ECE445

SENIOR DESIGN LABORATORY

DESIGN DOCUMENT

Voice-Controlled Robotic Study Assistant

Team #46
Jiaxuan He [he59]
Qi Jin [qjin6]
Shuohan Fang [shuohan5]
Yicheng Chen [yc69]

Advisor: Hua Chen
TA: Xihe Shao

March 31, 2026

Abstract

This design document presents the design and verification of a voice-controlled robotic study assistant, engineered to provide hands-free interaction with physical reading materials for users with upper-limb motor impairments. The system accommodates standard academic documents ranging from ISO A5 to ISO A4 sizes, with a maximum binding thickness of 50 mm. The physical architecture features a single auto-adapting document workstation that dynamically centers and secures materials without manual intervention. A closed-loop robotic page-turning mechanism executes a three-step non-destructive sequence: the adaptive paperweights apply a 2.0 N to 5.0 N normal force to secure the document margins, a 15 mm silicone vacuum suction cup lifts the top page using a regulated 1.5 N to 3.0 N vertical force, and a 12 V NEMA 17 stepper motor sweeps the page 180 degrees across the binding within 4 seconds.

System control and user interaction are managed by a central Raspberry Pi microcontroller. The audio interface processes predefined natural language voice commands with a targeted response latency of under 3 seconds. An overhead camera provides closed-loop visual feedback to detect and recover from mechanical faults (e.g., multi-page pickups) and supplies page images for text extraction. Utilizing the PaddleOCR framework, the vision module extracts machine-readable text with an expected accuracy of ≥ 90 %, which is subsequently processed by a text-to-speech (TTS) engine for audio playback. Power is distributed from a 12 V, 5 A main supply to dedicated 5 V and 3.3 V regulated rails to ensure thermal and electrical stability across all logic and electromechanical peripherals. Ultimately, the system targets a continuous operational success rate of ≥ 80 %, providing a reliable, autonomous, and comprehensive assistance reading platform.

Content

1. Introduction

This proposal outlines the design and verification of a voice-controlled robotic system engineered to assist users in interacting with physical reading materials. The system is intended to provide hands-free access to printed documents for individuals with upper-limb motor impairments. By integrating voice command recognition, a closed-loop electromechanical page-turning mechanism, vision-based monitoring, optical character recognition (OCR), and text-to-speech (TTS) technologies, the device enables users to turn pages, switch between document workstations, and listen to the contents of the current page autonomously.

The proposed system is designed for academic and professional settings in which users must interact directly with physical textbooks, notes, and other printed materials. Instead of replacing these materials with digital alternatives, the system provides a real-time assistive interface that supports document navigation, page monitoring, and audio access to printed text.

1.1 Problem Statement

Users with upper-limb mobility limitations encounter significant physical barriers when interacting with traditional printed materials such as bound textbooks, stapled lecture notes, and printed manuals. Basic actions such as turning pages, holding documents in place, or accessing different sections of a book often require continuous physical effort or external assistance.

Although digital accessibility tools are widely available, a large portion of academic and professional materials remains in physical form. Documents containing handwritten annotations, non-standard formatting, or mixed media are often difficult to digitize or inconvenient to access through conventional electronic devices. As a result, users are frequently unable to independently engage with these materials in real time.

Therefore, there is a need for an assistive system that enables direct interaction with physical documents. Such a system must provide reliable page manipulation, intuitive hands-free control, and accessible content delivery without requiring users to manually handle the materials.

1.2 Solution Overview & Visual Aid

To address these challenges, this project proposes a voice-controlled robotic reading assistant that enables hands-free interaction with physical documents.

The system integrates multiple subsystems, including voice command recognition, a closed-loop electromechanical page-turning mechanism, a vision-based monitoring system, and an OCR-driven text-to-speech (TTS) module. Users can issue commands such as "next page," "previous page," and "read page," which are processed by the control unit to trigger corresponding actions. The robotic mechanism performs precise page manipulation, while the vision system verifies successful page transitions and detects potential errors. The OCR and TTS modules extract and vocalize the content of the current page, enabling auditory access to printed text.

Figure 1.1 illustrates a representative usage scenario of the system. A user seated at a desk interacts with the device through voice commands, while the system autonomously performs page turning and reading tasks. This design enables real-time interaction with physical materials without requiring manual operation.

Figure 1.1 Example usage scenario of the voice-controlled robotic reading assistant.

1.3 High-Level Requirements List

The system must meet the following quantitative, high-level functional and performance requirements to be considered successful:

1. **Voice Control and System Latency:** The system must accurately recognize and initiate the execution of predefined natural language voice commands (e.g., "next page," "previous page," and "read page") with a command recognition success rate of ≥ 80 % across 20 trials. The central microcontroller must maintain a processing latency of ≤ 3 s from the completion of the user's speech to the initiation of the corresponding mechanical or software response.
2. **Auto-Adapting Document Interface:** The single document workstation must autonomously center and secure physical reading materials without manual intervention. The mechanical linkages and Z-axis spring actuators must dynamically accommodate document footprints ranging from ISO A5 (148 mm × 210 mm) to ISO A4 (210 mm × 297 mm) and binding thicknesses from 1 mm up to 50 mm, applying a passive normal holding force strictly between 2.0 N and 5.0 N.
3. **Mechanical Page-Turning and Paper Integrity:** The robotic page-turning mechanism must successfully separate and flip a single page forward or backward within 4 s per cycle, maintaining an operational success rate of ≥ 80 % over 20 continuous trials. To prevent multi-page pickups and physical damage, the vacuum suction cup must apply a strictly

regulated vertical lifting force between 1.5 N and 3.0 N, ensuring it remains safely below the 5.0 N tearing threshold of standard 80 $\mathrm{g/m}^2$ printing paper.

4. **Vision-Based Monitoring, OCR, and Audio Output:** The vision subsystem must capture page images and successfully detect mechanical page-handling errors in ≥ 8 out of 10 fault trials. Additionally, the OCR and text-to-speech (TTS) pipeline must extract machine-readable text with an end-to-end character accuracy of $\geq 90\ \%$, and subsequently generate audible speech playback with a measured sound pressure level of $\geq 65\ dB$ (at a 0.5 m distance) within $\leq 2\ s$ of text processing completion.

2. Design and Requirements

2.1 System Block Diagram

The proposed voice-controlled robotic reading assistant comprises five primary subsystems: the central control module, the power distribution module, the voice command interface, the vision-based monitoring and OCR subsystem, and the electromechanical page-turning mechanism. An auto-adapting workstation is included to physically support and align the document during operation. Figure 2.1 illustrates the high-level system block diagram, detailing the flow of power, control signals, and feedback between these components.

The system follows a centralized control architecture. Upon receiving a voice command through the microphone interface, the central controller processes the input and determines the corresponding system action. For navigation commands such as page turning, the controller generates control signals to drive the motors and the vacuum-based end-effector, enabling reliable single-page manipulation. The document is positioned and stabilized by the auto-adapting workstation to ensure consistent mechanical interaction.

During operation, the vision subsystem captures images of the current page and provides feedback to the controller to verify successful page-turning actions or detect potential errors. For reading commands, the captured page image is processed by the OCR module, and the extracted text is delivered to the text-to-speech system for audio output.

The detailed design considerations, component specifications, and verification procedures for each subsystem are described in the following subsections.

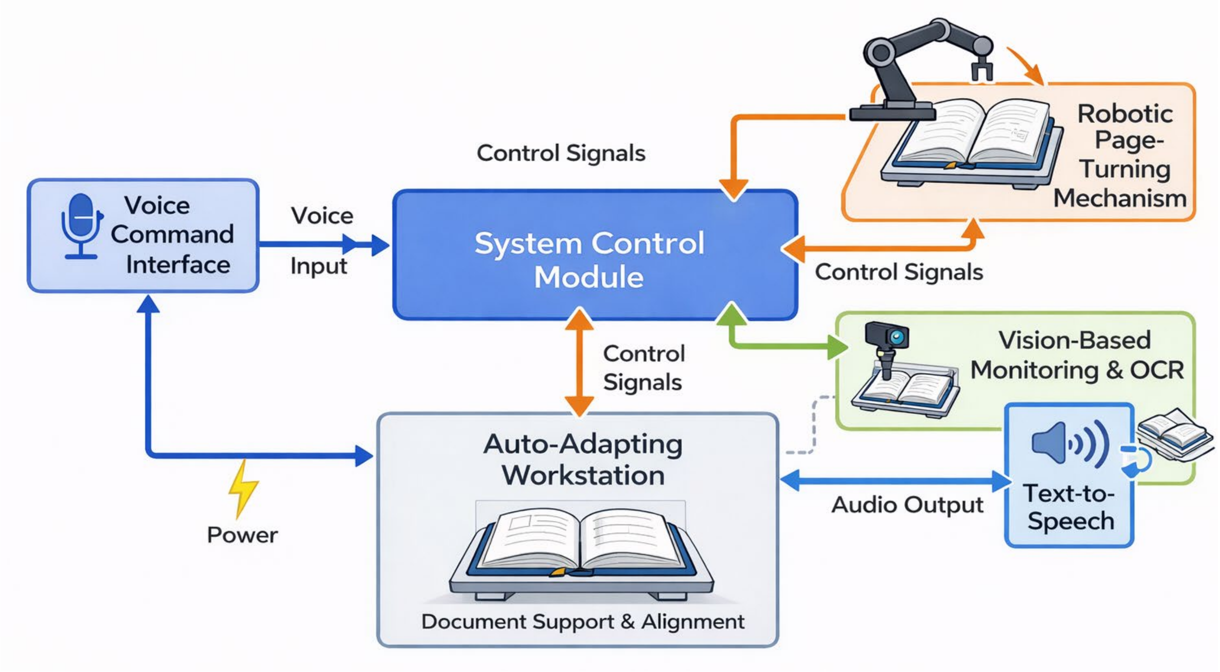


Figure 2.1 Page-Turning Robotic Assistant Block Diagram

2.2 Robotic Page-Turning Mechanism

The robotic page-turning mechanism executes the physical manipulation of the printed media. To accommodate various textbook sizes and binding types without causing damage, the system employs a coordinated three-step sequence. First, two actuated paperweights descend to secure the unlifted pages, applying approximately 2 N of downward force on both the left and right margins. Second, a vertically actuated robotic arm equipped with a 15 mm diameter silicone vacuum suction cup targets the bottom-center edge of the top sheet to initiate the lift. Third, a motorized swing arm rotates across the binding axis to push the lifted page to the opposite side. Figure 2.2 illustrates the spatial arrangement and physical integration of these actuators.

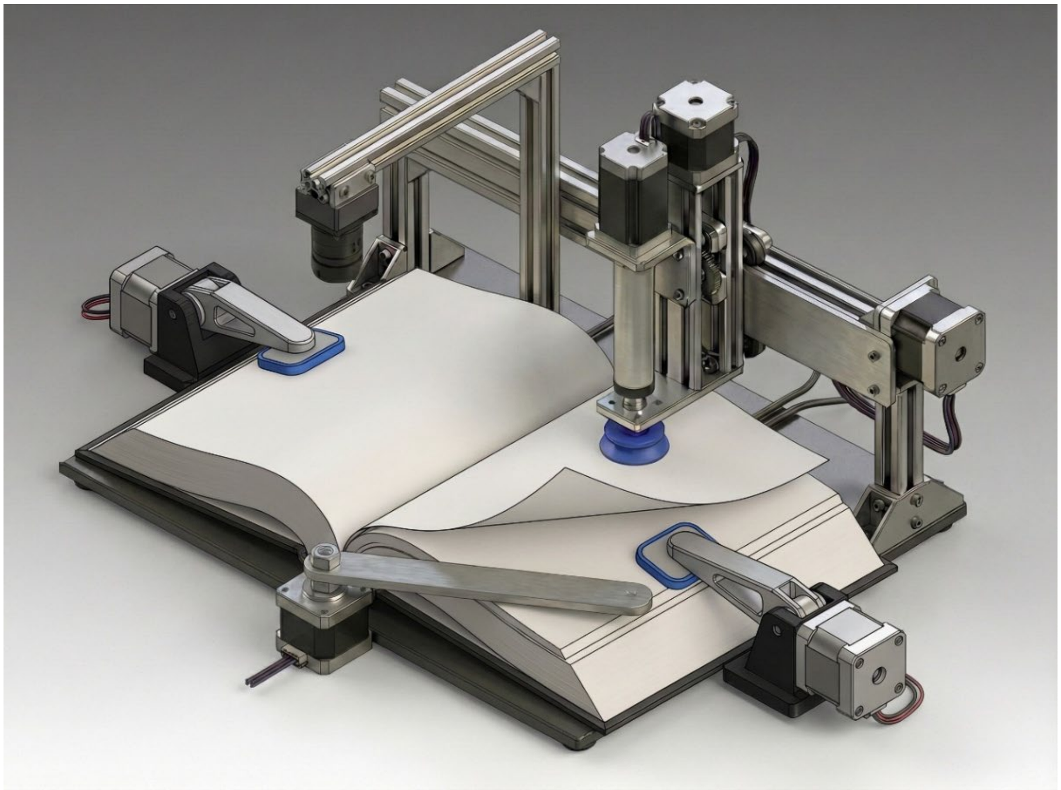


Figure 2.2 CAD Render of the Robotic Page-turning Mechanism and its Actuators

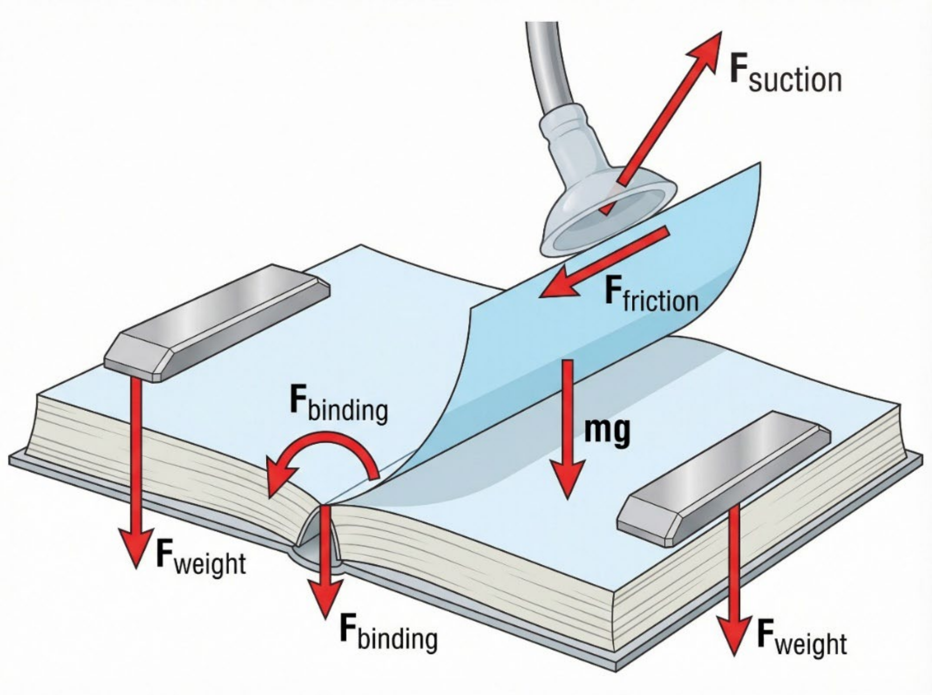


Figure 2.3 Free-body diagram of a single page during the suction lift phase.

To prevent multiple-page pickups and physical damage to the paper, the applied suction force must be precisely calibrated. Figure 2.3 depicts the free-body diagram of the top page during the initial lifting phase.

For a successful single-page lift, the vertical component of the suction force, $F_{suction}$, must exceed the sum of the opposing forces without exceeding the tensile limit of the paper. These opposing forces include the gravitational force on the page (mg), the vertical component of the static friction from the underlying page ($F_{friction}$), and the structural resistance at the book's binding ($F_{binding}$). This dynamic balance is defined in Equation (2.1):

$$F_{suction,y} > mg + F_{binding,y} + F_{friction,y} \tag{2.1}$$

where m is the mass of a single paper sheet, and g is the acceleration due to gravity. The suction force is generated by a 12 V DC micro vacuum pump and is governed by Equation (2.2):

$$F_{suction} = (P_{atm} - P_{vacuum})\pi r^2 \tag{2.2}$$

where P_{atm} is the ambient atmospheric pressure, P_{vacuum} is the absolute pressure within the suction system, and r is the effective radius of the suction cup (7.5 mm). Based on the mechanical properties of standard $80\ \mathrm{g/m^2}$ commercial printing paper, the maximum allowable localized force before fiber deformation or tearing occurs is approximately 5 N [1]. Therefore, the central microcontroller utilizes pulse-width modulation (PWM) to regulate the vacuum pump, maintaining $F_{suction,}$ strictly between 1.5 N and 3.0 N.

Once the suction cup elevates the page by a vertical clearance of 45 mm, the swing arm engages. The swing arm is driven by a NEMA 17 bipolar stepper motor. Operating at 12 V, this motor provides a holding torque of 0.4 N·m, which provides sufficient angular force to overcome the bending stress of the paper and complete the 180-degree sweep across the binding [2].

To ensure the reliability of the mechanical flip sequence and the physical integrity of the paper media, the lifting force and execution timing must be strictly controlled. The quantitative performance targets and the corresponding testing procedures for the page-turning assembly are detailed as shown in Table 2.1.

Table 2.1 Requirements and Verifications for the Vacuum-Based Page-Turning Mechanism

Requirements	Verification
1. The suction mechanism must apply a lifting force between 1.5 N and 3.0 N to separate a single sheet without exceeding the 5.0 N paper deformation threshold.	**1A.** Attach a digital force gauge to the suction cup and activate the vacuum pump at the operational PWM duty cycle. **1B.** Pull the gauge vertically until the suction seal breaks. Record the peak force and verify it falls within the [1.5 N, 3.0 N] range.
2. The motorized swing arm must complete a full 180-degree page turn within $t \le 4\ s$.	**2A.** Place a standard 500-page bound textbook on the workstation and send a "next-page" execution command via the microcontroller. **2B.** Measure the swing arm's travel time from motion start to its resting position using the internal timestamp. Verify $t \le 4$.
3. The mechanism must achieve a $\ge 80\%$ success rate for single-page forward/backward turns over 20 continuous trials, causing zero permanent damage.	**3A.** Execute 10 consecutive forward and 10 backward automated page turns on a bound textbook. **3B.** Visually inspect each turned page under standard 500 lx laboratory lighting. Verify a ≥ 80% success rate of single-page turns with absolutely no tears, wrinkles, or permanent creases.

2.3 Adaptable Document Interface

The system must reliably process various forms of printed media, primarily thick bound textbooks and thin stapled lecture notes. Because these document types possess distinct physical profiles—varying in footprint, thickness, and structural tension—the central reading platform and its associated actuators require dynamic spatial adaptability.

The physical workstation is dimensioned to accommodate document footprints ranging from standard ISO A5 (148 mm by 210 mm) to ISO A4 (210 mm by 297 mm) [3]. To support comprehensive academic materials, the platform allows for a maximum binding thickness of 50 mm. Bound textbooks inherently exhibit significant page curvature and inward binding tension, whereas stapled notes lie flat but are susceptible to lateral shifting during the page-turning cycle.

To counteract these varying physical states, the two actuated paperweights are mounted on vertical linear guide rails, providing a total stroke length of 55 mm. This extended stroke allows the paperweights to descend and secure a thin 1 mm packet of notes or rest atop a 50 mm textbook. The paperweights are driven by micro-servos utilizing active current-sensing feedback to apply a consistent normal force between 2.0 N and 5.0 N on the lateral margins of the document. This specific force range provides sufficient pressure to flatten the natural curvature of a thick textbook spine, minimizing geometric distortion for the overhead optical character recognition (OCR) camera, while avoiding abrasive damage to standard 80 g/m² printing paper. Furthermore, the contact pads of the paperweights are restricted to a width of 15 mm to ensure they secure the document without obscuring the text boundaries.

Similarly, the primary robotic page-turning arm incorporates dynamic Z-axis (vertical) compensation. Because the elevation of the top page changes depending on the document type and current page number, the stepper motor driving the vertical descent of the suction cup utilizes a pre-calibrated step sequence. This ensures the suction cup establishes a vacuum seal at an elevation of 50 mm for a fully opened textbook or descends completely to 1 mm for a flat stack of notes, preventing mechanical stalling and excessive downward force on the paper surface.

The single-workstation interface utilizes a dual-threaded lead screw and spring-loaded weights to provide autonomous spatial adaptation. The following requirements ensure that the mechanism consistently secures various document formats without manual adjustment, as shown in Table 2.2.

Table 2.2 Requirements and Verifications for the Auto-Adapting Document Interface

Requirements	Verification
1. The paperweights must dynamically accommodate document thicknesses from 1 mm to 50 mm, applying a consistent normal holding force between 2.0 N and 5.0 N.	**1A.** Place a 1 mm stack of stapled paper on the workstation. **1B.** Command the paperweights to engage. Insert a thin-film force-sensing resistor (FSR) between the paperweight pad and the paper. **1C.** Read the analog output of the FSR using a multimeter and convert it to Newtons to ensure it falls within the [2.0 N, 5.0 N] range. **1D.** Repeat steps 1A through 1C using a 50 mm thick bound textbook to verify force consistency across the full 55 mm vertical stroke.
2. The workstation must accommodate ISO A5 to ISO A4 document footprints, ensuring the 15 mm paperweight pads secure the pages without exceeding a 20 mm margin from the outer edge.	**2A.** Center an ISO A4 document (210 mm by 297 mm) on the workstation. **2B.** Engage the paperweights and measure the inward intrusion from the physical edge of the paper using digital calipers. **2C.** Verify the intrusion distance is ≤ 20 mm. **2D.** Replace the A4 document with an ISO A5 document (148 mm by 210 mm) and repeat steps 2B and 2C to ensure the actuators can successfully reach and secure narrower margins.

Table 2.2 (continued)

Requirements	Verification
3. The vertical (Z-axis) descent of the suction cup must successfully establish a vacuum seal on surface elevations ranging from 1 mm to 50 mm without causing the stepper motor to stall.	**3A.** Place a 1 mm document on the workstation and command the robotic arm to initiate a page lift. **3B.** Monitor the stepper motor current using an oscilloscope with a current probe. Verify the motor does not draw its maximum rated stall current (e.g., 1.5 A) upon contact. **3C.** Confirm visually that a successful vacuum seal is formed. **3D.** Repeat the procedure with a 50 mm thick book to verify successful seal formation at the higher elevation.

2.4 Power Module

The power module is responsible for receiving, stepping down, and distributing stable DC voltage from a standard 120 V AC mains supply to the corresponding 12 V, 5 V, and 3.3 V subsystems. As illustrated in Figure 2.4, the system receives an initial 12 V DC input via a 5 A AC-DC wall adapter.

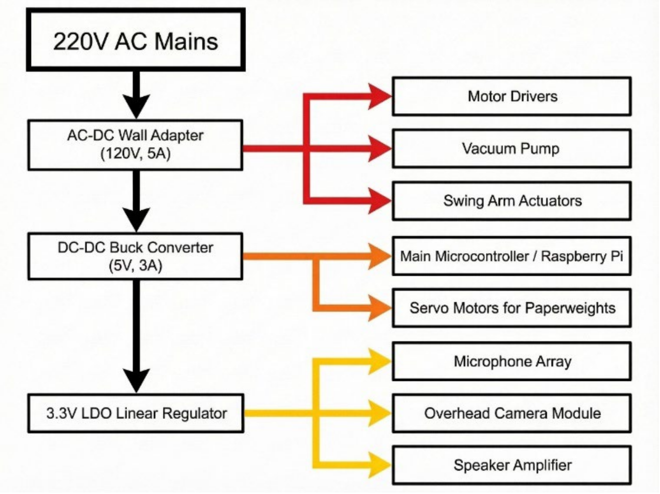


Figure 2.4 Power Distribution Tree Detailing the 12 V, 5 V, and 3.3 V Voltage Rails.

The 12 V rail directly supplies power to the high-power electromechanical components, which include the motor drivers for the NEMA 17 stepper motors, the DC vacuum pump, and the swing arm actuators. The theoretical peak current draw on the 12 V rail is estimated at 2.8 A during a simultaneous page-lift and swing-arm actuation event.

To power the logic controllers and intermediate-power peripherals, a DC-DC buck converter (e.g., LM2596) steps down the 12 V supply to a regulated 5 V rail capable of sourcing up to 3.0 A. This 5 V rail drives the main central processing unit (Raspberry Pi) and the micro-servo motors embedded in the mechanical paperweights. Because the voltage regulators must dissipate excess power as heat, thermal stability is a critical design constraint. The expected junction temperature T_J of the buck converter is calculated using Equation (2.3):

$$T_J = T_A + P_D\,\theta_{JA} \tag{2.3}$$

where T_A is the ambient temperature, P_D is the power dissipation, and θ_{JA} is the junction-to-ambient thermal resistance specified in the component datasheet [4].

Finally, a 3.3 V Low-Dropout (LDO) linear regulator steps down the 5 V supply to power the noise-sensitive logic and analog peripherals. These include the microphone array, the overhead camera module, and the speaker amplifier.

To ensure the selected voltage regulators maintain strict output regulation under continuous dynamic loads without excessive voltage ripple, the components will be evaluated using a custom constant-current test circuit, as depicted in Figure 2.5.

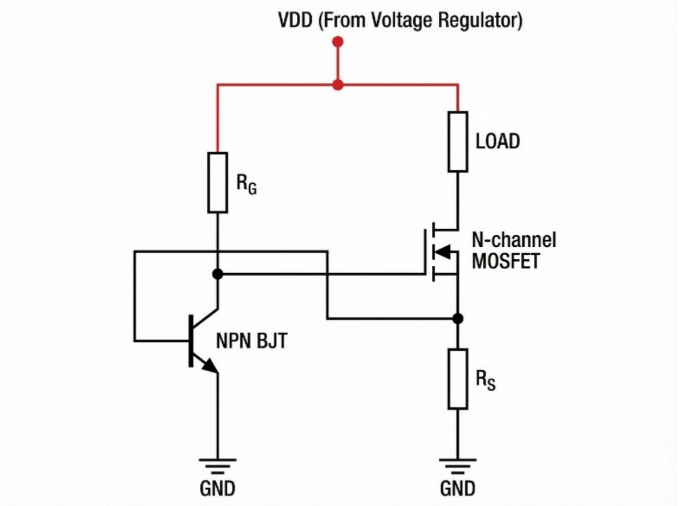


Figure 2.5 Constant-current Test Circuit for Verifying Voltage Regulator Load Stability

In this test configuration, the output of the voltage regulator is connected to the V_{DD} node. An N-channel MOSFET, biased by an NPN bipolar junction transistor (BJT), acts as an active electronic load. By adjusting the value of the sensing resistor R_S, the circuit sinks to a precise and constant current from the regulator, allowing for accurate empirical measurements of load regulation and thermal performance under maximum rated conditions.

To maintain logic integrity and prevent thermal-induced failures across the electromechanical subsystems, the power module must deliver stable, low-ripple voltage rails. The verification protocols for load regulation and thermal safety are detailed as shown in Table 2.3.

Table 2.3 Requirements and Verifications for the Power System

Requirements	Verification
1. The 5 V DC-DC buck converter must maintain a stable output voltage of 5 V ± 5% (4.75 V to 5.25 V) with a voltage ripple of $\leq$ 50 mV under a continuous current load of up to 2.5 A.	**1A.** Connect the 5 V output of the buck converter to the V_{DD} node of the constant-current test circuit shown in Figure 2.5. **1B.** Adjust the sensing resistor R_S to sink a continuous load of 2.5 A, verified using a digital multimeter in series with the load. **1C.** Connect an oscilloscope across the V_{DD} and GND nodes. Verify that the DC voltage remains within the [4.75 V, 5.25 V] range and the peak-to-peak ripple does not exceed 50 mV.
2. The 3.3 V LDO linear regulator must provide 3.3 V ± 5% (3.135 V to 3.465 V) under a 1.0 A load and maintain a surface package temperature below 85 °C after 15 minutes of continuous operation.	**2A.** Connect the 3.3 V output of the LDO to the V_{DD} node of the constant-current test circuit. **2B.** Adjust R_S to sink 1.0 A of current. **2C.** Monitor the output with an oscilloscope to verify the voltage stays within the [3.135 V, 3.465 V] range. **2D.** Maintain the 1.0 A load for 15 continuous minutes. Use an infrared thermometer to measure the IC package surface temperature and verify it is $\leq$ 85 °C.

2.5 System Control and Integration

Figure 2.6 illustrates the top-level state machine workflow that coordinates user interaction and page-turning operations. The overall system is modeled as a simplified state machine starting from the Initialization state. It transitions to the Idle state once all modules are successfully initialized. In the Idle state, the system waits for user voice commands and dispatches tasks accordingly.

Depending on the specific voice command, the system branches into different operational states. A “switch book” command triggers a transition to the Switch Book state, after which the system returns to Idle upon completion. Navigation commands such as “next page” and “previous page” lead to the Page Turn Forward and Page Turn Backward states, respectively, and return to Idle once the action is completed. Similarly, a “read page” command triggers the OCR state for text extraction, followed by the TTS Reading state for audio output, before successfully returning to Idle.

To ensure system reliability, error conditions that occur during workstation switching, mechanical page turning, or OCR processing cause an immediate transition to the Error Handling state. From this state, the system can either recover and return to Idle via a reset operation or transition to the Shutdown state in the case of critical hardware failures. This design ensures a clear, predictable control flow centered around the Idle state, providing unified handling of both user commands and unexpected system errors.

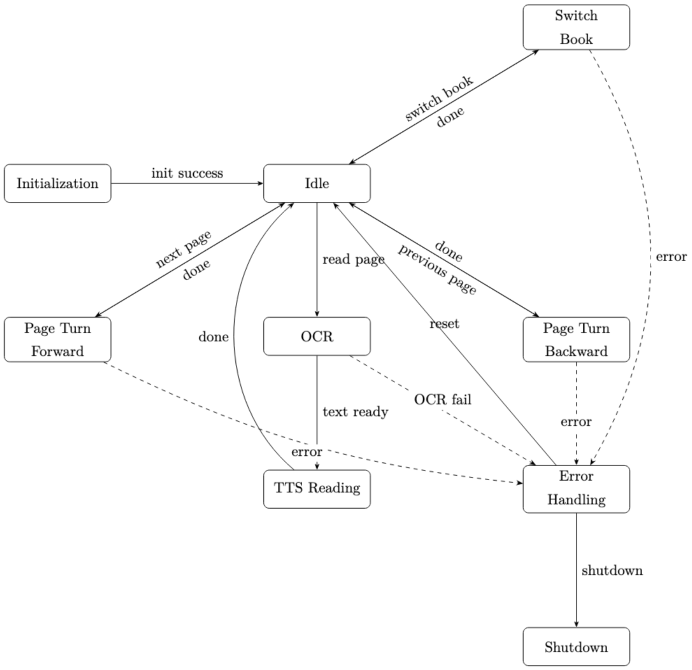


Figure 2.6 Top-Level State Machine Diagram for System Control and Integration

Moreover, Figure 2.7 illustrates the sub-state machine that models the core mechanical actions of a single page flip. The process begins in the Prepare State, where the system ensures that the page is ready for manipulation. It then transitions to the Align state, in which the system positions the suction cup at the edge of the target page.

Once aligned, the system enters the Suction state to activate the vacuum and attempt to pick up the page. If the suction is successful, the system proceeds to the Lift state, where the page is raised. This is followed by the Swing state to perform the flipping motion, and finally the Release state to place the page onto the other side of the document workstation.

In the event of a suction failure, the system transitions to the Fail state and performs a retry by returning to the alignment stage. This sub-state machine focuses on the most critical mechanical steps and isolates the primary failure point at the suction stage, enabling a simple yet effective control flow for reliable page turning.

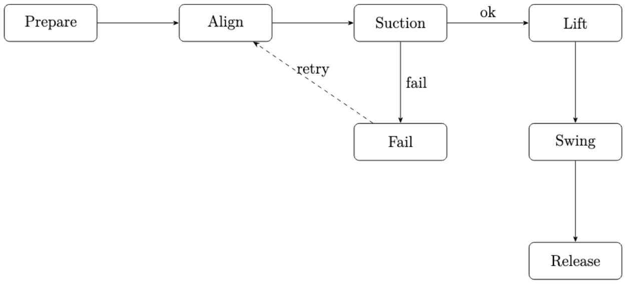


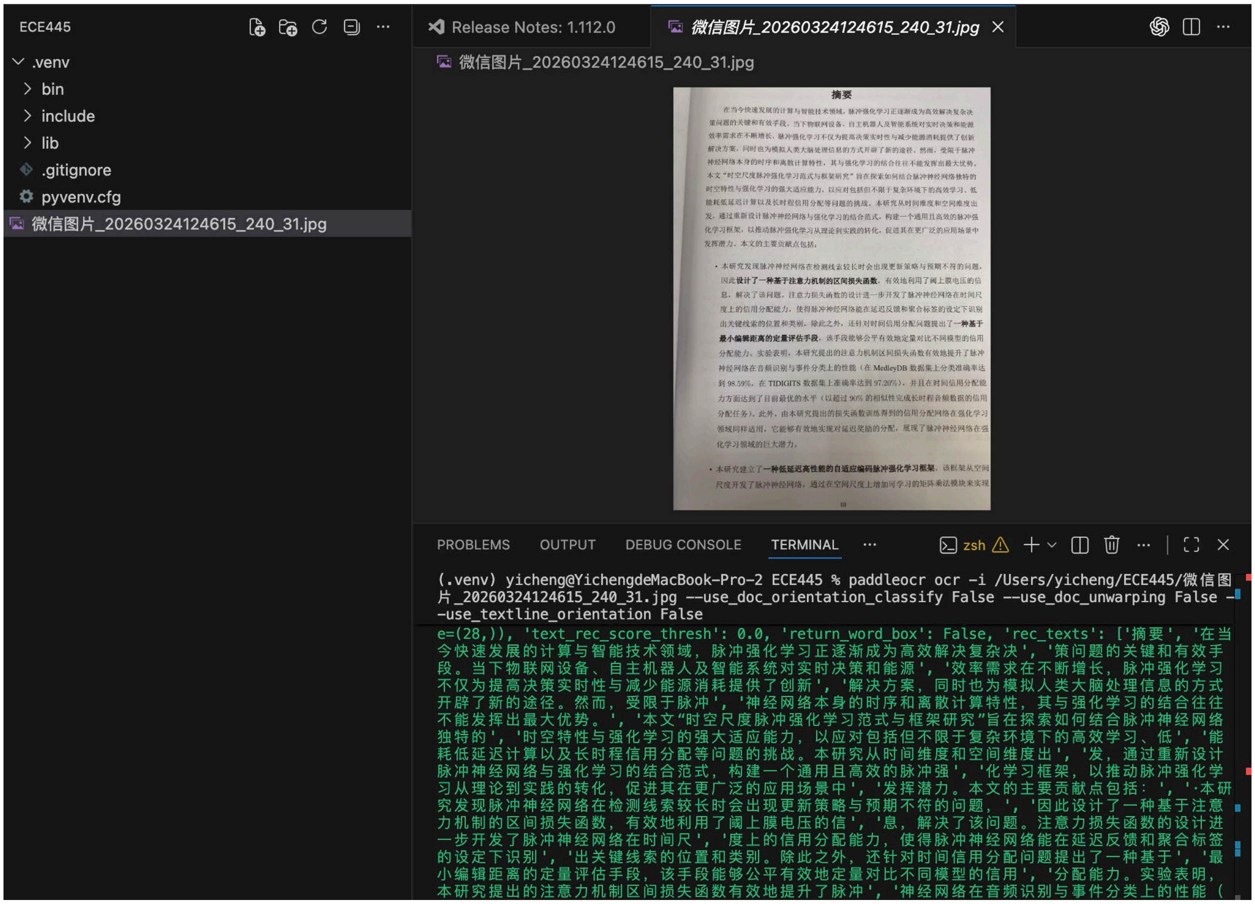
Figure 2.7 Sub-State Machine Diagram for the Mechanical Page-Turning Process

2.6 Vision-Based Page Monitoring and OCR

For the Optical Character Recognition (OCR) module, the system utilizes PaddleOCR, a mature open-source framework widely used for document image understanding. Once a page turn is completed or a voice command is received, the overhead camera captures the current page image. This captured image is first preprocessed to improve text visibility and recognition accuracy. The preprocessing pipeline includes grayscale conversion, page region cropping, contrast enhancement, and perspective correction, ensuring that the influence of page curvature, background noise, and camera viewing angles are minimized prior to text recognition.

The preprocessed page image is then sent to the PaddleOCR pipeline for text detection and recognition. In this design, the OCR module first locates the text regions on the page and then recognizes the printed content line by line, generating machine-readable text for downstream processing. Compared with a basic OCR-only approach, this two-stage pipeline is more suitable for book pages because it handles complex layouts, uneven text distribution, and partial geometric distortion caused by page flipping. Figure 2.8 demonstrates a preliminary test of the PaddleOCR engine accurately extracting printed text from a sample document.

After recognition, the extracted text is further post-processed to remove invalid symbols and obvious noise characters. Fragmented lines are merged into readable text blocks, and the final output is formatted into plain text. This processed text is delivered to the text-to-speech (TTS) module for audio playback and to the keyword search module for content retrieval. Thus, the OCR subsystem serves as a critical bridge between visual page understanding and interactive reading functions. While Tesseract OCR was considered as a lighter-weight alternative, PaddleOCR was ultimately selected for this system because it provides a more modern, document-oriented pipeline and superior practical support for multilingual printed page recognition.



```
(.venv) yicheng@YichengdeMacBook-Pro-2 ECE445 % paddleocr ocr -i /Users/yicheng/ECE445/微信图
片_20260324124615_240_31.jpg --use_doc_orientation_classify False --use_doc_unwarping False -
-use_textline_orientation False
e=(28,)), 'text_rec_score_thresh': 0.0, 'return_word_box': False, 'rec_texts': ['摘要', '在当
今快速发展的计算与智能技术领域，脉冲强化学习正逐渐成为高效解决复杂决', '策问题的关键和有效手
段。当下物联网设备、自主机器人及智能系统对实时决策和能源', '效率需求在不断增长，脉冲强化学习
不仅为提高决策实时性与减少能源消耗提供了创新', '解决方案，同时也为模拟人类大脑处理信息的方式
开辟了新的途径。然而，受限于脉冲', '神经网络本身的时序和离散计算特性，其与强化学习的结合往往
不能发挥出最大优势。', '本文“时空尺度脉冲强化学习范式与框架研究”旨在探索如何结合脉冲神经网络
独特的', '时空特性与强化学习的强大适应能力，以应对包括但不限于复杂环境下的高效学习、低', '能
耗低延迟计算以及长时程信用分配等问题的挑战。本研究从时间维度和空间维度出', '发，通过重新设计
脉冲神经网络与强化学习的结合范式，构建一个通用且高效的脉冲强', '化学习框架，以推动脉冲强化学
习从理论到实践的转化，促进其在更广泛的应用场景中', '发挥潜力。本文的主要贡献点包括：', '·本研
究发现脉冲神经网络在检测线索较长时会出现更新策略与预期不符的问题，', '因此设计了一种基于注意
力机制的区间损失函数，有效地利用了阈上膜电压的信', '息，解决了该问题。注意力损失函数的设计进
一步开发了脉冲神经网络在时间尺', '度上的信用分配能力，使得脉冲神经网络能在延迟反馈和聚合标签
的设定下识别', '出关键线索的位置和类别。除此之外，还针对时间信用分配问题提出了一种基于', '最
小编辑距离的定量评估手段，该手段能够公平有效地定量对比不同模型的信用', '分配能力。实验表明，
本研究提出的注意力机制区间损失函数有效地提升了脉冲', '神经网络在音频识别与事件分类上的性能（
```

Figure 2.8 Example of Text Extraction Using the PaddleOCR Engine on a Printed Document

The specific performance requirements and step-by-step verification procedures for the vision monitoring and OCR subsystem are detailed in Table 2.4.

Table 2.4 Requirements and Verifications for Vision-Based Monitoring and OCR

Requirements	**Verification**
1. The vision pipeline must capture, grayscale, crop, and geometrically dewarp a document image within ≤ 2 s.	**1A.** Place a document on the workstation with a 10° physical skew. **1B.** Trigger the preprocessing sequence and measure the execution time using the microcontroller's internal timestamp. Verify $t \leq 2\ s$. **1C.** Visually inspect the output buffer to confirm the image is successfully cropped to the text boundaries, grayscaled, and perspective-corrected to a true rectangular format.

Table 2.4 (continued)

Requirements	Verification
2. The OCR module must extract multilingual text (English and Chinese) with an end-to-end character accuracy of ≥ 90 % under 500 lx ambient lighting.	**2A.** Place an ISO A4 document containing exactly 500 known characters (a mix of English and Chinese) under 500 lx illumination, verified by a lux meter. **2B.** Execute the "read page" command and export the plaintext output. **2C.** Compare the output to the ground-truth text file. Verify that ≥ 450 characters are correctly identified and ordered (≥ 90% accuracy).
3. The vision system must correctly detect mechanical page-turn completions and physical faults (e.g., multi-page pickups) with an 80% success rate over 20 mixed trials.	**3A.** Execute 10 normal automated page turns. Verify via the serial monitor that the vision module flags all 10 as "successful." **3B.** Manually induce 10 mechanical faults (e.g., tape two pages together or physically block the swing arm). **3C.** Trigger the turn sequence for each fault. Verify that the vision module correctly identifies and flags the error state in ≥ 8 of the 10 induced fault trials.

2.7 Voice Command and Audio Output

The speech recognition system in this work is implemented using an offline voice recognition module integrated with the main control board. The module includes an onboard microphone and performs real-time recognition of predefined voice commands without requiring an internet connection. Compared with cloud-based solutions, this approach provides lower latency, improved reliability, and eliminates dependency on network connectivity, making it suitable for embedded applications. We are considering using Voice Recognition V3 to achieve this.

The voice recognition module operates in a command-driven mode, where a fixed set of voice commands is predefined and trained in advance. In this system, typical commands include “next page,” “previous page,” “read page,” “switch book,” and “shutdown.” By limiting the recognition space to a small set of task-specific commands, the system significantly reduces ambiguity and improves recognition accuracy in practical environments.

The module communicates with the microcontroller through a UART interface. When a voice command is recognized, the module outputs a corresponding command ID, which is transmitted to the control unit. The control program continuously monitors the serial input and maps each received command ID to a specific system action. These actions are then used to trigger transitions in the top-level state machine. For example, a “next page” command activates the forward page-turning state, while a “read page” command initiates the OCR and text-to-speech pipeline.

To ensure robustness, a simple validation mechanism is implemented to filter invalid or unrecognized inputs. If the received command does not match any predefined instruction, the system ignores the input or requests the user to repeat the command. This mechanism prevents unintended operations caused by misrecognition.

The specific performance requirements and verification procedures for the acoustic command interface and text-to-speech (TTS) output are detailed in Table 2.5.

Table 2.5 Requirements and Verifications for Voice Command and Audio Output

Requirements	Verification
1. The voice module must recognize 4 predefined valid commands ("next page," "previous page," "read page," "shutdown") and transmit the correct UART ID with $\geq 80\%$ accuracy and $\leq 3\ s$ latency.	**1A.** Position a user 0.5 m away in a 50 dB ambient noise environment. Vocalize the 4 valid commands 5 times each (20 total trials). **1B.** Monitor the UART RX buffer and internal timestamp. Verify the correct command ID is logged in $\geq$ 16 trials ($\geq$ 80%) and the execution latency is $\leq 3\ s$ per successful trial.
2. The system must reject untrained phrases and environmental noise without triggering false mechanical actuation, maintaining a $\geq 90\%$ rejection success rate.	**2A.** While the system is in the Idle state, vocalize 10 random, untrained phrases of typical command length from 0.5 m away. **2B.** Monitor the top-level state machine. Verify the system safely remains in Idle or requests a repeat in ≥ 9 out of 10 trials, with zero unintended motor actuations.
3. The text-to-speech (TTS) module must initiate audio playback within $\leq 2\ s$ of receiving a text string and produce a clear output volume of $\geq$ 65 dB at a 0.5 m distance.	**3A.** Transmit a 50-word plaintext string to the TTS pipeline. Use an oscilloscope to measure the time until the first analog audio waveform appears. Verify $t \leq 2\ s$. **3B.** Place a digital decibel meter 0.5 m from the speaker. Record the peak volume during playback and verify it reaches $\geq 65\ dB$.

2.8 Schematics

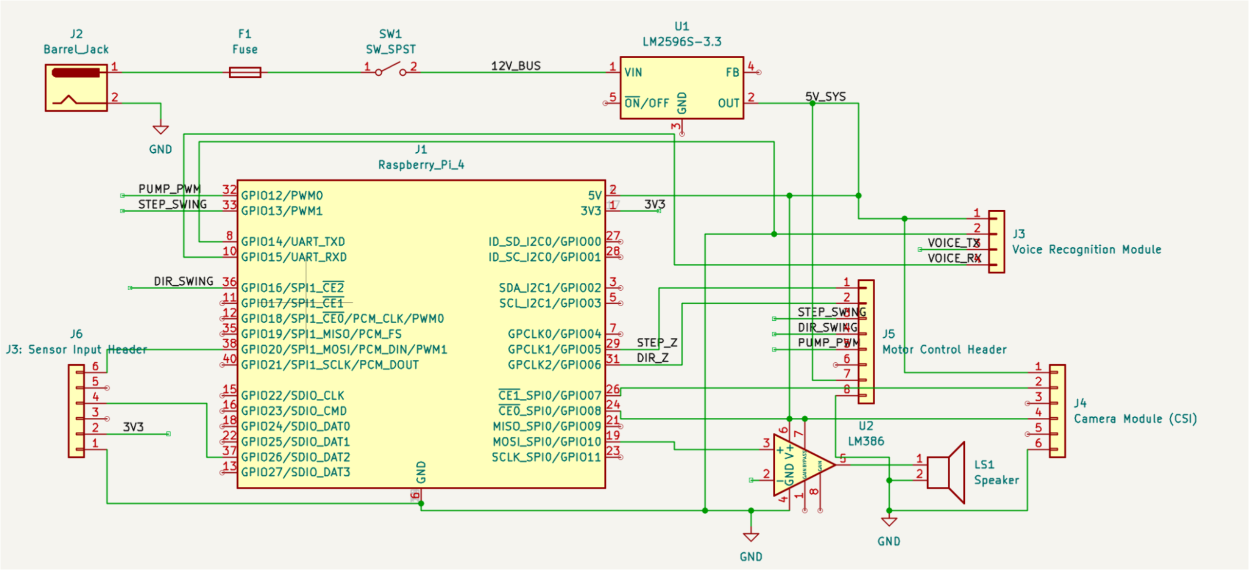


Figure 2.9 Complete system circuit schematic

As illustrated in the complete system schematic in Figure 2.9, the system is powered by a 12 V DC input supplied through a standard barrel jack connector. The input line is first protected by an inline fuse (F1) and an emergency stop switch (SW1), ensuring safe operation under fault conditions. The regulated 5 V supply required by the logic components is generated using a DC-DC buck converter (LM2596), which steps down the 12 V input to a stable 5 V rail (5V_SYS). All system components share a common ground to ensure signal integrity and reliable operation.

The central control unit is implemented using a Raspberry Pi, which coordinates all system-level operations, including voice command processing, state machine execution, image acquisition, and audio output. The Raspberry Pi interfaces with multiple subsystems through its GPIO, UART, and dedicated communication interfaces.

The voice recognition module is connected to the Raspberry Pi via a UART interface, where the module transmits recognized command IDs through the TX/RX lines. This allows the control unit to interpret user commands and trigger corresponding actions such as page turning and text reading.

Mechanical control signals are routed through a motor control header, which provides GPIO-based outputs including STEP and DIR signals for stepper motor drivers, as well as a PWM signal for regulating the vacuum pump. These signals enable precise control of the page-turning mechanism.

Sensor inputs, including limit switches for the Z-axis and swing arm, are connected through a dedicated sensor input header. These sensors provide real-time feedback to the Raspberry Pi, enabling closed-loop control and fault detection during operation.

For audio output, the system utilizes an LM386-based amplifier circuit driven by a GPIO audio signal from the Raspberry Pi. The amplified signal is delivered to a speaker, allowing the system to perform text-to-speech (TTS) playback of recognized document content.

Additionally, a camera interface is reserved for the vision-based OCR subsystem. The camera module connects to the Raspberry Pi through a dedicated interface, enabling image capture for page monitoring and text recognition.

Overall, the schematic integrates power regulation, signal routing, and subsystem interfaces into a unified architecture, supporting reliable interaction between the voice input, mechanical actuation, vision processing, and audio output modules.

2.9 Tolerance Analysis

The most critical tolerance issue in this design is reliable single-page pickup. If the suction force is too low, the top sheet will not separate from the stack or the book binding; if it is too high, multiple pages may be lifted, or the paper may crease or tear.

For the 15 mm diameter suction cup in Section 2.2, the effective contact area is $A = \pi(7.5\ \text{mm})^2 = 1.77 * 10^{-4}\ \text{m}^2$, so the required force window of 1.5 N to 3.0 N corresponds to a pressure differential of approximately 8.5 kPa to 17.0 kPa. This pressure range is modest for a miniature 12 V vacuum source and can therefore be regulated with pulse-width modulation around a nominal operating point. An $80\ \text{g/m}^2$ A4 page has a mass of roughly 5 g, so its gravitational load is only about 0.05 N [1], meaning that the dominant uncertainty comes from inter-page friction and binding resistance rather than page weight. A nominal suction force of about 2.2 N therefore leaves a margin of 0.7 N above the minimum lifting requirement while remaining 2.8 N below the conservative 5 N paper damage threshold for standard printing paper [1]. Likewise, the NEMA 17 swing actuator provides 0.4 N·m holding torque [2], which is significantly larger than the torque needed once a single page has been separated, so the dominant tolerance risk lies in pickup rather than transport.

Based on this analysis, the most critical tolerances are the suction-force regulation and the Z-axis contact position; the control system should maintain the effective suction force within approximately ± 0.5 N of its nominal value and the vertical contact error within approximately ± 2 mm to keep the mechanism inside a safe and reliable operating window.

3. Cost

Table 3.1 summarizes the estimated prototype cost based on the hardware currently specified in Sections 2.2 through 2.8 and current April 2026 prices that were checked on domestic public retail or marketplace pages. Taobao, JD, and Pinduoduo were treated as the preferred sources during collection; however, when the present environment could not stably expose a public platform price, equivalent domestic retail pages were used instead and the corresponding lines are marked accordingly. For items whose exact model has not yet been frozen, such as the paperweight servos, the suction cup set, and parts of the custom mechanical structure, a conservative estimate is retained in the notes column.

For the current single-book prototype, the estimated material cost is RMB 1,484, excluding shipping and common laboratory tools. Several lines still depend on unresolved final SKUs, especially the servos, suction cup set, vacuum pump, and some fabrication hardware, so Table 3.1 should be updated once those retail parts are finalized.

Table 3.1 Prototype Bill of Materials and Estimated Cost

Part / Module	Qty	Unit Cost (RMB)	Subtotal (RMB)	Source / Note
Raspberry Pi 4 Model B development board	1	487	487	Current domestic public retail price from LCSC for Raspberry Pi 4 Model B, 4 GB.
Offline voice recognition module	1	105	105	Current DFRobot China retail price for Gravity offline voice recognition module.
1080p USB webcam	1	31	31	Current Chinagoods marketplace price for a 1080p USB webcam configuration.
NEMA 17 stepper motor (0.4 N·m class)	2	88	176	Current domestic retail price from LeiSai Shop for a 42 mm, 0.42 N·m class stepper motor.
Metal-gear micro servo for paperweights	2	35	70	Exact servo model is not yet frozen; current value remains a conservative retail estimate.
12 V micro vacuum pump	1	38	38	Exact suction-pump SKU is still unresolved; current value remains a conservative retail estimate for a small 12 V pump.
Silicone suction cup and tubing set	1	20	20	The 15 mm suction cup is specified in the design text, but the exact retail SKU has not yet been selected.
12 V / 5 A power adapter	1	31	31	Current Chinagoods marketplace price for a 12 V / 5 A adapter; retail quantity and shipping may increase the final paid cost.

Table 3.1 (continued)

Part / Module	Qty	Unit Cost (RMB)	Subtotal (RMB)	Source / Note
LM2596 buck module	1	41	41	Current domestic public retail price for a ready-made LM2596 step-down module.
Limit switches	2	18	36	Current domestic search-result retail reference for a small mechanical travel or limit switch, approximately RMB 17.65 each.
Emergency-stop switch	1	12	12	Current domestic marketplace price for a 22 mm emergency-stop button.
T8 lead screw kit	2	23	46	Current domestic marketplace price for a T8 lead screw with nut.
2020 aluminum profile and brackets	1 set	121	121	Based on the present frame concept and a conservative allowance for profile stock and connectors.
Control PCB, drivers, amplifier, regulators, connectors, and passives	1 set	150	150	Exact BOM depends on the final schematic in Section 2.8, so this line remains a consolidated estimate.
Acrylic sheet, 3D-printed brackets, springs, and fasteners	1 set	120	120	Exact quantity depends on the final CAD dimensions in Section 2.3.
Total	-	-	1484	Estimated prototype material cost only.

The labor cost is calculated based on the standard formula provided by the ECE 445 guidelines: $Ideal\ Hourly\ Rate \times Actual\ Hours\ Spent \times 2.5$. Considering a localized undergraduate research stipend rate of $¥\,10/\mathrm{hour}$ and an estimated 40 hour of dedicated engineering work per team member, the labor cost is detailed in Table 3.2. The 2.5 multiplier is applied to account for university facility overhead, administrative expenses, and equipment utilization.

Table 3.2 Estimated Labor Cost

Team Member	Hourly Rate	Total Hours	Multiplier	Total Cost
Jiaxuan He	¥ 10	40	2.5	¥ 1,000
Qi Jin	¥ 10	40	2.5	¥ 1,000

Table 3.2 (continued)

Team Member	Hourly Rate	Total Hours	Multiplier	Total Cost
Shuohan Fang	￥10	40	2.5	￥1,000
Yicheng Chen	￥10	40	2.5	￥1,000
Total				**￥4,000**

4. Schedule

The schedule in Table 4.1 reflects the remaining work after completion of the design document and intentionally front-loads low-risk preparation work, such as benchmarking, vendor comparison, documentation cleanup, and interface review, before the team commits significant time to fabrication and integration. This ordering reduces rework risk while keeping the remaining build effort achievable within the available time.

Table 4.1 Remaining Project Schedule and Division of Labor

Week	Jiaxuan He	Qi Jin	Shuohan Fang	Yicheng Chen
4/6	Review the current CAD concept against representative textbooks and lecture notes, then consolidate key mechanical dimensions and reference photos.	Collect sample page images and compare several lightweight preprocessing settings for OCR on curved and flat pages.	Clean up the BOM spreadsheet, compare vendor options, and standardize the verification-record templates.	Organize the controller pin map, interface matrix, and candidate software libraries for motors, sensors, and serial devices.
4/13	Build a simple low-cost mock-up of the workstation layout to confirm basic spacing before formal fabrication.	Prepare a small internal image dataset and summarize the most common OCR failure cases for later testing.	Finalize the procurement priority list, check lead times, and prepare a part-receiving checklist.	Implement a software-only state-machine skeleton and logging framework so later hardware can be plugged in with minimal restructuring.
4/20	Freeze the first mechanical revision and release the fabrication drawings for the frame, paperweights, and suction arm.	Lock the camera position and preprocessing pipeline to be used during subsystem bring-up.	Place orders for the long-lead components and prepare the requirement-verification sheets for integration tests.	Freeze the electrical interfaces and power-tree details for the Raspberry Pi, drivers, amplifier, and limit switches.
4/27	Fabricate and assemble the frame, lead-screw stage, and paperweight structure.	Integrate camera capture with the OCR pipeline on the Raspberry Pi and verify basic page-image export.	Receive and inspect purchased parts, then update the BOM and issue log based on actual deliveries.	Bench-test the 12 V and 5 V rails, motor-driver outputs, and basic sensor inputs.

Table 4.1 (continued)

Week	Jiaxuan He	Qi Jin	Shuohan Fang	Yicheng Chen
5/4	Install the suction arm, vacuum pump, and swing mechanism, then verify manual and semi-automatic single-page flipping.	Implement page-turn success or failure detection and pass status flags to the controller.	Set up verification procedures for page-turn success rate, page damage, and OCR repeatability.	Integrate the low-level motor routines, voice-recognition input, and controller state transitions.
5/11	Tune paperweight force, Z-axis travel, and page-contact position for the selected single-book demonstration setup.	Integrate OCR output with TTS playback and stabilize the current-page text pipeline.	Run the first full round of requirement checks and compile the rework list for the final integration stage.	Close the loop between limit-switch feedback, fault handling, and recovery actions after failed page turns.
5/18	Lock the mechanical assembly, prepare spare parts, and perform repeated page-turning trials before the demo.	Perform final OCR and camera checks under the actual demonstration lighting conditions.	Polish the document set, update the cost table if procurement changed, and prepare the final demo script.	Complete full-system integration, bug fixing, and end-to-end rehearsal for the final presentation.

5. Ethics and Safety

5.1 Ethical Considerations

Developing an assistive robotic platform for users with upper-limb motor impairments necessitates strict adherence to ethical principles regarding accessibility, technical honesty, and data privacy. In accordance with Section I, Item 1 of the IEEE Code of Ethics—"to hold paramount the safety, health, and welfare of the public"—the system design must prioritize the physical and emotional well-being of the user [5]. Furthermore, following Section I, Item 3 regarding the honest disclosure of claims, the team must strictly report verification data without exaggerating the system's capabilities. The device must not be marketed as a universal reading machine if it has only been validated on predefined 148 mm to 270 mm documents under optimal laboratory lighting conditions.

A project-specific ethical concern involves data privacy. Because the system integrates an open microphone array and an overhead camera, it inherently risks capturing sensitive user speech or private desktop environments. To mitigate this ethical risk and comply with privacy protection standards, all audio and optical data must be processed locally on the Raspberry Pi microcontroller. The system is designed to operate without transmitting raw image or voice data to external cloud servers, and all volatile memory buffers containing page images or speech queries will be overwritten immediately after the text-to-speech (TTS) execution is completed.

5.2 Safety Concerns

The primary safety hazards associated with this electromechanical system include mechanical pinch points, unintended robotic actuation, and electrical thermal overloads. The 55 mm vertical stroke of the actuated paperweights and the 180-degree sweep of the motorized swing arm introduce moving interfaces capable of trapping fingers or loose clothing. In compliance with the Occupational Safety and Health Administration (OSHA) standard 29 CFR 1910.212 for general machine guarding, hazardous motion envelopes must be physically isolated or heavily monitored to protect the operator [6].

Accordingly, the physical prototype incorporates an acrylic enclosure around the primary motor linkages and clearly demarcated keep-out zones on the document workstation. An accessible hardware emergency-stop (E-stop) button is integrated to instantaneously sever the 12 V power plane to all high-torque actuators. Electrically, while operating on regulated 12 V and 5 V DC rails, stalled stepper motors can draw currents exceeding 1.5 A, creating localized heating and fire risks. To mitigate these electrical hazards, the power distribution module includes 3.0 A in-line fuses and active current-sensing limits on all actuator branches. During operation, the central controller is programmed to revert to a safe, de-energized Idle state upon detecting an OCR fault, a loss of optical feedback, or an unexpected motor stall, thereby ensuring user safety and preventing physical document damage.

References

[1] J. A. Bristow and P. Kolseth, *Paper Structure and Properties*. New York, NY: Marcel Dekker, Inc., 1986.

[2] NEMA 17 Stepper Motor Datasheet, Pololu Corporation, Las Vegas, NV, 2021. Available at: https://www.pololu.com/file/0J714/SY42STH38-1684A.pdf.

[3] *Writing paper and certain classes of printed matter—Trimmed sizes—A and B series, and indication of machine direction*, ISO 216:2007, Int. Org. for Standardization, Geneva, Switzerland, 2007.

[4] *LM2596 SIMPLE SWITCHER Power Converter 150-kHz 3-A Step-Down Voltage Regulator*, Texas Instruments, Dallas, TX, 2020. Available at: https://www.ti.com/lit/ds/symlink/lm2596.pdf.

[5] IEEE Code of Ethics, IEEE Policies, IEEE, Piscataway, NJ, 2020. Available at: https://ewh.ieee.org/cmte/substations/posted_documents/ieee_codeofethics.pdf.

[6] 1910.212 - General requirements for all machines, web page. Occupational Safety and Health Administration, Washington, DC. Available at: https://www.osha.gov/laws-regs/regulations/standardnumber/1910/1910.212. Accessed March 2026.