

自然言語処理に関する最近の話題

2023年11月9, 10日

中央大学工学部ビジネスデータサイエンス学科
難波英嗣

講演内容

- 学術論文への国際特許分類の自動付与
- 自然言語処理研究の動向

講演内容

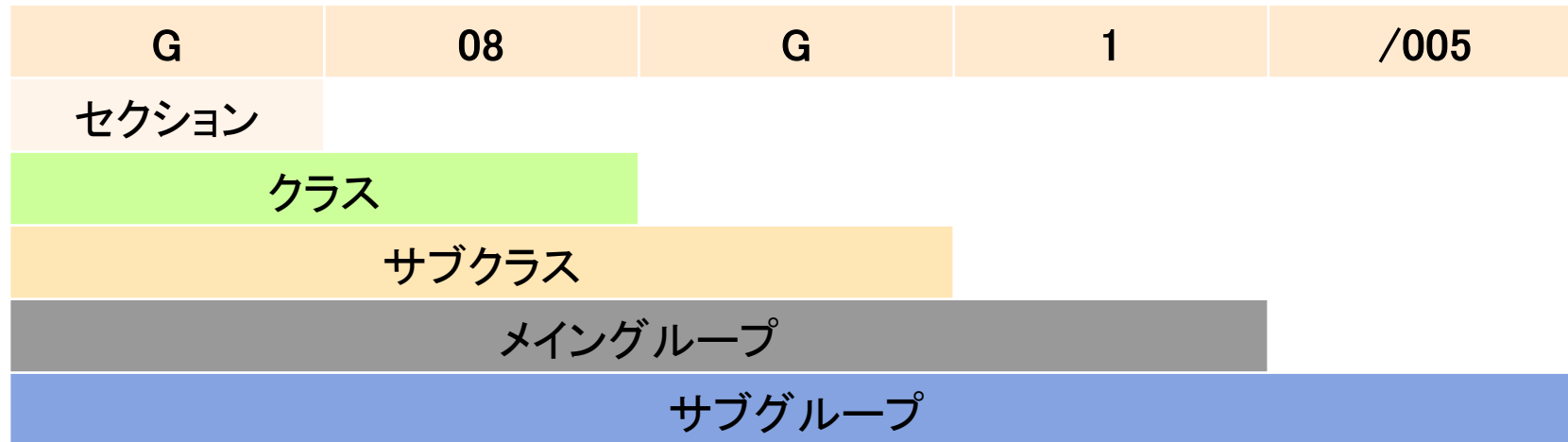
- 学術論文への国際特許分類の自動付与
- 自然言語処理研究の動向

JDreamIIIの検索結果の例

No.	標題	類似検索
☐ 1	マイコン制御の自動運転カー製作記 第12回 移動経路導出アルゴリズムの比較実験...逐次最短距離探索法とダイクストラ法 インターフェース Vol.49 No.2 Page.160-165 (2023.02.01) *ユーザインタフェース, *マイクロコンピュータ, *ブラシレスDCモータ, プリント基板, 位置測定, 地図, 位置決め, 障害物回避, *マイコン, マイコンボード, SLAM 【移動ロボット】	JDream III 特許
☐ 2	自動運転/先進運転支援システムの進化を支える三次元センシング技術 画像ラボ Vol.34 No.1 Page.16-21 (2023.01.10) *自動運転, *運転支援システム, *センサ, *三次元, 自動車, 走行, カメラ, 立体画像, シミュレーション, ステレオカメラ, *先進運転支援システム, 走行環境	JDream III 特許
☐ 3	カメラとLiDARが運動する自動運転向け前方検知 サイバネティクス Vol.28 No.1 Page.10-13 (2023.01.01) 鉄道, *自動列車運転, *カメラ, *LiDAR, センサ, 検出法, ステレオカメラ, 検知技術, 測距センサ, *列車自動運転	JDream III 特許
☐ 4	Focus!JR vol.17 JR西日本「自動運転・隊列走行BRT」公開 JR Gazette Vol.430 Page.56-57 (2023.01.01) 私企業, 西日本, 鉄道, *自動運転, *走行, 自動車, 自動操縦, *バス優先システム, *バス輸送, 高速輸送機関, プロジェクト, 開発, 路線バス, JR西日本, *隊列走行, *BRT, 開発プロジェクト	JDream III 特許
☐ 5	未来志向の駐車場 車両搬送ロボット"Stan"を用いた自動バレーパーキング 交通工学 Vol.58 No.1 Page.42-43,1(2) (2023.01.01) *駐車, *自動化, ロボット, 産業用ロボット, 駐車場, 製造業, 移送ロボット, 三菱重工業, *搬送ロボット	JDream III 特許
☐ 6	LiDARで3D/2D計測 第3部 5000円ToFカメラ・モジュールで体験!距離の画像処理 第3章 自律走行車での障害物検知 インターフェース Vol.49 No.1 Page.87-90 (2023.01.01) *ユーザインタフェース, *LiDAR, レーザ応用, レーザビーム, *画像処理, *自動運転, 無作為抽出, 雑音, RANSAC, ToFカメラ, ノイズ, ラズベリーパイ, レーザ光, 自律走行車	JDream III 特許
☐ 7	LiDARで3D/2D計測 第2部 1万円LiDARとマイコンで自動運転に挑戦 ラズベリー・パイとROSで地図作成&走行シミュレーション インターフェース Vol.49 No.1 Page.70-75 (2023.01.01) *ユーザインタフェース, *LiDAR, レーザ応用, レーザビーム, *自動運転, マイクロコンピュータ, 地図作成, 移動ロボット, 計算機シミュレーション, 地図, マイコン, ラズベリーパイ, レーザ光	JDream III 特許
☐ 8	LiDARで3D/2D計測 第2部 1万円LiDARとマイコンで自動運転に挑戦 第3章 経路計画と台車の走行制御実験 インターフェース Vol.49 No.1 Page.65-69 (2023.01.01) *ユーザインタフェース, *LiDAR, レーザ応用, レーザビーム, *自動運転, マイクロコンピュータ, *経路探索, アルゴリズム, 評価試験, フィードバック制御, 地図作成, PID制御, ダイクストラ法, マイコン, レーザ光, *経路計画	JDream III 特許
☐ 9	LiDARで3D/2D計測 第2部 1万円LiDARとマイコンで自動運転に挑戦 第2章 1つ前と今の点群データの差分から移動量を推定するICPアルゴリズム インターフェース Vol.49 No.1 Page.60-64 (2023.01.01) *ユーザインタフェース, *LiDAR, レーザ応用, レーザビーム, *自動運転, マイクロコンピュータ, 位置測定, 計測システム, 最急降下法, SLAM 【移動ロボット】, マイコン, レーザ光	JDream III 特許

資料種別で絞り込む	—
☐ 逐次刊行物 (13,999件)	
☐ 会議録 (8,927件)	
☐ ニュースレター (47件)	
☐ 年次報告 (2件)	
言語で絞り込む	—
☐ 日本語	
☐ 英語	
検索対象で絞り込む	—
☐ 抄録付き文献のみ	
用語で絞り込む	—
☐ 自動運転	
☐ 自動化	
☐ 自動制御	
☐ 制御装置	
☐ マイクロコンピュータ	
IPC(機械付与)で絞り込む	—
☒ G08G1	
☒ G05D1	
☐ G06F17	

国際特許分類(IPC)

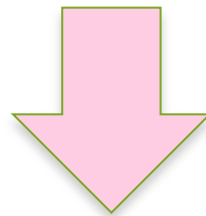


G	物理学
G08	信号
G08G	交通制御システム
G08G1	道路上の車両に対する交通制御システム
G08G1/005	歩行者用の誘導案内表示装置

論文に国際特許分類を付与した例

タイトル：携帯型磁界暴露量計の開発と暴露実測

概要：近年,各種電気設備からの電磁界による電子機器や人間の健康影響への関心が高まっている。本報告は,生活環境における商用周波磁界の暴露実態把握のための携帯型磁界暴露量計の開発と電力設備近傍での磁界暴露実測結果を紹介したものである。商用周波数帯の直交3軸成分の磁束密度と位相及び周波数を1秒以上の間隔で,測定でき,内部ICメモリーに記憶できる。0.1mG(ミリガウス)から50Gの広いダイナミックレンジを持つ。電界強度の高い変電所構内や鉄塔上でも十分安定に動作することを確認した。また測定データの諸統計量は携帯方法にほとんど影響されないことも判明した。

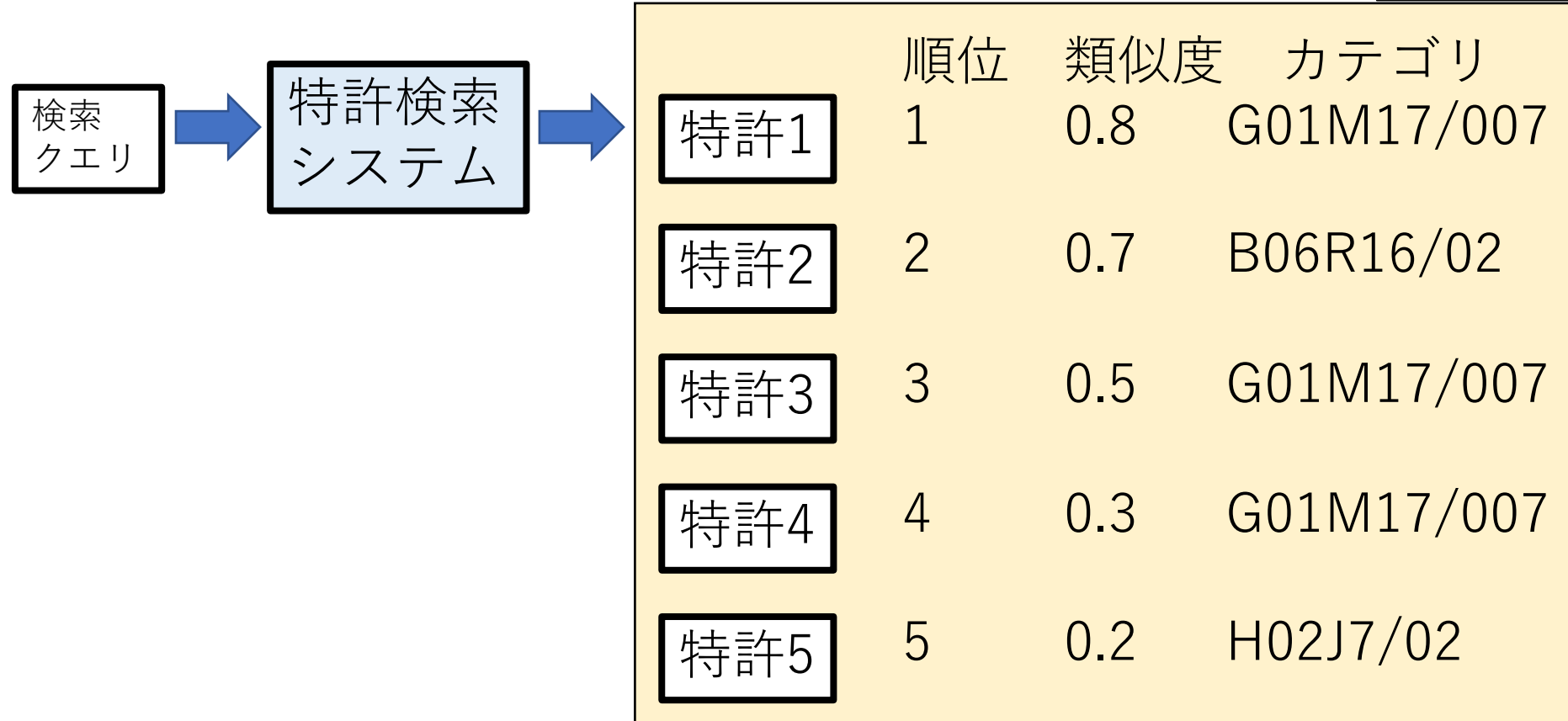


自動分類技術

国際特許分類：G01R33/00 (磁気的変量の測定)

k近傍法による文書分類

検索結果



k近傍法による文書分類

検索結果

学術
論文



特許検索
システム



	順位	類似度	カテゴリ
特許1	1	0.8	G01M17/007
特許2	2	0.7	B06R16/02
特許3	3	0.5	G01M17/007
特許4	4	0.3	G01M17/007
特許5	5	0.2	H02J7/02



分類結果

順位	カテゴリ	スコア
1	G01M17/007	$0.8+0.5+0.3=1.6$
2	B06R16/02	0.7
3	H02J7/02	0.2

国際特許分類を用いた論文検索



- 2019年2月よりJDream IIIでサービス提供開始
- 国際特許分類第8版(2022年6月) メイングループレベル
- 2006年以降の公開公報
- 約70%の分類精度。IPC付与率78%(抄録付き文献は88%)

<https://jdream3.com/>

中央大学難波研究室の
技術を提供しています！

講演内容

- 学術論文への国際特許分類の自動付与
- 自然言語処理研究の動向

自然言語処理とは

私達が日常的に使うことば(自然言語)をコンピュータで処理させるための技術で、人工知能(AI)における代表的な研究分野のひとつ。

自然言語処理の主な応用

- 日本語入力(かな漢字変換)
- 情報検索
- 対話システム
- 機械翻訳
- 文書校正
- 文書分類
- 文書要約
- テキストマイニング



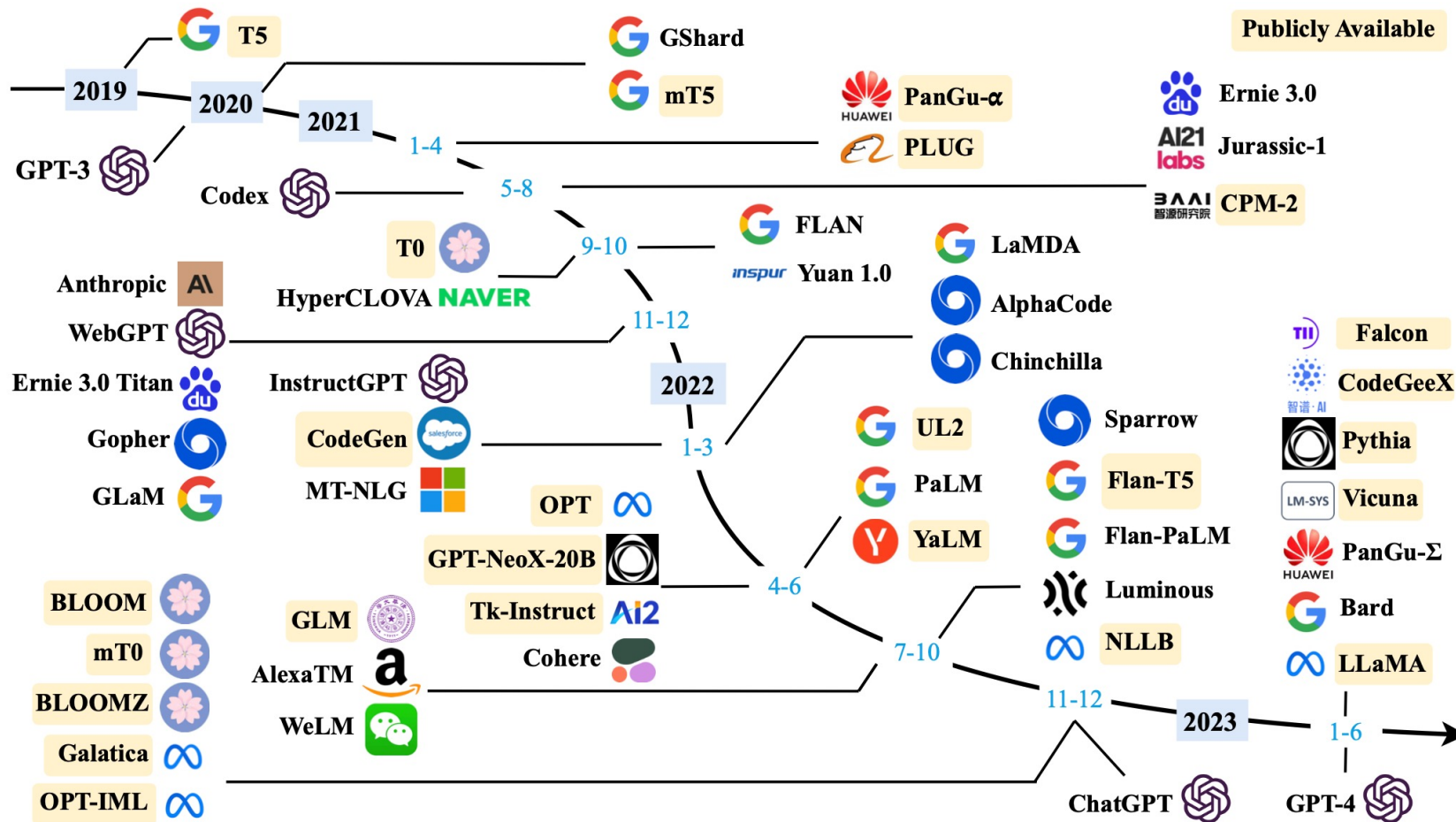


Fig. 2: A timeline of existing large language models (having a size larger than 10B) in recent years. The timeline was established mainly according to the release date (*e.g.*, the submission date to arXiv) of the technical paper for a model. If there was not a corresponding paper, we set the date of a model as the earliest time of its public release or announcement. We mark the LLMs with publicly available model checkpoints in yellow color. Due to the space limit of the figure, we only include the LLMs with publicly reported evaluation results.

言語モデル(1)

言語モデルとは、単語の並びに対する確率分布のこと。

$P(\text{“ツル が 飛ん で いる 。”}) \rightarrow \text{高い}$

$P(\text{“クレーン が 飛ん で いる 。”}) \rightarrow \text{低い}$

今、機械翻訳で“A crane is flying.”という英文を和訳したい。

“crane”の訳語 → ツル、クレーン

言語モデルの利用：音声認識、文字認識

言語モデル(2)

- 単語の並びが自然であるかどうかをコンピュータに認識させるにはどうすれば良いか？
- 人間は、過去に多くの文章を読んできた経験に基づいて単語の並びの自然さを判断できる。
- コンピュータに大量のテキストを与えることで、人間と同じような判断をさせることができるのでは？

言語モデル(3)

昔々あるところに、[?]

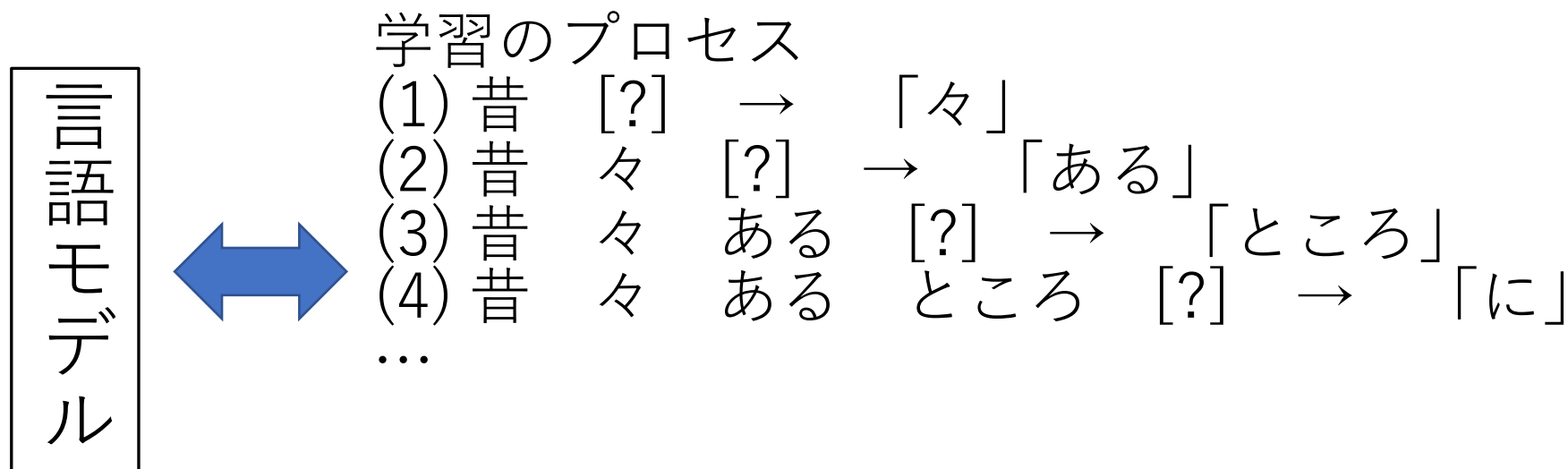
おじいさん	0.31
小さな	0.12
お爺さん	0.10
魚	0.10
翼	0.07

$P(\text{“飛ん”} | \text{“ツル が”}) \rightarrow$ 高い

$P(\text{“飛ん”} | \text{“クレーン が”}) \rightarrow$ 低い

GPTの学習

昔々あるところに、おじいさんとおばあさんが住んでいました。
おじいさんは山に芝刈りに、おばあさんは川に洗濯に行きました。
...



GPT-3の場合、Webページ 570GB(新聞記事に換算すると3200年分)+書籍を使って学習。

GPT-4を用いた生成結果

- 「昔々あるところに、」の続きを書かせると…

巨大な竜が住んでいました。その竜は、周りに住む人々を脅かし、村々を焼き尽くす恐ろしい存在でした。人々は、竜を倒すために勇者を探すことにしました。

そして、ある日、勇者が現れました。彼は、勇敢な心と強い意志を持ち、竜を倒す決意をしました。勇者は、村人たちに別れを告げ、竜が住む洞窟に向かいました。

竜は、勇者を見つけると、火を吹いて襲いかかりました。しかし、勇者は、自分が持っていた特別な剣で竜を切りつけ、勝利を手に入れました。

村人たちは、勇者を迎え入れ、竜から解放された喜びを分かち合いました。この物語は、勇気と決意があれば、どんな困難も克服するこりができるとの重要性も伝えていきます。また、勇者が多くの人々を助け、平和を取り戻すことの重要性も伝えていきます。

ChatGPTとGPT

ChatGPTは、GPTの基本的なアーキテクチャを使用しつつ、**追加の訓練ステップ**を経ることで、より対話型の応答を生成する能力を獲得。

NA 以下の文を、文意を変えずに言い換えてください。
ChatGPTは、GPTの基本的なアーキテクチャを使用しつつ、追加の訓練ステップを経ることで、より対話型の応答を生成する能力を獲得します。

 ChatGPTは、GPTの基礎的な設計を活用しながら、追加的な訓練工程を通じて、対話的な応答を作り出す能力を身につけます。

 Regenerate response



ChatGPTは、GPTの基本的な設計を基に、追加の訓練過程を経て、より対話的な反応を生成する能力を発展させます。

Was this response better or worse?



Better



Worse



Same



NA

配列とwhile文を使ったpythonの課題と模範解答を3つ作成してください。



課題1:

1から10までの整数が格納された配列を用意し、while文を使って配列の要素を順番に出力するプログラムを作成してください。

python

 Copy code

```
arr = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
i = 0
while i < len(arr):
    print(arr[i])
    i += 1
```

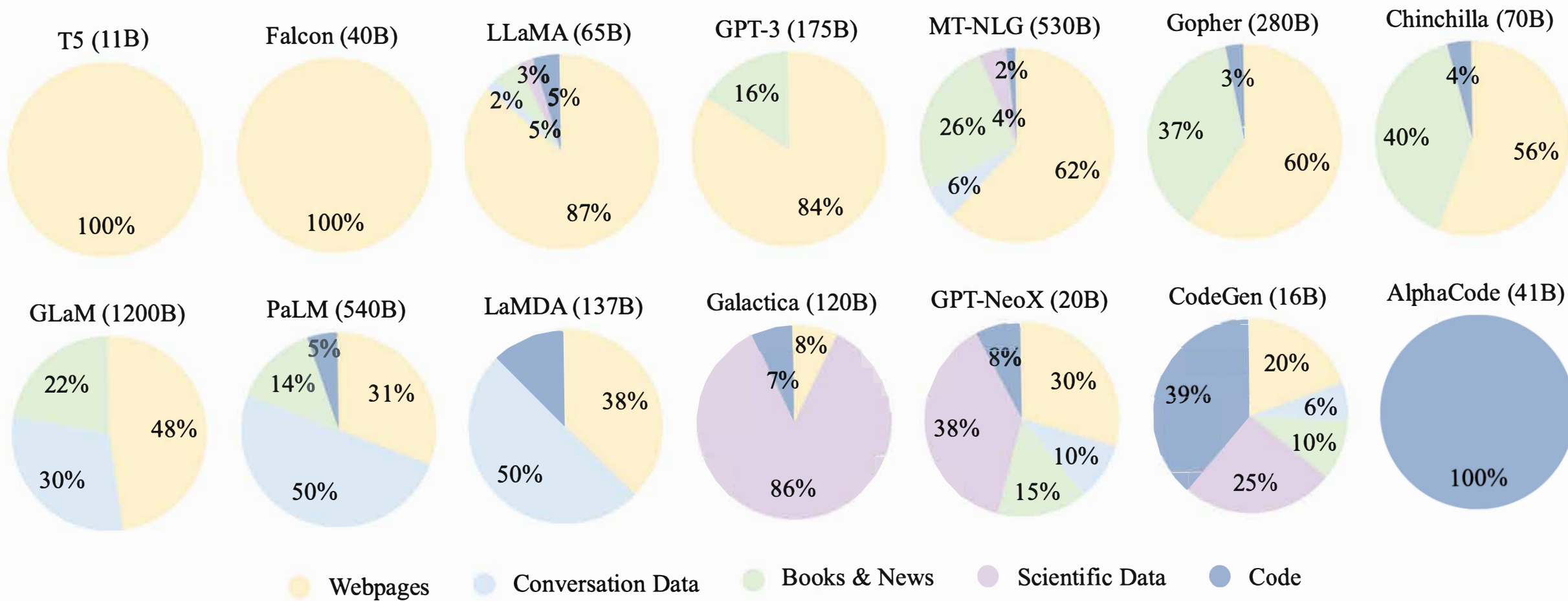


Fig. 5: Ratios of various data sources in the pre-training data for existing LLMs.

LLMの力比べ

Linxin Song et al., NLPBench: Evaluating Large Language Models on Solving NLP Problems, arXiv:2309.15630 [cs.CL]

Table 3: Experimental results in terms of accuracy (%) on our proposed dataset. The best average scores in each type of question are highlighted in **red bold**, and the best average scores for each model in a specific type of question are underlined in blue. Results marked with - denote the incomplete experiment caused by exceeding context length or other prompting errors.

Model	Setting	Multiple Choice			Short Answer			Math			Overall Acc.	
		w/ Context	w/o Context	Average	w/ Context	w/o Context	Average	w/ Context	w/o Context	Average		
LLAMA-2 (13b)	ZS	Orig.	20.00	20.83	20.75	39.23	37.10	38.54	20.00	0.00	4.00	28.72
		+SYS	26.67	34.03	33.33	43.85	27.42	38.54	0.00	0.00	0.00	33.77
		+CoT	26.67	19.44	20.13	22.31	9.68	18.23	0.00	0.00	0.00	17.82
		+CoT+SYS	33.33	27.08	27.67	23.08	9.68	18.75	0.00	0.00	0.00	21.28
	FS	Orig.	-	31.25	28.30	-	29.03	9.38	-	-	0.00	16.76
		+SYS	-	38.19	34.59	-	30.65	9.90	-	-	0.00	19.68
		+CoT	-	30.56	27.67	-	32.26	10.42	-	-	0.00	17.02
		+CoT+SYS	-	36.81	33.33	-	35.48	11.46	-	-	0.00	19.95
LLAMA-2 (70b)	ZS	Orig.	40.00	22.22	23.90	53.85	38.71	48.96	9.09	0.00	8.00	35.64
		+SYS	40.00	23.61	25.16	54.62	46.77	52.08	9.09	0.00	8.00	37.77
		+CoT	33.33	21.53	22.64	32.31	12.90	26.04	0.00	0.00	0.00	22.87
		+CoT+SYS	40.00	38.19	38.36	33.08	25.81	30.73	0.00	0.00	0.00	31.91
	FS	Orig.	33.33	29.17	29.56	48.46	38.71	45.31	9.09	0.00	19.38	36.93
		+SYS	26.67	34.72	33.96	46.92	40.32	44.79	0.00	0.00	0.00	37.23
		+CoT	26.67	31.94	31.45	38.46	51.61	42.71	0.00	0.00	0.00	35.11
		+CoT+SYS	26.67	38.19	37.11	35.38	48.39	39.58	4.55	0.00	4.00	36.17
PaLM-2	ZS	Orig.	66.67	37.50	40.25	49.23	35.48	44.79	13.64	33.33	16.00	40.96
		+SYS	66.67	45.83	47.80	51.54	37.10	46.88	4.55	33.33	8.00	44.68
		+CoT	60.00	36.81	38.99	47.69	37.10	44.27	18.18	33.33	20.00	40.42
		+CoT+SYS	53.33	41.67	42.77	40.00	30.65	36.98	13.64	0.00	12.00	37.77
	FS	+ToT	-	4.86	4.40	-	0.00	0.00	-	-	0.00	1.86
		Orig.	53.33	38.89	40.25	57.69	33.87	50.00	4.55	33.33	8.00	43.08
		+SYS	53.33	39.58	40.88	56.15	38.71	50.52	4.55	0.00	4.00	43.35
		+CoT	53.33	40.28	41.51	49.23	38.71	45.83	0.00	0.00	0.00	40.96
GPT-3.5	ZS	+CoT+SYS	40.00	38.89	38.99	53.85	40.32	49.48	0.00	0.00	0.00	41.75
		+ToT	-	10.42	9.43	-	1.61	0.52	-	-	0.00	4.25
		Orig.	40.00	52.05	50.64	75.38	58.73	69.99	36.36	33.33	36.00	59.55
		+SYS	46.67	40.69	41.51	71.54	62.90	68.75	13.64	33.33	16.00	53.72
	FS	+CoT	53.33	52.74	52.83	63.85	33.33	54.19	18.18	100.00	28.00	51.87
		+CoT+SYS	46.67	39.58	40.25	66.92	59.68	64.58	18.18	0.00	16.00	51.06
		+ToT	-	31.25	28.30	-	0.00	0.00	-	-	0.00	11.97
		Orig.	53.33	36.81	38.36	66.15	64.52	65.62	18.18	33.33	20.00	51.06
GPT-4	ZS	+SYS	46.67	44.44	44.65	66.15	54.84	62.50	18.18	0.00	16.00	51.86
		+CoT	40.00	40.28	40.25	64.62	62.90	64.06	13.64	0.00	12.00	50.53
		+CoT+SYS	40.00	46.53	45.91	66.15	64.52	65.62	18.18	0.00	16.00	53.99
		+ToT	-	30.56	27.67	-	56.45	18.23	-	-	0.00	21.01
	FS	Orig.	86.67	70.55	72.25	78.46	69.84	75.42	22.73	33.33	24.00	70.66
		+SYS	86.67	57.93	60.38	83.85	79.03	82.29	18.18	0.00	16.00	68.62
		+CoT	86.67	72.60	74.10	74.62	57.14	68.65	13.64	100.00	24.00	67.99
		+CoT+SYS	86.67	56.25	59.12	73.08	75.81	73.96	27.27	66.67	28.00	64.63
GPT-4	ZS	+ToT	-	60.42	54.72	-	0.00	0.00	-	-	0.00	23.14
		Orig.	86.67	62.50	64.78	77.69	75.81	77.08	22.73	0.00	20.00	68.08
		+SYS	86.67	59.03	61.64	81.54	79.03	80.73	13.64	33.33	16.00	68.35
		+CoT	86.67	60.42	62.89	78.46	75.81	77.60	36.36	0.00	32.00	68.35
	FS	+CoT+SYS	86.67	60.42	62.89	80.00	74.19	78.12	13.64	66.67	20.00	67.82
		+ToT	-	60.42	54.72	-	75.81	24.48	-	-	0.00	35.64

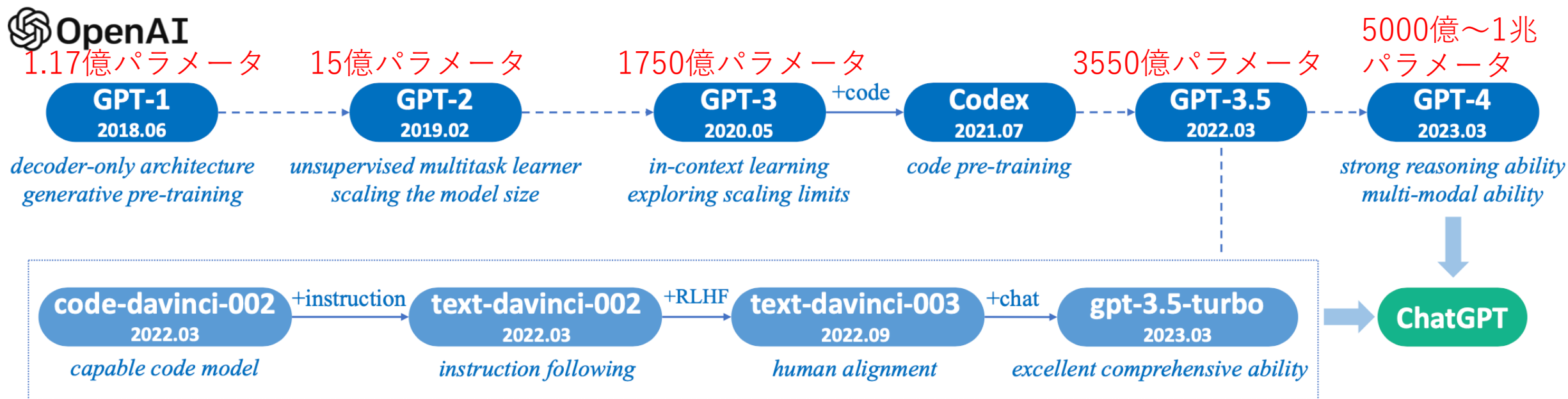


Fig. 3: A brief illustration for the technical evolution of GPT-series models. We plot this figure mainly based on the papers, blog articles and official APIs from OpenAI. Here, *solid lines* denote that there exists an explicit evidence (e.g., the official statement that a new model is developed based on a base model) on the evolution path between two models, while *dashed lines* denote a relatively weaker evolution relation.

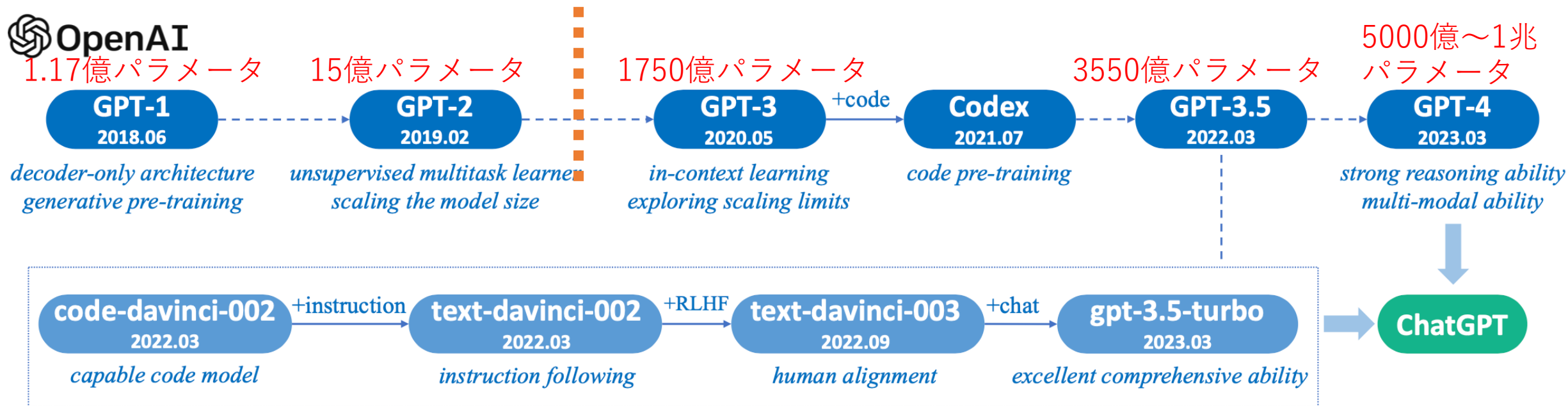


Fig. 3: A brief illustration for the technical evolution of GPT-series models. We plot this figure mainly based on the papers, blog articles and official APIs from OpenAI. Here, *solid lines* denote that there exists an explicit evidence (e.g., the official statement that a new model is developed based on a base model) on the evolution path between two models, while *dashed lines* denote a relatively weaker evolution relation.

様々な言語モデル(3)

	GPT-1	GPT-2	GPT-3	BERT	GPT-Neo	T5	GPT-3.5-T	Llama 2	GPT-4
学習データ	書籍7000冊 (5GB)	Web800万ページ (40GB)	Webページ (570GB)、 書籍 / 0.3T (token)	書籍 11000冊、 Wikipedia (16GB)	Web、論文、特許等 (825GB)	Webページ (745GB)		2T(token)	
パラメータ数	1.17億	15億	1750億	3億4000万	17億5000万	110億	3550億	700億	5000億～1兆(推定)
公開	オープンソース	オープンソース	APIのみ	オープンソース	オープンソース	オープンソース	APIのみ	オープンソース	APIのみ
その他			学習： 4億9000万円					学習： 25億円(推定)	

教師有り学習の例

感情の強度推定 [Kajiwara 2023]

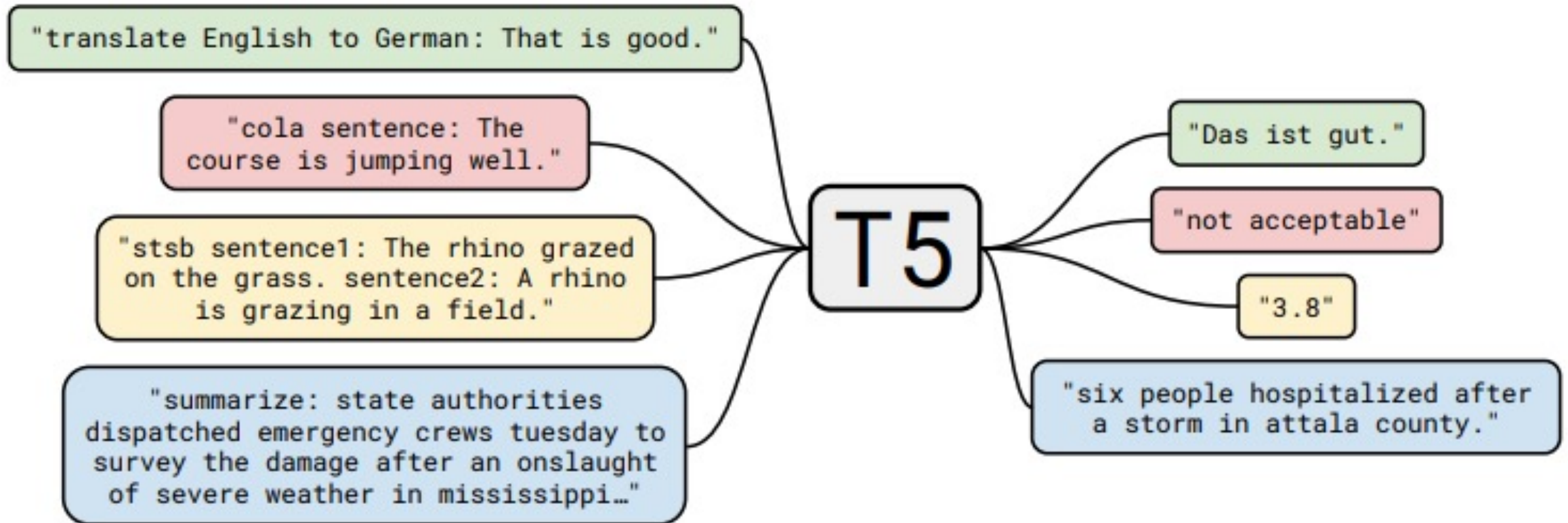
大学の大声で馬鹿騒ぎするノリが嫌いすぎて学校行くのが嫌になる



喜び:0, 悲しみ:0, 期待:0, 驚き:0, **怒り:3**, 恐れ:0, **嫌悪:3**, 信頼:0

SNSから収集した35,000件のテキストデータに8つの感情が4段階で付与されている

様々な言語モデル T5



教師有り学習 vs ゼロショット学習

感情の強度推定の比較

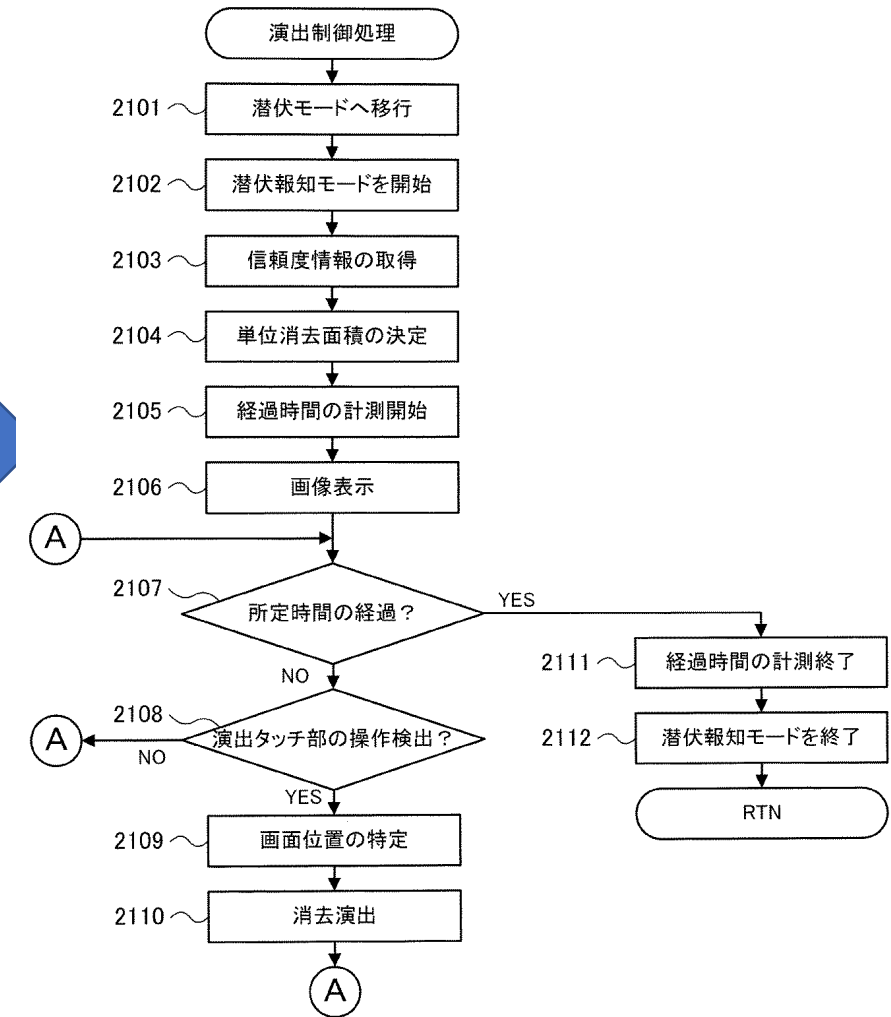
モデル	Quadratic Weighted Kappa (QWK)
GPT-4 (ゼロショット)	0.2919
T5 (教師有り)	0.5451

GPT-4: 以下のテキストの感情「喜び」の強度は4段階でいくつですか？

T5: 35,000件のテキスト+**人手**で8つの感情が4段階で付与されたデータを使って機械学習。

ChatGPTで説明文からフローチャートを生成

潜伏モードへ移行後(s2101)潜伏報知モードを開始する(s2102)。信頼度の情報を取得すると(s2103)、それを基に変倍率を決定し(s2104)、経過時間の計測を開始し(s2105)、所定の画像を画像表示部に表示する(s2106)。所定時間が未経過のときに(s2107でno)演出タッチ部の操作を検出すると(s2108でyes)、演出タッチ部の接触位置に対応する画像表示部の画面上の画面位置を特定し(s2109)、変倍率の情報を用いて表示倍率変更演出を行う(s2110)。所定時間が経過後は(s2107でyes)、経過時間の計測を終了し(s2111)、潜伏報知モードを終了する(s2112)。



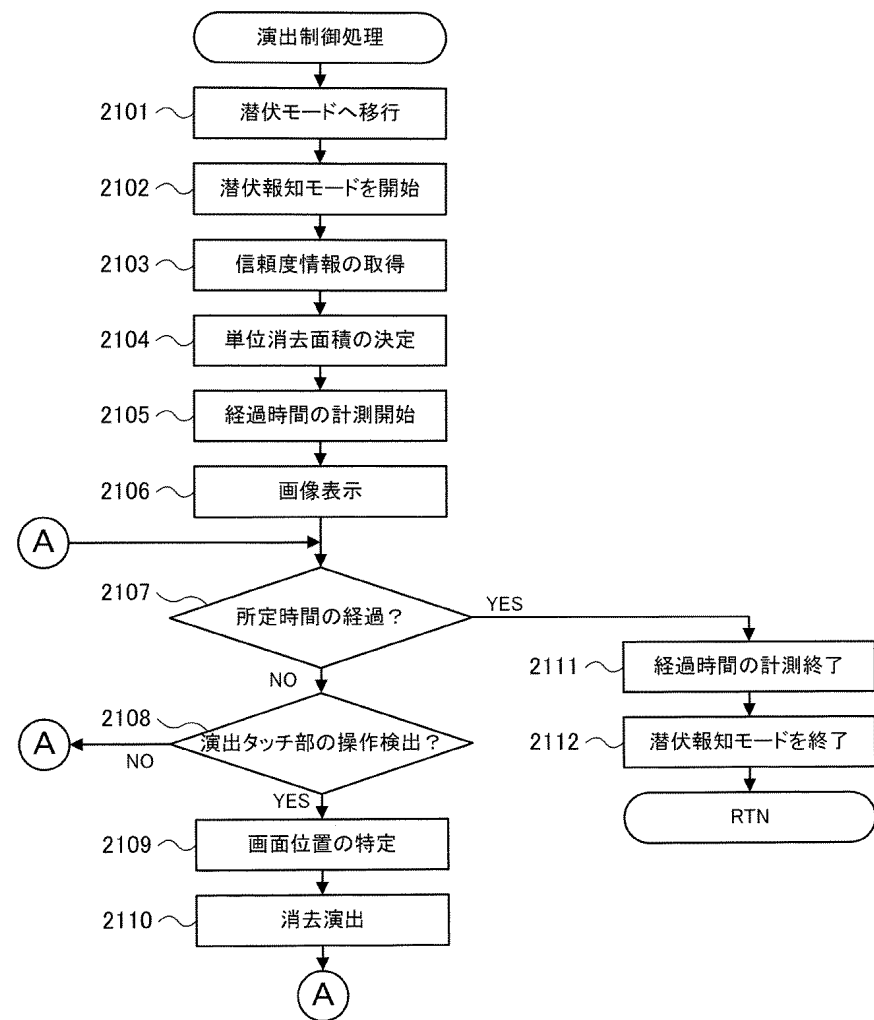
ChatGPTで説明文からフローチャートを生成



mermaid

Copy code

```
graph TD;
  A((s2101: 潜伏モードへ移行)) --> B((s2102: 潜伏報知モード開始));
  B --> C((s2103: 信頼度情報取得));
  C --> D((s2104: 変倍率決定));
  D --> E((s2105: 経過時間計測開始));
  E --> F((s2106: 画像表示));
  F -->|所定時間未経過| G((s2107: 演出タッチ部操作検出));
  G -->|yes| H((s2108: 画面位置特定));
  H --> I((s2109: 表示倍率変更演出));
  G -->|no| I;
  F -->|所定時間経過| J((s2107));
  J -->|yes| K((s2111: 経過時間計測終了));
  K --> L((s2112: 潜伏報知モード終了));
  L --> RTN([RTN]);
```



ChatGPTを用いたブーリアン検索式の自動生成

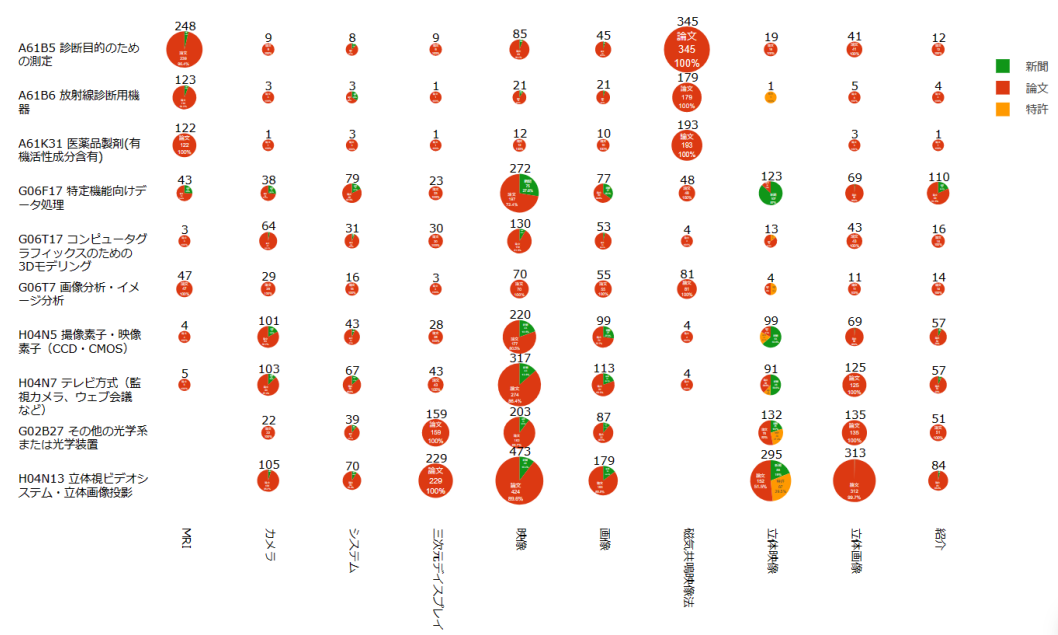
- Shuai Wand et al., Can ChatGPT Write a Good Boolean Query for Systematic Review Literature Search?, SIGIR '23: Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1426-1436, 2023.

Table 3: Example designed guided prompt for query formulation.

Step	Prompt	Example Answer
Step1	Follow my instructions precisely to develop a highly effective Boolean query for a medical systematic review literature search. Do not explain or elaborate. Only respond with exactly what I request. First, Given the following statement and text from a relevant study, please identify 50 terms or phrases that are relevant. The terms you identify should be used to retrieve more relevant studies, so be careful that the terms you choose are not too broad. You are not allowed to have duplicates in your list. statement: "Prevalence of Differentiated Thyroid Cancer in Autopsy Studies Over Six Decades: A Meta-Analysis" Text: Ten occult carcinomas of the thyroid gland were found in 274 unselected autopsies at the Pathology Service, Hospital José Joaquín Aguirre, between December 1980 and March 1983. This is the lowest incidence among the most recent published series. The smallest papillary microcarcinomas presented most frequently as nonencapsulated nonsclerosing tumors without inflammatory infiltrate or fibrosis, suggesting that they may represent the early stages of development. Our data show a relatively high and similar frequency of papillary microcarcinomas in surgical and autopsy series, but do not demonstrate risk factors for clinical evolution.	1. Differentiated thyroid cancer 2. Prevalence 3. Autopsy studies 4. Meta-analysis 5. Occult carcinomas 6. Thyroid gland 7. Incidence 8. Environmental factors 9. Etiology 10. Low risk ... 50. Parameters.
Step 2	For each item in the list you created in Step 1, classify it into three categories: terms relating to health conditions (A), terms relating to a treatment (B), terms relating to types of study design (C). When an item does not fit one of these categories, mark it as (N/A). Each item needs to be categorised into (A), (B), (C), or (N/A).	1. (A) Differentiated thyroid cancer 2. (N/A) Prevalence 3. (C) Autopsy studies 4. (C) Meta-analysis 5. (A) Occult carcinomas 6. (A) Thyroid gland 7. (N/A) Incidence 8. (N/A) Environmental factors 9. (N/A) Etiology 10. (N/A) Low risk ... 50. (N/A) Parameters.
Step 3	Using the categorised list you created in step 2, create a Boolean query that can be submitted to PubMed which groups together items from each category. For example: ((itemA1[Title/Abstract] OR itemA2[Title/Abstract] OR itemA2[Title/Abstract]) AND (itemB1[Title/Abstract] OR itemB2[Title/Abstract] OR itemB3[Title/Abstract]) AND (itemC1[Title/Abstract] OR itemC2[Title/Abstract] OR itemC3[Title/Abstract]))	((differentiated[Title/Abstract] OR thyroid[Title/Abstract] OR carcinoma[Title/Abstract] OR papillary[Title/Abstract] OR microcarcinoma[Title/Abstract]) AND (cancer[Title/Abstract] OR incidence[Title/Abstract] OR etiology[Title/Abstract] OR risk[Title/Abstract] OR gender[Title/Abstract] OR hormonal[Title/Abstract] OR nodular[Title/Abstract] OR goiter[Title/Abstract] OR Hashimoto's[Title/Abstract] OR malignancy[Title/Abstract] OR concomitant[Title/Abstract] OR tumor[Title/Abstract] OR infiltrate[Title/Abstract] OR fibrosis[Title/Abstract] OR development[Title/Abstract] OR frequency[Title/Abstract])) AND (autopsy[Title/Abstract] OR surgical[Title/Abstract] OR material[Title/Abstract] OR series[Title/Abstract] OR specimens[Title/Abstract] OR cases[Title/Abstract])).
Step 4	Use your expert knowledge to refine the query, making it retrieve as many relevant documents as possible while minimising the total number of documents retrieved. Also add relevant MeSH terms into the query where necessary, e.g., MeSHTerm[MeSH]. Retain the general structure of the query, however, with each main clause of the query corresponding to a PICO element. The final query still needs to be executable on PubMed, so it should be a valid query.	((differentiated thyroid cancer[MeSH] OR "differentiated thyroid"[All Fields] OR "thyroid carcinoma"[All Fields] OR "papillary microcarcinoma"[All Fields]) AND (prevalence[All Fields] OR incidence[MeSH] OR "etiology of"[All Fields] OR "risk factors"[All Fields] OR gender[All Fields] OR hormonal[All Fields] OR "nodular goiter"[All Fields] OR "Hashimoto's thyroiditis"[MeSH] OR malignancy[MeSH] OR "concomitant lesion"[All Fields] OR tumor[All Fields] OR infiltrate[All Fields] OR fibrosis[All Fields] OR "early stages of development"[All Fields] OR frequency[All Fields])) AND (autopsy[MeSH] OR surgical[All Fields] OR material[All Fields] OR series[All Fields] OR specimens[All Fields] OR cases[All Fields]))

ChatGPTの応用事例

JDream Innovation Assistの図表への説明文の自動付与



以下のCSVについて、論文は研究段階化どうか、特許は技術の応用度がどのくらいか、新聞はどのくらい市場化が進んでいるか、技術ごとにそれぞれ、簡潔に説明してください。

ChatGPTにより生成された説明文

以下は、各技術について論文の研究段階化、特許の技術応用度、新聞の市場化の進み具合を示した表です。

技術	論文	特許	新聞
MRI	研究段階	低	低
カメラ	低	低	中
システム	中	低	中
三次元ディスプレイ	研究段階	低	低
映像	高	低	高
画像	中	低	中
磁気共鳴映像法	研究段階	低	低
立体映像	中	中	高
立体画像	中	低	低
紹介	中	低	低

MRI: 論文の数が比較的多いですが、特許や新聞記事はほとんどありません。MRI技術はまだ研究段階にあり、特許や市場化は進んでいません。

カメラ: 論文や特許の数は少なく、新聞記事もあまりありません。カメラ技術はまだ研究段階にあり、特許や市場化もあまり進んでいません。

システム: システムに関連する論文の数は中程度で、特許や新聞記事は比較的少ないです。システム技術は一部で研究が進んでおり、特許や市場化も一部進んでいます。

三次元ディスプレイ: 三次元ディスプレイに関連する論文の数は比較的多いですが、特許や新聞記事はほとんどありません。三次元ディスプレイ技術はまだ研究段階にあり、特許や市場化は進んでいません。

映像: 映像技術に関連する論文数は高く、特許も少ないです。一方、新聞記事は比較的多いです。映像技術は研究が進んでおり、市場化も一部進んでいます。

まとめ

- 学術論文への国際特許分類の自動付与
- 自然言語処理研究の動向
 - ChatGPTとは、言語モデルのひとつであるGPTを使用したチャットボット。
 - 文書生成だけでなく、機械翻訳やプログラムの生成なども可能。
 - 学習用データが存在する場合には、小規模な言語モデルでもGPT-4を上回ることがある。
 - 一方で、未知の問題に関してはLLMが役立つ。