
CONTRACT UNDERSTANDING ATTICUS DATASET

Contract Understanding Atticus Dataset (CUAD) v1 is a corpus of more than 13,000 labels in 510 commercial legal contracts that have been manually labeled to identify 41 categories of important clauses that lawyers look for when reviewing contracts in connection with corporate transactions.

CUAD is curated and maintained by The Atticus Project, Inc. to support NLP research and development in legal contract review. Analysis of CUAD can be found at <https://arxiv.org/abs/2103.06268>. Code for replicating the results and the trained model can be found at <https://github.com/TheAtticusProject/cuad>.

FORMAT

The files in CUAD v1 include 1 CSV file, 1 SQuAD-style JSON file, 28 Excel files, 510 PDF files, and 510 TXT files.

- 1 master clauses CSV: a 83-column 511-row file. The first column is the names of the contracts corresponding to the PDF and TXT files in the "full_contracts_pdf" and "full_contracts_txt" folders. The remaining columns contain (1) text context (sometimes referred to as clause), and (2) human-input answers that correspond to each of the 41 categories in these contracts. See a list of the categories in "Category List" below. The first row represents the file name and a list of the categories. The remaining 510 rows each represent a contract in the dataset and include the text context and human-input answers corresponding to the categories. The human-input answers are derived from the text context and are formatted to a unified form.
- 1 SQuAD-style JSON: this file is derived from the master clauses CSV to follow the same format as SQuAD 2.0 (<https://rajpurkar.github.io/SQuAD-explorer/explore/v2.0/dev/>), a question answering dataset whose answers are similarly spans of the input text. The exact format of the JSON format exactly mimics that of SQuAD 2.0 for compatibility with prior work. We also provide Python scripts for processing this data for further ease of use.
- 28 Excels: a collection of Excel files containing clauses responsive to each of the categories identified in the "Category List" below. The first column is the names of the contracts corresponding to the PDF and TXT files in the "full_contracts_pdf" and "full_contracts_txt" folders. The remaining columns contain (1) text context (clause) corresponding to one or more Categories that belong in the same group as identified in "Category List" below, and (2) in some cases, human-input answers that correspond to such text context. Each file is named as "Label Report - [label/group name] (Group [number]).xlsx"
- 510 full contract PDFs: a collection of the underlying contracts that we used to extract the labels. Each file is named as "[document name].pdf". These contracts are in a PDF format and are not labeled. The full contract PDFs contain raw data and are provided for context and reference.
- 510 full contract TXTs: a collection of TXT files of the underlying contracts. Each file is named as "[document name].txt". These contracts are in a plaintext format and are not labeled. The full contract TXTs contain raw data and are provided for context and reference.

We recommend using the master clauses CSV as a starting point. To facilitate work with prior work and existing language models, we also provide an additional format of the data that is

similar to datasets such as SQuAD 2.0. In particular, each contract is broken up into paragraphs, then for each provision category a model must predict the span of text (if any) in that paragraph that corresponds to that provision category.

=====

DOWNLOAD

Download CUAD v1 at www.atticusprojectai.org/cuad.

=====

CATEGORIES AND TASKS

The labels correspond to 41 categories of legal clauses in commercial contracts that are considered important by experienced attorneys in contract review in connection with a corporate transaction. Such transactions include mergers & acquisitions, investments, initial public offering, etc.

Each category supports a contract review task which is to extract from an underlying contract (1) text context (clause) and (2) human-input answers that correspond to each of the categories in these contracts. For example, in response to the "Governing Law" category, the clause states "This Agreement is accepted by Company in the State of Nevada and shall be governed by and construed in accordance with the laws thereof, which laws shall prevail in the event of any conflict.". The answer derived from the text context is Nevada.

To complete the task, the input will be an unlabeled contract in PDF format, and the output should be the text context and the derived answers corresponding to the categories of legal clauses.

Each category (including context and answer) is independent of another except as otherwise indicated in "Category List" "Group" below.

33 out of the 41 categories have a derived answer of "Yes" or "No." If there is a segment of text corresponding to such a category, the answer should be yes. If there is no text corresponding to such a category, it means that no string was found. As a result, the answer should be "No."

8 out of the 41 categories ask for answers that are entity or individual names, dates, combination of numbers and dates and names of states and countries. See descriptions in the "Category List" below. While the format of the context varies based on the text in the contract (string, date, or combination thereof), we represent answers in consistent formats. For example, if the Agreement Date in a contract is "May 8, 2014" or "8th day of May 2014", the Agreement Date Answer is "5/8/2014".

The "Expiration Date" and the "Effective Date" categories may ask for answers that are based on a combination of (1) the answer to "Agreement Date" or "Effective Date" and/or (2) the string corresponding to "Expiration Date" or "Effective Date".

For example, the "Effective Date" clause in a contract is "This agreement shall begin upon the date of its execution". The answer will depend on the date of the execution, which was labeled as "Agreement Date", the answer to which is "5/8/2014". As a result, the answer to the "Effective Date" should be "5/8/2014".

An example of the "Expiration Date" clause is "This agreement shall begin upon the date of its execution by MA and acceptance in writing by Company and shall remain in effect until the end of the current calendar year and shall be automatically renewed for successive one (1) year

periods unless otherwise terminated according to the cancellation or termination clauses contained in paragraph 18 of this Agreement. (Page 2)." The relevant string in this clause is "in effect until the end of the current calendar year". As a result, the answer to "Expiration Date" is 12/31/2014.

A second example of the "Expiration Date" string is "The initial term of this Agreement commences as of the Effective Date and, unless terminated earlier pursuant to any express clause of this Agreement, shall continue until five (5) years following the Effective Date (the "Initial Term"). The answer here is 2/10/2019, representing five (5) years following the "Effective Date" answer of 2/10/2014.

Each category (incl. context and answer) is independent of another except otherwise indicated under the "Group" column below. For example, the "Effective Date", "Agreement Date" and "Expiration Date" clauses in a contract can overlap or build upon each other and therefore belong to the same Group 1. Another example would be "Expiration Date", "Renewal Term" and "Notice to Terminate Renewal", where the clause may be the same for two or more categories.

For example, the clause states that "This Agreement shall expire two years after the Effective Date, but then will be automatically renewed for three years following the expiration of the initial term, unless a party provides notice not to renew 60 days prior the expiration of the initial term." Consequently the answer to Effective Date is 2/14/2019, the answer to Expiration Date should be 2/14/2021, and the answer to "Renewal Term" is 3 years, the answer to "Notice to Terminate Renewal" is 60 days.

Similarly, a "License Grant" clause may also correspond to "Exclusive License", "Non-Transferable License" and "Affiliate License-Licensee" categories.

=====

CATEGORY LIST

Category (incl. context and answer)

Description

Answer Format

Group

1

Category: Document Name

Description: The name of the contract

Answer Format: Contract Name

Group: -

2

Category: Parties

Description: The two or more parties who signed the contract

Answer Format: Entity or individual names

Group: -

3

Category: Agreement Date

Description: The date of the contract

Answer Format: Date (mm/dd/yyyy)

Group: 1

4

Category: Effective Date

Description: The date when the contract is effective

Answer Format: Date (mm/dd/yyyy)

Group: 1

5

Category: Expiration Date

Description: On what date will the contract's initial term expire?

Answer Format: Date (mm/dd/yyyy) / Perpetual

Group: 1

6

Category: Renewal Term

Description: What is the renewal term after the initial term expires? This includes automatic extensions and unilateral extensions with prior notice.

Answer Format: [Successive] number of years/months / Perpetual

Group: 1

7

Category: Notice to Terminate Renewal

Description: What is the notice period required to terminate renewal?

Answer Format: Number of days/months/year(s)

Group: 1

8

Category: Governing Law

Description: Which state/country's law governs the interpretation of the contract?

Answer Format: Name of a US State / non-US Province, Country

Group: -

9

Category: Most Favored Nation

Description: 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?

Answer Format: Yes/No

Group: -

10

Category: Non-Compete

Description: 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?

Answer Format: Yes/No

Group: 2

11

Category: Exclusivity

Description: 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).

Answer Format: Yes/No

Group: 2

12

Category: No-Solicit of Customers

Description: 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

Group: 2

13

Category: Competitive Restriction Exception

Description: This category includes the exceptions or carveouts to Non-Compete, Exclusivity and No-Solicit of Customers above.

Answer Format: Yes/No

Group: 2

14

Category: No-Solicit of Employees

Description: 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)?

Answer Format: Yes/No

Group: -

15

Category: Non-Disparagement

Description: Is there a requirement on a party not to disparage the counterparty?

Answer Format: Yes/No

Group: -

16

Category: Termination for Convenience

Description: Can a party terminate this contract without cause (solely by giving a notice and allowing a waiting period to expire)?

Answer Format: Yes/No

Group: -

17

Category: Right of First Refusal, Offer or Negotiation (ROFR/ROFO/ROFN)

Description: 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?

Answer Format: Yes/No

Group: -

18

Category: Change of Control

Description: 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?

Answer Format: Yes/No

Group: 3

19

Category: Anti-Assignment

Description: Is consent or notice required of a party if the contract is assigned to a third party?

Answer Format: Yes/No

Group: 3

20

Category: Revenue/Profit Sharing

Description: Is one party required to share revenue or profit with the counterparty for any technology, goods, or services?

Answer Format: Yes/No

Group: -

21

Category: Price Restriction

Description: Is there a restriction on the ability of a party to raise or reduce prices of technology, goods, or services provided?

Answer Format: Yes/No

Group: -

22

Category: Minimum Commitment

Description: 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?

Answer Format: Yes/No

Group: -

23

Category: Volume Restriction

Description: Is there a fee increase or consent requirement, etc. if one party's use of the product/services exceeds certain threshold?

Answer Format: Yes/No

Group: -

24

Category: IP Ownership Assignment

Description: 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?

Answer Format: Yes/No

Group: -

25

Category: Joint IP Ownership

Description: Is there any clause providing for joint or shared ownership of intellectual property between the parties to the contract?

Answer Format: Yes/No

Group: -

26

Category: License Grant

Description: Does the contract contain a license granted by one party to its counterparty?

Answer Format: Yes/No

Group: 4

27

Category: Non-Transferable License

Description: Does the contract limit the ability of a party to transfer the license being granted to a third party?

Answer Format: Yes/No

Group: 4

28

Category: Affiliate IP License-Licensor

Description: Does the contract contain a license grant by affiliates of the licensor or that includes intellectual property of affiliates of the licensor?

Answer Format: Yes/No

Group: 4

29

Category: Affiliate IP License-Licensee

Description: Does the contract contain a license grant to a licensee (incl. sublicense) and the affiliates of such licensee/sublicensor?

Answer Format: Yes/No

Group: 4

30

Category: Unlimited/All-You-Can-Eat License

Description: Is there a clause granting one party an "enterprise," "all you can eat" or unlimited usage license?

Answer Format: Yes/No

Group: -

31

Category: Irrevocable or Perpetual License

Description: Does the contract contain a license grant that is irrevocable or perpetual?

Answer Format: Yes/No

Group: 4

32

Category: Source Code Escrow

Description: 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.)?

Answer Format: Yes/No

Group: -

33

Category: Post-Termination Services

Description: 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?

Answer Format: Yes/No

Group: -

34

Category: Audit Rights

Description: Does a party have the right to audit the books, records, or physical locations of the counterparty to ensure compliance with the contract?

Answer Format: Yes/No

Group: -

35

Category: Uncapped Liability

Description: 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.

Answer Format: Yes/No

Group: 5

36

Category: Cap on Liability

Description: 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.

Answer Format: Yes/No

Group: 5

37

Category: Liquidated Damages

Description: 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)?

Answer Format: Yes/No

Group: -

38

Category: Warranty Duration

Description: What is the duration of any warranty against defects or errors in technology, products, or services provided under the contract?

Answer Format: Number of months or years

Group: -

39

Category: Insurance

Description: Is there a requirement for insurance that must be maintained by one party for the benefit of the counterparty?

Answer Format: Yes/No

Group: -

40

Category: Covenant Not to Sue

Description: 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?

Answer Format: Yes/No

Group: -

41

Category: Third Party Beneficiary

Description: 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?

Answer Format: Yes/No

Group: -

=====

SOURCE OF CONTRACTS

The contracts were sourced from EDGAR, the Electronic Data Gathering, Analysis, and Retrieval system used at the U.S. Securities and Exchange Commission (SEC). Publicly traded companies in the United States are required to file certain contracts under the SEC rules. Access to these contracts is available to the public for free at <https://www.sec.gov/edgar>. Please read the Datasheet at <https://www.atticusprojectai.org/> for information on the intended use and limitations of the CUAD.

=====

CATEGORY & CONTRACT SELECTION

The CUAD includes commercial contracts selected from 25 different types of contracts based on the contract names as shown below. Within each type, we randomly selected contracts based on the names of the filing companies across the alphabet.

Type of Contracts: # of Docs

Affiliate Agreement: 10

Agency Agreement: 13

Collaboration/Cooperation 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/No-Solicit/Non-Disparagement Agreement: 3

Outsourcing Agreement: 18

Promotion Agreement: 12

Reseller Agreement: 12

Service Agreement: 28

Sponsorship Agreement: 31

Supply Agreement: 18

Strategic Alliance Agreement: 32

=====

REDACTED INFORMATION AND TEXT SELECTIONS

Some clauses in the files are redacted because the party submitting these contracts redacted them to protect confidentiality. Such redaction may show up as asterisks (***), underscores (_), or blank spaces. The dataset and the answers reflect such redactions. For example, the answer for "January _ 2020" would be "1/_/2020".

For any categories that require an answer of "Yes/No", annotators include full sentences as text context in a contract. To maintain consistency and minimize inter-annotator disagreement, annotators select text for the full sentence, under the instruction of "from period to period".

For the other categories, annotators selected segments of the text in the contract that are responsive to each such category. One category in a contract may include multiple labels. For example, "Parties" may include 4-10 separate text strings that are not continuous in a contract. The answer is presented in the unified format separated by semicolons of "Party A Inc. ("Party A"); Party B Corp. ("Party B")".

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.

To address the foregoing limitations, annotators manually deleted the portion that is not responsive, replacing it with the symbol "" to indicate that the two text segments do not appear immediately next to each other in the contracts. For example, if a "Termination for Convenience" clause starts with "Each Party may terminate this Agreement if" followed by three subparts "(a), (b) and (c)", but only subpart (c) is responsive to this category, we manually delete subparts (a) and (b) and replace them with the symbol "". Another example is for "Effective Date", the contract includes a sentence "This Agreement is effective as of the date written above" that appears after the date "January 1, 2010". The annotation is as follows: "January 1, 2010 This Agreement is effective as of the date written above."

Because the contracts were converted from PDF into TXT files, the converted TXT files may not stay true to the format of the original PDF files. For example, some contracts contain inconsistent spacing between words, sentences and paragraphs. Table format is not maintained in the TXT files.

=====

LABELING PROCESS

Our labeling process included multiple steps to ensure accuracy:

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.

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.

=====

LICENSE

CUAD is licensed under the Creative Commons Attribution 4.0 (CC BY 4.0) license and free to the public for commercial and non-commercial use.

We make no representations or warranties regarding the license status of the underlying contracts, which are publicly available and downloadable from EDGAR.

Privacy Policy & Disclaimers

The categories or the contracts included in the dataset are not comprehensive or representative. We encourage the public to help us improve them by sending us your comments and suggestions to info@atticusprojectai.org. Comments and suggestions will be reviewed by The Atticus Project at its discretion and will be included in future versions of Atticus categories once approved.

The use of CUAD is subject to our privacy policy <https://www.atticusprojectai.org/privacy-policy> and disclaimer <https://www.atticusprojectai.org/disclaimer>.

=====

CONTACT

Email info@atticusprojectai.org if you have any questions.

=====

ACKNOWLEDGEMENTS

Attorney Advisors

Wei Chen, John Brockland, Kevin Chen, Jacky Fink, Spencer P. Goodson, Justin Haan, Alex Haskell, Kari Krusmark, Jenny Lin, Jonas Marson, Benjamin Petersen, Alexander Kwonji Rosenberg, William R. Sawyers, Brittany Schmeltz, Max Scott, Zhu Zhu

Law Student Leaders

John Batoha, Daisy Beckner, Lovina Consunji, Gina Diaz, Chris Gronseth, Calvin Hannagan, Joseph Kroon, Sheetal Sharma Saran

Law Student Contributors

Scott Aronin, Bryan Burgoon, Jigar Desai, Imani Haynes, Jeongsoo Kim, Margaret Lynch, Allison Melville, Felix Mendez-Burgos, Nicole Mirkazemi, David Myers, Emily Rissberger, Behrang Seraj, Sarahginy Valcin

Technical Advisors & Contributors

Dan Hendrycks, Collin Burns, Spencer Ball, Anya Chen