

Vox PopulAI

Lessons from a global
law firm's exploration
of generative AI

contents

Overview	4
----------	---

Ashurst GenAI trial insights	6
------------------------------	---

1. The initial value of GenAI for legal is in helping lawyers create first drafts quicker and more efficiently	7
--	---

2. GenAI-generated output can be difficult to distinguish from human output when legally correct	8
--	---

3. Quality in a legal context is multi-dimensional, with subjective and objective elements	9
--	---

4. Beyond legal and client work, GenAI can make day-to-day life easier and help people feel more prepared for the future	10
--	----

5. The 'jagged frontier' of AI in the legal industry needs clarification to derive value	11
--	----

Ashurst trial design approach and methodology	12
---	----

Recommendations for running GenAI trials	15
--	----

In conclusion: "My eyes are more open"	16
--	----

Appendix	17
----------	----

Common Assessment Criteria	17
----------------------------	----

Acknowledgements	18
------------------	----

Endnotes	18
----------	----

Contacts and Authors	19
----------------------	----

Overview

The speed at which organisations are adopting generative AI technology (GenAI) shows no sign of slowing. Business leaders in every industry are seeking to identify the best use cases for GenAI and where it can add the most value to their organisations. This is as true in law firms as it is in other industries. At Ashurst, our focus has been on taking this a step beyond use case identification and value prediction to truly understanding and also measuring how GenAI might impact the way we work and serve our clients.

From November 2023 to March 2024, Ashurst's Office of the Chief Digital Officer (the CDO team) led three global GenAI trials involving 411 partners, lawyers and staff representing all of our practice areas and business services functions across 23 offices in 14 countries in our global network. Through the trials, we sought to prove or disprove a series of working hypotheses concerning whether, how and how much our people would engage with and find value in GenAI. The trials were designed to capture a mix of qualitative and quantitative assessment data through a blind study, controlled experiments with small groups of participants, and feedback surveys, giving our people a voice in how GenAI could benefit them. Outcomes were achieved using only publicly available data, not client data, and by adopting rigorous guardrails throughout.

Our findings uncovered five key insights:

- 1. Value:** The initial value of GenAI for legal is in helping lawyers create first drafts quicker and more efficiently.
- 2. Accuracy:** GenAI-generated output can be difficult to distinguish from human output when legally correct.
- 3. Quality:** Quality in a legal context is multi-dimensional, with subjective and objective elements.
- 4. Scope:** Beyond legal and client work, GenAI can make day-to-day life easier and help people feel more prepared for the future.
- 5. Adoption:** The '*jagged frontier*' of AI in the legal industry needs clarification to derive value.

This report explores these insights and the detail of Ashurst's approach and methodology applied to the conduct of GenAI trials, as well as the specific outcomes of certain experiments and the blind study, with the aim of helping other organisations navigate their own GenAI exploration. In this report, we also provide some recommendations based on our experience for other organisations wishing to conduct their own experiments with GenAI and drive improved experiences for their people.

As we continue to experiment with this technological capability as it matures, our understanding, insights and approach will evolve, and we hope to learn further lessons from future exercises that will help us serve our clients and our staff even better.



Ashurst GenAI trial insights

1

The initial value of GenAI for legal is in helping lawyers create first drafts quicker and more efficiently

Understanding and measuring the value of GenAI in assisting with common legal work types was a key research objective.

The trial data indicated that the greatest initial value for GenAI in a law firm context is in helping lawyers to create first drafts quicker and more efficiently. The value of GenAI was not solely in generating content, but also using GenAI to accelerate collecting and understanding the source material used to inform the first draft. Feedback indicated that participants immediately saw value in using GenAI to get a rough first draft, which they could then reshape. Across the tools used and attempts to draft multiple types of documents, an average of 77% of post-trial survey respondents agreed or strongly agreed that usage of GenAI helped them get to a first draft quicker.

In our controlled experiments, we measured approximate time savings of 80% to draft UK corporate filings requiring review and extraction of information from company articles of association, 59% to draft industry/sector-specific research reports that require reviewing and extracting key information from public company filings (Form 10-Ks), and 45% on creating first draft legal briefings.

One participant from the blind study (see [Ashurst trial design approach and methodology](#) below for more information) noted that, *“the GenAI case summary was a useful starting point as it was easier to have...rather than a blank page”*. This sentiment was echoed in other participants’ feedback. GenAI, therefore, was very helpful in enabling participants to overcome so-called “blank page syndrome”.

It is worth noting that we also engaged trial participants in preparing first drafts of non-legal content, such as job descriptions, policy documents, and social media posts. In our controlled experiments, we measured approximate time savings of up to two hours per participant per task for this type of content and a very high overall quality scoring (compared to the overall scoring of legal content – see the next two insights for further detail). GenAI helped participants not just to draft the content, but it helped pull the relevant source material into one place for them to review. Participants noted that GenAI output was appropriate, useable and having a first draft to react to enabled them to move forward more effectively.

It is Ashurst’s view that using GenAI in this way has the potential to impact the drafter’s workflow by freeing them up to conduct higher-value work or other development activities. Producing first drafts quicker may also mean the reviewer receives a first cut sooner and, ultimately, may increase the delivery speed of legal services externally. However, it is important to add a healthy caution: potential time savings are always subject to the content being legally correct. If the initial steer provided by GenAI is incorrect, there is a risk that the drafter is sent along an incorrect path, which may result in more time being taken to correct the work once checked. We, therefore, found that potential value is tied to the accuracy of the GenAI output. For this reason, it is also imperative that appropriate safeguards are put in place to ensure that work generated by GenAI is thoroughly and properly reviewed by appropriate experts.

GenAI-generated output can be difficult to distinguish from human output when legally correct

It has been well established that large language models themselves do not produce output that can be relied upon as being legally correct.ⁱ Therefore, we sought to understand whether and how accuracy might be improved when using both generic ‘assistant’ type tools and tools designed for specific legal use cases, in each case enabling contextualising with legal data.

This was investigated by our blind study, where common assessment criteria were designed to allow Ashurst to benchmark the GenAI output of different tools against the outputs of our lawyers working without GenAI tools. Accuracy formed part of that criteria (see the [Appendix](#)). The expert panel were asked to assess if output (in this instance, case summaries) was “*legally correct, precise, misguided and/or contradictory*” and apply a score out of 5. Quantifying output in this way provided measurable and comparable data points between the output from the GenAI tools and the work of Ashurst lawyers.

Our data indicated that when judged on accuracy alone:

- **GenAI-generated content had a greater variance in accuracy:** the output accuracy scores for GenAI-generated first draft content spanned 1 to 4 out of 5 versus 3 to 4 out of 5 for human-generated first draft content; and
- **Human output still scored higher on average:** the average output accuracy score for GenAI-generated content was 3 out of 5 versus 3.5 out of 5 for human-generated content.

Our expert panel was also asked to predict if the output was GenAI- or human-generated. We asked this question because the pre-trial survey data – in which staff sentiment and expectations were baselined – suggested that there were certain preconceptions around GenAI-generated output. A total of 65% of respondents for one trial expected either good- or high-quality output on the basis of reading about GenAI and little or moderate use of ChatGPT in their personal lives.

Interestingly, the expert panel correctly identified *all* outputs generated by our lawyers as human-generated. However, 50% of the output generated by GenAI was

either misidentified as human-produced output or our experts couldn’t tell whether the output was human- or AI-generated output. In 67% of cases, the accuracy score for the GenAI-generated output matched or exceeded the scoring for human-generated first draft output. No human-generated output was misidentified as having been created by GenAI.

We reviewed the expert commentary accompanying the study scoring to understand what had made some of the GenAI content virtually indistinguishable from human output and whether accuracy was the only obvious marker. More broadly, Ashurst wanted to understand if there were clear hallmarks of human- versus GenAI-produced work to help organisations distinguish between the two.

Based on the trials, it seems likely that a mix of language choice, tone and structure all provide indications of source – our experts specifically referenced language and writing techniques as clear indicators of human work. Writing techniques referenced as human included well-structured executive summaries and analysis of issues in addition to pitching vocabulary correctly. Commentary also focussed on how the author applied the judgment and whether broader legal implications were explored. Interestingly, inconsistencies in grammar and spelling were also called out as human traits too. This is a topic that Ashurst is particularly interested in and will be continuing to explore as we advance our GenAI research and development activities.

The study commentary was, however, useful in uncovering an initial set of expectations from reviewers of GenAI output as to what *quality* might mean in the context of legal work.

Quality in a legal context is multi-dimensional, with subjective and objective elements

While various studies have reported on the accuracy and speed of GenAI in performing certain legal-based tasks,ⁱⁱ few have specifically reported on the quality of GenAI outputs in a legal context.

In our trials, we tested whether the “quality” of GenAI content could be assessed based on the overall score across our assessment criteria. However, in seeking to understand the expert commentary accompanying our blind study scoring, it became clear that quality could not be reduced to a simple mathematical average of scores as these were ultimately influenced by a number of objective and subjective factors. Indeed, while Ashurst’s lawyers rigorously supervise, review and ensure excellent quality in the delivery of our legal services, attaching a numerical value to each specific piece of legal work is not an established way of working. Rather, they often intuit the quality based on their experience, expertise and preferences, none of which are easily replicable by any digital technology, including GenAI.

For example, our blind study expert panel were able to provide definitive scoring across our assessment criteria. However, in reviewing their commentary it became clear that each expert had their own additional preferences and sometimes bugbears when reviewing legal output. For example, one expert expressed frustration at inconsistent grammar/spelling, whereas another focused on how comprehensive the output was. By contrast, another focused on vocabulary and whether the language used was archaic.

Each expert was allocated a different case and so the variables of case complexities and subject matter could account for the differing commentary. However, the feedback indicates that determining quality can be very subjective, despite their placing a numerical value on it. For one expert, quality was absolute. In this instance, the GenAI hallucinated and cited an incorrect verdict. The expert immediately recognised this, dismissed the remaining output and applied critical scoring across the board (each criterion scoring 1 out of 5). Whereas, for other participants in the wider GenAI trials, minor mistakes were more forgivable, given the efficiencies gained:

i. **“Acceptable” errors:** In one experiment, a participant experienced *“the odd blip”* when using GenAI to interrogate a large internal policy document. This was referred to as *“an acceptable level of failure”* when interviewed, with the implication being that the error would be found and corrected upon review.

ii. **Still worth having:** In another experiment, a participant used GenAI to produce training material and extract key contractual elements from public body guidelines. The output was deemed unusable (scoring 2 out of 5 for usability) and saved no legal time; yet, the participant would still recommend using GenAI to a colleague on the basis of their overall experience of using it for other tasks and still found the relevant tool *“really impressive”*.

iii. **The redline’s fault, not the AI’s:** For another participant, GenAI was used to analyse a redline of evolving financial regulation. It scored averagely (3.2 out of 5), but the participant partly blamed the lower quality on the redline itself being *“not the best”* and would still strongly recommend using GenAI to a colleague.

The generally tolerant sentiment of the examples above was echoed more broadly in experiments where participants cited hallucinations and critiqued the tone of GenAI output, but 67% of post-trial survey respondents still recommended we onboard the associated tool for its overall perceived advantages, being holistic efficiency gains, a starting point to react to rather than a blank page and a “second pair of eyes”.

The common assessment criteria worked well in the trial setting to provide an initial indication of output quality, but commentary demonstrated an additional subjective dimension to determining quality that didn’t form part of the original scorecard.

Early evidence, therefore, indicates that quality is a multi-dimensional concept and one that needs further exploration. In a future where there is potential for lawyers to be using GenAI to create high-quality legal output, and the quality of outputs of potential GenAI tools needs to be evaluated and measured to inform an investment decision, organisations should consider quality by applying both numerical and subjective measures.

4

Beyond legal and client work, GenAI can make day-to-day life easier and help people feel more prepared for the future

The trials were instrumental in uncovering not just pockets of value in legal work and potential client-facing solutions, but also showing the potential impact of GenAI on our people's day-to-day work and lives. Interestingly, staff across the organisation, including legal and consulting personnel and business services professionals, reported ways in which GenAI made their daily tasks easier.

A key point participants cited was increased productivity in meetings. Previous research exercises indicated that meetings, calls and follow-up activities take up a significant proportion of staff time across the whole of the firm. In an open experiment, 21 participants used GenAI to summarise and recap meetings and to create action points arising from them. On average, participants reported circa 10 minutes saved per 30-minute call. The output received an average score of 3.8/5. Participants were confident in the output, noting it allowed them to be more present in meetings as their attention wasn't diverted by note-taking.

GenAI also made day-to-day tasks easier for some participants by providing a useful sense check for their work. For some, GenAI was considered a "second pair of eyes" or a "spare pair of hands". In one instance, the GenAI presented such a novel angle on a legal point in a court judgment that the participant refused to believe it could be correct (which it was). While for some GenAI became a verification tool, Ashurst's view remains that users need to proceed with caution. Participants reported frequent hallucinations across all tools trialled, which strongly supports the need to keep "the human in the loop".

Part of the trial objectives was to understand the level of training and support that staff would need in using GenAI. Interestingly, the trials uncovered a use for GenAI itself as an effective support mechanism for staff, with some participants voicing a hope that in a future where GenAI

is integral to how they approach their work this might free up some space and time in their day to engage in more learning and direct client-facing activity.

For example, 61% of post-trial survey respondents agreed that using the GenAI tool in question would help them feel more supported in managing their workload. Feedback from one-on-one sessions also uncovered a window into the working world of some of our lawyers, with one participant explaining, "sometimes [when] I need something done [when] no one else is online, it's nice to be able to quickly do something". GenAI provided consistent, desk-side assistance for the participant outside of regular business hours and it allowed them to move forward with their work, subject to checking the output with a colleague later.

The value uncovered for GenAI during the trials was therefore far broader than we originally anticipated. Tellingly, 88% of respondents at the end of the biggest trial said that using GenAI technology helped them to feel more prepared for the future.

Delivering and embedding a GenAI capability for Ashurst staff was thus about more than just saving time – it was about preparing them to meet and stay ahead of both client and market demands as they evolve.

The ‘jagged frontier’ of AI in the legal industry needs clarification to derive value

It has been proposed that the boundary between what GenAI is capable of doing well and what it is not capable of doing well (or at all) forms a “jagged technological frontier” that we must navigate carefully in order to derive value from it.ⁱⁱⁱ In particular, there is a danger in trying to leverage GenAI on the wrong side of the frontier because it can cause the human operator to perform worse than without GenAI.

This danger is particularly acute for the legal industry, where the main currency of lawyers is the actual and perceived quality of their work.

In designing the trials, we sought to explore the relevance of the jagged frontier metaphor by engaging with an extremely broad participant group – covering a range of roles, tenures and geographies – and undertaking a number of trial activities aimed at testing GenAI on as many different tasks and work types as possible within the trial parameters set. Through this approach, it became clear very quickly that the performance of GenAI for the legal profession was indeed quite “jagged”.

While the three trials were useful in exposing some of the “edges” of the frontier for the legal industry, it is our view that not only is the industry only at the start of its exploration journey, but merely identifying the boundary and understanding what sits on each side is not enough to derive value from GenAI. Collating lists of use cases is, on its own, not going to progress the industry or any individual organisation forward. Lawyers and legal professionals need to learn to truly engage with and use GenAI for their day-to-day work. This requires a sustained and focussed strategy and effort around embedding GenAI as a capability that the whole of the organisation can easily tap into and exploit to their benefit.

At the outset of the trials, we set a hypothesis that our people expect that using GenAI will be intuitive without the need for training or support. This was partially evidenced in the trial feedback. For example, when asked what they liked most about a specific GenAI tool, “ease of use” was the most commonly cited. However, ease of use is not directly correlated to the value to be derived from GenAI.

Putting GenAI technology and capability directly into the hands of our people, and setting them on completing legal work known to be within the frontier did not guarantee high-quality, legally correct outputs. In our case, the risk was less that the lawyer performed worse than they would had they done the work on their own, but that they would be frustrated and discouraged by the experience.

Over the course of the trials, it became clear that Ashurst needed not only to help our people to gain a real understanding of the impact and possibilities within their day-to-day tasks and the range of potential applications of GenAI to their work, but that we also needed to facilitate a continuous dialogue about their experience. It would not have been possible to accurately predict and then uncover the myriad of ways in which GenAI added value, or didn't add value in some cases, without both hands-on experience and continuous dialogue. It also, importantly and somewhat surprisingly, helped our people to better understand and articulate the value that they bring to certain types of work on their own.

Furthermore, GenAI opened the door to greater understanding of the digital literacy and development support needs of our people and what is required to achieve broader digital transformation at Ashurst. No technology exists or is used in isolation, especially in the workplace. By giving our people the opportunity to access and experience GenAI, we uncovered opportunities to upskill people in both technical and soft skills knowledge and training. By the end of the trials, we were able to form a clearer view of what additional learning and development would help to properly embed new digital capabilities across the firm. This did not cover just GenAI, but also a broader range of skills and training alongside a suite of enablement and support measures to be developed and delivered in partnership with our knowledge, expertise, and learning and development colleagues.

Ashurst trial design approach and methodology

While running technology trials was not a new exercise for Ashurst, running multiple, concurrent trials on a global scale had not been attempted before at Ashurst.

Previous trials (eg Ashurst's [recent implementation of business diagramming tool Jigsaw](#)) had primarily focussed on testing whether a particular tool met a singular set of user needs with a purposefully limited trial base of users. By contrast, the GenAI trials needed to explore and measure the impact of a potential new *capability*. The trials did involve the use of certain GenAI-powered technology tools, so we were simultaneously evaluating the tools themselves alongside seeking to understand and measure the value GenAI could bring to the whole of Ashurst's business.

Starting with a hypothesis

Before starting the trials, Ashurst developed a set of six working hypotheses around GenAI's potential impact within Ashurst. These hypotheses were set in the early stages of the trial design work and allowed us to scope the approach, frame research objectives and develop a set of common assessment criteria.

Next, we mapped these to six core areas of measurement to focus on: *accuracy; audience readiness; search speed; trust; value; and experience*. For each, an accompanying research objective was set, which formed the basis of how we sought to prove or disprove the relevant hypothesis.

Defining our hypotheses also lent a more-scientific lens to the trials which, when combined with our user-centred trial design approach, allowed Ashurst to collect significant amounts of quantitative and qualitative data around specific work being done within Ashurst. Having real-time, measurable data to substantiate the trial findings was invaluable for the CDO team when formulating recommendations and next steps for GenAI for the firm. However, it is important to stress that – given the sample size – much of the data cannot be simply extrapolated to get to definitive answers on the actual time and costs savings that the firm might achieve through the use of GenAI.

For this reason, Ashurst's view remains that the industry is still some years away from being able to determine or implement any necessary changes to law firm business models as a direct result of the adoption of GenAI.

A multi-modal approach

We captured quantitative and qualitative assessment data using the following approach, consisting of a blind study, controlled experiments and feedback surveys:

(i) The blind study

Ashurst ran a discrete study that involved a blind comparison of various UK case summaries prepared manually by Ashurst lawyers against the output from the GenAI tools being trialled. A panel of four expertise lawyers across three practice areas suggested four cases ranging in length and complexity. The human- and GenAI-produced output was anonymised prior to assessment by the expert panel so that there were no formatting clues as to the origin of the content. This part of our study's objectives was meant to:

1. analyse and score the output to provide a meaningful indication of GenAI's impact and value; and
2. collect initial data and insights on whether and how GenAI-generated output meets Ashurst lawyers' expectations.

To achieve Objective 1, the CDO team worked closely with the expert panel to develop common assessment criteria that could be used to quantitatively evaluate all current and future GenAI-produced output (see the [Appendix](#) for an overview of the assessment criteria).

When combined with qualitative feedback from the study's participants, evaluating output in this consistent, measured way enabled the CDO team to make strategic recommendations for the firm that were grounded in objective data. The study data also informed whether the hypotheses established the right areas for investigation.

(ii) Small group experiments

The trials were experiment driven with varying degrees of complexity, rigour and structure. All participants were encouraged to explore the GenAI tools being trialled independently and select individuals participated in experiments undertaken in controlled conditions. The CDO team co-designed, with participants, experiments that were tailored to how a specific legal task was performed in reality.

While some experiments were open to all regardless of practice area/function, the majority related to specific legal tasks from practice areas. Some were designed to find efficiencies in daily work, whereas others trialled new approaches and opportunities both for Ashurst and our clients. Approximately 25% of participants (from across a range of practice areas and business support functions) took part in at least one structured experiment in addition to their own "free play".

Ashurst's user-centred approach helped to uncover more than just measurable data. In one case, where GenAI ultimately failed to improve the quality and efficiency of the legal task being undertaken, our lawyers were able to clarify and better articulate their own value in what was seen to be not just a simple 'task', but rather a complex and nuanced exercise.

Also, and very importantly, the small group experiments allowed us to gain a better understanding of the digital literacy and development needs of staff and what would be required to achieve not just embedding of GenAI as a capability, but broader digital transformation at Ashurst. This understanding is now informing how Ashurst approaches our efforts to provide support and drive a digital-led culture across the firm.

Finally, exploring GenAI through experiments with concrete use cases helped make conversations around daily challenges and our lawyers' requirements less theoretical and more tangible. This had the added benefit of clearly demonstrating to the firm's leadership the enthusiasm for, and advantages of, investing in new technologies.

(iii) Creating a culture of experimentation: community forums; feedback sessions; and surveys

Achieving the aims of the trials could not be done solely through observation and data recording. Ashurst designed the trial environment to facilitate a continuous dialogue with participants about their experience. This helped to create an engaging, experimental culture where participants could share insights and tips, and troubleshoot issues. Within each of these trials, leads were appointed to nominate participants from their respective practice areas/functions. A wide range of geographies, tenures and roles across the firm was considered to broaden the data set.

While we had a notion of potential use cases for GenAI, there was consensus that the malleability of the technology meant it could be used in a myriad of ways. Fast-paced 'art of the possible' sessions were designed to create an environment for trial leads to explore the impact of GenAI during their day-to-day work and the range of potential applications in their area of expertise. These sessions were integral to our trials as they provided us with a bank of use cases from which to design the experiments.

Conducting the trial through a collaborative forum meant that participants could follow the trial's progress and lessons in real time while also having easy access to demonstrations and session recordings, briefing packs, educational material and trial guidelines. The forum also helped to create a community atmosphere.

The CDO team also worked closely with an internal governance group (including the firm's general counsel, risk and information security teams) to develop a set of strict guidelines for each trial. These acted as a 'guardrail' to protect against the risk of non-permitted data being used in the trial.

We also monitored participants' expectations, sentiments and experiences of GenAI before, during and after the trial. This was achieved through surveys and feedback sessions designed and facilitated by the CDO team. In particular, feedback indicated that imposing time limits on the trial and holding participants accountable with friendly competition and challenges increased engagement.

Testing GenAI through an experiential lens helped Ashurst to understand the approach that we should take to support our people in the most-effective ways as we adopt this new capability. Applying this approach also provided us with measurable data to substantiate the trial's findings.

Recommendations for running GenAI trials

We have set out below key lessons from running GenAI trials that we hope will help others to navigate their own GenAI exploration and will help them to understand how engaging in similar trials can provide improved experiences for their own people.

-
- 1. Consider the design of both trial activities and the trial environment** – Consider how to structure the trial not just to collect data to meet your research objectives but also to allow participants to engage in the easiest way possible. This could involve the use of collaboration forums, one-on-one discussions, mid-trial and post-trial feedback sessions and surveys.

 - 2. Ensure direct access for experimenting with GenAI** – Put GenAI in as many people's hands as possible directly (within appropriate safeguards). Consider the nature of the trial, the spread of people involved (include a mix of geographies, role and tenure) and then strategically allocate the number of licences available to get an appropriate range.

 - 3. Use trial leads to guide the process** – Appoint trial leads spread across multiple teams and with deep knowledge of your workforce to suggest the right people to trial the products. Ensure that you pay attention to creating a diverse group of people with different skills and levels of understanding of GenAI (ie not just people who are already enthusiastic users), and get representation from across all business functions. This approach will help to uncover potential challenges you may have in rolling out GenAI to the broader workforce.

 - 4. Run 'art of the possible' sessions** – People needed more support with understanding how to apply GenAI rather than technical support on the tools themselves. Provide educational packs ahead of interactive sessions that help people to think of ideas for using GenAI in their practice area and daily tasks. Include quick demonstrations of the specific features to keep the session fast paced. Use the feedback to design specific experiments you might wish to run.

 - 5. Hold people to account and create competition** – Shorter, more time-bound trials seemed to create more of a buzz and focus than longer, more open-ended trials. Creating friendly competition amongst teams and offices by sharing information about usage statistics and designing challenges seemed to really motivate lawyers who, as it turned out, were very competitive. If people are not using it, then find out the reason why and consider reallocating licences to others on waiting lists.
-

In conclusion: “My eyes are more open”

Ashurst’s trials approach is just one of many ways the legal profession is seeking to understand and capture the value that GenAI may bring to bear. Ultimately, each organisation needs to take the approach that works best for them and their people. By putting our people at the centre of the trials, we were able to move them past the hype to real experience and value with one participant remarking “*my eyes are more open*” for having been part of the trials. This sense of understanding the realistic uses and limitations of GenAI was an important finding of the trials for many of the participants.

Fortunately, the approach has paid us dividends. It allowed the firm to collate a mass of qualitative and quantitative data from which further studies and investigations can spring. The trials conducted so far might not answer all of the questions we had in November 2023 – in fact they’ve generated more questions than answers – but they have provided valuable insights on what Ashurst’s initial approach should look like. This has ultimately allowed us to make investment decisions based on measurable data and to move forward with designing our own GenAI policies and implementing its use in our firm.

Critical to the success of the trials was the level of engagement and enthusiasm that staff demonstrated for the trials, which far exceeded Ashurst’s initial hopes. Participants had a willingness to experiment, learn, and in some instances fail, on a scale we haven’t experienced previously. GenAI was not the answer to all challenges experienced across the organisation, but it opened the door to a better understanding of pain points, user needs and our lawyers’ daily experience.

A key trial lesson we learned was that an experienced approach was crucial in delivering outcomes that would resonate with staff. The key to benefitting from GenAI is understanding the users within the organisation and how they would want to use GenAI, not just understanding what the technology can actually do. Ashurst is not implementing GenAI with an “if you build it they will come” mindset. The firm has made its people the essential stakeholders in the process while also designing a framework to be able to continually assess and reassess the tools available in the market and to act on opportunities that will differentiate us in the legal marketplace.

There is much more testing to be done to understand how Ashurst can leverage GenAI for more-complex work, as well as ways to improve our approach to onboarding new technological capabilities. This report has attempted to lift the lid on our methodology and findings so far, rather than presenting something definitive and complete. Moving forward, sharing and collaborating in this way will be crucial to navigating the noise around GenAI and promoting meaningful, beneficial change in the legal industry.

Appendix

Common Assessment Criteria

Our criteria had to balance being detailed enough to be constructive, but without being too prescriptive or onerous on the expert panel and trial participants, and also generic enough to apply across different GenAI tools as well as human-produced output. The scoring produced by the trial criteria also had to be meaningful enough to allow Ashurst to determine whether we had proved or disproved any of our original working hypotheses.

The common assessment criteria worked well in the trial setting to provide an initial indication of output quality, but user and expert commentary demonstrated an additional subjective dimension to determining quality which didn't form part of the original scorecard.

Criteria	Guidance provided to expert panel	Score
Accuracy	<i>When answering this, think about whether it is correct (including, where relevant, legally correct), precise, misguided and/or contradictory.</i>	/5
Appropriateness	<i>When answering this, think about whether it is relevant to the task (both in content and tone), is right for the intended audience and contains the right level of detail.</i>	/5
Completeness	<i>When answering this, think about whether it covers all of the required content, retrieves and makes connections between relevant information and meets the scope of the task.</i>	/5
Usability	<i>When answering this, think about whether it requires reformatting/tidying, is well structured and is fit for purpose.</i>	/5
Confidence	<i>When answering this, think about whether it is shareable with others (internally and/or externally), something that takes you forward in your work and is something you feel is representative of the quality of work expected of you/Ashurst.</i>	/5
Average score	<i>Each of the above criteria were marked out of five. Each of the five scores were then averaged to produce an average output score.</i>	/5

Acknowledgements

In addition to thanking everyone from all of the Ashurst offices who took part in these GenAI trials, Ashurst would like to acknowledge the following members of the Office of the Chief Digital Officer for their invaluable contributions to the report: Sarah Chambers; Sophia Slade; Richard Keaney; and Chris Boulter.

Endnotes

- i See for example, Linklaters LLP (2023, 30 Oct), *The LinksAI English Law Benchmark*. Available at: <https://techinsights.linklaters.com/post/102irc3/the-linksaienglish-law-benchmark-can-you-rely-on-ai-for-legal-advice#:~:text=The%20benchmark%20comprises%2050%20questions.The%20questions%20are%20hard>
- ii See for example, Lauren Martin, Nick Whitehouse, Stephanie You, Lizzie Catterson, Rivindu Perera, AI Centre of Excellence, Onit Inc., New Zealand. *Better Call GPT, Comparing Large Language Models Against Lawyers*. Available at: <https://arxiv.org/pdf/2401.16212.pdf>
- iii Dell'Acqua, Fabrizio and McFowland III, Edward and Mollick, Ethan R. and Lifshitz-Assaf, Hila and Kellogg, Katherine and Rajendran, Saran and Kraymer, Lisa and Candelon, François and Lakhani, Karim R., *Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality* (September 15, 2023). Harvard Business School Technology & Operations Mgt. Unit Working Paper No. 24-013. Available at SSRN: <https://ssrn.com/abstract=4573321>

Contacts



Tara Waters
Partner, London
Chief Digital Officer
T +44 20 7859 2755
tara.waters@ashurst.com



Hilary Goodier
Partner, Melbourne
Global Head of Ashurst Advance
T +61 3 9679 3095
hilary.goodier@ashurst.com

Authors



Tara Waters
Partner, London
Chief Digital Officer
T +44 20 7859 2755
tara.waters@ashurst.com



Sarah Chambers
Director, London
Digital Experience
T +44 20 7859 2394
sarah.chambers@ashurst.com



Sophia Slade
Senior Manager, London
Digital Experience
T +44 20 7859 3981
sophia.slade@ashurst.com

Ashurst