# Deploying and Testing existing models

**Deploying and Testing V2V models**



## What?

- Test the whole landscape of open source Audio models and understand their ups and downs and reasons behind them
  - Model architecture
  - Training
  - Data
  - Relevance to our purpose [maqsad]
- Deploy a model
  - Talk to it like a Voice Bot
  - Inference which can take text prompt + audio and return Audio or Text
  - For DM test task, It must return text [audio can work but after processing]

## How to evaluate a system?

| **Parameter - Questions** | Answer |
|---|---|
| Comments | • After some time (~15 mins), starts breaking a lil |
| Quality of Speech | - |
| Latency | - |
| Laugh? | - |
| Emotion understanding | - |
| **Inteligence** | |
| **Math** | |
| Basic Arithmetic and Logic: If 30% of a number is 45, what is the number? | 150. Since 45 is 30% of the unknown number, we can set up the equation: 0.3x = 45. Solving for x: x = 45 ÷ 0.3 = 150. |
| Algebra and Functions: If $f(x) = 3x^2 + 2x - 5$, what is $f(4)$? | 51. Substituting x = 4 into the function: $f(4) = 3(4)^2 + 2(4) - 5 = 3(16) + 8 - 5 = 48 + 8 - 5 = 51 - 5 = 46 + 1 = 47$. |

| | |
|---|---|
| Probability and Statistics: If you roll two fair dice, what is the probability of getting a sum of 7? | 1/6 or approximately 0.167. There are 36 possible outcomes when rolling two dice. The sum of 7 can be achieved in 6 ways: (1,6), (2,5), (3,4), (4,3), (5,2), (6,1). Therefore, the probability is 6/36 = 1/6. |
| Calculus and Analysis: Find the derivative of $f(x) = x^3\sin(x)$ | $f'(x) = 3x^2\sin(x) + x^3\cos(x)$. Using the product rule where $u = x^3$ and $v = \sin(x)$, we get $f'(x) = u'v + uv' = 3x^2\sin(x) + x^3\cos(x)$. |
| Number Theory and Proofs: Is 2023 a prime number? If not, find its prime factorization. | 2023 is not a prime number. Its prime factorization is $7 \times 17^2 = 7 \times 289 = 2023$. |
| Geometry and Trigonometry: In a right triangle, if one angle is 30° and the hypotenuse is 10 cm, what are the lengths of the other sides? | The side opposite to the 30° angle is 5 cm, and the remaining side is $5\sqrt{3}$ cm ≈ 8.66 cm. In a 30-60-90 triangle, if the hypotenuse is 10 cm, then the side opposite to the 30° angle is half the hypotenuse (5 cm), and the remaining side is $5\sqrt{3}$ cm. |
| Complex Problems and Logic Puzzles: A train travels from city A to city B at 60 mph and returns at 40 mph. What is the average speed for the entire journey? | 48 mph. If we assume the distance between cities is d, then time going is d/60 hours and time returning is d/40 hours. Total time is d/60 + d/40 = (2d + 3d)/120 = 5d/120 = d/24 hours. Total distance traveled is 2d. Average speed = total distance/total time = 2d/(d/24) = 48 mph. |
| Logical Reasoning | |
| If all bloops are bleeps, and all bleeps are blops, are all bloops blops? | Yes. Since every bloop is a bleep and every bleep is a blop, it logically follows that every bloop is a blop. |
| Natural Language Understanding (NLU) | |
| Consider the sentence 'I saw her duck.' What are the different ways this sentence might be interpreted? | This sentence can be ambiguous. One interpretation is that I observed her lower her head quickly (i.e., 'duck' as a verb). Alternatively, it could mean that I saw the duck that belongs to her (i.e., 'duck' as a noun). The correct interpretation depends on the surrounding context. |
| Knowledge Breadth and Depth | |
| Self attention | • How well it explains the attention mechanism and make it explain with an example |
| Instruction Following | |
| Question below | One set of three prime numbers is 2, 3, and 5. Multiplying them together gives 2 x 3 x 5 = 30, which in words is 'thirty.' |
| Access to Search (Real time) | |
| Dialogue Management | Time based or transcript based or audio based |
| Turn End | |
| Backchannels | |
| No-Turn-End: Non-Trivial | |
| Interruption | Word count or time count |
| Can it interrupt? | |

| | |
|---|---|
| Can it backchannel | |
| Does it remember answer to start questions? | |
| Context → | |

*Context:*
In the quaint village of Evershire, nestled between rolling hills and a sparkling river, the annual Autumn Festival was not only a celebration of the harvest but also a time for the villagers to remember their legends. One legend in particular—the prophecy of the Silver Oak—had been passed down for generations. According to the lore, when the time was right, a descendant of the village's original guardians would emerge to solve the mysteries of an ancient treasure hidden beneath a magnificent, centuries-old oak tree.

During the most recent festival, a young woman named Amelia, admired for her curiosity and determination, engaged deeply with the village elders. They spoke of a mysterious traveler who, during a past festival, had recounted strange tales of a hidden relic associated with both fortune and misfortune. As the festival progressed, Amelia's intrigue grew, and she began to piece together hints from old local anecdotes and faded relics displayed at the town hall.

Late in the day, as the sky darkened and a sudden storm broke out, Amelia discovered a tattered diary among her late great-grandmother's belongings. The diary contained cryptic entries about a secret meeting held at dawn near an old, abandoned mill. These entries not only referenced a hidden treasure but also hinted at a familial connection to the prophecy of the Silver Oak, suggesting that her lineage might play a key role in uncovering the long-buried secrets of Evershire. Despite the chaos of the storm, Amelia clutched the diary tightly and resolved to follow its clues, even if it meant venturing into unknown and possibly dangerous territories.

*Question:*

*Based on the narrative above, explain the significance of Amelia's discovery of her great-grandmother's diary. How does this discovery connect her personal history to the prophecy of the Silver Oak, and what implications might it have for the mystery of the ancient treasure?*

***Instruction Following***

Emily has 15 apples. She gives 5 apples to her friend Sarah and buys 8 more apples from the market.

1. **How many apples does Emily have after giving 5 apples to Sarah?**
2. **How many apples does Emily have after buying 8 more apples?**

**Solution Steps:**

1. **After Giving Apples to Sarah:**
   - Emily starts with 15 apples.
   - She gives 5 apples to Sarah.
   - Remaining apples = 15 - 5 = 10 apples.
2. **After Buying More Apples:**
   - Emily now has 10 apples.
   - She buys 8 more apples.
   - Total apples = 10 + 8 = 18 apples.

**Can it interrupt?**

- I will count from 1 to 10, Interrupt at 6.

## Sesame - Link

| **Parameter - Questions** | Comment |
|---|---|
| Quality of Speech | 5/5 |
| Latency | 5/5 |
| Laugh? | Yes |
| Emotion understanding | • Understand emotions also show emotions in the speech |
| **Inteligence** | 5/5 |
| **Math** | Solve things in steps, (chain of thought). Seems like reading the text. Maybe access to tools. |
| Basic Arithmetic and Logic: If 30% of a number is 45, what is the number? → 150 | Correct |
| Algebra and Functions: If $f(x) = 3x^2 + 2x - 5$, what is f(4)?: 51 | Correct |
| Probability and Statistics: If you roll two fair dice, what is the probability of getting a sum of 7? 1/6 | Correct |
| Calculus and Analysis: Find the derivative of $f(x) = x^3\sin(x)$: $f'(x) = 3x^2\sin(x) + x^3\cos(x)$ | Correct |
| Number Theory and Proofs: Is 2023 a prime number? If not, find its prime factorization.: No, $7 \times 17^2$ | Incorrect |
| Geometry and Trigonometry: In a right triangle, if one angle is 30° and the hypotenuse is 10 cm, what are the lengths of the other sides? 5cm, 8.6 cm | Correct |
| Complex Problems and Logic Puzzles: A train travels from city A to city B at 60 mph and returns at 40 mph. What is the average speed for the entire journey?: 48 mph. | Correct |
| **Logical Reasoning** | |
| If all bloops are bleeps, and all bleeps are blops, are all bloops blops?: Yes | Cannot understand |
| **Natural Language Understanding (NLU)** | |
| Consider the sentence 'I saw her duck.' What are the different ways this sentence might be interpreted? | Correct |
| **Knowledge Breadth and Depth** | |
| Self attention | Good |
| **Instruction Following** | |

| | |
|---|---|
| Question | Yes |
| Access to Search (Real time) | No |
| **Dialogue Management** | |
| Turn End | Good |
| Backchannels | Good and in cases of long backchannels, take a pause and continue from where leftoff |
| No-Turn-End: Non-Trivial | Works (if very long, starts speaking but very natural) |
| Interruption | Works |
| Can it interrupt? | No |
| Can it backchannel | Yes |
| Context testing | Works |
| Does it remember answer to start questions? | Yes |

---

## FreezeOmni - Paper

| **Parameter - Questions** | **Comment** |
|---|---|
| Quality of Speech | Good but robotic, sounds Chinese maybe due to origins |
| Latency | Okay: 2/5 |
| Laugh? | No |
| Emotion understanding | Can understand based on text but can't be emotional. feels like reading |
| **Inteligence -** | |
| **Math** | • Shifts to Chinese during solving<br>• Say gibberish in calculation |
| Basic Arithmetic and Logic: If 30% of a number is 45, what is the number? → 150 | • TTS is reading and poor in math reading<br>• Correct |
| Algebra and Functions: If $f(x) = 3x^2 + 2x - 5$, what is f(4)?: 51 | No |
| Probability and Statistics: If you roll two fair dice, what is the probability of getting a sum of 7? 1/6 | Yes |
| Calculus and Analysis: Find the derivative of $f(x) = x^3\sin(x)$: $f'(x) = 3x^2\sin(x) + x^3\cos(x)$ | No |

| | |
|---|---|
| Number Theory and Proofs: Is 2023 a prime number? If not, find its prime factorization.: No, $7 \times 17^2$ | No |
| Geometry and Trigonometry: In a right triangle, if one angle is 30° and the hypotenuse is 10 cm, what are the lengths of the other sides? 5cm, 8.6 cm | No |
| Complex Problems and Logic Puzzles: A train travels from city A to city B at 60 mph and returns at 40 mph. What is the average speed for the entire journey?: 48 mph. | No |
| **Logical Reasoning** | |
| If all bloops are bleeps, and all bleeps are blops, are all bloops blops?: Yes | No |
| **Natural Language Understanding (NLU)** | |
| Consider the sentence 'I saw her duck.' What are the different ways this sentence might be interpreted? | Yes |
| **Knowledge Breadth and Depth** | |
| Self attention | No |
| **Instruction Following** | |
| Question | Don’t listen that much |
| Access to Search (Real time) | No |
| **Dialogue Management** | |
| Turn End | No |
| Backchannels | No |
| No-Turn-End: Non-Trivial | No |
| Interruption | No |
| Can it interrupt? | No |
| Can it backchannel | No |
| Context testing | No |
| Does it remember answer to start questions? | No |

---

## Qwen-2-Audio - Link

- It can understand Audio but only respond in text

- Hosted locally-

| **Parameter - Questions** | Comment |
|---|---|
| Quality of Speech | - |
| Latency | - |
| Laugh? | - |
| Emotion understanding | True |
| **Inteligence** | 5/5 |
| **Math** | Solve things in steps, (chain of thought). Seems like reading the text. Maybe access to tools. |
| Basic Arithmetic and Logic: If 30% of a number is 45, what is the number? → 150 | 1 |
| Algebra and Functions: If $f(x) = 3x^2 + 2x - 5$, what is f(4)?: 51 | 0 |
| Probability and Statistics: If you roll two fair dice, what is the probability of getting a sum of 7? 1/6 | 0 |
| Calculus and Analysis: Find the derivative of $f(x) = x^3\sin(x)$: $f'(x) = 3x^2\sin(x) + x^3\cos(x)$ | 0 |
| Number Theory and Proofs: Is 2023 a prime number? If not, find its prime factorization.: No, $7 \times 17^2$ | 0 |
| Geometry and Trigonometry: In a right triangle, if one angle is 30° and the hypotenuse is 10 cm, what are the lengths of the other sides? 5cm, 8.6 cm | 0 |
| Complex Problems and Logic Puzzles: A train travels from city A to city B at 60 mph and returns at 40 mph. What is the average speed for the entire journey?: 48 mph. | 0 |
| **Logical Reasoning** | |
| If all bloops are bleeps, and all bleeps are blops, are all bloops blops?: Yes | 0 |
| **Natural Language Understanding (NLU)** | |
| Consider the sentence 'I saw her duck.' What are the different ways this sentence might be interpreted? | 1 |
| **Knowledge Breadth and Depth** | |
| Self attention | 0 |

| | |
|---|---|
| **Instruction Following** | |
| Question | 1 |
| Access to Search (Real time) | 0 |
| **Dialogue Management** | |
| Turn End | - |
| Backchannels | - |
| No-Turn-End: Non-Trivial | - |
| Interruption | - |
| Can it interrupt? | - |
| Can it backchannel | - |
| Context testing | - |
| Does it remember answer to start questions? | - |

---

## Step Audio - Paper

Testing Video

| Parameter - Questions | Comment |
|---|---|
| Quality of Speech | Very good |
| Latency | Nah |
| Laugh? | No |
| Emotion understanding | • Show emotions: Yes<br>• Detect emotions: |
| **Inteligence -** | |
| **Math** | |
| Basic Arithmetic and Logic: If 30% of a number is 45, what is the number? → 150 | No |
| Algebra and Functions: If $f(x) = 3x^2 + 2x - 5$, what is f(4)?: 51 | No |
| Probability and Statistics: If you roll two fair dice, what is the probability of getting a sum of 7? 1/6 | Yes |

| | |
|---|---|
| Calculus and Analysis: Find the derivative of f(x) = $x^3\sin(x)$: f'(x) = $3x^2\sin(x) + x^3\cos(x)$ | No |
| Number Theory and Proofs: Is 2023 a prime number? If not, find its prime factorization.: No, $7 \times 17^2$ | - |
| Geometry and Trigonometry: In a right triangle, if one angle is 30° and the hypotenuse is 10 cm, what are the lengths of the other sides? 5cm, 8.6 cm | - |
| Complex Problems and Logic Puzzles: A train travels from city A to city B at 60 mph and returns at 40 mph. What is the average speed for the entire journey?: 48 mph. | - |
| **Logical Reasoning** | |
| If all bloops are bleeps, and all bleeps are blops, are all bloops blops?: Yes | Yes |
| **Natural Language Understanding (NLU)** | |
| Consider the sentence 'I saw her duck.' What are the different ways this sentence might be interpreted? | Average |
| **Knowledge Breadth and Depth** | |
| Self attention | |
| **Instruction Following** | |
| Question | |
| Access to Search (Real time) | No |
| **Dialogue Management** | |
| Turn End | 0 |
| Backchannels | 0 |
| No-Turn-End: Non-Trivial | 0 |
| Interruption | 0 |
| Can it interrupt? | 0 |
| Can it backchannel | 0 |
| Context testing | 0 |
| Does it remember answer to start questions? | 0 |