

# Ideation Deliverable

**Team:** I1

**Case title:** City of Amsterdam - CiteConnect: Linking Legal Advice to Case Law Using RAG

**Date:** 31 October 2025

**Members:** Shreya Aggarwal, B Caissotti di Chiusano, Matas Jurkevičius, Xue Wang

**FigJam link:** <https://www.figma.com/board/zgti0ju2pVS31xNw90kHxr/DSP-Workshop---Group-I---Project-1?node-id=3-125&p=f&t=JIFv436zrVIvmWl5-0>

# 1 Core Report

## 1.1 Reframed Problem Statement

Municipal legal advisors are required to reference case law when drafting legal advice letters, using European Case Law Identifier (ECLI) citations. However, retrieving relevant cases is often slow and inefficient due to poor metadata, inconsistent keywording, and limited search functionalities in current databases. This project aims to apply Retrieval-Augmented Generation (RAG) to enhance the speed, reliability, and relevance of ECLI retrieval.

## 1.2 Research Question

To what extent does RAG help legal advisors efficiently retrieve relevant ECLI cases without compromising the accuracy of citations?

## 1.3 Primary Persona and Empathy Insights

### Persona

Maggie, 25, is recently graduated and works in a high-pressure environment with tight deadlines and high expectations. She wants a support system that can save her time, streamline her workflow, and introduce new perspectives. She currently relies on manual citation, a time consuming process which does not always lead to finding the most relevant cases.

### Empathy Insights

- What the user **thinks**: Will I save time or end up wasting time with the tool? Will I be able to retrieve all the relevant citations?
- What the user **feels**: Relieved when the tool reduces manual effort, and is frustrated when the output is inconsistent.
- What the user **says**: I want a winnable case and the most relevant citation.
- What the user **does**: I already have a legal advice letter, I go through all the arguments, and find citations for them.

## 1.4 Ecosystem Highlights

- Users must work under pressure, and navigate unfamiliar citations, influenced by their tech literacy and habits.
- The system should deliver formal, unbiased outputs, aligned with the tone of a legal office, and adaptable to different users and preferences.
- The interface must be professional, transparent, and easy to use, supporting accurate and confident decision-making.

## 1.5 Idea Generation Summary

The Crazy Eights brainstorming method was used to gather ideas for the project scope. Ideas were then summarized and grouped into 3 clusters: Core App Features, Personalization, and Usability. **Core App Features** include "RAG suggestions" which rank retrieved ECLI citations, "Co-Citation Hints" display cases that are commonly cited together, "Import & Export" to upload/download advice letters with inserted citations, a "Chatbot" section for explainability on the retrieved cases, and a "Citation Checker" for accuracy. Under **Personalisation**, users can add missing citations with the "Manual Add" and access their "Library" with frequently used ECLIs. Finally, **Usability** includes an "Accessible UI" with keyboard shortcuts and light/dark themes, and a "Help" section for explanations and instructions.

## 1.6 Decision and Justification

As shown in by the average scores in Table 1.1, **Core features** scored the highest, as they satisfy the project brief by reducing the time and effort required to cite relevant cases using RAG. The explainability offered by introducing a Chatbot keeps the users in control and makes the product understandable. **Usability** features ensure that the system remains ethical and safe. Finally, even though **Personalization** features reflect user control from the empathy map, they remain less relevant for the project brief and therefore will have lower priority during the implementation process. Table 4 in Appendix B displays the complete scoring for each idea.

Idea	Fit brief	Empathy fit	Ecosystem fit	Principles fit	Practicality	Avg
Personalization	3.3	4.4	4.2	4.1	3.9	4
Core Features	4.7	4.9	4.5	4.3	4.5	4.6
Usability	4.0	4.4	4.0	4.4	4.4	4.3

Table 1.1: Scoring of ideas (1 low, 5 high)

## 1.7 Chosen Concept Direction

The concept sketch (Figure 1) in Appendix B shows how the proposed solution can help legal advisors retrieve relevant ECLI citations within their drafting workflow. The storyboard reflects the user's frustration of the slow, repetitive, and cognitively heavy process of citation gathering and evaluation, especially under time pressure, which turns to relief after efficiently adding references to the legal advice letter using RAG. It complies with the municipal ecosystem by using existing documents and leveraging on current systems and processes, rather than replacing them. The design follows our principles by keeping humans in control (manual confirmation, retry) and using a simple and understandable user interface to reduce cognitive load.

## 1.8 Risks, Assumptions, Next Checks

- **Language risk:** all data and output must be in Dutch. *Next Check:* test a small Dutch example with the RAG pipeline and ask a municipal lawyer to validate correctness.
- **Metadata quality:** inconsistent or missing ECLI metadata may break retrieval. *Next check:* run exploratory profiling on the ECLI dataset and log patterns of missingness & inconsistency.
- **Legal adoption risk:** advisors might not trust AI suggestions without explainability. *Next check:* include rationale fields (summary, verdict, law) in the prototype and validate trust with user interviews.

## 1.9 Inclusion and Ethics Checklist

- **Aligned with personas and the reframed brief:** Yes, the solution addresses municipal legal advisors' workflow when drafting ECLI-based letters.
- **Understandable, reduces cognitive load:** Yes, suggestions appear inside the drafting process, so users do not need to switch tools or perform manual searching.
- **Accessibility and equity considerations are explicit:** Yes, the system displays ECLI numbers with short summaries and relevance scores in a clear, text-based interface after the user uploads a draft.
- **Avoids harmful patterns and biased assumptions:** Yes, the tool only presents case summaries and ECLI numbers; users decide which to include, ensuring neutrality.
- **Consent and privacy constraints recorded if user data is involved:** Yes, we work with anonymized documents.

## Appendix A: Roles and Responsibilities

Responsibility or task	Owner(s)	Due date	Evidence location
Reframed problem draft	Xue, Shreya	26/09/2025	Section 1.1
Persona and empathy update	Shreya, B, Xue	29/10/2025	Section 1.3
Ecosystem highlights	Shreya	2/10/2025	Section 1.4
Crazy Eights clustering export	B	10/10/2025	Section 1.5
Scoring table snapshot	B	10/10/2025	Section 1.6
Concept sketch and labels	Matas	10/10/2025	Section 1.7 and Figure 1
Inclusion and ethics checklist	Xue	29/10/2025	Section 1.9
Final two-page PDF assembly	Xue, B, Matas, Shreya	31/10/2025	

## Appendix B: Evidence Figures and Tables

Section	Details
Name	Frank van de Lent
Biography	Frank is a 47-year-old legal professional in Gemeente Amsterdam. He specializes in administrative law and regulations, is experienced in drafting legal advice letters, and supervises juniors. Frank is cautious with AI tools and spends extra time verifying results.
Motivations	Frank noticed some juniors work efficiently with AI tools. He wants to save time searching and is looking for a more efficient ECLI search engine.
Frustrations	AI-generated results are sometimes untrustworthy. He does not want to spend too much time learning a tool that is not proven efficient or accurate.
Empathy: Thoughts	“Will I spend even more time verifying AI-generated results?” “How trustworthy are these results?”
Empathy: Actions	Draft legal advice letters, search on ECLI websites, guide juniors on ECLI citation, and identify the most relevant ECLI cases.
Empathy: Sayings	“I know people use AI a lot, but I find AI-generated results are not reliable in my work.” “If I need to spend more time learning this tool and verifying results, why use it at all?”
Empathy: Feelings	Untrustworthy, but looking for an efficient way to handle ECLI citation.

Table 2: Persona and Empathy Highlights: Frank van de Lent

Dataset	Size	Scope	Key Fields
Advice letters	567	Bicycle-towing objections	letter_id, zaaktype, onderwerp, dictum, anonymized_text, unique_words, cited_ecli
ECLI cases	2,447	Dutch court rulings	ecli, land, rechtbank, jaar, code, full_text, length

Table 3: Provided datasets

Category	Idea	Fit brief	Empathy fit	Ecosystem fit	Principles fit	Practical ity	Avg
Core Features	RAG Suggestions	5	5	5	4	4	4.6
	Co-Citation Hints	4	5	4	5	4	4.4
	Import & Export	5	4	4	4	5	4.4
	Chatbot	3	5	5	5	5	4.6
	Citation Checker	5	4	4	5	4	4.4
Personalisation	Manual Add	4	4	4	4	4	4
	My Library	3	4	4	4	4	3.8
Usability	Accessible UI	4	5	4	5	4	4.4
	Help / Guidance	4	4	4	4	4	4

Table 4: Scoring of all ideas for each category



Figure 1: Concept Sketch

Workshop 1

Team leader: S. Aggarwal

Clients : B., Xue

Interviewer : Mates (interviewer)  
(Designer) Shreyas (notes)

- ① target audience → COA lawyers
- ② security aspect → law/powers (separate)
- ③ what kind of product (→ clear or not)
- ④ security issues of data (separate) → safe
- ⑤ (→) linking + doing both or just running

**Clients and Interviewers**

- ① most important : Accuracy . Fast . Easy to understand. Instructions | B., Xue
- ② Outcomes : Law citation + ECLI citation  
if two ECLI cases are contradict, raise a flag
- ③ Which types : Website? App?
- ④ Priorities : User can choose which one is most important  
relevance? new cases? own/lose cases?

④ Interview-stakeholder (15 top) (main name, Kait) → AI specialist from main org.

① interested in ECLI - every week/month  
② anything will be provided (start with small dataset). 2-3 diff subjects  
③ mostly English (multilingual LLMs) (try to translate), English.  
④ maybe more other data we need to find sources  
⑤ just more data will be provided → many English cases  
⑥ data AI+X, some other might be good for batch, but translate to English first  
⑦ app user since/prompt - which can be jurisdiction for bc work  
⑧ user friendly, streamlit has pleasant controls  
⑨ can be for admin, tellers, data opportunities, function - some info of some file  
⑩ legacy, internal tool out of now  
⑪ retrieval - not just numbers, but summary or link/features

Figure 2: Stakeholders, Clients and Interviews

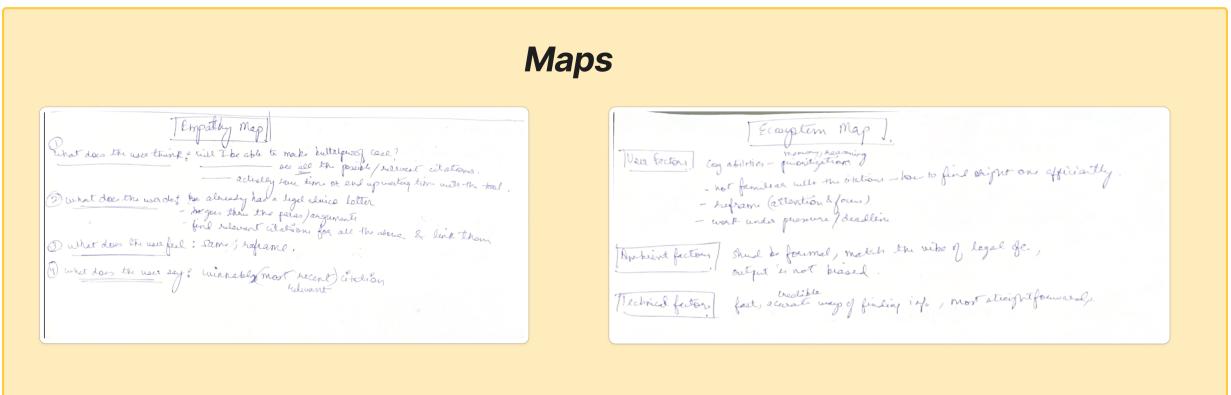


Figure 3: Draft: Empathy / Ecosystem Map

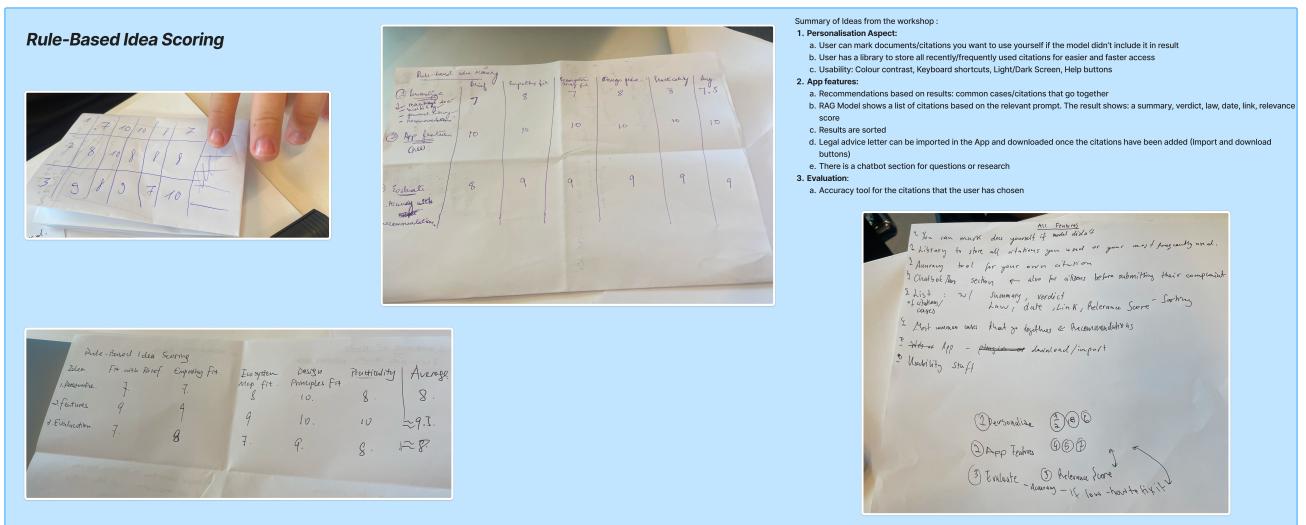


Figure 4: Rule-Based Idea Scoring

**Gemeente Amsterdam**

**UNIVERSITEIT VAN AMSTERDAM**

**CiteConnect**

Linking Legal Advice to Case Law Using RAG  
Data Systems Project  
Supervisor: Miles Glanze, Koenraad Verheyen, Agnieszka Kowalska

**Abstract**

Municipal legal advisors must cite relevant ECCL cases when drafting legal advice letters, but due to poor metadata, inconsistent key-wording, and limited search functions, retrieving the right case is slow and frustrating. We focus on Dutch ECCL cases where the municipality is a party. We apply a RAG-based solution to improve ECCL retrieval during drafting, reducing the average time needed to identify a relevant ECCL citation by at least 30% while maintaining or improving perceived relevance in a user evaluation.

**Persona**

<b>Maggie</b>	<b>Frankie</b>
<b>Background</b> 25 y.o. recent law graduate Tech savvy but small previous experience Working in a high stress environment. Lots of expectations and deadlines	<b>Background</b> 55 y.o. legal professional in Gemeente Amsterdam Experienced in legal advice letter drafting and supervising juniors Juniors she supervises work with AI
<b>Motivation &amp; Frustrations</b> Looking for a support system that saves her time and introduces new perspectives Current citation system is old and time-wasting	<b>Motivation &amp; Frustrations</b> Wants to keep up with the latest technology Does not trust AI - spends extra time verifying results Does not want to spend time learning a new difficult tool

**Empathy Map**

**Ecosystem Map**

<b>User factors</b> User's ability for memory / recall / prioritization / task switching Not familiar with all the citations Ability to work under pressure / deadlines to meet User's tech literacy, habits and expectations	<b>Ambient factors</b> User is sitting in a legal office with stacks of paper and a computer screen User is generally collaborating with others	<b>Technical factors</b> How fast and accurately users can achieve their goals Output should be format: match the vibe of legal office Output is not biased The system's ability to adjust to different users & preferences
---	---	---

**Terminology**

- European Case Law Identifier (ECCL):** uniform identifier for judicial decisions
- Legal Advice Letter:** provides concise and practical advice, addressing immediate concerns and guides the client into making informed decisions

**Data**

Dataset	Size	Scope	Key Fields
Advice letters	567	Bicycle-towing objections	letter id, zaaltype, onderwerp, dictum, anonymized text, unique words, cited eccl, land, rechtbank, jaar, code, full text, length
ECCL cases	2,417	Dutch court rulings	full text, length

**Key Design Features**

**Core App Features and Evaluation**

- RAG-based citation suggestions with: summary · verdict · law · date · link · relevance score
- Sorted results with recommended co-cited cases
- Import & export: upload draft letter, add citations, download updated version
- Built-in chatbot for follow-up questions or research
- Evaluation: Accuracy checking tool for user-selected citations

**Personalisation**

- User can manually mark missing relevant citations
- User has a personal library of recent / frequently used ECCL for quick reuse

**Usability**

- Accessible UI: colour contrast, light/dark mode, keyboard shortcuts
- Help / guidance buttons for onboarding and support

**Research**

- LegalBench-RAG: A Benchmark for Retrieval-Augmented Generation in the Legal Domain**
- LEGAL-BERT: The Magpie's straight out of Law School**
- CLUEC: A Dataset for Legal Case Retrieval and Retrieval-Augmented Analysis Generation**

**Concept model sketch**

**References**

- https://huggingface.co/
- https://www.semanticscience.org/
- https://www.semanticscience.org/
- https://www.semanticscience.org/
- https://www.semanticscience.org/

Figure 5: Ideation Poster