Data-driven projects

A few thoughts

東京万博
~Tokyo Burn Punk~
2024.10.4(金)

BASEMENTBAR & THREE

開場17:30/開演18:00
¥3,500(D別)

はじめちゃん × Simmer Pine pre.

最高な日!!

Date — 20240927
Venue — BASEMENTBAR&THREE Shimokitazawa
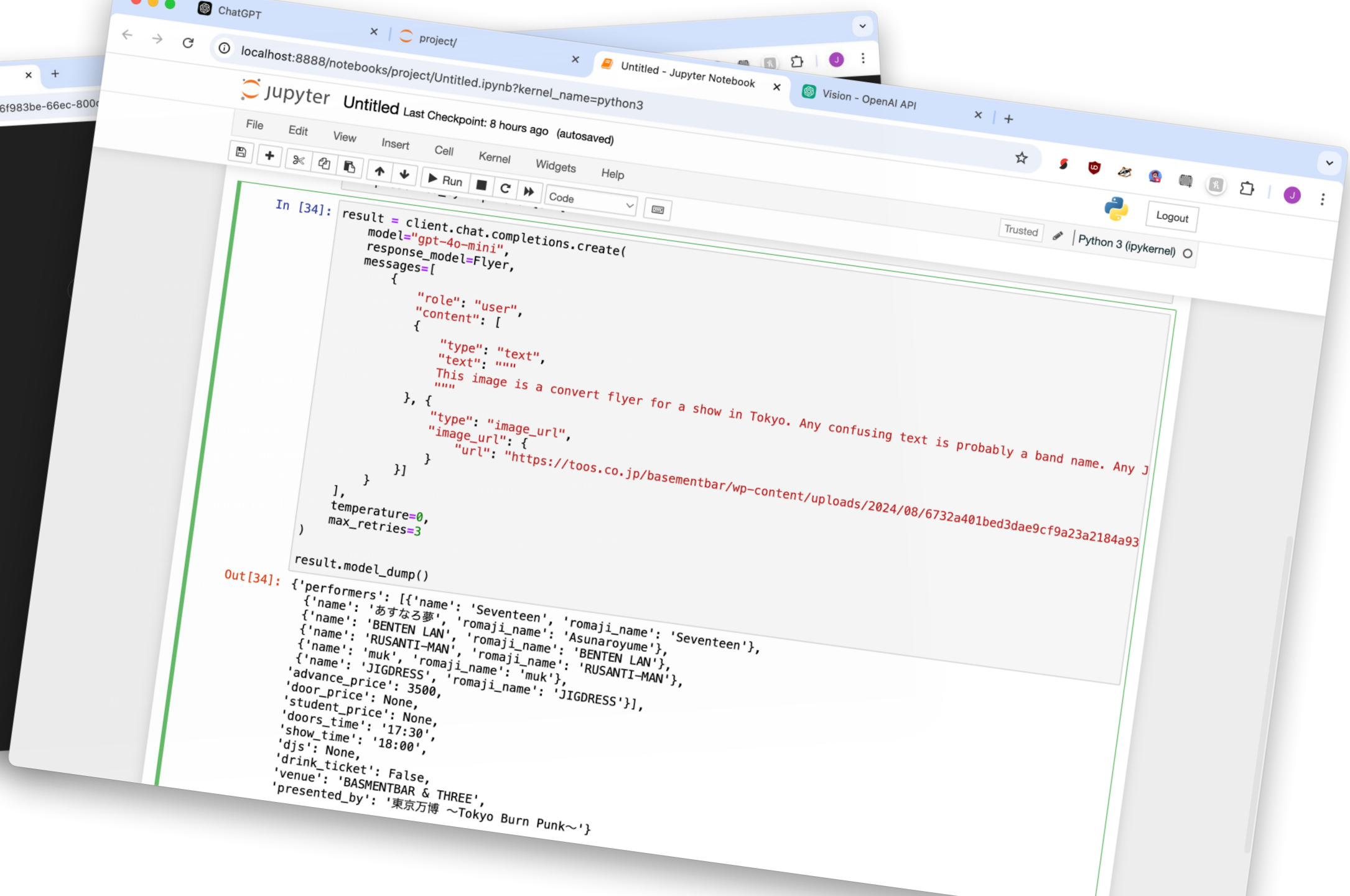Price — ¥3,000 (1drink include)
Open&Start — 6:30 pm

LIVE
Simmer Pine
Newdums
Salmon Pink
三つ峠

DJ
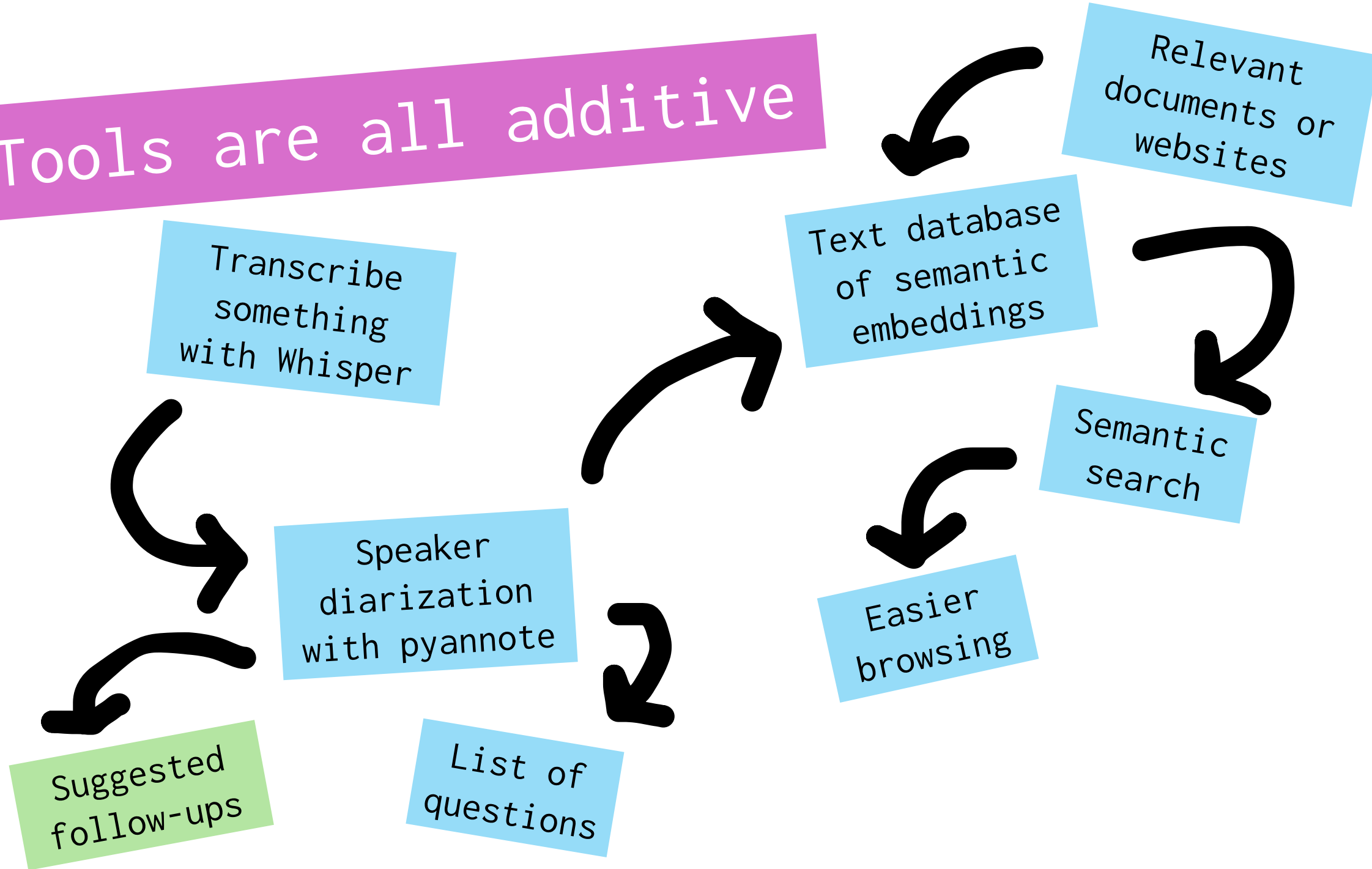GaMoN
Koshin
B.i.C
JORDAN

FOOD
こちゅ飯

ChatGPT
project/
localhost:8888/notebooks/project/Untitled.ipynb?kernel_name=python3
Untitled - Jupyter Notebook
Vision - OpenAI API
バンド名の特定
chatgpt.com/c/66f983be-66ec-800
ChatGPT 4o

jupyter **Untitled** Last Checkpoint: 8 hours ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help

Code

Trusted | Python 3 (ipykernel) O

Logout

```python
In [34]: result = client.chat.completions.create(
             model="gpt-4o-mini",
             response_model=Flyer,
             messages=[
                 {
                     "role": "user",
                     "content": [
                         {
                             "type": "text",
                             "text": """
                             This image is a convert flyer for a show in Tokyo. Any confusing text is probably a band name. Any J
                             """
                         }, {
                             "type": "image_url",
                             "image_url": {
                                 "url": "https://toos.co.jp/basementbar/wp-content/uploads/2024/08/6732a401bed3dae9cf9a23a2184a93
                             }
                         }]
                 }
             ],
             temperature=0,
             max_retries=3
         )

         result.model_dump()
```

```
Out[34]: {'performers': [{'name': 'Seventeen', 'romaji_name': 'Seventeen'},
           {'name': 'あすなろ夢', 'romaji_name': 'Asunaroyume'},
           {'name': 'BENTEN LAN', 'romaji_name': 'BENTEN LAN'},
           {'name': 'RUSANTI-MAN', 'romaji_name': 'RUSANTI-MAN'},
           {'name': 'muk', 'romaji_name': 'muk'},
           {'name': 'JIGDRESS', 'romaji_name': 'JIGDRESS'}],
          'advance_price': 3500,
          'door_price': None,
          'student_price': None,
          'doors_time': '17:30',
          'show_time': '18:00',
          'djs': None,
          'drink_ticket': False,
          'venue': 'BASMENTBAR & THREE',
          'presented_by': '東京万博 〜Tokyo Burn Punk〜'}
```

Projects

- Managing data projects and coding projects is **the worst thing**
- **Estimates**: all wrong
- **Data**: all misunderstood
- I hope you discovered that while you were working on this project!

Ideas

- You have data. What can it be **joined with?**
  - Good target for automation: joining with data that frequently changes
  - "What columns would make this job more effective?"
- Things that need to be done repeatedly
- Things that are done by new hires
- Sell something to others that benefits you, they'll pay you to improve it

Tools are all additive

Relevant documents or websites

Text database of semantic embeddings

Semantic search

Transcribe something with Whisper

Speaker diarization with pyannote

Easier browsing

Suggested follow-ups

List of questions

Cost

- Lots of venture capital = **artificially cheap**
- Price is not related to ability
- You will need ongoing support and adjustments
  - Don't ignore this!!!!
- Using a platform is better for big, quick changes
  - Journalism example: using an online Whisper tool vs. one built into in-house CMS infrastructure

Expertise

- You **(maybe)** have the skills to prototype
- Hire **single-time-fee consultant** for guidance
  - They have experience, most important early on
  - They know what is possible vs impossible
  - Best if not someone who can develop for you
  - Ask "what do people use AI for that they should not?"
- Don't trust what any company says :)

Estimating

- Estimating how long it takes is **very hard**
- Estimating what you'll need+want is **very hard**
  - You'll get 80% of the way there in 20% of the time
  - Edge cases are the problem!
  - AI is good in "easy" cases but sometimes the hard cases are almost impossible
- Never trust internal or external estimates.
  - Think days, weeks or month. Not a specific number.

Be open to change

- Have a product that can be used at 80% ability
- How can it be used with people instead of fully automated?
- Data is often **misunderstood**, and you realize flaws later on. Can it still be used?
- Beautiful visualization can hide bad analysis

Evaluations

- Create **number-based evaluations** early on
- Don't just measure "accuracy"
  - What happens when it goes wrong?
- If it's human-in-the-loop, quality can be lower
- Test with **people who are suspicious of AI**
- Be sure to measure and test over time (weekly, monthly), **models change**

tokyogigguide.com/en/

tokyo 東京
gig guide ギグガイド

⌂ GIGS ⌄ LIVE HOUSES ⌄ TICKETS LOG IN/JOIN ? 🔍 ⌄ 🌐 EN/日本語 ⌄

UPCOMING GIGS

Takahashi Yasuyuki (tb) + Ito Keita (b) ...
2024 / 09 / 30 (Mon), 19.00 @ Knuttel House Iriya 🏷 Avant garde, Free Jazz, Instrumental, Jazz

Spacious Recollection (鈴木直人, 石川広行, 山本裕之)
2024 / 09 / 30 (Mon), 19.15 @ velvetsun Ogikubo 🏷 Jazz

中尾憲太郎 [ex. NUMBER GIRL] (bass), 類家心平 (tr...
2024 / 09 / 30 (Mon), 19.30 @ POLARIS tokyo Ogawamachi 🏷 All Mix

Kevin Sommer, Silke Strahl, Misaki Moto...
2024 / 09 / 30 (Mon), 19.30 @ Bar Isshee Sendagi 🏷 Improvised

Terrificus (from Australia), VOQ, 広本晋, 陰...
2024 / 10 / 01 (Tue), 18.30 @ Oriental Force Koenji 🏷 Avant garde, World, Improvised, Experimental, Alternative

The True Three Per: Whoacho, Semba Kiyoh...
2024 / 10 / 01 (Tue), 18.30 @ Jirokichi Koenji 🏷 Experimental, Funk, Improvised

more...

Data-driven projects

A few ideas