Europe

Generative AI ('GenAI') in Outgoing Disclosure

A Companion to the ILTA Active Learning Best Practices Guide

1.1 The purpose of this GenAI Guide (the '**Guide**') is to provide practical industry guidance on the use of Generative AI tools or other advanced review technologies such as GPT-based or proprietary Large Language Models ('**LLMs**') (together '**GenAI**') and approaches in supporting court-ordered document review exercises ('**disclosure**') under Practice Direction 57AD ('**PD57AD**'). That said, the principles and best practices set out in this Guide are intended to be adaptable to other disclosure workflows beyond the parameters of PD57AD where GenAI tools support document review exercises. The Guide is therefore deliberately broad in scope to accommodate evolving use cases and procedural contexts, making it adaptable to reflect the specific requirements of a given matter, including those within applicable procedural rules and directions issued by courts or other adjudicative bodies.

1.2 The Guide is intended to serve as a companion to the ILTA Active Learning Best Practices Guide ('**A-L Guide**'), building upon and complementing its contents. The Guide references but does not replicate the A-L Guide.

1.3 The Guide assumes that GenAI can be effectively deployed in disclosure alongside existing **Active Learning** methodologies, however, it recognises that GenAI may also be used outside of Active Learning workflows. GenAI use cases in disclosure are evolving rapidly, therefore the Guide is drafted in intentionally board terms to support adaptability over time, and flexible, defensible practices across a range of matters which may be subject to external scrutiny or judicial interrogation. The appropriate adoption of GenAI in any given matter should be considered and agreed by the Parties on a case-by-case basis, noting that additional factors beyond those addressed in the Guide may also be relevant to those circumstances.

1.4 The Guide is not prescriptive. Parties are encouraged to apply and adapt it to reflect the specific circumstances of their matter, including the nature of the disclosure, applicable procedural rules, and any relevant court directions. Where GenAI is used under PD57AD disclosure, parties should cooperate and, where appropriate, agree on the intended scope and parameters of GenAI use, which should be documented in an amended Section 2 of the **Disclosure Review Document** ('**DRD**').

1.5 GenAI is a tool to assist practitioners, not a decision-maker. Practitioners remain fully responsible for any decisions, certifications, or representations made in the course of legal practice, including any content signed, submitted to the court, or relied upon in proceedings. GenAI outputs must therefore be treated as aids to legal analysis and must be used in tandem with practitioner legal judgment and verification.

2 DEFINITIONS AND BACKGROUND

2.1 The definitions below are provided to help readers understand the scope and context of GenAI as used in the Guide. Parties should assess whether a legal or technical definition applies in their matter and determine whether the tools they intend to use fall within the Guide's scope. Where relevant, those tools should be clearly identified as GenAI systems to which this Guide's standards and recommendations apply.

2.2 According to the *UK Courts and Tribunals Judiciary Guidance for Judicial Officer Holders*, Generative AI is:

> *A form of AI which generates new content, which can include text, images, sounds and computer code. Some generative AI tools are designed to take actions.*[1]

2.3 For the purposes of the Guide, the following features are particularly relevant:

(a) A **GenAI workflow** refers to the defined process by which GenAI tools are used to support or automate one or more steps in a disclosure exercise. This may include the selection or design of prompts, the application of the GenAI tool to specific datasets, and the handling, review, and validation of GenAI outputs. A GenAI workflow may operate independently of, or alongside, other review technologies such as Active Learning, and typically includes human oversight, quality assurance, and prompt management mechanisms. GenAI workflows should be appropriately documented, monitored, and updated as the disclosure project progresses.

(b) **Automation bias** refers to the human tendency to over-rely on, or uncritically accept, the information or decisions provided by automated or AI systems. This can lead to reduced independent critical evaluation of the system's outputs with the consequential risk of inaccurate decisions.

(c) GenAI models are based on **foundation models**. Foundation models are: '*Machine learning models trained on very large amounts of data that can be adapted to a wide range of tasks.*'[2] These models are general purpose; models may differ, for example, in their design, the data they have been trained on, and their purpose.[3] As a result, a preferred model may be relevant to which GenAI system is chosen, how it is integrated into a workflow, and how it is used and tested. The GenAI element used in disclosure may not be integrated into the document review system and may instead operate as a separate or external component. The document review platform itself may incorporate additional workflows, whether or not GenAI-enabled, which influence how outputs are produced, and affect the associated risks and available mitigations.[4]

1. AI Judicial Guidance
2. A pro-innovation approach to AI regulation: government response - GOV.UK
3. For example, reasoning models such as Chat GPT-01 compared to GPT models like Chat GPT-5. Further, Large Language Models are a type of Foundation Model, however there are other types of Foundation Models, such as Multi-Modal Models.
4. For example, guardrails against hallucination risk.

(d) A GenAI system requires **prompts** (i.e. an input) which will in turn produce responsive **output** by the GenAI system. Users may take multiple additional steps between prompting and output (e.g. adopting technical guardrails that limit what prompts can be used and what outputs can be generated; '*grounding*' outputs so they are based on one or multiple specific document(s), etc.). Iterating prompts to achieve the optimum output is known as **prompt engineering**.

(e) GenAI differs from **Active Learning**. Active Learning uses deterministic algorithms[5] that are generally publicly available[6], have been subject to academic research and are tailored to electronic document review tasks. In contrast, GenAI relies on non-deterministic[7], stochastic foundation models[8]. These models are typically developed and owned by third parties rather than the document review platform provider and are therefore: (i) subject to additional intellectual property and commercial protections; (ii) not designed for the purpose of electronic document review; and (iii) are not always transparent. An eDisclosure/eDiscovery vendor may not provide explicit notification of the model in use, and models may be updated or replaced in the background without user awareness. The practical implications for the purposes of the Guide are that: (i) GenAI demands a greater need to consider workflow design and testing; and (ii) it may only be possible to evaluate performance by assessing inputs and outputs, rather than by examining the design of the foundation model itself. These differences do not preclude the use of GenAI, but they do give rise to additional considerations that parties should be aware of and seek to manage appropriately. There is an inherent risk that GenAI systems may change over time (due to model updates or retraining, for example) which may affect outputs and consistency.

(f) A **feedback loop** refers to a structured process by which insights or results from the review stage (e.g. machine feedback and suggestion, reviewer decisions, coding accuracy, or error patterns) are communicated. Feedback loops enable continuous improvement of the workflow, which may include prompt optimisation, model calibration, or other workflow adjustments. They can be formal (e.g. via validation reports or structured quality control reviews) or informal (e.g. reviewer comments or SME[9] flags) and are particularly useful in identifying issues such as automation bias, inconsistent coding, or evolving relevance criteria.

(g) **Technical guardrails** refer to system-level constraints, controls, or safeguards that are implemented within a GenAI workflow to prevent inappropriate, inaccurate, or unverified outputs. These may include: (i) limiting the scope of prompts; (ii) restricting access to certain datasets; (iii) enforcing input/output validation rules; (iv) grounding outputs in specified source documents; or (v) requiring a '*nil response*' where the system lacks sufficient confidence to generate a reliable output. Technical guardrails are critical to managing risks associated with potential fabricated output/'hallucination', bias, overreach and automation error, and should be tailored to the specific GenAI tool, workflow, and intended use case. They should also be documented and monitored throughout the review lifecycle.

5. Whereby the same input will result in the same output.
6. Therefore are transparent, measurable and mimicable.
7. Whereby the same inputs could (but do not necessarily) result in different outputs.
8. Therefore are not measurable or mimicable.
9. Defined at paragraph 5.1 (k) (iii)

3 CIRCUMSTANCES APPROPRIATE FOR GENAI

3.1 The explanation at Section 3 of the A-L Guide is applicable here.

4 POTENTIAL USE CASES

4.1 The following is a non-prescriptive, non-exhaustive list of current potential use cases for GenAI. It is possible that one or more of the below may be used in combination with an Active Learning workflow:

Processing stage

(a) *Data processing and clean-up* – Enhancing text recognition in documents to enable more effective text-based searching, analytics, and manual review.

(b) *Converting documents to text for search and analytics* – Transcribing video, audio or image files into text to enable text-based searches (such as keyword or concept searches) or supporting analytics (such as clustering).

Review

(c) *Issue identification and categorisation* – Identifying conceptual issues and grouping thematically similar documents. This can assist in identifying key themes more efficiently and may reveal patterns or connections that might not otherwise have been identified.

(d) *Document Categorisation & Prioritisation* – Enhancing Active Learning and TAR processes by prioritising potentially high-value/relevant documents, key themes, custodial patterns and relationships between entities.

(e) *Multi-document summarisation* – Analysing multiple documents to generate a narrative from multiple documents, and (/or) for individual documents.

(f) *First-pass relevance review* – GenAI may be used to replace first-tier review entirely, allowing potentially relevant documents to be escalated directly to second-tier review. It can also be used to support a human first-pass relevance review by providing reviewers with additional information and/or context to inform their decision-making. For example, GenAI may generate a relevance score to offer a rationale for why a document may (or may not) be relevant in response to a given prompt, which could include references to exemplar documents[10]. In these scenarios, the final relevance determination remains with the reviewer.

10. For example, via Retrieval Augmented Generation.

(g) *Privilege identification* – Supporting the identification and grouping of potentially privileged material. GenAI may assist by identifying key privilege indicators such as the involvement of legal personnel, references to legal terminology, or the context of actual or anticipated litigation or legal advice.

(h) *Redaction Assistance* – Automating or enhancing the identification of content requiring redaction (e.g. by identifying information that is both not relevant and confidential such as personal data, bank account details, or other generic identifiers). GenAI may assist by flagging such content for further human review, supporting defensible redaction decisions.

(i) *Quality checking and sampling* – Supporting quality control by identifying documents or groups of documents for targeted review. This may include cases where GenAI's suggested output differs from a human reviewer's decision (e.g., a document flagged as relevant by GenAI but not by a reviewer), or where sampling is required to test the accuracy or consistency of review outcomes for validation or audit purposes.

(j) *Sentiment Analysis* – Analysing text to identify and assess the sentiment expressed within documents (e.g., positive, negative, neutral). This may assist in collating, categorising or prioritising documents that reflect emotional tones or attitudes, which can be relevant to certain issues or themes.

(k) *Chain of Inquiry Analysis* – Identifying documents that may not be directly relevant themselves, but which lead to further lines of inquiry.

(l) *Anomaly and Pattern Detection* – Detecting outliers and anomalies in communication patterns.

(m) *Foreign Language Review* – Supporting the identification, translation and analysis of non-English (or other non-primary review language) content. GenAI may assist by detecting and flagging foreign language material (including mixed-language documents), producing machine translations to provide reviewers with an initial understanding of the content, and highlighting key terms or passages for targeted human translation where accuracy is critical. It may also identify potential linguistic nuances or idiomatic expressions that could affect legal interpretation.

5 PRACTITIONER BEST PRACTICE

5.1 The following key takeaways, drawn from current industry practice, are intended to support legal teams:

(a) **Accountability and responsibility** - GenAI is a support tool, not a substitute for legal judgment. The party using a GenAI workflow should be accountable and responsible for its use at all times.

(b) **Technical input** – Parties should seek appropriate technical expertise to support the design, testing, and deployment of GenAI workflows.

(c) **Co-operation** – Where the use of GenAI is proposed, parties are encouraged to engage in early-stage (and continuing, where necessary) discussions to agree on its use, scope, and associated guardrails. As set out at paragraph 1.4, these discussions should be clearly recorded in procedural documents (e.g. the DRD) or case management correspondence as appropriate.

(d) **Consider the tool's pricing model and costs** – There are different pricing structures for different GenAI tools. The pricing structure can have a material impact on overall disclosure costs or otherwise shift the point at which costs are incurred during a disclosure process.

(e) **Joint technology session** – In some cases where GenAI will be used in a disclosure workflow, it may be useful at the earliest opportunity[11] for parties to hold a joint technology session. Such a session, distinct from formal CCMCs, CMCs, or Disclosure Guidance Hearings, can facilitate common ground on technical aspects and lead to more effective cooperation and agreement on rules governing GenAI use.

(f) **Identifying the purpose of GenAI** – As set out at paragraphs 1.4 and 5.1(c) of the Guide, the intended use of GenAI should be clearly set out in *Section 2* of the DRD, including how it will be deployed, its role in the disclosure workflow, and the specific use cases (e.g., classification, privilege review, and/or redaction). Where custom prompts or workflows are used, the methodology for prompt design, testing, and refinement should be outlined, along with any prompt repositories maintained. *Section 2* should be treated as a living record and updated as the methodology evolves, with changes promptly communicated to the other party(ies) and the court where necessary.

(g) **Identifying the dataset** – Parties should describe the dataset(s) over which GenAI tools will be deployed. Different tools and prompts may be appropriate for different types of data within the overall dataset, depending on the nature and purpose of the analysis.

(h) **Identifying documents or issues not suitable for GenAI** – Parties should assess whether there are specific documents or categories of documents for which GenAI may not be appropriate. This determination will depend on the technology available and the nature of the data and may evolve over the life of a disclosure project as tools develop. For example, parties should consider file type and file size. Parties should then consider alternative workflow(s) for use on documents identified as not suitable for GenAI.

(i) **Identifying appropriate benchmarks** – As with Elusion Testing in Active Learning, parties should agree on appropriate benchmarks to assess the accuracy, consistency, and reliability of GenAI outputs. These may include: (i) alignment with human-coded ground truth datasets; (ii) consistency of outputs across similar documents; (iii) alignment of privilege or issue tagging with known exemplars; and (iv) stability and reproducibility of outputs across prompt iterations. Benchmarks should be established during workflow design, incorporated into validation protocols, and reviewed periodically as GenAI tools or prompts are refined. Where possible, results should be recorded to demonstrate defensibility.

11. This should be conducted before a CCMC, ideally.

(j) **Iterative testing to design workflows** – Parties should consider iterative testing (i.e. repeatedly testing and refining) for each GenAI workflow. This may include: (i) sample testing to check the suitability of a workflow for a larger dataset; or (ii) iteratively testing prompts and measuring accuracy against pre-defined benchmarks. Parties should consider keeping a record of such testing, observations about outcomes, and what additional steps they took as a result.

(k) **Reviewer instructions** – Parties should consider what information they provide to reviewers about GenAI's purpose(s), intended use and safeguards. It may be beneficial to have a feedback loop to those responsible for GenAI workflow design and operation. Parties should also be aware of the risks of automation bias.

(l) **Effective feedback loops between the AI system and reviewers** – where documents are only classified by GenAI, there is a risk that GenAI's coding decisions differ from those of human reviewers. Parties should consider: (i) how to keep human reviewers informed of GenAI outputs; (ii) how to incorporate reviewer observations and output verifications into GenAI workflows; and (iii) how to ensure that incorrect or inconsistent GenAI outputs are actively identified and fed back into the workflow to improve review quality.

(m) **Validate the output** – Parties should consider at what point to validate output. This could be on an ongoing basis to provide a feedback loop for the GenAI workflow (and reviewers more generally), and/or at the end of a review workflow. Validation methods could include:

(i) *Defining failure thresholds and fallback plans* – Set these before deploying GenAI. If the threshold set is surpassed, document the result and switch to a manual or alternative workflow until the issue is resolved.

(ii) *Random Sampling* – Reviewing a statistically significant sample of GenAI-classified documents to validate consistency and accuracy through further manual review. If the AI system provides GenAI-produced summaries, parties should consider validating output without reference to these summaries.

(iii) *Elusion Testing* – Elusion testing should be conducted where applicable, ensuring that no relevant documents remain excluded from review due to GenAI misclassification. In this way GenAI and Active Learning can be used in tandem to enhance the defensibility of GenAI.

(iv) *SME Testing* – Applying independent second-tier review to ensure that the AI's classifications align with subject-matter expert ('**SME**') expectations.

(v) *Precision and Recall Metrics* – Calculating precision (the proportion of relevant documents retrieved) and recall (the proportion of all relevant documents identified) to assess GenAI efficacy.

(vi) *Active Learning* – Using statistics to identify (and manually re-review) documents on which GenAI, Active Learning and/or a human reviewer disagreed.

(n) **Ending a review** – GenAI can be used in combination with Active Learning to determine a reasonable point at which to end a review. Parties should consider in advance how and when they will be able to identify such a point, as the timing of ending a review has a direct impact on manual review costs. However, parties should also validate that decision at the time of ending the review.

(o) **Audit trail** – In addition to recording the intended GenAI workflow in the DRD, parties may also consider maintaining a separate audit log of prompts which may include information such as: (i) what prompts were used; (ii) for what purposes; (iii) when; (iv) over which part of the dataset; (v) the rationale for each prompt; (vi) any amendments to those prompts; and (vii) explanations for changes and the timing of those changes. Parties should also consider recording what GenAI model was used (such as vendor, name and version), when, and under what parameters.

(p) **Appropriate understanding of prompts** – Parties (or in practice, their SMEs) should seek to understand and be able to explain how a prompt works technically. If that is not feasible, parties should at least be prepared to explain the results of iterative testing or sampling. Parties should also seek to understand, explain and justify the technical guardrails in place throughout a GenAI workflow. For example, some GenAI systems are designed so that they should give a '*nil response*' where they are unable to produce an accurate response, rather giving an inaccurate response.

(q) **Appropriate transparency** – Parties may be subject to transparency requirements in a case, for example, because of court directions or an agreed methodology. In the absence of such a requirement, it is for parties using or intending to use GenAI to consider how/when it provides appropriate transparency to another party or the court. Transparency is crucial, particularly in respect of: (i) the intended or actual use of GenAI; (ii) the purposes for which GenAI is and is not used; (iii) GenAI workflow design and how and when it is tested. The Guide uses the phrase '*appropriate transparency*' because what transparency looks like will depend on the circumstances of a given case. The Guide acknowledges that the general need for transparency may be subject to limitations, such as a party's rights or obligations to withhold certain information (for example, due to legal privilege).

(r) **Appropriate explainability** – Parties may be required to explain the intended or actual use of GenAI, such as how the GenAI workflow is designed, operates and has been tested. Parties should seek to put the court in an informed position to be able to exercise necessary supervision of the disclosure process and the parties' compliance with applicable procedural rules and directions. Parties should consider in advance what steps they need to take in order to explain their process as and when required (e.g. (i) identifying an appropriate SME at an early stage; (ii) maintaining a structured audit log).

(s) **Disclosure Certificate** – Parties can summarise the proposed or agreed GenAI workflow in the Disclosure Certificate.