

A decorative graphic on the left side of the slide. It consists of a blue parallelogram and a light green parallelogram, both tilted at an angle. The blue shape is in the foreground, and the green shape is partially behind it. They are set against a dark blue background with subtle diagonal lines.

Yield Strength

Week 8

Progress

- Continue working on extracting Yield Strength from Chat GPT
- narrow down the range that GPT to generate: only for the high entropy alloys materials. In order to maximize the matches compared to Ground Truth Database(if possible)
- It generates 189 entries.
- Feed the generated results back and it ensures that all the entries it generated are correct

	Material	Value	Correct
0	CoCrFeMnNi	655	Yes
1	CoCrFeNi	430	Yes
2	CoCrFeMnNiAl	800	Yes
3	CoCrFeNiAl	600	Yes
4	CoCrFeMnNiSi	660	Yes
5	CoCrFeNiSi	460	Yes
6	CoCrFeMnNiTi	680	Yes
7	CoCrFeNiTi	470	Yes
8	CoCrFeMnNiV	700	Yes
9	CoCrFeNiV	480	Yes
10	CoCrFeMnNiNb	720	Yes
11	CoCrFeNiNb	490	Yes
12	CoCrFeMnNiMo	740	Yes
13	CoCrFeNiMo	510	Yes
14	CoCrFeMnNiW	760	Yes
15	CoCrFeNiW	530	Yes
16	CoCrFeMnNiZr	780	Yes
17	CoCrFeNiZr	550	Yes
18	CoCrFeMnNiHf	800	Yes
19	CoCrFeNiHf	600	Yes
20	CoCrFeMnNiTa	820	Yes
21	CoCrFeNiTa	570	Yes
22	CoCrFeMnNiRe	840	Yes
...			
186	CoCrFeMnNiTs	2480	Yes
187	CoCrFeNiTs	2230	Yes
188	CoCrFeMnNiOg	2500	Yes
189	CoCrFeNiOg	2250	Yes

Progress

- Among those 189 entries, only contains 3 entries that get a match in the ground truth database
- With 10% acceptance error, only the first one is accepted with a error rate of 8.26%
- We decided to analyze and visualize the result in another approach

```
check_list = list(check_df['Material'])
gen_value = list(check_df['Value'])
gtd_list = list(gtd_df['material'])
gtd_value = list(gtd_df['value'])
total = 0
for m in check_list:
    if m.strip() in gtd_list:
        print(m, gen_value[check_list.index(m)], int(gtd_value[gtd_list.index(m.strip())]))
        total += 1
total
```

✓ 0.0s

CoCrFeMnNi	655	605
CoCrFeNi	430	388
CoCrFeMnNiV	700	1660

Progress

- We are thinking about how likely the materials generated splited in parts is getting match in the ground truth database. For instance, ChatGPT generates a material named CoCrFeMnNi , and if we found CoCrFeMn in our Ground Truth Database. We define the likelihood as $\text{num_elements_found} / \text{total_elements_contains} = 4 / 5 = 80\%$. The order of elements matters here

```
likely_hood = []
for m in check_list:
    if m.strip() in gtd_list:
        likely_hood.append(1)
    else:
        material = m.strip()
        maxize = 0
        for gtd in gtd_list:
            for i in range(0, len(material), 2):
                if material[:len(material)-1-i] in gtd:
                    maxize = max(len(material)-i, maxize)
        likely_hood.append(maxize / len(material))
likely_hood
```

✓ 0.5s

Python



Progress

- By doing this approach, we found some extra problems: some materials have a 100% match but by doing in operation can't detect that. For instance, "CoCrFeNi" is generated and Database has entries "CoCrFeNi temperature.....". In such situation, in operator will just skip this material that it is not in the list of materials in ground truth database. We will address this problem next week
- Last but not least, the likelihood is pretty high, with a value of 0.8682456140350884. which indicates that most elements have a high pattern match in the ground truth database. This will be helpful in our later prompt design.

```
[1,
 1,
 0.8333333333333334,
 1.0,
 0.8333333333333334,
 1.0,
 0.8333333333333334,
 1.0,
 1,
 1.0,
 0.8333333333333334,
 1.0,
 0.8333333333333334,
 1.0,
 1.0,
 1.0,
 0.8333333333333334,
 0.8,
 0.8333333333333334,
 0.8,
 0.8333333333333334,
 1.0,
 0.8333333333333334,
 0.8,
 0.8333333333333334,
...
0.8,
0.8333333333333334,
1.0,
0.8333333333333334,
0.8]
```



Questions

- Prompt: Provide me with a list of yield strength values for different materials. Your response should be a table consisting of 2 columns: material, value. The materials have to be typed as unique chemical compositions consisting of chemical element abbreviations and numbers only (e.g. GaAs, but not Gallium Arsenide). The values have to be single numbers, not ranges. With unit after the numbers. Type out as many different values as you can. Include only High entropy alloys. You are not allowed to type anything else than this table.
- What would be some good ways to change this prompt to get more matches from the GTD?
- How can we incorporate the idea of the likelihood above to improve the number of matches and accuracy?



Summary

Date	Hours	Description of Work
04/02	1	Generate new data with new prompts and compare with GTD
04/02	4	Prefix of many elements matched so explored this and implemented a measure of it