

---

# CUAD: An Expert-Annotated NLP Dataset for Legal Contract Review—Supplementary Materials

---

## A Appendix

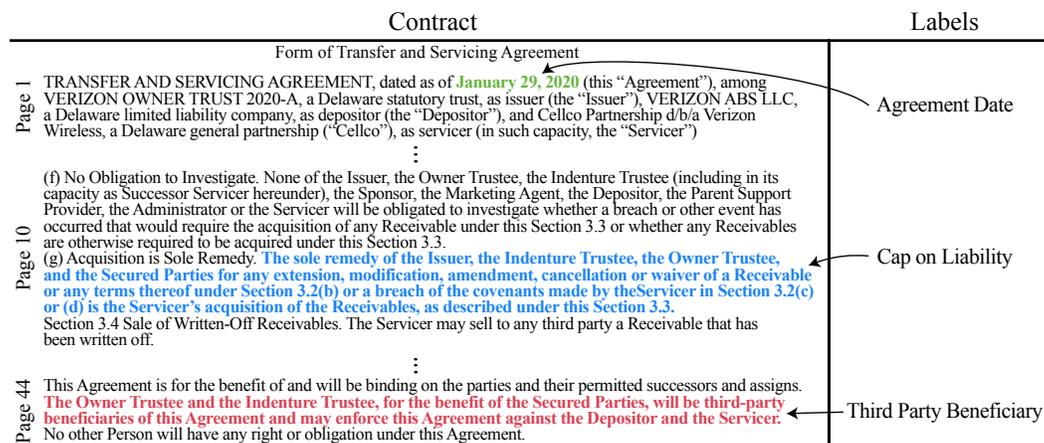


Figure 1: Our dataset consists of over 500 contracts, each carefully labeled by legal experts to identify important clauses, which models can then learn to extract from contracts. Our dataset covers a diverse set of contracts, including 25 different contract types. It can be tedious and expensive for legal professionals to manually find important clauses, especially from long contracts such as this one with over 100 pages long.

### A.1 Special Cases

The one small exception during metric computation is for the Parties label, which (unlike for the other labels) often has several very small extracted segments of text in a given contract. We relax what counts as a match for the Parties label by also counting as a match any case when the ground truth segment is a *substring* of a predicted extraction of text. This is reasonable in practice because our predicted extractions are bounded by to be at most about a paragraph in length. Another exception is that the Price Restrictions provision did not have examples in the test set due to randomization in our train-test split, so performance for that class was ignored in this paper.

### A.2 Dataset Details

**Labeling Process Details.** The steps of our dataset creation process is as follows.

1. *Law Student training.* Law students attended training sessions on each of the categories that included a summary, video instructions by experienced attorneys, multiple quizzes and workshops. Students were then required to label sample contracts in eBrevia, an online contract review tool. The initial training took approximately 70-100 hours.
2. *Law Student Label.* Law students conducted manual contract review and labeling in eBrevia.

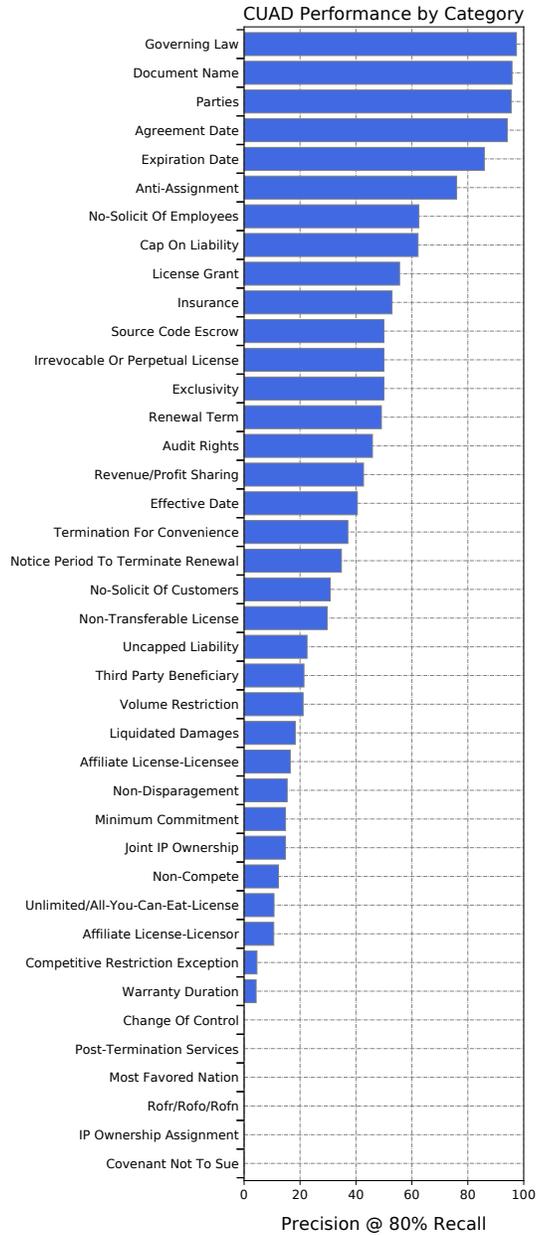


Figure 2: Comparison of Precision @ 80% Recall for DeBERTa-xlarge across different label categories. While performance is high for some labels, it has much room for improvement for other labels.

Contract Type	Number of Contracts
Affiliate Agreement	10
Agency Agreement	13
Collaboration Agreement	26
Co-Branding Agreement	22
Consulting Agreement	11
Development Agreement	29
Distributor Agreement	32
Endorsement Agreement	24
Franchise Agreement	15
Hosting Agreement	20
IP Agreement	17
Joint Venture Agreement	23
License Agreement	33
Maintenance Agreement	34
Manufacturing Agreement	17
Marketing Agreement	17
Non-Compete Agreement	3
Outsourcing Agreement	18
Promotion Agreement	12
Reseller Agreement	12
Service Agreement	28
Sponsorship Agreement	31
Supply Agreement	18
Strategic Alliance Agreement	32
Transportation Agreement	13
Total	510

Table 1: A breakdown of contract types and their count.

3. *Key Word Search.* Law students conducted keyword search in eBrevia to capture additional categories that have been missed during the “Student Label” step.
4. *Category-by-Category Report Review.* Law students exported the labeled clauses into reports, review each clause category-by-category and highlight clauses that they believe are mislabeled.
5. *Attorney Review.* Experienced attorneys reviewed the category-by-category report with students comments, provided comments and addressed student questions. When applicable, attorneys discussed such results with the students and reached consensus. Students made changes in eBrevia accordingly.
6. *eBrevia Extras Review.* Attorneys and students used eBrevia to generate a list of “extras,” which are clauses that eBrevia AI tool identified as responsive to a category but not labeled by human annotators. Attorneys and students reviewed all of the “extras” and added the correct ones. The process is repeated until all or substantially all of the “extras” are incorrect labels.
7. *Final Report.* The final report was exported into a CSV file. Volunteers manually added the “Yes/No” answer column to categories that do not contain an answer.

**Redacted Information.** Some clauses in the files are redacted because the party submitting these contracts redacted them to protect confidentiality. Such redaction may show up as \*\*\* or \_\_\_ or blank space. The dataset and the answers reflect such redactions. For example, the answer for “January \_\_\_ 2020” would be “1/[ ]/2020”).

Some sentences in the files include confidential legends that are not part of the contracts. An example of such confidential legend is as follows: THIS EXHIBIT HAS BEEN REDACTED AND IS THE SUBJECT OF A CONFIDENTIAL TREATMENT REQUEST. REDACTED MATERIAL IS MARKED WITH [\* \* \*] AND HAS BEEN FILED SEPARATELY WITH THE SECURITIES AND EXCHANGE COMMISSION. Some sentences in the files contain irrelevant information such as footers or page numbers. Some sentences may not be relevant to the corresponding category. Some

sentences may correspond to a different category. Because many legal clauses are very long and contain various sub-parts, sometimes only a sub-part of a sentence is responsive to a category.

**Contract Types.** We provide a list of each of the 25 contract types, along with the number of contracts in CUAD of each type, in Table 1.

**Label Category Details.** We provide descriptions of every label category in Tables 2 and 3.

### A.3 Conversion to SQuAD 2.0 Format

In the question answering literature, some datasets have answers that are spans of given input text, similar to us. A particularly notable dataset that shares this format is SQuAD 2.0 (Rajpurkar et al., 2018), a reading comprehension dataset with questions that have spans of the passage as answers.

To facilitate the use of prior work on datasets such as SQuAD 2.0, we format our dataset in the same format. In particular, we first segment a contract into different paragraphs typically range from one to five sentences. Then for each label category and each such paragraph, we format the question as follows:

“Highlight the parts (if any) of this clause related to "<Label Category>". Details: <Label Category Description>”

where the label category descriptions are the same as in Tables 2 and 3.

The answer is then the span of text of the given passage that should be highlighted, or the empty string if nothing should be highlighted as relevant to that label category, along with the character position where that span begins.

## References

Pranav Rajpurkar, Robin Jia, and Percy Liang. Know what you don’t know: Unanswerable questions for squad. *ArXiv*, abs/1806.03822, 2018.

Category	Description
Document Name	The name of the contract
Parties	The two or more parties who signed the contract
Agreement Date	The date of the contract
Effective Date	On what date is the contract is effective?
Expiration Date	On what date will the contract's initial term expire?
Renewal Term	What is the renewal term after the initial term expires? This includes automatic extensions and unilateral extensions with prior notice.
Notice to Terminate Re- newal	What is the notice period required to terminate renewal?
Governing Law	Which state/country's law governs the interpretation of the contract?
Most Favored Nation	Is there a clause that if a third party gets better terms on the licensing or sale of technology/goods/services described in the contract, the buyer of such technology/goods/services under the contract shall be entitled to those better terms?
Non-Compete	Is there a restriction on the ability of a party to compete with the counterparty or operate in a certain geography or business or technology sector?
Exclusivity	Is there an exclusive dealing commitment with the counterparty? This includes a commitment to procure all "requirements" from one party of certain technology, goods, or services or a prohibition on licensing or selling technology, goods or services to third parties, or a prohibition on collaborating or working with other parties), whether during the contract or after the contract ends (or both).
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)?
Competitive Restriction Exception	This category includes the exceptions or carveouts to Non-Compete, Exclusivity and No-Solicit of Customers above.
No-Solicit of Employees	Is there a restriction on a party's soliciting or hiring employees and/or contractors from the counterparty, whether during the contract or after the contract ends (or both)?
Non-Disparagement	Is there a requirement on a party not to disparage the counterparty?
Termination for Conve- nience	Can a party terminate this contract without cause (solely by giving a notice and allowing a waiting period to expire)?
ROFR/ROFO/ROFN	Is there a clause granting one party a right of first refusal, right of first offer or right of first negotiation to purchase, license, market, or distribute equity interest, technology, assets, products or services?
Change of Control	Does one party have the right to terminate or is consent or notice required of the counterparty if such party undergoes a change of control, such as a merger, stock sale, transfer of all or substantially all of its assets or business, or assignment by operation of law?
Anti-Assignment	Is consent or notice required of a party if the contract is assigned to a third party?
Revenue/Profit Sharing	Is one party required to share revenue or profit with the counterparty for any technology, goods, or services?
Price Restriction	Is there a restriction on the ability of a party to raise or reduce prices of technology, goods, or services provided?
Minimum Commitment	Is there a minimum order size or minimum amount or units per-time period that one party must buy from the counterparty under the contract?
Volume Restriction	Is there a fee increase or consent requirement, etc. if one party's use of the product/services exceeds certain threshold?
IP Ownership Assign- ment	Does intellectual property created by one party become the property of the counterparty, either per the terms of the contract or upon the occurrence of certain events?
Joint IP Ownership	Is there any clause providing for joint or shared ownership of intel-lectual property between the parties to the contract?

Table 2: Label categories and their descriptions (part 1/2).

Category	Description
License Grant	Does the contract contain a license granted by one party to its counterparty?
Non-Transferable License	Does the contract limit the ability of a party to transfer the license being granted to a third party?
Affiliate IP License-Licensors	Does the contract contain a license grant by affiliates of the licensor or that includes intellectual property of affiliates of the licensor?
Affiliate IP License-Licensee	Does the contract contain a license grant to a licensee (incl. sublicensee) and the affiliates of such licensee/sublicensor?
Unlimited/All-You-Can-Eat License	Is there a clause granting one party an “enterprise,” “all you can eat” or unlimited usage license?
Irrevocable or Perpetual License	Does the contract contain a license grant that is irrevocable or perpetual?
Source Code Escrow	Is one party required to deposit its source code into escrow with a third party, which can be released to the counterparty upon the occurrence of certain events (bankruptcy, insolvency, etc.)?
Post-Termination Services	Is a party subject to obligations after the termination or expiration of a contract, including any post-termination transition, payment, transfer of IP, wind-down, last-buy, or similar commitments?
Audit Rights	Does a party have the right to audit the books, records, or physical locations of the counterparty to ensure compliance with the contract?
Uncapped Liability	Is a party’s liability uncapped upon the breach of its obligation in the contract? This also includes uncap liability for a particular type of breach such as IP infringement or breach of confidentiality obligation.
Cap on Liability	Does the contract include a cap on liability upon the breach of a party’s obligation? This includes time limitation for the counterparty to bring claims or maximum amount for recovery.
Liquidated Damages	Does the contract contain a clause that would award either party liquidated damages for breach or a fee upon the termination of a contract (termination fee)?
Warranty Duration	What is the duration of any warranty against defects or errors in technology, products, or services provided under the contract?
Insurance	Is there a requirement for insurance that must be maintained by one party for the benefit of the counterparty?
Covenant Not to Sue	Is a party restricted from contesting the validity of the counterparty’s ownership of intellectual property or otherwise bringing a claim against the counterparty for matters unrelated to the contract?
Third Party Beneficiary	Is there a non-contracting party who is a beneficiary to some or all of the clauses in the contract and therefore can enforce its rights against a contracting party?

Table 3: Label categories and their descriptions (part 2/2).