

Email Label Guide

Each data point represents an entire conversation.

Under the “**messages**” section, you can see all the exchanged emails in order.

Goal of Labeling

1. **Provide examples** of how to respond in specific situations.
2. **Remove all sensitive data** from the emails.

Let's look at an example:

```
1
2 {
3     // Do not change
4     "id": 4,
5
6     // Should always be <NAME_1>
7     "sender_name": "<NAME_1>",
8
9     // Should always be <EMAIL_1>
10    "sender_email_address": "<EMAIL_1>",
11
12    // Do not change (this is not private)
13    "recipient_name": "DATASPORT-Info",
14
15    // Should always be <EMAIL_2>
16    "recipient_email": "<EMAIL_2>",
17
18    // Do not change
19    "billing_category": null,
20
21    // Do not change
22    "priority": "Normal",
23
24    // These are the exchanged emails
25    "messages": [
26        {
27            ### THIS IS A LABEL ###
28            // The AI model should respond to this email because it's a simple question.
29            "preferred_reaction": "respond",
30
31            ### THIS IS A LABEL ###
32            "category": "question",
33
34            // Do not change
35            "direction": "incoming",
36
37            // Do not change
38            "subject": "Kontakt Info",
39
40            ### ANONYMIZE ###
41            // For incoming messages, only anonymize
42            "body": "
43                Event: Lausanne Marathon ### THE EVENT NAME IS NOT PRIVATE!!! ###
44                First Name: <FIRST_NAME_1> ### THIS IS PRIVATE!!! MAKE SURE TO USE THE CORRECT ID ###
45                Last Name: <LAST_NAME_1>
46                Date of Birth: <BIRTH_DATE_1>
47                Address: <ADDRESS_1>
48                Email: <EMAIL_1>
49                Phone: <PHONE_NUMBER_1>
50                Message: Hello Team,
51                I have registered for Pontresina Run on <DATE_1>. This is the first time I'm traveling to Switzerland.
52                It would be nice if I could connect with some of the runners or running teams to gain confidence for this run.
53                Please guide me with it.
54                Regards, <FIRST_NAME_1>"
55        },
56        {
57            // For outgoing messages, do not change
58            "preferred_reaction": null,
59
60            // For outgoing messages, do not change
61            "category": null,
62
63            // Do not change
64            "direction": "outgoing",
65
66            // Do not change
```

```

67         "subject": "Re: Kontakt Info",
68
69         "body": "We are glad that you are participating. Unfortunately, we cannot share other participants' information due to data protection policies. Have a n
70     }
71 },
72
73     // Do not change
74     "language": "en",
75
76     // Do not change
77     "subject": "Kontakt Info"
78 }
79

```

Labels [↗](#)

The main labels are **preferred_reaction** and **category**. These fields must always be filled because the system has no way of processing them without your input.

Category [↗](#)

This label identifies the type of incoming message:

1. **question**: The sender is asking for information but not requesting an action.
 - Example: *"What is DataSport's phone number?"*
 - This question can be answered using publicly available information, such as the DataSport website or event site.
2. **request**: The sender is asking for an action to be taken.
 - Example: *"Refund my ticket purchase."*
 - This message requires an action to fulfill the request.
3. **info**: The sender is sharing information or expressing sentiment, and no direct response is needed.
 - Example: *"Thank you for your email, I am excited for the event!"*
 - A brief reply like *"Good luck during the event!"* can be sent, but it's not required.
4. **spam**: Unwanted or irrelevant messages.
 - Example: Unsolicited promotions, unrelated offers, or obvious spam content.

Preferred Reaction [↗](#)

This label describes how the AI should handle the message:

- **respond**: The AI should respond directly to the email.
 - Example: *"When does the Lausanne Marathon start?"*
 - Use for questions that can be answered with publicly available information or for simple requests that don't require human intervention.
- **Warning**: Not every question belongs in this category, and some requests might be answered instead of forwarded!
 - **Example 1**:
"Official registration closed two weeks ago. Can you register me via this email?"
 - Although it seems like a request, it can be answered: *"Sorry, we can't register anyone after the deadline."*
 - **Example 2**:
"A question that requires contacting someone from another company"
- **forward_to_datasport**:
 1. The AI should forward the email to a datasport member.
 - Example: *"Refund my ticket, I don't want to go."*
 - Use when the message needs a data sport team member to take action.

Same warning applies as to respond.
- **forward_to_organizer**
 1. The AI should forward the email to an organizer.
 - Example: *"What do i do with the luggage at the event?"*
 - Use when the message needs information outside the reach of datasport.
- 1. **nothing**: No action or response is needed.
 - Example: *"Thank you for your response."*
 - The email is acknowledged, but no further action is required, and it should not be forwarded either.

Anonymization: [↗](#)

Replace all private information with placeholders, except for the event name, as it adds important context to the email and is not considered private. Some fields may already be filled in, but always double-check for accuracy.

- **Anna** → `<FIRST_NAME_1>`
 - Note: The number "1" is an ID. If there are multiple names related to the sender, use `<FIRST_NAME_2>` for the second name.

Some Common Placeholders: [↗](#)

- `<NAME_1>` → Full name
- `<ADDRESS_1>` → Physical address
- `<EMAIL_1>` → Email address
- `<PHONE_NUMBER_1>` → Phone number

- <BANK_ACCOUNT_NUMBER_1> → Bank account number
- <IP_ADDRESS_1> → IP address
- <WEB_LINK_1> → Website link
- <VOUCHER_NUMBER_1> → Ds voucher numbers
- <INSURANCE_CODE_1> → Q42.JYR04.13-0021

Removal of Low-Quality Emails [↗](#)

If the conversation does not make sense at all due to parsing errors or other issues, remove it from the list.

If the conversation is missing emails remove it.

Changing Conversations [↗](#)

- If, due to parsing errors, two separate messages are combined into a single message, create a new message. Remove the second message from the first one and copy it into the newly created message.
- If you notice unnecessary data in the message (e.g., footer or included image data), remove it.
- If the order of messages is incorrect but makes sense when read in reverse, change the order accordingly.
- If you see a response that could have been handled better, revise it.

Example Found in Current Data: [↗](#)

Incoming: "Can you send me the phone number and email of DataSport?"

Outgoing: "Sure, here is our email: [x@y.com](#)"

Incoming: "I also want the phone number."

Outgoing: "Sorry, here is that too: +41...."

Revised Version:

Change the first response to: "Sure, here is our email: [x@y.com](#) and our phone number: +41...."

Cheat Sheet: [↗](#)

Label name	
category:	request ; question ; info ; spam
preferred_reaction:	respond ; forward_to_datasport forward_to_organizer ; nothing
direction:	incoming ; outgoing ; forwarded