From Knowledge Management to Intelligence Engineering - A practical approach to building AI inside the law-firm using open-source Large Language Models

Uwais Iqbal[1]

[1]Simplexico Limited, 33 Carnarvon Road, London, E10 6DW, United Kingdom

Abstract

Open-source foundational language models unlock a new opportunity for building AI inside the law firm. In this paper, we explore the different options in the AI build vs buy equation facing law firms and outline four postures across the spectrum of building AI. We motivate a particular posture that leverages open-source foundational models in a way that both mitigates data privacy and security concerns, while enabling customisation of these models with internal data. We explore the different ways in which these models can be fine-tuned and present a novel addition of *intelligence engineering* to the traditional knowledge management process that involves instruction fine-tuning language models to infinitely scale access to explicit knowledge. We provide a practical demonstration of this technical approach with a proof of concept using an open-source foundational model based on the GPT-3 architecture and an open-source dataset of contracts. We also provide a qualitative analysis of results.

Keywords

Large Language Models, Legal AI, Natural Language Processing, Generative AI, Foundational Language Model, Knowledge Management, Intelligence Engineering

1. Introduction

The recent wave of hype surrounding Generative AI and Large Language Models (LLMs) has captured the collective imagination of the legal industry. Following the launch of ChatGPT [1], GPT-3, and GPT-4, [2] [3] there has been considerable interest in exploring how this new wave of AI could bring transformation to the legal industry. Reports and predictions estimate that up to 40% of legal work could be displaced by AI systems [4].

Despite the newfound enthusiasm of the legal sector around AI, the practical constraints around actually doing technology in the legal sector still remain as relevant as ever. In a conservative and risk-averse industry like legal, concerns around data privacy, security and confidentiality dictate the pace of adoption, transformation and innovation [5]. There is a balance to be struck in driving innovation with technology in the sector while staying faithful to these justified concerns.

The scope of this paper is to explore whether Large Language Models could be used in law firms, and if so, how this may be realised technically. This paper attempts to provide a practical middle path to the future of AI in legal by facilitating the recent advancements around Generative AI and LLMs in a way that mitigates concerns around data privacy, security and confidentiality. In particular, we explore how open-source foundational language models can be brought inside the law firm to avoid the issues raised by sharing sensitive data with external vendors.

Rather than exploring the largest most performant models which demonstrate the state-of-the-art in terms of performance, we take a pragmatic approach and instead explore the *state-of-the-feasible* through models with parameter sizes that can be run on commodity hardware to set the floor. We test this approach using an open-source contracts dataset as a proxy for internal data within a firm and share results on this experiment.

2. AI Build vs Buy - The Options

The build vs buy distinction is actually more of a spectrum when it comes to AI. With traditional software engineering there were two extremes - either build the product yourself or go buy it from the market. Developing software is a relatively simple process that doesn't involve lots of moving parts.

AI is fundamentally different. With AI there are a number of core components involved across three different contexts. There are three main components: a) the underlying code for the algorithm, b) the data and c) the compute resources where all of these elements are combined to create the model. The three parts come together to give rise to the derived product; the model.

In: *Proceedings of the Third International Workshop on Artificial Intelligence and Intelligent Assistance for Legal Professionals in the Digital Workplace (LegalAIIA 2023),* held in conjunction with ICAIL 2023, June 19, 2023, Braga, Portugal.
uwais.iqbal@simplexico.ai (U. Iqbal)
https://www.simplexico.ai/ (U. Iqbal)
0009-0008-3384-6383 (U. Iqbal)

There are also three different stages to creating AI which are relevant to the different build vs buy options:

1. **Pre-training** - In the pre-training stage, the code for the algorithm is used with data in a compute resource to train a model.
2. **Fine-tuning** - In the fine-tuning stage, the pre-trained model is further refined with additional data using the code for the algorithm in a compute resource to create a more refined model focused for a particular domain or for a particular task.
3. **Serving** - In the serving stage, the model is served in a compute environment and packaged in an API so it can be called on by other software services and the AI capability can be consumed.

2.1. Buying AI

Buying AI can be understood as purchasing a specific AI-enabled application from a vendor. Just as one might purchase an electrical appliance like a toaster or a kettle from a supermarket. These AI-enabled applications perform particular tasks and fulfill a certain defined set of needs. Examples of such AI-enabled applications in the legal sector include offerings from vendors such as Harvey, Spellbook, Kira etc.

2.2. The AI Build Spectrum

Building AI can be understood as going one level higher to interact with the underlying technology through APIs or actual code. AI can be built through four different postures across a spectrum that spans from one extreme of a consumer posture to another extreme of a creator posture. The four postures are as follows:

1. The Consumer Posture
2. The Consumer Customiser Posture
3. The Creator Customiser Posture
4. The Creator Posture

Figure 1 outlines a visual representation of the postures across the AI build spectrum and how the different components of data, code, compute and model are distributed across open-source, vendor and internal management.

2.2.1. The Consumer Posture

In the Consumer Posture, an organisation acts as a consumer of AI and the pre-training, fine-tuning and serving stages are all managed by a vendor. The Consumer Posture enables an organisation to just get on with consuming and integrating AI into their applications. The vendor's job is to worry about training and serving the AI system. The organisation just acts as a consumer and can build applications around it. A typical use case with

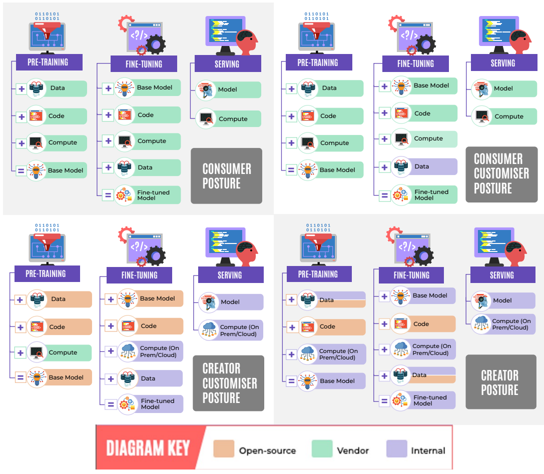

Figure 1: Visual demonstration of the various AI postures across the AI build spectrum

the consumer posture would involve plugging into APIs offering AI services such as Azure Cognitive Services or Open AI to use a model for a particular task.

The Consumer Posture is easy to get started with and no AI skills are needed. The services are all managed by the vendor so there are no technical or infrastructure concerns to worry about. However, under the Consumer Posture there is no control over the AI process so there is no effective way to mitigate bias and risk. The models available through such services tend to be too generic and general purpose to be effective and useful for specialised tasks in the legal domain.

2.2.2. The Consumer Customiser Posture

The Consumer Customiser Posture is similar to the Consumer Posture. The vendor provides fine-tuning as a service offering and allows an organisation to customise and fine-tune the underlying model with their own data. The vendor still takes care of everything from pre-training, fine-tuning and serving. A typical use with the Consumer Customiser Posture would involve fine-tuning a model with an organisation's internal data through an API service like Open AI so the model is more familiar with specific domain language.

Much like the Consumer Posture, the Consumer Customiser Posture is easy to get started with and is offered as a managed service meaning all the technical infrastructure work is taken care of by the vendor. However, this posture involves sharing internal data with a vendor so raises risks and concerns around data privacy and security. It also raises concerns and questions around the IP ownership of the fine-tuned model managed by the vendor.

2.2.3. The Creator Customiser Posture

In the Creator Customiser Posture, the posture changes from one of a consumer to a creator. In this posture, open-source pre-trained models are utilised for commercial purposes. The open-source base models are brought inside an organisation and further fine-tuned on a combination of open-source and internal data. Under this posture the pre-training is performed by another party who then makes their pre-trained model publicly available under an open-source license. The fine-tuning and serving phases are managed internally so infrastructure is needed to facilitate the compute resources in the form of either on-premise or cloud resources.

A typical use case for the Creator Customiser Posture would involve bringing an open-source foundational model in house and fine-tuning it on internal data and serving it internally for internal applications.

The Creator Customiser Posture is attractive in that it doesn't require the sharing of data with a third party. There is more control over the AI creation process so it becomes easier to mitigate for bias and risk as well as put in place controls for safety and governance. The IP of fine-tuned models is owned by the organisation. However, this posture requires infrastructure to be managed and maintained internally for the compute resources in the fine-tuning and serving phases. It also requires some level of specialised skill set around AI.

2.2.4. The Creator Posture

The Creator Posture goes a step further than the Creator Customiser Posture. Instead of relying on a third party to perform the pre-training phase to create the pre-trained model, in the Creator Posture the pre-training phase is brought inside the organisation. There is no dependence on a third party and all three phases of pre-training, fine-tuning and serving are managed internally.

The Creator Posture is the polar opposite of the Consumer Posture and sits on the other end of the AI build spectrum. There is complete control over the process of creating AI so there the ability to have full provenance of data. There is a complete transparency and the digital footprint of the models can be traced back to their origins. Mitigating for bias, risk and putting in place controls for safety and governance is made much more accessible.

Just like the Creator Customiser Posture, a specialised skillset around AI is needed and the infrastructure for pre-training, fine-tuning and serving all have to be managed internally.

These four postures present the different ways an organisation can go about *building* AI. The options of buying AI and the four postures of building AI are not mutually exclusive. An organisation can develop a complimentary strategy across buying and the four postures of building to explore general purpose AI applications as well as building their own AI systems with internal data.

3. The State of Affairs - Open-source Large Language Models

Over recent months, there has been something of a revolutionary movement in the open-source community. Since the release of ChatGPT, the open-source community have been active in replicating the capabilities of closed-source models. The data and hardware requirements for creating foundational language models were previously significant barriers to entry. Only organisations in a privileged position could create such foundational language models and then proceed onto the later phase of development. The ability to create such models were in the hands of the few.

The recent wave of activity in the open-source community has significantly changed this dynamic. With the release of open-source foundational model suites like LLaMa [6], pythia [7], Cerebras-GPT [8], StabilityLM [9], and MPT-7B [10] these base models are now publicly accessible. The ability to customise models and continue through phases of development are in the hands of everyone. While some of the open-source base models have been released under licenses only permitting academic use, some models are available for commercial use. Table 1 outlines the licenses for some of these open-source foundational models and whether they are available for commercial use.

Name	Provider	Parameters	License	Commercial Use
LLaMa	Facebook Research	65B	Academic Use Only	No
pythia	EleutherAI	70M-12B	Apache 2.0	Yes
Cerebras-GPT	Cerebras	111M-13B	Apache 2.0	Yes
StabilityLM	StabilityAI	3B/7B	CC BY-SA 4.0	Yes
MPT-7B	MosaicML	7B	Apache 2.0	Yes

Table 1
Open-source foundational language models with their parameter sizes and licenses

The allure of open-source models for law firms is that they can be used with the Creator Customiser Posture.

These models can be brought inside the law firm and developed further on internal data mitigating the risks around data privacy and security while still enabling access to cutting-edge technology. These developments around open-source foundational language models now mean that AI can be brought inside the law firm to build AI systems that power use cases inside the law firm.

4. Layers of Fine-tuning - Customising Open-source Language Models

An open-source foundational language model can be further fine-tuned to customise the model with domain-specific and internal data. There are three distinct layers of fine-tuning that can be performed with language models.

1. **Unsupervised Fine-Tuning** - With a domain-specific corpus of raw unstructured text, the language model can be fine-tuned to learn the particular nuances and quirks of legal language. This can be made even more specific by focusing on a particular practice area or a particular area of law. The data requirements for unsupervised fine-tuning are not restrictive. All that is needed is a corpus of raw text documents which should be relatively easy to find within a law firm. The data does not even need to be structured as only the raw text is needed.
2. **Instruction Response Fine-tuning** - Recasting structured data into tasks consisting of instruction response pairs results in language models being able to generalise across unseen tasks really well [11]. Table 4 in the appendix shows some examples of instruction response pairs. The base language model can be fine-tuned on pairs of instructions and responses to create an ability to follow particular instructions and commands [12]. This creates an ability for the model to generalise across different tasks by learning how to complete the instruction on unseen data in a zero-shot fashion. There are some requirements around the data for instruction response fine-tuning since the data has to be structured from raw text into instruction response pairs.
3. **Reinforcement Human Learning Feedback (RHLF)** - The third phase involves creating a more human-like interface to the language model by using Reinforcement Learning to teach the model how to converse as a human [13]. This creates a conversational layer with the language model that allows it to be interacted with as a chat bot.

The first two levels of fine-tuning create domain-specific functionality while the last is more aesthetic. These layers of fine-tuning can be performed with a combination of internal and open-source datasets for maximum learning. There are already a growing number of such open-source datasets for instruction response fine-tuning and RHLF available for commercial use [14]. Open-source datasets can be combined with internal datasets and *stacked* in a modular fashion to create the desired intelligence capabilities within the language model.

4.1. From Knowledge Management to Intelligence Engineering

The Instruction Response fine-tuning approach is of particular importance for law firms. Knowledge management (KM) can be defined as the "tools, techniques, and strategies to retain, analyse, organise, improve, and share business experience" [15]. Within the context of a law firm, knowledge management involves "a firm's ability to identify, capture, and leverage the internal knowledge of individuals" to "enhance the ability of all law firm staff to create and share knowledge across the firm and to provide excellent client services and to compete in an increasingly aggressive professional legal services environment" [16].

Knowledge management is based on three fundamental concepts: a) data, b) information, and c) knowledge [17]. Data is understood as the raw resource without context. Information is understood as data with context that is able to provide value. Knowledge is understood as information combined with understanding and capability. The distinction between knowledge and information can be clarified as "knowledge being a personal subjective process emerging from previous experiences, while information is objective data about the environment" [18]. Knowledge lives in the minds of people and is anthropomorphic while information is not.

Knowledge can further be broken down into two main types: a) Tacit Knowledge, and b) Explicit Knowledge [15]. Tacit knowledge refers to personal knowledge embedded in individual experience while explicit knowledge refers to tacit knowledge that has been documented. One of the challenges in knowledge management is the difficulty in capturing tacit knowledge. One of the key functions of a knowledge management strategy is to make the tacit explicit so that it can be easily transferred and communicated from one individual to another.

We can introduce a fourth related concept of *intelligence* to the fundamental concepts of knowledge management. Intelligence can be defined as the ability to acquire knowledge and skills. As such, intelligence can be possessed by humans as natural intelligence and by machines in the form of artificial intelligence (AI). By including intelligence within the fundamental concepts of knowledge management, AI can be adopted to achieving and enhance the objectives of knowledge management.

Knowledge management within law firms has traditionally focused on capturing the tacit knowledge from the minds of highly skilled and experienced individuals into explicit knowledge in the form of written content. The explicit knowledge captured in content allows this expertise to be shared with and accessed by other individuals in the firm. However, one of the practical challenges with knowledge management is how to make this explicit knowledge easily accessible, retrievable and consumable to other individuals within the organisation.

Traditionally knowledge management seeks to take tacit knowledge from the mind of a human and capture it as explicit knowledge in the form of written content so that it can be consumed by another human. The process of knowledge transfer goes from human (tacit knowledge) to content (explicit knowledge) to human (tacit knowledge). With the advent of AI, large language models and instruction response fine-tuning, we propose *intelligence engineering* as an additional step in the process. Rather than going from human to content to human, we propose introducing a machine into the process; going from human (tacit knowledge) to content (explicit knowledge) to machine (intelligence) to human (tacit knowledge).

Practically, this would involve much of the same processes around knowledge management as before but with a few additional steps. Tacit knowledge from the mind of a human can be captured as explicit knowledge in the form of an instruction response dataset. Existing explicit knowledge content can easily be restructured into an instruction response format. Such a dataset can be used to fine-tune a large language model with the Creator Customiser posture to create protected and privileged intelligence within a law firm. Then, individuals in the firm can interface with the large language model to retrieve, access and query the intelligence captured. This additional step of *intelligence engineering* removes existing bottlenecks around accessing and retrieving explicit knowledge. In the context of knowledge management within a law firm, a large language model effectively creates infinite scale in providing access to explicit knowledge. While explicit knowledge has always been static in the form of content, large language models transform this content to intelligence that is dynamic, scalable and easily accessible.

We explore and evaluate this approach by layering unsupervised fine-tuning and instruction fine-tuning with the Creator Customiser Posture using an open-source foundational language model and a publicly available dataset of contracts.

5. Experimental Results and Findings

5.1. Experimental Setup

To demonstrate how an open-source foundational language model can be leveraged inside a law firm we practically demonstrate how the Creator Customiser Posture can be used with layered unsupervised and instruction response fine-tuning.

We first select a base foundational model. The Cerebras-GPT model suite contains 7 GPT-3 models ranging from 111M up to 13B in parameter size [8]. The models were created by training on The Pile dataset [19]. These models are licensed under the Apache 2.0 license and are available for commercial use.

Taking a pragmatic approach we focus on the smaller parameter size variants of the Cerebras-GPT model suite. In particular we work with the 590M parameter variant. Our goal is to provide a pragmatic demonstration of the approach rather than optimising for performance.

In order to proxy law firm internal data, we use two open-source datasets related to contracts. Both are in the public domain and are available under the CC-BY 4.0 license. These datasets are as follows:

1. **Contract Understanding Atticus Dataset (CUAD)** [20] - The CUAD dataset consists of 510 commercial legal contracts with over 13,000+ labels that have been manually labelled under the supervision of experienced lawyers. The annotations identify legal clauses that are considered important in contract review in connection with a corporate transaction, including mergers acquisitions, etc.
2. **Merger Agreement Understanding Dataset (MAUD)** [21] - The MAUD dataset consists of 152 merger agreements with over 47,000+ labels that have been manually labeled under the supervision of experienced lawyers to identify 92 questions in each agreement used by the 2021 American Bar Association (ABA) Public Target Deal Points Study.

We combine the raw text from the contracts and agreements in the CUAD and MAUD datasets to create a dataset for unsupervised fine-tuning. The resulting combined dataset consists of 662 documents with 1.8M tokens. This dataset acts as a proxy for documents and unstructured text which may sit inside a document management system at a law firm.

Taking inspiration from FLAN [11], we mine the labels in the CUAD dataset to produce a collection of 8,000+ instruction response pairs. The pairs span a number of legal specific tasks including drafting, classification and extraction. This dataset acts as a proxy for internal data contained in a precedent bank or within explicit knowledge content and practice notes. With a little mining, any internal structured textual data can be reconstructed in the form of instruction response pairs for use with large language models. Table 4 in the appendix shows examples of these instruction response pairs created from the CUAD dataset. We perform the experiments on commodity hardware in the form of a single Nvidia A100 GPU with 80GB of RAM.

5.2. Results

Using the base model, we perform two stages of fine-tuning:

1. **Unsupervised Fine-Tuning** - Using the dataset of 1.8M tokens created by combining the raw text from the CUAD and MAUD datasets, we fine-tune the model in an unsupervised manner. Through this stage the model learns the nuances of legal language.
2. **Instruction Fine-Tuning** - Using the 8,000+ instruction response pairs created from mining the CUAD data, we further fine-tune the model with these pairs. The model learns how to perform these legal specific tasks.

For both datasets we use an 80:20 split to create the training and testing sets. Using the 590M parameter variant from the Cerebras-GPT model suite as the base model, we run the unsupervised fine-tuning with the combined text from the CUAD and MAUD datasets.

We compare the quality of the fine-tuned language models by using perplexity as an intrinsic evaluation metric. The perplexity score captures the average number of words that can be encoded. In more concrete terms, a perplexity score of 4 means that when trying to guess the next word, the model is as confused as if it had to pick between 4 different words. A lower perplexity score means that the language model is more precise at predicting words and is better.

We report the following results after 3 epochs of unsupervised fine-tuning:

Perplexity on Test Set (Before fine-tuning)	8.33
Train Loss	1.64
Test Loss	1.54
Perplexity on Test Set (After fine-tuning)	4.68
Duration	66 minutes
Cost	$6.10

Table 2
Results after performing unsupervised fine-tuning for 3 epochs using the combined text from the CUAD and MAUD datasets

We take the fine-tuned model and use it to further perform instruction fine-tuning using the 8,000+ instruction response pairs mined from the CUAD dataset. We report the following results after 3 epochs of instruction response fine-tuning:

Perplexity on Test Set (Before fine-tuning)	14.09
Train Loss	1.45
Test Loss	1.23
Perplexity on Test Set (After fine-tuning)	3.43
Duration	30 minutes
Cost	$3.05

Table 3
Results after performing instruction response fine-tuning for 3 epochs using the 8,000+ instruction response pairs mined from the CUAD dataset.

Table 5 in the appendix details some sample outputs from the fine-tuned model with their associated prompts.

5.3. Findings

The significant difference in perplexity scores before and after fine-tuning indicate that massive amounts of domain-specific data isn't needed to effectively fine-tune open-source models. This also demonstrates that fine-tuning can be performed with reasonable volumes of data at a reasonable cost in a reasonable time frame. We can estimate order of magnitude data requirements for the different fine-tuning layers. For unsupervised fine-tuning on the order of hundreds of documents are needed while for instruction response fine-tuning on the order of thousands of instruction response pairs are needed.

We posit that since the language in the domain is more specific and standardised, it is easier for the model to learn the nuances of legal language. As opposed to generic pre-training datasets like Pile [19], there is less variability in the language so it is easier for the model to learn. Further work is needed to evaluate and quantify the effectiveness of such fine-tuned large language models by way of extrinsic evaluation to quantify performance on downstream tasks. Further investigation is also needed to understand the effect of parameter size on the performance of fine-tuning on legal domain specific language and tasks.

The results also indicate that fine-tuning open-source foundational models is feasible and practical from a variety of operational perspectives: data, cost and time. The data requirements are not unreasonable - most law firms have access to thousands of documents that can be used for unsupervised fine-tuning. Following our approach, internal knowledge content, structured databases and practice notes can be mined to create instruction response pairs for instruction response fine-tuning. The cost of running such fine-tuning experiments is very cheap which means that the operational expense of experimentation is not a barrier to innovation. The time requirements for performing fine-tuning are not prohibitive to limit rapid prototyping and iterative development.

The approach outlined by taking the Creator Customiser Posture with open-source foundational models and performing unsupervised and instruction-response fine-tuning is a methodology that can be adopted for intelligence engineering to create an intelligence layer to make explicit knowledge more accessible so it can be readily be consumed by other humans. In effect, this methodology can immediately unlock value by advancing knowledge management objectives within law firms through intelligence engineering.

6. Conclusion

We have introduced the various options available when it comes to the buy vs build question for AI. We have outlined the four postures across the AI build spectrum and demonstrated that the Creator Customiser posture is the most appealing for law firms looking to leverage internal data while mitigating risks around data privacy and security. We have assessed the opportunities open-source foundational language models present and outlined the various ways in which these models can be further refined on internal data. We have also presented a practical way and method in which large language models can be introduced into knowledge management workflows in law firms through intelligence engineering.

We have carried out an early evaluation of this approach using an open-source foundational language model and an open-source contract dataset. While the experimental analysis in this paper is limited to a qualitative analysis of results, early findings indicate the value and technical feasibility of this approach in unlocking opportunities for building AI within the law firm. Further experimentation is required to quantitatively evaluate this approach on downstream tasks to provide an objective assessment of performance. The impact of larger parameter sizes on downstream performance is an area for further investigation.

References

[1] Y. Liu, T. Han, S. Ma, J. Zhang, Y. Yang, J. Tian, H. He, A. Li, M. He, Z. Liu, Z. Wu, D. Zhu, X. Li, N. Qiang, D. Shen, T. Liu, B. Ge, Summary of chatgpt/gpt-4 research and perspective towards the future of large language models, 2023. arXiv:2304.01852.

[2] OpenAI, Gpt-4 technical report, 2023. arXiv:2303.08774.

[3] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, D. Amodei, Language models are few-shot learners, 2020. arXiv:2005.14165.

[4] J. Hatzius, J. Briggs, D. Kodani, The potentially large effects of artificial intelligence on economic growth, 2023. URL: https://www.key4biz.it/wp-content/uploads/2023/03/Global-Economics-Analyst_-ThePotentially-Large-Effects-of-Artificial-Intelligence-on-Economic-Growth-Briggs_Kodnani.pdf.

[5] W. D. Henderson, Innovation diffusion in the legal industry, Dickinson L. Rev. 122 (2017) 395.

[6] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, A. Rodriguez, A. Joulin, E. Grave, G. Lample, Llama: Open and efficient foundation language models, 2023. arXiv:2302.13971.

[7] S. Biderman, H. Schoelkopf, Q. Anthony, H. Bradley, K. O'Brien, E. Hallahan, M. A. Khan, S. Purohit, U. S. Prashanth, E. Raff, A. Skowron, L. Sutawika, O. van der Wal, Pythia: A suite for analyzing large language models across training and scaling, 2023. arXiv:2304.01373.

[8] N. Dey, G. Gosal, Zhiming, Chen, H. Khachane, W. Marshall, R. Pathria, M. Tom, J. Hestness, Cerebras-GPT: Open Compute-Optimal Language Models Trained on the Cerebras Wafer-Scale Cluster, 2023. URL: http://arxiv.org/abs/2304.03208. doi:10.48550/arXiv.2304.03208, arXiv:2304.03208 [cs].

[9] StabilityAI, Stability ai launches the first of its stablelm suite of language models, 2023. URL: https://stability.ai/blog/stability-ai-launches-the-first-of-its-stablelm-suite-of-language-models, accessed: 2023-05-17.

[10] M. N. Team, Introducing mpt-7b: A new standard for open-source, ly usable llms, 2023. URL: www.mosaicml.com/blog/mpt-7b, accessed: 2023-05-17.

[11] J. Wei, M. Bosma, V. Y. Zhao, K. Guu, A. W. Yu, B. Lester, N. Du, A. M. Dai, Q. V. Le, Finetuned Language Models Are Zero-Shot Learners, 2022. URL: http://arxiv.org/abs/2109.01652. doi:10.48550/arXiv.2109.01652, arXiv:2109.01652 [cs].

[12] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. L. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, J. Schulman, J. Hilton, F. Kelton, L. Miller, M. Simens, A. Askell, P. Welinder, P. Christiano, J. Leike, R. Lowe, Training language models to

follow instructions with human feedback, 2022. URL: http://arxiv.org/abs/2203.02155. doi:10.48550/arXiv.2203.02155, arXiv:2203.02155 [cs].
[13] Y. Bai, A. Jones, K. Ndousse, A. Askell, A. Chen, N. DasSarma, D. Drain, S. Fort, D. Ganguli, T. Henighan, N. Joseph, S. Kadavath, J. Kernion, T. Conerly, S. El-Showk, N. Elhage, Z. Hatfield-Dodds, D. Hernandez, T. Hume, S. Johnston, S. Kravec, L. Lovitt, N. Nanda, C. Olsson, D. Amodei, T. Brown, J. Clark, S. McCandlish, C. Olah, B. Mann, J. Kaplan, Training a helpful and harmless assistant with reinforcement learning from human feedback, 2022. arXiv:2204.05862.
[14] Databricks, Free dolly: Introducing the world's first truly open instruction-tuned llm, 2023. URL: https://www.databricks.com/blog/2023/04/12/dolly-first-open-commercially-viable-instruction-tuned-llm, accessed: 2023-05-17.
[15] T. Groff, T. Jones, Introduction to Knowledge Management: KM in Business, Taylor & Francis Group, 2003. URL: https://books.google.co.uk/books?id=pmJv09aw4rMC.
[16] T. du Plessis, A. du Toit, Knowledge management and legal practice, International Journal of Information Management 26 (2006) 360–371. URL: https://www.sciencedirect.com/science/article/pii/S0268401206000636. doi:https://doi.org/10.1016/j.ijinfomgt.2006.06.003.
[17] A. Terrett, Knowledge management and the law firm, Journal of knowledge management (1998).
[18] J. Roos, L. Edvinsson, N. C. Dragonetti, Intellectual capital: Navigating the new business landscape, Springer, 1997.
[19] L. Gao, S. Biderman, S. Black, L. Golding, T. Hoppe, C. Foster, J. Phang, H. He, A. Thite, N. Nabeshima, S. Presser, C. Leahy, The pile: An 800gb dataset of diverse text for language modeling, 2020. arXiv:2101.00027.
[20] D. Hendrycks, C. Burns, A. Chen, S. Ball, Cuad: An expert-annotated nlp dataset for legal contract review, 2021. arXiv:2103.06268.
[21] S. H. Wang, A. Scardigli, L. Tang, W. Chen, D. Levkin, A. Chen, S. Ball, T. Woodside, O. Zhang, D. Hendrycks, Maud: An expert-annotated legal nlp dataset for merger agreement understanding, 2023. arXiv:2301.00876.

A. Examples of Instruction Response Data

The following table provides some examples from the instruction response dataset that was created by mining the CUAD dataset.

Instruction	Response	Instruction Type
Draft a renewal term clause for a marketing affiliate agreement between Birch First Global Investments Inc. ("Company") and Mount Knowledge Holdings Inc. ("Marketing Affiliate", "MA") with a renewal term of successive 1 year	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 provisions contained in paragraph 18 of this agreement.	Drafting
Classify the following clause as either a renewal term clause or a notice period to terminate renewal clause or a governing law clause .. or a covenant not to sue clause or a third party beneficiary clause: "This agreement may be terminated by either party at the expiration of its term or any renewal term upon thirty (30) days written notice to the other party."	This is a notice period to terminate renewal clause	Classification
Extract the jurisdiction of the governing law of the agreement from the following text: "This Agreement shall be governed by, and any dispute arising hereunder shall be determined in accordance with, the laws of State of New York"	New York	Extraction

Table 4
Examples of the instruction response data created from the CUAD dataset

B. Example Outputs of Model

The following table outlines examples of outputs from the models with the associated prompts:

Model Input	Model Output
Extract the jurisdiction of the governing law of the agreement from the following text: This Agreement shall be governed by, and construed in accordance with, the Laws of the State of New York, applicable to contracts executed in and to be performed entirely within that state.	New York
Draft a notice period to terminate renewal clause for a video-on-demand content license agreement between Rogers Cable Communications Inc. ("rogers") and Euromedia Holdings Corp. ("licensor") with a notice period of 60 days	Subject to the terms and conditions of this agreement, during the term hereof, and for a period of one (1) year thereafter, Rogers shall have the right to terminate this agreement at any time upon thirty (30) days written notice to the other party.

Table 5
Examples of generated output from the fine-tuned model