

## Introduction

### Can computational linguistics be used to accurately answer legal questions?

We can test this by training Llama 2 on real estate law text files and comparing the accuracy of the model's answers before and after its fine-tuning. If the accuracy is significantly greater post-training, this experiment demonstrates the potential of neural networks in the legal field.

Ideal Chat Example:

Q: What can happen to the real property if there are unpaid taxes or special ad valorem levies?

A: The real property can be liable for sale pursuant to law.

We hypothesize that a language model trained on New York law will be able to answer legal questions with more accuracy than the base model.



## Background

There's currently no widely available legal chat bot for official New York Law. The law consists of long, hundred page articles which are scarcely read outside of the legal field.

Data is widely available and many legal questions and answers are accessible online, however a chat-bot which can answer these questions will save the layman time in seeking legal advice.

Information is important for people who rent, buy, sell, or lend in regards to the real estate sector— a lot of people violate the law without knowing!

Core training:

1. Official NY Real Estate Law
2. Instruction set derived from NY Real Estate Law
3. Q&A from certified lawyer forums

**RPPL - Real Property**

Article 1 (1-2) SHORT TITLE; DEFINITIONS  
Article 2 (10 - 18) TENURE OF REAL PROPERTY  
Article 3 (119 - 125) USES AND TRUSTS  
Article 4-A (124 - 130-K) TRUSTS; INTERESTS AND INTERESTS THEREIN  
Article 5 (130 - 139) DEEDS AND CONVEYANCES  
Article 6 (139 - 144-A) Easements and Tenants  
Article 7 (144 - 149) Easements and Tenants  
Article 8 (149 - 154) Easements and Tenants  
Article 9 (154 - 159) Easements and Tenants  
Article 10 (159 - 164) Easements and Tenants  
Article 11 (164 - 169) Easements and Tenants  
Article 12 (169 - 174) Easements and Tenants  
Article 13 (174 - 179) Easements and Tenants  
Article 14 (179 - 184) Easements and Tenants  
Article 15 (184 - 189) Easements and Tenants  
Article 16 (189 - 194) Easements and Tenants  
Article 17 (194 - 199) Easements and Tenants  
Article 18 (199 - 204) Easements and Tenants  
Article 19 (204 - 209) Easements and Tenants  
Article 20 (209 - 214) Easements and Tenants  
Article 21 (214 - 219) Easements and Tenants  
Article 22 (219 - 224) Easements and Tenants  
Article 23 (224 - 229) Easements and Tenants  
Article 24 (229 - 234) Easements and Tenants  
Article 25 (234 - 239) Easements and Tenants  
Article 26 (239 - 244) Easements and Tenants  
Article 27 (244 - 249) Easements and Tenants  
Article 28 (249 - 254) Easements and Tenants  
Article 29 (254 - 259) Easements and Tenants  
Article 30 (259 - 264) Easements and Tenants  
Article 31 (264 - 269) Easements and Tenants  
Article 32 (269 - 274) Easements and Tenants  
Article 33 (274 - 279) Easements and Tenants  
Article 34 (279 - 284) Easements and Tenants  
Article 35 (284 - 289) Easements and Tenants  
Article 36 (289 - 294) Easements and Tenants  
Article 37 (294 - 299) Easements and Tenants  
Article 38 (299 - 304) Easements and Tenants  
Article 39 (304 - 309) Easements and Tenants  
Article 40 (309 - 314) Easements and Tenants  
Article 41 (314 - 319) Easements and Tenants  
Article 42 (319 - 324) Easements and Tenants  
Article 43 (324 - 329) Easements and Tenants  
Article 44 (329 - 334) Easements and Tenants  
Article 45 (334 - 339) Easements and Tenants  
Article 46 (339 - 344) Easements and Tenants  
Article 47 (344 - 349) Easements and Tenants  
Article 48 (349 - 354) Easements and Tenants  
Article 49 (354 - 359) Easements and Tenants  
Article 50 (359 - 364) Easements and Tenants  
Article 51 (364 - 369) Easements and Tenants  
Article 52 (369 - 374) Easements and Tenants  
Article 53 (374 - 379) Easements and Tenants  
Article 54 (379 - 384) Easements and Tenants  
Article 55 (384 - 389) Easements and Tenants  
Article 56 (389 - 394) Easements and Tenants  
Article 57 (394 - 399) Easements and Tenants  
Article 58 (399 - 404) Easements and Tenants  
Article 59 (404 - 409) Easements and Tenants  
Article 60 (409 - 414) Easements and Tenants  
Article 61 (414 - 419) Easements and Tenants  
Article 62 (419 - 424) Easements and Tenants  
Article 63 (424 - 429) Easements and Tenants  
Article 64 (429 - 434) Easements and Tenants  
Article 65 (434 - 439) Easements and Tenants  
Article 66 (439 - 444) Easements and Tenants  
Article 67 (444 - 449) Easements and Tenants  
Article 68 (449 - 454) Easements and Tenants  
Article 69 (454 - 459) Easements and Tenants  
Article 70 (459 - 464) Easements and Tenants  
Article 71 (464 - 469) Easements and Tenants  
Article 72 (469 - 474) Easements and Tenants  
Article 73 (474 - 479) Easements and Tenants  
Article 74 (479 - 484) Easements and Tenants  
Article 75 (484 - 489) Easements and Tenants  
Article 76 (489 - 494) Easements and Tenants  
Article 77 (494 - 499) Easements and Tenants  
Article 78 (499 - 504) Easements and Tenants  
Article 79 (504 - 509) Easements and Tenants  
Article 80 (509 - 514) Easements and Tenants  
Article 81 (514 - 519) Easements and Tenants  
Article 82 (519 - 524) Easements and Tenants  
Article 83 (524 - 529) Easements and Tenants  
Article 84 (529 - 534) Easements and Tenants  
Article 85 (534 - 539) Easements and Tenants  
Article 86 (539 - 544) Easements and Tenants  
Article 87 (544 - 549) Easements and Tenants  
Article 88 (549 - 554) Easements and Tenants  
Article 89 (554 - 559) Easements and Tenants  
Article 90 (559 - 564) Easements and Tenants  
Article 91 (564 - 569) Easements and Tenants  
Article 92 (569 - 574) Easements and Tenants  
Article 93 (574 - 579) Easements and Tenants  
Article 94 (579 - 584) Easements and Tenants  
Article 95 (584 - 589) Easements and Tenants  
Article 96 (589 - 594) Easements and Tenants  
Article 97 (594 - 599) Easements and Tenants  
Article 98 (599 - 604) Easements and Tenants  
Article 99 (604 - 609) Easements and Tenants  
Article 100 (609 - 614) Easements and Tenants

## Experimental setup

Compiled raw text from the official website for New York Law: Sections Real Property, Real Property Actions and Proceedings, Real Property Tax. (34k rows)

Fed batches of 20-25 pages of text to GPT 3.5-Turbo (under the 4000 token limit) to create our instruction dataset. 80% Train, 20% Test. (13.5k rows)

Model: Llama2-7B  
Evaluation: Rouge and BLEU scores



## Results

Currently experiencing difficulties regarding the size of the model and the (literal) cost of training it.

## Summary + Conclusion

We have a LLM capable of answering questions regarding NY State Real Estate Law!

Some limitations:

- Hallucinations: very costly if this project was ever going to go live. We could get sued for bad hallucinations.
- A quick glance at our instruction set reveals the rigidity of the answers in training. We will likely need to add a layer of reinforcement learning.
- Currently only trained on a small section of NY Real Estate Law
- Scraped data has some problems. After cleaning, it is a very small dataset.

## Future directions

1. Train on the remaining articles of NY State Real Estate Law (raw text and instruction set)
2. Train on city and county level data in NY — append each question and answer with the relevant county and city name.
3. Try other small, specialized models for Q&A (Mistral)
4. Find ways to cut down on hallucinations (RAG?)
5. Replicate the process for other states
6. Build a LangChain app to allow users to upload documents, summarize sections of the document, ask specific questions about the document.

## References

Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., Bikel, D., Blecher, L., Ferrer, C. C., Chen, M., Cucurull, G., Esiobu, D., Fernandes, J., Fu, J., Fu, W., ... Scialom, T. (2023, July 19). Llama 2: Open Foundation and fine-tuned chat models. arXiv.org. <https://arxiv.org/abs/2307.09288>