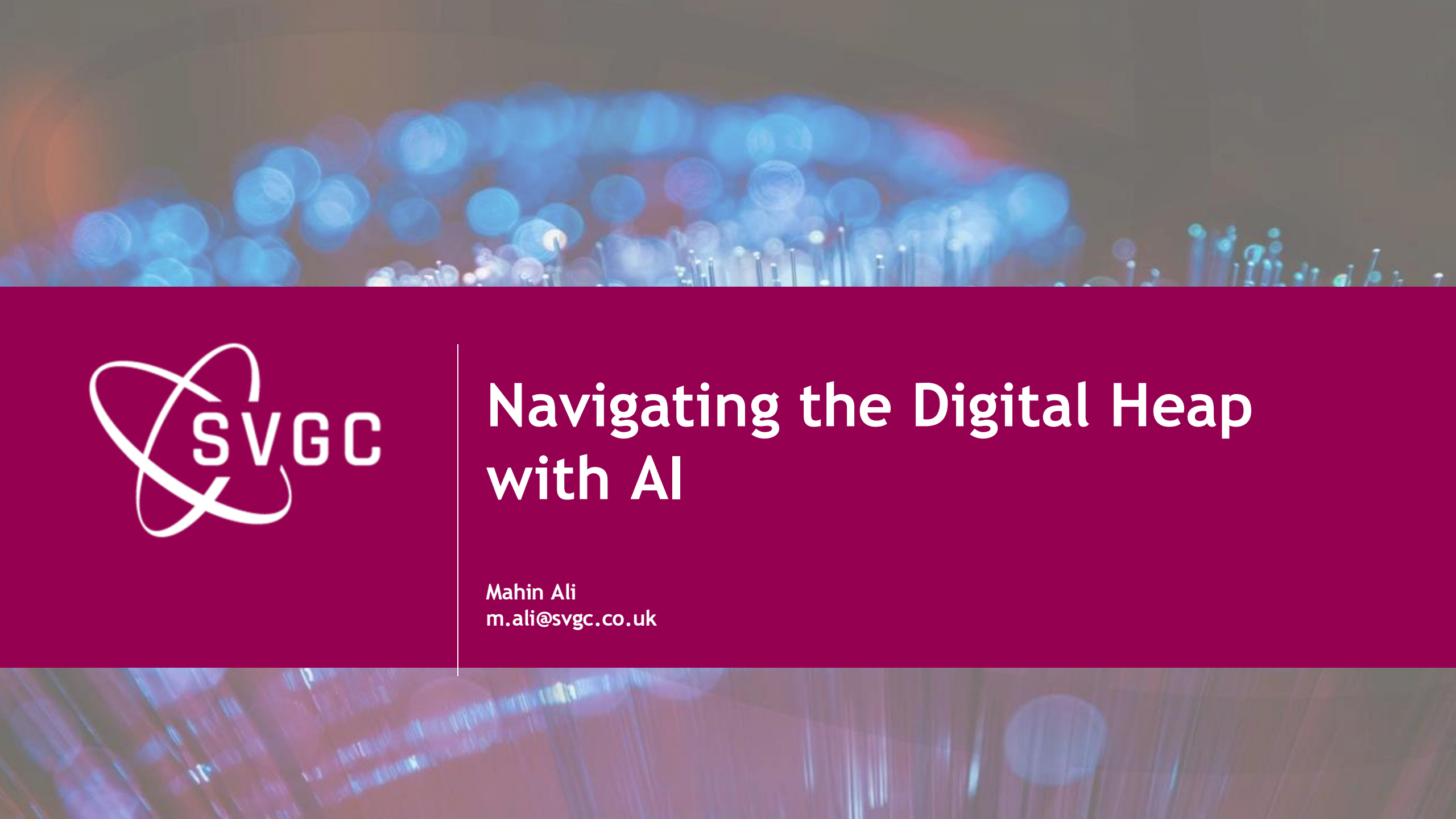


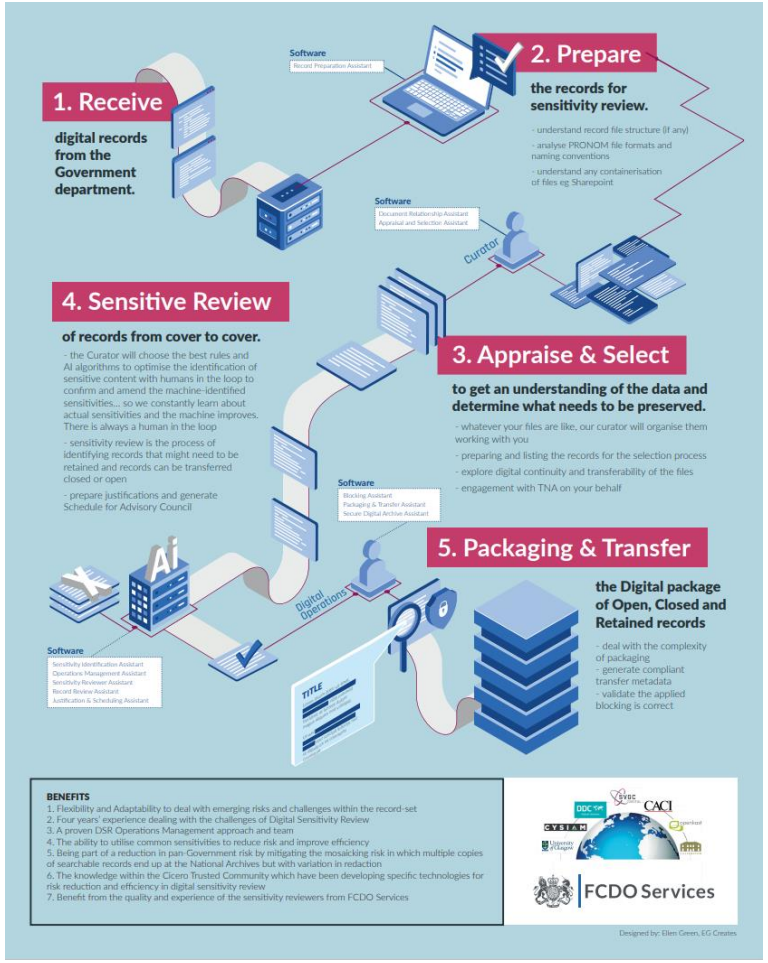
Navigating the Digital Heap with AI

Mahin Ali
m.ali@svgc.co.uk

Who We Are

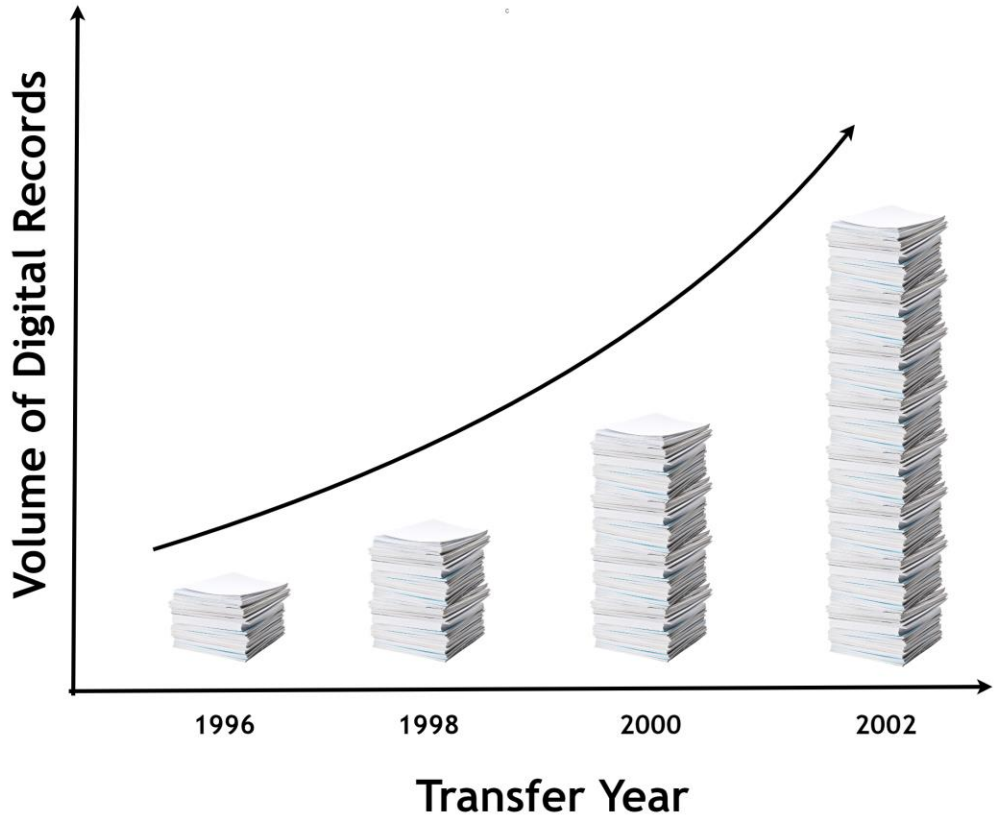
Background

- Over 28 years of history supporting defence, security and government clients as an independent systems and digital engineering consultancy.
- Collaborating with FCDO Services to deliver Digital Sensitivity Review (DSR) solutions across government departments since 2018, facilitating delivery of records to The National Archives.
- Experienced in onboarding AI solutions from both industry and academia to tackle challenges relating to managing government collections through the Cicero Trusted Community.

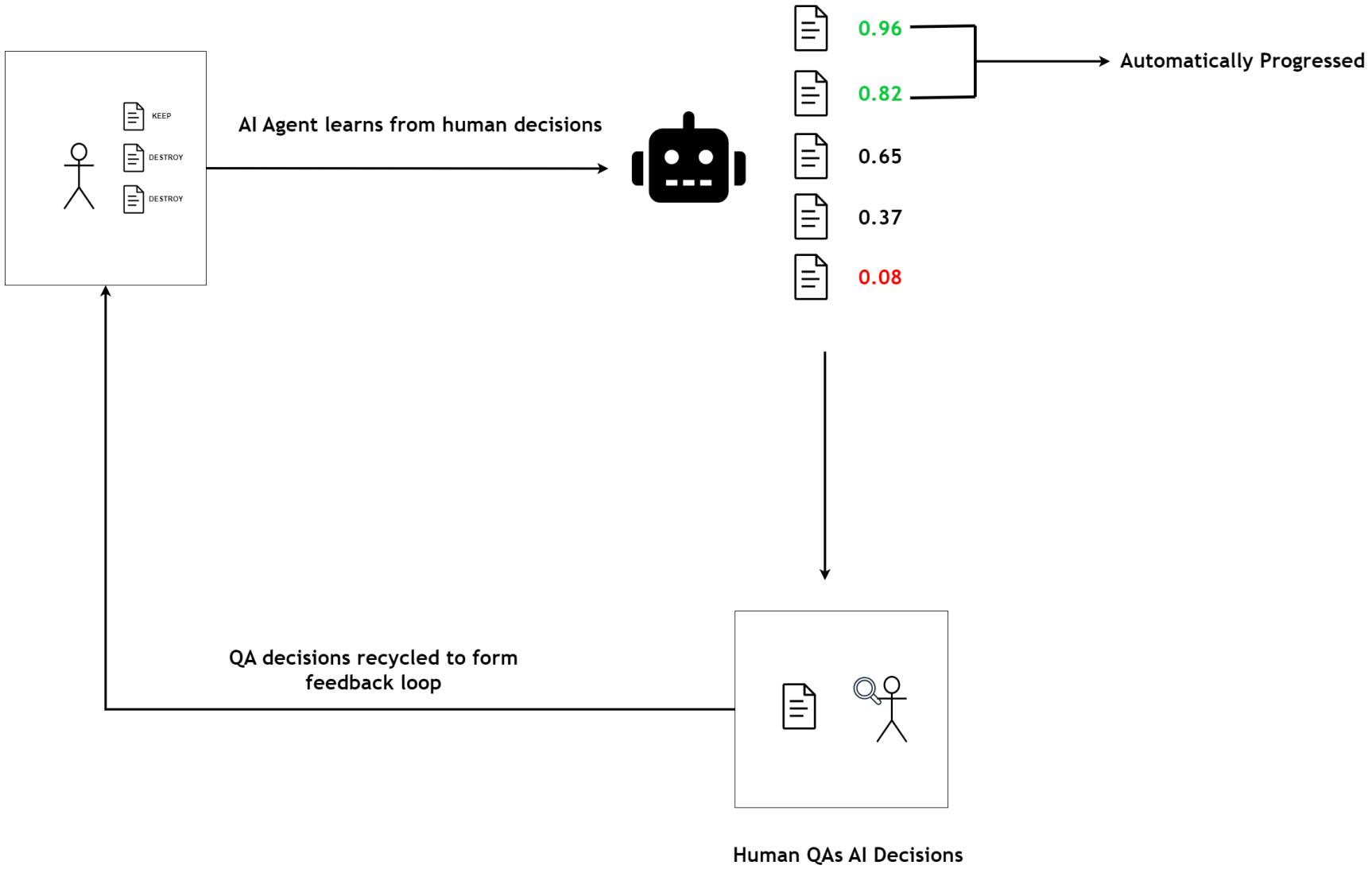


Problem Statement

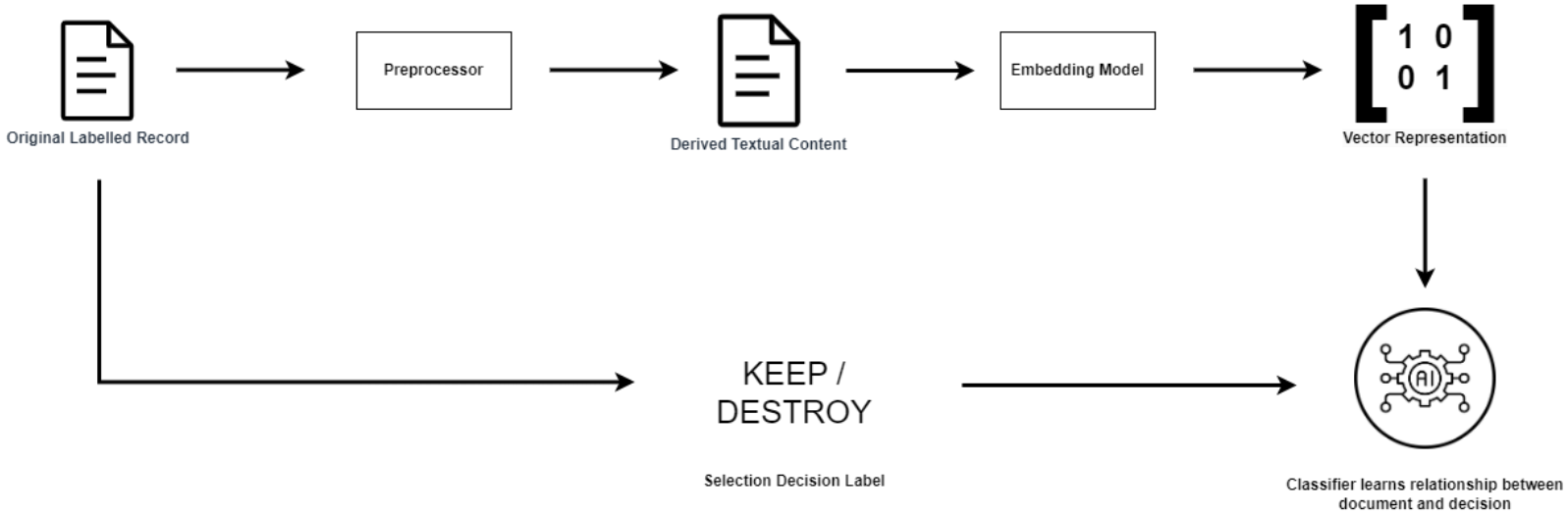
- In 2013, the transfer deadline of public records in the UK was reduced from 30 years to 20 years from the creation of a record.
- However, the volume of records in the digital age grows rapidly year-on-year.
- To preserve compliance with the transfer deadline without drastically increasing the spend on reviewer resource requires technological advancement to automate parts of the process.
- The first stage of review of a record prior to transfer is selection review, the scale of the records at this stage is in the hundreds of thousands.



Our AI Solution (Semi-Automated Selection Framework)



Our AI Solution (Feature Extraction / Training)



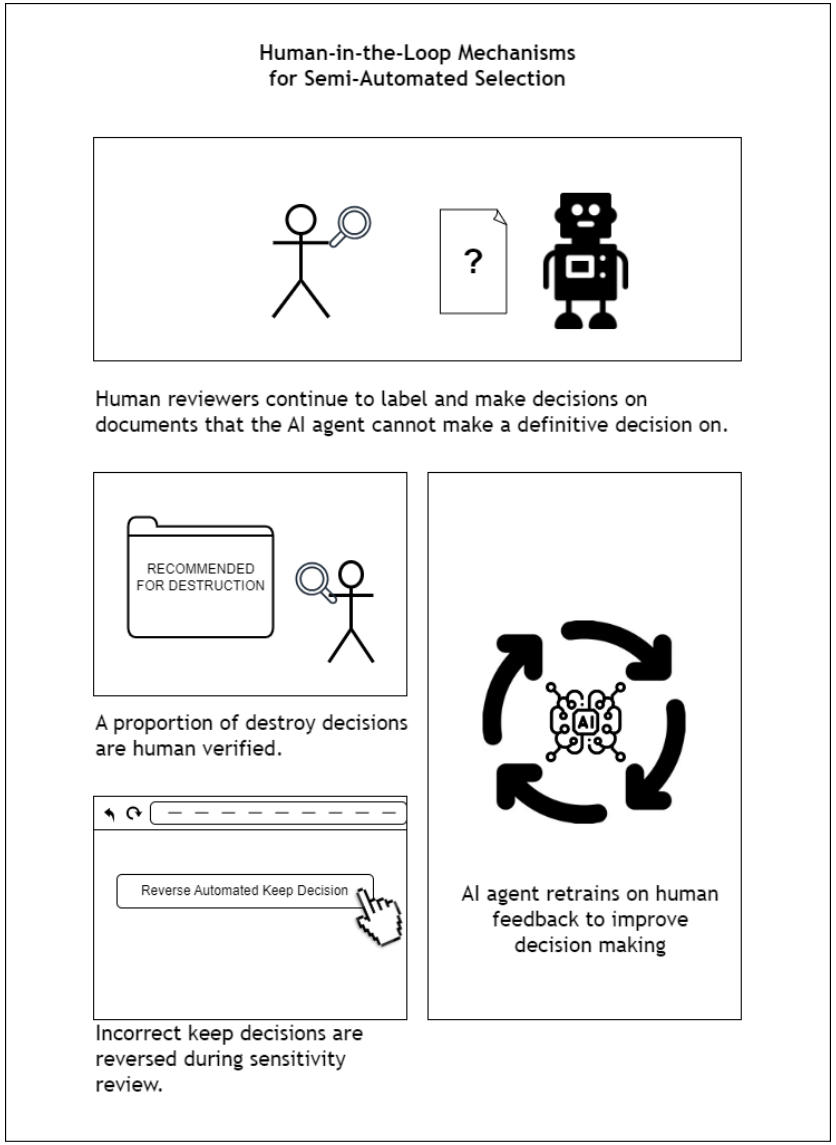
Mitigating Risk (Human-in-the-Loop)

Keeping human reviewers in the loop to review a sampled proportion of automated decisions allows for:

- A continuous stream of labelled training data
- Verification of performance in production
- Detection of data-drift
- Feedback to be collected on model decisions

The QA process is:

- Randomized, to minimise bias.
- Stratified, such that each score threshold and collection has the appropriate level of representation.

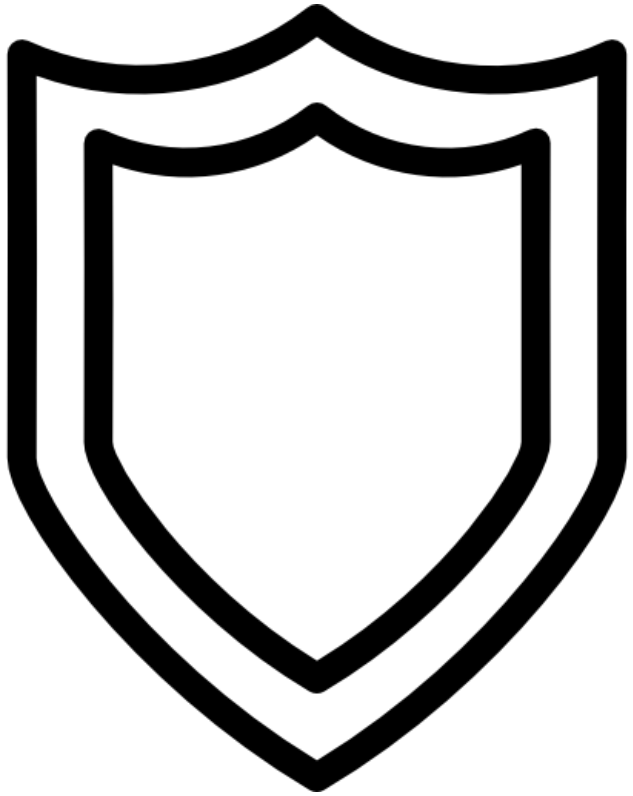


Mitigating Risk (Guardrailing)

Guardrails are put into place such that model inference isn't applied to:

- Records from years where there is a lack of sufficient training data for.
- Transfer series that hasn't been sufficiently represented in the training data.

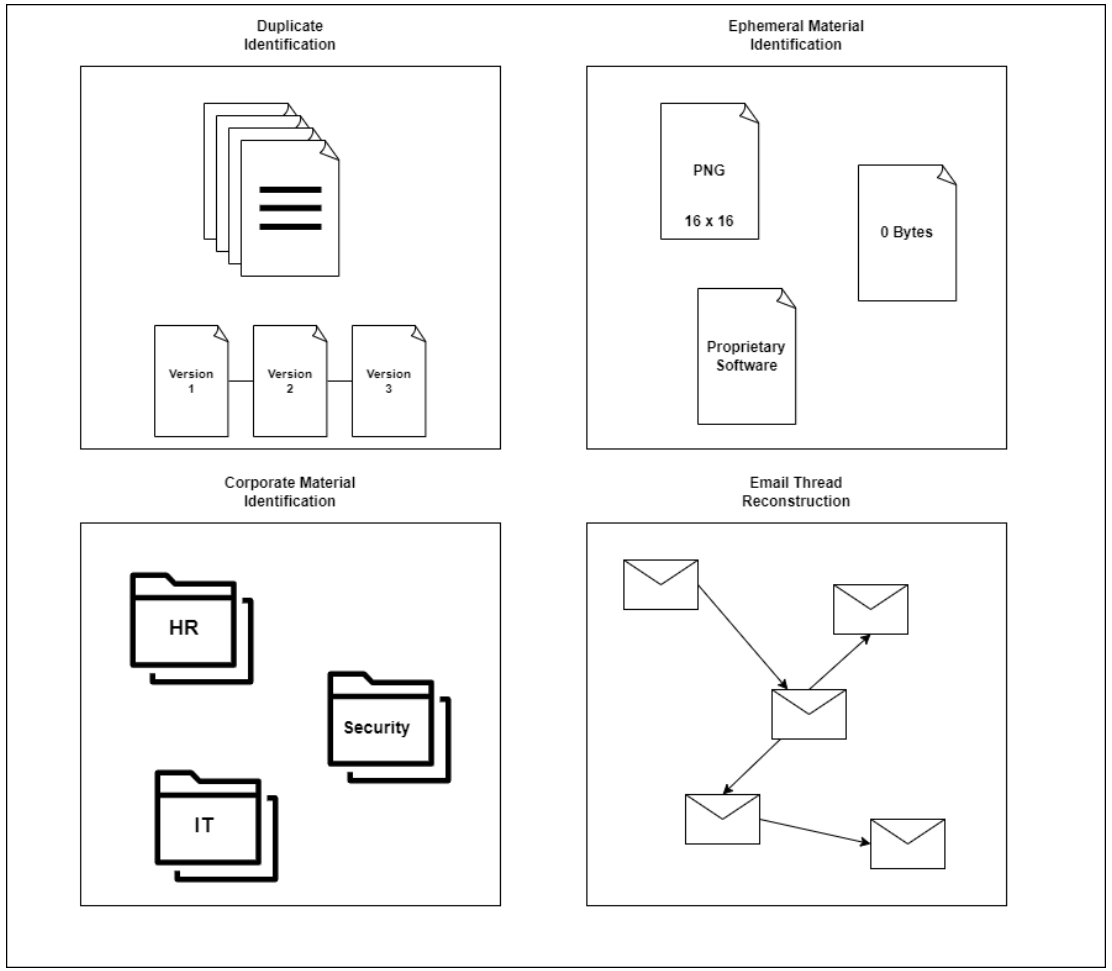
This ensures that the human reviewers remain in the loop for all records that we cannot confidently extend the model policy to, until sufficient evidence in our monitoring processes allow for it.



Mitigating Risk (Deterministic Techniques)

Utilising deterministic techniques in tandem with AI ensures:

- ❑ A reduced human QA burden
- ❑ Higher quality training data and performance increase†
- ❑ A lower dependency on probabilistic outcomes



† *Lee, K. et al. (2022) 'Deduplicating Training Data Makes Language Models Better', arXiv [cs.CL] [Preprint]. Available at: http://arxiv.org/abs/2107.06499.*

Conclusion

Real Impact so far:

80,000+ Decisions Automated

Over 40 Weeks of Reviewer Effort Saved

Verified Compliance with Agreed Level of Risk

Thank you for listening

Mahin Ali
m.ali@svgc.co.uk

svgc.co.uk