

Spotlight AI case study

A look into the impact AI can have on eDiscovery efficiency.



Overview

This report is based on results from a test intended to benchmark Spotlight AI against human reviewers. Hanzo's award-winning Spotlight AI is an automated data relevancy engine that helps legal teams navigate vast datasets for faster, smarter, better-informed decision-making. Spotlight AI is the first Legal Tech product to solve the challenge of automating collaboration data review using innovative AI.

The data was based on Slack messages for an employment discrimination lawsuit and consisted of messages in both public and private channels. Collaboration data like this is especially difficult and slow to review because most short messages are brief, informal, and with multiple different topics within a conversation. For this particular matter, unlike the human review data set, Spotlight AI was run across a data set that did **not** include keywords.

Case details

This matter was an employment discrimination case. The original matter consisted of over **8.8 million** messages across **21 custodians** and **47,110 channels**. . Using traditional date culling, the original data set was reduced to **2.1 million** messages which Spotlight AI was then run across.

For human review, keywords were applied, further reducing the message population to **513,469 messages**. After deduplication, the total final message population was **499,495 messages**.

The traditional human review was performed by batching the messages out into review batches and performing a first-pass review for matter relevance with traditional relevant/non-relevant tagging (and issue tags).

Spotlight AI was performed independently within Hanzo Illuminate, where the originating data resides. The client, outside counsel, and review team agreed upon a case description, which was copied into Illuminate, whereby the Spotlight AI process was run end-to-end.

Executive summary

Spotlight AI vs. human review for relevancy assessment

Key findings

Speed:

Hanzo's Spotlight AI had a turnaround time of approximately 12 hours for the original dataset with 40 questions asked. Traditional first pass review was 191.5 hours for human reviewers. Spotlight AI was **16 times faster** to perform this process.

Conclusion:

Hanzo is significantly faster than human review, with a processing speed of over 166,000 messages per hour. *The time required will always depend on the complexity and volume of a matter.*

Accuracy:

Hanzo's relevancy recall is **more than 82%**, which is above the industry-accepted 70%-80%¹. This achievement is particularly notable given the fragmented nature of Slack conversations, which makes it challenging to delineate the start and end of discussions accurately, as well as the variation in the methodology used for the manual review.

Culling:

The ability to effectively cull data is crucial, especially given the low percentage of responsive messages (1.63%). The primary goal is to safely exclude as many non-relevant messages as possible from manual review. In this context, **precision is less critical** because even a **low precision rate allows for significant data reduction**. Using the needle-in-a-haystack analogy, providing 20 items with only one relevant needle (5% precision) is still highly beneficial as it reduces the need to sift through the entire haystack. The same principle applies here. Hanzo culling achieved 57% with 82% recall, which means 57% of the messages could be safely removed from manual review.

Cost savings:

At an average of \$45/hour for human review, this matter would cost over \$8,600 in manual review. Using Spotlight AI to assess relevancy before any manual review would have **saved 57%** of that, approximately \$4900. The cost incurred to run Spotlight AI in this instance would be approximately \$300, which entails a **net savings of \$4600** out of the \$8600.

¹ In re Diisocyanates Antitrust Litig, the court considered the acceptable recall rates for TAR. The parties agreed that a recall rate between 70% and 80% was generally acceptable. In Lawson v. Spirit AeroSystems, Inc., the court addressed the use of Technology Assisted Review (TAR) in discovery. The case held that an 85% recall rate was reasonable and typical for TAR, and the defendant agreed to an 80% recall rate after initial results showed less than 70% recall.



Conclusion

Spotlight AI shows significant promise in efficiently and accurately assessing relevancy, far **surpassing the speed of human** review while maintaining high precision. Hanzo's approach results in **substantial cost savings**, with direct savings exceeding 57% of a manual review budget, even when accounting for the cost of running Spotlight AI.

When combined with a hybrid workflow that utilizes keywords and traditional filters such as channel types, custodians, and dates, this method already allows for the **rapid identification of relevant content** while significantly reducing the volume of irrelevant data typically reviewed in a traditional process.

Data set statistics

Process time details

Hanzo is currently processing complex facets (presented to the user as questions) across 1,000,000 (1M) messages in approximately 6 hours. Scaling up to the originating data set of 2.1 million messages using 40 facets, Hanzo has a **turnaround time of ~12 hours**, given the cluster scale used for the matter. As Hanzo continues to optimize Spotlight AI, we anticipate that processing time will continue to become shorter.

Human review time for this matter was recorded at **~191.5 hours**. Based on the time and setup of this matter and Spotlight AI enhancement run, Hanzo is currently **16 times faster** at evaluating content for relevancy than human review. Advantages to this process are numerous and include the ability for LLM's to run 24 hours a day without human distractions or bias.

| Data input type | Human review message population After traditional culling via keyword, date, etc. | Human review message population After deduplication | Number of relevant messages | Total relevant Percentage | Time to process 20 questions |
|-----------------|--|--|-----------------------------|---------------------------|------------------------------|
| Slack | 513,469 | 499,495 | 8,138 | 1.63% | ~ 6 hours (.25 days) |

Recall/rejection results

In order to calculate Recall and Culling or Rejection, Hanzo creates "Documents" consisting of grouped messages. Extensive data science optimization in this area has occurred and continues. Additionally, LLM capabilities continue to improve allowing for larger context windows and better throughput.

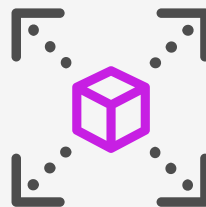
| Document size | Message recall | Document recall | Rejection |
|--------------------------|---------------------|-------------------|-----------------------|
| 85 Messages per document | 7,049/8,008 = 88.0% | 386 / 469 = 82.3% | 3,121 / 5,409 = 57.7% |

Spotlight observations & takeaways



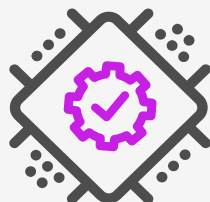
Broad contextual analysis offers efficiency and reliability.

Hanzo's Spotlight AI identified 386 of the 469, (82.3%) of the "documents (groups of messages)" quickly (**16x faster** than traditional methods) and without using keywords when performing its relevancy assessment demonstrating the effectiveness of broad contextual analysis.



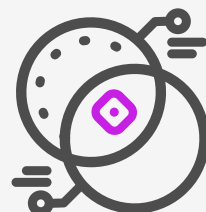
Term identities are important for identification.

Entity definitions are critical (like user names, aliases, etc.). For instance, Suzy Queue appeared as "Suzy," "Queue," or "SQ." This type of information is crucial for the system to identify relevant messages accurately.



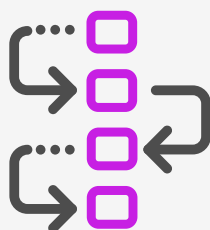
Optimized "document" structure is effective for Recall and Culling.

Hanzo's "document" creation consists of varying numbers of messages to measure recall and culling effectiveness. The team tested for the best balance between maximizing data culling and maintaining high recall accuracy.



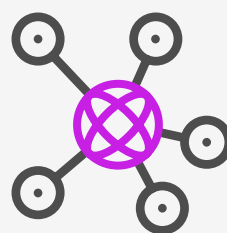
Internal system terminology can affect results, so include it in the questions.

Internal system names and business slang require additional attention and should be identified where possible in the Spotlight AI questions to ensure maximum efficiency.



Question structure is important

It is important to note that both the case description as well as the structure of questions is important to the process. While we developed the system to be user friendly, it is important to understand that the way you structure your questions and case descriptions can affect the relevancy outcome.



Traditional culling methods can be valuable for fine-tuning.

As mentioned, the human-reviewed dataset was culled using traditional methods like keywords and date filters. It is important to note that Spotlight AI **was not** prompted to search for specific keywords when performing its relevancy assessment.

Addendum

Relevancy detection methodology & details summary

- **Spotlight AI review:** Hanzo's Spotlight AI analyzed 40 reviewer-generated questions for responsiveness across the dataset of messages.
- **Relevancy labels:** Messages deemed as relevant or potentially relevant are labeled (Relevant or Potentially Relevant) based on their relevance to the questions. "Potentially Responsive" indicates messages near (within 10–20 messages) of relevant content. Messages that are neither relevant or potentially relevant are not labeled within Illuminate.

Label examples:

Consider a case where David Jones is accused of crashing an Audi in the company parking lot on Tuesday, June 25th. A Spotlight AI question may be, "David Jones is known as a David, Dave, Jones, and DJ. Did Jones crash the car on Tuesday?"

⊗ Not relevant message example:

Alice:
What time did you get in today?

Bob:
I was running late, so 9:15.

Alice:
Crap, we have a call in 5 minutes!

Bob:
K. I'll be there right away.

Bob:
See you for coffee later?

☑ Relevant example:

One of the messages in the following conversation would be marked as **relevant**:

Alice:
I'm doing an inventory on the cars.

Bob:
'First time since last year, right?'

Alice:
Sure, but I'm short one car.

Bob:
Well I'm sure you heard David Jones cashed that Audi on Tuesday.

Alice:
Right. Ok, thanks for the info.

🔍 Potentially relevant:

In the following conversation, relevance cannot be constrained to a single message, so all messages get marked as Potentially Relevant:

Alice:
Hey Bob, I need to know where that Audi is, I'm doing inventory.

Bob:
It's been out of action since Tuesday.

Alice:
Ok, I see DJ had the car that day. What happened?

Bob:
He managed to crash it in the parking lot near the cafeteria..

Alice:
Ouch. Lots of paperwork!

Contact us
today

Sales
contact@hanzo.co