Propose a model monitoring pipeline and describe how you would track model drift in 500 words.

Audio Speech Recognition (ASR) is used for many functions, but at the core of it, its function is to convert speech to text, i.e. transcription of audio. As with all Machine Learning models, model drift may occur, hence a model monitoring pipeline would be required to ensure the ASR model is able to perform as required. This model monitoring pipeline would cover multiple aspects, including data collection, feature drift detection, computation of evaluation metrics, and retraining

The initial stage of the monitoring pipeline would cover data collection. For ASR models, this involves not only recording the transcription pairs but also capturing acoustic features inherent to audio signals, such as pitch, speed, duration, volume and signal to noise ratio. This information would be stored in a centralized database to allow for a systematic and scalable approach to track the evolution of the ASR model's input characteristics over time. Metadata related to timestamps, sources, and contextual information are also logged, providing a wide overview of the data.

Given that ASR models heavily rely on acoustic features, there is a need to monitor the statistical properties of the audio inputs. Feature drift, characterized by shifts in the distribution of these acoustic characteristics, can significantly impact the model's performance. Techniques like cepstral mean and variance normalization are applied to maintain the consistency of feature distributions. Monitoring changes in these features ensures that the model's representations of audio remain consistent over time. If audio features are observed to have a changing trend over time, that would mean a drift in the data. Detection of outliers is also important to ensure the data does not affect the model training. Hence, regularly comparing the current acoustic feature distributions with historical data enables the early detection of drift, ensuring timely intervention.

To assess the performance of the ASR model, it is essential to define and compute relevant evaluation metrics. Text-based metrics such as Word Error Rate (WER) and Character Error Rate (CER) offer a good understanding of the model's transcription accuracy as they are direct comparisons between ground truth text and the transcribed output text. Regularly computing these metrics over time allows for the tracking of the model's performance and identification of potential drift. This step ensures that the ASR system aligns with the desired accuracy levels, meeting the required level of performance.

The fourth crucial component of the model monitoring pipeline is the implementation of a feedback loop for retraining. When drift is detected, it's important to initiate automatic retraining processes. Leveraging the most recent audio data, the ASR model is updated to adapt to the evolving patterns within the input distribution. This dynamic approach ensures that the model remains robust and responsive to changes, preventing long-term performance degradation. The feedback loop also involves documenting the corrective actions taken during retraining, contributing to a knowledge base for future model improvement.

In conclusion, a well-structured model monitoring pipeline for Audio Speech Recognition covers comprehensive data collection, feature drift detection, continuous computation of evaluation metrics, and an adaptive retraining mechanism. By addressing these four critical categories, the pipeline ensures the resilience and sustained accuracy of ASR models, enabling them to deliver reliable transcription services.