

Understanding and Supporting Formal Email Exchange by Answering AI-Generated Questions

Yusuke Miura

Waseda University

Chi-Lan Yang

The University of Tokyo

Masaki Kuribayashi

Waseda University

Keigo Matsumoto

The University of Tokyo

Hideaki Kuzuoka

The University of Tokyo

Shigeo Morishima

Waseda Research Institute
for Science and Engineering

Overview

Propose and Evaluate QA-based Approach to Make Reply Work Efficiently

Propose QA-based Approach

The screenshot shows an email reply interface. On the left, an "Incoming Mail" section displays an email from Ken Tanaka. The main area is titled "Questions by AI" and contains several questions with input fields: "Which research lab are you ultimately aiming to join for your graduation project?" (with "Tanaka Lab" selected), "Are you willing to accept this award?" (with "Yes" selected), "What is your GPA?" (with "3.5" entered), and "What are your current future aspirations?" (with "Research Career" selected). A "Generate Reply" button is at the bottom. On the right, a "Customize Reply" panel allows users to select relationship (Sender/You), formality (Informal/Neutral/Formal), tone (Friendly/Professional/Encouraging), and length (Long/Middle/Short).

Evaluate through User Studies

Controlled Experiment (N=12)



Reply Tasks



Questionnaire

Field Study (N=8)



System Usage



Interview

Users can create replies **by answering “Questions from LLM”**

Overview

Propose and Evaluate QA-based Approach to Make Reply Work Efficiently

Propose QA-based Approach

The screenshot shows an email interface with three main sections:

- Incoming Mail:** A message from Ken Tanaka regarding a selection process for a graduation project award. It lists four criteria: research lab, willingness to accept an award, GPA, and future aspirations.
- Questions by AI:** A central panel with a blue border containing AI-generated questions and user inputs:
 - Question: "Which research lab are you ultimately aiming to join for your graduation project?" with input "Tanaka Lab".
 - Question: "Are you willing to accept this award?" with input "Yes".
 - Question: "What is your GPA?" with input "3.5".
 - Question: "What are your current future aspirations?" with input "Research Career".
 - A "New option" field with an "Add" button.
 - A question: "Can you send your response within the next day or two?" with a mouse cursor over the text.
- Customize Reply:** A panel on the right for personalizing the reply:
 - Relationship:** Sender and You fields.
 - Formality:** Informal, Neutral, Formal buttons.
 - Tone:** Friendly, Professional, Encouraging buttons.
 - Length:** Long, Middle, Short buttons.
 - Additional Request:** A text area for extra notes.
 - Generate Reply:** A blue button at the bottom.

Evaluate through User Studies

Controlled Experiment (N=12)



Reply Tasks



Questionnaire

Field Study (N=8)



System Usage



Interview

Users can create replies **by answering “Questions from LLM”**

Overview

Propose and Evaluate QA-based Approach to Make Reply Work Efficiently

Propose QA-based Approach

The screenshot shows an email interface with three main sections:

- Incoming Mail:** Contains an email from Ken Tanaka to Sato. The email text includes: "Dear Sato, This email is being sent to the relevant members with Bcc. The Ministry of Education has informed me that only one individual can be awarded per examination. Therefore, we will need to select one candidate from among you. Rather than resorting to a random draw, I believe it would be more appropriate to make an objective decision based on the following criteria: 1. The research lab you are ultimately aiming to join for your graduation project 2. Your willingness to accept this award 3. Your GPA 4. Your current future aspirations I would appreciate receiving your response within the next day or two. Thank you for your cooperation. Best regards, Ken Tanaka".
- Questions by AI:** A central panel with a refresh icon. It contains four questions with input fields: "Which research lab are you ultimately aiming to join for your graduation project?" (input: Tanaka Lab), "Are you willing to accept this award?" (input: Yes), "What is your GPA?" (input: 3.5), and "What are your current future aspirations?" (input: Research Career). Below these is a "New option" field with an "Add" button and a question "Can you send your response within the next day or two?".
- Customize Reply:** A panel on the right with options for "Relationship" (Sender, You), "Formality" (Informal, Neutral, Formal), "Tone" (Friendly, Professional, Encouraging), and "Length" (Long, Middle, Short). There is an "Additional Request" text area and a "Generate Reply" button at the bottom.

Evaluate through User Studies

Controlled Experiment (N=12)



Reply Tasks



Questionnaire

Field Study (N=8)



System Usage



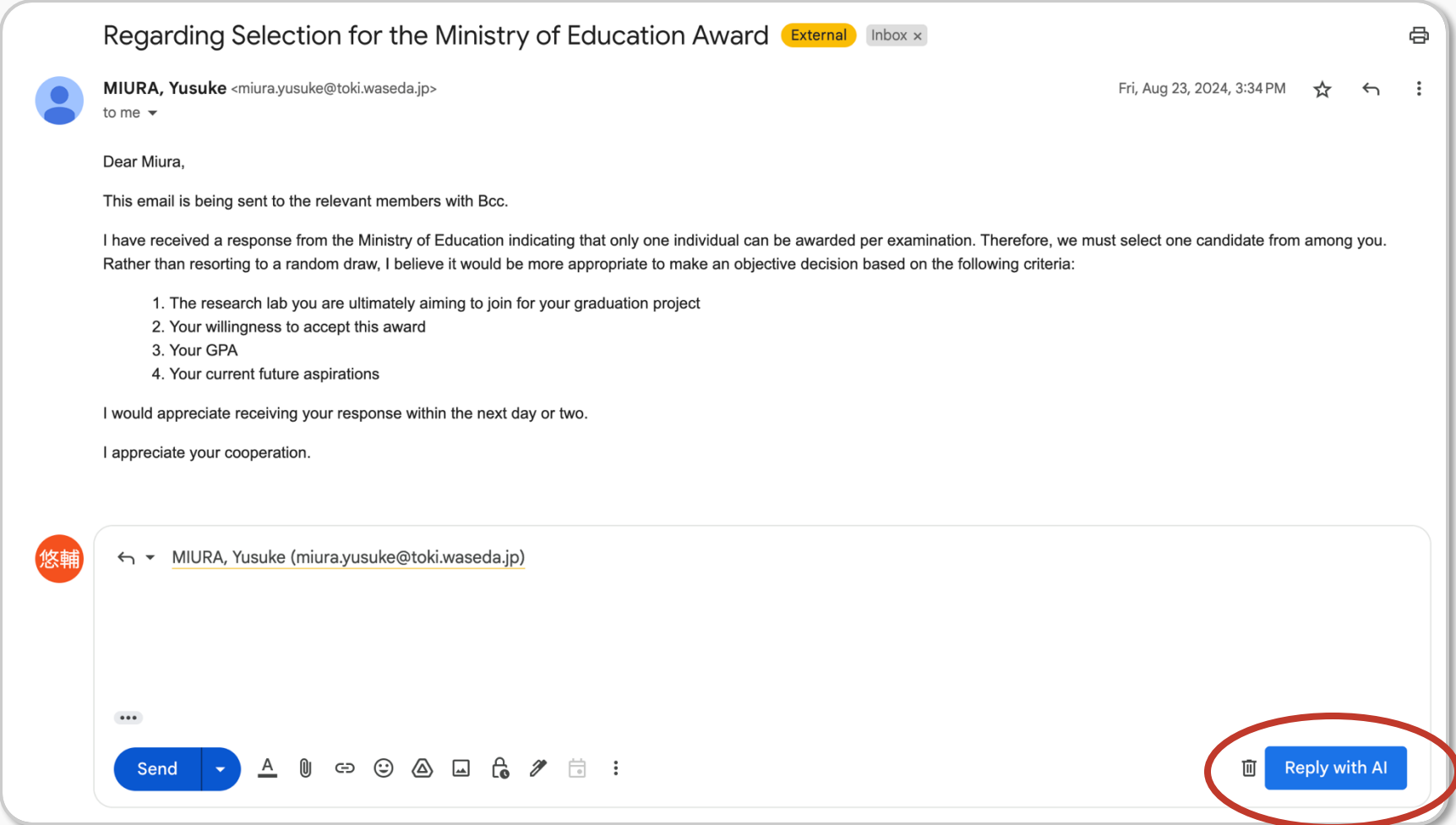
Interview

Users can create replies **by answering “Questions from LLM”**

Overview

Streamline Email Reply Tasks with “Questions from LLM”

Step 1 For the email users want to reply to, press the “Reply with AI” button



Overview

Streamline Email Reply Tasks with “Questions from LLM”

Step 2 Check the **questions and options generated by the LLM** and provide answers

The screenshot displays a three-pane interface for email management. The left pane, titled "Incoming Mail", shows an email received on Sep 6, 2024, at 7:30 PM. The subject is "Regarding Selection for the Ministry of Education Award" and the sender is "Tanaka, Ken". The email body addresses "Dear Sato" and discusses a selection process for an award, listing four criteria: 1. The research lab you are ultimately aiming to join for your graduation project, 2. Your willingness to accept this award, 3. Your GPA, and 4. Your current future aspirations. It requests a response within the next day or two and is signed by Ken Tanaka. The middle pane, titled "Questions by AI", contains a single black dot, representing a question generated by the LLM. The right pane, titled "Customize Reply", offers options for "Relationship" (Sender, You), "Formality" (Informal, Neutral, Formal), "Tone" (Friendly, Professional, Encouraging), and "Length" (Long, Middle, Short). There is also a text area for "Additional Request" and a "Generate Reply" button at the bottom.

Overview

Streamline Email Reply Tasks with “Questions from LLM”

Step 3 Adjust the tone if necessary, and then click the “Generate Reply” button

The screenshot shows a web interface for replying to an email. It is divided into three main sections: Incoming Mail, Questions by AI, and Customize Reply.

- Incoming Mail:** Contains the original email text. The recipient is Ken Tanaka, and the sender is Ken Tanaka. The email discusses a selection process for a graduation project and lists four criteria: 1. The research lab you are ultimately aiming to join for your graduation project, 2. Your willingness to accept this award, 3. Your GPA, and 4. Your current future aspirations. The sender requests a response within the next day or two.
- Questions by AI:** A list of questions generated by AI based on the email content. Each question has a corresponding answer field:
 - Question: "Which research lab are you ultimately aiming to join for your graduation project?" Answer: "Tanaka Lab"
 - Question: "Are you willing to accept this award?" Answer: "Yes"
 - Question: "What is your GPA?" Answer: "3.5"
 - Question: "What are your current future aspirations?" Answer: "Research Career" (with an "Add" button for a "New option")
 - Question: "Can you send your response within the next day or two?" (This question is highlighted with a mouse cursor)
- Customize Reply:** A panel for adjusting the reply's characteristics:
 - Relationship:** Sender and You (input fields)
 - Formality:** Informal, Neutral, Formal (radio buttons)
 - Tone:** Friendly, Professional, Encouraging (radio buttons)
 - Length:** Long, Middle, Short (radio buttons)
 - Additional Request:** A text area for extra notes.
 - Generate Reply:** A prominent blue button at the bottom.

Background

Email is Widely Used, but Replying is Time-Consuming and Demanding

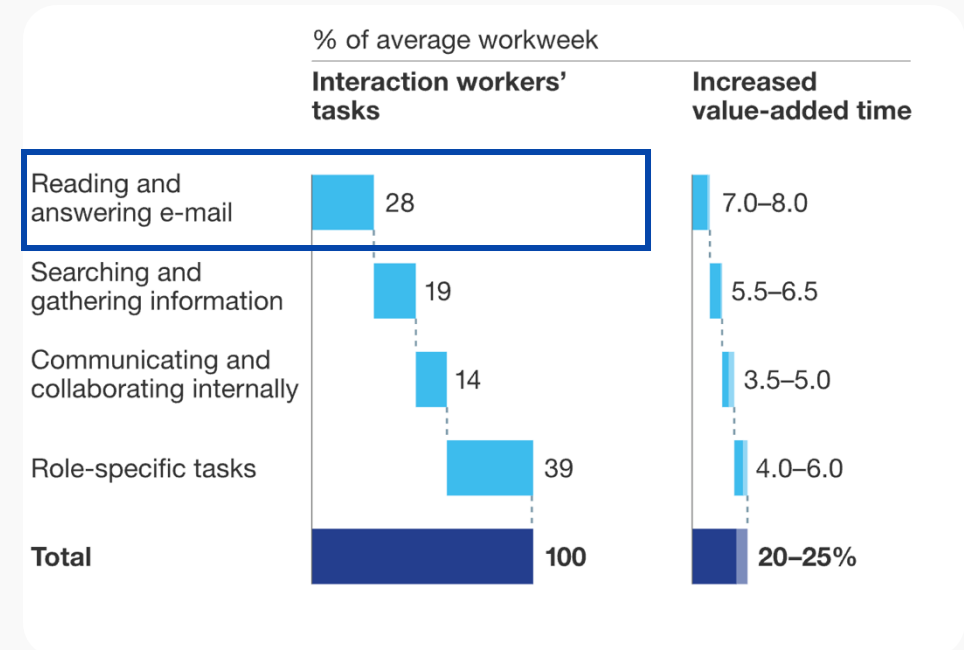
Examples of Situations Where Email is Used



Information Sharing [1]

Task Management [2]

High Workload for Replying



Workers spend 28% of their working hours checking and replying to emails [3]

[1] Nelson, Les, et al. "Mail2tag: Augmenting email for sharing with implicit tag-based categorization." 2011 International Conference on Collaboration Technologies and Systems (CTS). IEEE, 2011.

[2] Bellotti, Victoria, et al. "Quality versus quantity: E-mail-centric task management and its relation with overload." Human-Computer Interaction 20.1-2 (2005): 89-138.

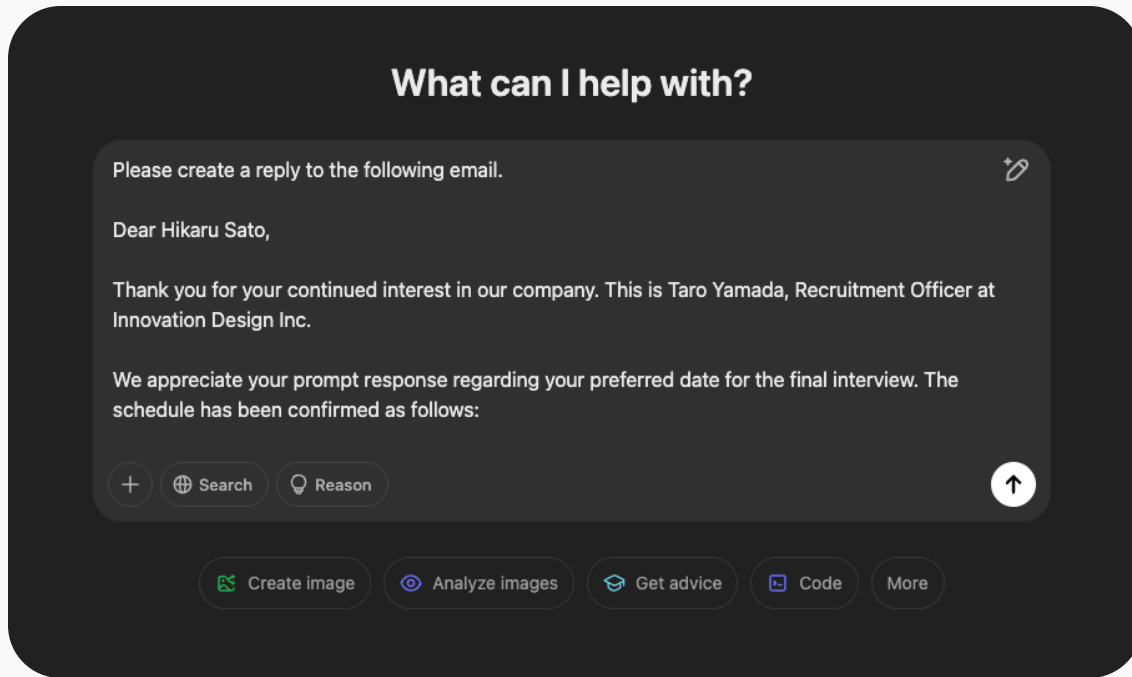
[3] Chui, Michael, et al. "The social economy: Unlocking value and productivity through social technologies." (2012).

Background

LLM-Based Reply Tools can Ease the Burden of Replying

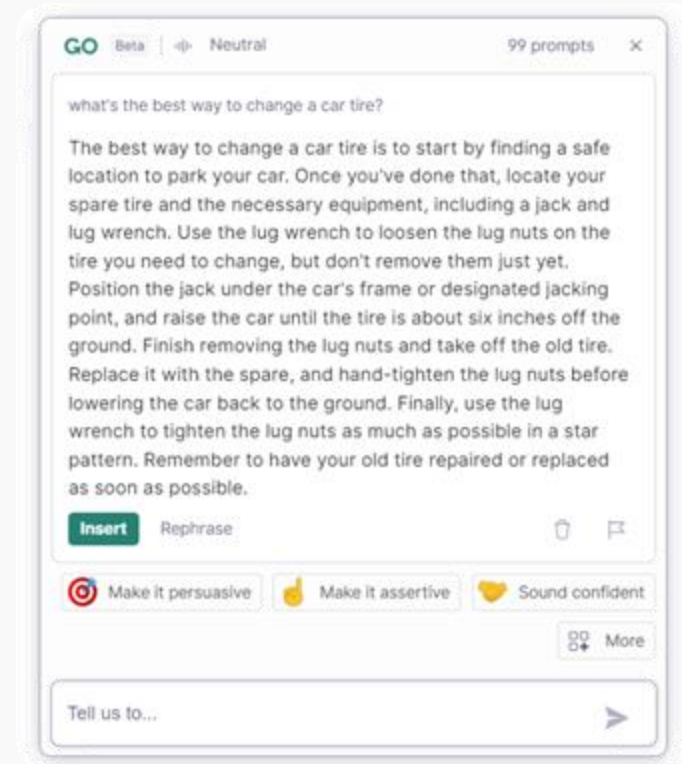
Examples of LLM-based Reply Assistance Tools

ChatGPT



Natural text responses and information are generated based on user-entered questions or commands

GrammarlyGO [4]



Email reply drafts are automatically generated

Background

Users Obtain Desired Reply Drafts by Providing Prompts to LLM

Reply Process Using Conventional Reply-Support Tools

1. Receive Email



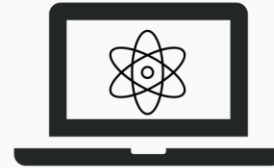
2. Create Detailed Prompt

I want to convey that I agree to take on the Fixed Assets Committee.
I would also like to convey that I have no problem with the annual inventory response.

:



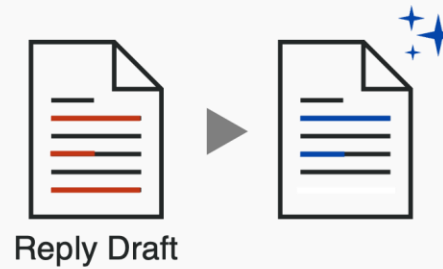
3. Get Draft by AI



Generate



4. Check & Revise



Reply Draft

5. Send Reply



Send



Background

However, Crafting Effective Prompts Remains a Challenge [4]

Reply Process Using Conventional Reply-Support Tools

1. Receive Email



2. Create Detailed Prompt

I want to convey that I agree to take on the Fixed Assets Committee.

I would also like to convey that I have no problem with the annual inventory response.

⋮



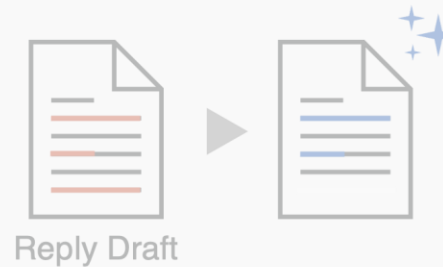
3. Get Draft by AI



Generate



4. Check & Revise



Reply Draft

5. Send Reply



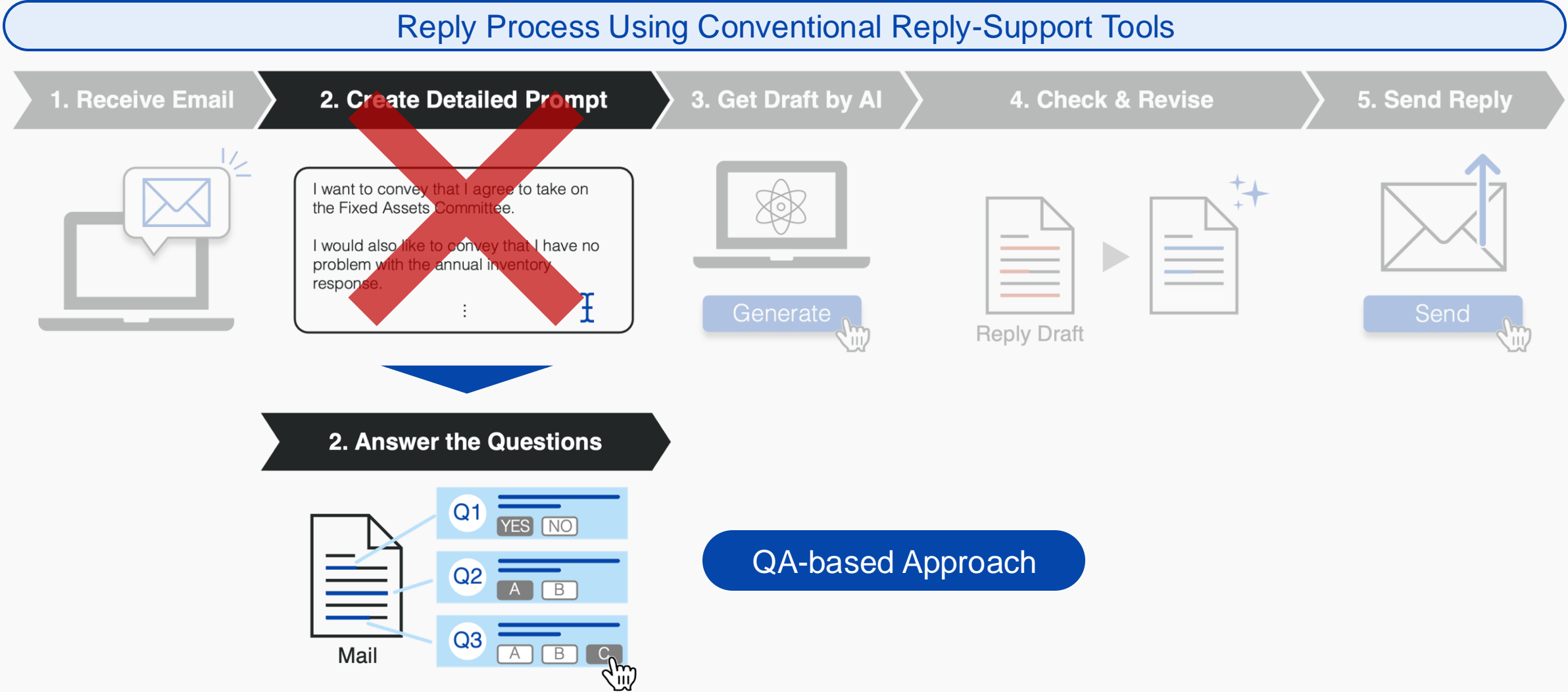
Send



I can't get the output I want...

Our Approach: AI-driven QA-based Approach

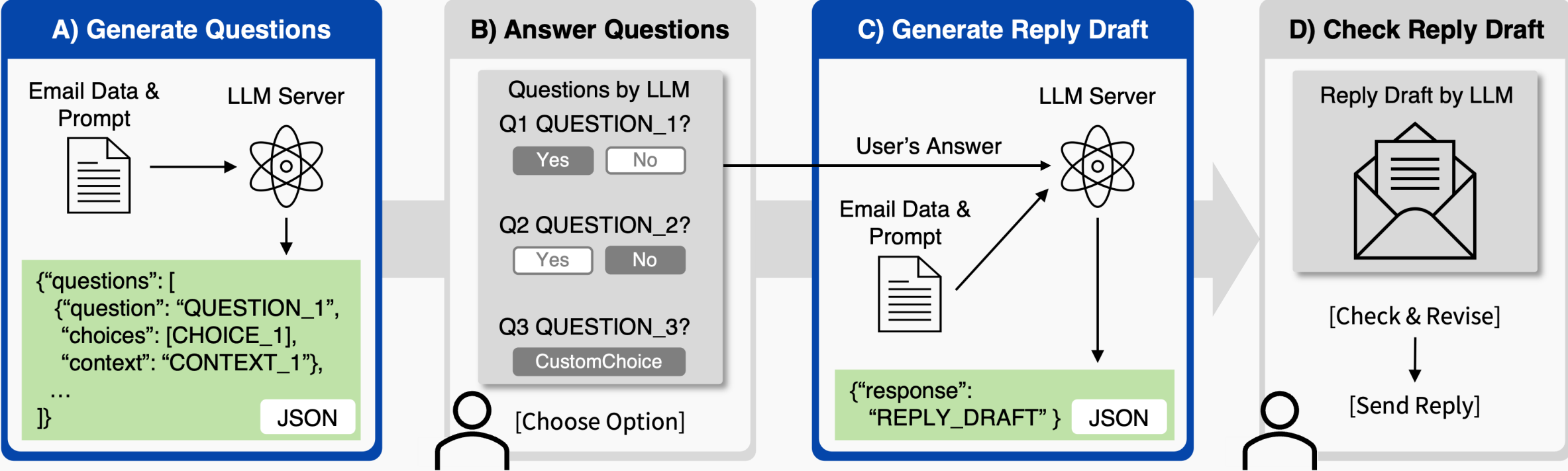
Instead of Writing Prompts, Users Respond to LLM-Generated Questions



Our Approach: AI-driven QA-based Approach

Streamline Email Reply Tasks with “Questions from LLM”

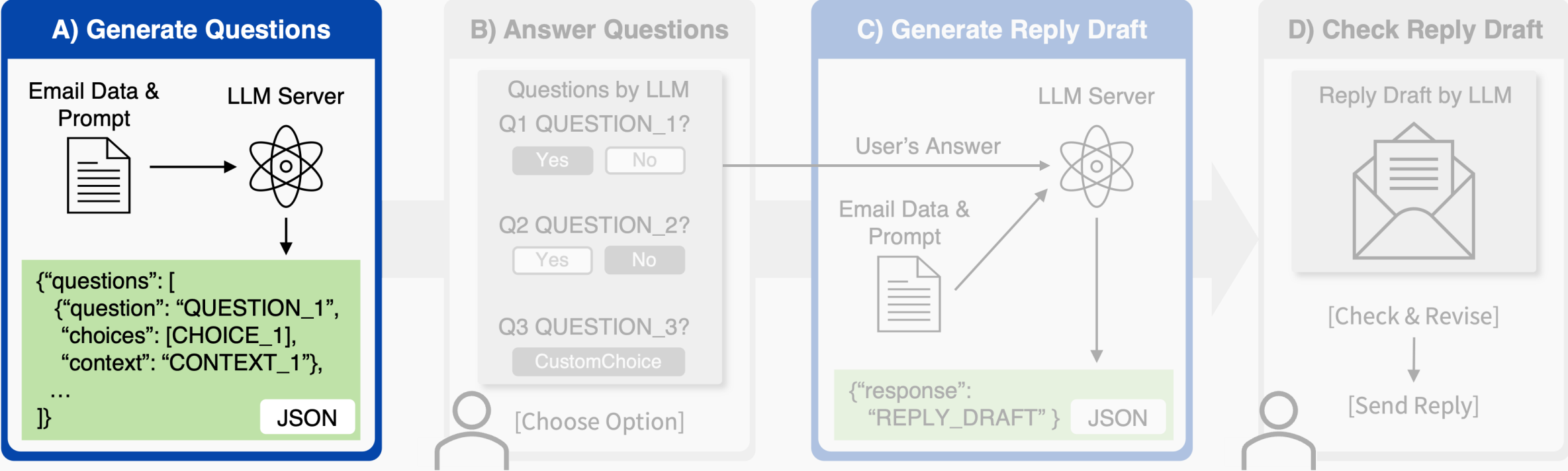
Reply Process Using Our QA-based Approach



Our Approach: AI-driven QA-based Approach

Streamline Email Reply Tasks with “Questions from LLM”

Reply Process Using Our QA-based Approach

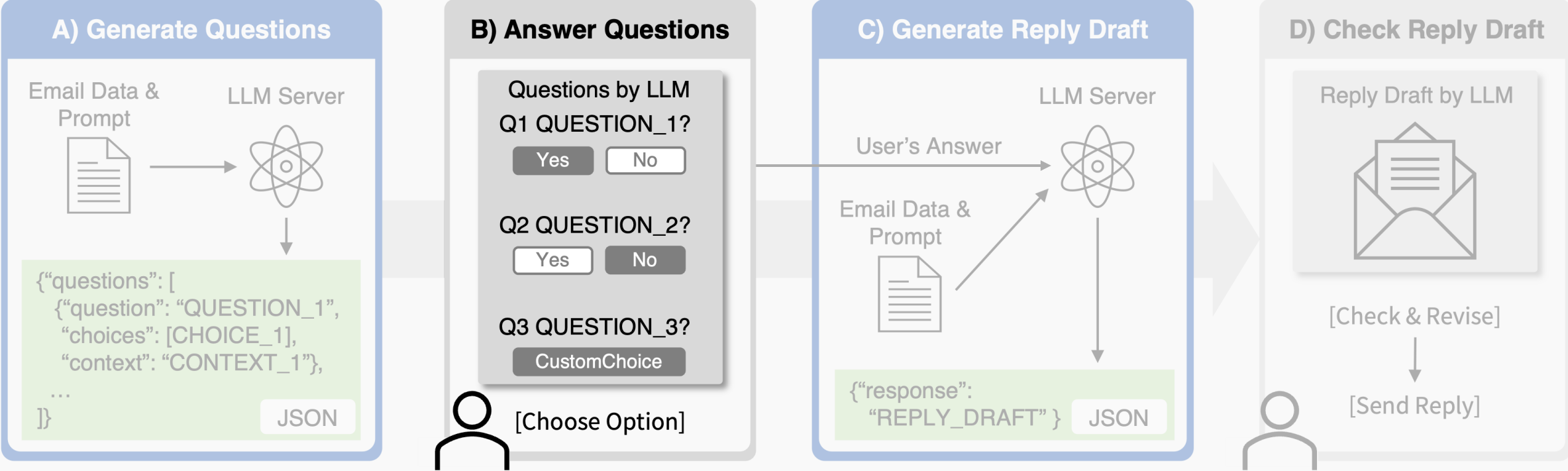


LLM receives the **email data** and **prompt**, then generates questions to elicit user's intention

Our Approach: AI-driven QA-based Approach

Streamline Email Reply Tasks with “Questions from LLM”

Reply Process Using Our QA-based Approach

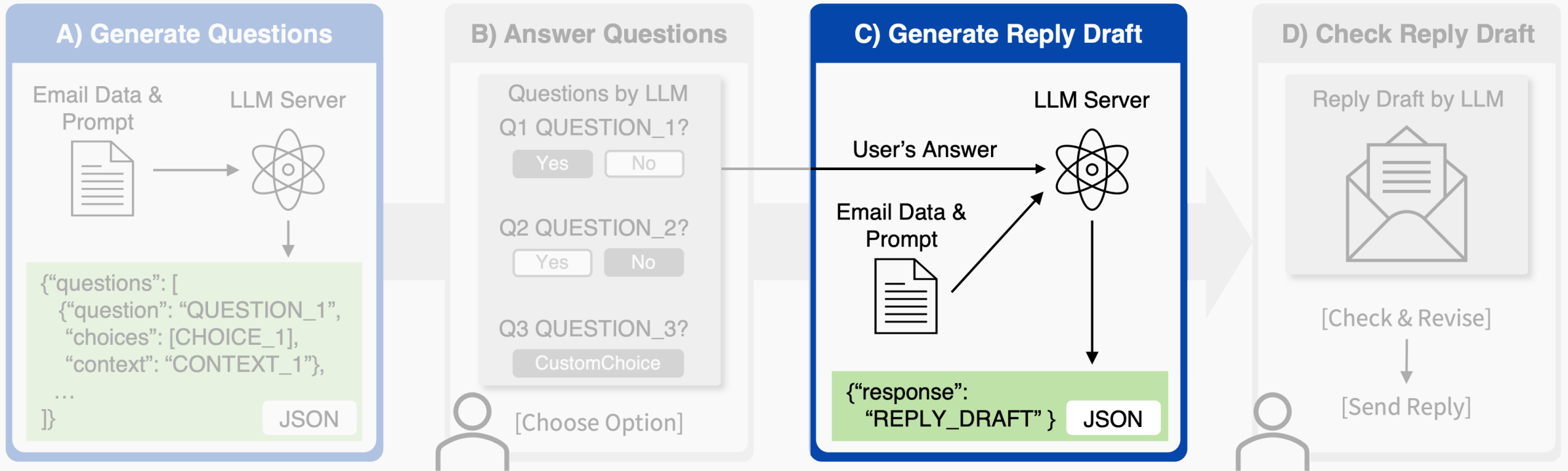


User answers the questions from the LLM

Our Approach: AI-driven QA-based Approach

Streamline Email Reply Tasks with “Questions from LLM”

Reply Process Using Our QA-based Approach

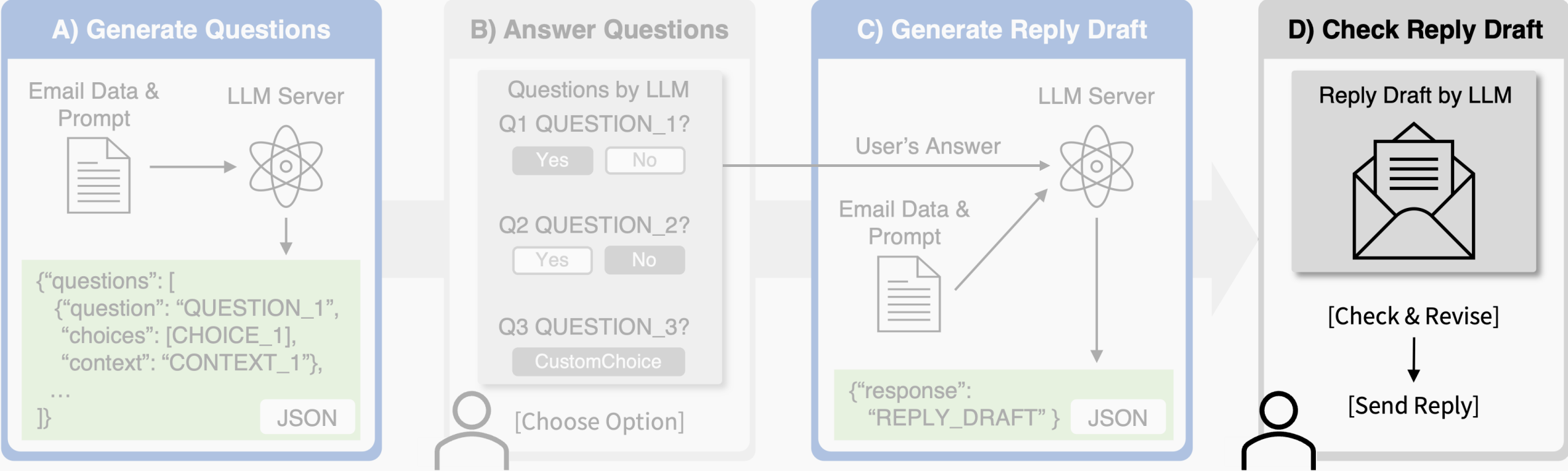


LLM receives **questions**, **user's answers**, **email data**, and **prompt**, then creates a reply draft

Our Approach: AI-driven QA-based Approach

Streamline Email Reply Tasks with “Questions from LLM”

Reply Process Using Our QA-based Approach

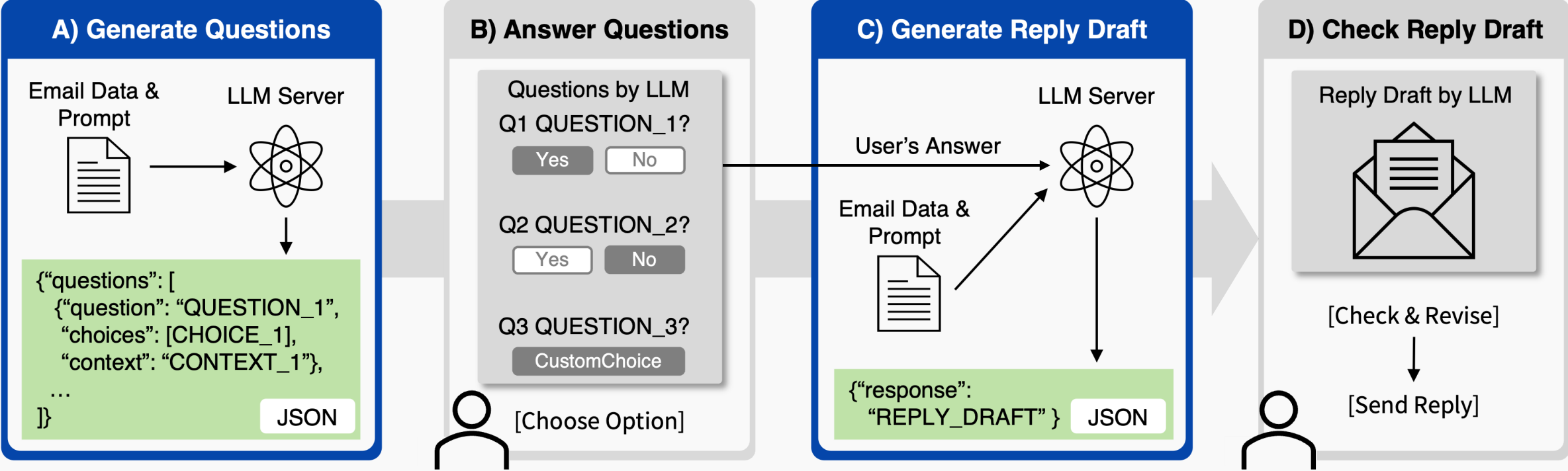


User reviews and revises the draft, then sends the final response

Our Approach: AI-driven QA-based Approach

Streamline Email Reply Tasks with “Questions from LLM”

Reply Process Using Our QA-based Approach



Reduces the burden of prompt creation and enhances email comprehension

Our Approach: AI-driven QA-based Approach

However, QA-based Approach could Negatively Affect...



...

Our Approach: AI-driven QA-based Approach

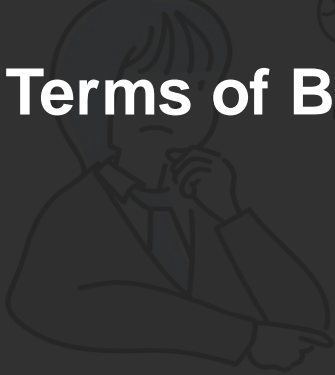
However, QA-based Approach could Negatively Affect...

Sense of Agency

Relationship with the Counterpart

Research Question (RQ):

**How Does a QA-based Approach Impact Formal Email Communication
in Terms of Benefits and Potential Risks?**



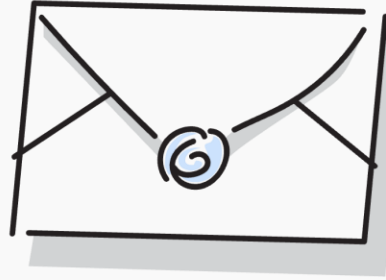
Our Goal

To Comprehensively Investigate How the QA-based Approach Affects...

RQ 1. Process



RQ 2. Quality



RQ 3. Relationship



Hypotheses

Formulated Eight Hypotheses across Three Research Questions:

RQ 1. Process



H1-a. Work Efficiency (↑)

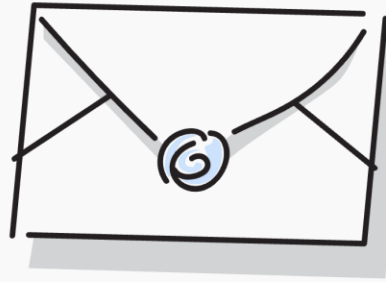
H1-b. Cognitive Load (↓)

H1-c. Work Satisfaction (↑)

H1-d. Difficulty in Task Initiation (↓)

H1-e. Sense of Agency (↓)

RQ 2. Quality



H2. Email Quality (↑)

RQ 3. Relationship



H3-a. User's Impression (↑)

H3-b. Psychological Distance (↑)

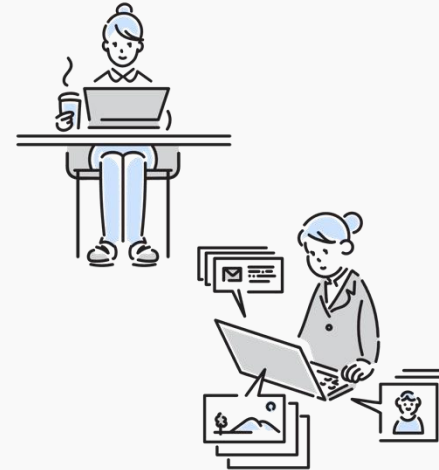
User Study

Conducted Two Types of User Studies to Validate the Hypotheses

Study 1. Controlled Experiment (N=12)

Study 2. Field Study (N=8)

Procedure



Replying Tasks

Questionnaire

5-Day System Usage

Interview

Purpose

Quantitatively understand
the proposed QA-based approach

Qualitatively understand
the proposed QA-based approach

User Study

Conducted Two Types of User Studies to Validate the Hypotheses

Study 1. Controlled Experiment (N=12)

Study 2. Field Study (N=8)

Procedure

Purpose



Replying Tasks

Questionnaire



5-Day System Usage

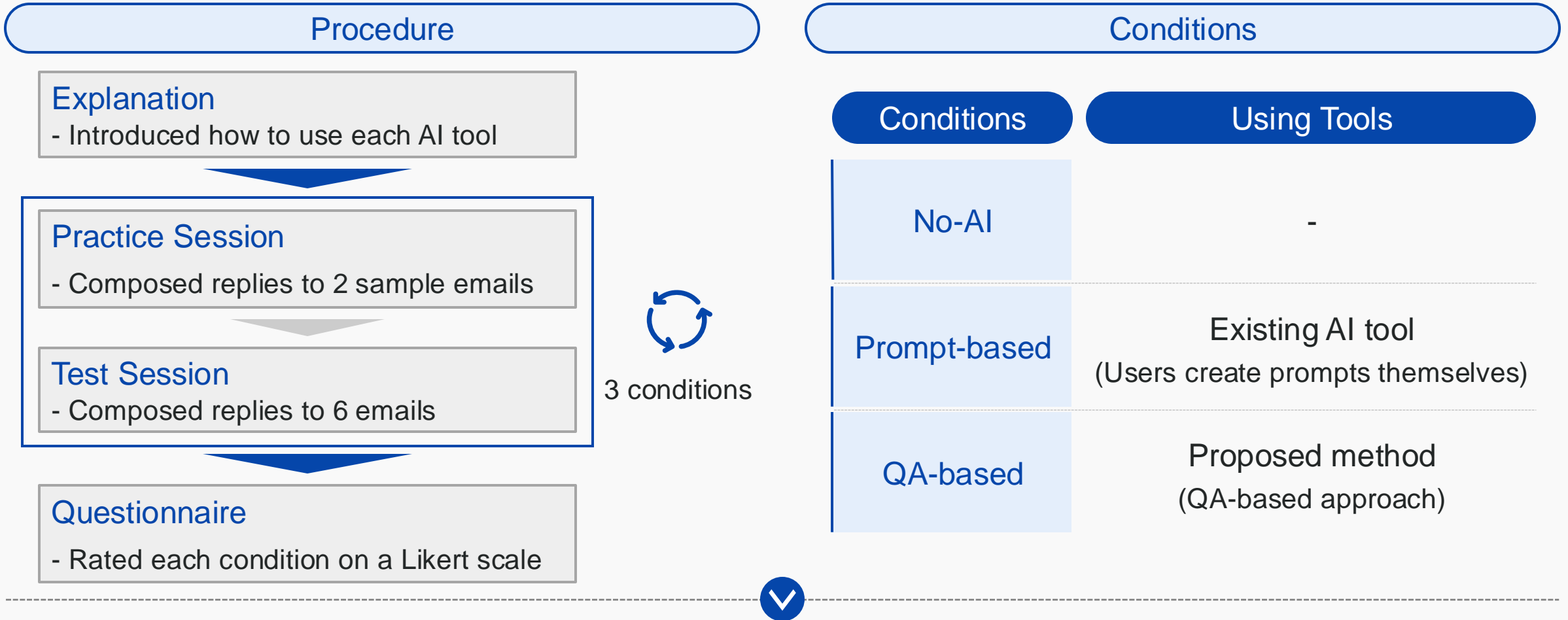
Interview

Quantitatively understand
the proposed QA-based approach

Qualitatively understand
the proposed QA-based approach

Study 1 (Controlled)

Participants Composed Replies to Fictional Emails Under Three Conditions



After the above experiment, additional evaluators (N=18) were recruited to assess the **quality and impression of the emails** created by participants

User Study

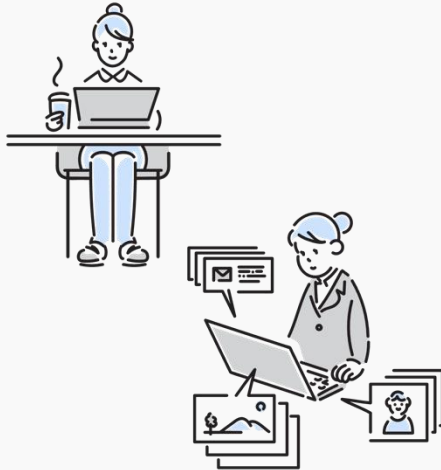
Conducted a Five-Day Field Study to Gain Further Insights

Study 1. Controlled Experiment (N=12)

Study 2. Field Study (N=8)

Procedure

Purpose



Replying Tasks

Questionnaire

5-Day System Usage

Interview

Quantitatively understand the proposed QA-based approach

Qualitatively understand the proposed QA-based approach

Study 2 (Field Study)

Participants Used the Prototype to Respond to Emails in Their Daily Lives

Procedure

Explanation

- Explained how to install and use the prototype

5-Day Prototype Usage

- Freely used the prototype to draft email replies for school or work

Interview

- Asked participants about their impressions

Interface of Our Prototype System

The screenshot displays the prototype interface for responding to an email. It is divided into three main sections: 'Incoming Mail', 'Questions by AI', and 'Customize Reply'.

- Incoming Mail:** Shows an email received at 7:30 PM from Ken Tanaka. The subject is 'Regarding Selection for the Ministry of Education Award'. The body of the email asks for information on four criteria: 1. Research lab, 2. Willingness to accept the award, 3. GPA, and 4. Future aspirations. A 'Reply' button is at the bottom.
- Questions by AI:** Lists the four criteria from the email as questions. The first question is 'Which research lab are you ultimately aiming to join for your graduation project?' with a dropdown menu showing 'Tanaka lab'. The second question is 'Are you willing to accept this award?' with 'Yes' and 'No' buttons. The third question is 'What is your current GPA?' with a text input field containing '3.55'. The fourth question is 'What are your current future aspirations?' with a dropdown menu showing 'Research career'. A 'New option' field with an 'Add' button is also present.
- Customize Reply:** Allows users to select their relationship (Professor or Student), formality (Informal, Neutral, Formal), tone (Friendly, Professional, Encouraging), and length (Long, Middle, Short). There is an 'Additional Request' text area and a 'Generate Reply' button.

Implemented and distributed as a Chrome extension

Results of Study 1 (Controlled)

Key Metrics and Results in Study 1

	Metric	Definition of Metric	Respondent	Result
RQ 1	Efficiency	How much text (that appears in the final email) is typed per second	-	Improved
	Cognitive Load	NASA-TLX score	Participants	Decreased
	Satisfaction	How satisfied users are with their efficiency and the quality of the email	Participants	Improved
	Difficulty in Task Initiation	How much of a psychological hurdle users feel before starting the task	Participants	Decreased
	Sense of Agency	How strongly users feel they created the reply	Participants	Decreased
RQ 2	Quality of Email	Politeness / Readability / Meeting Demands	Evaluators	Improved (vs. No-AI)
RQ 3	User's Impression	How favorable or kind an impression users create	Evaluators	No Difference
	Psychological Distance	How much psychological distance users feel toward the recipient	Participants	Increased (vs. No-AI)

Results of Study 1 (Controlled)

Improved Efficiency, Cognitive Load, Satisfaction, Difficulty in Task Initiation

	Metric	Definition of Metric	Respondent	Result
RQ 1	Efficiency	How much text (that appears in the final email) is typed per second	-	Improved
	Cognitive Load	NASA-TLX score	Participants	Decreased
	Satisfaction	How satisfied users are with their efficiency and the quality of the email	Participants	Improved
	Difficulty in Task Initiation	How much of a psychological hurdle users feel before starting the task	Participants	Decreased
	Sense of Agency	How strongly users feel they created the reply	Participants	Decreased
RQ 2	Quality of Email	Politeness / Readability / Meeting Demands	Evaluators	Improved (vs. No-AI)
RQ 3	User's Impression	How favorable or kind an impression users create	Evaluators	No Difference
	Psychological Distance	How much psychological distance users feel toward the recipient	Participants	Increased (vs. No-AI)

Results of Study 2 (Field Study)

Improved Perceived Efficiency, Workload, and Difficulty in Task Initiation



Improved Efficiency Reduced Workload

- Question-based organization of email points enabled a **faster grasp of the email content**
- Less manual input makes drafting replies easier, **like just filling out a questionnaire**



Lower Barrier to Start Work

- With reduced effort, participants were **less likely to postpone replying**

Results of Study 1 (Controlled)

QA-based Approach Improved Quality of the Email (vs. No-AI)

	Metric	Definition of Metric	Respondent	Result
RQ 1	Efficiency	How much text (that appears in the final email) is typed per second	-	Improved
	Cognitive Load	NASA-TLX score	Participants	Decreased
	Satisfaction	How satisfied users are with their efficiency and the quality of the email	Participants	Improved
	Difficulty in Task Initiation	How much of a psychological hurdle users feel before starting the task	Participants	Decreased
	Sense of Agency	How strongly users feel they created the reply	Participants	Decreased
RQ 2	Quality of Email	Politeness / Readability / Meeting Demands	Evaluators	Improved (vs. No-AI)
RQ 3	User's Impression	How favorable or kind an impression users create	Evaluators	No Difference
	Psychological Distance	How much psychological distance users feel toward the recipient	Participants	Increased (vs. No-AI)

Results of Study 2 (Field Study)

Participants Reported Improved Phrasing, Structure, Typos, and Content



Better Phrasing & Structure

- Offered more careful wording than usual



No Concerns About Typos

- Provided confidence that no typos were present



Fewer Missed Requests

- Lowered the chance of overlooking recipients' requirements

Results of Study 2 (Field Study)

Participants Reported Improved Phrasing, Structure, Typos, and Content



Better Phrasing & Structure

- Offered more careful wording than usual



No Concerns About Typos

- Provided confidence that no typos were present



Fewer Missed Requests

- Lowered the chance of overlooking recipients' requirements



QA-based Approach can serve as a **mediator for users' email communication**

Results of Study 1 (Controlled)

However, Users' Sense of Agency Decreased in the QA-based condition

	Metric	Definition of Metric	Respondent	Result
RQ 1	Efficiency	How much text (that appears in the final email) is typed per second	-	Improved
	Cognitive Load	NASA-TLX score	Participants	Decreased
	Satisfaction	How satisfied users are with their efficiency and the quality of the email	Participants	Improved
	Difficulty in Task Initiation	How much of a psychological hurdle users feel before starting the task	Participants	Decreased
	Sense of Agency	How strongly users feel they created the reply	Participants	Decreased
RQ 2	Quality of Email	Politeness / Readability / Meeting Demands	Evaluators	Improved (vs. No-AI)
RQ 3	User's Impression	How favorable or kind an impression users create	Evaluators	No Difference
	Psychological Distance	How much psychological distance users feel toward the recipient	Participants	Increased (vs. No-AI)

Results of Study 2 (Field Study)

3 out of 8 Participants Reported Their Sense of Agency Decreased



Sense of Agency Decreased

- Felt more like an editor or supervisor
- Hesitant to use AI in **important situations**



Sense of Agency Remained

- **Carefully checked** if their intentions and wording were accurately reflected

Results of Study 2 (Field Study)

5 out of 8 Participants Reported They Maintained Their Sense of Agency



Sense of Agency Decreased

- Felt more like an editor or supervisor
- Hesitant to use AI in **important situations**



Sense of Agency Remained

- **Carefully checked** if their intentions and wording were accurately reflected

Results of Study 2 (Field Study)

5 out of 8 Participants Reported They Maintained Their Sense of Agency



Sense of Agency Decreased

- Felt more like an editor or supervisor
- Hesitant to use AI in **important situations**



Sense of Agency Remained

- **Carefully checked** if their intentions and wording were accurately reflected



High AI mediation can still **maintain users' agency** if they **actively engage with the AI's output**

Results of Study 1 (Controlled)

No Difference in Impression, Increased Psychological Distance (vs. No-AI)

	Metric	Definition of Metric	Respondent	Result
RQ 1	Efficiency	How much text (that appears in the final email) is typed per second	-	Improved
	Cognitive Load	NASA-TLX score	Participants	Decreased
	Satisfaction	How satisfied users are with their efficiency and the quality of the email	Participants	Improved
	Difficulty in Task Initiation	How much of a psychological hurdle users feel before starting the task	Participants	Decreased
	Sense of Agency	How strongly users feel they created the reply	Participants	Decreased
RQ 2	Quality of Email	Politeness / Readability / Meeting Demands	Evaluators	Improved (vs. No-AI)
RQ 3	User's Impression	How favorable or kind an impression users create	Evaluators	No Difference
	Psychological Distance	How much psychological distance users feel toward the recipient	Participants	Increased (vs. No-AI)

Results of Study 2 (Field Study)

Changes in Psychological Distance Varied from User to User



Psychological Distance Reduced after Giving a Positive Impression

- **Quick & high-quality replies** created a favorable impression
- **The recipient also replied more quickly**, boosting a sense of closeness



Psychological Distance Increased

- Replying on AI made exchanges feel **superficial**
- Users **forgot the exchanges** became more frequent

Results of Study 2 (Field Study)

Changes in Psychological Distance Varied from User to User



Psychological Distance Reduced after Giving a Positive Impression

- **Quick & high-quality replies** created a favorable impression
- **The recipient also replied more quickly**, boosting a sense of closeness



Psychological Distance Increased

- Replying on AI made exchanges feel **superficial**
- Users **forgot the exchanges** became more frequent

Results of Study 2 (Field Study)

Changes in Psychological Distance Varied from User to User



Psychological Distance Reduced after Giving a Positive Impression

- **Quick & high-quality replies** created a favorable impression
- **The recipient also replied more quickly**, boosting a sense of closeness



Psychological Distance Increased

- Replying on AI made exchanges feel **superficial**
- Users **forgot the exchanges** became more frequent

Results of Study 2 (Field Study)

Changes in Psychological Distance Varied from User to User



Psychological Distance Reduced after Giving a Positive Impression

- **Quick & high-quality replies** created a favorable impression
- **The recipient also replied more quickly**, boosting a sense of closeness



Psychological Distance Increased

- Replying on AI made exchanges feel **superficial**
- Users **forgot the exchanges** became more frequent



The method and extent of AI mediation should be tailored to the **relationship and context**

Summary

Propose & Evaluate QA-based Approach with LLM to Support Email Replies

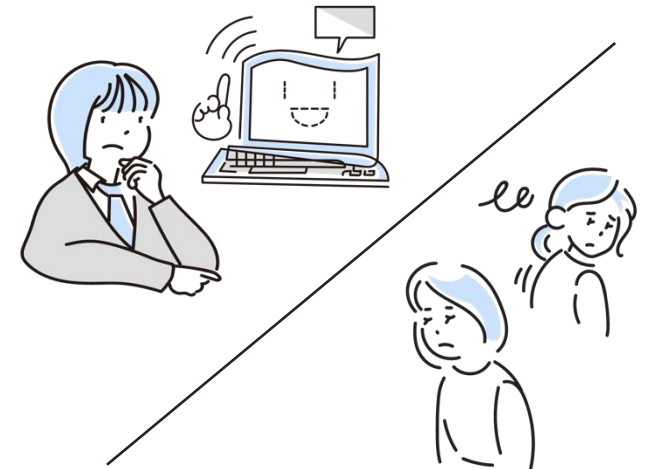
Our two user studies (controlled and field) revealed...



Higher Efficiency
Lower Workload
Reduced Procrastination



High-Quality Email Creation



**Potential Negative Impact on
Agency & Psychological Distance**



Future system designs should **balance efficiency with user agency**, adapting to each situation

Understanding and Supporting Formal Email Exchange by Answering AI-Generated Questions

Yusuke Miura

Waseda University

Chi-Lan Yang

The University of Tokyo

Masaki Kuribayashi

Waseda University

Keigo Matsumoto

The University of Tokyo

Hideaki Kuzuoka

The University of Tokyo

Shigeo Morishima

Waseda Research Institute
for Science and Engineering