

UCAI '21: Workshop on User-Centered Artificial Intelligence
Ingolstadt, Germany, 5th September 2021

Noise over Fear of Missing Out

Legal professionals prefer recall over precision
for AI-assisted legal information extraction tasks

Johannes Schleith, Nina Hristozova,
Brian Cechmanek, Carolyn Bussey,
Leszek Michalak
Thomson Reuters Labs

Context

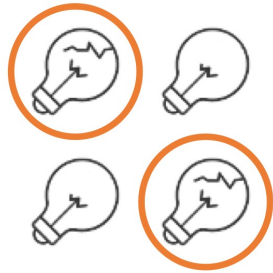
User Experience Metrics \leftrightarrow Data Science Metrics

Research Question

For AI-assisted legal information extraction tasks
what is the right balance between **Precision and Recall**
from an end-users point of view?*

Precision and Recall

Task: Identify all faulty light bulbs¹



"All information"

- Perfect Precision
- Perfect Recall



"Too little information"

- High Precision = All bulbs that have been found are faulty
- Low Recall = Not all faulty bulbs have been found

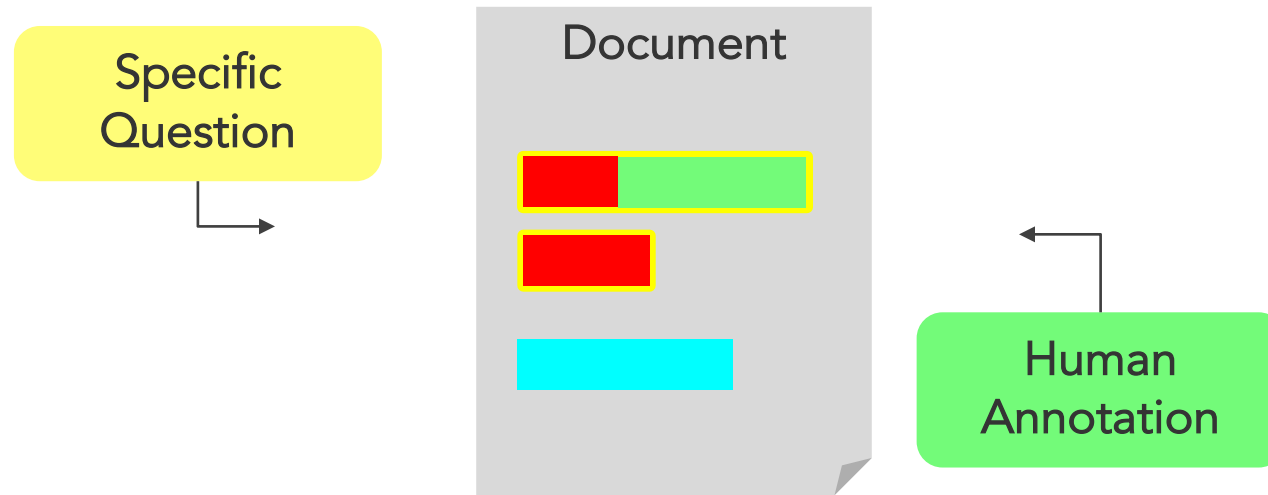


"Too much information"

- Low Precision = Not all bulbs that have been found are faulty
- High Recall = All faulty bulbs have been found

¹ Faulty Bulb Analogy by Laura Skylaki

Method



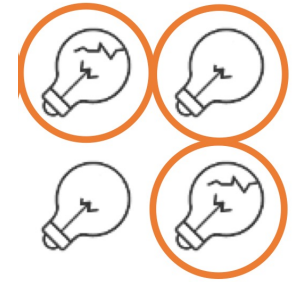
System & Annotation Task

What is the
landlord's address?



COMMERCIAL LEASE

THIS LEASE AGREEMENT is made and entered into on December 1, 2012, by and between **Temple CB, LLC**, whose address is **4350 Temple City Boulevard, El Monte, California 91731** (hereinafter referred to as "Landlord"), and Okra Energy, Inc., whose address is **4350 Temple City Boulevard, El Monte, California 91731** (hereinafter referred to as "Tenant").



*"Too much
information"*



Human Annotation

Example Source: Public SEC filings

System & Annotation Task

What are the repair obligations?



ARTICLE VII - CONSTRUCTION

[...] neither Tenant nor any third party may construe the permission granted Tenant hereunder to create any responsibility on the part of the Landlord to pay for any improvements, alterations or repairs occasioned by the Tenant.

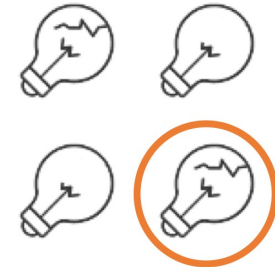
ARTICLE VIII - OBLIGATIONS FOR REPAIRS

Section 1. Landlord's Repairs. [...] the Landlord shall be required to repair only defects, deficiencies, deviations or failures of materials or workmanship in the building. [...]

Section 2. Tenant's Repairs. The Tenant shall repair and maintain the Leased Premises in good order and condition, except for reasonable wear and tear, [...]

ARTICLE XVI – DEFAULT

[...] payment of any costs and expenses of such reletting, including, without limitation, brokerage fees and attorney's fees and costs of alterations and repairs;



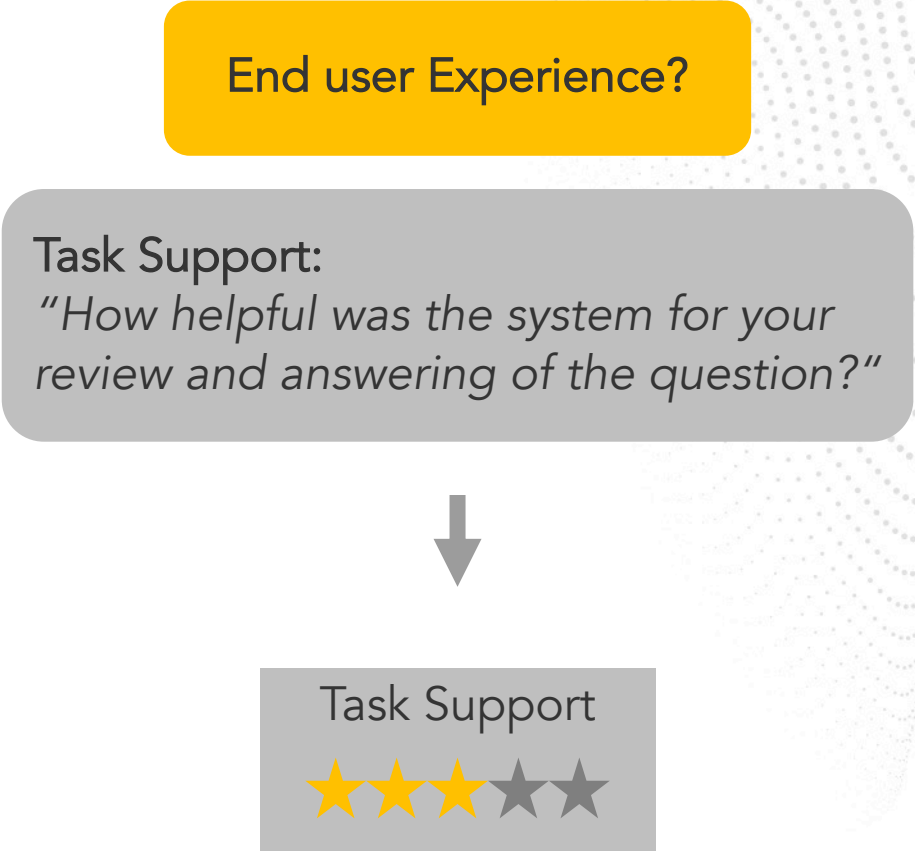
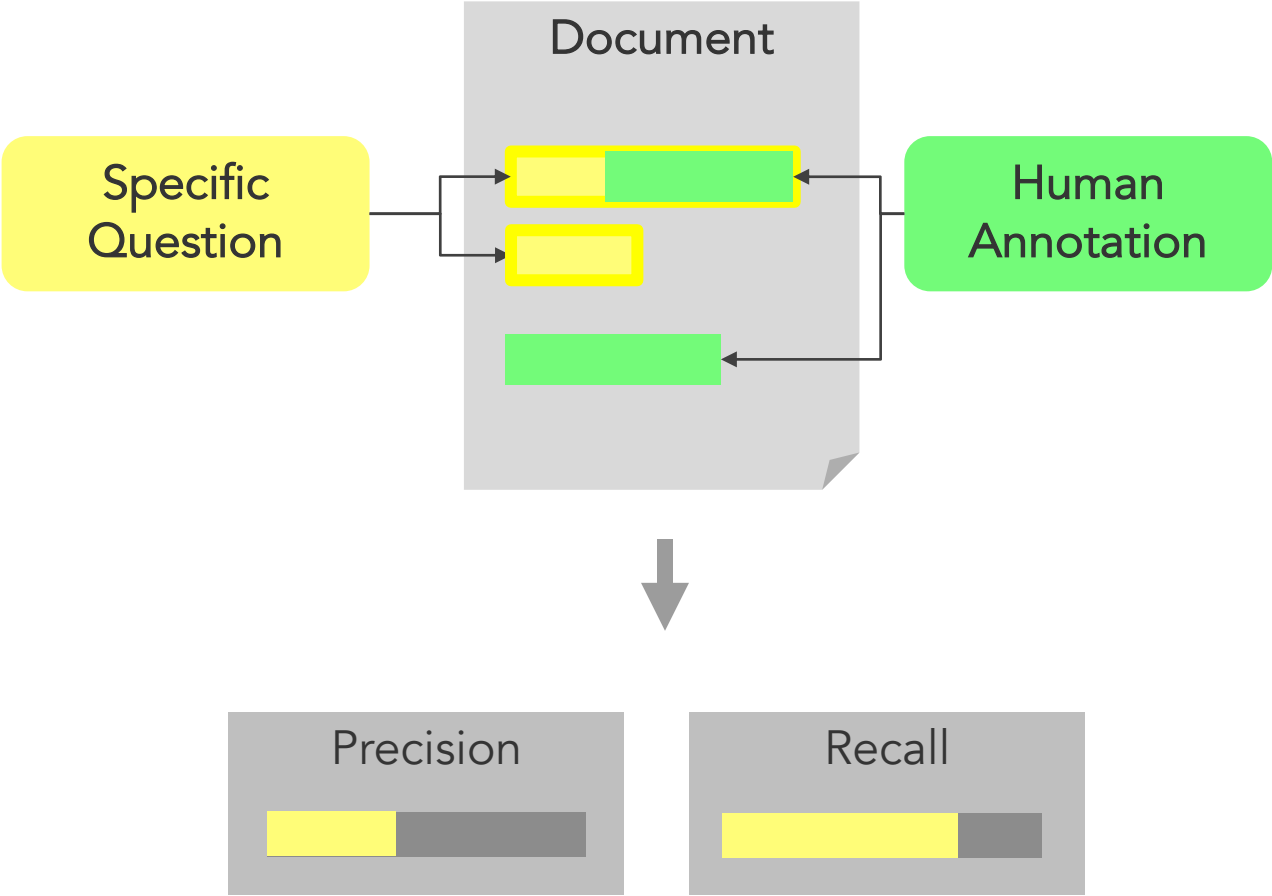
"Too little information"



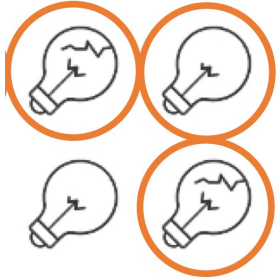
Human Annotation

Example Source: Public SEC filings

Metrics



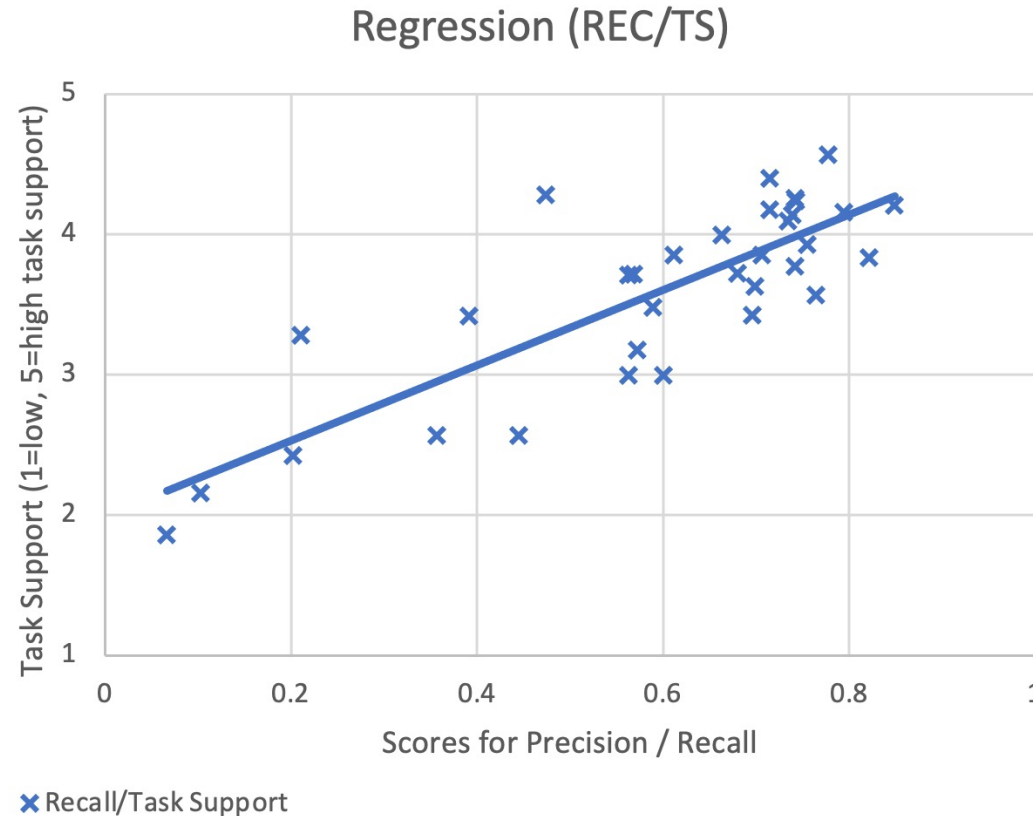
Results



Preference for “too much information”

Recall significantly predicted Task Support ($B = 2.68, p < 0.01$)*.

Specific Questions with high Recall were perceived as “more helpful” than questions with low Recall.



Completeness

Given the novelty of extraction methods in legal, participants felt a need for ‘complete coverage’. They often preferred to manually double check everything (6 of 10)**.

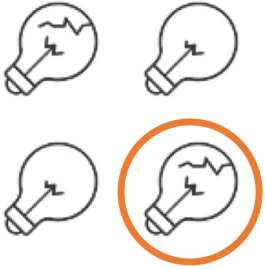
Example Quotes

“I would still want to comb through the document. I'd have to do that to consider myself a ‘competent lawyer’ ”

* Recall, Precision and Task Support scores for 33 questions, based on 95 reviews by 20 participants

** Qualitative insight from 10 follow up interviews.

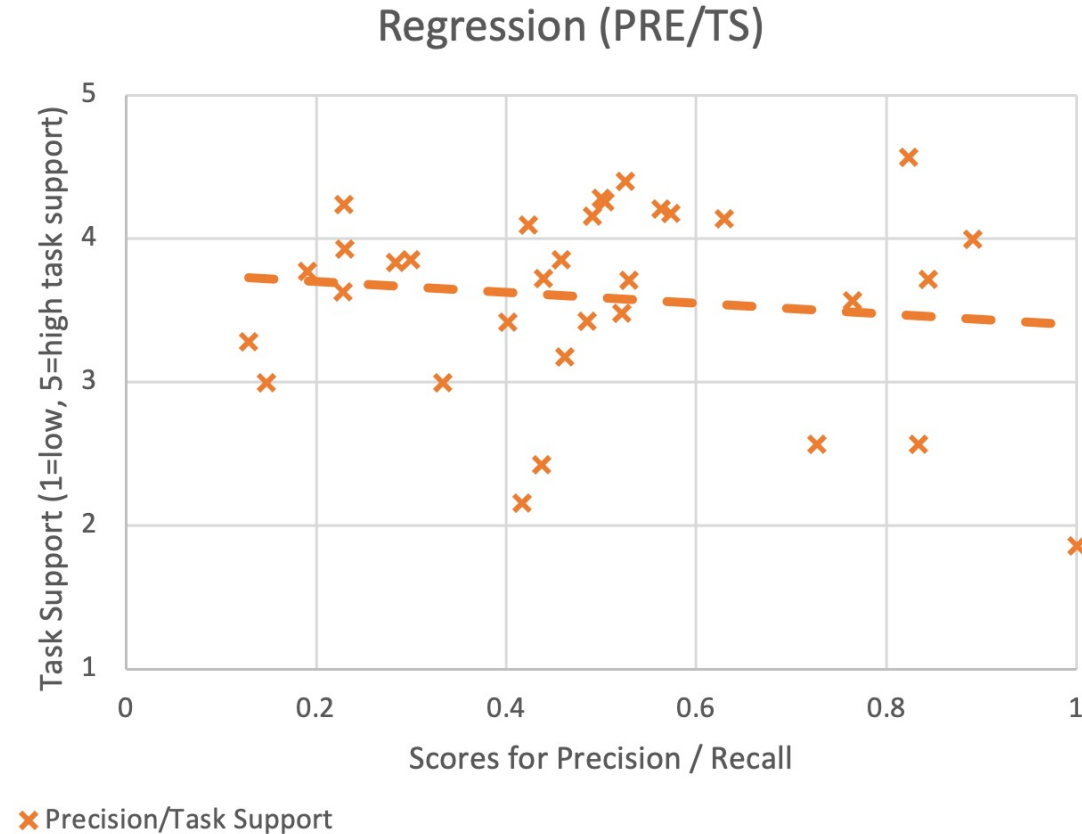
Results



No preference for
“too little information”

Precision did not significantly
predict Task Support ($B = -0.01, p = 0.97$)*.

Low or high Precision did not
have an impact on Task
Support.



Fear of Missing Out

Participants were **afraid of missing something**. When extractions methods did not yield any relevant text, just showing ‘no results’ made them anxious (6 of 10 participants)**.

Example Quotes

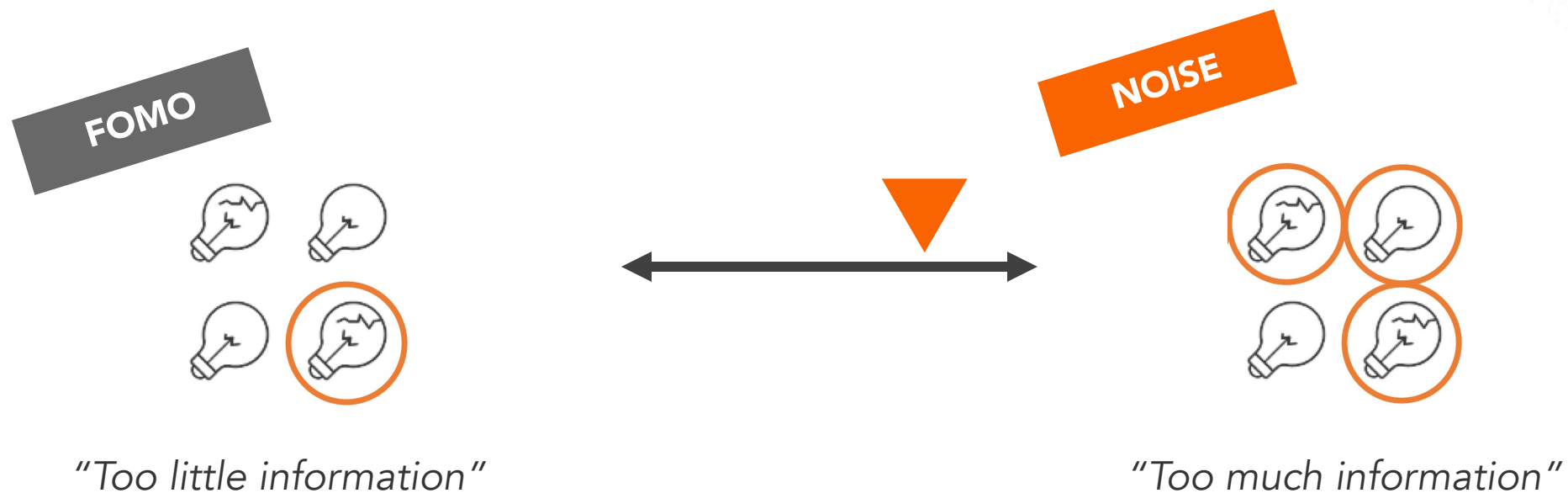
“I’m scared that it’s missing something.”

* Recall, Precision and Task Support scores for 33 questions, based on 95 reviews by 20 participants

** Qualitative insight from 10 follow up interviews.

Discussion

Legal professionals prefer Recall over Precision
for AI-assisted legal information extraction tasks



Discussion

Human-AI Collaboration

Participants were positive about positioning the system as **a review assistant** – rather than automation.

Quotes

"It could be helpful [to] have the software do a full first review"

Explainability

Participants **misinterpreted** the system and lacked explanation

Quotes:

I don't understand why information is highlighted or extracted (6 of 10)

I think it's learning from my inputs (2 of 10)

I think it's working like Boolean/Regex search (2 of 10)

UCAI '21: Workshop on User-Centered Artificial Intelligence
Ingolstadt, Germany, 5th September 2021

Thank you!
Questions?

Johannes.schleith@tr.com
@jshlth

Johannes Schleith, Nina Hristozova,
Brian Cechmanek, Carolyn Bussey,
Leszek Michalak
Thomson Reuters Labs