INSTRUCTURE

INTELLIGENT
INSIGHTS

ARCHITECTURE
OVERVIEW

Engineering, Security, and Operations

October 2024

Table of Contents

◆ INSTRUCTURE

Introduction

Intelligent Insights

In today's data-driven world, the ability to analyze, and derive intelligent and meaningful insights from data has become a cornerstone of success across education. As institutions strive to improve student outcomes, personalize learning experiences, and optimize resource allocation, the role of data analytics has become increasingly crucial.

At Instructure, we recognize the power of data lies not just in its collection, but in the intelligent analysis that transforms raw information and statistics into actionable insights. This is why Intelligent Insights was born—a sophisticated product that harnesses the latest advancements in artificial intelligence and machine learning to provide educators and administrators with unprecedented visibility into student performance, engagement, and potential.

At its core is the innovative AI-powered 'Ask Your Data' feature that allows users to interact with their data in natural language. This means educators can ask complex questions about student trends, correlations, and predictions, receiving instant, data-backed responses that would traditionally require extensive manual analysis.

By leveraging this intelligent analytics tool, educational institutions can uncover hidden patterns, identify at-risk students early, tailor interventions, and make data-informed decisions that significantly impact student success.

The following supplemental document describes the Intelligent Insights architecture for those curious technical types who love getting into the detail of just how we improve educational outcomes for all.

Hosting

Instructure's product family, including Intelligent Insights, is hosted in the cloud by Instructure and delivered over the internet through the world's most trusted public cloud provider, Amazon Web Services (AWS). The basic building blocks of AWS include services such as Elastic Compute Cloud (EC2), Elastic Load Balancing (ELB), Auto Scaling Groups (ASG), Simple Storage Service (S3), Elastic Block Store (EBS), Virtual Private Cloud (VPC), Simple Email Service (SES), and Identity and Access Management (IAM). We also use advanced AWS platform capabilities including Amazon Kinesis, AWS Lambda, AWS Fargate, AWS Elastic Kubernetes Service ("EKS"), and Amazon Relational Database Services ("RDS"). Instructure's products are designed to make full use of the real-time redundancy and capacity capabilities offered by AWS, running across multiple availability zones in regions throughout the world. Primary storage is provided by Amazon S3, which is designed for durability exceeding 99.99999999%.

Within Intelligent Insights, the Ask Your Data functionality is hosted by a third-party vendor which utilizes Google Cloud Platform (GCP). GCP is a leading cloud provider known for its advanced infrastructure and innovation in areas like machine learning, artificial intelligence, and data analytics. Google's expertise in these domains is a significant advantage to delivering the Ask Your Data functionality within Intelligent Insights.

Hosting Regions

For US customers, Intelligent Insights uses two Amazon Web Services (AWS), ensuring that client data is not stored outside of the United States:

- AWS US East (Northern Virginia, us-east-1)

- AWS US West (Oregon, us-west-2)

For international clients, we use the following AWS regions:

- AWS Canada Central (Montreal, ca-central-1)

- AWS EU West (Ireland, eu-west-1)

- AWS EU Central (Frankfurt, eu-central-1)

- AWS Asia Pacific (Sydney, ap-southeast-2)

- AWS Asia Pacific (Singapore, ap-southeast-1)

In each region we operate, we utilize three (3) Availability Zones (AZ) for redundancy.

Security Program

Intelligent Insights is included as part of Instructure's robust information security program that runs on a continuous, PDCA-cycle. It was created based on guidance provided by ISO/IEC 27000:2018 and controls described in ISO/IEC 27001:2013 and is managed by Instructure's Chief Information Security and Privacy Officer (CISPO). The security program is attested to by a number of current security certifications including ISO 27001, SOC 2, SOC 3, and UK Cyber Essentials Plus.

Intelligent Insights has a software development lifecycle (SDLC) that incorporates secure coding practices and controls. All code goes through a developer peer-review process before it is merged into the code base repository. The code review includes security auditing based on the Open Web Application Security Project (OWASP) secure coding and code review documents (including the OWASP Top Ten) and other community sources on best security practices.

Instructure's Security Team regularly performs vulnerability scans on Intelligent Insights using a number of internal and external tools and techniques and we make available to customers the results of our annual third-party penetration tests because we believe that being open about all things - security included - enables us to build the best possible products.

In addition to these measures, the Amazon Web Services infrastructure on which Intelligent Insights is hosted has a variety of formal accreditations. Some of the many certifications include:

DoD SRG • FedRAMP • FIPS • IRAP • ISO 9001 • ISO 27001 • ISO 27017 • ISO 27018 • MLPS Level 3 • MTCS • PCI DSS Level 1 • SEC Rule 17-a-4(f) • SOC 1 • SOC 2 • SOC 3 • UK Cyber Essentials Plus

For additional information about Amazon Web Services security certifications and standards compliance, please refer to:

http://aws.amazon.com/security
http://aws.amazon.com/compliance/

Product Security

The following is an overview of Intelligent Insights product security measures:

- All data is encrypted in transit with TLS v1.2 or higher.

- All data is stored at rest within AES-256-bit-encrypted volumes.

- The API is secured by OAuth.

- All environments are deployed into an AWS Virtual Private Cloud (VPC) within secure private networks. NAT Gateways are used to ensure that instances do not have routable IP addresses. Each component is protected by a security group with an appropriate, restrictive rule set. The only device that has access to the public internet is the Elastic Load Balancer (ELB).

- A Web Application Firewall further protects the application from potential exploits, filtering traffic before it even reaches the ELB.

- Host-based vulnerability scans are performed at least quarterly on all external-facing systems. Critical and high vulnerabilities are tracked to remediation.

- Instructure maintains a Data Protection Policy that is reviewed at least annually.

- Instructure is compliant with the EU's national data privacy and protection law, the General Data Protection Regulation ("GDPR").

Personally Identifiable Information (PII)

Personally Identifiable Information (PII) comes from customer data within Canvas LMS. This is via the datasets (tables) of Canvas Data 2. Typically, the only PII captured during use from a user is the user's full name, display name and email address, as well as the Canvas LMS instance's LMS id (UUID) and the LTI id (lti_lms_id), identifying the user in Canvas. More information on PII can be found at: https://www.instructure.com/policies/product-privacy-policy.

More information on the Canvas Data 2 (DAP) datasets can be found at:

https://api-gateway.instructure.com/doc/

Accessibility

Ensuring an accessible and pleasant experience to all users, regardless of disability, is a key focus of Instructure and we are committed to W3C's Web Accessibility Initiative, Section 508 guidelines, and ensuring our products are accessible and inclusive.

We target WCAG 2.1 AA standards in our development and have established our own NFR ladders regarding accessibility. We also recognize that software accessibility is often a journey more than it is a destination, and so we acknowledge and commit to continuous improvement.

Accessibility is ingrained from Product Design through Engineering, and we have dedicated accessibility staff, subject matter experts, and Product Managers. Internally, from all aspects of the business, we maintain a number of accessibility advocates who meet regularly to raise issues, discuss trends, and maintain skills and knowledge in the accessibility space.

Accessibility training is conducted for all engineering staff. Instructure is committed to all new features being fully accessible before being deployed to production. Current commitments to Accessibility standards are documented publicly on our website.

Ask Your Data

Introduction

The Ask Your Data functionality of Intelligent Insights is a conversational AI which offers institutions the ability to query their Canvas data with natural language, using AI and large language model (LLM) technologies. Users can interact and query their data on the fly and quickly visualize them, without any technical knowledge.

Data teams and BI Analysts can quickly explore Canvas data, generate queries, and improve internal BI systems.

Third-Party Provider

Doowii is a third-party provider that powers the 'Ask Your Data' feature of Intelligent Insights. This integration provides institution administrators with access to advanced data analytics capabilities directly within the Canvas LMS environment. The following relates specifically to Ask Your Data and our third-party processor, Doowi.

Data flows from Canvas LMS to Doowii through the Canvas DAP client, which is separate from the Canvas LTI (Learning Tools Interoperability) integration and associated APIs. The Canvas DAP client ensures seamless data exchange and integration between the two platforms. The data is updated every 6 hours.

Hosting

Customer data for Ask Your Data will always be stored and hosted in the same region as a customer's Canvas data.

For US customers, Ask Your Data customer data is hosted in the following region by Google Cloud Platform (GCP):
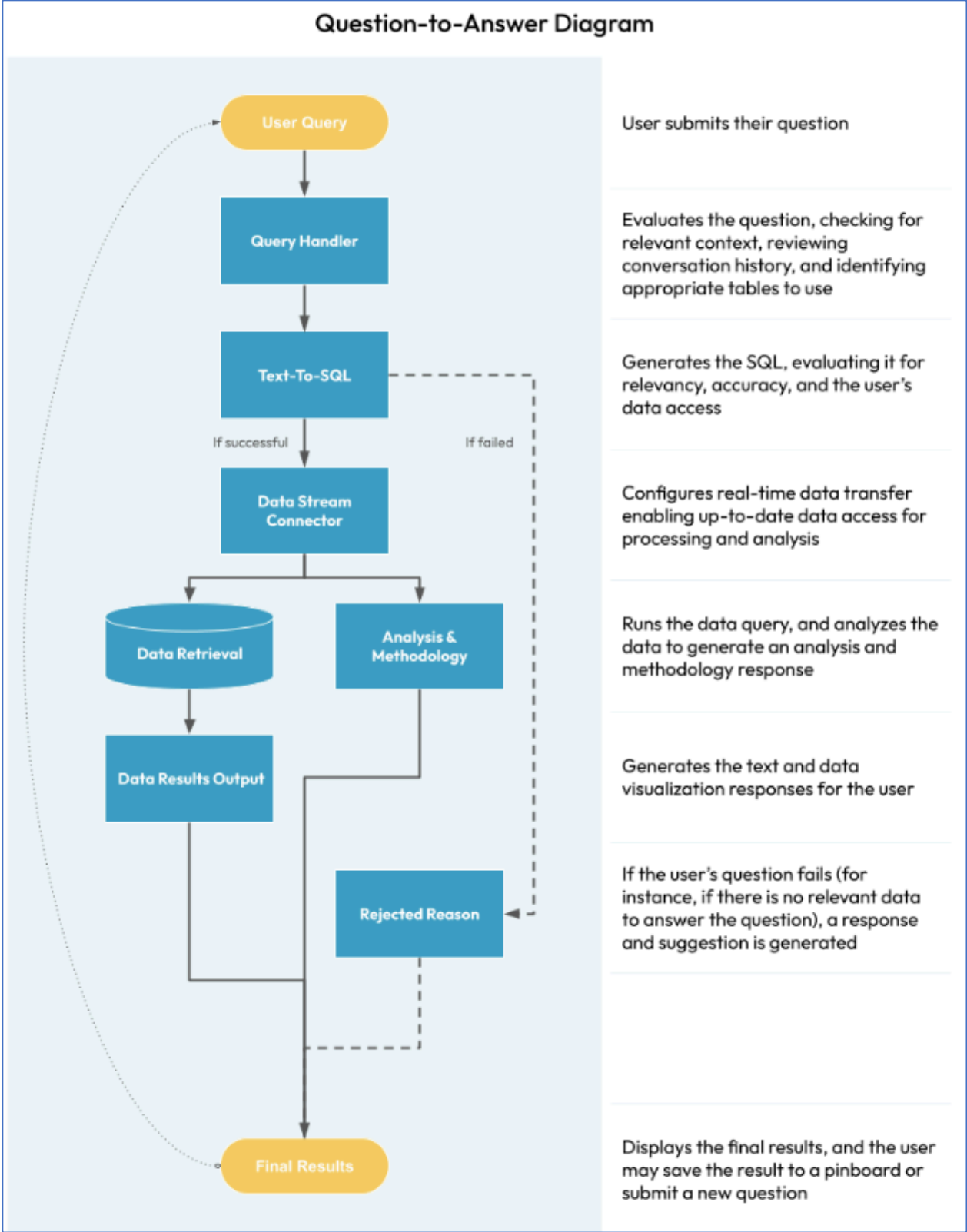
- GCP US, California (us-west2)

For international clients, we use the following GCP regions:

- GCP Canada, Montreal (northamerica-northeast1)

- GCP EU, Frankfurt (europe-west3)

- GCP Asia Pacific, Singapore (asia-southeast1)

- GCP Asia Pacific, Sydney (australia-southeast1)

Architecture Diagram

Question-to-Answer Diagram

User Query — User submits their question

Query Handler — Evaluates the question, checking for relevant context, reviewing conversation history, and identifying appropriate tables to use

Text-To-SQL — Generates the SQL, evaluating it for relevancy, accuracy, and the user's data access

If successful / If failed

Data Stream Connector — Configures real-time data transfer enabling up-to-date data access for processing and analysis

Data Retrieval / **Analysis & Methodology** — Runs the data query, and analyzes the data to generate an analysis and methodology response

Data Results Output — Generates the text and data visualization responses for the user

Rejected Reason — If the user's question fails (for instance, if there is no relevant data to answer the question), a response and suggestion is generated

Final Results — Displays the final results, and the user may save the result to a pinboard or submit a new question

AI 'Nutrition' Facts

AI Nutrition Facts

Ask Your Data in Intelligent Insights

Description

AI-powered query tools enable Canvas admins to ask a natural language question and receive a textual or visual response with a detailed explanation of the methodology.

Model & Data

Base Model	Chat GPT-4o, GPT-3.5 Turbo, and Claude 3.5 Sonnet
Trained with User Data	No
Data Shared with Model	Other

Prompt, Summary Statistics

Privacy & Compliance

Data Retention
Until account termination

Data Logging	Logs Data

Methodology description, ability to view the generated SQL

Regions Supported	NORAM, EMEA, APAC, LATAM

PII	Exposed

Prompt, Summary Statistics

Outputs

AI Settings Control	Yes
Human in the Loop	Yes

Methodology description, ability to view the generated SQL

Guardrails
Question scoped to domain only, highly ambiguous terms ask for clarification

Expected Risks
Incorrect interpretation of the question. Inaccurate SQL may be generated. Suggested questions may not be relevant.

Intended Outcomes
Provide accurate data retrieval and analysis through natural language.

Instructure has developed nutrition fact labels for AI-enabled products to increase transparency and improve decision making.

Language Model

Ask Your Data uses several large language models to provide the service. These include:

- GPT4 turbo preview
- GPT4o
- Llama 2

- GPT3.5 turbo
- Fine-tuned GPT
- GPT4

Privacy (PII)

PII

Ask Your Data utilizes all data that is available within the Canvas Data 2 Data Access Platform datasets including Canvas LMS and Canvas Catalog data. For a complete list of PII that is stored and processed to provide Ask Your Data functionality, please refer to the Data Access Platform Datasets resource at:

https://api-gateway.instructure.com/doc/

Data Collection

Doowii ingests Canvas customer data from the via the Data Access Platform (DAP) API. Customer metadata includes information such as emails and user activities within the same region that a customer is located in. Doowii uses this data to answer Ask Your Data questions and provide features such as chat threads. Questions and answers from Ask Your Data are collected for analytical purposes as well as product fine-tuning.

Deletion Requests

Doowii removes all the customer data within a week once we are notified that the customer decides to no longer use the Ask Your Data functionality.

Termination

At termination, Doowii will remove customer data after 30 days.

Security

Instructure has completed an extensive security audit and review of our third-party processor that delivers the Ask Your Data functionality within Intelligent Insights. Instructure takes responsibility for the acts and omissions of our sub-processors contractually and ensures that they meet our security/privacy standards.

The following is an overview of Ask Your Data security measures:

- All data is encrypted in transit.

- All data is encrypted at rest.

- The API is secured by OAuth.

- No customer data is used to train third-party provider language models (see *Nutrition Facts*)

For additional information about cloud provider security certifications and standards compliance, please refer to:

http://aws.amazon.com/security
http://aws.amazon.com/compliance/, and
https://www.google.com/about/datacenters/data-security/
https://cloud.google.com/security/compliance/compliance-reports-manager

Conclusion

It should now be clear to all—the future of impactful educational lies in our ability to harness data effectively. Intelligent Insights isn't just a tool, it's your partner in achieving exceptional student outcomes. With the AI-powered 'Ask Your Data', you're not just looking at data; you're engaging with it, uncovering insights that were previously hidden, and making decisions with confidence.

This is only the first step in embracing the transformative potential of Artificial Intelligence in education. As we continue to innovate, the ability to predict student challenges, personalize learning paths, and provide effective learning for everyone will become even more sophisticated. In today's educational landscape, this level of insight isn't just an advantage – it's a necessity for staying ahead.

The era of AI-driven, data-informed education is here, and it promises a future where every student can reach their full potential.

INSTRUCTURE