледі немесе аномалия табады. Нығайтпалы оқытуда агент әрекет жасап, сыйақы арқылы стратегиясын жақсартады.

Терең оқыту - көп қабатты нейрондық желілер арқылы күрделі белгілерді автоматты үйренетін машиналық оқыту бағыты. Ол кескінді тануда, сөйлеуді өңдеуде, аудармада және үлкен тілдік модельдерде маңызды. Бірақ жоғары тестілік дәлдік әр контексте қауіпсіздікке тең емес: модель жаңа популяцияда, басқа семестрде немесе өзгерген дерек ағынында нашарлауы мүмкін.

Анық емес логика

Классикалық логикада тұжырым ақиқат немесе жалған болады. Анық емес логикада объект ұғымға белгілі дәрежеде жатады. Мысалы, 72 балл «орташа үлгерім» жиынына 0,7, ал «жоғары үлгерім» жиынына 0,3 дәрежеде тиесілі болуы мүмкін. Бұл тәсіл температура, тәуекел, жүктеме сияқты шекарасы қатаң емес ұғымдарды модельдеуге ыңғайлы. Нәтиже мүшелік функцияларына және ережелерге тәуелді болғандықтан, оларды пән мамандарымен негіздеу қажет.

NLP, үлкен тілдік модельдер және RAG

Табиғи тілді өңдеу (NLP) жүйеге мәтінді бөлу, мағыналық ұқсастықты өлшеу, объектілерді тану, жіктеу, аудару және сұраққа жауап беру мүмкіндігін береді. Үлкен тілдік модельдер кең мәтіндік деректерде үйреніп, нұсқаулық пен контекст негізінде жаңа мәтін құрастырады. Олар ықтимал тілдік жалғастықты есептейді; сондықтан сенімді естілетін, бірақ фактілік қате жауап беруі мүмкін.

Retrieval-Augmented Generation (RAG) - пайдаланушы сұрағына қатысты құжат үзінділерін алдымен іздеп, содан кейін сол контекст негізінде тілдік модельге жауап жасататын архитектура. Қарапайым ағын: сұрақты векторға айналдыру, ұқсас үзінділерді табу, оларды промптқа қосу, жауап пен дереккөзді көрсету. RAG модельді қайта оқытпай-ақ ұйымның жаңартылатын құжаттарымен жұмыс істеуге көмектеседі, бірақ іздеу сапасын, құжат нұсқасын және қолжетімділік құқығын міндетті түрде бақылау керек.

Тексеру сұрағы: *Ереже жиі өзгеріп, жауап құжатқа сілтеме беруі тиіс болса, RAG неге тиімді? Күтілетін жауап: ол өзекті дереккөзді тауып, жауапты соған негіздейді және тексеруге жағдай жасайды.*

Қолдану салалары

ИАЖ құндылығы алгоритмнің күрделілігімен емес, нақты пайдаланушыға өлшенетін пайдалы нәтиже беруімен анықталады. Бір салада қате ұсыныс тек

ыңғайсыздық туғызуы мүмкін, ал медицинада немесе көлік басқаруда адам өміріне қауіп төндіреді. Сондықтан қолдану саласы архитектураға, тексеру деңгейіне және адам бақылауына тікелей әсер етеді.

Сала	ИАЖ міндеті	Ерекше бақылау
Білім беру	Жеке оқу траекториясы, академиялық тәуекел, автоматты кері байланыс, білім іздеу	Студентті таңбалау, дерек құпиялығы, педагогикалық жауапкершілік
Медицина	Диагностикалық қолдау, кескін талдау, пациентті сұрыптау	Клиникалық валидация, түсіндірмелілік, дәрігердің соңғы шешімі
Қаржы	Алаяқтықты анықтау, кредиттік скоринг, тәуекел бағалау	Кемсітушілік, жалған дабыл, реттеу және аудит
Өндіріс	Болжамды техникалық қызмет, сапаны бақылау, роботтандыру	Сенсор сапасы, қауіпсіз тоқтату, жабдықтың тозуы
Сауда және сервис	Ұсыным, сұраныс болжамы, чат-бот, динамикалық қор	Манипуляция, filter bubble, тұтынушы дерегі
Мемлекеттік сектор	Өтініштерді жіктеу, ресурсты жоспарлау, қызмет навигациясы	Құқық, ашықтық, шағымдану мүмкіндігі, цифрлық теңсіздік
Киберқауіпсіздік	Аномалия, зиянды мінез-құлық, оқиға басымдығы	Қарсыластың бейімделуі, false positive, құпия логтар

ИАЖ әзірлеудің өмірлік циклі

1. Мәселені және пайдаланушыны анықтау. «ЖИ қолданамыз» емес, қандай шешімді кім және қай кезде қабылдайтынын нақтылау.
2. Табыс өлшемін белгілеу. Техникалық метрикамен бірге уақыт үнемі, қате құны, пайдаланушы сенімі және әлеуметтік әсер өлшенеді.
3. Дерек пен білімді дайындау. Дерек көздері, сапа, рұқсат, өкілдік, белгілеу және жаңарту жиілігі тексеріледі.
4. Әдіс пен архитектураны таңдау. Қарапайым ереже жеткілікті болса, күрделі модель енгізудің қажеті жоқ.
5. Оқыту немесе білім базасын құру. Нұсқалар, параметрлер, дерек жиындары және сараптамалық шешімдер құжатталады.
6. Валидация және қауіп тесті. Дәлдік, тұрақтылық, топтар бойынша әділдік, қауіпсіздік және түсіндірмелілік бағаланады.
7. Пилот және адам бақылауы. Шектеулі ортада нақты жұмыс процесімен тексеріліп, кері байланыс жиналады.
8. Мониторинг және жаңарту. Дерек пен модель drift-і, қателер, шағымдар, нұсқа және қайта оқыту триггерлері бақыланады.

Кейс: «Академиялық навигатор»

Кейс міндеті: ***Университет студенттің семестрді сәтті аяқтамау қаупін ертерек байқап, кураторға дер кезінде және негізделген қолдау әрекетін ұсынғысы келеді. Жүйе жазалау немесе оқудан шығару шешімін автоматты қабылдамайды.***

Мәселені дұрыс тұжырымдау

Нашар тұжырым: «Студенттерді ақылды алгоритммен бағалау». Дұрыс тұжырым: «Алдағы төрт аптада академиялық қиындыққа тап болуы ықтимал студентті ерте анықтап, кураторға тексерілетін себептер мен қолдау нұсқасын ұсыну». Екінші нұсқада уақыт көкжиегі, пайдаланушы, әрекет және адам жауапкершілігі анық.

Элемент	Жобалық шешім	Неліктен маңызды
Кіріс деректері	Сабаққа қатысу динамикасы, тапсырма өткізу, бағалар тренді, LMS белсенділігі, пән жүктемесі	Тек қажетті дерек; медициналық/жеке мәліметті себепсіз қоспау
Интеллектуалды ядро	Түсіндірілетін ML моделі + маңызды академиялық ережелер	Модель тәуекелді бағалайды, ереже міндетті шектерді және ерекше жағдайды тексереді
Шығыс	Тәуекел деңгейі, негізгі 3 фактор, сенімділік, ұсынылатын байланыс	«Сәтсіз студент» сияқты таңба емес, әрекетке бағытталған бейтарап тіл
Адам бақылауы	Куратор деректі тексеріп, студентпен сөйлесіп, әрекетті бекітеді	Автоматты санкция жоқ; студент түсініктеме және түзету сұрай алады
Кері байланыс	Ұсыныс қабылданды ма, қандай қолдау берілді, нәтиже өзгерді ме	Модельді ғана емес, интервенция тиімділігін бағалау

Гибридті шешімнің мысалы

Модель студент үшін p(тәуекел) = 0,78 есептеді делік. Түсіндіру модулі негізгі факторлар ретінде соңғы үш тапсырманың берілмеуін, қатысудың екі аптада төмендеуін және аралық бақылау нәтижесінің нашарлауын көрсетті. Ережелік қабат «егер ресми академиялық демалыс белгісі болса, автоматты хабарлама жіберме» деген шектеуді тексереді. Куратор деректі қарап, студентпен байланысуды немесе белгі дұрыс емес болса, ұсынысты қабылдамауды таңдайды.

Бұл жерде «0,78» ақиқаттың 78 пайызы емес. Ол модельдің берілген дерек пен үйретілген үлгі жағдайында тәуекел класына берген бағасы. Оны шекті мәнмен салыстыруға болады, бірақ шек қателік құнына тәуелді. Шекті төмендетсек, көбірек тәуекелді студентті табамыз, сонымен бірге жалған дабыл да көбейеді.

Нәтиже	Мағынасы	Кейстегі салдары
True Positive	Тәуекел бар, жүйе анықтады	Қолдау дер кезінде ұсынылады
False Positive	Тәуекел жоқ, жүйе бар деді	Артық араласу, студент сенімінің төмендеуі

Нәтиже	Мағынасы	Кейстегі салдары
False Negative	Тәуекел бар, жүйе байқамады	Қолдау қажет студент назардан тыс қалады
True Negative	Тәуекел жоқ, жүйе жоқ деді	Ресурс қажетсіз жұмсалмайды

Бағалау метрикалары міндетке сай таңдалады. Precision - жүйе тәуекел деп көрсеткендердің қаншасы шынымен тәуекелде болғанын, recall - тәуекелде болғандардың қаншасын жүйе тапқанын көрсетеді. F1 осы екеуін теңгерімдейді. Бірақ метрика соңғы мақсатты толық ашпайды: қолдау алғаннан кейін студент нәтижесі жақсарды ма, әртүрлі топтарға жүйе тең әсер етті ме, кураторлар түсіндіруді пайдалы деп тапты ма - осылар да өлшенуі тиіс.

1 минуттық топтық тапсырма: *Үш адамнан шағын топ құрыңыз. Бірінші адам бір қосымша дерек көзін, екінші адам бір ықтимал bias тәуекелін, үшінші адам бір түсіндіру элементін ұсынсын. 40 секундтан кейін екі топтың жауабын тыңдаңыз.*

Мүмкін жауаптар: қосымша дерек ретінде оқытушының құрылымдалған кері байланысы; bias тәуекелі ретінде интернетке қолжетімділігі төмен студенттің LMS белсенділігін қабілетпен шатастыру; түсіндіру ретінде «қатысу соңғы 14 күнде 25 пайыздық тармаққа төмендеді» деген нақты, әрекетке болатын себеп.

Сапа, тәуекел және жауапты ИАЖ

Интеллектуалды жүйе «дұрыс болжады ма?» деген бір сұрақпен бағаланбайды. Ол мақсатқа сай ма, қауіпсіз бе, түсіндіріле ме, деректі заңды және орынды пайдалана ма, өзгерген ортада тұрақты ма, қате болғанда кім жауап береді - осының бәрі сапаның бөлігі.

Өлшем	Жобалау кезіндегі бақылау сұрағы
Мақсат және пропорционалдық	Бұл міндетке ЖИ шынымен қажет пе? Қарапайым ереже жеткілікті емес пе?
Дерек құпиялығы	Қандай дерек жиналады, не үшін, қанша уақыт сақталады, кім көреді?
Әділдік	Әртүрлі топтар үшін қате жиілігі мен нәтиже сапасы қалай өзгереді?
Түсіндірмелілік	Пайдаланушы нәтижеге әсер еткен себеп пен шектеуді түсіне ала ма?
Сенімділік	Жүйе белгісіз жағдайда сенімділігін төмендетіп, адамға жібере ала ма?
Қауіпсіздік	Prompt injection, дерек уландыру, рұқсатсыз қолжетімділік және модельді теріс пайдалану қалай тежеледі?
Адам бақылауы	Қандай шешімді адам бекітеді? Пайдаланушыға шағымдану және түзету арнасы бар ма?
Мониторинг және жауапкершілік	Қате, drift және шағымды кім бақылайды? Қай кезде жүйе тоқтатылады немесе жаңартылады?

Негізгі қауіптер

- Деректегі bias: тарихи теңсіздік модельге заңдылық ретінде көшеді.
- Distribution shift және model drift: нақты орта оқу дерегінен өзгереді немесе уақыт өте өзгереді.
- Галлюцинация: генеративті модель дереккөзде жоқ ақпаратты сенімді түрде құрастырады.
- Automation bias: адам жүйе ұсынысын тексермей қабылдайды, әсіресе интерфейс нәтижені тым сенімді көрсетсе.
- Feedback loop: жүйе шешімі болашақ деректі өзгертіп, өз болжамын жалған түрде «растауы» мүмкін.
- Қауіпсіздік шабуылдары: зиянды енгізу, prompt injection, модельді немесе құпия контексті шығару әрекеттері.

Маңызды қағида: *Жоғары тәуекелді шешімде ИАЖ адамның орнын жай алмастырмайды; ол адамның шешімін негізді, уақтылы және тексерілетін етуге тиіс. Соңғы өкілеттік пен жауапкершілік анық белгіленеді.*

Қорытынды және бекіту

Бес негізгі тұжырым

1. ИАЖ - ақпараттық жүйе мен ЖИ әдісінің қосындысы ғана емес; ол дерек, білім/модель, инференс, интерфейс, адам әрекеті және мониторингтен тұратын әлеуметтік-техникалық жүйе.
2. Интеллектуалды функция деректен заңдылық табу, белгісіздікте қорытынды жасау, контекстке бейімделу немесе білімге сүйеніп түсіндірілетін ұсыным беру арқылы көрінеді.
3. Әдіс міндетке сай таңдалады: тұрақты және түсінікті білімге - ереже; тарихи заңдылыққа - ML; күрделі сенсорлық дерекке - DL; бұлыңғыр ұғымға - fuzzy; құжатқа негізделген диалогқа - RAG.
4. Техникалық метрика жеткіліксіз: қателік құны, пайдаланушыға пайда, әділдік, құпиялық, түсіндірмелілік және адам бақылауы бірге бағаланады.
5. ИАЖ іске қосылған сәтте аяқталмайды. Дерек, модель, пайдаланушы мінез-құлқы және орта өзгеретіндіктен, үздіксіз мониторинг пен жауапкершілік қажет.

Ауызша бекіту сұрақтары

№	Сұрақ	Қысқа жауап
1	Дәстүрлі ақпараттық жүйе мен ИАЖ арасындағы негізгі айырмашылық қандай?	ИАЖ білім/модель арқылы талдау, болжау немесе түсіндірілетін ұсыным жасайды.
2	Ережелік жүйе қай кезде ML-ден тиімді болуы мүмкін?	Ереже анық әрі тұрақты, дерек аз, аудит пен түсіндіру талабы жоғары болғанда.
3	RAG тілдік модель қатесін толық жоя ма?	Жоқ. Ол дереккөзбен негіздеуді жақсартады, бірақ іздеу және генерация қатесі қалады.

№	Сұрақ	Қысқа жауап
4	False negative кейсте нені білдіреді?	Қолдау қажет студентті жүйенің байқамай қалуы.
5	Model drift неге мониторингті қажет етеді?	Уақыт өте нақты дерек өзгеріп, бастапқы сапа төмендеуі мүмкін.

Exit ticket:
Студенттерден қағазға немесе LMS-ке үш қысқа жауап жазуды сұраңыз:

- Бүгін түсінген бір жаңа ұғым.
- ИАЖ-ға айналдыруға болатын бір нақты ақпараттық жүйе мысалы.
- Сол жүйедегі ең маңызды бір тәуекел және оны азайтудың бір тәсілі.

Студенттің өзіндік жұмысына ұсыныс
Кез келген саладағы бір ИАЖ мысалын таңдаңыз. 1 бетте оның пайдаланушысын, кіріс дерегін, интеллектуалды әдісін, шығысын, бір сапа метрикасын және екі этикалық тәуекелін көрсетіңіз. Соңында «адам қай жерде шешім қабылдауы керек?» сұрағына жауап беріңіз.

Негізгі терминдер глоссарийі

Термин	Анықтамасы
Инференс	Бар дерек пен білім/модельден жаңа қорытынды алу процесі.
Білім базасы	Пәндік сала фактілері, ұғымдары және ережелері формалды сақталатын қор.
Онтология	Пәндік ұғымдар, қасиеттер және олардың байланыстарының формалды сипаттамасы.
Белгі (feature)	Модель кіріс ретінде пайдаланатын объект қасиеті немесе одан есептелген көрсеткіш.
Модель	Деректегі заңдылықты параметрлер арқылы бейнелеп, жаңа кіріске нәтиже беретін математикалық объект.
Precision	Оң деп болжанғандардың ішіндегі шынайы оң нәтижелер үлесі.
Recall	Барлық шынайы оң нәтижелердің ішінен модель тапқан үлес.
Model drift	Уақыт өте модель мен нақты орта сәйкестігінің нашарлауы.
Bias	Дерек, модель немесе процесс салдарынан кей топтарға жүйелі әділетсіз әсер етуі мүмкін ауытқу.
RAG	Алдымен релевантты дереккөзді тауып, кейін сол контекстпен генеративті жауап құратын архитектура.
Human-in-the-loop	Шешім процесінде адам тексеруі, бекітуі немесе түзетуі сақталатын тәсіл.

Ұсынылатын әдебиеттер

- Russell, S., Norvig, P. Artificial Intelligence: A Modern Approach. 4th ed. Pearson, 2021.
- Mitchell, T. M. Machine Learning. McGraw-Hill, 1997.
- Goodfellow, I., Bengio, Y., Courville, A. Deep Learning. MIT Press, 2016.
- Giarratano, J. C., Riley, G. D. Expert Systems: Principles and Programming. 4th ed. Thomson, 2005.
- ISO/IEC 22989:2022. Information technology - Artificial intelligence - Artificial intelligence concepts and terminology.
- ISO/IEC 42001:2023. Information technology - Artificial intelligence - Management system.
- NIST. Artificial Intelligence Risk Management Framework (AI RMF 1.0). 2023.