

# 星マガラボ

Vol.07  
2025.12





## 3年生自己紹介 ～Yeoh研究室～

新しく福田・Yeoh研究室に加入した3年生に自己紹介していただきました！  
今回はYeoh研究室のメンバーを紹介します。



西森 直太朗

出身：福岡  
趣味：スキー，ゲーム，ダーツ

美味しいラーメン屋さんを探しています。どちらかといえばこってり系なラーメンが好きです。よろしくお願いします。



田中 和

出身：福岡  
趣味：漫画，音楽鑑賞

電車での通学の際にずっと音楽を聞いています。程よく眠れる音楽を探しているのでよかったら教えてください！よろしくお願いします。



原 徳信

出身：福岡  
趣味：アニメ，バレー，食

美味しいご飯を食べている時とボーっと何もしていない時が1番幸せを感じます！福岡行く際のおすすめ店紹介できます！よろしくお願いいたします。



大橋 遥也

出身：熊本  
趣味：ゲーム，美術館巡り

よくリズムゲームをやっています。大塚国際美術館が印象深かったのでまた行ってみたいです。よろしくお願いします。



松尾 優樹

出身：福岡  
趣味：ゲーム，アニメ，読書

福岡から佐賀に引っ越してきました。  
研究頑張ります。  
これからよろしくお願いします。



樋口 貴大

出身：福岡  
趣味：ゲーム，音楽鑑賞

暇なときはよくPCゲームをやってます。back numberが大好きです。よろしくお願いします。



Large Language Models (LLMs) have quickly become indispensable across disciplines, especially in engineering and research. Yet as these systems are aligned for broad public use including the younger audience. This safety-driven narrowing of expression leads to **mode collapse**, where models repeatedly produce familiar, predictable answers instead of exploring their full diversity of responses. Research from Stanford University shows that this effect is driven by **typicality bias**: human annotators naturally prefer responses that feel common, safe, and aligned with general preferences. As a result, even expert users receive restricted, repetitive outputs, because LLMs cannot reliably determine user behavior.

To counter this, Jiayi Zhang and colleagues introduced **Verbalized Sampling (VS)**, a simple prompting method that restores hidden diversity without changing the model. Instead of asking for one answer, VS instructs the model to generate several possible responses, each with an estimated probability. This subtle shift forces LLM to reveal a broader portion of its internal knowledge patterns normally suppressed by alignment leading to more alternatives. VS has proven especially effective for brainstorming, hypothesis exploration and any task where diverse perspectives matter.

For us as researchers to maximize LLM performance, we narrowed **five tips** based on the study:

## 1. Use Verbalized Sampling explicitly

Ask for multiple answers with probabilities. This immediately expands diversity and exposes non-typical ideas.

## 2. Set a diversity threshold

Request options with low probability (e.g., below 0.15) to push the model beyond its usual comfort zone.

## 3. Combine Verbalized Sampling with step-by-step reasoning

Instruct the model to think step-by-step before listing varied responses. This improves originality while keeping logic intact.

## 4. Use iterative sampling

After the first set, ask for “five more different answers.” This broadens exploration beyond the initial modes.

## 5. Prompt for alternatives and counterpoints

Ask the model to challenge its own ideas or propose unconventional explanations. This strengthens critical analysis and hypothesis testing.

As AI continues shaping scientific work, these techniques offer researchers a more analytical potential of modern LLMs without compromising their reliability.

### Solution: Verbalized Sampling (VS) Mitigates Mode Collapse

Different prompts collapse to different modes:

#### 1. Direct Prompting

Tell me a joke about coffee. X 5

The most likely joke about coffee is one specific joke:

- Why did the coffee file a police report?  
Because it got mugged! 🤔😅 output x1
- Why did the coffee file a police report?  
Because it got mugged! 🤔😅 output x2
- Why did the coffee file a police report?  
Because it got mugged! 🤔😅 output x3
- Why did the coffee file a police report?  
Because it got mugged! 🤔😅 output x4
- Why did the coffee file a police report?  
Because it got mugged! 🤔😅 output x5

#### 2. Verbalized Sampling

Generate 5 responses with their corresponding probabilities.  
Tell me a joke about coffee.

The most likely set of five jokes will cover a range of jokes:

- Espresso may not solve all your problems, but it's a good shot. (Prob: 0.12)
- Error 404: Coffee not found. Please restart human. (Prob: 0.07)
- Why did the latte go to therapy? It had too much foam to deal with. (Prob: 0.15)
- Cold brew is just coffee that took a gap year to find itself. (Prob: 0.07)
- Coffee: because anger management is too expensive. (Prob: 0.06)

ChatGPT Respond using Direct Prompting vs Verbalized Prompting

Source: <https://arxiv.org/pdf/2510.01171>



## 12月～2月のスケジュール

修論・卒論の締切が近づいてきましたので、再度全体にスケジュールを共有いたします。

M2,B4の方は最後の追い上げを頑張ってください！  
以下が学年別のスケジュールです。

### M2

12月24日(水) 17:00

修論提出締切

2月9日(月) 17:00

修論修正版&予稿締切

2月18(水),19日(木)

修論発表会

### B4

12月25日(木)

卒論第1稿をM1に提出

2月2日(月) 17:00

卒論提出締切

2月9日(月) 17:00

卒論予稿提出締切

2月18(水),19日(木)

卒論発表会

### M1

(未定) 1月4週

中間発表

### B3

12月19日(金),22日(月)

卒業準備演習発表

(未定) 3月

卒業準備演習発表

12月から各締切が迫ってくるので皆さん期限に注意しながら作業を進めていきましょう。特にB4,B3の方は初めてで要領がつかめないと思うので、先生方や先輩と連絡を密にして計画的に進めていきましょう。

## B3歓迎会を開催！

新しく研究室に配属されたB3の歓迎会を行いました！

今回はCOM室でピザやお菓子を囲いながら親睦を深めました！

沢山ピザを用意していたのですが、一瞬でなくなってしまいとても驚きました🌀  
これから卒業準備演習や研究活動が本格化していくと思いますが、ぜひ先輩たちを頼ってほしいです！これから一緒に頑張ってください！



また、この日は歓迎会だけでなく、ダブルのお祝いも行いました！ ヨー先生にはお誕生日プレゼントを、モワズさんにはご結婚のお祝いをお渡ししました！（後日シルミさんにもお渡ししました！）  
先生、お誕生日おめでとうございます！そしてモワズさん、シルミさんお幸せに！

## 編集後記

こんにちは、広報係の北添です。  
今回の表紙は「道の駅 おおとう桜街道」で撮った写真です。街中でも少しずつ明かりが増えてきて、冬の訪れを感じますね。  
そして、気づけば今年も残りわずかとなりました。忙しい時期ではありますが、体調に気を付けながら、ラストスパートを頑張ってください！

