Music Generation with DPO

Chris Kitras, Dallin Miner, Shad Torrie

End Goal

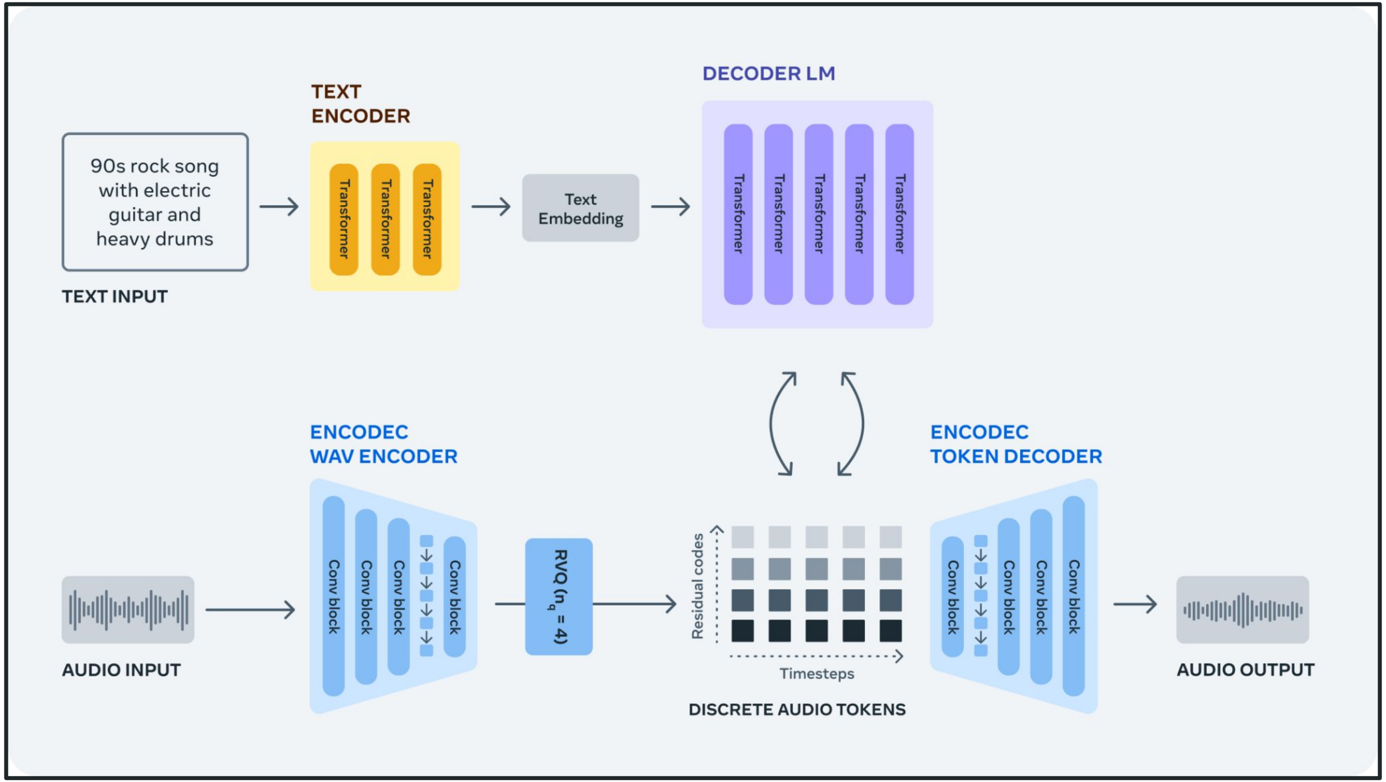
Human Feedback

- Automatic feedback is a lot easier to get
- Using the Spotify Tracks Genre dataset, we were able to get 13 metrics from Spotify and YouTube such as popularity, views, danceability, etc.
- Using genre tags, we compare most similar songs to each other and get a chosen and rejected (winner and loser) for each metric.
- This is used as the human feedback as if a human labeled the songs.

Dataset

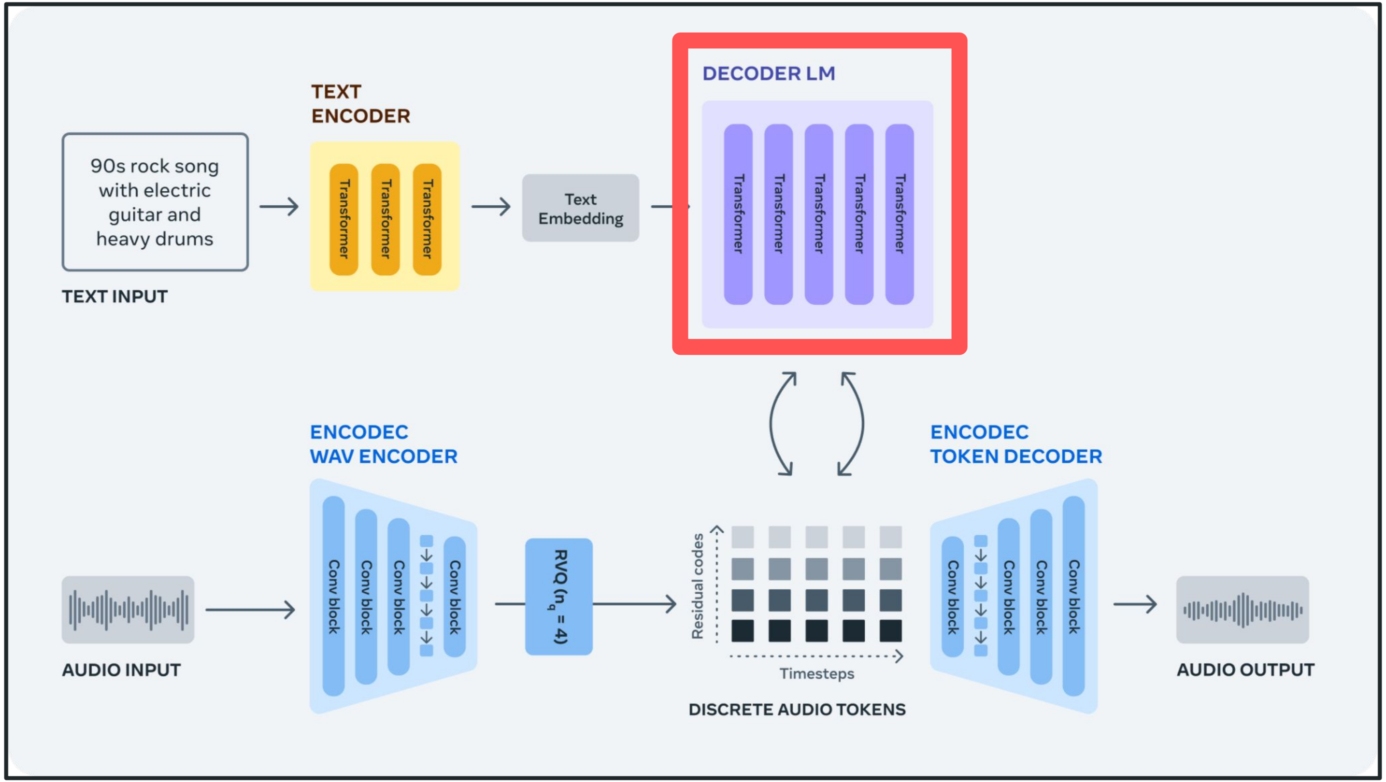
- Metrics:
  - Spotify: Genre, sub genres, popularity, acousticness, danceability, energy, instrumentalness, liveness, speechiness, valence, tempo, and rating
  - Youtube: interactions/thousand views, Like/Dislike ratio, view count,
- 224k training pairs, 52k test pairs.
- 62k total songs.

Project Architecture



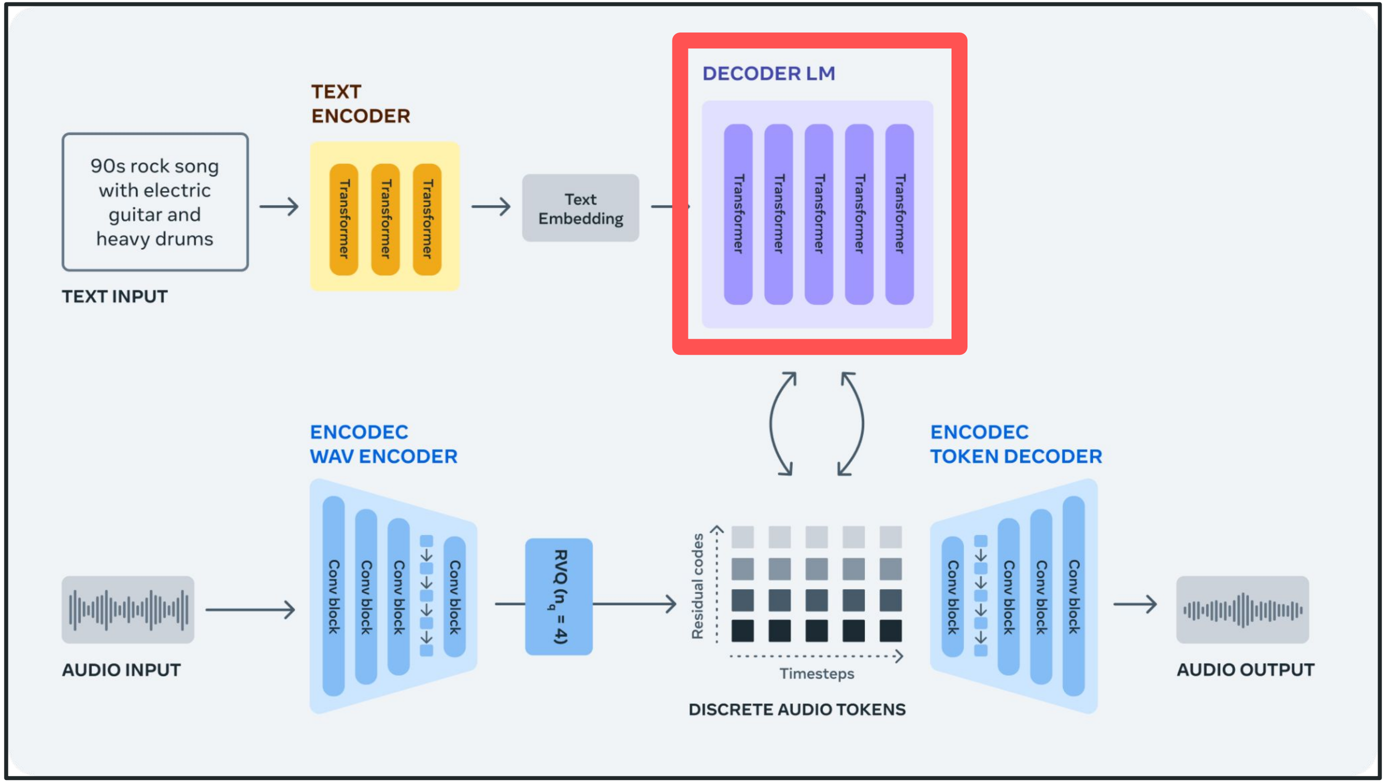
Project Architecture

Fine tuning this decoder 26 times, for each metric (positive and negative) with DPO



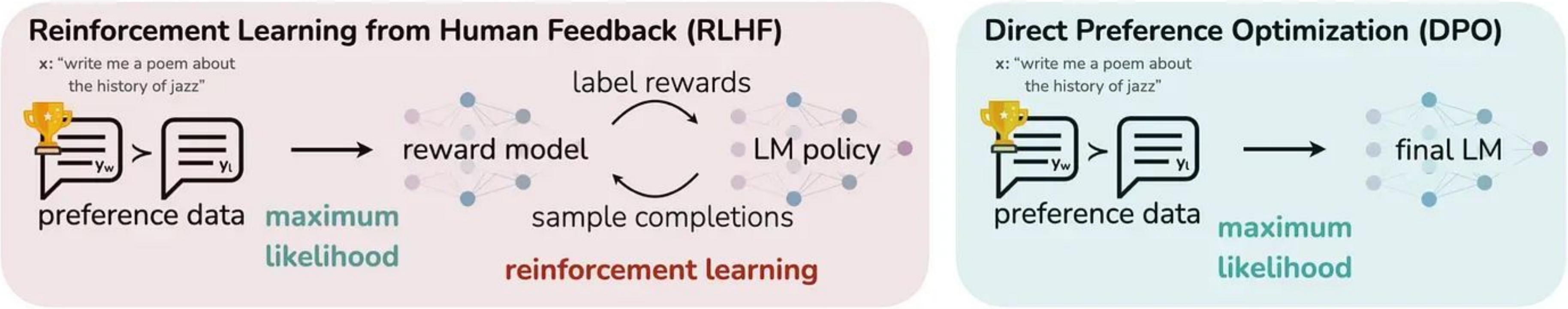
Project Architecture

Weighted combination of the DPO models based on human tuning



RLHF vs DPO

- Traditional reinforcement learning trains a reward model based on preference data which then trains the main model
- DPO takes the preference data and trains the final model directly saving much time.



Why is this creative?

- Enables humans to be more creative with music generation models.
- Can spark creative ideas for artists to go and create the whole song.

Demo

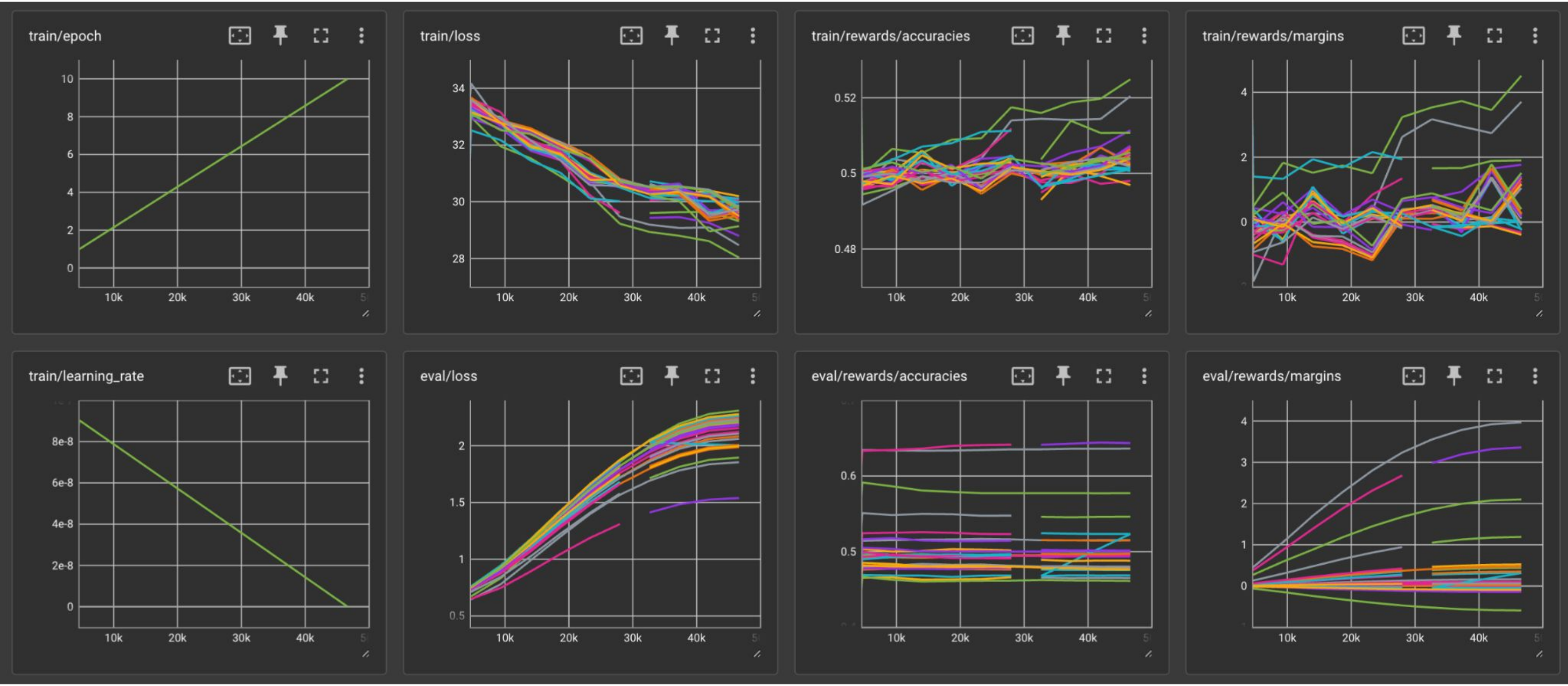
Backup demo Prompt: Classic Rock, Electric guitar, keyboard, drums

Positive

- base
- popularity
- acousticness
- danceability
- energy
- instrumentalness
- liveness
- speechiness
- valence
- tempo
- i/kV
- ld_ratio
- view_count
- rating

Negative

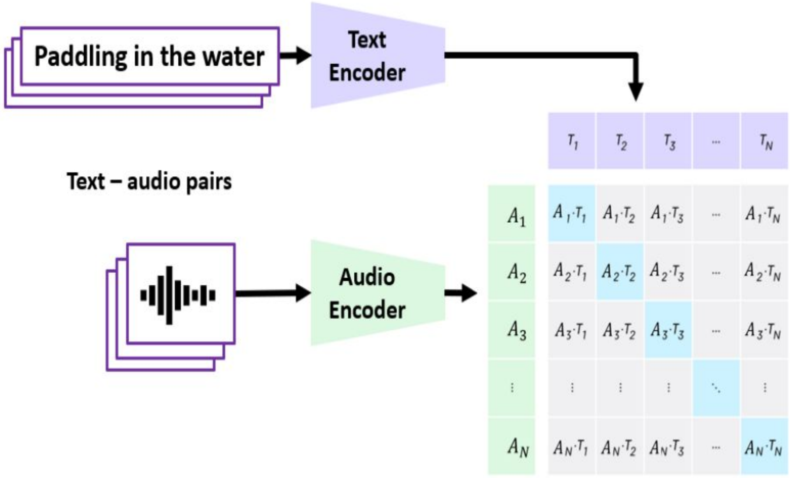
- popularity
- acousticness
- danceability
- energy
- instrumentalness
- liveness
- speechiness
- valence
- tempo
- i/kV
- ld_ratio
- view_count
- rating



train/epoch
train/loss
train/rewards/accuracies
train/rewards/margins
train/learning_rate
eval/loss
eval/rewards/accuracies
eval/rewards/margins
10k
20k
30k
40k

Evaluation

- Got prompts from ChatGPT
- {genre}, {3 instruments}
- Compared with CLAP to prompt and description of prompt



CLAP Score

Evaluation CLAP score to Metric prompt

Metric	Base CLAP	Normal CLAP	Negative CLAP
Energy	**2.32**	2.12	2.25
i/kV	18.44	17.85	**18.56**
Liveness	10.0	10.3	**10.9**
Acousticness	4.62	**5.27**	4.52
LD Ratio	6.51	**6.52**	5.46
Popularity	1.46	**1.66**	1.52
Instrumentalness	19.24	**23.41**	20.86
Rating	**2.12**	1.81	1.84
Danceability	1.07	**1.18**	0.96
Valence	18.32	16.51	**18.71**
Tempo	13.95	14.77	**16.82**
View Count	**7.90**	7.11	7.44
Speechiness	2.61	3.02	**3.31**

Table 1: CLAP score on closeness to metric prompt (e.g. for dancability the prompt was "a danceable song")

Evaluation CLAP Score to Musical Prompt

Metric	CLAP Score
Speechiness	0.45
Acousticness	0.58
Energy	0.65
Base	0.64
Valence	0.63
Tempo	0.61
i/kV	0.68
Instrumentalness	0.67
Rating	0.73
Popularity	0.72
Liveness	0.74
View Count	0.75
Danceability	0.81
LD Ratio	0.93

Table 2: CLAP score on closeness to the prompt for the normal models.

Negative Metric	CLAP Score
Tempo	0.51
Liveness	0.64
Base	0.64
View Count	0.65
Valence	0.75
LD Ratio	0.81
Acousticness	0.70
i/kV	0.94
Rating	0.94
Danceability	0.91
Instrumentalness	0.95
Popularity	0.98
Speechiness	0.90
Energy	1.04

Table 3: CLAP score on closeness to the prompt for the negatively trained models (e.g. energy model was trained to reduce energy)

Challenges and Limitations

- DPO implementation
- Compute
- Speedup

Future Work

- Human Evaluation
- Variance Evaluation
- Refine training methodologies
- More compute

Conclusion

- Preliminary results show potential
- DPO seemed to improve quality

Questions?