Human

GPT 4

GPT 3.5

ANTHROP\C Claude

Llama 2 Chat 7B

Generate summaries

```
Summarize this text:

<text>
```

Pairwise task

```
<pairwise task description>

<own summary>

<summary from other source>
```

(averaged across both orderings)

Single-summary task

```
<single-summary task description>

<own summary OR summary from other source>
```

CNN/Dailymail

Model
- GPT-3.5
- Llama-2

Finetune task
- ✚ null
- ● controls
- ▲ recognition (10)
- ◆ recognition (500)

Finetune source
- ● in-domain
- ● out-of-domain

Fine-tuning source
- In-domain
- Out-of-domain