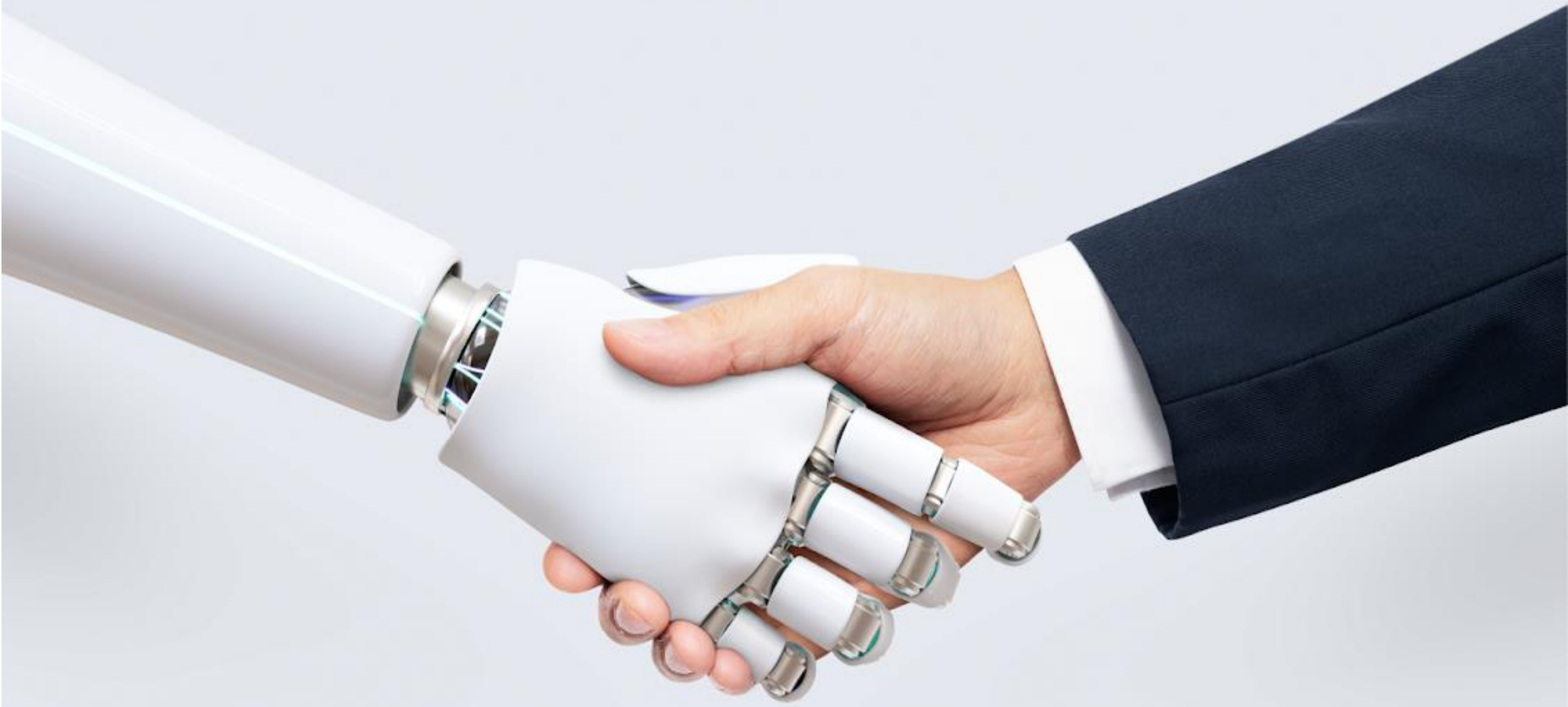
# SARA AI



## Incorporating AI in Education

## Table of Contents

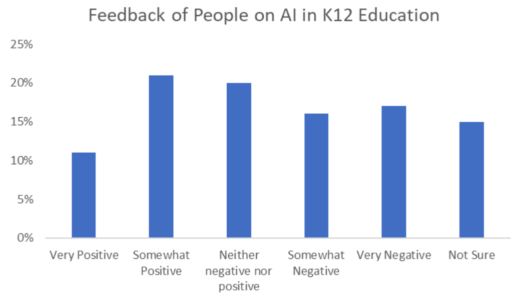
# Problem Identification

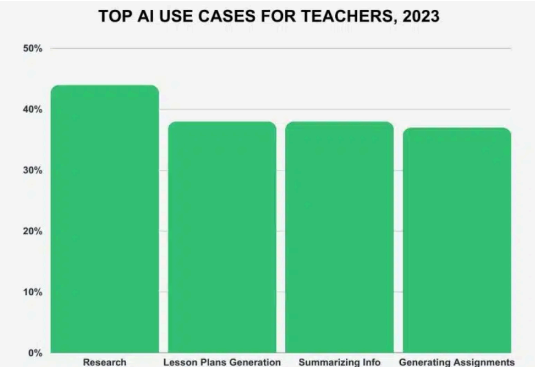
## AI & EdTech Revolution:

It has been fast-tracking disruption in education, starting from personalized learning and automated grading to intelligent curriculum design, among others. The funding of EdTech and the demand for future-ready skills have boosted the artificial intelligence in education market towards a boom.

## Key Statistics for AI in The Education Market

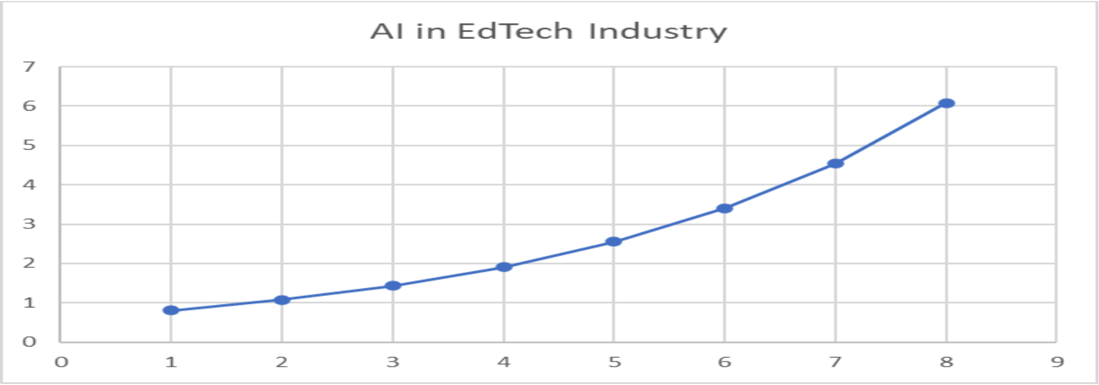
- The AI in the education market was valued at $2.5 billion in 2022 and is projected to surge to $88.2 billion by 2032 at a CAGR of 43.3%.
- So, for well over a third of the people, AI in education is positive: 37% of men, and 27% of women agree that the impact of AI in education is somewhat beneficial, and 53% of men and 51% of women agree that schools must teach children how to use AI.





- The market for AI in personalized learning is forecast to grow at a CAGR of 44.3% to reach $48.7 billion by 2030 – up from $5.2 billion in 2022.
- AI helps 73% of students understand the learning material better and 63% study more efficiently.

AI is automating tasks, from administration to complex educational processes. By 2025 it will be central to learning. Virtual training, common in adult education, is expanding to schools. AI tool are streamlining educators' work provides ready-made, tailored curriculum resources.



AI is shifting the tectonic plates in the sphere of education, turning around even the very key elements, by some digital classroom assistants powered by data-driven insights. Other promised potentials for educative AI are better learning experiences, decreasing administrative burdens, and future-readiness. But most importantly, the ethical aspects of keeping privacy, securing data, avoiding biases in algorithms, and striking a balance between human interaction and technology could help Artificial Intelligence to be at its best.

### Gap Analysis of AI Solutions in EdTech Sector

The AI technologies are fast developing in the EdTech sector for solutions that enhance teaching and learning experiences. Despite huge progress, there are gaps that still exist and for that to be attained in respect of the AI potential within education. Below is a gap analysis focused on the current AI-driven solutions available to teachers and students.

| Category | Current Solutions | Existing Gaps | Opportunities for Improvement |
|---|---|---|---|
| **Personalized Learning** | AI-driven platforms tailor learning experiences to individual student needs, adapting content based on real-time performance data. | **Limited Customization**: Many solutions provide only surface-level personalization, lacking depth in adapting to unique learning styles, cultural contexts, and individual interests. | Develop AI systems that incorporate a broader range of data points, including emotional and social factors, to create more holistic and deeply personalized learning experiences. |
| **Curriculum Generation** | AI tools generate curriculums aligned with educational standards, providing content, activities, and assessments linked to learning outcomes. | **One-Size-Fits-All**: Current AI-generated curricula often lack flexibility to accommodate diverse educational environments, such as under-resourced schools or different pedagogical approaches. | Enhance AI tools to allow more customizable and adaptable curriculum design that can be tailored to different teaching contexts and educational philosophies. |

| | | | |
|---|---|---|---|
| **Grading and Assessment** | Automated grading systems use AI to evaluate student work, providing instant feedback and reducing the administrative burden on teachers. | **Bias and Fairness Issues**: AI grading systems reinforce biases present in training data, leading to unfair assessments. Additionally, AI may struggle with subjective areas like creativity and critical thinking. | Improve the transparency and fairness of AI grading algorithms by integrating diverse datasets and incorporating human oversight in the grading process, particularly in subjective assessments. |
| **Student Engagement and Motivation** | AI-driven gamification and interactive platforms are used to boost student engagement and motivation by making learning more engaging and fun. | **Sustainability of Engagement**: While initial engagement may be high, maintaining student interest over time is challenging. Additionally, these tools often do not address intrinsic motivation or deeper cognitive engagement. | Design AI tools that incorporate adaptive challenges and rewards, fostering long-term engagement while also encouraging deeper cognitive processing and intrinsic motivation. |
| **Equity and Accessibility** | AI solutions aim to make education more accessible by offering personalized learning paths and support for students with diverse needs, including those with disabilities. | **Digital Divide**: AI tools are unevenly distributed, particularly in low-income or rural areas, and often lack optimization for students with special needs. | Invest in infrastructure to support equitable access to AI tools, and develop more inclusive AI technologies that cater to a wider range of disabilities and learning challenges. |

In this respect, although AI-driven solutions in the EdTech domain quickly deliver improvement with considerable value to teachers and students, there still remain some of the most important gaps to address in aspects of personalization, bias and fairness, equity and accessibility, and data privacy and security. Simply put, the gaps are indicative of opportunities: targeted activities on those gaps can help the EdTech sector create more effective, inclusive, and safe AI-driven solutions that will better serve the diverse needs of the education community.

### Knowledge Limitations and Lack of Real-Time Updates

Current chatbots, including LLMs, are trained once on large but static datasets. In other words, their knowledge is essentially frozen at the date of their last training or update. They don't access or use new information in real-time, which could leave them open to potential inaccuracies or strong lagging in responses related to recent developments.

Models and algorithms contributing to this limit:

- **Static Knowledge Base:** Because the training data of chatbots is too old, they are not capable of providing any proper information about recent events or updates.
- **No Real-time Learning:** The actual models lack adaptability since they do not support real-time learning without extensive retraining.
- **Limited Contextual Understanding:** This refers to the incapability of chatbots to understand subtly nuanced contexts, hence leading to irrelevant or inaccurate responses.

**Interaction Limitations**

- **Stateless conversations:** Chatbots look at each interaction as completely separate from any other, and can't build upon or learn from previous conversations.
- **Non-personalization:** Most of the time, chatbots provide very broad responses and are incapable of personalizing information according to users or individual preferences.
- **Limited Sensory Response:** Text-based user interfaces lower interaction richness, reducing the level of emotional engagement and making complex information more difficult to understand.

**Bias and Fairness**

- **Biased Training Data:** These chatbots can reflect social biases from training data and come up with somewhat prejudicial or unfair outputs while the development process ensues.
- **Lack of Diversity:** The percentage of certain groups is almost always low in the training data, hence leading to biased models.

These challenges collectively hinder the development of truly intelligent and human-like chatbots. Addressing these issues is crucial for creating more effective and reliable conversational AI systems

**Limited Sensory & Lack of Human Touch:**

One big limitation of chatbots—present language models, for that matter—is that their interaction is limited to text. This definitely puts an extremely high ceiling on the richness of interaction and richness of information—nuanced information—that humans naturally exchange through multiple sensory channels. Impersonal interactions, literally and figuratively devoid of human touch, lack emotional depth and forfeit critical nonverbal cues.

Key aspects of this problem:

- **Weak Sensory Infrastructures:** Without the ability to interpret nonverbal stimuli, a chatbot is weak in understanding and empathy.
- **Ineffective Communication:** The transmission of intricate information and building trust is an unmanageable issue without human-like interaction.
- **Artificial Experience:** The text-based interaction cannot replicate the feel of a human conversation.

# SARA AI

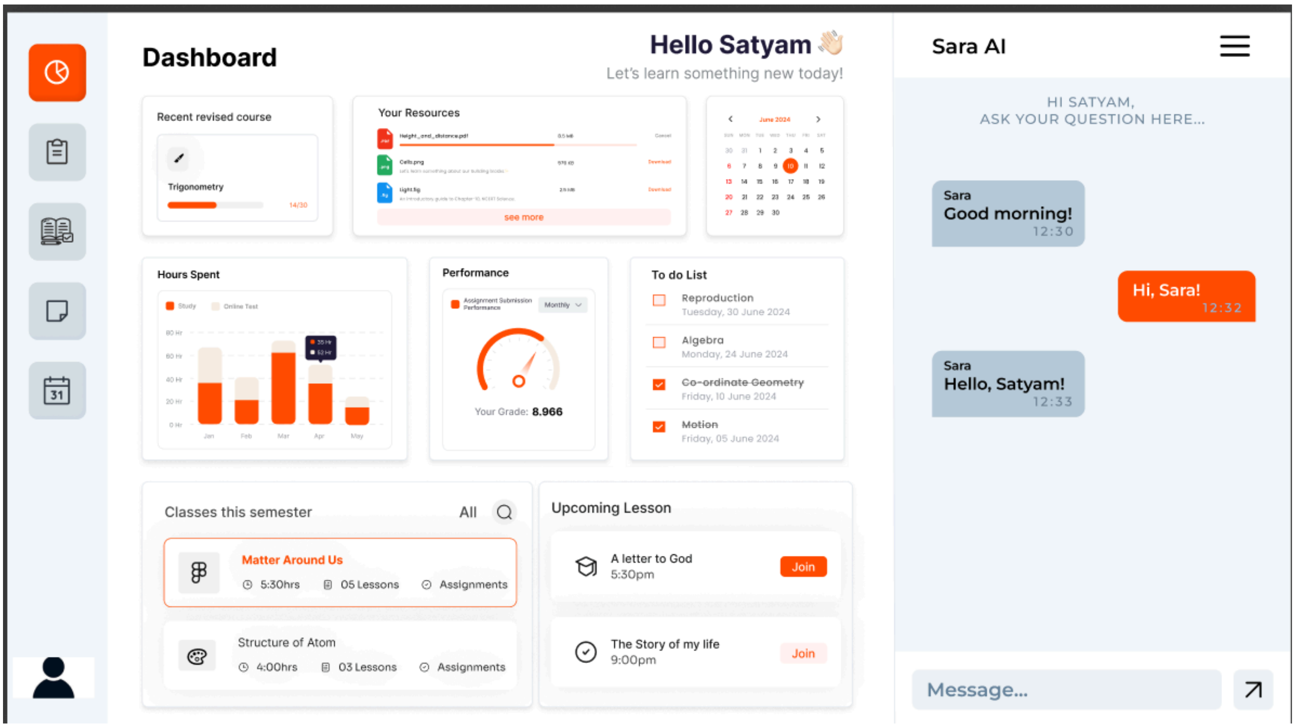
I have designed a Tutor AI, called **SARA** to provide a rich set of features that include the following: track student learning progress, an assignment copilot that will help in the completion of their assignments, a question generator to generate questions relevant to the learner's needs, test generation, and performance analysis from test results to indicate areas that need improvement.

Furthermore, to aid in learning, the AI is instilled with the ability to be able to search on the internet for learning materials to ensure that relevant materials are available to the learner. This system will be also integrated with the Gemma summary generator model for helping to create short summaries from big topics. I have trained the AI model in such a manner that, with all these functionalities, it executes optimally; hence versatile, very effective for students, and lecturers alike.

**Key Features:**

1) **Progress Tracker**: It keeps monitoring the evaluation process of testing students' learning journeys continuously and recognizing the areas in which a student is weak, thus realigning the progress targets based on the level of IQ and pace of learning of an individual. This kind of tailored monitoring helps provide support where students most need it.
2) **Assignment Copilot**: The Assignment Copilot will be equipped with generative AI capabilities in its direction of the student toward their everyday tasks, with deeper guidance, hints, and resources. This will facilitate the smooth completion of tasks and hence make learning easier.
3) **Test Creator**: Students should be enabled to upload their curriculum and set the tests to their syllabus. Relevant tests born out of AI enable students to properly test their knowledge and readiness with appropriate practice pertinent for each learner in their education.
4) **Performance Analysis**: The AI includes the detailed scorecards of strengths and weaknesses after every test. Beyond this, there would be further actionables in the form of insights and study plans given by the AI, helping students improve systematically over their performance and understanding of weak areas.
5) **Knowledge Extraction from the Web**: It retrieves and curates top-rated educational resources from the web pertaining to the queries of the user. It affords the luxury of deep learning, since study can be done more effectively with the help of quality materials.

**Intuitive User Interface:** SARA AI comes with an application-based user interface designed to make the integration of AI functions easier. It provides a single window for monitoring academic progress, planning, and getting help personalized to individual needs. This interface brings together the ease of use and high-AI capability by providing students with real-time updates and actionable insights while enjoying a seamless and engaging experience in learning.



## Dashboard

**Hello Satyam** 👋

Let's learn something new today!

Recent revised course

Trigonometry 14/30

Your Resources

- Height_and_distance.pdf 0.5 MB Cancel
- Cells.png 570 KB Download<br>Let's learn something about our building blocks
- Light.fig 2.5 MB Download<br>An introductory guide to Chapter-10, NCERT Science.

see more

June 2024

| SUN | MON | TUE | WED | THU | FRI | SAT |
|---|---|---|---|---|---|---|
| 30 | 31 | 1 | 2 | 3 | 4 | 5 |
| 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 13 | 14 | 15 | 16 | 17 | 18 | 19 |
| 20 | 21 | 22 | 23 | 24 | 25 | 26 |
| 27 | 28 | 29 | 30 | | | |

Hours Spent

Performance

Assignment Submission Performance — Monthly

Your Grade: **8.966**

To do List

- □ Reproduction<br>Tuesday, 30 June 2024
- □ Algebra<br>Monday, 24 June 2024
- ☑ ~~Co-ordinate Geometry~~<br>Friday, 10 June 2024
- ☑ ~~Motion~~<br>Friday, 05 June 2024

Classes this semester — All

- **Matter Around Us** — 5:30hrs · 05 Lessons · Assignments
- Structure of Atom — 4:00hrs · 03 Lessons · Assignments

Upcoming Lesson

- A letter to God — 5:30pm — Join
- The Story of my life — 9:00pm — Join

Sara AI

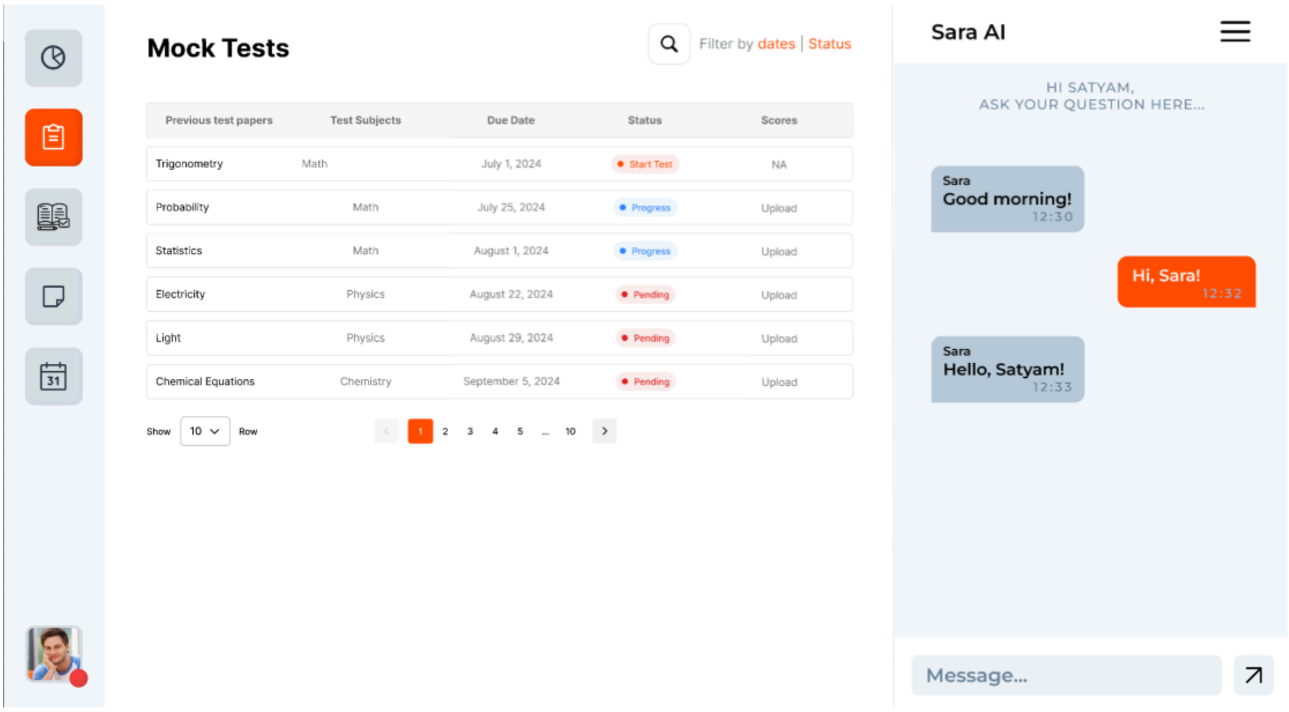
HI SATYAM,
ASK YOUR QUESTION HERE...

Sara: Good morning! 12:30

Hi, Sara! 12:32

Sara: Hello, Satyam! 12:33

Message...



## Mock Tests

Filter by dates | Status

| Previous test papers | Test Subjects | Due Date | Status | Scores |
|---|---|---|---|---|
| Trigonometry | Math | July 1, 2024 | ● Start Test | NA |
| Probability | Math | July 25, 2024 | ● Progress | Upload |
| Statistics | Math | August 1, 2024 | ● Progress | Upload |
| Electricity | Physics | August 22, 2024 | ● Pending | Upload |
| Light | Physics | August 29, 2024 | ● Pending | Upload |
| Chemical Equations | Chemistry | September 5, 2024 | ● Pending | Upload |

Show 10 Row

< 1 2 3 4 5 ... 10 >

Sara AI

HI SATYAM,
ASK YOUR QUESTION HERE...

Sara: Good morning! 12:30

Hi, Sara! 12:32

Sara: Hello, Satyam! 12:33

Message...

# Key Models & Algorithms Used

### RAG (Retrieval-Augmented Generation)

RAG combines retrieval and generation to enhance language generation. It retrieves relevant information from a knowledge base to improve the quality of generated responses.

**Usage in Code:** Though RAG is not directly reflected in any of the code pieces above, it is subtly implicit in the larger scheme of things within the project. For example, generating hints and resources or answering questions could make use of a RAG model if integrated. In that respect, SARA would be able to obtain extraneous textual data and utilize this information in trying to come up with more accurate and contextually relevant responses for assignments and questions.

### Meta-Llama-3.1

Meta-Llama-3.1 is a large language model used for various natural language processing tasks. In the provided code, it's used to generate responses and assist with assignments.

- Usage in Code:
  - Model Initialization: The FastLanguageModel from the unsloth library loads the pre-trained "Meta-Llama-3.1-8B" model. This model is configured with specific settings like maximum sequence length and 4-bit quantization for efficient processing.
  - Inference: The model is used in the assignment_copilot function to generate detailed steps and hints for solving assignments. It tokenizes the assignment description and uses the model to produce a response.
  - Text Streaming: For generating responses, the TextStreamer is used to handle output generation and streaming efficiently.

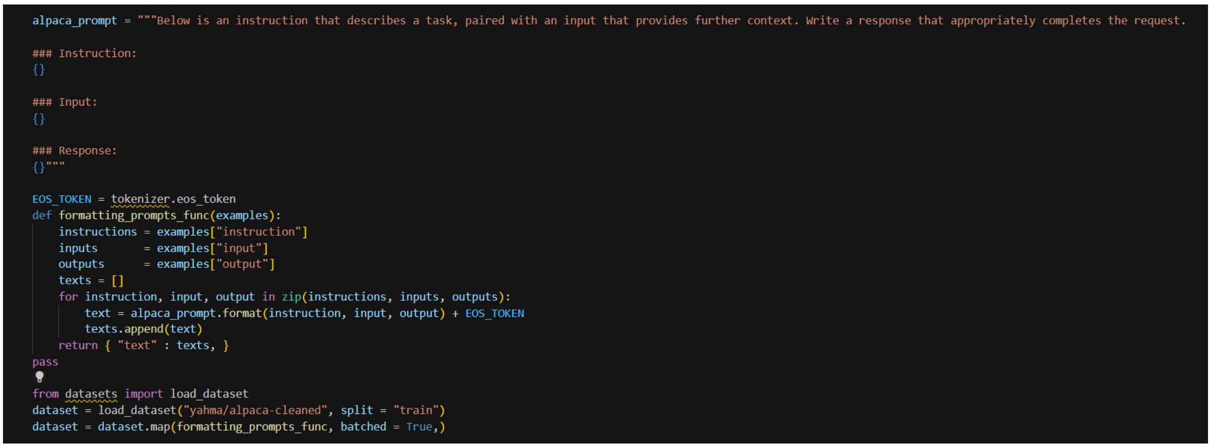
### NLP Techniques and Algorithms

1. **Tokenization**
   - Purpose: Tokenization converts text into a format that the model can process, typically splitting text into smaller units like words or subwords.
   - Usage: In the provided code, tokenization is handled by the tokenizer from the FastLanguageModel. It prepares the input text for the model to generate responses.
2. **Text Generation**
   - Purpose: Text generation involves producing human-like text based on a given prompt.
   - Usage: The Meta-Llama-3.1 model generates responses and hints for assignments. It uses the prompt to produce relevant text, streamlining the assignment assistance process.

3. **Evaluation Algorithms**
    - Purpose: Evaluation algorithms judge the quality of the answers and provide feedback.
    - Usage: The evaluate_answer function compares the user answer with the correct answer and provides feedback for better understanding.
4. **Prompt Engineering**
    - Purpose: Crafting prompts to get the desired responses from language models.
    - Usage: The standard template format of alpaca_prompt when generating the hints and resources for assignments makes sure that the model produces consistent and useful outputs.
5. **Model Fine-Tuning**
    - Purpose: Fine-tuning is used to adapt a pre-trained model to the concrete tasks to become better in these specific tasks.
    - Usage: Fine-tune the Meta-Llama-3.1 model on formatted educational data using this SFTTrainer to optimize it to produce correct relevant responses.
6. **Text Streaming**
    - Purpose: This is for quick handling and generation of vast volumes of text.
    - Usage: This is then fed into a TextStreamer, which processes the output from the model to allow for continuous generation of text, and finally controls the flow of responses.

**Code Snippets and Functionalities:**

In the project, this code formats training data using the alpaca_prompt template, which structures inputs, instructions, and outputs for the model. The formatting_prompts_func function standardizes the dataset and appends an End of Sequence (EOS) token to each entry, ensuring proper sequence recognition during training. The preprocessed "yahma/alpaca-cleaned" dataset is then formatted for consistency. This process is crucial for fine-tuning the language model, enhancing its ability to generate accurate and relevant responses for educational tasks.



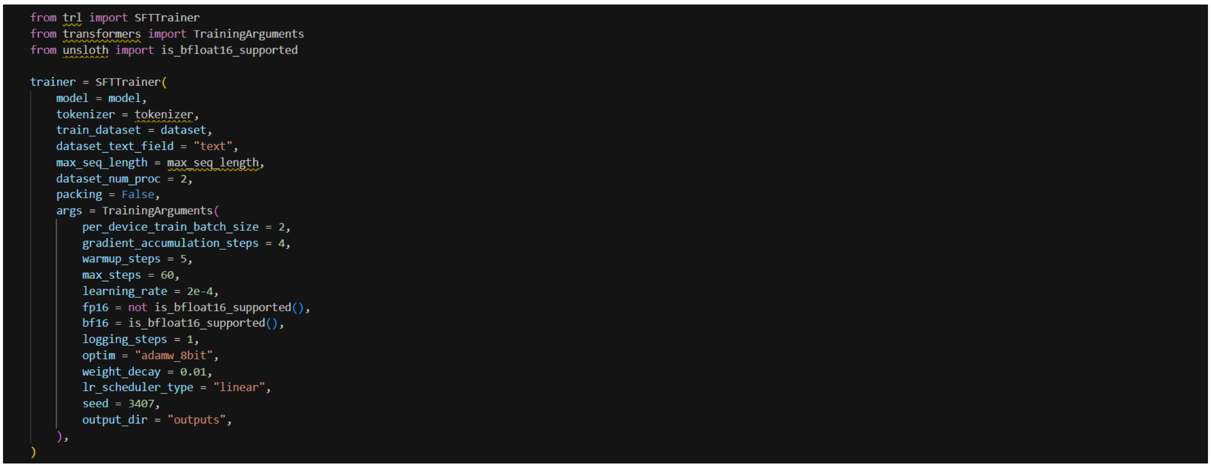
```
alpaca_prompt = """Below is an instruction that describes a task, paired with an input that provides further context. Write a response that appropriately completes the request.

### Instruction:
{}

### Input:
{}

### Response:
{}"""

EOS_TOKEN = tokenizer.eos_token
def formatting_prompts_func(examples):
    instructions = examples["instruction"]
    inputs       = examples["input"]
    outputs      = examples["output"]
    texts = []
    for instruction, input, output in zip(instructions, inputs, outputs):
        text = alpaca_prompt.format(instruction, input, output) + EOS_TOKEN
        texts.append(text)
    return { "text" : texts, }
pass

from datasets import load_dataset
dataset = load_dataset("yahma/alpaca-cleaned", split = "train")
dataset = dataset.map(formatting_prompts_func, batched = True,)
```

This script refines the language model of SARA AI using class SFTTrainer in the trl library. Configuration works with a pre-trained model in conjunction with a pre-trained tokenizer to train on a dataset formatted with educational prompts.

Key technical aspects include:

- **Batch Size and Gradient Accumulation**:  Accumulation: Batch size is 2, with gradient accumulation over 4 steps.
- **Learning Rate and Steps**: The learning rate is 2e-4, with a linear scheduler, 5 warmup steps, and 60 total steps.
- **Precision**: It uses bfloat16 if supported; otherwise, fp16 for reduced usage of memory.
- **Optimization:** It uses adamw_8bit for reducing memory usage.
- **Logging and Output**: Log every step and save in the directory mentioned.

This fine-tuning ensures that the model can generate responses of high quality and relevance to the context of the subject under discussion, performing exemplary in educational tasks.

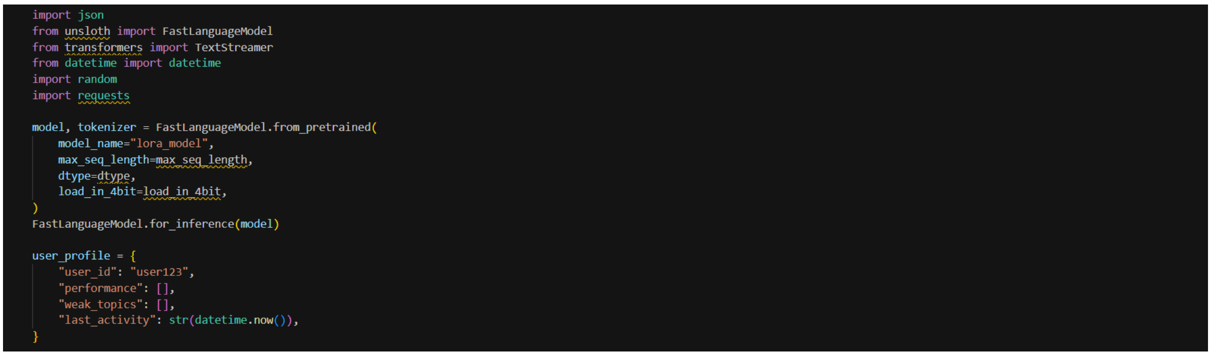


```
from trl import SFTTrainer
from transformers import TrainingArguments
from unsloth import is_bfloat16_supported

trainer = SFTTrainer(
    model = model,
    tokenizer = tokenizer,
    train_dataset = dataset,
    dataset_text_field = "text",
    max_seq_length = max_seq_length,
    dataset_num_proc = 2,
    packing = False,
    args = TrainingArguments(
        per_device_train_batch_size = 2,
        gradient_accumulation_steps = 4,
        warmup_steps = 5,
        max_steps = 60,
        learning_rate = 2e-4,
        fp16 = not is_bfloat16_supported(),
        bf16 = is_bfloat16_supported(),
        logging_steps = 1,
        optim = "adamw_8bit",
        weight_decay = 0.01,
        lr_scheduler_type = "linear",
        seed = 3407,
        output_dir = "outputs",
    ),
)
```

This is the code snippet in running the SARA AI for initialization of the environment toward response generation and user profile handling.

- **Model Loading**: The FastLanguageModel is loaded with the pre-trained "lora_model," configured with specific settings concerning sequence length, data type, and 4-bit quantization. The model is readied for inference: This allows it to generate responses based on user inputs.
- **User Profile Initialization**: Defining here a dictionary user_profile where every single user has to be kept. This will contain:
    - user_id: An identification through which a user can be identified.
    - performance: This list will be used to store how the user has performed over the period of use
    - weak_topics: Empty list to initialize the areas upon which the user is falling weak
    - last_activity: Timestamp to denote the last activity of that user, initialized to date and current time.

This setup allows the AI to generate responses and track user interactions effectively, enhancing its ability to provide personalized educational support and monitor progress.



```
import json
from unsloth import FastLanguageModel
from transformers import TextStreamer
from datetime import datetime
import random
import requests

model, tokenizer = FastLanguageModel.from_pretrained(
    model_name="lora_model",
    max_seq_length=max_seq_length,
    dtype=dtype,
    load_in_4bit=load_in_4bit,
)
FastLanguageModel.for_inference(model)

user_profile = {
    "user_id": "user123",
    "performance": [],
    "weak_topics": [],
    "last_activity": str(datetime.now()),
}
```

This code snippet provides personalized assignment support and interaction using the Tutor AI project:

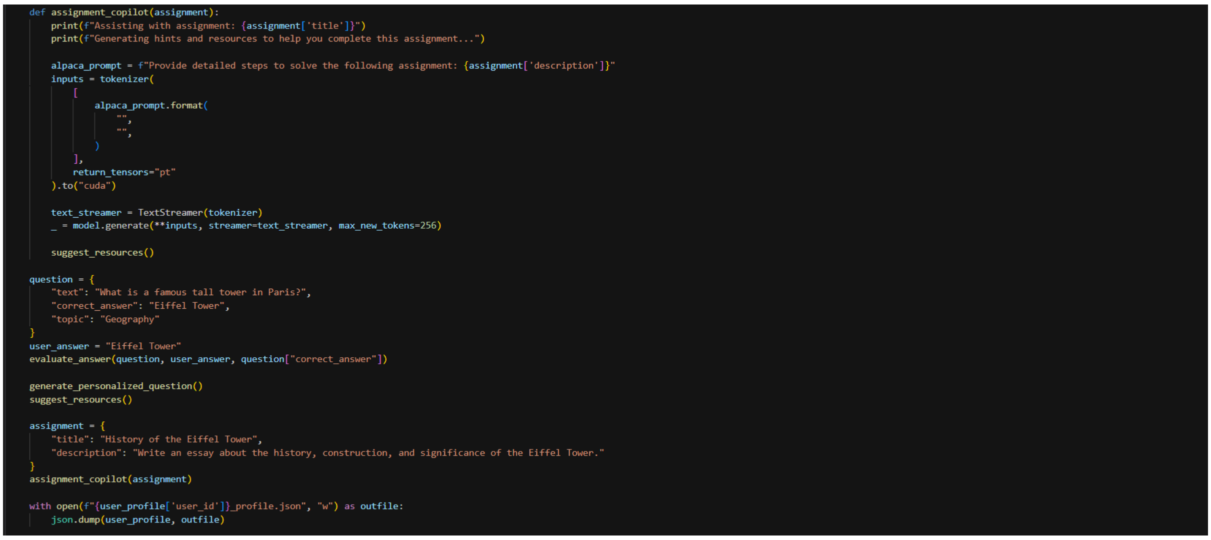
**Deduction control:** The assignment_copilot function incorporates the model to offer hints and resources through a provided assignment description for an even better experience during the completion of assignments.

**Question evaluation:** Offers feedback regarding the user's response and interaction as well as personal question creation.

**Assignment processing:** Processes a sample assignment and gives some guidance on how to complete the essay.

**User Profile Management:** This saves a user's profile, for example, his performance and activity data, in a JSON file to track progress.

This setup improves personalized support and progress monitoring within the Tutor AI system.



```
def assignment_copilot(assignment):
    print(f"Assisting with assignment: {assignment['title']}")
    print(f"Generating hints and resources to help you complete this assignment...")

    alpaca_prompt = f"Provide detailed steps to solve the following assignment: {assignment['description']}"
    inputs = tokenizer(
        [
            alpaca_prompt.format(
                "",
                "",
            )
        ],
        return_tensors="pt"
    ).to("cuda")

    text_streamer = TextStreamer(tokenizer)
    _ = model.generate(**inputs, streamer=text_streamer, max_new_tokens=256)

    suggest_resources()

question = {
    "text": "What is a famous tall tower in Paris?",
    "correct_answer": "Eiffel Tower",
    "topic": "Geography"
}
user_answer = "Eiffel Tower"
evaluate_answer(question, user_answer, question["correct_answer"])

generate_personalized_question()
suggest_resources()

assignment = {
    "title": "History of the Eiffel Tower",
    "description": "Write an essay about the history, construction, and significance of the Eiffel Tower."
}
assignment_copilot(assignment)

with open(f"{user_profile['user_id']}_profile.json", "w") as outfile:
    json.dump(user_profile, outfile)
```

# Future Scope

There is huge potential for growth and impact on the educational landscape with the SARA AI project. Further future developments in this area could concern the following: **Enhanced Personalization**: Front-load more sophisticated machine learning algorithms with an increased range of sources of data, which would provide SARA with the capability to offer increasingly tailored educational experiences that adapt to the particular learning styles, preferences, and needs of each student.

**Multilingual Support**: Extending the AI to support multiple languages would open up the tool to a global audience, bridge the gap of the use of languages, and provide quality education to students not fluent in English.

**AI-driven Content Creation**: Further releases could allow creation by the AI of tailor-made learning materials, interactive modules, and multimedia—further enhancing the learning experience to serve these diverse learning modalities.

**AR and VR integration**: Use of AR and VR technologies would afford students an immersive learning experience, especially in subjects that are hands-on or require visualization, like science experiments and historical simulations.

**Improvements in accessibility**: Newer versions can make the system more accessible to more students with special needs, either visually, auditorily, or even cognitively impaired, by using specialized interfaces and adaptive learning strategies on content formats that are inclusive.

**Collaboration and Scalability**: This AI could be further developed as a collaborative learning environment where students work on projects together, share resources, and learn from each other through AI-driven insights and feedback.

**Ethical AI and Great Data Privacy**: On growing the system, there needs to be an enormous component of ethical AI practices and robust data privacy mechanisms that give trust the technology is being used responsibly and transparently.

**Integration with Educational Institutions**: SARA AI could integrate well into existing educational systems, providing support to teachers and institutions through automated administrative tasks and performance analytics in real-time, and recommending improvements in the curriculum based on data-driven insights.