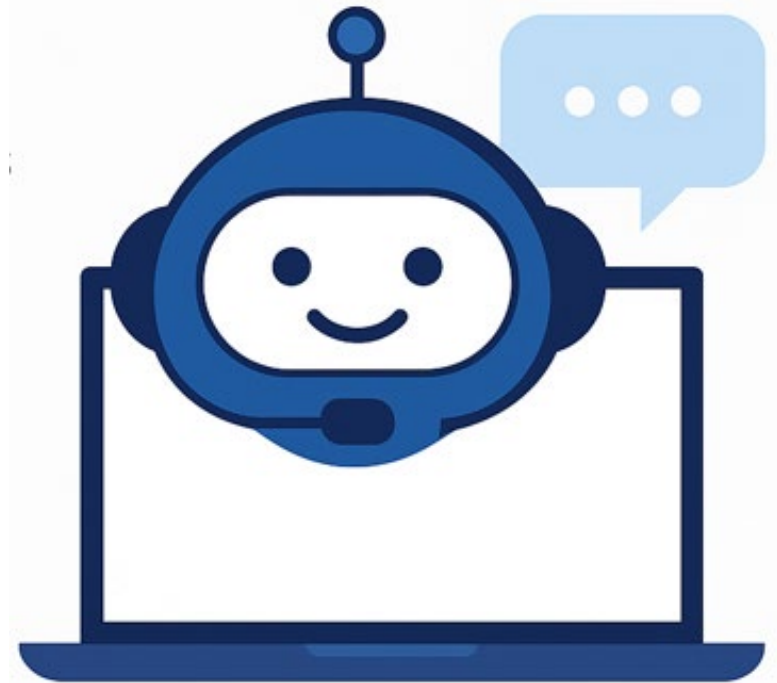


NJIT

New Jersey Institute
of Technology

Designing an AI-Powered IR Chatbot - fostering cross-departmental collaboration and data access

Xilin Zhang, PhD
Dong Dinh, PhD
Eugene Deess, PhD



Team



- **Xilin Zhang** Associate Director of Institutional Research
- **Keywords:** AI Product Development, System Design



- **Dong Dinh** Institutional Data Analyst
- **Keywords:** LLM Model, System Engineering



- **Perry Deess** Executive Director of Institutional Effectiveness
- **Keywords:** User Experience, Higher Education Management

Where Is That Information When You Need It?

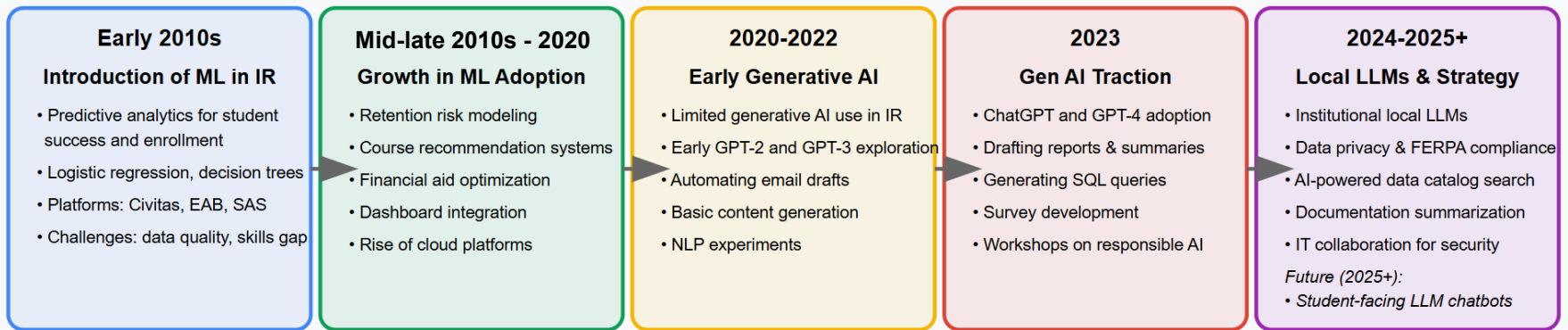


- Higher education staff spend an average of 8.8 hours weekly searching for information (Microsoft, 2023)
- Institutional data requests are repetitive yet require manual retrieval each time (Yanosky R., 2009)

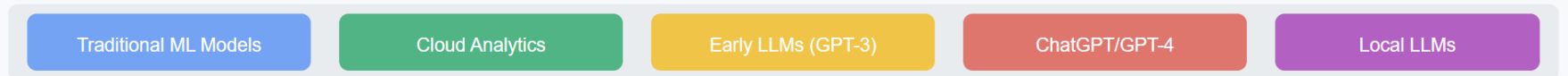
• Microsoft, 2023. Will AI Fix Work? [Source: <https://www.microsoft.com/en-us/worklab/work-trend-index/will-ai-fix-work>]
• Yanosky, R., 2009. Institutional Data Management in Higher Education. ECAR Key Findings. EDUCAUSE.

History

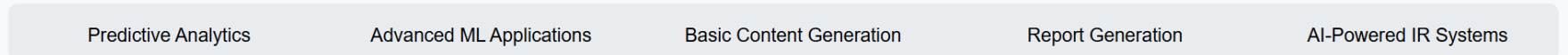
Evolution of AI in Institutional Research (2010-2025+)



Technology Evolution



Primary Use Cases

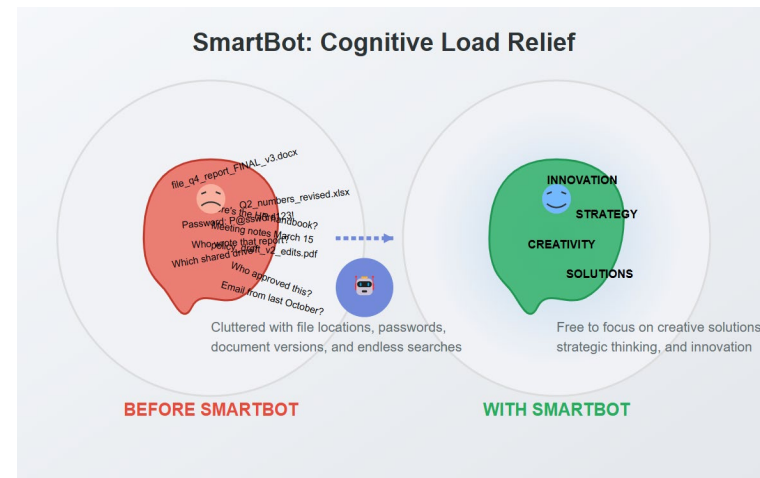
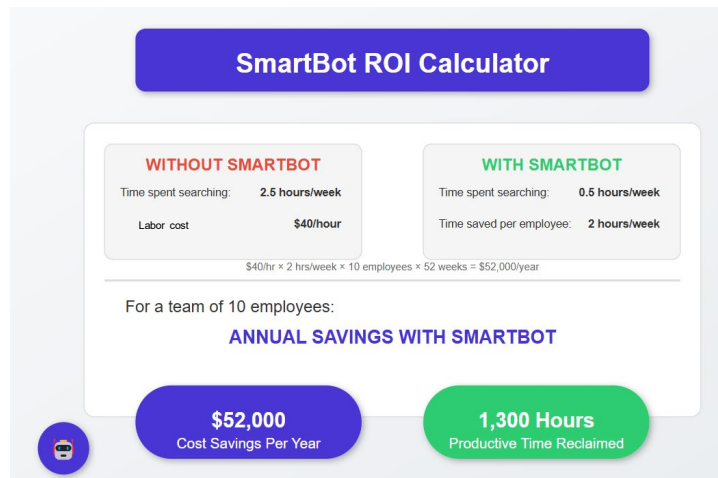
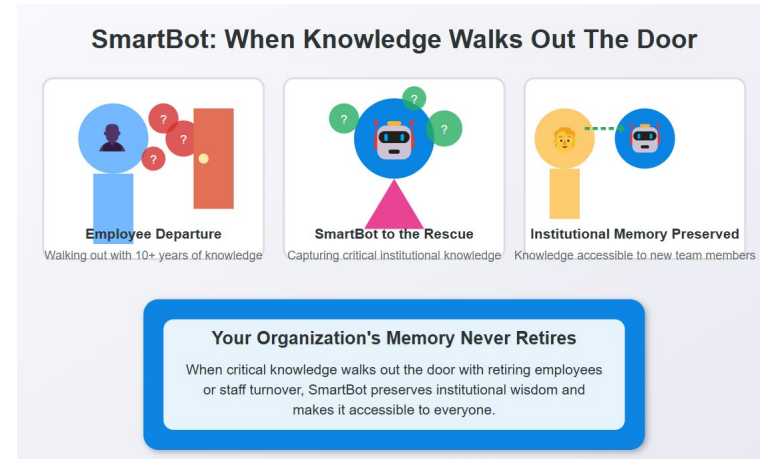
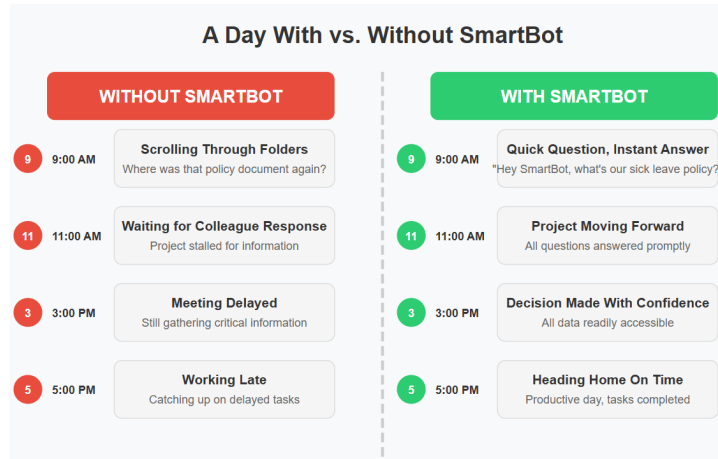


- Jones, D. and Ross, L. E. (2024). The of Use of Generative Artificial Intelligence in Institutional Research/Effectiveness. Brief. Association for Institutional Research. <http://www.airweb.org/community-surveys>.

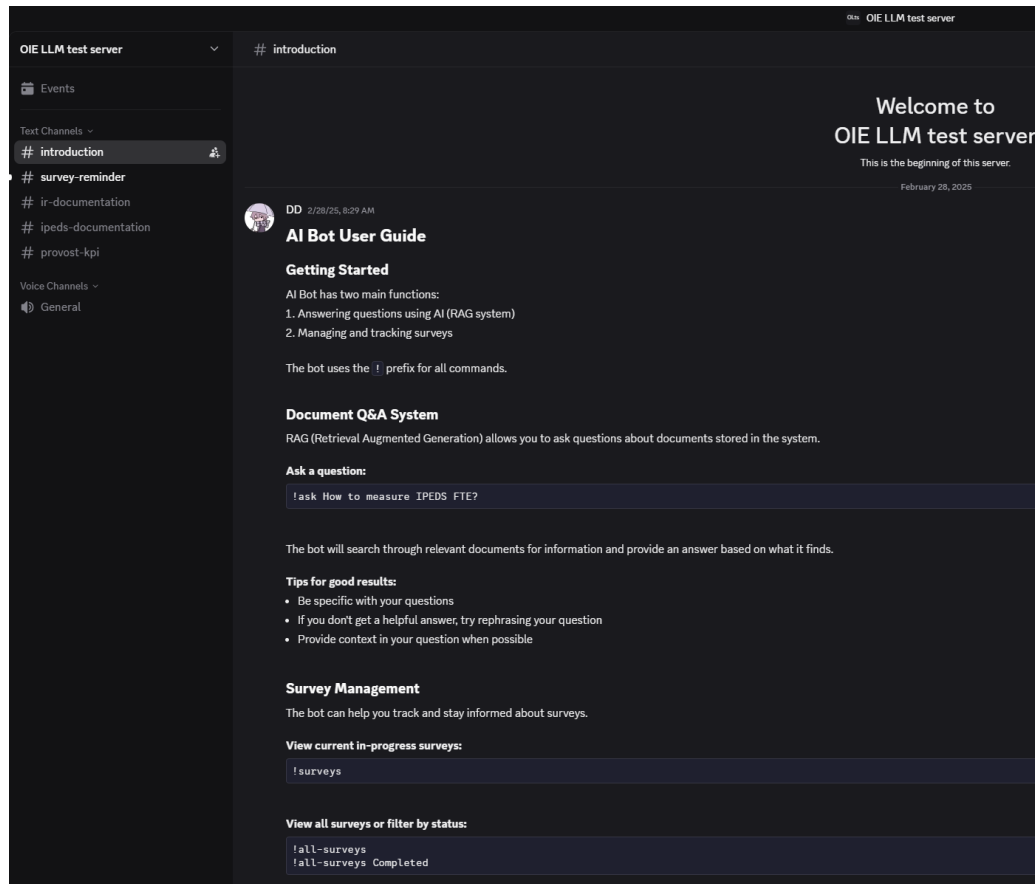
Quick Poll - Accessing Institution's AI Readiness

- Level 1: Fragmented Data
 - No centralized data or governance
 - Reliant on IR professionals for knowledge
- Level 2: Breaking down Silos
 - 60%+ data defined, through individual or small group efforts
 - Data definitions may not be aligned across departments
- Level 3: Modern and AI Ready
 - Data integrated with clear business definitions
 - Centralized data governance
- Level 4: AI-Accelerating
 - Centralized structured and unstructured data with clear metadata for AI use

Introducing IR SmartBot - Digital Knowledge Assistant

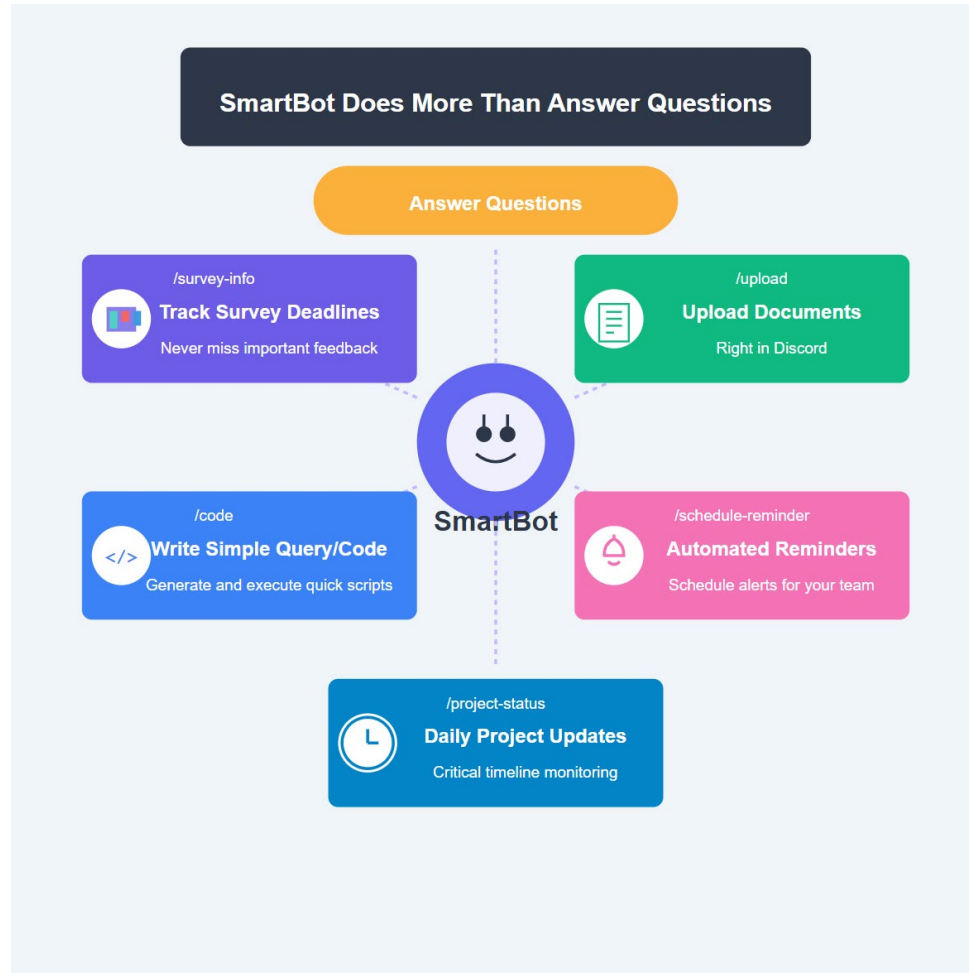


Key Features of AI-Powered IR chatbot



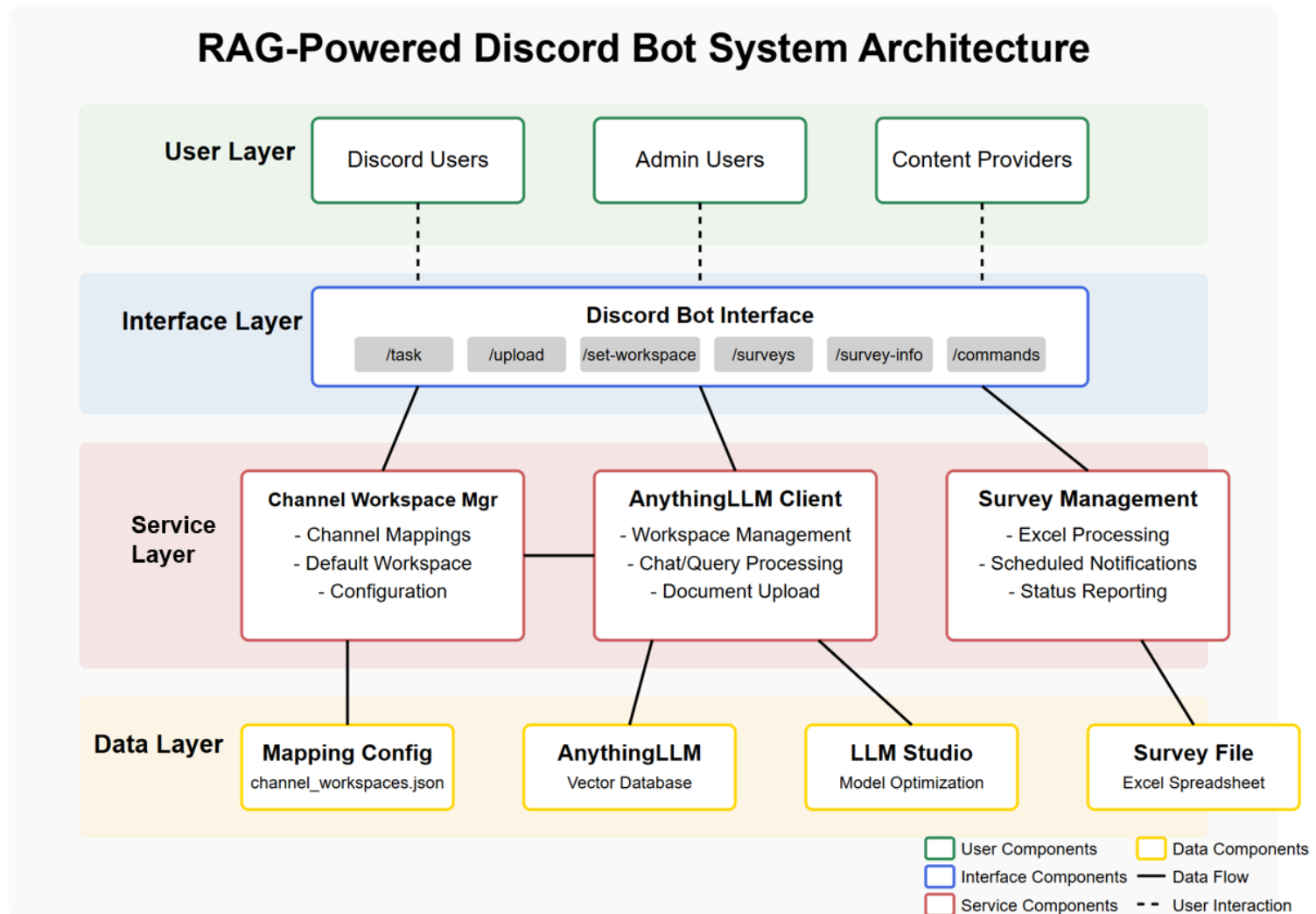
- Natural Language Processing
- Cross-departmental Data Integration
- Real-time collaboration

Smart Bot Does More Than Answer Questions

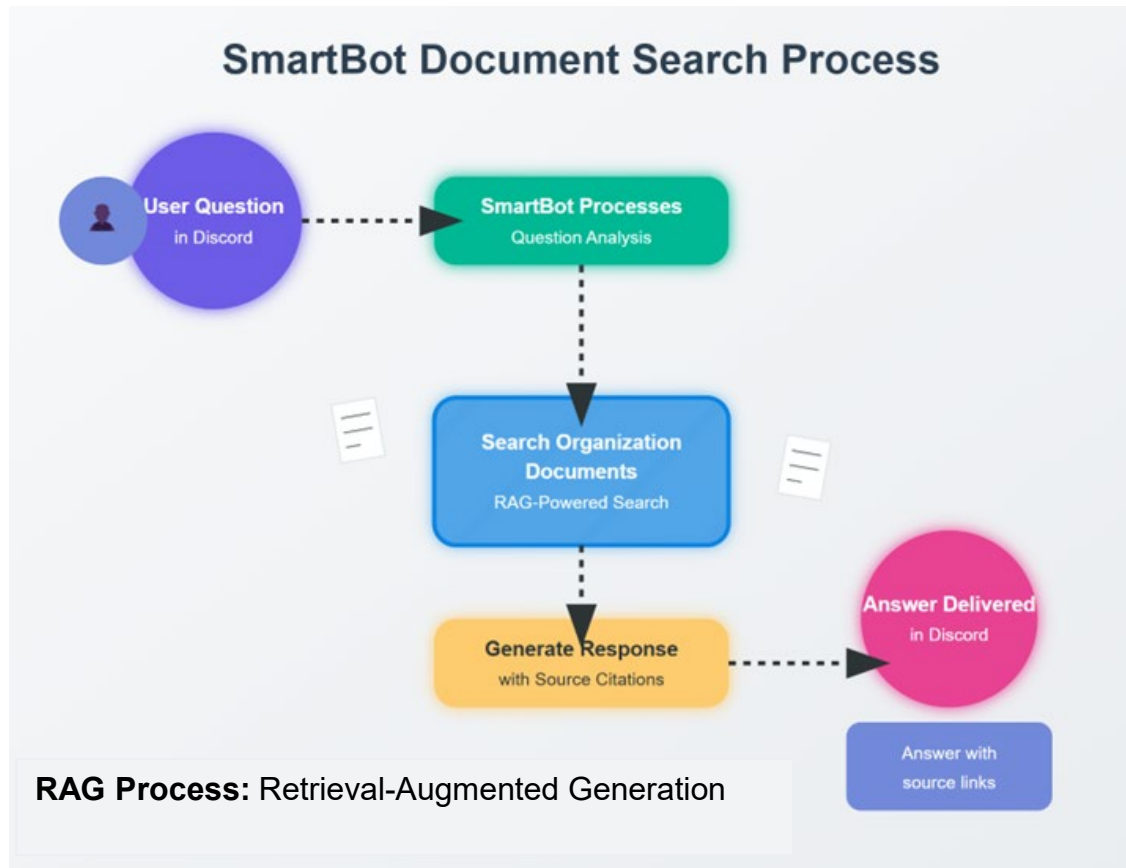


- **Task Reminder** Channel
(Project coordinator)
- **IR-Documentation** Channel
(Secretary)
- **IPEDS-documentation** channel
(Domain expert)
- **Provost-KPI** channel
(IR library)
- **Banner Support** channel
(Coding assistant)

System Design



How SmartBot Works - The Magic Made Simple



- Think of it as having a research assistant
 - Who has read every document in your organization
 - 24/7 available

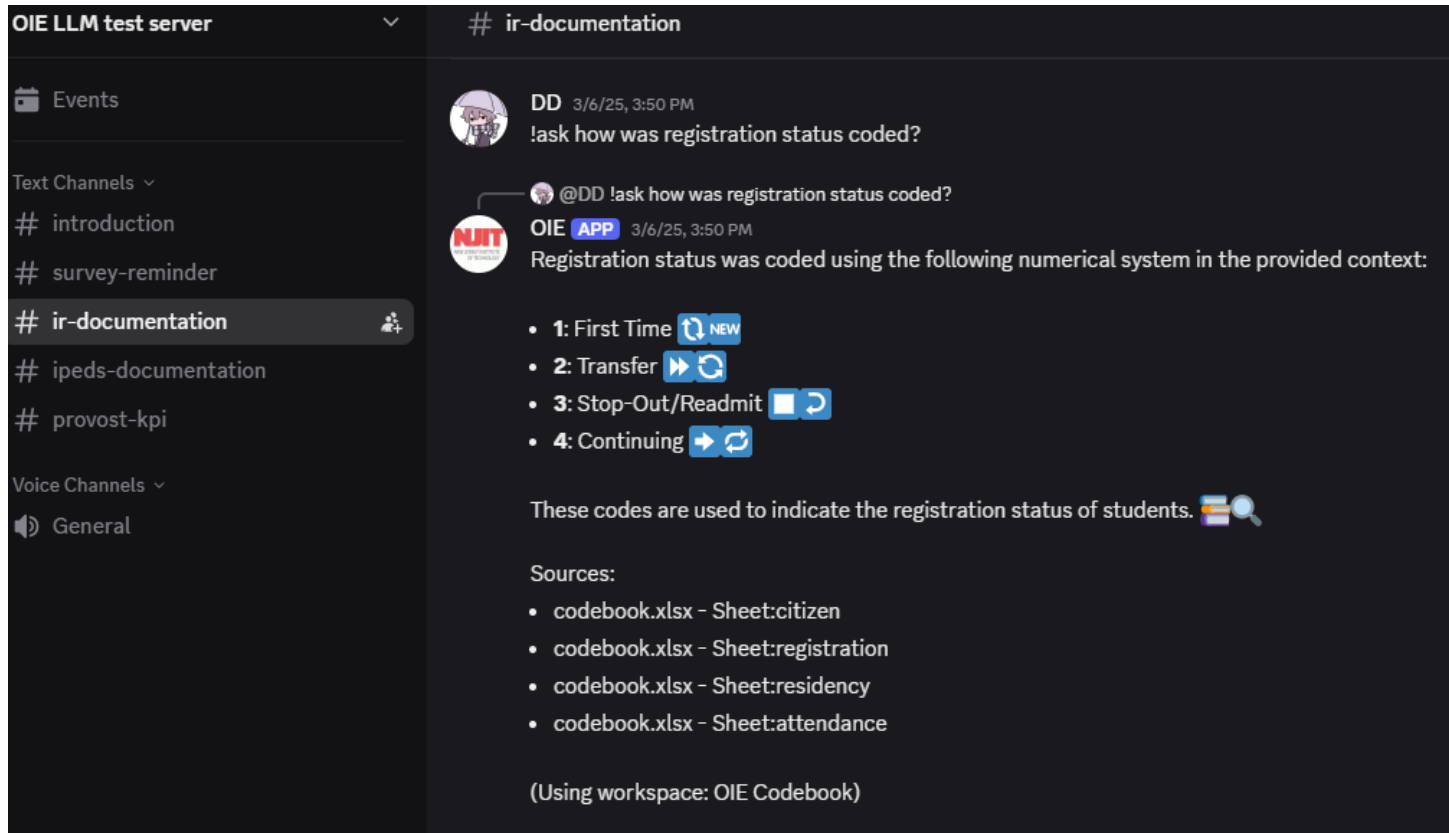
Let's See SmartBot in Action (Demo)

■ Daily task reminder to the team

The screenshot shows a Slack interface with a sidebar on the left containing a list of channels: # introduction, # survey-reminder (selected), # ir-documentation, # ipeds-documentation, # provost-kpi, and a General voice channel. The main area displays a message from SmartBot (OIE APP) titled "Daily Survey Status Update" dated 3/26/25, 8:56 AM. The message lists three surveys: Survey ID: 964 (Federal Web-Based Spring 2025 Collection, Source: Integrated Postsecondary Education Data System (IPEDS), Deadline: 03/28/2025), Survey ID: 977 (World University Rankings, Source: Times Higher Education (THE), Deadline: 03/30/2025), and Survey ID: 979 (SURE Spring Enrollment File, Source: New Jersey Office of the Secretary of Higher Education (OSHE), Deadline: 04/01/2025). It also states "Total in-progress surveys: 3" and "Next deadline: 03/28/2025". A second identical message is shown below, dated "Yesterday at 8:56 AM". The bottom of the screen shows a message input field with the text "#survey-reminder".

Let's See SmartBot in Action (Demo)

- Immediate answers with relevant context and sources
- New Employee Training
- Standardized data definition to the group access



The screenshot shows a Discord chat window for the 'OIE LLM test server'. The left sidebar lists text channels: # introduction, # survey-reminder, # ir-documentation (selected), # ipeds-documentation, and # provost-kpi. The main chat area shows a message from user 'DD' asking 'ask how was registration status coded?'. The SmartBot 'OIE APP' responds with the following text:

Registration status was coded using the following numerical system in the provided context:

- 1: First Time  NEW
- 2: Transfer 
- 3: Stop-Out/Readmit 
- 4: Continuing 

These codes are used to indicate the registration status of students. 

Sources:

- codebook.xlsx - Sheet:citizen
- codebook.xlsx - Sheet:registration
- codebook.xlsx - Sheet:residency
- codebook.xlsx - Sheet:attendance

(Using workspace: OIE Codebook)

Let's See SmartBot in Action (Demo)

Official document glossary

OIE LLM test server

Events

Text Channels

introduction

survey-reminder

ir-documentation

ipeds-documentation

provost-kpi

Voice Channels

General

ipeds-documentation

Xilin 11:17 AM

ask How to calculate FTE? Provide SQL code

@Xilin ask How to calculate FTE? Provide SQL code

OIE APP 11:17 AM

To calculate Full-Time Equivalent (FTE) using SQL, you can use one of two methods based on the data available. Below are examples for both methods:

Method 1: Using Fall Student Headcounts

Assuming you have tables `full_time_enrollment`, `part_time_undergraduate_enrollment`, and `part_time_graduate_enrollment` with columns `count`:

```
SELECT
  (SELECT count FROM full_time_enrollment) AS full_time_count,
  (SELECT count FROM part_time_undergraduate_enrollment) * 0.403543 AS part_time_undergrad_fte,
  (SELECT count FROM part_time_graduate_enrollment) * 0.361702 AS part_time_grad_fte,
  ((SELECT count FROM full_time_enrollment) +
   (SELECT count FROM part_time_undergraduate_enrollment) * 0.403543 +
   (SELECT count FROM part_time_graduate_enrollment) * 0.361702) AS total_fte
FROM dual;
```

Method 2: Using 12-Month Instructional Activity

Assuming you have tables `undergraduate_credit_hours` and `graduate_credit_hours` with columns `total_hours`:

```
SELECT
  (SELECT total_hours FROM undergraduate_credit_hours) / 30 AS undergrad_fte,
  (SELECT total_hours FROM graduate_credit_hours) / 24 AS grad_fte,
  ((SELECT total_hours FROM undergraduate_credit_hours) / 30 +
   (SELECT total_hours FROM graduate_credit_hours) / 24) AS total_fte
FROM dual;
```

Replace the table and column names with those that match your database schema. The `dual` table is used in Oracle SQL for calculations without referencing a value).

Sources:

- IPEDSGlossary.csv

Let's See SmartBot in Action (Demo)

- Immediate answers with relevant context and sources

The screenshot shows a Discord chat window for the server "OIE LLM test server". The left sidebar lists several text channels, with "# provost-kpi" selected. The main chat area shows a message from user "Deess" asking "ask How many students are in the honors college?". Below this, a response from the "OIE APP" bot is displayed. The response provides a list of student counts for each year from 2015 to 2024, followed by a "Sources:" section listing "Provost KPI.csv" four times. At the bottom, it notes "(Using workspace: Provost KPI)".

Deess 3/19/25, 11:01 PM
ask How many students are in the honors college?

@Deess ask How many students are in the honors college?

OIE APP 3/19/25, 11:01 PM
The number of students enrolled in the Honors College is as follows:

- 2015: 714
- 2016: 708
- 2017: 727
- 2018: 733
- 2019: 718
- 2020: 648
- 2021: 623
- 2022: 556
- 2023: 626
- 2024: 760

Sources:

- Provost KPI.csv
- Provost KPI.csv
- Provost KPI.csv
- Provost KPI.csv

(Using workspace: Provost KPI)

Let's See SmartBot in Action (Demo)

- Coding support with the knowledge of your student information system(SIS)

OIE LLM test server banner-support

Xilin 6/4/25, 2:40 PM
!ask write SQL code to extract email information from Banner

@DD !ask wrote sql code

OIE APP 6/4/25, 2:40 PM
Certainly! Below is an example of SQL code that extracts email data from the Banner system. This example assumes that the email addresses are stored in a table named `SPRIDEN`, which is common in Banner systems.

SQL Query to Extract Email Data

```
SELECT
  SPRIDEN.PIDM AS Student_ID,
  SPRIDEN.EMAIL_ADDRESS AS Email_Address,
  SGBSTDN.SGBSTDN_STST_CODE AS Student_Status,
  SGBSTDN.SGBSTDN_LEVL_CODE AS Level_Code,
  STV_COLL.STV_COLL_DESC AS College_Name,
  STVMAJR.STVMAJR_DESC AS Major_Name
FROM
  SPRIDEN
JOIN
  SGBSTDN ON SPRIDEN.PIDM = SGBSTDN.SGBSTDN_PIDM
JOIN
  STV_COLL ON SGBSTDN.SGBSTDN_COLL_CODE_1 = STV_COLL.STV_COLL_CODE
JOIN
  STVMAJR ON SGBSTDN.SGBSTDN_MAJR_CODE_1 = STVMAJR.STVMAJR_CODE
WHERE
  SPRIDEN.EMAIL_ADDRESS IS NOT NULL;
```

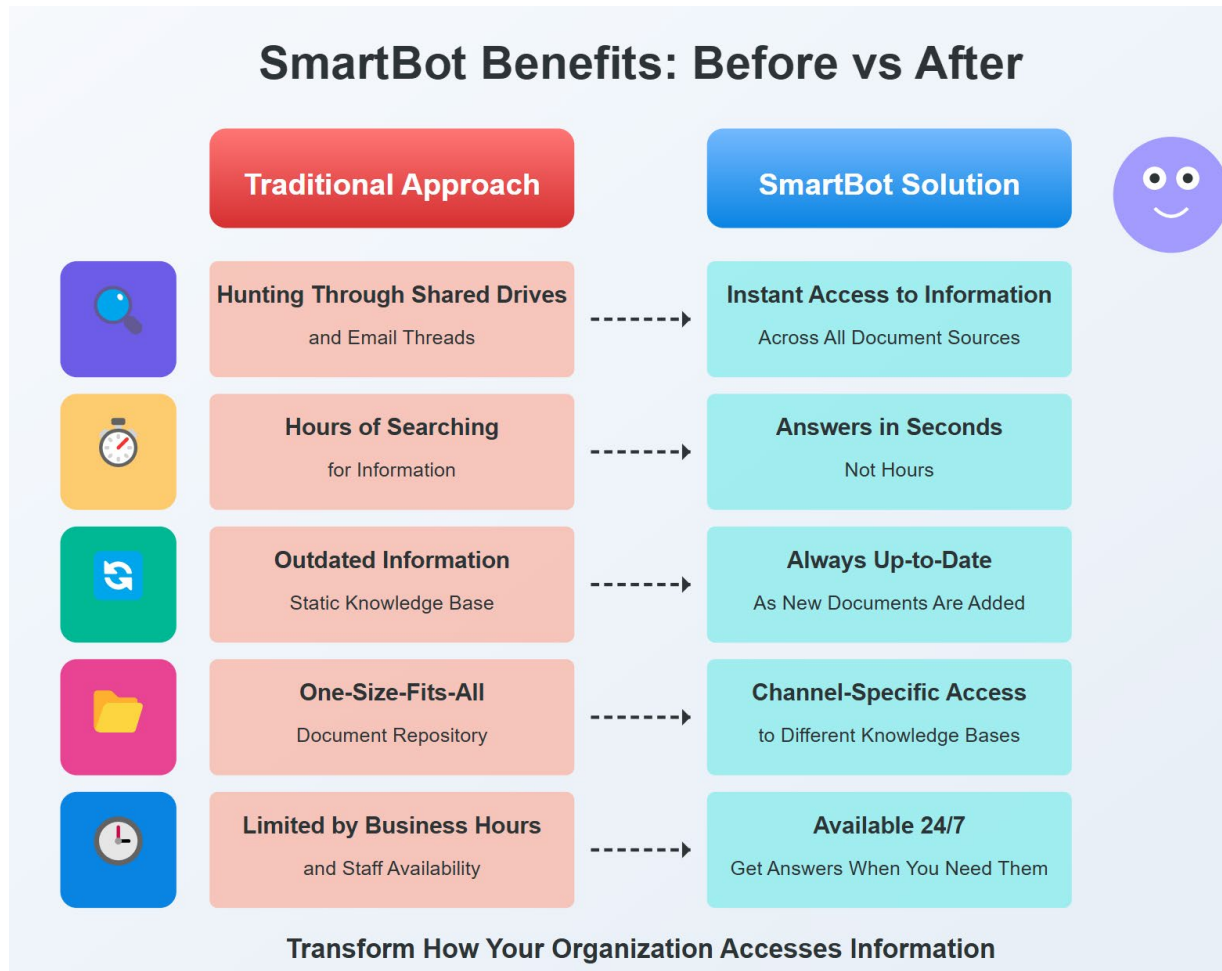
Explanation:

- **SPRIDEN:** This table typically contains student demographic information, including email addresses.
- **SGBSTDN:** This table contains student status and level information.
- **STV_COLL:** This table contains college descriptions.
- **STVMAJR:** This table contains major descriptions.

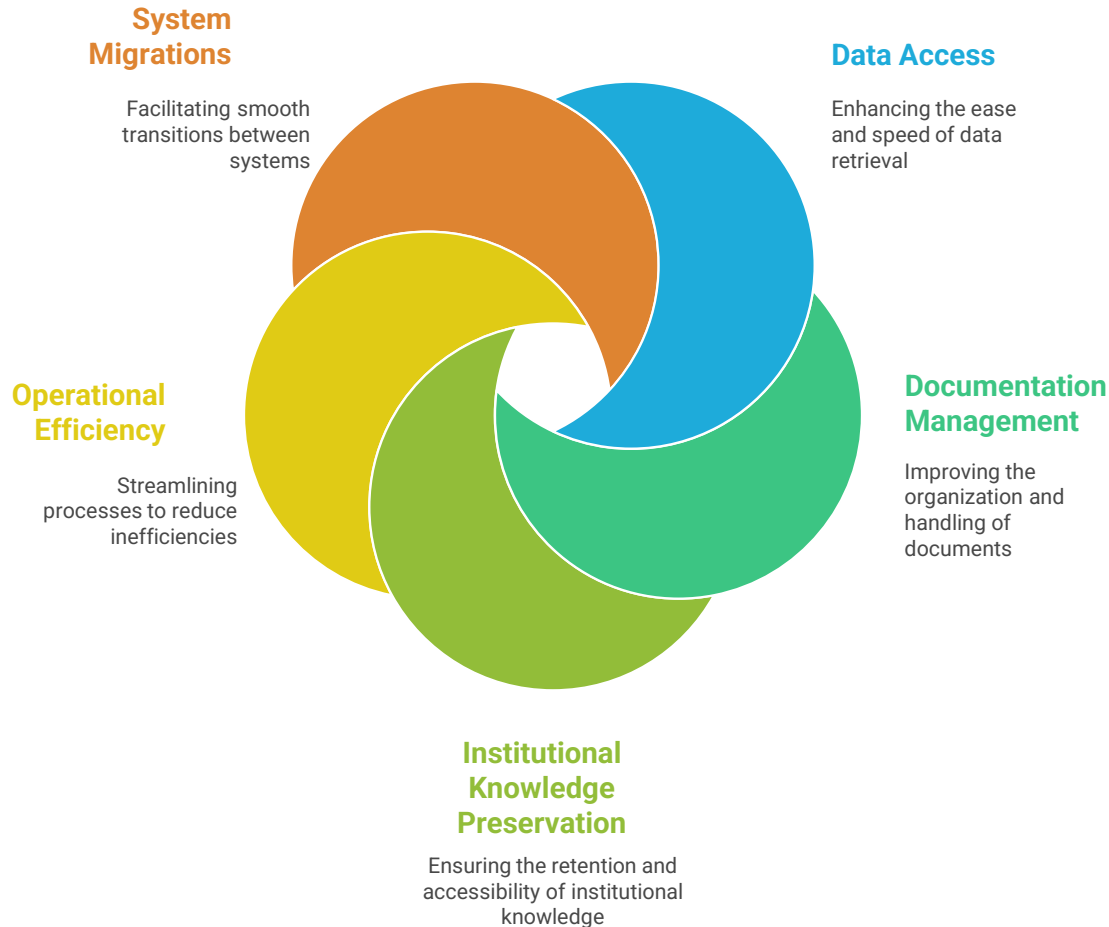
Additional Considerations:

1. **Effective Terms:** If your data includes effective terms, you might need to adjust the query to handle these effectively.
2. **Security:** Always use bind variables in your SQL queries to prevent SQL injection attacks.
3. **Performance:** Ensure that appropriate indexes are in place on the columns used in the `JOIN` and `WHERE` clauses to improve performance.

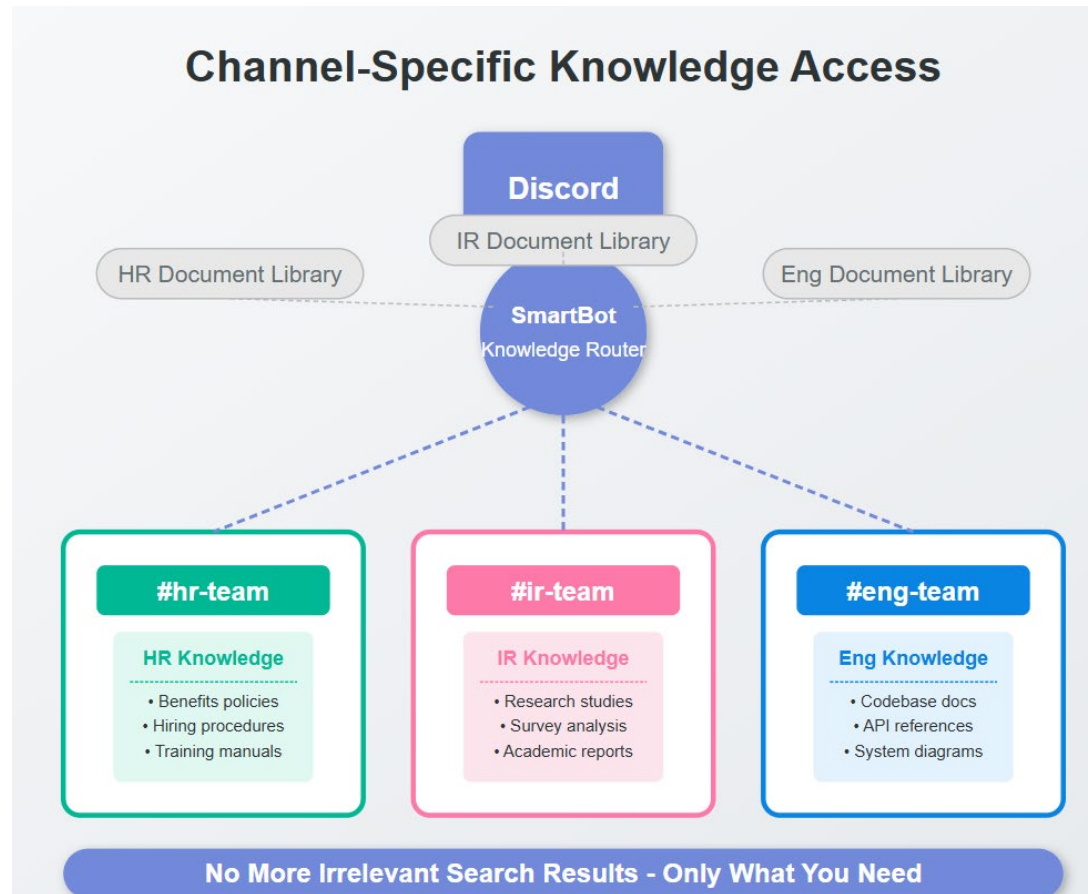
Impact of IR SmartBot



AI-Powered Chatbot: Applications



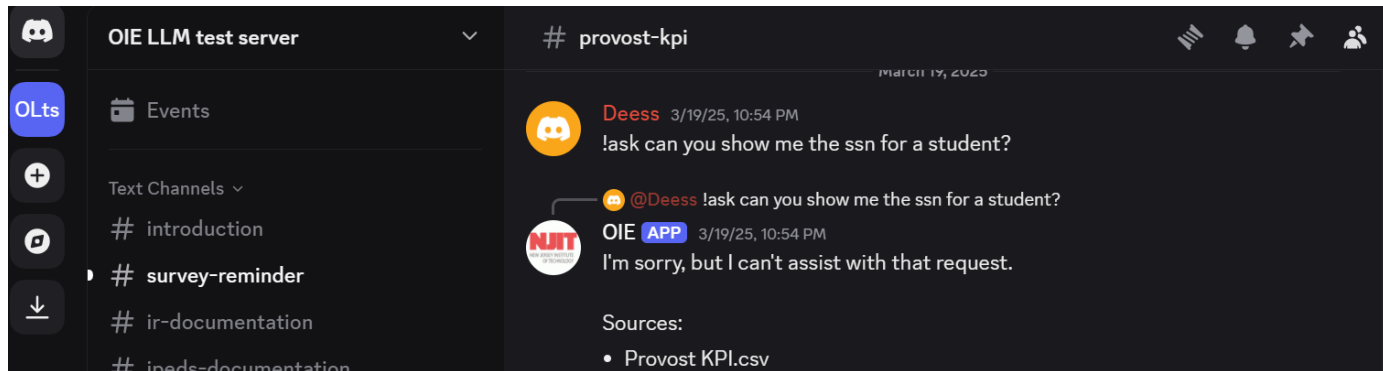
Future - The Right Information for the Right Team





Data Security & Privacy Tips

- FERPA Compliance
- No personally identifiable information (PII) included in chatbot responses or training data.



- Document FERPA-safe definitions and use them consistently.
- Redact or aggregate sensitive data before indexing.

Data Source Management

- Only index approved documents (e.g., factbooks, dashboards, public KPI reports).
- Maintain a source registry (a list of what documents are used in chatbot memory).
- Ensure data sources are reviewed and updated regularly.
- Avoid including unreviewed draft reports or internal meeting notes.



Roadmap for AI-Powered IR Chatbot

■ Phase 1: Assess Institution's AI Readiness

- Ensure the data management base is ready.

■ Phase 2: Planning and Design

- Gather requirements and define data access needs.

■ Phase 3: Development and Integration

- Develop chatbot features and integrate with existing systems.

■ Phase 4: Testing and Deployment

- Conduct testing and deploy across departments.

Interested in IR SmartBot?

- A general version of the IR Chatbot that is adaptable for use for various institutions
- Receive updates on our product's progress - Sign Interest in IR SmartBot [form](#) (shared in the chat)
- Opportunity to be the first-generation user of the Beta version and gain early access to this innovative tool.
- Join the AI in IR community by receiving an invitation to collaborate and share insights.

Share Your Thoughts

