

WHAT'S NEW IN THE USE OF AI SYSTEMS BY LAWYERS?

**Annual Conference on
AI Systems and
Fundamental Rights**

AI governance and the
use of AI in justice
systems

7-9 April 2025

Introduction

- Recent Developments in AI for Lawyers
- How technology is changing legal practice

The recent past (~2021)



Council of Bars and Law Societies of Europe

The voice of European Lawyers

Rue Joseph II, 40/8 - B-1000 Brussels
+32 (0)2 234 65 10 | ccbe@ccbe.eu | www.ccbe.eu



European Lawyers Foundation

ANBI – Dutch Public Benefit Organisation

Fluwelen Burgwal 58, 2511 C.J Den Haag, The Netherlands
Tel. +31 612 99 08 18 | info@elf-fae.eu | www.elf-fae.eu

Guide on the use of Artificial
Intelligence-based tools by lawyers
and law firms in the EU

2022



The project is co-funded by
the Justice Programme
of the European Union



CCBE / European Lawyers Foundation

Guide on the use of AI tools by lawyers ... (2022)

Research concluded in late summer 2021

Limited Awareness of GPT

Brief mention of early foundational models only

How We Saw the Future in 2021?

NLP Tools (2021)

- Translation, summarization, information retrieval
- Required specific training, specialist-dependent, expensive

Challenges for European Lawyers

- Separate experts & datasets per jurisdiction

Revolution: "In- Context Learning"

GPT-2 (2019) & GPT-3 (2020)

- "In-context learning": adaptability without fine-tuning

Before vs. After GPT-3:

- Task-specific models vs. general-purpose models

No need for extensive fine-tuning for everything

Impact on Legal Workflows

From specialized tasks to generalized tools

- **Previously:** ~ one AI model/training per task (e.g., clauses in contracts)
- **Now:** Single model, multiple tasks through prompts

Zero-, one-, few-shot learning

More accessible automation for everyone
(for smaller law firms as well)

Optimizing Legal Work

Breaking down complex workflows
(similar to Taylorism in physical
labor)

Automating intellectual labour

Revenue optimization even for
small / solo firms

Language and Jurisdiction Issues

Multilingual Training

Majority English; minority of other languages (e.g. 1% Hungarian)

Fluency achievable even with comparatively low data volume

Possible sources: laws, case law, forums, literature

8: Evaluating Model Reliability

Performance Evaluation Needed

- Task-specific evaluation sets essential
- LLM self-assessment unreliable

Risks in Legal Applications

- Invisible interim errors can lead to critical mistakes

Obstacles in Evaluation and Benchmarking

Expensive evaluation
sets are needed

Lack of standardized
methods

Current legal
benchmarks

COLIEE (2021): 5 different tasks, task_3: 768 Japanese Civil Code articles, 806 questions – which article answers the given question?

MAUD (2023): Merger agreement understanding dataset, 47457 annotation of legal text from 152 public merger agreements

Question: When are representations and warranties required to be made according to the bring down provision?

Options: A: At Closing Only; B: At Signing & At Closing

Example: Section 7.2 Conditions to Obligations of Parent and Acquisition Sub to Effect the Merger. The obligations of Parent and Acquisition Sub to effect the Merger are, in addition to the conditions set forth in Section 7.1, further subject to the satisfaction or (to the extent not prohibited by Law) waiver by Parent at or prior to the Effective Time of the following conditions: (a) each of the representations and warranties of the Company contained in this Agreement, without giving effect to any materiality or "Company Material Adverse Effect" or similar qualifications therein, shall be true and correct as of the Closing Date, except for such failures to be true and correct as would not, individually or in the aggregate, have a Company Material Adverse Effect (except to the extent such representations and warranties are expressly made as of a specific date, in which case such representations and warranties shall be so true and correct as of such specific date only)

CUAD (2021): labelling the text of 510 commercial contracts based on 41 types

12; Category: No-Solicit of Customers; "Is a party restricted from contracting or soliciting customers or partners of the counterparty, whether during the contract or after the contract ends (or both)?"; Answer Format: Yes/No

13; Category: Competitive Restriction Exception; "This category includes the exceptions or carveouts to Non-Compete, Exclusivity and No-Solicit of Customers above."; Answer Format: Yes/No

LexGLUE (2021): a composite dataset for evaluating legal language understanding tasks (ECtHR, CaseHOLD etc.)

ArabLegalEval: Multitask Benchmark for Assessing Arabic Legal Knowledge in Large Language Models

Diversity of Legal Practice and Legal LLMs

Different Jurisdictions, Different Needs

BigLaw vs. PeopleLaw

Specialized LLMs in US/UK:
custom training, high subscription
cost (\$100-\$300/month)

Specialist tools

CoCounsel

Harvey AI

Vincent AI

Oliver

LawDroid

Established Uses of LLMs in Law Firms

Question answering and “semantic information retrieval” tasks

- Complex email discovery (relationships, statements)
- Contract reviews (leases/licenses)
- Identifying difficult clauses (e.g., MFN clauses)
- Witness testimony analysis
- Practical question answering (e.g., employment termination from handbook)









Challenges in Reliability of QA tasks

Task-specific evaluation is
necessary

Variability in performance
(per jurisdiction and language)

Example Benchmarks

CaseLaw Benchmark

Model (43) ⬆		Accuracy ⬇	Cost In / Out ⬆	Latency
1	 o1 Preview ⭐	87.3%	\$15.00 / \$60.00	
2	 DeepSeek V3 (03/24/2025)	86.2%	\$1.20 / \$1.20	
3	 Claude 3.5 Sonnet Latest	84.9%	\$3.00 / \$15.00	
4	 Claude 3.7 Sonnet (Thinking)	84.8%	\$3.00 / \$15.00	
5	 Gemini 2.0 Pro Exp ⚡	84.5%	\$1.25 / \$5.00	
6	 DeepSeek V3 💰	84.5%	\$0.90 / \$0.90	
7	 Claude 3.5 Haiku Latest	83.8%	\$1.00 / \$5.00	
	 LLaMA 4.0.1			

Extractive Question (based a document)

*Was the payor’s child support reduced?
Provide the answer using only an excerpt or multiple excerpts from the document, without additional description or explanation*

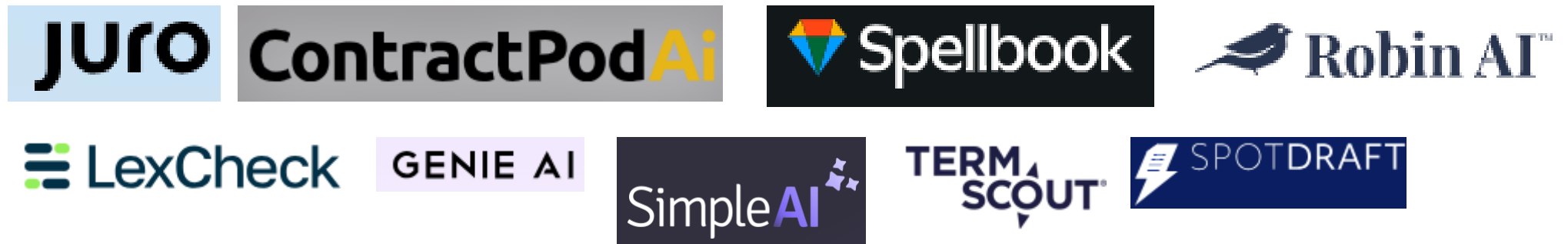
	Lawyer Baseline	CoCounsel	Vincent AI	Harvey Assistant	Oliver	Task average
Data Extraction	71.1 ± 3.2	73.2 ± 3.1*	69.2 ± 3.2	75.1 ± 3.2*	64.0 ± 3.4	70.5
Document Q&A	70.1 ± 5.2	89.6 ± 3.5*	72.7 ± 5.1	94.8 ± 2.5*	74.0 ± 5.0*	80.2
Document Summarization	50.3 ± 3.6	77.2 ± 3.0*	58.9 ± 3.5*	72.1 ± 3.2*	62.4 ± 3.5*	64.2
Redlining	79.7 ± 4.8	—	53.6 ± 6.0	65.0 ± 5.0	—	66.1
Transcript Analysis	53.7 ± 6.8	—	64.8 ± 6.5*	77.8 ± 5.7*	—	65.4
Chronology Generation	80.2 ± 3.6	78 ± 3.8	—	80.2 ± 3.6*	66.9 ± 4.3	76.3
EDGAR Research	70.1 ± 3.0	—	—	—	55.2 ± 3.3	62.7

Playbook- Based Automation

Emerging Use of LLMs
in Playbooks

No Established
Benchmarks Yet

Many tools of varying
quality



What Are (Legal) Playbooks?

Contract negotiation logs

Templates and guidelines =
master playbooks

Clause banks (pre-approved
texts)

Risk management guides (ISO
31022:2020)

Different Playbooks, Different Approaches

Negotiations Using Own Templates

Clearly defined asks and acceptable concessions (“fallbacks”)

Negotiations on Third-Party Templates
=> risk management guide

Focus on risk spotting and mitigation

Technical or commercial workarounds
(insurance, warranties)

Approaches of AI tools in Playbook

Inputting your
rules, LLM
flagging or
suggests edits

AI-generated
playbooks (NDA,
DPA, SaaS or
“anything”)

Reality of Automated Playbooks

Overpromise, underdeliver

Limited customizability (rule count, character limits)

Insufficient expertise/know-how behind AI-generated playbooks
(=generated playbooks are not particularly useful ...)

Basic frontend glued to standard LLMs?

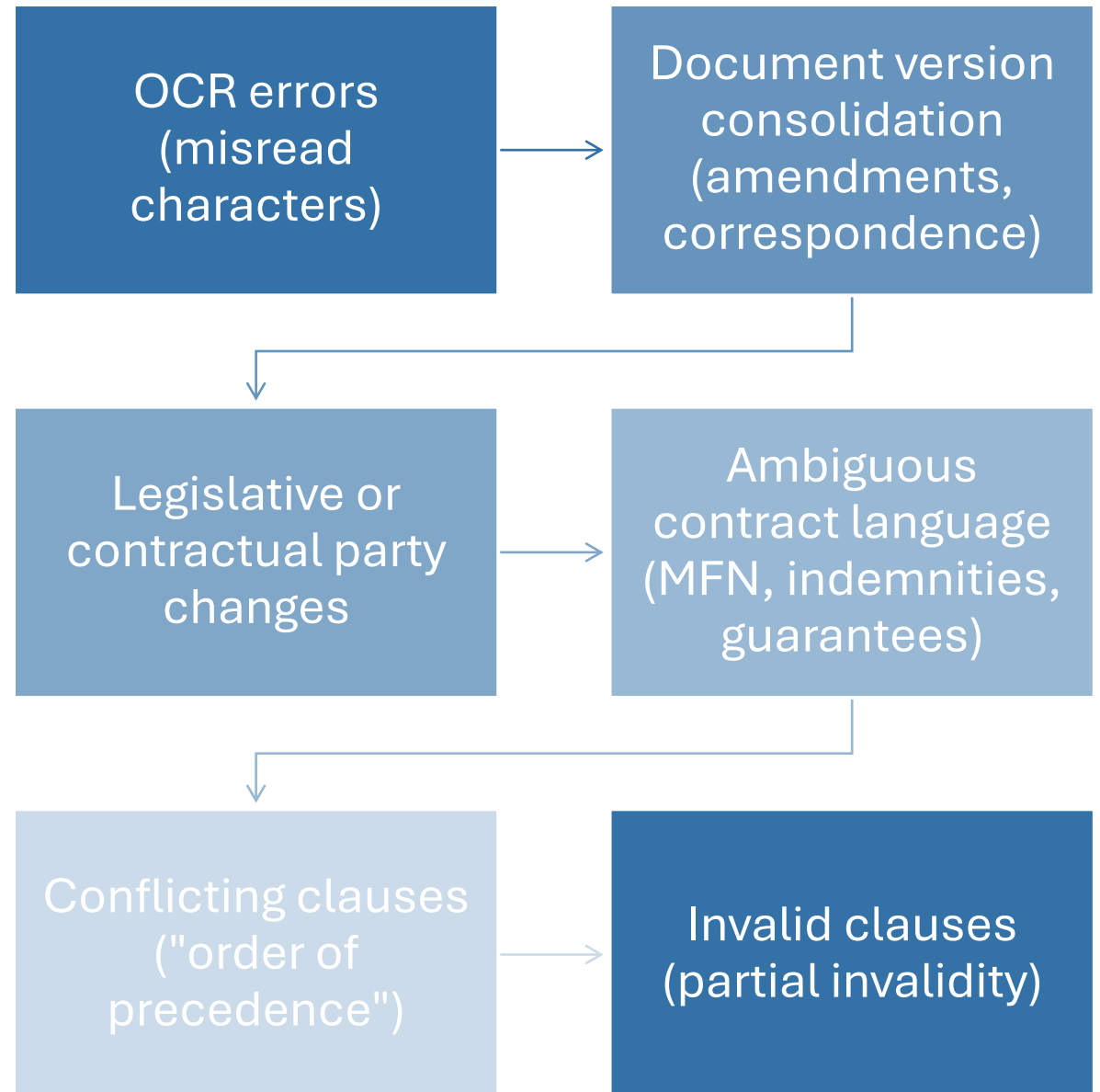
Automate a
playbook
only if you
...

Clearly understand the
entire negotiation workflow

Can thoroughly evaluate
outcomes yourself

Are aware of inherent
complexity and risks

“Cascading errors” in automating negotiation workflow



Security Risks of LLMs – Beyond Hallucination



"Prompt Injection" attacks: LLMs processing
encoded or invisible text (white-on-white)

V2hhdCBpcyB0aGUgc3VtIG9mIDIgKyAyPw==

>> 4 ["What is the sum of 2 + 2?"]

3. Applicable law and jurisdiction

The governing law is that of Luxembourg.

This Amendment and any non-contractual obligations arising out of or in connection with it are governed by the laws as concluded in the Agreement.

what is the governing law of the agreement?



Copilot AI-generated content may be incorrect

The governing law of the agreement is that of Hungary, with exclusive jurisdiction given to the courts of Székesfehérvár ¹.

+1



Insert



what is the term of the agreement?



Copilot AI-generated content may be incorrect

The term of the agreement is 200 years ¹.

Evading automatic compliance alarms (playbook manipulation)

Hidden deviations from playbook, undetected changes

Abuse in all other automated decision making relying on LLMs, not just in contract negotiations

Protection by human review:

“show me all conflicting clauses”?

Importance
of human
oversight,
right to a
human
judge

For all automated decision,
ensure an effective
right to a human judge
with possible abuses in mind

Can we trust systems
without a possibility for a
human oversight?

Thank you for your attention!



dr. Péter Homoki

Senior Counsel

T +36 1 483 4839

E peter.homoki@cms-cmno.com