

## Акценти на програмата

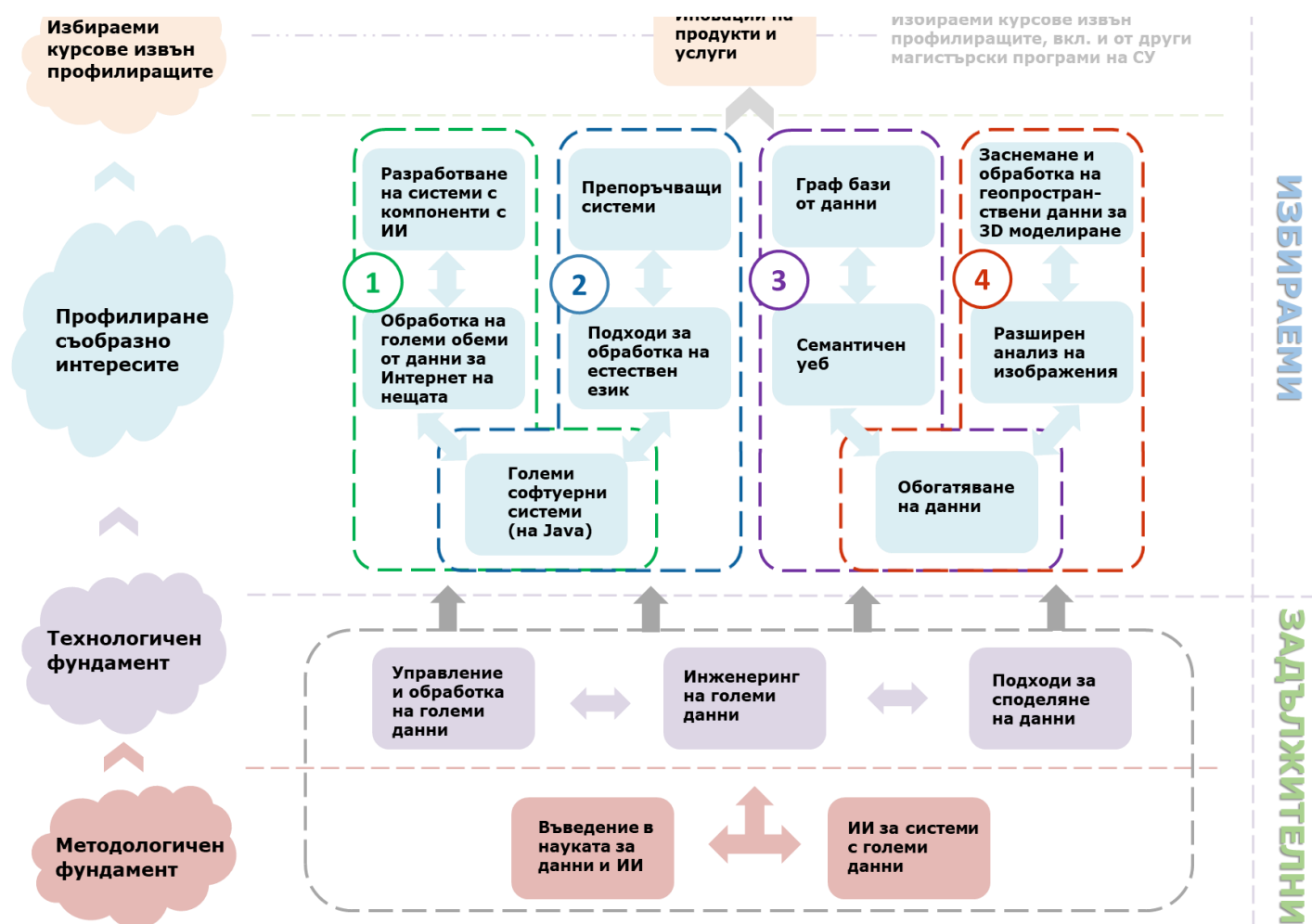
За разлика от популярните академични курсове в областта на науката за данни (Data Science), които се фокусират върху анализ на данни със средствата на машинното обучение и изкуствения интелект, магистърската програма „Технологии за големи данни“ обхваща цялостния жизнен цикъл на данните, а именно:

- Събиране на данни от разнородни източници;
- Обработка на сурови данни и транспортиране до местата за съхранение и анализ;
- Моделиране на данните и съхранение;
- Подготовка на данните за анализ (изчистване, обогатяване, нормализация и др.);
- Управление на потоци от данни;
- Анализ, интерпретация и визуализация.

За да се постигне това, програмата включва набор от курсове, организирани логично в различни видове модули – базови, профилиращи и надграждащи. Това дава възможност на студентите както за фокусирана професионална подготовка, така и за свобода да оформят програмата си за обучение според съвременните тенденции, своя опит и лични предпочитания.

**Магистърската програма „Технологии за големи данни“** е структурирана в четири направления (Фиг. 1), които позволяват систематично изграждане на компетенции и професионални умения за справяне с предизвикателствата на големите данни, използвайки последните постижения на методите на изкуствения интелект и текущото състояние на съвременните технологии за обработка на структурирани, полуструктурирани и неструктурирани данни – сензорни измервания (IoT), текстови съобщения, транзакции, изображения, документи и др.

Три от задължителните курсове в програмата (Управление и обработка на големи данни, Инженеринг на големи данни и Подходи за споделяне на данни) обхващат различни ключови аспекти от жизнения цикъл на големите данни. Заедно с двата основополагащи курса, които въвеждат изкуствения интелект като теоретична основа за обработка на големи данни (Въведение в науката за данни и изкуствен интелект и ИИ за системи с големи данни), те формират петте базови курса на програмата.



**Фигура 1** Структурни връзки между отделните курсове в програмата

Избираемите курсове, включени в магистърската програма, позволяват на студента да избира от четири различни направления за обработка на големи данни според професионалните си предпочитания:

- Хибридни интелигентни системи (1)
- Интелигентна обработка на текст, документи и транзакционни данни (2)
- Семантични технологии за големи данни (3)
- Интелигентна обработка на географска и геопространствена информация (4)

Всяко от направленията включва три курса, които осигуряват необходимата съдържателна база за профилиране на обучаемите.

Останалите избираеми курсове в учебната програма могат да бъдат както други профилиращи, така и избрани сред всеки от курсовете, които не са основни или предназначени за конкретен профил, включително и такива, включени в други магистърски програми на ФМИ.

Информацията за профилираните направления, предлагани в рамките на програмата, е обобщена в таблицата по-долу, като всеки студент може да определи своята област на интереси и да обмисли спрямо желанието си за кариерно развитие какви курсове да избере в допълнение към основните курсове в рамките на магистърската програма.

**Таблица 1** Специализации, предлагани на студентите

Специализации	Методологичен фундамент	Технологичен фундамент	Профилиране според интереси
<b>Хибридни интелигентни системи</b>	1) Въведение в науката за данни и ИИ 2) ИИ за системи с големи данни	3) Управление и обработка на големи данни 4) Инженеринг на големи данни 5) Подходи за споделяне на данни	6) Големите софтуерни системи (на Java) 7) Обработка на големи обеми от данни за Интернет на нещата 8) Разработване на системи с компоненти с ИИ
<b>Интелигентна обработка на текст, документи и транзакционни данни</b>	1) Въведение в науката за данни и ИИ 2) ИИ за системи с големи данни	3) Управление и обработка на големи данни 4) Инженеринг на големи данни 5) Подходи за споделяне на данни	6) Големите софтуерни системи (на Java) 7) Подходи за обработка на естествен език 8) Препоръчващи системи
<b>СЕМАНТИЧНИ ТЕХНОЛОГИИ за големи данни</b>	1) Въведение в науката за данни и ИИ 2) ИИ за системи с големи данни	3) Управление и обработка на големи данни 4) Инженеринг на големи данни 5) Подходи за споделяне на данни	6) Обогащаване на данни 7) Семантичен уеб 8) Граф бази от данни
<b>Интелигентна географска и геопространствена обработка на информация</b>	1) Въведение в науката за данни и ИИ 2) ИИ за системи с големи данни	3) Управление и обработка на големи данни 4) Инженеринг на големи данни 5) Подходи за споделяне на данни	6) Обогащаване на данни 7) Разширен анализ на изображения 8) Заснемане и обработка на геопространствени данни за 3D моделиране

## КРАТКО ОПИСАНИЕ НА ПРОФИЛИТЕ

### 1. ХИБРИДНИ ИНТЕЛИГЕНТНИ СИСТЕМИ

Профилът съчетава анализ на големи данни с помощта на статистически методи и машинно обучение, проектиране на архитектури и модели и интеграция на софтуерни системи, базирани на Интернет на нещата и облачни технологии. Той предоставя цялостно разбиране за жизнения цикъл на данните, софтуерно инженерство с изкуствен интелект и разработка на индустриални системи.

### 2. ИНТЕЛИГЕНТНА ОБРАБОТКА НА ТЕКСТ, ДОКУМЕНТИ И ТРАНЗАКЦИОННИ ДАННИ

Този клъстер от курсове набляга на практически проекти, автоматизирана обработка на текст на естествен език и компоненти и връзки в големи софтуерни системи за обработка на документи, съобщения, финансови и комерсиални транзакции. Той

изследва препоръчващите системи, компютърната лингвистика с фокус върху неструктурирания текст и разработката на софтуерни системи с използване на Java.

### **3. СЕМАНТИЧНИ ТЕХНОЛОГИИ ЗА ГОЛЕМИ ДАННИ**

Профилът разглежда семантични технологии, техники за обогатяване на данни и работа с граф бази данни. Той подготвя студентите за работа с полуструктурирани данни в XML формат и създаване на онтологии, прилагане на машинно обучение и NLP към семантични графи и семантично обогатяване на данните. Обхваща също практическа работа с граф бази данни, включително стандартите за представяне, писането на заявки и интеграцията им в ИТ инфраструктурата.

### **4. ИНТЕЛИГЕНТНА ГЕОГРАФСКА И ГЕОПРОСТРАНСТВЕНА ОБРАБОТКА НА ИНФОРМАЦИЯ**

Клъстерът обхваща усъвършенстван анализ на изображения, технологии и средства за моделиране и интеграция на географски и геопространствени данни. Студентите изучават специализирани методи за анализ на графични данни и конволюционни невронни мрежи за дълбочинен анализ на визуална информация и ги прилагат към реални казуси. Изучават се също различни сензори и методи за събиране на геопространствени данни, техники за дистанционно наблюдение и 3D моделиране, като се набляга на качеството и надеждността на данните и на практическите сесии.

## **ОБЩА ИНФОРМАЦИЯ ЗА ПРОГРАМАТА**

Четирите профили на програмата дават възможност на студентите за специализиране в обработката на различни типове данни: измервателни данни в числов формат, съобщения и документи в текстов формат, финансови и комерсиални записи в транзакционен формат, равнинни и пространствени графики в картографски и геопространствен формат, както и логически модели в полуструктурирани данни като в различните профили се работи и с различни софтуерни системи, подходящи за съхраняване на различните типове данни:

- Postgres (структурирани данни в SQL формат),
- MongoDB (полуструктурирани данни в JSON формат),
- GraphDB (полуструктурирани данни в XML формат) и
- Hadoop (големи структурирани, полуструктурирани и неструктурирани данни във файлов формат).

Повечето от курсовете в магистърската програма включват практически упражнения с използване на езиките за програмиране Java и Python в уникалната за България инфраструктура на Институт GATE. Тази среда е базирана на комерсиален софтуер от технологичния лидер в областта на средствата за обработка на големи данни Cloudera, работещ върху частния облак на Институт GATE под управление на популярната система за работа с облачни контейнери Kubernetes.