1 - General Introduction to Handwritten Text Recognition
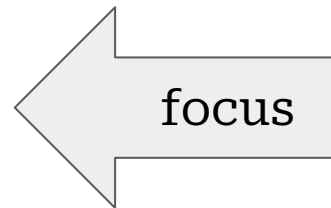
What is HTR?

Document Image → ATR → Machine Transcription
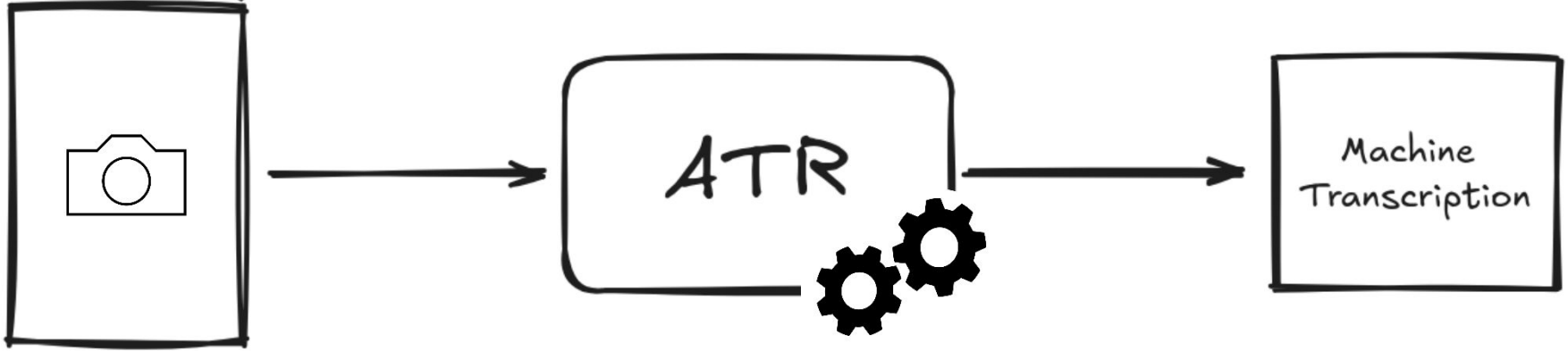
Areas of Application

- academia:
  - facilitate research in other areas, e.g. history, linguistics
  - library indexing/cataloguing

focus

- industry:
  - digitise manually-filled in documents, e.g. insurance forms
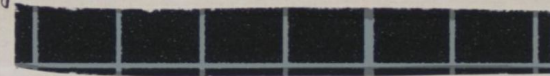 - convert handwritten tablet notes

Typical Challenges

- various forms of ink/background disturbances
  - stains, holes, noise
  - ink bleed-through, fading

Typical Challenges

Typical Challenges

- various forms of ink/background disturbances
  - stains, holes, noise
  - ink bleed-through, fading

- handwriting variability

- orthographic variability

- high cost of annotations
  - special expertise, depending on script

HTR Pipeline

Data acquisition
(images, annotations)

Segmentation

Preprocessing

Text Recognition
(Training, inference)

Postprocessing