AIEdison: Generative AI for Augmented Invention

Generating and Evaluating Patent Claims Using Large Language Models Fine-Tuned with Selected Patents and Patent Applications for Augmented Invention

黃楷鈞 Kai-Jun Huang 台北市懷生國中 Taipei Municipal HuaiSheng Junior High School

Advisor: Jieh-Sheng Lee, Assistant Professor, National Yang Ming Jiao Tung University

Keywords: Large Language Models, Patents, Claim Languages, Technical Features, Augmented Invention

Project Goal

AIEdison is a large language models system for augmented invention, which means using generative AI to help innovators to invent good solutions.

Motivation

▪Large language models (LLMs) such as GPT (Generative Pre-Trained Transformer) have made tremendous progress and generated great applications for real world – This project uses it to create AIEdison for Augmented Invention.

▪ Most inventions made in modern history have been applied for patents, for example, according to the World Intellectual Property Organization more than 3.3M patent applications were filed last year.

▪ Over the years, we have enormous volume of patent language documents, which are ideal for large language models training and fine-tuning.

▪ I start with GPT-2 since it is both open source and free, but I plan to use larger models such as GPT-3 later to further improve AIEdison.

Precise Industry and Technical Focus

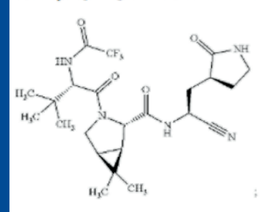
▪ Previously researchers have used general patent data to fine-tune & pre-train large language models.

▪ In contrast, I use patent data with precise technical focus to fine-tune GPT-2 since all inventors have specific technical field focus and irrelevant patents can be noise or toxic to their desired results.

▪ For example, autonomous vehicle patents are strikingly different from pharma patents in both language and patentability requirements. (Toyota vs. Pfizer)

1. A method of a vehicle including an Advanced Driver Assistance System ("ADAS system") and a communication unit, the method comprising:
receiving, by the communication unit of the vehicle, a wireless message from a wireless network, wherein the wireless message includes optimization settings data describing how to modify an operation of the ADAS system of the vehicle based on (1) one or more preferences of a first user for the operation of the ADAS system and (2) a characteristic behavior of the ADAS system as determined based on one or more digital simulations including a digital version of the vehicle, wherein the wireless message is transmitted to the vehicle responsive to the first user reserving the vehicle for their use; and
modifying, by a processor of the vehicle, one or more control parameters of the ADAS system based on the optimization settings data so that the operation of the ADAS system conforms with the one or more preferences of the first user.

1. The compound (1R,2S,5S)-N-{(1S)-1-Cyano-2-[(3S)-2-oxopyrrolidin-3-yl]ethyl}-6,6-dimethyl-3-[3-methyl-N-(trifluoroacetyl)-L-valyl]-3-azabicyclo[3.1.0]hexane-2-carboxamide having the structure



or a solvate or hydrate thereof.

Research Method – Software & Hardware

Some key information of how I conduct the experiments and create AIEdison:
I use GPT-2 Medium (345M parameters), which is open source and free from Hugging Face's transformers library.
GPT-2 has been pre-trained with a 40GB dataset called WebText by OpenAI.
I fine-tune GPT-2 with 2,000, 12,000 & 40,000 sets of Toyota Patents & Applications dated between September 1, 2012 to 2022 & I use claim 1 text only because it is the most important claim of all.
My fine-tuning tasks are done on GPU provided by Google Colab Pro+ and I set Epochs: 5, Learning Rate: 5e-4, 3 Outputs per Input, Maximum Output at 300 tokens

Research Method – Quantitative Assessment

Randomly select 3 Toyota patents & 3 applications (a total of 6) that were newly published and not included in our 3 fine-tuning datasets.
Use the beginning text of claim 1 from these 6 selected patents/application as the prompts for input into the fine-tuned models of 2,000, 12,000 and 40,000, and produce 6 results.
Conduct three rounds of the tests using the same prompts as input then calculate the average word number of three outputs per test to get a total of 18 results for comparison.

US 20220297722
US 20220301319
US 20220301203
US 11447092
US 11451974
US 11451693

Results - Average Output Word Counts

ROUND 1	2000	12000	40000
US 20220297722	108	153	235
US 20220301319	118	127	172
US 20220301203	147	205	140
US 11447092	119	126	252
US 11451974	108	213	220
US 11451693	107	106	122
AVG.	117.8	154.0	190.2

ROUND 2	2000	12000	40000
US 20220297722	136	153	185
US 20220301319	85	127	117
US 20220301203	166	205	218
US 11447092	116	126	237
US 11451974	147	213	205
US 11451693	92	106	121
AVG.	123.7	154.0	180.0

ROUND 3	2000	12000	40000
US 20220297722	122	165	193
US 20220301319	146	152	191
US 20220301203	110	142	171
US 11447092	173	126	211
US 11451974	98	183	144
US 11451693	119	112	182
AVG.	128.0	146.7	182.0

Comparing Averages Between Datasets

AVG.	2000	12000	40000
ROUND 1	117.8	154.0	190.2
ROUND 2	123.7	157.7	180
ROUND 3	128.0	146.7	182
AVG.	123.2	152.8	184.1

Quantitative Results

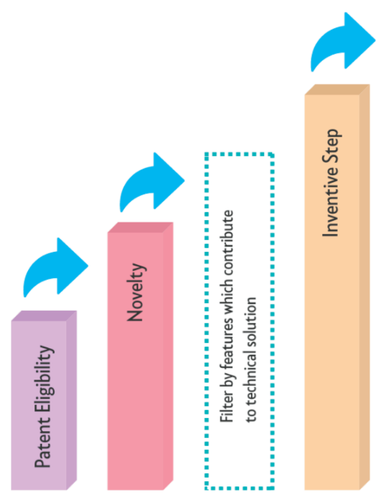
I count the total number of words and technical features in claims generated by GPT-2 fine-tuned with different size of datasets (2,000, 12,000 and 40,000).

▪ I found that larger the fine-tune dataset, longer the generated claims (15/18 - 83.3%).

▪ I also found that the longer generated claims contain more technical features than the shorter claims (83.3%).

▪ Technical features are technical elements that determine the contribution made to the prior art by the invention (inventiveness).

▪ New technical features generated by GPT-2 may become a source of inspiration for inventors (inventive ness and helpfulness).



Qualitative Survey

▪ However, it is difficult to assess qualitatively readability (whether the generated claims are sensible), inventiveness (whether the technical features in the generated claims are creative) and helpfulness (whether the technical features generated are inspiration).

▪ Thus, I design quality tests asking three experienced patent lawyers to review all the claims generated by GPT-2 fine-tuned with 2,000, 12,000 and 40,000 datasets in 3 rounds of tests.

▪ Each patent lawyer rate all the claims in terms of readability, inventiveness and helpfulness with a score from 1 to 10 one by one.

▪ Claims generated by GPT-2 fine-tuned with larger datasets score higher on readability (72.2%), inventiveness (66.7%) and helpfulness (77.8%).

Results and Further Improvements

▪ Quantitative result shows that GPT-2 fine-tuned with larger datasets generate longer claims containing more technical features (83.3%).

▪ Qualitative survey shows that claims generated by GPT-2 fine-tuned with larger datasets score higher on readability (77.2%), inventiveness (66.7%) and helpfulness (77.8%).

▪ Next improvement steps are fine-tuning with even larger datasets of up to 100,000 by including vehicle companies other than Toyota and using bigger and more powerful large language models such as GPT-3.

▪ I hope that AIEdison using large language models fine-tuned by patents with a precise technical focus will help innovators in the technical field to invent good solutions just like Thomas Edison's research lab helped innovators to produce great inventions.

References

[1] Farhan, Rey. "GPT-2 Fine-Tuning Tutorial with PyTorch & Huggingface in Colab." colab.research.google.com/drive/13dZVYEOMhXhkXWfvSMVM1TTtUDrT6Aeh.
[2] Lee, Jieh-Sheng, and Jieh Hsiang. "Patent Claim Generated by Fine-Tuning OpenAI GPT-2." World Patent Information - Elsevier, 2020, https://doi.org/10.48550/arXiv.1907.02052.
[3] Lee, Jieh-Sheng. "Evaluating Generative Patent Language Models." arXiv preprint arXiv:2206.14578, 2022, https://doi.org/10.48550/arXiv.2206.14578.
[4] Radford, Alec. "Better Language Models and Their Implications." OpenAI, 14 February 2019, openai.com/blog/better-language-models.
[5] Patent Public Search. United States Patent and Trademark Office, www.uspto.gov/patft/index.html.