The references in this model card refer to the associated paper:

Anticipatory Music Transformer.
John Thickstun, David Hall, Chris Donahue, and Percy Liang.
Preprint Report, 2023.

Table 11: Model Card [111] - Anticipatory Music Transformer.

Model Details	
Organization Developing the Model	Stanford Center for Research on Foundation Models
Model Date	June 2023
Model Type	Autoregressive Causal Transformer
Additional Modeling Details	See Section 3
License	BigScience Open RAIL-M License
Correspondence	jthickstun@cs.stanford.edu
Intended Use	
Primary Intended Uses	Collaborative co-composition between a human composer and an Anticipatory Music Transformer. The role of the anticipatory model in this collaboration could include, e.g., infilling tedious/low-entropy details (productivity enhancement) and suggesting possible continuations (creative ideation).
Primary Intended Users	Artists, musicians, and composers.
Out-of-Scope Uses	**Long-Context Generation.** These models cannot generate full-length song structures without human control. The models have a context length of 1024 tokens (331 events). At 68 tokens/second (the average for Lakh MIDI; see Appendix D) this corresponds to approximately 15 seconds of context. Models conditioned on more than 331 events will only use the most recent 331 events (including anticipated events) to predict the next event. **Music Metadata.** These models do not explicitly model or generate metadata, including: metrical structure, key signature, tempo, note-value (eighth-note, quarter-note etc.). **Extended Music Vocabulary.** These models generate sequences with a narrow vocabulary of notes, instruments, and timings. They do not model or generate other aspects of music, including: dynamics, articulations, or lyrics.
Factors	
Western Bias	These models are trained on the Lakh MIDI dataset, a collection of predominantly Western music. See Section 7.3 for further discussion.
Metrics	
Automatic Metrics	Next-event perplexity (defined in Table 1) and bits per seconds (defined in Appendix E).
Human Evaluation	Pairwise human preferences between generated music and reference compositions.
Decision Thresholds	For human evaluation, we generated samples from anticipatory models using nucleus sampling with $p = 0.95$. See Section 4 for further discussion.
Approaches to uncertainty and variability	We report p-values for pairwise comparisons between music generated by different models and ground truth music using the Wilcoxon signed-rank test. Due to computational constraints, we do not account for variability in the model training process, such as dataset splits or the random seed for optimization.

Datasets	
Training Data	The `0`–`d` splits of the Lakh MIDI dataset, augmented using anticipation (see Section 3) with the prior distribution over controls described in Appendix C.
Validation Data	The `e` split of the Lakh MIDI dataset.
Test data	The `f` split of the Lakh MIDI dataset.
Out-of-Distribution Data	We do not evaluate out-of-distribution performance.
Preprocessing	Preprocessing and filtering of the Lakh MIDI dataset is described in Appendix D.
Motivation	We chose to work with the Lakh MIDI dataset because it is the largest collection of symbolic music data currently in use by the machine learning community.
Quantitative Analyses	
Aggregated Analysis	Our analysis of aggregate results based on automatic metrics and human evaluation are presented in Section 4.1 and Section 4.2 respectively. Key findings include: • Anticipatory training does not interfere with autoregressive model performance, as measured by perplexities of comparable anticipatory and autoregressive models. • Accompaniments generated by an Anticipatory Music Transformer have similar musicality to ground truth accompaniments according to human evaluators.
Disaggregated Analysis	We do not perform a disaggregated analysis of the Anticipatory Music Transformer. One obstruction to conducting such an analysis is a lack of metadata associated with the Lakh MIDI dataset.
Ethical Considerations	
Labor Displacement	We are broadly concerned by the transient disruptions of labor markets caused by the introduction of new productivity-enhancing and automative technologies. See Section 7.1 for a discussion of the possible disruptive effects of generative music models on the creative economy.
Copyright	The Lakh MIDI dataset contains large quantities of copyrighted music. The copyright status of models trained on this data—and music sampled from these models—is an open legal question. See Section 7.2 for further discussion.