

Course Correction: Steering AI for Language Learning

Alex Hollinghead

McGraw Center for Teaching & Learning, Princeton University

Hisae Matsui

Department of East Asian Studies, Princeton University



Project Goals: JPN101 Chatbot

- Build a chatbot for beginner-level Japanese practice (JPN101).
- Chat conversations should be structured to reinforce grammar and vocabulary from the curriculum.
- No unexpected grammar or unfamiliar vocabulary.
- Build a replicable, open-source system that can be adapted for other courses

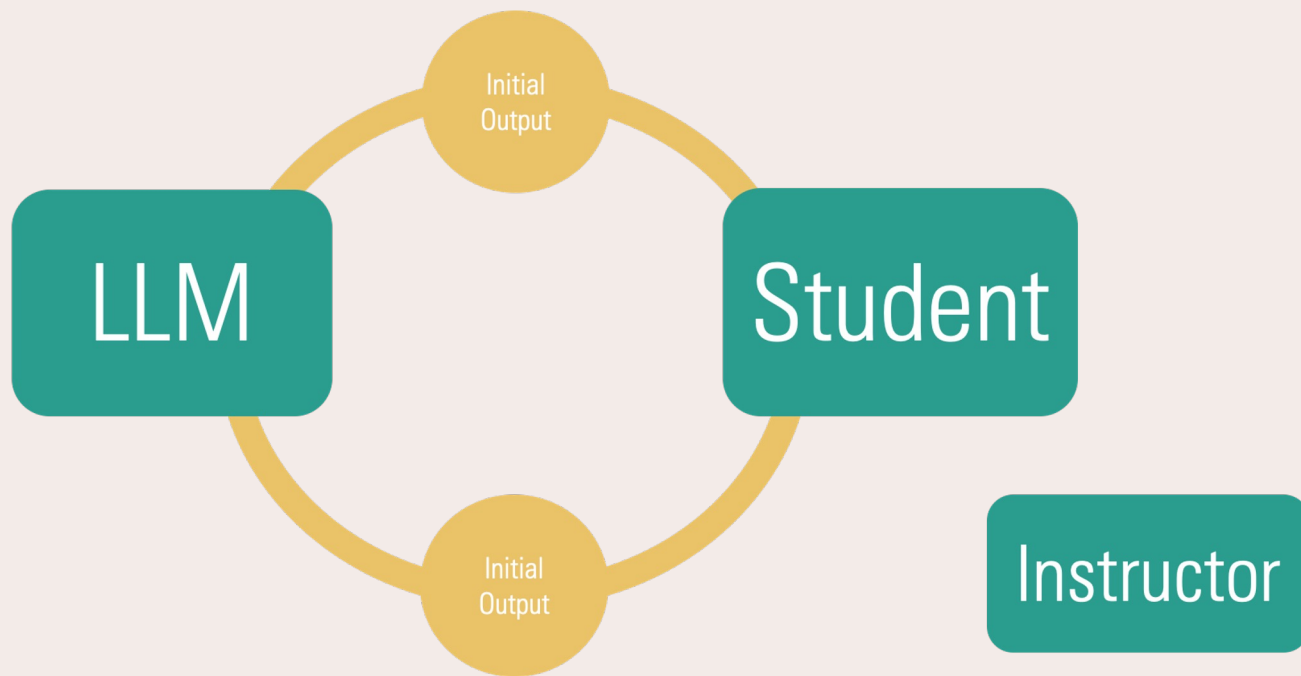


The Challenge: Beginner-Level Learners

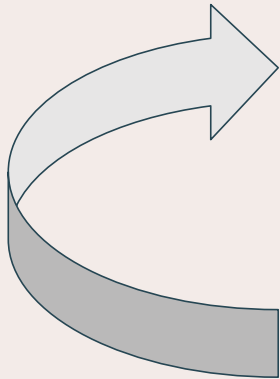
- LLM outputs need to be level-appropriate and align with the curriculum.
- Attenuating grammar and vocabulary levels of LLM outputs to specific targets is difficult.



The Challenge: Instructor Out-of-the-Loop



Two Approaches to Control Outputs

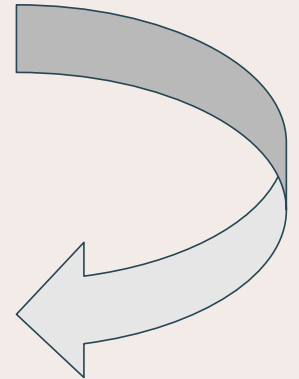


1. Fine-Tuning

Train the model on message-response pairs that exemplify grammar from each chapter.

2. Prompt Engineering

Use carefully designed prompts to align the model's output with each activity's learning goals.



Fine-Tuning Strategy

```
1 {  
2   "messages": [{  
3     "role": "system",  
4     "content": "Use は and です in statement sentences."  
5   }, {  
6     "role": "user",  
7     "content": "アリスさんは アメリカじんです。"  
8   }, {  
9     "role": "assistant",  
10    "content": "そうですか。わたしは にほんじんです。"  
11  }]  
12 } {  
13   "messages": [{  
14     "role": "system",  
15     "content": "Use は and です in statement sentences."  
16   }, {  
17     "role": "user",  
18     "content": "よこたさんは にほんじんです。"  
19   }, {  
20     "role": "assistant",  
21     "content":  
22       "そうですか。マルティネスさんは アメリカ人です。"  
23   }]  
24 } {  
25   "messages": [{  
26     "role": "system",  
27     "content": "Use は and です in statement sentences."  
28   }, {  
29     "role": "user",  
30     "content": "うえださんは だいがくせいです。"  
31   }, {  
32     "role": "assistant",  
33     "content":  
        "そうですか。デュランさんは だいがくせいじゃありません  
        。"  
    }]  
34 }
```

- Vocabulary is explicit, so no need to finetune
- Grammar: message-response pairs that exemplify “Grammar Points”
- Examples produced with an LLM and reviewed by human instructor
- Evaluated outputs with uniform set of input prompts
- We found diminishing returns after ~15 training examples per grammar point

Grammar Points

1	system	user	assistant
2	Use は and です in statement sentences.	アリスさんは アメリカ人です。	そうですか。わたしは にほんじんです。
3	Use は and です in statement sentences.	よこたさんは にほんじんです。	そうですか。マルティネスさんは アメリカ人です。
4	Use は and です in statement sentences.	うえださんは だいがくせいじゃありません。	そうですか。デュランさんは だいがくせいじゃありません。
5	Use は and です in statement sentences.	すぎうらさんは だいがくせいじゃありません。	そうですか。ウーさんは だいがくせいじゃありません。
6	Use は and です in statement sentences.	たなかさんは さんねんせいじゃありません。	そうですか。わたしは いちねんせいじゃありません。
7	Use は and です in statement sentences.	うえださんは せんせいじゃありません。	そうですか。わたしは がくせいじゃありません。
8	Use は and です in statement sentences.	たしろさんは にほんじんです。	そうですか。ホホワイトさんは カナダじんです。
9	Use は and です in statement sentences.	よしのさんは にねんせいじゃありません。	そうですか。あかぎさんは さんねんせいじゃありません。
10	Use は and です in statement sentences.	ロベスさんは スペインじんです。	そうですか。ほりさんは にほんじんです。
11	Use は and です in statement sentences.	リーさんは だいがくせいじゃありません。	そうですか。わたしは だいがくいんせいじゃありません。
12	Use は and です in statement sentences.	たしろさんは がくせいじゃありません。	そうですか。わたしは せんせいじゃありません。
13	Use は and です in statement sentences.	アリさんは ちゅうごくじんじゃありません。	そうですか。リーさんは かんこくじんじゃありません。
14	Use は and です in statement sentences.	トーマスさんは カナダじんです。	そうですか。ハリスさんは イギリスじんです。
15	Use は and です in statement sentences.	たなべさんは いちねんせいじゃありません。	そうですか。わたしは にねんせいじゃありません。
16	Use は and です in statement sentences.	さかもとさんは せんせいじゃありません。	そうですか。よしださんは がくせいじゃありません。
17	Use は and です in statement sentences.	ホンさんは たいわんじんです。	そうですか。ワンさんは ちゅうごくじんじゃありません。
18	Use は and です in statement sentences.	いとうさんは いちねんせいじゃありません。	そうですか。ジャンさんは さんねんせいじゃありません。
19	Use は and です in statement sentences.	シュミットさんは アメリカ人じゃありません。	そうですか。ミュラーさんは カナダじんです。
20	Use は and です in statement sentences.	なかむらさんは にねんせいじゃありません。	そうですか。わたしは よねんせいじゃありません。
21	Use は and です in statement sentences.	ウォーカーさんは りゅうがくせいじゃありません。	そうですか。わたしは りゅうがくせいじゃありません。
22	When you are asked a question using '～は～ですか', answer using 'はい/いいえ' followed by the appropriate response.	アメリカ人じゃありません。	いいえ、そうじゃありません。
23	When you are asked a question using '～は～ですか', answer using 'はい/いいえ' followed by the appropriate response.	アリスさんは せんせいじゃありません。	いいえ、せんせいじゃありません。
24	When you are asked a question using '～は～ですか', answer using 'はい/いいえ' followed by the appropriate response.	いちねんせいじゃありません。	はい、そうです。
25	When you are asked a question using '～は～ですか', answer using 'はい/いいえ' followed by the appropriate response.	うえださんは いちねんせいじゃありません。	はい、そうです。
26	When you are asked a question using '～は～ですか', answer using 'はい/いいえ' followed by the appropriate response.	がくせいは アメリカ人じゃありません。	はい、アメリカ人じゃありません。
27	When you are asked a question using '～は～ですか', answer using 'はい/いいえ' followed by the appropriate response.	がくせいは カナダじんです。	いいえ、カナダじんじゃありません。
28	When you are asked a question using '～は～ですか', answer using 'はい/いいえ' followed by the appropriate response.	がくせいは せんせいじゃありません。	いいえ、せんせいじゃありません。
29	When you are asked a question using '～は～ですか', answer using 'はい/いいえ' followed by the appropriate response.	がくせいは ちゅうごくじんじゃありません。	はい、ちゅうごくじんじゃありません。
30	When you are asked a question using '～は～ですか', answer using 'はい/いいえ' followed by the appropriate response.	がくせいは にほんじんです。	いいえ、にほんじんじゃありません。
31	When you are asked a question using '～は～ですか', answer using 'はい/いいえ' followed by the appropriate response.	がっこうは ウェストサイドだいがくですか。	はい、そうです。
32	When you are asked a question using '～は～ですか', answer using 'はい/いいえ' followed by the appropriate response.	キムさんは かんこくじんじゃありません。	はい、そうです。
33	When you are asked a question using '～は～ですか', answer using 'はい/いいえ' followed by the appropriate response.	こちらは つむらさんですか。	はい、そうです。
34	When you are asked a question using '～は～ですか', answer using 'はい/いいえ' followed by the appropriate response.	こちらは てづかさんですか。	いいえ、てづかさんじゃありません。
35	When you are asked a question using '～は～ですか', answer using 'はい/いいえ' followed by the appropriate response.	さとうさんは いちねんせいじゃありません。	いいえ、そうじゃありません。

Our Prompt Components

We assemble prompts for the LLM in a modular fashion:

Mode Prompt

"You are engaging in roleplaying conversations to help college students practice learning the Japanese language. You respond to the messages..."

The "mode prompt" describes the way the LLM is being used

Activity Prompt

"You are a university student in Japan. You are having breakfast at your dormitory's common dining area, talking to a Japanese roommate..."

The activity prompt describes the setting and context of the roleplaying scenario

Materials

- Grammar points that the student has seen so far
- Vocabulary that the student has learned so far

Grammar points and vocabulary materials that the students have been exposed to already

Objectives

- The student uses わたしは～ to start a sentence
- The student Uses time expressions like 'ごぜん7じに

Objectives that reflect desired learning outcomes to guide the conversation.

Conversation Objectives

- We leverage the “structured outputs” feature offered by many LLM APIs
- The LLM response contains a field “objective_complete”
- If the model determines that the current objective is satisfied, the flag switches to true
- The UI checks off the objective and our system provides a new objective to the model
- This keeps the LLM on task!

```
{  "name": "objective_tracking",

  "schema": {
    "type": "object",
    "properties": {
      "content": {
        "type": "string"
      },
      "objective_complete": {
        "type": "boolean"
      }
    },
    "required": [
      "content",
      "objective_complete"
    ],
    "additionalProperties": false
  },
  "strict": true
}
```

Activity Designer

Select Model* FT - GPT 4.1 ▼ Select Class* JPN 101 ▼

Vocabulary	0 active	<
Grammar	0 active	<

Select Activity [Create new](#)

Activity Prompt
Enter new activity prompt here

Conversation Objectives
Complete these objectives in sequence

Message Send

Overview of Evaluation Criteria

- Evaluated the performance of four AI models in a Japanese roleplay scenario designed for beginner-level college students.
- Models Analyzed:
 - FT_4.1, Base_4.1, FT_4o, Base_4o
- Number of Transcripts:
 - 10 conversations per model
- The evaluation focuses on:
 - Vocabulary Adherence
 - Grammar Adherence
 - Adherence to Specific Instructions (e.g., Kanji Usage)
 - Parroting
 - Other Errors (e.g., Semantic or Severe Syntax Errors)



Results

- Improvements:
 - Simpler, shorter, easier to understand responses.
 - Strayed from selected vocabulary less often.
 - Better adherence to grammar points.
- ❖ Fine-tuned models still occasionally included unfamiliar items.
- ❖ Fine-tuned models tended to “parrot”

Metric	4o mini	4o mini Finetuned	4.1	4.1 Finetuned
Total AI Conversational Turns	110	114	96	80
Vocabulary Violations	151	82	26	11
Grammar Violations	68	32	10	0
Kanji Use Violations	283	39	5	0
Parroting Instances	3	13	16	25
Other Errors (Semantic/Severe Syntax)	2	2	1	1



Parroting <FT 4.1>

1. user: こんにちは
2. assistant: こんにちは、まついさん。きのうは なにを しましたか。
3. user: にほんごをべんきょうしました
4. assistant: そうですか。きのうは にほんごを べんきょうしましたね。まついさんは まいにち なにを しますか。
5. user: まいにちべんきょうします。そしてテニスをします
6. assistant: そうですか。まいにち べんきょうしますね。そして、テニスを しますか。たいてい なんじに おきますか。
7. user: たいてい いろいろじにおきます
8. assistant: そうですか。たいてい いろいろじに おきますね。まついさんは たいてい どこで べんきょうしますか。
9. user: たいてい としょかんでべんきょうします
10. assistant: そうですか。たいてい としょかんで べんきょうしますね。しゅうまつは なにを しますか。
11. user: ほんをよみます
12. assistant: そうですか。しゅうまつは ほんを よみますね。わたしは しゅうまつ えいがを みます。まついさんは えいがが すきですか。
13. user: はい。すみません。クラスがあります
14. assistant: はい、わかりました。じゃあ、また。
15. user: じゃあ、また
16. assistant: じゃあ、また。

1. user: Hello.
2. assistant: Hello, Matsui-san. What did you do yesterday?
3. user: I studied Japanese.
4. assistant: I see. You studied Japanese yesterday. What do you do every day, Matsui-san?
5. user: I study every day. And I play tennis.
6. assistant: I see. You study every day. And you play tennis, huh? What time do you usually get up?
7. user: I usually get up at six o'clock.
8. assistant: I see. You usually get up at six. Where do you usually study, Matsui-san?
9. user: I usually study at the library.
10. assistant: I see. You usually study at the library. What do you do on weekends?
11. user: I read books.
12. assistant: I see. You read books on weekends. I watch movies on weekends. Do you like movies, Matsui-san?
13. user: Yes. Sorry, I have class.
14. assistant: Okay, understood. See you later.
15. user: See you later.
16. assistant: See you later.



(Not) Parroting <Base 40>

1. assistant: まついさんは、平日には何をしますか？日常のルーティンを教えてください。
2. user: あさ、おきます。そして、あさごはんをたべます。それからだいがくにいきます
3. assistant: いいですね！大学で何をしますか？授業や勉強ですか？
4. user: にほんごをべんきょうします
5. assistant: それは素晴らしいです！日本語の授業の後、何をしますか？たとえば、友達と遊ぶとか、宿題をするなどですか？
6. user: ほんをよみます
7. assistant: 本を読むのは良いですね！どんな本を読みますか？
8. user: ミステリーのはんをよみます
9. assistant: ミステリーの本はおもしろいですね！まついさんは、どのくらいの頻度で本を読みますか？例えば、毎日とか、週に何回とかですか？
10. user: まいにちはんをよみます
11. assistant: 毎日本を読むのは素晴らしい習慣ですね！では、週末には何をしますか？友達と遊ぶ予定などありますか？
12. user: レストランにいきます
13. assistant: レストランに行くのは楽しみですね！どんな料理を食べる予定ですか？
14. user: ピザをたべます
15. assistant: ピザはおいしいですね！どんなトッピングのピザが好きですか？
16. user: ソーセージです
17. assistant: ソーセージのトッピングはとてもおいしいですね！ピザを食べるとき、飲み物は何を飲みますか？
18. user: みずをのみます。

1. assistant: Matsui-san, what do you do on weekdays? Please tell me about your daily routine.
2. user: In the morning, I wake up. Then, I eat breakfast. After that, I go to the university.
3. assistant: Sounds good! What do you do at the university? Classes or studying?
4. user: I study Japanese.
5. assistant: That's wonderful! What do you do after Japanese class? For example, do you hang out with friends or do homework?
6. user: I read books.
7. assistant: Reading books is great! What kind of books do you read?
8. user: I read mystery books.
9. assistant: Mystery books are interesting! How often do you read books? For example, every day or a few times a week?
10. user: I read books every day.
11. assistant: Reading every day is a wonderful habit! Then, what do you do on the weekends? Do you have plans to hang out with friends?
12. user: I go to restaurants.
13. assistant: Going to restaurants sounds fun! What kind of food are you planning to eat?
14. user: I'll eat pizza.
15. assistant: Pizza is delicious! What kind of toppings do you like on your pizza?
16. user: Sausage.
17. assistant: Sausage toppings are very tasty! What do you drink when you eat pizza?
18. user: I drink water.



Nonsensical sequence <Base 4o>

1. assistant: . . . 他には、大学でどんなことをしますか？
2. user: にほんごをべんきょうします
3. assistant: 日本語を勉強するんですね！とても大事ですね。高校生たちはどうですか？他に何かその後にしますか？
4. user: テニスをします

1. assistant: ...What else do you do at the university?
2. user: I study Japanese.
3. assistant: You study Japanese! That's very important. How about the high school students? Do you do anything else after that?
4. user: I play tennis.



Current Challenges

Some barriers to incorporating this tool into coursework without the instructor in the loop:

- Even the best models still occasionally deviate from vocabulary restrictions
- Evaluating models for allowable outputs will become more complex as allowable grammar grows
- Some models occasionally take conversations in nonsensical/meandering directions
- Parroting responses interrupt conversation flow and make the conversation unnatural.



Potential Ways Forward

- Leverage structured outputs for Chain of Thought?
- Reduce parrot responses via prompt modifications
- Could we train a classifier model to identify outputs that violate grammatical restrictions?
- Filtering model outputs against vocabulary list

Questions & Discussion

