Jojo Performance Review

TL; DR

- Jojo does too much explaining and doesn't answer the user's question.
    - Shorten responses
    - Provide tl;dr answer at the top of responses
    - Train more beyond "What is X thing" questions
- Jojo loves huge lists of bullet points
    - Diagrams and flow charts are a very good replacement
    - Increase diagram use, especially for long responses
    - Make diagrams even for non-technical subjects

Methodology

Used sample size of 10,000 chats. An LLM-as-a-Judge evaluated Jojo's conversations and user sentiment, found trends and key examples, and combined this with traditional statistical analysis to identify the root causes of user frustration.

In other words, ChatGPT and statistics go through the database of Jojo's chats and figure out what made users upset.

Root Problems

Too Much Yapping

Jojo does not get to the point or answer the core question quickly enough.

- 77% of chats are intended to be a single-question interaction. A huge majority of users just want the answer to one question from Jojo. In these chats:
    - The user only ever tries to get an answer to their initial question
    - If the conversation goes beyond a single prompt, it is out of frustration
    - Ideally, these should be resolved within one Jojo response
- However, only 62% of these cases are actually answered in a single response.
- In the remaining 38% of cases, users tend to re-ask the same question or request a more concise/clear explanation.
- The root cause is that Jojo does not focus on answering the question and yaps too much

Analysis

Many chats which should be one response (the student is just asking for a single question or simple task) drag on.

Chats Identified as "Should be 1 Response"
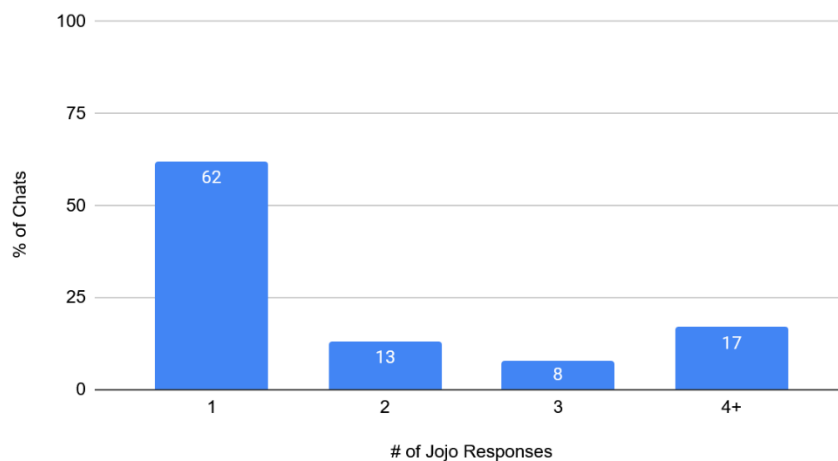
As chats get longer, the frustration sentiment spikes up around 50% compared to chats that end in one response. The following graph shows the average frustration for chats that should be one response. The frustration score is out of 100, with 0 being no frustration.

Frustration Sentiment as Responses Grow

# of Jojo Responses	Average Frustration
1	23
2	58
3	35
4+	39
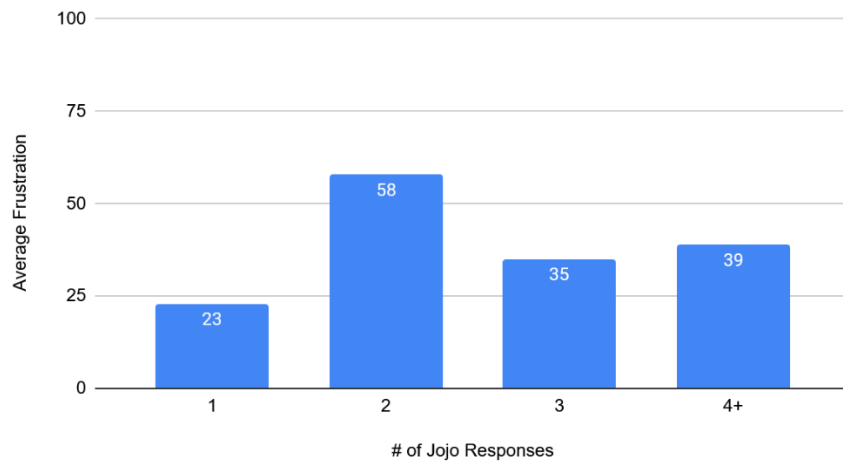
From these 2 graphs, we conclude that when Jojo fails to answer in one response, thus forcing the user to ask more follow ups, frustration grows.

Example

A typical example of how Jojo fails to answer in one response and how it plays out from there.

Chat ID: 6d986d74-e5e7-4433-8caa-817a530e7940

Question is asking why. Although vague, it's clear the student want the motivation behind data ratio, not just what it is

why is data ratio

JOJO AI

Let me help you understand the concept of ratio data.

📄 Found 5 results

3.6.1 Efficiency Ratios ↗
Business Mana...

Measurement and data processing ↗
Chemistry

1.1b.3 Design Guidance ↗
Design Technol...

3.5.1 Profitability Ratios ↗
Business Mana...

Jojo loves repeating what it is doing before and after getting sources. This happens in almost every response

Let me explain ratio data and its characteristics.

What is Ratio Data? Ratio data is the highest level of measurement that has all

Not what the student is asking plus a true zero point. This means that not

only can we measure the differences between values, but we can also make

meaningful statements about their ratios. **Very long yapping sentence that contributes basically nothing**

Key Characteristics of Ratio Data:

1. **True Zero Point** **In a "why" question, it's reasonable to explain what ratio data is, but you don't need to focus so much on it**

- Has a meaningful zero point where zero represents the complete absence of the measured quantity

- For example: Height, weight, revenue - these can all be zero and it means "none"

2. **Equal Intervals**

- The difference between any two consecutive units is the same

I cut out more yapping, Jojo's response is very long

Follow-on is not relevant, since the student asked a "why" question, it's better to ask if they want further explanation on something

Would you like me to provide some practice questions to test your understanding of ratio data?

in much much more simple terms help

Frustration

JOJO AI

Let me explain ratio data in a much simpler way.

Ratio data is numbers where: **Still yapping about what ratio data is, not why it exists, why it's used, or anything closer to what was being asked**

1. Zero means "nothing"

2. You can do any math with the numbers

3. You can say "this is twice as much as that"

More yapping below

Actionable Steps

- Make Jojo's responses shorter. Very few responses are too short, but many are too long.
  - It's better to let the user ask for more explanation if they need it, rather than info dumping on them at the beginning.
  - Shortening the response length also lets students focus on where exactly they want to take the conversation, instead of shotgunning everything about the topic.
- Have Jojo a short tl;dr answer at the top
- Maybe Jojo is overfitting to "explain what X thing is" type questions, so look for biases in the training set.
- Broaden training on common types of questions (what, why, how, etc.).
- Make Jojo's follow-up questions more relevant to what the user asks.
- GPT-5 is much more concise and focused than Sonnet 3.5, so this is already improving.

Walls of Bullet Points

Jojo loves to make huge lists with bullet points. When it tries to just list stuff out or doesn't show relationships visually, it can cause frustration. But when Jojo draws its own diagrams, it is very effective.

- Extremely few chats have Jojo making diagrams.
- You almost always must ask Jojo directly to make a diagram. I'm not sure if users even know this is a feature since it isn't mentioned anywhere.
- The cases where it does do make diagrams are very effective.
- Jojo shouldn't be scared to draw diagrams even for non-technical subjects like TOK, since visual learning aids are always good.
  - Usually, the diagrams are flow charts or text boxes anyways, which help with every subject.

Analysis

In general, frustration in chats with diagrams is much lower. However, the sample size of chats with diagrams is very small. This graph proves that diagrams are effective, but not necessarily that they are better than just listing stuff out without them.

Frustration vs Diagrams
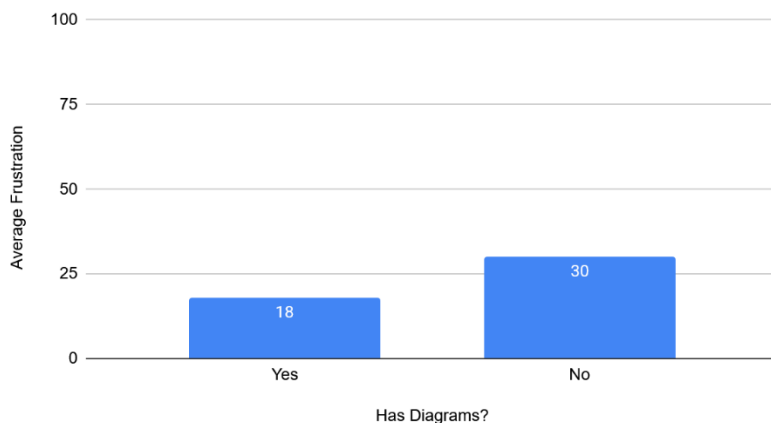
However, many chats would benefit from having diagrams. 31% of chats could benefit from diagrams, which is a huge percentage.

% of all Chats vs Benefit from Diagrams
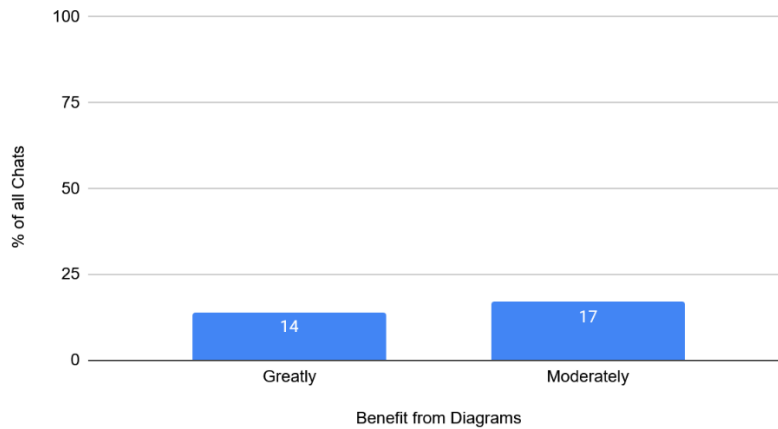
These stats show that more diagrams will help with the common issue of users not wanting to read a big boring wall of text.

Examples

A great place to use diagrams is when Jojo would normally list out a bunch of points. We can instead make a diagram to show relationships between concepts visually, rather than having a wall of bullet points.

Chat ID: bf50b7f0-91d8-4584-82bf-c86c7c98b40a (second message)

Prompt: "tell me the caracteristics of free trade"

Jojo's response (just a bunch of lists):

Conditions for Successful Free Trade:

1. **Market Efficiency**

- Competitive markets
- Good information flow
- Low transaction costs

2. **Economic Stability**

- Sound monetary policies
- Stable exchange rates
- Strong financial systems

3. **Infrastructure**

- Efficient transportation

Benefits of Free Trade:

1. **Economic Efficiency**

- Countries specialize in their comparative advantage
- More efficient allocation of resources
- Higher overall productivity

2. **Consumer Benefits**

- Lower prices due to competition
- Greater variety of products
- Better quality goods and services

3. **Market Benefits**

- Access to larger markets
- ~~Economies of scale~~

Potential Drawbacks:

1. **Short-term Challenges**

- Structural unemployment in some sectors
- Domestic industry adjustment
- Income inequality concerns

2. **Economic Dependencies**

- Countries may become overly dependent on imports
- Vulnerability to global market fluctuations
- Possible trade deficits

We could group these subpoints together in a flowchart, connecting the subpoints to boarder key concepts. It is much nicer and easier to read than so many bullet points.

Actionable Steps

- Tell users that they can ask for diagrams on the UI
- Tell Jojo to make diagrams more frequently, especially when it would otherwise just make a huge list
- Expand diagrams to non-technical subjects
- Could also have users set preferences for if they are a visual or hands-on learner to further customize their experience
    - E.g. if they select visual, they would be given more diagrams. If they select a different learning style, their interactions with Jojo would look different
    - More of a general suggestion, not sure how feasible or useful it is

Other Problems

List of more minor or statistically small problems found during the data gathering process.

- Jojo gives broken latex
    - LLMs are pretty good at latex in general, so I'm not sure what's causing this
    - Answers are usually still understandable with the broken latex
- Jojo leaves out units (e.g. m/s^2)
- "Tip", "Note", etc. text boxes get cut off

💡 Tip

For English B writing tasks, remember to:

- Use appropriate paragraph breaks
- Include clear transitions between ideas
- Maintain consistent punctuation
- Use bullet points for lists
- Include appropriate spacing between sections