Mainstay's Use of Generative AI

Mainstay AI Positioning Statement / Philosophy

Artificial Intelligence (AI) is a transformational technology with the potential to meaningfully impact higher education. At Mainstay, we believe AI's best use is to power conversations that help people access and achieve higher learning. Our human-centered, AI-supported approach to conversation is designed to coach learners with trusted, personalized, and empathetic guidance while opening pathways to human support. That means that information must be accurate, reliable, and presented in a way that encourages each learner to take the next step in their journey.

Mainstay's generative AI features also support a more robust user experience for learners, including enhanced knowledge matching that provides more accurate responses to student questions, expanded use of approved content, and a more conversational experience between learners and the chatbot.

Mainstay AI Data Security Statement

Mainstay does not perform any conclusions or assessments about individual students. Our platform uses AI for operational efficiency (examples include Knowledge Base design, Fuzzy Question generation) or to make suggestions related to questions during the escalation processes. We do not utilize AI in any capacity to make determinations related to an individual student or the information to which we have access.

Additionally, Mainstay anonymizes Personally Identifiable Information (PII) related to a student prior to any use of AI. For example, if a student sends a message with a phone number, email address, physical address, SSN, bank account number, sensitive document ID like a passport or taxpayer ID, that information is anonymized before using any generative AI service or API.

Lastly, Mainstay maintains the following certifications to ensure that we maintain and operate at the highest levels of data security protection and privacy, including SOC2 and TX–RAMP.

What Type of AI does Mainstay use?

Mainstay uses both Natural Language Processing (NLP) and Generative Pre-trained Transformers (GPT), to augment and enhance our partners' ability to conduct meaningful conversations at scale. We believe this provides the best quality and accuracy for our partners and their students.

Mainstay utilizes Natural Language Processing in the following manner:

- To assist in matching incoming learner questions with pre-curated answers within the knowledge base
- To find relevant information from knowledge sources provided by the organization to assist with answering a student question

Mainstay utilizes GenAI in the following manner:

- To offer suggestions for response during live chat scenarios and escalation emails
- To summarize conversations for advising staff to prepare for interactions with students
- To generate answers to questions from defined knowledge sources provided by the organization using the platform
- To offer learners more fluid, unstructured texting and voice conversations where learners are free to ask questions outside of the normal branching logic of a Script, but without deviating from the relevant topic

Is private or public technology used for Generative AI?

Mainstay leverages OpenAI, which is a publicly available service, but utilizes proprietary Mainstay-owned code for data anonymization and processing.

Mainstay maintains an Enterprise Data Agreement with OpenAI which does not allow them to use any of the content sent over by Mainstay to be used in training their models.

How does Mainstay ensure privacy policies around learner data is upheld?

Mainstay anonymizes any Personally Identifiable Information (PII) related to a student prior to any use of AI. For example, removing phone numbers, email addresses, physical addresses, SSNs, bank account numbers, sensitive document IDs like a passport or taxpayer IDs that might have been submitted during a conversation.

Additionally, partners have control over what data is imported into the Mainstay platform, as well as which data points are available as context for the Generative AI.

How do we prevent unauthorized data leakage?

Mainstay is SOC2 and TX-RAMP certified. Additionally, we perform quarterly penetration testing of our platform to ensure, to the best of our ability, that any issues that might compromise partner data are identified and remedied. Partners have access to our SOC reports at https://trust.mainstay.com

Additionally, through user roles and permissions settings, partners can control who on their team has access to what data and even which learners' data.

Here are some typical concerns that partners have and how Mainstay mitigates these risks:

Risk	Mitigation measures
Sharing of sensitive data or student PII	Student's personally identifiable information (PII) or other sensitive content will not be shared directly with the AI technology leveraged by the platform (Open AI). Mainstay has developed a detection algorithm to identify and mask any PII shared by a student during their conversation with the chatbot. (See below: "Mainstay Platform Security Measures"). As mentioned above, partners have control over what data is imported into the Mainstay platform, as well as which data points are available as context for the Generative AI.
Bias	Measures taken to identify, manage, and mitigate bias in the chatbot's training data and algorithms include consistent reviews of the models, prompts used to generate responses, and regularly training employees on bias mitigation tactics to ensure effective human oversight. Additionally, much of the research work that Mainstay has conducted with its partners is specifically focused on generating content that avoids bias, which has further helped to ensure the efficacy of our training.
Content validity and reliability	Mainstay deploys standard prompts that instruct the Generative AI to: • Only use the content provided by the partner organization to answer learner's questions • Not agree to take action based on any student instructions • Only answer questions related to the topics for the use case that the platform has been deployed to handle (i.e. college access, college success, post-secondary pathways, etc.) Additionally, partner organizations are given the option to test and provide additional instructions for the Generative AI to follow in order to adhere to partner-specific policies, etc.

	In addition to those guardrails, when the Mainstay platform processes partner-approved content (websites and documents) it automatically ensures those are up-to-date (or warns partners that static documents may be out of date), and then searches that data for relevant information for the student's question. There is a safety check on URLs to ensure the generative AI is only giving links that are valid and sourced from the provided context.
Platform security and resilience	Mainstay ensures that the chatbot can withstand unexpected adverse events or changes in usage. The Mainstay platform's architecture is designed for high-availability and high-security. All of the core components are clustered, and the databases leverage read replicas to ensure swift cut-over in the event of an unexpected network or hardware failure. All data is encrypted in-motion and at rest, following industry standards and best practices. Mainstay has earned SOC 2 Type II compliance certification and is TX-ramp certified, demonstrating their commitment to data security and privacy. Other data security details to note: ● Mainstay is FERPA compliant; FERPA data is always encrypted with a minimum 128-bit SSL (TLS) in motion and encrypted at rest in partner's database. ● Mainstay is able to comply with data-removal policies and remove any personally identifiable information upon request. ● Mainstay is able to supply an internal data breach policy and data privacy standards (available on request). ● Mainstay commits to quarterly penetration testing of its web service and servers.
Transparency	Partner organizations are encouraged and guided by Mainstay best practices to ensure that learners understand how the chatbot works. All proactive content and knowledge base answers are designed to clearly communicate to learners that they are interacting with an AI chatbot and not a human. When needed or requested, learners are given the option to be connected with a human for help or the escalations are automated for sensitive situations.

Mainstay Platform Security Measures

The Mainstay team has established processes in place to ensure that learner data shared within their platform, and particularly any sensitive data or PII, remains secure and protected at all times.

This enables the team to operate generative-AI based tools in the Mainstay platform without compromising student safety.

The Mainstay platform is able to identify any Personally Identifiable Information (PII) by using a detection algorithm that runs only on Mainstay's computers. When we identify PII such as names, birthdates, or addresses, we substitute this information with a placeholder value prior to sending the messages over to OpenAI / ChatGPT.

What a contact messages in	What OpenAI sees
my ss number is 123-45-6789, and my child's is 987-65-4321	*my ss number is [Social Security #1], and my child's is [Social Security #2]*
My credit card is 1234-5678-9010-1112	*My credit card is [Credit Card #1].*
My email is james@mainstay.com.	*My email is [Email #1].*

This strategy keeps information private, but also allows the AI to understand the full context of the message when responding.

If an answer does need to incorporate a piece of PII, say, a date, the AI is instructed to respond using the placeholder it received. Once we get the message back, we'll substitute the placeholder out for the real date, keeping the interaction seamless.

Mainstay Generative AI Features Overview

Each Generative AI feature of the Mainstay platform is described in detail on a separate support article that is listed in this article:

Generative AI Settings & Bot Personality Controls