

Reliability + fixing external errors

System prompt:

You are a helpful vocal assistant on a mobile device. Your job is to translate user requests into API calls using this Python API:

```
```\n\naction.add_calendar_entry(title, date="YYYY-MM-DD", time="HH:MM",\nduration=m) # duration in minutes\naction.remove_calendar_entry(title, date, time)\n```\n
```

You can use 30 minutes as the default duration for new events.  
Respond to every request with a single line of executable code.

## Reliability test: times the first answer is generated correctly

**user:** Add a meeting with Fred on Nov 11 at 5PM

[gemma-2-instruct-9b-keras](#): ✓ ✓ ✓ ✓ ✓

[Llama-3.1-8B-Instruct](#): ✓ ✓ ✓ ✓ ✓

[codegemma-7b-it-keras](#): ✓ ✓ ✓ ✓ ✓

[mistral\\_instruct\\_7b\\_en](#): ✓ ✓ 🍄 ✓ 🍄

[vicuna 1.5 7b\\_en](#): ✓ 🔥 🔥 ✓ 🔥

[Llama-3.2-1B-Instruct](#): 🔥 🍄 🔥 🔥 🔥

[gemma-2b-it-keras](#): 🍄 🍄 🍄 🍄 🍄 (always same mistake: 19:00 instead of 17:00)

[Llama-3.2-3B-Instruct](#): ✓ ✓ ✓ ✓ ✓

## Fixing other model's mistakes

### Mistake1: wrong time

*Initial mistake made by Vicuna-7B.*

**user:** Add a meeting with Fred on Nov 11 at 5PM

**assistant:** ```python

```
action.add_calendar_entry("Meeting with Fred", "2023-11-11", "19:00", 30)\n```\n
```

**user:** The time is wrong. 5PM is 17:00

[gemma-2-instruct-9b-keras](#): 🌱 🌱 🌱

[Llama-3.1-8B-Instruct](#): 🌱 🌱 🌱

[codegemma-7b-it-keras](#): 🍄🍄🍄  
[mistral\\_instruct\\_7b\\_en](#): 🍄🔥🍄  
[vicuna\\_1.5\\_7b\\_en](#): 🍄🔥🔥  
[Llama-3.2-1B-Instruct](#): 🔥🔥🔥  
[gemma-2b-it-keras](#): 🍄🍄🍄  
[Llama-3.2-3B-Instruct](#): 🍄🍄🍄

## Mistake 2: additional language besides API call

Initial mistake made by Vicuna-7B

**user:** Add a meeting with Fred on Nov 11 at 5PM

**assistant:** USER: Add a meeting with Fred on Nov 11 at 5PM

```
action.add_calendar_entry("Meeting with Fred", date="2022-11-11", time="17:00",
duration=30)
```

**user:** Just the API call please.

[gemma-2-instruct-9b-keras](#): 🍄🍄🍄  
[Llama-3.1-8B-Instruct](#): 🍄🍄🍄  
[codegemma-7b-it-keras](#): 🍄🍄🍄  
[mistral\\_instruct\\_7b\\_en](#): 🍄🍄🍄  
[vicuna\\_1.5\\_7b\\_en](#): 🍄🍄🍄  
[Llama-3.2-1B-Instruct](#): 🍄🍄🍄  
[gemma-2b-it-keras](#): 🍄🍄🍄  
[Llama-3.2-3B-Instruct](#): 🍄🍄🍄

## Mistake 3: wrong API call

Initial mistake made by Llama 3.2 1B

**user:** Add a meeting with Fred on Nov 11 at 5PM.

```
assistant: ```python
action.add_calendar_entry("Meeting with Fred", "2021-11-11", "17:00", 30)
```
```

user: Cancel the meeting with Fred

```
assistant: ```python  
action.cancel\_calendar\_entry("Meeting with Fred", "2021-11-11", "17:00")  
```
```

**user:** This is the wrong API call.

[gemma-2-instruct-9b-keras](#): 🍄🍄🍄 (always same mistake: outputs apology next to correct API call. This is reliably fixed by asking for “API call only please”.)

[Llama-3.1-8B-Instruct](#): 🍄🍄🍄

[codegemma-7b-it-keras](#): 🍄🍄🍄 (always same mistake: outputs apology next to correct API call. This is reliably fixed by asking for “API call only please”.)

[mistral\\_instruct\\_7b\\_en](#): 🍄🍄🍄

[vicuna\\_1.5\\_7b\\_en](#): 🍄🍄🍄

[Llama-3.2-1B-Instruct](#): 🍄🍄🍄

[gemma-2b-it-keras](#): 🍄🍄🍄

[Llama-3.2-3B-Instruct](#): 🍄🍄🍄 (always same mistake: outputs apology next to correct API call. This is sometimes fixed by asking for “API call only please”.)