

A closer look: Electronic records in the digital world

Use of Artificial Intelligence in document imaging



With the increase of cost of physical space, many organisations are opting to save storage space by digitising (i.e. scanning) hardcopy documents. If you were to visit one of these scanning centres, you would find piles and piles of paper waiting to be scanned. The process of physically preparing the documents for scanning and running the papers through the scanners create dust that perpetuate the scanning room. No wonder one of the biggest challenges faced by scanning centres is the difficulty of retaining staff!

To date, process automation can only partially reduce manpower needs. Most parts of the scanning process are still manual, such as preparing the documents (i.e. removing staples, unfolding dog-ears, and ensuring correct paper orientation, etc.), feeding the documents into the scanners, performing indexing and quality checks on the images, re-assembling the files, etc. On indexing process, although Optical Character Recognition (OCR) could help, manual intervention is often still necessary because of the inaccuracies in the technology. Similarly, performing quality checks is still a largely manual process as no good automated solution is available.

However, with the advances in Artificial Intelligence (AI), an increasing number of organisations are exploring if AI can be used in the imaging process and how it might affect their Evidence Act certification.

Can AI-enabled process be Evidence Act certified?

According to section 116A(5) of the Evidence Act, and section 10 (1) of the Evidence (Computer Output) Regulations, an imaging process can be certified if it complies with all of the Compliance Criteria for Image Systems under the First Schedule of the Evidence (Computer Output) Regulations. There is nothing in the Act or the Regulations that stipulates what technology ought to be used or ought not be used.

Therefore, as long as the AI-enabled process meets the compliance criteria, it can be certified.

What can be AI-enabled?

The two process steps that are most likely to adopt AI-enabled solutions are indexing and quality checking.

OCR tools have long been used to assist in the indexing of imaged documents. This is the process of having machines read and identify text in specific areas of the documents and propose them as indices. However, traditional OCR tools have limitation. They, for example, could have difficulties in differentiating alphabet “o” and numeric “0”, to tell whether a text string is a sentence or a name.

New generation OCR tools with built-in AI are changing this. These AI-OCR tools apply contextual information to make “intelligent” guesses to improve OCR reading accuracy. Moreover, AI-enabled image recognising software could allow photographs be indexed automatically (at least in theory).

Traditionally, scanning centres also check each scanned image manually to ensure that it meets quality standards. Typically, they would check that the images are good representations of the original documents such as no missing pages, dog-ears, etc. This is usually a tedious and time-consuming process.

New AI tools are being developed to help. Reading the images, the tools could determine the probability that there are errors on the pages, and when such probability exceeds a threshold (which can be determined using big data), they will route the images for human check.

Is manual intervention still required?

Unfortunately, technology (available to us commercially) has not reach a state of perfection. Even AI-enabled OCR and quality checking tools can make mistakes. Hence, at least for the foreseeable future, error detection and correction processes are still necessary. What we can hope for is that AI tools will improve the quality of the work, reduce the volume of errors and rework, and achieve overall productivity improvement.

The economies of using AI

When we talk about document imaging, we are usually talking about large quantity of documents. It is not uncommon to talk about imaging in terms of millions of pages, if not more.

If one can save 1 second of manpower per page for 1 million pages, that is a saving of 1 million seconds, which translates to 277 man-hours, or \$4,155 at \$15 per hour.

Typically, it would not be a stretch at all to say it takes a human at least 2 seconds to scan an image page for error, and 3 seconds, to correct an error, say change an alphabet "o" to numeric "0" on an index field.

If this process can be assisted by AI such that there is lesser need for human intervention and fewer errors to be corrected manually, the savings could quickly add up and justify the cost of implementing AI (for that matter, any technology that helps).

Taking the aforementioned into consideration, you might wish to take a closer look at the following areas highlighted below.

1. Is your document captured in its entirety?
2. Is the key content in your document is captured accurately?
3. How do you ensure that the index is entered completely and accurately?

In summary, while creating efficiency for your imaging process, do ensure that your imaging systems and their use are compliant with the Compliance Criteria for Image Systems

For a deeper discussion

Chia Peiru

Senior Manager

9628 9054

peiru.chia@sg.pwc.com

The information contained in this document is of a general nature only. It is not meant to be comprehensive and does not constitute the rendering of legal, tax or other professional advice or service by PwC. PwC has no obligation to update the information as law and practices change. The application and impact of laws can vary widely based on the specific facts involved. Before taking any action, please ensure that you obtain advice specific to your circumstances from your usual PwC client service team or your other advisers. No representation or warranty (express or implied) is given as to the accuracy or completeness of the information contained in this document, and, to the extent permitted by law, PwC does not accept or assume any liability, responsibility or duty of care for any consequences of you or anyone else acting, or refraining to act, in reliance on the information contained in this document or for any decision based on it.

© 2018 PwC. All rights reserved. Not for further distribution without the permission of PwC. "PwC" refers to the network of member firms of PricewaterhouseCoopers International Limited (PwCIL), or, as the context requires, individual member firms of the PwC network. Each member firm is a separate legal entity and does not act as agent of PwCIL or any other member firm. PwCIL does not provide any services to clients. PwCIL is not responsible or liable for the acts or omissions of any of its member firms nor can it control the exercise of their professional judgment or bind them in any way. No member firm is responsible or liable for the acts or omissions of any other member firm nor can it control the exercise of another member firm's professional judgment or bind another member firm or PwCIL in any way.