Project Proposal for master's in information systems

This project Proposal entitled "AudioBookify" by:

1. Yuraja Kadri
2. Yash Jivani
3. Shrutika Gajbhiye
4. John Melo
5. Dipen Patel
6. Vicky Jadhav

Date: 02/25/2024

CONTENTS

Abstract

Title: "AudioBookify: Bridging Text to Sound for the Visually Impaired"

The Audiobookify project addresses the pressing need for universal access to written content in the digital age. According to WHO (World Health Organization) With approximately 253 million people globally experiencing vision impairment, there's a growing demand for innovative solutions to facilitate access to literature and educational materials. Our project aims to bridge this accessibility gap by developing Audiobookify, an advanced application that converts images and PDFs into audiobooks.

While initially conceived to cater to the needs of blind and visually impaired individuals, Audiobookify also serves a broader audience, including students and individuals who prefer listening to reading. Through seamless integration of Optical Character Recognition (OCR) technology, Audiobookify transforms text from images or PDFs into audio format, enabling users to engage with written content conveniently and efficiently.

The Audiobookify application offers a user-friendly interface and intuitive functionality, allowing users to access their favorite books, notes, or documents anytime, anywhere. Whether commuting, exercising, or relaxing at home, users can immerse themselves in their chosen reading material through auditory means, enhancing both productivity and leisure time.

By democratizing access to written content, Audiobookify promotes inclusivity and equal opportunities for knowledge acquisition. It empowers individuals with visual impairments to engage with literature and educational resources independently, while also catering to the diverse reading preferences of a wider audience.

In summary, the Audiobookify project represents a significant step towards creating a more accessible and inclusive digital reading experience. Through its innovative features and user-centric design, Audiobookify sets a new standard for universal access to written content in the digital era.

Introduction

In today's digitally driven world, written content serves as a fundamental pillar for learning, communication, and entertainment. Yet, for individuals with visual impairments or those who favor auditory learning, accessing written material poses daunting obstacles. To bridge this accessibility gap, our project, Audiobookify, endeavors to craft a comprehensive solution that seamlessly converts images and PDFs into audio books. By catering to the needs of visually impaired individuals and those who prefer listening over reading, Audiobookify aims to revolutionize access to written content.

Powered by advanced Optical Character Recognition (OCR) technology, Audiobookify conducts intricate analyses of images and PDFs, discerning and extracting text, symbols, tables, diagrams, and coding language content. This sophisticated OCR process facilitates the transformation of diverse forms of written material into accessible text format. Subsequently, the extracted text undergoes synthesis into audio format, thereby fashioning an audiobook readily accessible to users.

The significance of Audiobookify transcends mere convenience; it empowers visually impaired individuals to navigate written content autonomously, enriching their educational and recreational pursuits. Furthermore, by accommodating individuals who favor auditory learning—such as commuters or those with busy lifestyles—Audiobookify fosters inclusivity and serves as a valuable tool for lifelong learning.

Through the integration of cutting-edge technology and a user-centric approach, Audiobookify endeavors to redefine the interaction between individuals and written material. By rendering literature, educational resources, and information more accessible through audio format, Audiobookify seeks to elevate the quality of life for visually impaired individuals while offering a convenient alternative for all who relish the auditory experience over traditional reading.

Preliminary Literature Review

Introduction/Background Information:

Text to speech synthesis was first attempted in 1779 by the Russian professor Christian Kratzenstein. Kratzenstein documented and explained the difference between the sounds of the five long vowels (A/E/I/O/U). To mimic these sounds, he created a machine that mirrored the structure of human vocal system. This machine used vibrating reeds to mimic the sounds of these vowels. It is important to note that the I vowel was not produced with a vibrating reed. It was produced by blowing air into the machine like a flute. Text to speech synthesis was later attempted in Vienna in 1791 by the "Acoustic-Mechanical Speech Machine" invented by Wolfgang von Kempelen. Like Kratzenstein's machine, Von Kemplen's machine mimicked the human vocal system and was able to produce individual sounds and a few sound combinations (Lemmetty, 1999).

Text-to-speech technology (TTS) took a significant leap in 1968 with the development of the first full text-to-speech English system by Noriko Umeda in the Electrotechnical Laboratory in Japan. This system used an articulatory model to articulate sounds along with a syntactic analysis module. This allowed the system to generate intelligible English speech. However, the speech that was generated did not include prosody and was monotonous (Lemmetty, 1999). Finally, in 1939 the first speech synthesizer known as VODER (Voice Operating Demonstrator) was introduced by Homer Dudley. The VODER synthesizer worked by analyzing speech into acoustic parameters which would synthesize a speech signal. The signal would then pass through ten filters where the output was manipulated by fingers. Although the output of the VODER synthesizer was mediocre, this device set the foundation for modern-day TTS technologies (Lemmetty, 1999).

Currently, many institutions like universities have limited resources for visually impaired students. To digest information, visually impaired individuals rely on braille or on verbal translation from their peers. To help visually impaired individuals digest information more efficiently, a text-to-audio converter system will need to be designed. To design this system, we will need to identify factors such as the appropriate API, the type of texts (books, articles, PDFs) the system will convert, and design the database for storing the audible material.

Body of the Review/Discussion of Sources:

Audiobooks have been proven to be just as effective as traditional reading methods in helping people retain information. The University of Oregon conducted an experiment that showed comparable competencies between listening to and reading materials. The experiment showed that the test subjects remembered and forgot the same portions of the material regardless of the medium. The reason for this is because the part of the brain that is responsible for language comprehension works the same regardless of whether the individual is listening to or reading the material (Best, 2020). Since audiobooks are comparable to traditional reading for information retention, audiobooks are especially beneficial to individuals with vision impairments.

Advancements in text-to-speech technology (TTS) has single-handedly led to the recent success of audiobooks. Recent advancements in TTS deep learning have improved the ability to recreate natural – sounding audio that mimics human speech. For example, it was recorded that the United States experienced annual audiobook sales revenue of about $1.3 billion in the year 2020 alone (Pethe et al., 2023). Additionally, by using advanced text-to-speech technology, the National Library Service for the blind has expanded their audiobook library by recording tens of thousands of books. Despite these recent TTS advancements, TTS systems still have drawbacks when it comes to prosody and intonation variation. For example, machine – generated audio stays within a small constant range of pitch and volume, making the speech monotone. In order to develop a more natural prosody mechanism, the system must contain text-character/prosody analysis prediction capabilities. Text-character/prosody analysis will determine how textcharacters are read and the pitch at which they are read.

Most text-to-speech synthesizers contain certain modules that perform separate important functions. One module is called the Natural Language Processor (NLP). The NLP creates a phonetic transcription of the chosen text with prosody. The second module is called the Digital Signal Processor (DSP). The DSP module converts symbolic text from the NLP module into an intelligible audio medium (Isewon et al., 2014). In addition, NLP has two main functions. The first main function is text analysis. Text analysis works by segmenting text into tokens. The second main function is implementation of pronunciation rules. Once text analysis is successfully completed, pronunciation rules can be implemented. Finally, the product of the NLP module is sent to the Digital Signal Processor (DSP) module for processing (Isewon et al., 2014).

Advancements in TTS technology have made audiobooks a staple resource for people with visual impairments. Experiments have shown that there is no difference between information retention rates from reading text versus listening to text. Having an adequate TTS system prevents visually impaired individuals from being at a disadvantage. To develop an efficient TTS system, we will need to design our natural language processor module along with the digital processor module.

Project Gains and Contributions

- **Improved Accessibility for Visually Impaired Individuals:**
  The development of this software application will significantly enhance accessibility for visually impaired individuals by providing them with a reliable and efficient tool for converting PDF documents into audiobooks. This enables them to access a wide range of written content across various domains, including education, literature, and professional documentation, thereby promoting inclusivity and equal access to information.

- **Enhanced Educational and Professional Opportunities:**
  Access to educational materials, scholarly articles, and professional documentation in audiobook format enhances educational and professional opportunities for visually impaired individuals, enabling them to pursue higher education, advance their careers, and engage in lifelong learning. This facilitates greater academic and economic inclusion, with potential long-term benefits for individuals and society.

- **Empowerment Through Technology:**
  By bridging the gap between written text and auditory content, the software empowers visually impaired individuals to independently access and engage with digital content in a format that suits their needs and preferences. This empowerment fosters greater autonomy, confidence, and participation in academic, professional, and recreational activities, ultimately enhancing their quality of life.

- **Advancement of Assistive Technology:**
  The project contributes to the advancement of assistive technology by integrating state-of-the-art Optical Character Recognition (OCR) and Text-to-Speech (TTS) technologies into a user-friendly software application tailored for visually impaired users. This innovation not only improves the accessibility of digital content but also showcases the potential of technology to address real-world challenges and improve the lives of individuals with disabilities.

- **Publishing Industry:**
  The publishing industry can leverage the software to expand the accessibility of their publications to a wider audience, including visually impaired readers. By enabling publishers to convert their PDF-based books and documents into audiobooks, the project contributes to making literary and educational content more inclusive and accessible, fostering greater participation and engagement in getting knowledge.

- **Content Creators and Authors:**
  Content creators and authors can reach a broader audience by making their written works accessible in audiobook format. The project empowers content creators and authors to ensure that their works are accessible to visually impaired readers, promoting diversity and inclusivity in literature and digital content consumption.

"So, What" test and Justification

- **Accessibility for All:**
  So, what if we develop software to convert PDF documents into audiobooks for visually impaired individuals? Well, by doing so, we're ensuring that individuals with visual impairments have equal access to information and educational resources, empowering them to engage with digital content on an equal footing with sighted individuals.

- **Inclusivity in Education:**
  So, what if educational institutions adopt this software? It means that visually impaired students can fully participate in classroom activities, access textbooks, and educational materials in audiobook format, and pursue academic success without facing the barriers posed by traditional written content.

- **Broadening Literary Access:**
  So, what if publishers utilize this technology? It expands the reach of literature and educational content to a wider audience, including those who prefer auditory learning or have difficulty reading printed text, thereby fostering a more inclusive literary landscape, and promoting lifelong learning.

- **Empowering Content Creators:**
  So, what if content creators make their works accessible in audiobook format? It means that authors, educators, and creators can ensure that their content reaches diverse audiences, including visually impaired readers, contributing to a more inclusive and equitable society where everyone can participate in cultural and educational discourse.

- **Advancing Assistive Technology:**
  So, what if components of this project are integrated into assistive technology solutions? It accelerates the development of innovative accessibility tools and devices, improving the quality of life for individuals with disabilities and driving forward the evolution of assistive technology to better meet the needs of diverse user groups.

Projected Research Methodologies and Techniques

We will devised a multi-phase approach for our research project, aimed at converting text from PDFs into audio books. This methodical procedure encompasses several steps, each leveraging various Python libraries. Specifically, we plan to utilize PyPDF2 for PDF parsing, Pillow for image processing, Pytesseract for OCR (image to text) functionality, and potentially pyttsx3 for text-to-speech synthesis.

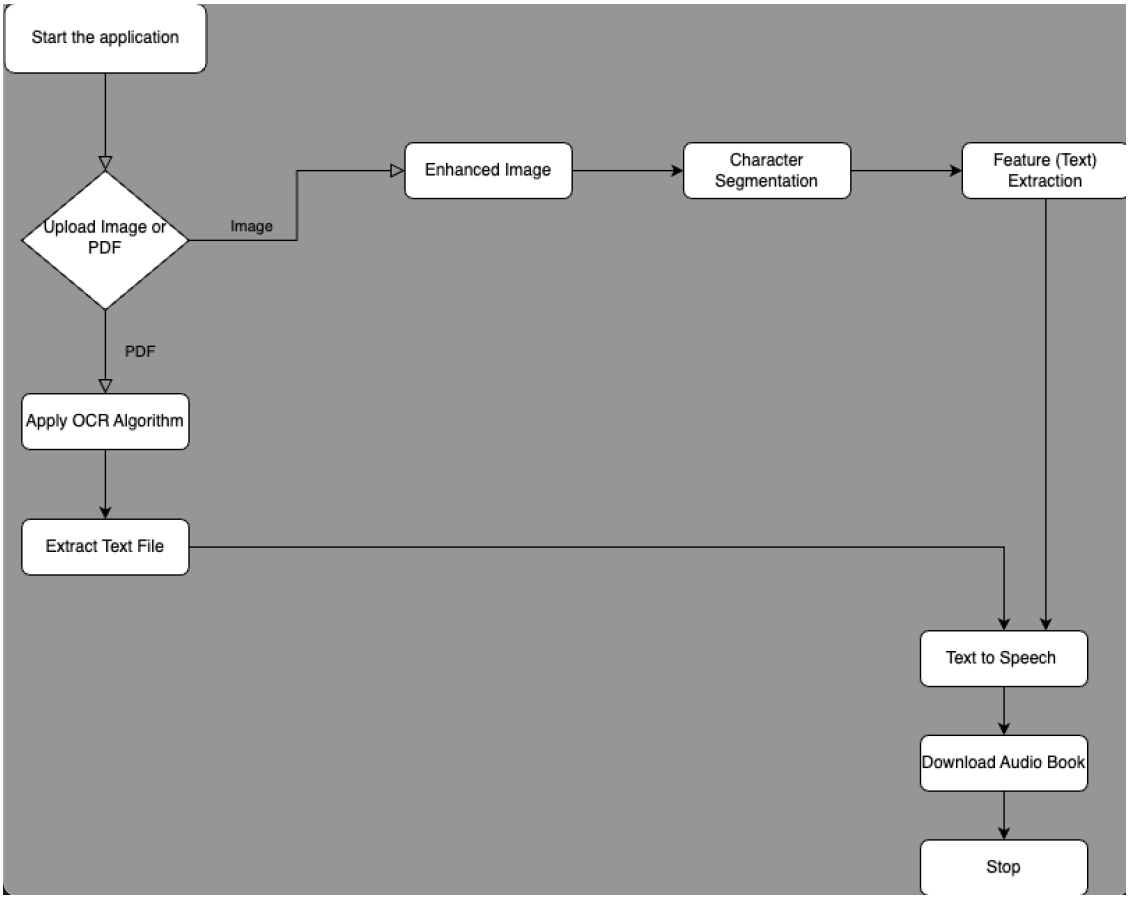
The project commences with the selection of an input file, which can either be a PDF or an image. Based on this file format, appropriate preprocessing and OCR text extraction techniques will be applied. A pivotal aspect of our approach is text recognition, which extends beyond conventional text to include recognition of handwritten text, tables, diagrams, and programming language code. These additional features are intended to be seamlessly integrated into our system.

Following text extraction, text-to-speech synthesis will be employed to convert the extracted text into audio format. The final output will be stored as an MP3 audio book file. The workflow of the project is designed to ensure user-friendly access to audio books, thereby promoting inclusivity and diversity by enhancing accessibility to various types of textual content. The flowchart for the given project involves several key steps:

Input Selection: The process begins with selecting an input file, which can be either a PDF or an image containing text.
File Format Check: Next, the system checks the format of the input file to determine whether it is a PDF or an image.
Preprocessing: If the input file is an image, preprocessing techniques such as converting it to grayscale or enhancing its quality may be applied.
Text Extraction: Using Optical Character Recognition (OCR) technology, text is extracted from the input file, including normal text, handwritten text, tables, diagrams, and programming code.

Text-to-Speech Conversion: The extracted text is then converted into audio format using text-to-speech synthesis.
Output Generation: Finally, the synthesized audio is saved as an MP3 audio book file, ready for user access and consumption.

Overall, the flowchart depicts a streamlined process of converting text from PDFs or images into audio books, enhancing accessibility and diversity in accessing textual content. The given below is the flow chart Diagram:



-Diagram-01 Flow Chart

To improve the accuracy of our OCR models under diverse conditions, we will incorporated a range of specialized training methods and datasets. This dedication to accuracy underscores our commitment to meeting the needs of individuals with vision impairments while also providing a valuable resource for learning.

Project Planning

We have chosen the Agile software development methodology for its iterative and incremental approach, allowing for flexibility and adaptability throughout the project lifecycle. Agile plan is given below:

Communication and Planning (Week 1-2):
Set up communication channels. Define project goals and objectives.
Conduct initial sprint planning.

Sprint 1 (Week 3-4):
Requirement gathering and analysis.
Design initial architecture and user interface.

Sprint 2 (Week 5-6):
Development of core functionalities (PDF parsing, image preprocessing).
Begin integration of OCR functionality.

Sprint 3 (Week 7-8):
Complete OCR integration.
Start development of text-to-speech synthesis.

Sprint 4 (Week 9-10):
Finish text-to-speech synthesis.
Begin user interface development.

Sprint 5 (Week 11-12):
Conduct testing (unit, integration, usability).
Address any bugs or issues identified.

Sprint 6 (Week 13-14):
Deployment to test environment.
Gather user feedback.

Sprint 7 (Week 15-16):
Incorporate user feedback into updates.
Finalize deployment.

Tasks & Schedule

The Project planning diagram is given below:

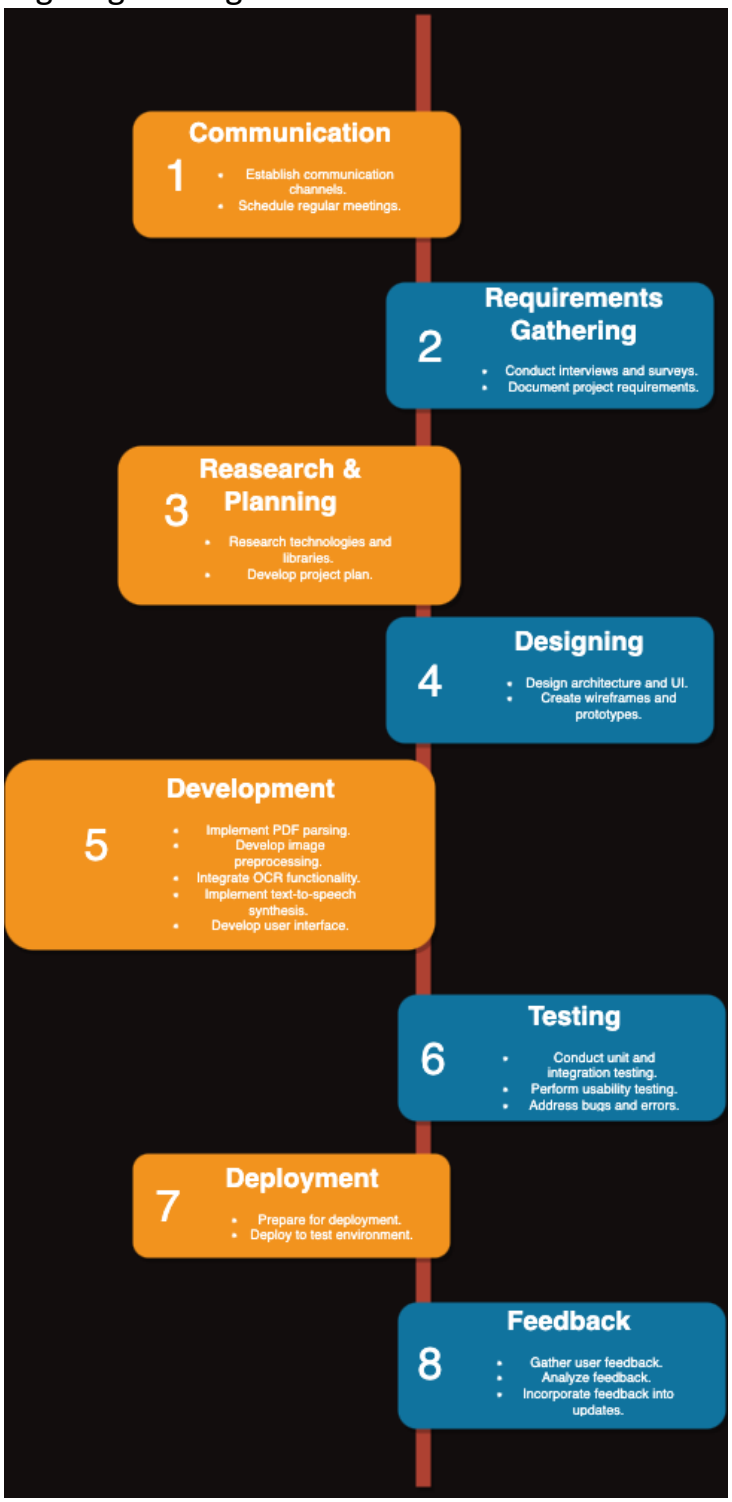


Diagram-02 Project plan

Creating a task and schedule is essential for project management as it provides a roadmap for the entire project lifecycle. It helps in organizing and prioritizing tasks, allocating resources effectively, and ensuring that the project stays on track to meet its objectives and deadlines. A well-defined schedule also helps in managing dependencies between tasks, identifying potential bottlenecks or risks, and adjusting as needed to mitigate issues and keep the project on schedule.

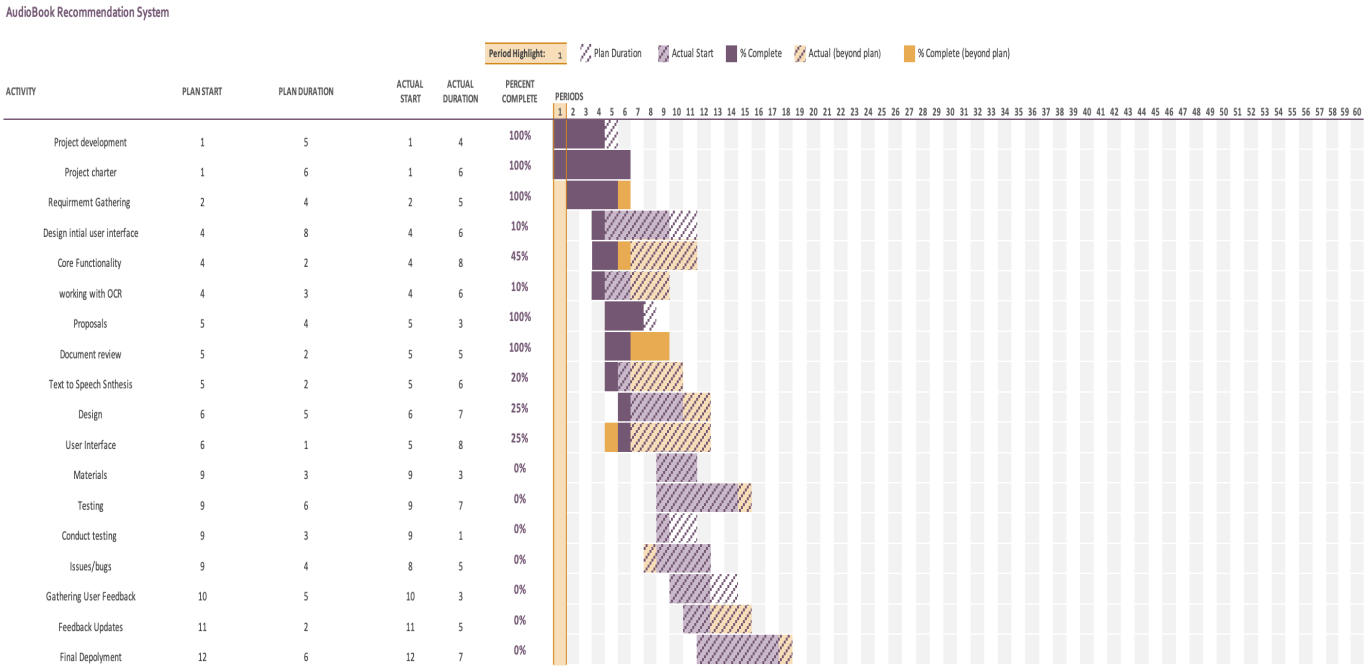


Diagram-03 Gantt Chart

The given above is Gantt Chart shows the status of the Project. Gantt chart helps in planning, scheduling, and tracking project progress efficiently.

References

1. Lemmetty, Sami. Review of Speech Synthesis Technology - SPA, Helsinki University of Technology, Mar. 1999, research.spa.aalto.fi/publications/theses/lemmetty_mst/thesis.pdf.

2. Best, Emily. Audiobooks and Literacy - A Rapid Review of the Literature. National Literacy Trust, Feb. 2020. Accessed Feb. 2024.

3. Pethe, Chart, et al. “Prosody Analysis of Audiobooks.” Cornell University, Stony Brook University, 10 Oct. 2023, https://arxiv.org/abs/2310.06930. Accessed Feb. 2024.

4. Isewon, Itunuoluwa, et al. Design and Implementation of Text to Speech Conversion for Visually Impaired People. International Journal of Applied Information Systems, Apr. 2014.Accessed Feb. 2024.

5. Rohit Barve, et al. International Journal of Research in Engineering, Science and Management Volume-3, Issue-2, February2020 Future Vision Technology:https://www.ijresm.com/Vol.3_2020/Vol3_Iss2_February20/IJRESM_V3_I2_186.pdf

6. OrCam: A New Vision for Machine Learning https://d3.harvard.edu/platform-rctom/submission/orcam-a-new-vision-for-machine-learning/

7. Optical Character Recognition (OCR) https://paperswithcode.com/task/optical-character-recognition

8. WHO|Visual impairment and blindness. WHO, 7 April 1948. http://www.who.int/ mediacentre/factsheets/fs282/en/