



[RU] AI Assistant Usage in Student Life — Аналитический отчет

Автор: Екатерина Сувалова — Junior Data Analyst

Дата: август 2025

Инструменты: Excel, Power BI (Power Query, DAX), GitHub, Notion

1) Краткое резюме

- Проанализировано **10 000 сеансов** использования ИИ-ассистента (далее — ИИ) студентами и учениками старших классов (далее — студенты) за **1 год**; основная аудитория — студенты **бакалавриата 60%**, **школьники** и **магистры** — по **20%**.
 - В **48%** сеансов задание успешно **выполняется**, в **29%** — появляется **идея**, в **16%** — студенты испытывают **затруднения**, а в **8%** — **прекращают** попытки.
 - **71%** пользователей **возвращаются** к ИИ после первого сеанса.
 - Наибольшая **эффективность** достигается при длительности сеанса **5–10 мин**; в низкоэффективных сеансах в **2–3× больше запросов**.
 - Пики **активности** приходятся на **пятницу** и **воскресенье**, а самые загруженные месяцы — **август, июнь** и **январь**.
 - По **количеству сеансов** лидируют **Writing** (письменные задачи) (3,1 тыс.) и **Studying** (учебная подготовка)/**Homework Help** (домашняя работа) (2,0 тыс.), а **Coding** (написание кода) демонстрирует наибольшую долю **успешно завершенных сеансов** — **60%**.
-

2) Цель проекта

Цель: Показать полноценный аналитический процесс на учебном датасете: от подготовки данных и расчета метрик до визуализации и интерпретации результатов. Проект демонстрирует владение Excel, Power BI (Power Query, DAX), а также умение формулировать выводы и описывать ограничения анализа в условиях синтетических данных.

Основные вопросы:

1. Какие группы студентов активнее используют ИИ?
2. Как распределяются сеансы по дисциплинам и задачам?
3. Как распределяются сеансы по уровням эффективности и как на эффективность влияют длительность и число запросов?

- 4. В какие месяцы и дни недели студенты используют ИИ чаще?
- 5. Каковы итоговые исходы сеансов в разрезе дней недели и типов задач?
- 6. Насколько устойчив интерес студентов к ИИ?

3) Данные и источники

- **Источник:** AI Assistant Usage in Student Life (Kaggle).
- **Объем:** 10 000 строк (сеансов) за **полный календарный год** (24 июня 2024 — 24 июня 2025).
- **Цель создания датасета:** смоделировать **реалистичные взаимодействия студентов с ИИ-инструментами** (например, ChatGPT) в учебных целях. Датасет полностью **синтетический**, но построен так, чтобы отражать реальную логику и поведение. Подходит для исследовательского анализа данных, экспериментов с машинным обучением, построения признаков и визуализаций.
- **Основные поля:**

StudentLevel	школьник, бакалавр, магистр
Discipline	Дисциплина (например, биология, психология и др.)
SessionDate	Дата и время сеанса
SessionLengthMin	Длительность сеанса в минутах
TotalPrompts	Количество запросов в сеансе
TaskType	Тип задачи (например, Writing, Coding, Research)
AI_AssistanceLevel	Оценка полезности ИИ по шкале от 1 до 5
FinalOutcome	Итог сеанса: выполнено задание, появилась идея, запутался, сдался
UsedAgain	Использовал ли ИИ снова
SatisfactionRating	Оценка удовлетворенности по шкале от 1 до 5

- **Производные метрики:**

Prompt density 10min	частота запросов за 10 минут сеанса	Формула: Total prompts / (Session length min / 10)
Efficiency score	индекс эффективности использования ИИ: отношение оценки полезности к частоте запросов. Чем меньше запросов нужно для высокой оценки полезности, тем выше эффективность	Формула: AI Assistance Level / Prompt density 10min

4) Используемые инструменты

Excel:

- первичная очистка данных
- проверка типов, нормализация дат/времени, поиск аномалий

- расчет производных показателей
- создание сводных таблиц

Power BI:

- анализ и визуализация данных,
- создание мер и вычисляемых колонок в DAX, добавление столбцов в Power Query
- разработка и оформление интерактивных дашбордов с tooltips (всплывающими подсказками)

GitHub/Notion: публикация и презентация проекта (RU/EN)

5) Методология

1. **Импорт и первичная очистка данных в Excel:** удаление дубликатов, заполнение пропусков, выравнивание форматов дат и времени.
 2. **Проверка структуры:** анализ распределений, выявление логических связей между полями.
 3. **Подготовка данных в Power Query:** настройка форматов и единиц измерения, переименование полей для удобства восприятия, создание условных столбцов (conditional columns), формирование сортировочных ключей, группировка показателей по категориям (бины по Efficiency score, Session length, Prompt density).
 4. **Создание DAX-мер:** расчет дополнительных показателей (DayOfWeek, High Efficiency Share %, Outcome % и др.; см п.6).
 5. **Анализ и интерпретация данных:** выявление закономерностей и ключевых различий по длительности сеансов, количеству запросов, типам задач, показателям эффективности и тд.
 6. **Разработка дизайна дашбордов:** подбор информативных визуализаций, добавление tooltips с краткими выводами, настройка цветовой палитры и приведение к единому стилю.
 7. **Формулирование итоговых выводов:** обобщение результатов анализа и визуализаций, подготовка краткого текстового резюме для презентации проекта.
 8. **Презентация проекта:** подготовка визуальных материалов (скриншоты, GIF-анимации), структурирование и оформление проекта в Notion для наглядной и удобной демонстрации результатов.
-

6) DAX: вычисляемые столбцы и меры

Помимо производных метрик, рассчитанных в Excel и условных колонок в Power Query, в DAX были созданы дополнительные меры и вычисляемые столбцы.

DayOfWeek

Вычисляемый столбец, переводящий **дату сеанса в название дня недели** (Monday, Tuesday и т.д.). Используется для анализа распределения количества сеансов по дням

недели. Используется в графике Number of Sessions by Day of Week на дашборде Time & Behavior.

```
DayOfWeek = FORMAT([Session date], "dddd")
```

DayOfWeekNum

Вспомогательный вычисляемый столбец, который присваивает дню недели порядковый номер 1–7. Используется как **сортировочный ключ** для поля DayOfWeek, чтобы дни отображались в календарном порядке.

```
DayOfWeekNum = WEEKDAY([Session date], 2)
```

High Efficiency Share %

Мера вычисляет **долю сеансов с высоким показателем эффективности** от общего числа сеансов. Используется в KPI карточке на дашборде Efficiency & Prompt Analysis.

```
High Efficiency Share % =  
DIVIDE(  
    CALCULATE(DISTINCTCOUNT('ai_assistant_usage_student_life'[Sessions]), 'ai_assistant_usage_student_life'[EfficiencyScoreGroup] = "High (> 2.2)"),  
    CALCULATE(DISTINCTCOUNT('ai_assistant_usage_student_life'[Sessions]), REMOVEFILTERS('ai_assistant_usage_student_life'[EfficiencyScoreGroup]))  
)
```

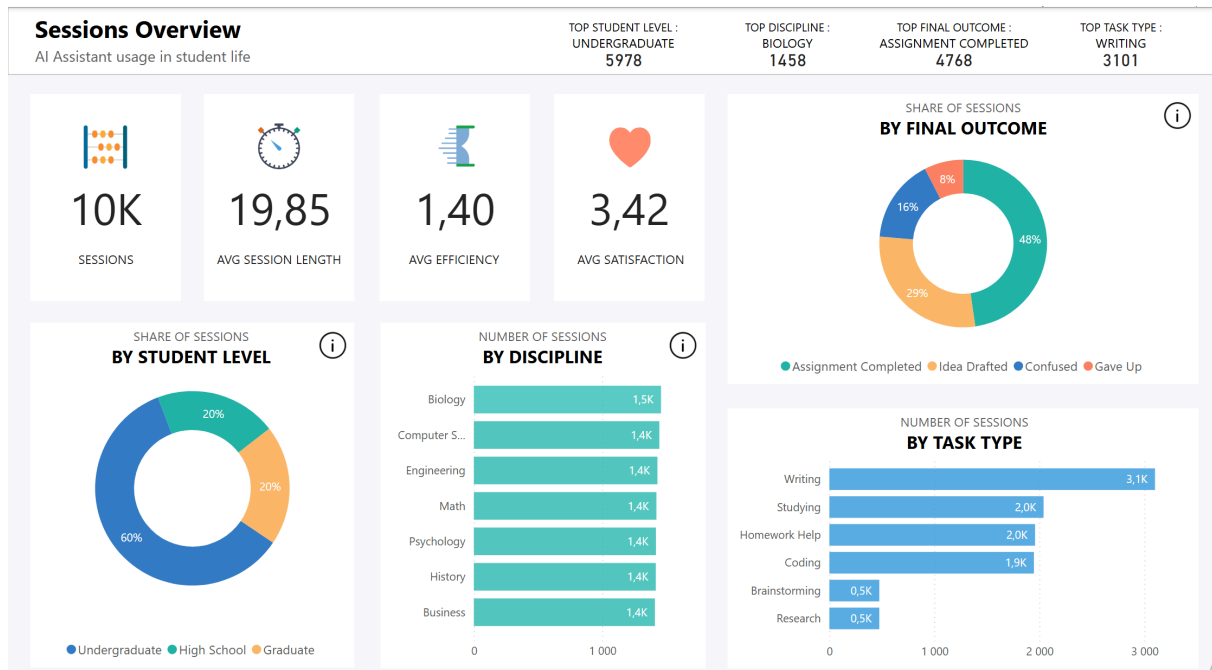
Outcome %

Мера рассчитывает **процентное распределение исходов** (outcomes) **внутри каждой группы задач** (Task Type). Показывает, какую долю занимают разные результаты в рамках одного типа задач. Используется в графике Share of Tasks by Final Outcome на дашборде Outcome Analysis.

```
Outcome % =  
VAR Numerator =  
    [Sessions (distinct)]  
VAR Denominator =  
    CALCULATE (  
        [Sessions (distinct)],  
        ALLEXCEPT (  
            'ai_assistant_usage_student_life',  
            'ai_assistant_usage_student_life'[Task type]  
        )  
    )  
RETURN  
    DIVIDE ( Numerator, Denominator )
```

7) Выводы по дашбордам

7.1 Overview



- **Share of Sessions by Final Outcome**

Почти половина всех сеансов (48%) завершается успешным выполнением задания (**Assignment Completed**), что свидетельствует о высокой результативности использования ИИ. Еще 29% пользователей используют систему для генерации идей (**Idea Drafted**), превращая ее в источник вдохновения. Вместе эти две категории составляют более 3/4 всех исходов, демонстрируя значительный образовательный потенциал ИИ. При этом в 16% случаев студенты испытывают затруднения (**Confused**), главным образом в задачах Studying (учебная подготовка) и Homework Help (домашняя работа). Это может указывать на ограничения модели: она не всегда способна разъяснить сложные концепции или ее ответы оказываются недостаточно понятными для учащихся. Лишь 8% пользователей полностью прекращают работу (**Gave Up**), что подчеркивает редкость полных отказов и сохраняющуюся вовлеченность, особенно с учетом того, что 71% студентов возвращаются к ИИ повторно.

- **Share of Sessions by Student Level**

На студентов бакалавриата приходится 60% активности, что делает их основной группой пользователей ИИ. Школьники и магистры показывают одинаковую долю (по 20%), обращаясь к ИИ заметно реже. Это может означать, что школьники пока сталкиваются с меньшим числом задач, где применение ИИ оправдано, а магистры чаще полагаются на собственные исследовательские навыки. Таким образом, именно бакалавриат выступает наиболее вовлеченной группой, хотя ИИ сохраняет потенциал и среди других групп.

- **Number of Sessions by Discipline**

Биология немного опережает другие дисциплины по числу сеансов (≈1,5 тыс.),

однако в целом **распределение остается достаточно равномерным**: компьютерные науки, инженерия, математика, психология, история и бизнес также активно представлены (по $\approx 1,4$ тыс.). Это показывает, что **ИИ находит применение во всех сферах обучения** — от гуманитарных до точных и прикладных наук. Такой баланс подчеркивает их универсальность и ценность для студентов разных направлений.

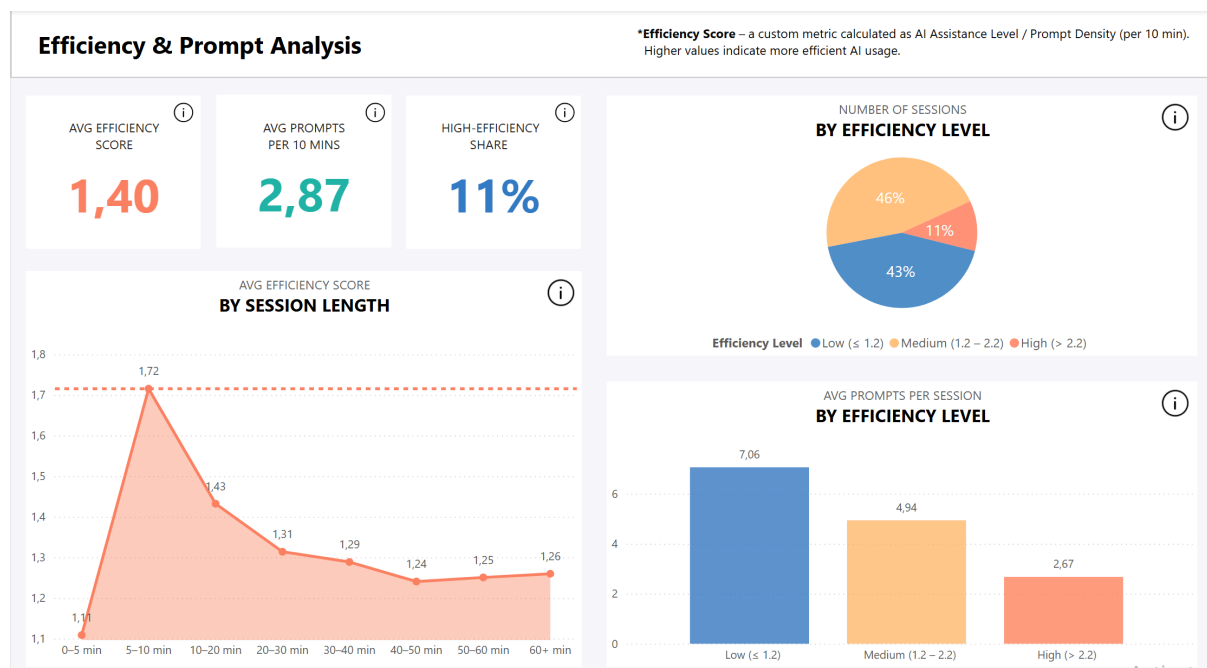
- **Number of Sessions by Task Type**

Письменные задачи (Writing) заметно доминируют — более 3,1 тыс. сеансов, что составляет почти **треть всей активности**. Это ожидаемо, ведь работа с текстами является ключевым элементом академической деятельности. Далее следуют **Studying** и **Homework Help** (по 2,0 тыс.), а также **Coding** (1,9 тыс.), что подчеркивает востребованность ИИ как для подготовки к учебе, так и для практических заданий. Значительно **реже ИИ используется для Brainstorming и Research** (по 0,5 тыс.), что говорит о восприятии его **скорее как помощника в решении конкретных задач, чем как средство для поиска идей или проведения исследований**.

Что это значит:

ИИ постепенно становится неотъемлемой частью учебного процесса, пользуясь особой **популярностью среди студентов бакалавриата** и при выполнении **письменных заданий**. Он показывает **хорошую результативность и универсальность**: используется и в гуманитарных, и в точных, и в прикладных дисциплинах. Хотя часть студентов (16%) сталкивается с трудностями, **низкий уровень полных отказов (8%)** и **высокая доля повторных обращений** показывают, что **ИИ воспринимается как надежный и полезный инструмент в учебе**.

7.2 Efficiency & Prompt Analysis



- **Avg Efficiency Score by Session Length**

Короткие сеансы (5–10 минут) показывают наибольшую эффективность (1,72), что указывает на то, что быстрые и сфокусированные взаимодействия с ИИ приносят

лучшие результаты. **Более длительные сеансы (10+ минут) характеризуются устойчиво более низкой эффективностью** ($\approx 1,2-1,4$), возможно, из-за утомляемости или менее целенаправленного использования. **Самые короткие (0–5 минут) сеансы оказываются наименее продуктивными** (1,11), что, вероятно, связано с пробными запросами или прерванными попытками.

Однако важно помнить, что неравномерность показателей связана не только с качеством взаимодействия. Например, исследовательские и творческие задачи требуют больше времени и большего количества запросов, поэтому их показатель эффективности может быть ниже, чем у прикладных задач.

Для упрощения визуального восприятия минимальное значение на графике установлено на 1,1.

- **Number of Sessions by Efficiency Level**

Показатель Efficiency Score разделен на три группы по терцилям: низкий ($<1,2$), средний ($1,2-2,2$) и высокий ($>2,2$). **Основная часть сеансов приходится на низкий (43%) и средний (46%) уровни.** При этом почти половина случаев попадает в среднюю категорию, что отражает сбалансированное использование ИИ большинством студентов. **Высокая эффективность встречается лишь в 11% сеансов,** что может быть связано как с особенностями метрики Efficiency Score, так и с самим характером учебных задач: в большинстве случаев требуется несколько уточняющих запросов, и это естественным образом удерживает студентов в «средней» зоне.

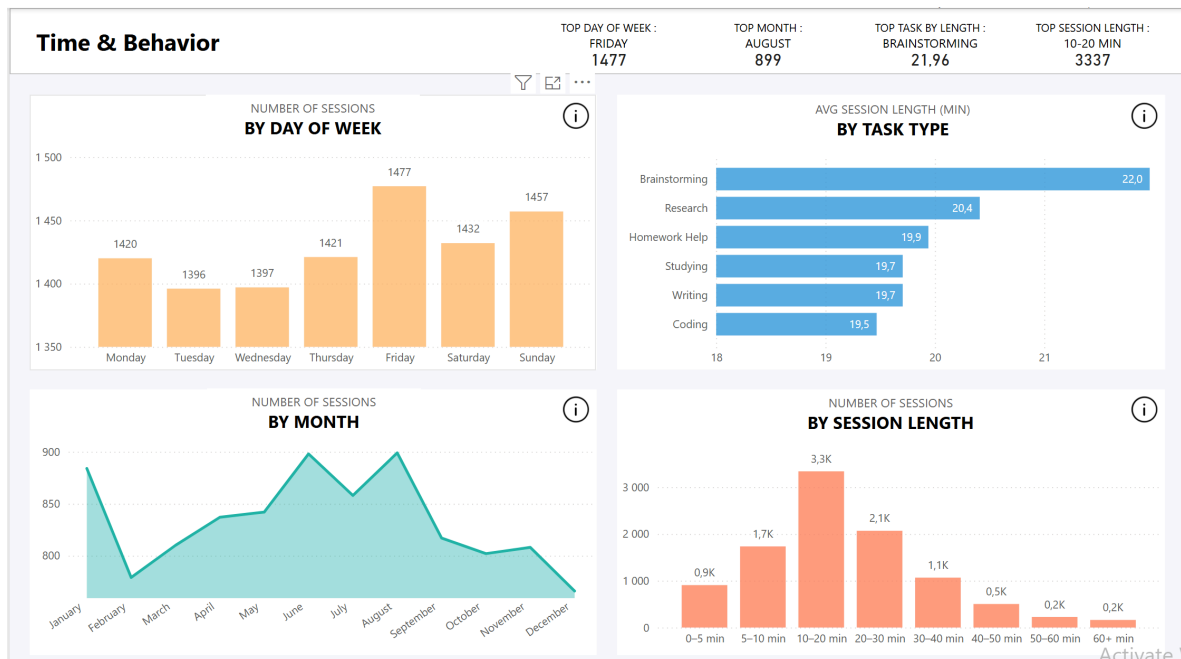
- **Avg Prompts per Session by Efficiency Level**

График наглядно показывает логику метрики Efficiency Score: **чем меньше запросов требуется для достижения результата, тем выше эффективность.** Так, у пользователей с **низкой эффективностью** в среднем **7,06 запросов** на сеанс, тогда как у **высокоэффективных** — всего **2,67**.

Что это значит:

Данные показывают, что **эффективность работы с ИИ зависит от сочетания небольшого числа продуманных запросов и сфокусированного взаимодействия.** **Наиболее продуктивными оказываются короткие сеансы (5–10 минут)** с ограниченным количеством четко сформулированных запросов. При этом **высокий уровень эффективности встречается сравнительно редко (11%),** что связано как с особенностями метрики Efficiency Score, так и с характером учебных задач.

7.3 Time & Behavior



- **Number of Sessions by Day of Week**

Наибольшее количество сеансов приходится на **пятницу (1477), воскресенье (1457) и субботу (1432)**. **Минимальные значения** наблюдаются **во вторник (1396) и среду (1397)**. Такая динамика может объясняться учебным ритмом: в середине недели у студентов более плотный учебный график, и времени на самостоятельную учебу остается меньше. **К концу недели активность возрастает** — вероятно, пик выполнения домашних заданий приходится на пятницу и выходные. Интересно, что в пятницу популярностью пользуются coding-задачи, тогда как в выходные чаще встречаются studying и homework help.

Для упрощения визуального восприятия минимальное значение на графике установлено на 1 350.

- **Avg Session Length (in minutes) by Task Type**

Наибольшая средняя продолжительность сеансов характерна для **Brainstorming (22 мин) и Research (20,4 мин)** — задач, требующих поиска идей и анализа информации. Более прикладные и структурированные виды активности — **Homework Help (19,9 мин), Studying (19,7 мин), Writing (19,7 мин), Coding (19,5 мин)** — занимают **чуть меньше 20 минут**. Это отражает различие между творческими и более конкретными задачами.

Для упрощения визуального восприятия минимальное значение на графике установлено на 18.

- **Number of Sessions by Month**

Пики активности приходятся на июнь (898), август (899) и январь (884) и совпадают с периодами экзаменов и подготовки к поступлениям. Высокая активность в августе может быть также связана с сезоном найма в IT, что подтверждается ростом числа сеансов по Coding. **Самый низкий показатель наблюдается в феврале (779)**, вероятно из-за зимних каникул. В целом динамика указывает на то, что **активность студентов**

возрастает в периоды академических и карьерных пиковых нагрузок и снижается во время каникул.

Для упрощения визуального восприятия минимальное значение на графике установлено на 750.

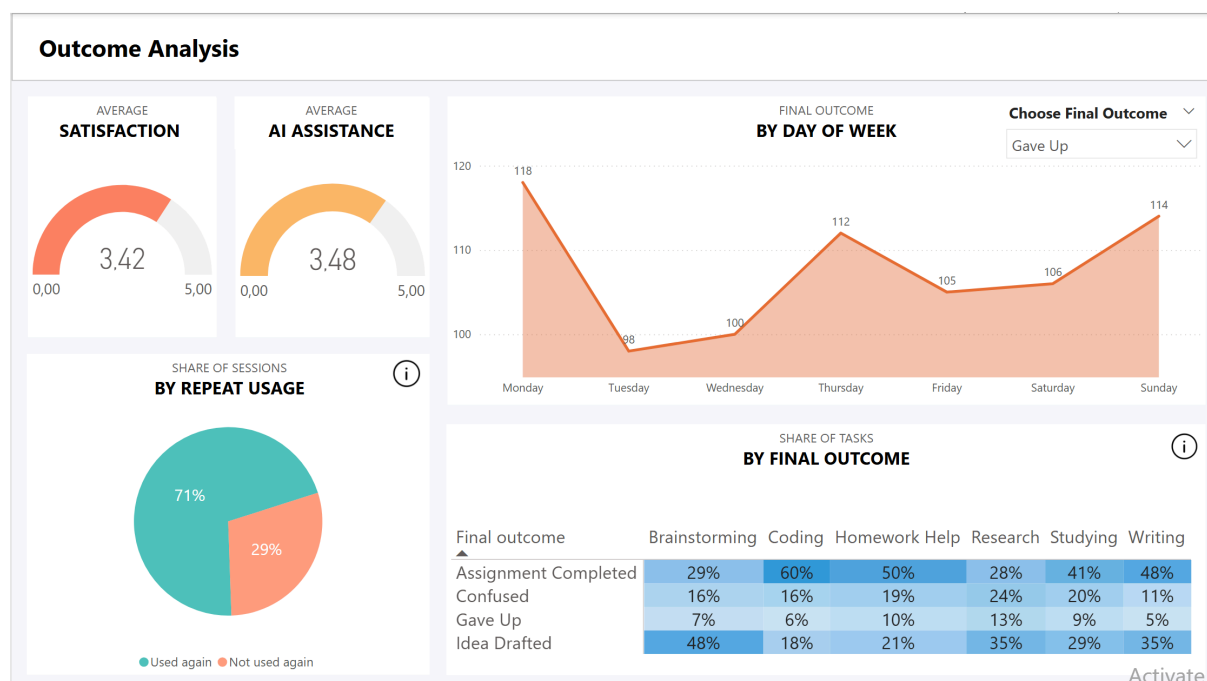
- **Number of Sessions by Session Length**

Большинство сеансов длится 10–20 минут (3,3 тыс.). Также распространены форматы 20–30 минут (2,1 тыс.) и 5–10 минут (1,7 тыс.). Более продолжительные взаимодействия (30+ минут) встречаются значительно реже, а самые короткие (<5 минут) составляют всего 0,9 тыс. случаев. В целом основная часть сеансов укладывается в диапазон 5–30 минут, что отражает характер студенческих задач: от быстрых уточнений до работы над небольшими фрагментами учебного материала.

Что это значит:

Студенты чаще всего используют ИИ для кратких и средних по длительности сеансов — 5–30 минут, что соответствует типичному формату учебной работы: решение задач или поиск идей. Самые длинные сеансы характерны для исследовательских и творческих задач (Research, Brainstorming), тогда как практические активности укладываются в более сжатые временные рамки. В течение недели активность распределяется относительно равномерно, но усиливается в пятницу и на выходных. Сезонная динамика указывает на всплески использования ИИ в периоды летних и зимних экзаменов. Таким образом, ИИ выступает в роли гибкого помощника, поддерживающего студентов и в повседневной работе, и в периоды экзаменационной и карьерной нагрузки.

7.4 Outcome Analysis



- **Share of Sessions by Repeat Usage**

71% пользователей возвращаются к использованию ИИ после первого сеанса. Это говорит о том, что студенты в целом воспринимают его как полезный инструмент,

оправдывающий ожидания. **Доля студентов, не обратившихся к ИИ повторно относительно невысокая** (29%) и может объясняться тем, что часть студентов использовала ИИ для решения разовой задачи или не получила достаточной пользы от взаимодействия.

- **Final Outcome by Day of Week**

Распределение исходов сеансов по дням недели показывает заметные различия в динамике. **Количество сеансов с успешным выполнением задания** (Assignment Completed) **достигает пика в пятницу**, что согласуется с наибольшим количеством сеансов в этот день и указывает на стремление студентов завершать задания к концу недели. **Показатель Idea Drafted наиболее выражен в воскресенье**, что может отражать склонность использовать выходные для генерации идей. **Доля затруднительных сеансов** (Confused) **заметно повышается в пятницу и субботу**: это совпадает с пиком активности по числу сеансов (см. Number of Sessions by Day of Week). При этом **в воскресенье**, несмотря на второе по величине количество сеансов, **показатель Confused снижается**, что может говорить о более сфокусированном и целенаправленном использовании ИИ. **Показатель Gave Up** (полное прекращение попыток) **остаётся на относительно стабильном уровне в течение недели**, с небольшими повышениями в понедельник, четверг и воскресенье, отражая естественные колебания вовлечённости. В целом график подчёркивает цикличность продуктивности и трудностей в зависимости от дня недели.

- **Share of Tasks by Final Outcome**

Сравнение типов задач по итоговым исходам показывает заметные различия в результативности. **Наибольшая доля успешного завершения** (Assignment Completed) **наблюдается в задачах Coding (60%), Homework Help (50%) и Writing (48%)**, что указывает на высокую эффективность ИИ при выполнении этих задач. **Brainstorming чаще всего приводит к формированию идеи** (Idea Drafted, 48%), что подтверждает его ценность как подготовительного этапа, а не конечного результата. **Research отличается наибольшей долей отказов** (Gave Up) **и более равномерным распределением остальных исходов**, что отражает сложность и неопределённость этого типа задач. **Задача Studying в основном выполняется успешно, но также нередко приводит к возникновению трудностей** (Confused) **или появлению идеи** (Idea Drafted), что указывает на то, что процесс обучения не всегда ведёт к прямому результату, а может включать этапы поиска и преодоления трудностей.

Что это значит:

Анализ исходов сеансов показывает, что **студенты в целом воспринимают ИИ как полезный инструмент и активно возвращаются к его использованию**. При этом **результат взаимодействия зависит от контекста**: задачи, связанные с практической работой (Coding, Homework Help, Writing), чаще всего завершаются успешно, тогда как исследовательские и более открытые форматы (Research, Brainstorming) приводят к возникновению идеи или прекращению попыток. Динамика по дням недели отражает закономерности в использовании ИИ: **в пятницу студенты успешно завершают больше заданий, в выходные чаще используют ИИ для генерации идей, а количество сеансов, при которых студенты столкнулись с затруднениями (Confused), либо прекратили попытки (Gave Up) колеблется, но остаётся в пределах естественной учебной нагрузки**. В целом данные подчёркивают, что **ИИ встроен в учебный процесс как инструмент для**

выполнения конкретных задач, поиска идей и поддержки при возникновении трудностей, при этом его восприятие и результативность зависят как от типа деятельности, так и от времени использования.

8) Выводы по проекту

ИИ встроен в учебную практику как гибкий помощник: чаще всего студенты используют его в **формате коротких и средних сеансов (5–30 минут)** в периоды повышенной учебной нагрузки — **к концу недели и в экзаменационные месяцы**. **Эффективность выше при более коротких и сфокусированных сеансах** с меньшим числом продуманных запросов. Итог взаимодействия связан с типом активности: **прикладные задачи** вроде написания кода, выполнения домашних заданий или написания текстов **чаще завершаются успехом**, тогда как **креативные и исследовательские задачи** (Brainstorming, Research) нередко **приводят к возникновению идеи или столкновению с трудностями**. При этом **большинство взаимодействий дает результат** (полный или частичный), а **доля прекращенных попыток остается невысокой** — что согласуется с **71% повторных обращений**.

В целом можно сказать, что ИИ стал для студентов **привычным рабочим инструментом**: они обращаются к нему, когда нужно **быстро справиться с заданием, найти идею или разобраться в сложной теме**. Даже если не всегда удастся добиться идеального результата, **опыт использования ИИ чаще оказывается полезным**, и именно поэтому большинство студентов возвращаются к нему снова.

9) Ограничения анализа

- **Метрика Efficiency Score — условный показатель**

Она отражает соотношение числа запросов и достигнутого результата, но интерпретация корректна только в рамках данной логики (меньше запросов = выше эффективность). В реальности продуктивность взаимодействия с ИИ может определяться и другими факторами (например, сложностью задания или стилем работы студента).

- **Ограниченность контекста**

В датасете нет информации о географии, уровне подготовки и мотивации студентов, а также об академических результатах и технических параметрах взаимодействия. Также не учитываются внешние обстоятельства, такие как расписание занятий, экзаменационные периоды, учебная нагрузка, которые могут влиять на активность студентов при использовании ИИ. Поэтому анализ отражает только характер сеансов, а не реальную успешность студентов в обучении.

- **Синтетическое происхождение данных**

Датасет создан искусственно для моделирования учебных сценариев. Это значит, что он не обязательно полностью отражает реальные паттерны поведения студентов, а лишь имитирует их на основе заданных правил. Поэтому результаты анализа стоит воспринимать скорее как моделирование сценариев, чем как исследование реального опыта.

- **Упрощение категорий исходов**

Все результаты сведены всего к четырем категориям (Assignment Completed, Idea Drafted, Confused, Gave Up). Это дает целостное представление, но не отражает более тонкие сценарии (например, частичный успех, неполное понимание или получение неверного ответа).

Таким образом, результаты анализа показывают **возможные закономерности в синтетически смоделированных данных**, но их основная ценность заключается в **демонстрации аналитического подхода и применения инструментов**, а не в точной интерпретации реального поведения студентов.

10) Как воспроизвести

1. **Скачайте** файлы .pbix и .xlsx из GitHub.
 2. **Откройте** .pbix в Power BI Desktop (только пользователи с установленным Power BI Desktop смогут запустить проект).
 3. **Переключайтесь между страницами отчета**; при наведении на визуализации появятся всплывающие подсказки с краткими пояснениями.
 4. Если Power BI недоступен, см. **отчеты (RU/EN) в Notion** — в них представлены скриншоты и gif-анимации дашбордов.
-

11) Ссылки и файлы

GitHub-репозиторий: [ai-assistant-student-life](#)

Excel: [ai_assistant_usage_student_life.xlsx](#)

Power BI: [ai_assistant_usage_student_life.pbix](#)

Notion презентация (RU): [\[RU\] AI Assistant Usage in Student Life — Аналитический отчет](#)

Notion презентация (EN): [\[EN\] AI Assistant Usage in Student Life — Analytical Report](#)

Источник данных: [Kaggle](#)