1. **Task (A) - Identification and Redaction of Sensitive Information:**
   * **Nature:**
     + Printed, Hand-written, Mixed documents.
     + Text from left to right, top to bottom.
   * Workflow:
     + Pre-processing the input picture
     + Convert image to text by [Microsoft Azure Vision Studio](https://portal.vision.cognitive.azure.com/gallery/featured)
     + Identifying sensitive information by NER model.
     + Redact the sensitive information.
   * **Key Challenges:**
     + Pre-processing required (cropping) e.g. figure 1 to figure 1a,1b,1c
     + Train NER model for specific entity (ID, certain user defined category).
   * **Risks:**
     + Hand-written text overlapped with printed text which is not likely to be correctly recognised.
2. **Task (B) - Identification of English Handwriting on Old Land Documents:**
   * **Nature:** 
     + Mixed documents.
     + Text from left to right, top to bottom.
     + Much formatted than Task A, Clearly line by line, less overlapped situation
   * Workflow:
     + Pre-processing the input picture
     + Convert image to text by [Microsoft Azure Vision Studio](https://portal.vision.cognitive.azure.com/gallery/featured)
   * **Key Challenges:**
     + Pre-processing required (cropping) e.g. figure 1 to figure 1a,1b,1c
   * **Risks:**
     + Abundant of Cursive handwritten text
3. **Task (C) - Optical Character Recognition for Old Press Releases:**
   * **Nature:**
     + English text – At most all printed
     + English text - Text from left to right, top to bottom.
     + Chinese text - Printed, Hand-written, Mixed documents.
     + Chinese text - Text from top to bottom, right to left.
   * Workflow:
     + Pre-processing the input picture
     + Convert image to text by [Microsoft Azure Vision Studio](https://portal.vision.cognitive.azure.com/gallery/featured)
   * **Key Challenges:**
     + Same challenges for English text in task A and B
     + Read Chinese text in correct orientation.
   * **Risks:**
     + The Chinese text is not likely to be read in correct orientation based on previous GRS result.
4. **Task (D) - Auto-generation of Subtitle and Corresponding Tagging for Motion Picture Films:**
   * **Nature:**
     + A mp4 video
   * Workflow:
     + Pre-processing soundtrack (noise suppression, voice bgm separation)
     + Convert voice to text by [Microsoft Azure Speech Studio](https://speech.microsoft.com/portal)
     + Create subtitle for the video.
   * **Key Challenges:**
     + Eliminate background noise.
     + Accurately generating subtitles.
   * **Risks:**
     + Risk of misinterpretation of context.
     + Address the risk of inaccuracies in subtitle generation.

Figure 1: Required to separate the picture into two or more pictures.

A close-up of a letter

Description automatically generated

Figure 1a, Figure 1b

A close up of a ticket

Description automatically generatedA close-up of a document

Description automatically generated

Figure 1c

A letter from a person

Description automatically generated with medium confidencec