Batchalign Web App Functionalities

The web application built using Django and Batchalign as the backend will provide an intuitive interface for processing audio files, generating transcripts, and managing outputs.

---

1. Audio File Upload and Management

- File Upload
  - Support for multiple audio file formats (e.g., MP3, WAV, MP4, M4A, MOV).
  - Drag-and-drop upload functionality.

- File Validation and Conversion
  - Validate that the uploaded file is a valid audio file.
  - Convert unsupported formats (e.g., M4A, MOV) into supported formats (e.g., WAV or MP3) using a backend conversion tool (e.g., FFmpeg).
  - Provide error messages for invalid or corrupted files.

- Audio File Management
  - List uploaded files with metadata (e.g., file name, size, upload date).
  - Allow users to delete or replace files.

- File Structures for Storing Outputs
  - Organize audio files, generated transcripts, diarization results, and other metadata in a structured directory system.
  - Maintain a clear mapping between audio files, processed transcripts, and associated data.

---

2. Batchalign Processing

- Transcription

  - Run Batchalign to generate transcripts from uploaded audio files.

  - Support for language selection (e.g., English, Spanish).

- Alignment and Diarization

  - Perform forced alignment to time-align audio with words or sentences.

  - Use additional APIs (e.g., Pyannote.ai) to enhance speaker diarization and confidence levels.

  - Store diarization results alongside transcripts for easy review and editing.

- Custom Configurations

  - Allow advanced users to tweak Batchalign parameters (e.g., language model or processing speed).

---

3. Transcript Review and Editing

- Transcript Display

  - Show transcripts in CHAT format, which is specialized for communication studies.

  - Visually highlight speaker diarization results and timestamped sections.

  - Indicate areas where the ASR model has low confidence or failed to detect audio, particularly for child speech or sounds.

- Editing Features

- Allow users to manually edit transcripts inline.

- Enable users to assign custom speaker labels (e.g., SPEAKER1 can be renamed to "CHILD" or "MOTHER").

- Provide quick editing options for inserting tags like [CHILD SOUND] or other annotations.

- Audio Integration

- Play audio alongside the transcript, with synchronized highlighting for the current sentence or word.

- Allow reviewers to jump to specific timestamps by clicking on the transcript.

---

4. Force Alignment for Existing Transcripts

- Handling Pre-Existing Transcripts

- Allow users to upload manually typed transcripts in CHAT format that lack timecodes.

- Use Batchalign's forced alignment to match the transcript with the audio file.

- Integrate Pyannote.ai's diarization results to fill gaps where ASR loses track (e.g., for child sounds).

- Improving Forced Alignment

- Enhance alignment by preprocessing transcripts and audio to account for missing child sounds.

- Highlight mismatched or unaligned sections for user review and manual adjustment.

---

5. Ease of Use and Deployment

- User-Friendly Interface

- Simplified design tailored for users in psychology and related fields, who may have limited technical skills.
  - Clear instructions and tooltips for all features.

- Simplified Deployment
  - Provide a one-click installation script or Docker container for easy setup.
  - Include clear configuration guides for server setup, dependencies, and basic usage.

- Pre-Configured Options
  - Ship the app with sensible defaults for Batchalign and diarization settings to minimize initial configuration effort.

---

6. Analytics and Insights
- Processing Metrics
  - Show word count, audio duration, and language distribution for processed files.
  - Provide logs for failed processing jobs with error details.

---

7. Additional Features
- Search and Filter
  - Search transcripts by keywords or file names.
  - Filter by date range, language, or processing status.