Development and Evaluation of an AI-powered Video Captioning and Transcription Web App

Thema:
: Development and Evaluation of an AI-powered Video Captioning and Transcription Web App

Art:
: BA

BetreuerIn:
: Thomas Schmidt

BearbeiterIn:
: Thi Ha My Pham

ErstgutachterIn:
: Christian Wolff

Status:
: in Bearbeitung

Stichworte:
: web development, video. transcription, aws, user interface design, subtitles, content creation, social media

angelegt:
: 2024-01-08

Antrittsvortrag:
: 2024-05-06

Hintergrund

With the growth of video-hosting-platforms like Youtube, Tiktok, and many other online courses, videos have become a predominant medium for communication, entertainment, and education. Therefore, making video content accessible to all users, including people with communication impairment, who may require captions, sign language interpretation, or other accommodations to access the content, is crucial. There are many transcription tools provide a solution to this challenge by transforming spoken content into text. However, they still have some limitations, especially when it comes to smoothly adding captions to video files permanently. The AWS Transcribe is chosen for the web app's transcription and captioning features, as it is a robust and well-established automatic speech recognition (ASR) service provided by Amazon Web Services. It is known for its high accuracy in transcribing spoken words into text and its transcription accuracy can also be improved using custom vocabularies and custom language models. Besides, AWS Transcribe supports over 100 languages, making it versatile for users with diverse language requirements.

Zielsetzung der Arbeit

- Develop an AI-powered Video Captioning and Transcription Web App to automate the process.
- Improve accessibility to video content for users with hearing impairments.
- Increase the efficiency and accuracy of video transcription.

- Create a user-friendly interface for easy navigation and interaction.
- Evaluate the effectiveness of the AI model in real-world scenarios.

Konkrete Aufgaben

Front-end development

- Create an intuitive and responsive user interface using HTML, CSS, and JavaScript, with a focus on accessibility and usability.
- Ensure a smooth and responsive experience across various devices.
- Integrate video playback functionality.
- Create a form for uploading video files.

AWS integration

- Set up AWS Transcribe and integrate it into the web app.
- Establish secure communication channels with AWS services.

Real-time Captioning

- Implement real-time captioning features using AWS Transcribe.

Functional and user testing

- Ensure all features work as intended.
- Validate the accuracy of the transcription and captioning.
- Gather feedback on the user interface and overall experience.

Erwartete Vorkenntnisse

- Web development in HTML, CSS and JavaScript
- Familiarity with integrating AWS services into web applications
- Usability Testing

Weiterführende Quellen

Mahoney, K. (2023, Juni 28). The Current State of Captioning: A Report by 3Play Media. 3Play Media. https://www.3playmedia.com/blog/the-current-state-of-captioning-a-report-by-3play-media/

Sheth, A. (2023, November 29). Speech Recognition: AWS Transcription Platform Embraces Generative AI. Prompts Daily. https://www.neatprompts.com/p/revolutionizing-speech-recognition-aws-transcription-platform-embraces-generative-ai

Guida, L. (2022, Juni 10). Use AWS AI and ML services to foster accessibility and inclusion of people with a visual or communication impairment | AWS Machine Learning Blog. https://aws.amazon.com/blogs/machine-learning/use-aws-ai-and-ml-services-to-foster-accessibility-and-inclusion-of-people-with-a-visual-or-communication-impairment/

Rajamani, S., & Penmatcha, R. (2021, März 10). Translate video captions and subtitles using Amazon Translate | AWS Machine Learning Blog. Trans-late Video Captions and Subtitles Using Amazon Translate. https://aws.amazon.com/blogs/machine-learning/translate-video-captions-and-subtitles-using-amazon-translate/

Guttikonda, S., & Saxman, P. (2023, Oktober 16). Generative AI in education: Build-ing AI solutions using course lecture content | AWS Public Sector Blog. https://aws.amazon.com/blogs/publicsector/generative-ai-education-building-ai-solutions-using-course-lecture-content/

Krishna, R., Hata, K., Ren, F., Fei-Fei, L., & Niebles, J. C. (2017). Dense-Captioning Events in Videos. 2017 IEEE International Conference on Computer Vision (ICCV), 706–715. https://doi.org/10.1109/ICCV.2017.83

Lin, K., Li, L., Lin, C.-C., Ahmed, F., Gan, Z., Liu, Z., Lu, Y., & Wang, L. (2022). SwinBERT: End-to-End Transformers with Sparse Attention for Video Captioning (arXiv:2111.13196). arXiv. http://arxiv.org/abs/2111.13196

Liu, Z., Ning, J., Cao, Y., Wei, Y., Zhang, Z., Lin, S., & Hu, H. (2022). Video Swin Transformer. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3192–3201. https://doi.org/10.1109/CVPR52688.2022.00320

Tang, M., Wang, Z., Liu, Z., Rao, F., Li, D., & Li, X. (2021). CLIP4Caption: CLIP for Video Caption. Proceedings of the 29th ACM International Conference on Multimedia, 4858–4862. https://doi.org/10.1145/3474085.3479207

Yang, B., Zhang, T., & Zou, Y. (2022). CLIP Meets Video Captioning: Concept-Aware Representation Learning Does Matter (arXiv:2111.15162). arXiv. http://arxiv.org/abs/2111.15162