

# ネットワークオペレーションにおける 生成AI技術の活用検討について

JANOG54

2024/07/04 15:30~17:00

JPRSホール (コンベンションホールA)

KDDI総合研究所

宮坂拓也

# このプログラムは・・・

## ■ LLM/生成AIとネットワーク運用に関するプログラム応募が(偶然)2件あり、そのプログラムを1個にまとめたものです！

↓元の2つの応募プログラム情報です ☺

<p><b>タイトル</b></p> <p><b>アブストラクト</b></p>	<p>LLMと生成AIを活用したネットワーク運用</p> <p>JANOG51において、「AI/MLによるネットワーク運用と自律型ネットワークに向けて[1]」という題目で、機械学習や深層学習といったAI技術のネットワーク運用について皆様と議論させていただきました。そのJANOG51が開催された2023年1月は、まさにリリースされたばかりのChatGPTが世界全体で話題となった時でした。そのような大規模言語モデル(LLM)/生成AIの流行は、一時のブームで終わらず、現在も社会全体で適用が試みられていると思います。</p> <p>ネットワーク運用においても、これらの技術の適用可能性は十分あるのではないのでしょうか？ネットワーク運用の業務内容は多種多様にわたり、顧客やり取り、回線手配、ネットワーク設計、コンフィグ作成、障害対応などの「自然言語」を用いたやり取りが多く、それらの技術の適用可能性は十分あるのではないのでしょうか？</p> <p>本発表では、LLM/生成AIのネットワーク運用への適用に関して皆様と話しあいたいと思います。発表においては、背後にある基礎的な技術・モデルの拡張手法(例：Fine-tuning, Retrieval-Augmented Generation(RAG))等を整理し、研究例も含めたLLM/生成AIの適用事例を紹介した上で、ネットワーク運用者の皆様と議論を行い、今後のより良いネットワーク運用を皆様と目指していきたいです。</p> <p>[1] <a href="https://www.janog.gr.jp/meeting/janog51/aim/">https://www.janog.gr.jp/meeting/janog51/aim/</a></p> <div data-bbox="810 968 947 1110"></div> <p>宮坂 拓也 株式会社KDDI総合研究所</p>	<p>ネットワークオペレーションにおける生成AI技術の活用検討について</p> <p>私たちの部署では、西日本地域におけるInternetを支える大規模なNW設備の運用保守を24時間365日行っています。30府県で提供するNW回線は、設計構築に協力会社も含めて数万人レベルにも及ぶメンバーにより日々構築されており、現在、約11万台の装置により構成されています。大規模かつ成長し続けるNWを限りある人的リソースで保守するため、自動化を含めた業務の生産性向上に継続して取り組んでいる中で、急速に普及している生成AI技術のNWオペレーションへの活用に大きな期待をしており現場で試行錯誤を重ねています。実運用に向けた検討の中でいくつかの課題が抽出されており、それらに対するアプローチは以下のようなものになります。</p> <p>まず生成AIの核となる大規模言語モデル(LLM)の環境構築に際し、クラウド上のLLMサービスの利用や、オンプレミスでLLM環境を構築するにはセキュリティポリシーに抵触したり、コストが膨大になるリスクがあります。フィージビリティ(実現性)から模索中の段階で、このような問題を解決できる大きなプロジェクトとし推進が可能な組織は多くないと思います。そこで、データサイズが小さく一般的なPCでも動作するモデル(Local LLM)を組み合わせて、セキュアなオンプレミス環境に比較的安価で用意できるマシンリソースの中で動作するアーキテクチャの構成を試み、精度・応答速度などを評価してみました。</p> <p>次に、NWオペレーションにおける判断には、過去のトラブルチケットの内容、NW装置のアラーム情報やコマンド応答内容など専門的な情報がほとんどの場合で必要になりますが、頻繁に更新されるこれらの情報をLocal LLMに直接学習させ続けることは難しいです。このため、外部の情報をベクトルデータベースとすることでLocal LLMが検索・参照できるようにする検索拡張生成(RAG)という手法を用いてNWオペレーション特有の専門的情報をLocal LLMが取り扱えるようにしました。</p> <div data-bbox="1421 968 2397 1135"><p>また、LLM 要なオペ ツールを 利用し、LL 私たちが よりオペ</p><div data-bbox="1559 968 1676 1088"></div><p>佐藤 亮介 株式会社NTTフィールドテクノ</p><div data-bbox="1778 968 1895 1088"></div><p>白井 嵩士 株式会社NTTフィールドテクノ</p><div data-bbox="1997 968 2114 1088"></div><p>田口 順史 株式会社NTTフィールドテクノ</p><div data-bbox="2206 968 2323 1088"></div><p>近藤 徳志 株式会社NTTフィールドテクノ</p></div> <p>このような取り組みの経過を、実際のアーキテクチャ・評価データなどを用いて紹介させていただき、生成AI技術を活用したNWオペレーションの実運用に向けて活発な議論をさせていただければと思います</p> <ul style="list-style-type: none"><li>・NWオペレーションにおいて、生成AIを活用したいユースケースについて</li><li>・生成AI活用の際する課題感について(セキュリティや環境構築のハードルなど)</li><li>・紹介した技術スタック及びフレームワーク(RAG・langchainなど)に関する議論など</li></ul>
<p><b>議論のポイント</b></p>	<p>本発表では、LLM/生成AIのネットワーク運用への適用に関して皆様と話しあいたいと思います。具体的には、以下の2つのポイント(現在と未来)に関して、実際のネットワーク運用者と議論をする予定です。</p> <ul style="list-style-type: none"><li>・現在：すでにLLM/生成AI活用している？</li><li>・未来：どんな業務にLLM/生成AI使えそう？</li></ul>	



以下は、ある技術系のカンファレンスに応募されたプログラムのアブストラクトです。  
この2つのプログラムの類似度を100点満点で教えてください。



## (1) 応募案1

JANOG51において、「AI/MLによるネットワーク運用と自律型ネットワークに向けて [1]」という題目で、機械学習や深層学習といったAI技術のネットワーク運用について皆様と議論させていただきました。そのJANOG51が開催された2023年1月は、まさにリリースされたばかりのChatGPTが世界全体で話題となった時でした。そのような大規模言語モデル(LLM)/生成AIの流行は、一時のブームで終わらず、現在も社会全体で適用が試みられていると思います。ネットワーク運用においても、これらの技術の適用可能性は十分あるのではないのでしょうか？ネットワーク運用の業務内容は多種多様にわたり、顧客やり取り、回線手配、ネットワーク設計、コンフィグ作成、障害対応などの「自然言語」を用いたやり取りが多く、それらの技術の適用可能性は十分あるのではないのでしょうか？本発表では、LLM/生成AIのネットワーク運用への適用に関して皆様と話しあいたいと思います。発表においては、背後にある基礎的な技術・モデルの拡張手法(例：Fine-tuning, Retrieval-Augmented Generation(RAG))等を整理し、研究例も含めたLLM/生成AIの適用事例を紹介した上で、ネットワーク運用者の皆様と議論を行い、今後のより良いネットワーク運用を皆様と目指していきたいです。[1]  
<https://www.janog.gr.jp/meeting/janog51/aiml/>

## (2) 応募案2

私たちの部署では、西日本地域におけるInternetを支える大規模なNW設備の運用保守を24時間365日行っています。30府県で提供するNW回線は、設計構築に協力会社も含めて数万人レベルにも及ぶメンバーにより日々構築されており、現在、約11万台の装置により構成されています。大規模かつ成長し続けるNWを限りある人的リソースで保守するため、自動化を含めた業務の生産性向上に継続して取り組んでいる中で、急速に普及している生成AI技術のNWオペレーションへの活用に大きな期待をしており現場で試行錯誤を重ねています。実運用に向けた検討の中でいくつかの課題が抽出されており、それらに対するアプローチは以下のようなものになります。まず生成AIの核となる大規模言語モデル(LLM)の環境構築に際し、クラウド上のLLMサービスの利用や、オンプレミスでLLM環境を構築するにはセキュリティポリシーに抵触したり、コストが膨大になるリスクがあります。フィージビリティ(実現性)から模索中の段階で、このような問題を解決できる大



これらの2つのアブストラクトは、どちらも生成AI技術（特に大規模言語モデル、LLM）のネットワーク運用への適用に焦点を当てていますが、異なる観点や具体的なアプローチを取っています。以下に類似点と相違点をまとめます。

## 類似点

1. **主題:** 両方とも生成AI技術（LLM）をネットワーク運用に適用することをテーマにしています。
2. **技術的要素:** LLMやRAG（Retrieval-Augmented Generation）などの技術に言及しています。
3. **目的:** より効率的なネットワーク運用を目指しています。
4. **議論の意図:** 両方とも、参加者との議論を通じて知見を深めることを目的としています。

## 相違点

1. **具体的なアプローチ:**
  - 応募案1は、主にLLMの基礎技術や研究例の紹介を通じて議論を行うことに重点を置いています。
  - 応募案2は、実際の運用環境でのLLM適用の具体的な試行と、その結果に基づく課題解決のアプローチを述べています。
2. **技術の導入方法:**
  - 応募案1は、LLM/生成AIの技術を広範に紹介し、今後の適用可能性について議論することを目指しています。
  - 応募案2は、具体的な運用方法（クラウド vs. オンプレミス、Local LLM、Langchain Agentなど）について詳細に述べています。

## 3. 対象読者:

- 応募案1は、ネットワーク運用全般に興味を持つ技術者や研究者を対象としています。
- 応募案2は、実際のネットワーク運用に密接に関わる技術者や管理者を対象としており、具体的な実装や評価に焦点を当てています。

## 類似度の評価

両方のアブストラクトは、同じ技術（LLM/生成AI）をネットワーク運用に適用するという共通のテーマを持っていますが、そのアプローチや具体的な内容には違いがあります。これらの点を考慮に入れた上で、類似度を評価すると以下ようになります：

**類似度: 70点 / 100点**

この評価は、両方のアブストラクトが同じテーマを扱っているものの、具体的なアプローチや対象とする読者が異なるためです。

内容	発表者	目安時間
<ul style="list-style-type: none"> <li>LLM概要とネットワーク運用への適用検討</li> </ul>	 <p>宮坂 拓也 株式会社KDDI総合研究所</p>	15分
<ul style="list-style-type: none"> <li>ネットワーク運用への活用検討事例 (KDDI編)</li> </ul>	 <p>仲松 匠 KDDI株式会社</p>	15分
<ul style="list-style-type: none"> <li>ネットワーク運用への活用検討事例 (NTT-FT編)</li> </ul>	    <p>佐藤 亮介 株式会社NTTフィールドテクノ</p> <p>白井 嵩士 株式会社NTTフィールドテクノ</p> <p>田口 順史 株式会社NTTフィールドテクノ</p> <p>近藤 優吉 株式会社NTTフィールドテクノ</p>	30分
<ul style="list-style-type: none"> <li>議論</li> </ul>	<p>全員 + 参加者 + 司会</p>	30分

内容	発表者	目安時間
• LLM概要とネットワーク運用への適用検討	 宮坂 拓也 株式会社KDDI総合研究所	15分
• ネットワーク運用への活用検討事例 (KDDI編)	 仲松 匠 KDDI株式会社	15分
• ネットワーク運用への活用検討事例 (NTT-FT編)	 佐藤 亮介 株式会社NTTフィールドテクノ	30分
• 議論	 白井 嵩士 株式会社NTTフィールドテクノ	30分
	 田口 順史 株式会社NTTフィールドテクノ	
	 近藤 健吉 株式会社NTTフィールドテクノ	
	全員 + 参加者 + 司会	

## ■ 名前: 宮坂拓也 (みやさかたくや)

## ■ 経歴

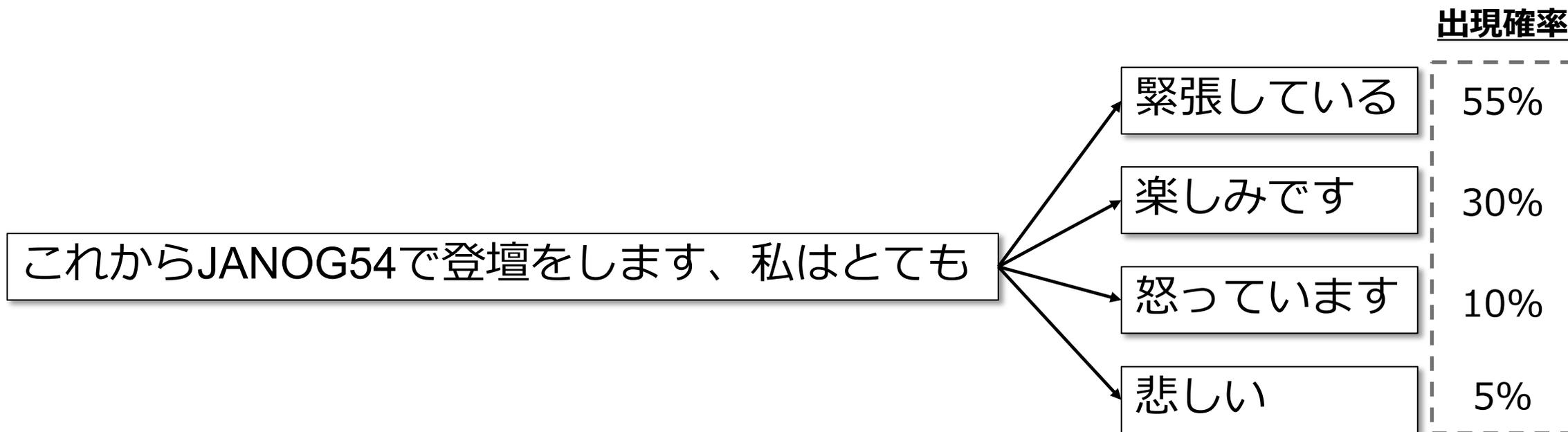
- **2011/4~2018/3: Network Engineer (KDDI)**
  - KDDIのバックボーンネットワークの開発業務 (ルーター導入)
  - IETFにおける標準化活動
- **2018/4~2021/9: Research engineer (KDDI総合研究所)**
  - 自動運転/コネクティッドカーのための通信インフラに関する研究開発
- **2021/10~ (現在): Senior manager (KDDI総合研究所)**
  - ネットワーク運用に関する研究開発をリード



本日の話題

## ■ Large Language Model (LLM)

- 大規模言語モデル
- 自然言語の意味の理解、文脈の理解を行うことができ、近年では大規模なニューラルネットワークと、学習データにより、非常に高精度なモデルが利用可能となっている
- 基本的には条件付き確率モデルで、前の数語から次の1語を予測するモデル



**GPT = Generative Pre-trained Transformer**

**生成的な + 事前学習された + Transformerモデル**

生成AI的な、入力情報や条件からオリジナルな答え（文章）を生成することができる。

大量の学習データ（Webサイト等）から大規模なニューラルネットワークを事前に学習。

後述するTransformer。Attention層のみで構築。

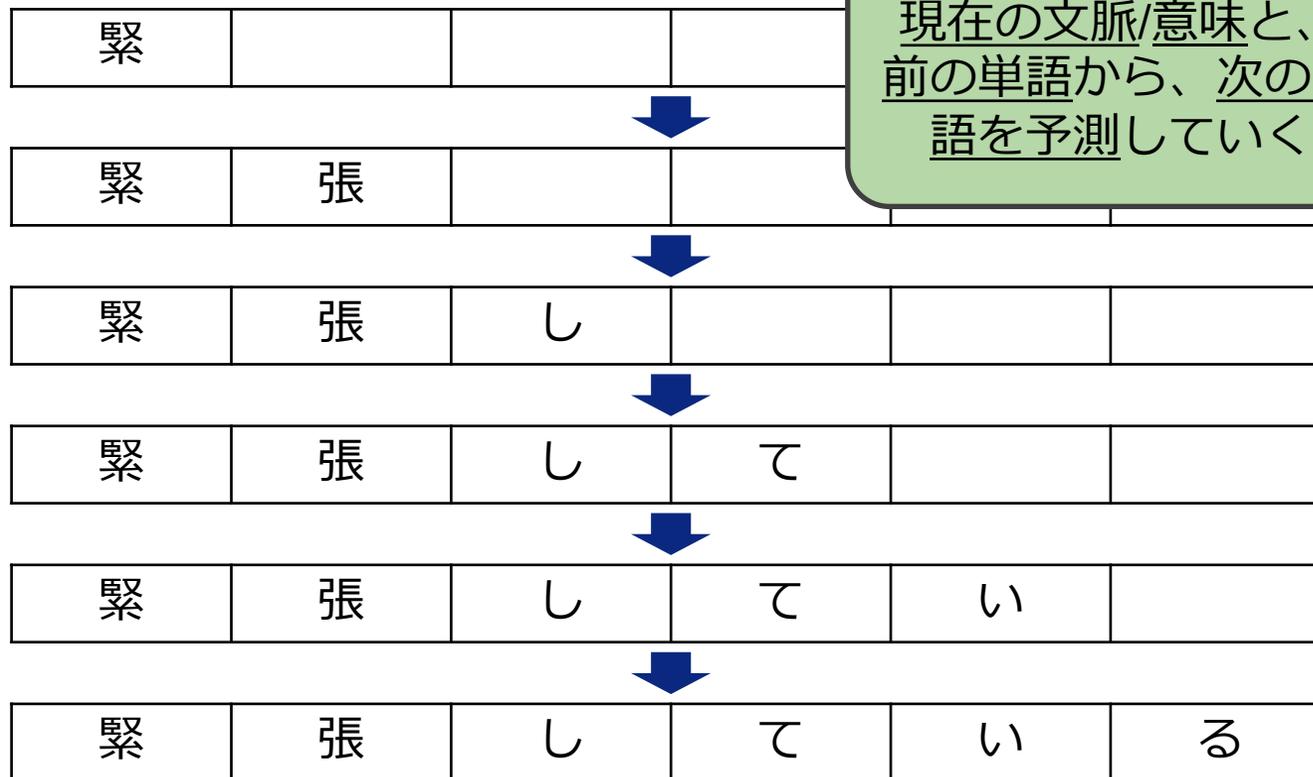
# GPT = Generative + Pre-trained + Transformer

これからJANOG54で登壇をします、私はとても

入力

LLM/GenAI Model (例 GPT)

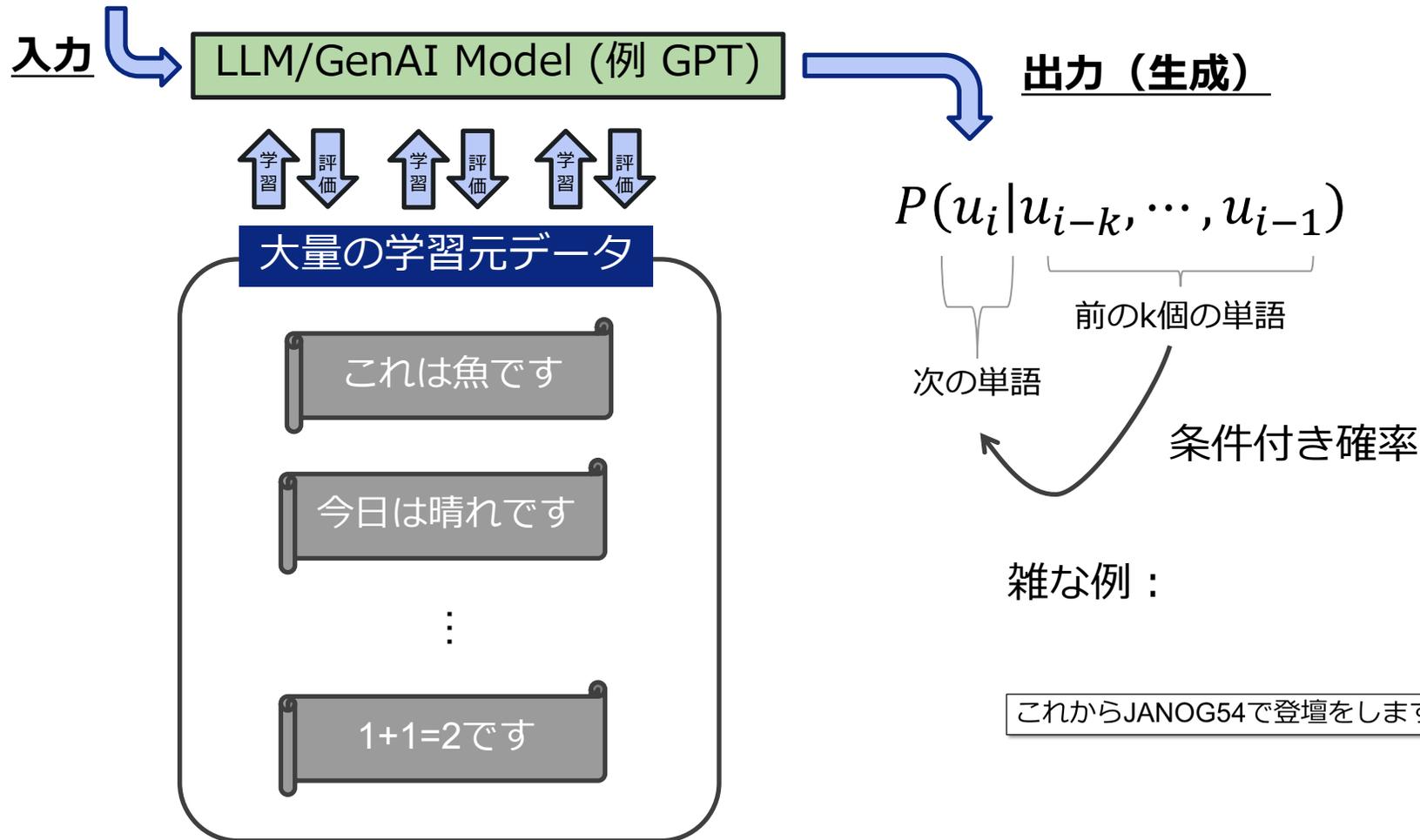
出力 (生成)



現在の文脈/意味と、  
前の単語から、次の単語  
を予測していく

# GPT = Generative + Pre-trained + Transformer

これからJANOG54で登壇をします、私はとても



大量のテキストデータから、自然言語における次単語の条件付き確率（言語モデル）を学習する

雑な例：

	出現確率
緊張している	55%
楽しみです	30%
怒っています	10%
悲しい	5%

## ■ Transformer

- 2017年にGoogleより発表された論文「Attention Is All You Need」にて紹介
- Attention層のみを用いた、機械翻訳モデルとして紹介
- 機械翻訳だけでなく、汎用的な自然言語処理、映像解析等に広く利用されている（2024年現在）

## ■ Attention、Multi-Head Attention

- Transformerの最大のポイント。単語間の類似度を学習し、どの単語に“注目”すれば良いか考えることができる

Scaled Dot-Product Attention

QとKの各ベクトルの内積  
→単語間の類似度を表す

$$\text{Attention}(Q, K, V) = \text{softmax} \left( \frac{QK^T}{\sqrt{d_k}} \right) V$$

Query, Key, Value

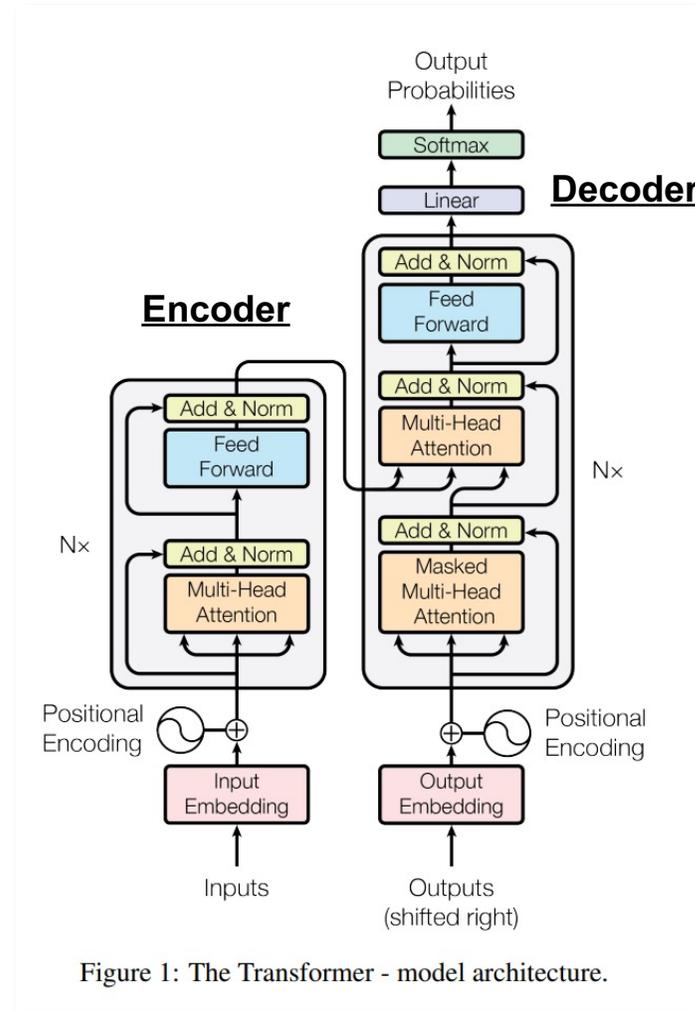
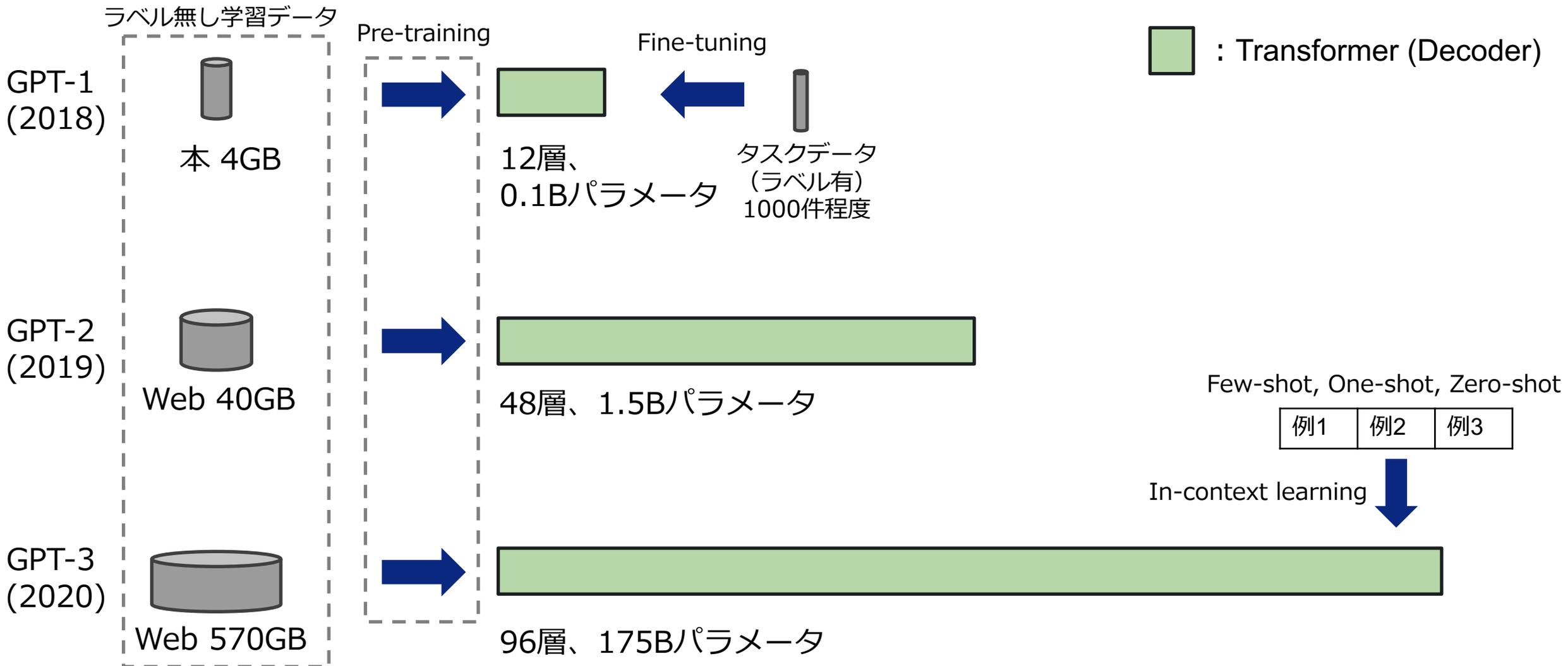
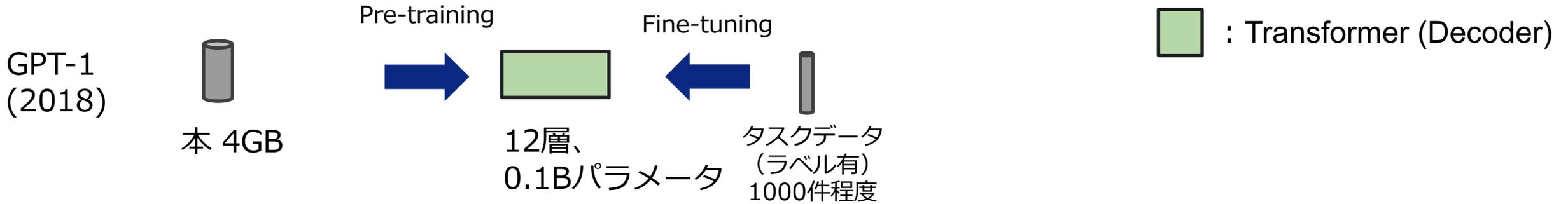


Figure 1: The Transformer - model architecture.

<https://arxiv.org/abs/1706.03762>

# GPT-1~3の各論文における比較





## ■ GPT-1(2018年)

### ● 技術的特徴

- Transformer(Decoder)を12層重ねたモデル
- **ラベル無しデータのみでPre-training** (次単語の予測モデル)
- Pre-trainingされたモデルを利用して、**本タスクをFine-tuning** (ラベル有りデータ)

### ● すごかったところ

- 1つのモデルで、多様なタスクに対応できる
- Transformerのすごさ
- ラベル無しデータで学習できること

<https://openai.com/index/language-unsupervised/>



## ■ GPT-2(2019年)

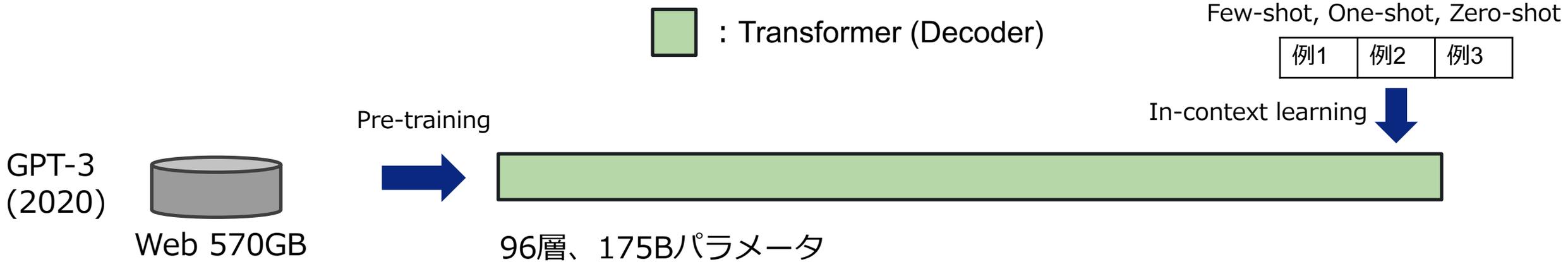
### ● 技術的特徴

- モデルとしては、GPT-1と基本的には変わらない
  - Layer Normalization層が一部追加
- **層の数を大きくした** : 12,24,36,48(1542Mパラメータ)
- **Pre-training用のデータを新しく作成&大量に用意**
  - Redditにおいて、3karma(いいね的なもの)以上ついた投稿のものを収集

### ● すごかったところ

- 大規模データ、大規模モデルの言語モデルの可能性を示した
  - Fine-tuning無しで、様々なタスクに利用可能
  - スケール則 : 規模が大きくなればなるほど精度が向上

<https://openai.com/index/better-language-models/>



## ■ GPT-3(2020年)

### ● 技術的特徴

- モデルとしては、GPT-2と基本的には変わらないが、Sparse Transformerを用いることで高速化を図る
- **層の数を更に大きくした** : 96層(175Bパラメータ)
- **Pre-training用のデータをs更に大量に用意** (570GB)
- **Few-shot** : 自然言語でタスクの例を数点、言語モデルに与えること (In-context learning)
  - 例を与えることで、タスクのコンテキストを提供する

### ● 当時すごかったところ

- 更なる大規模データ、大規模モデルによる高精度化
- 上記モデルに対して、数個の例 (Few-shot) を与えるだけで、どんなタスクにも対応できる

<https://openai.com/index/language-models-are-few-shot-learners/>

## Few-shot

このトラブルチケットの原因を教えてください。

<トラブルチケットのURL>

例として、

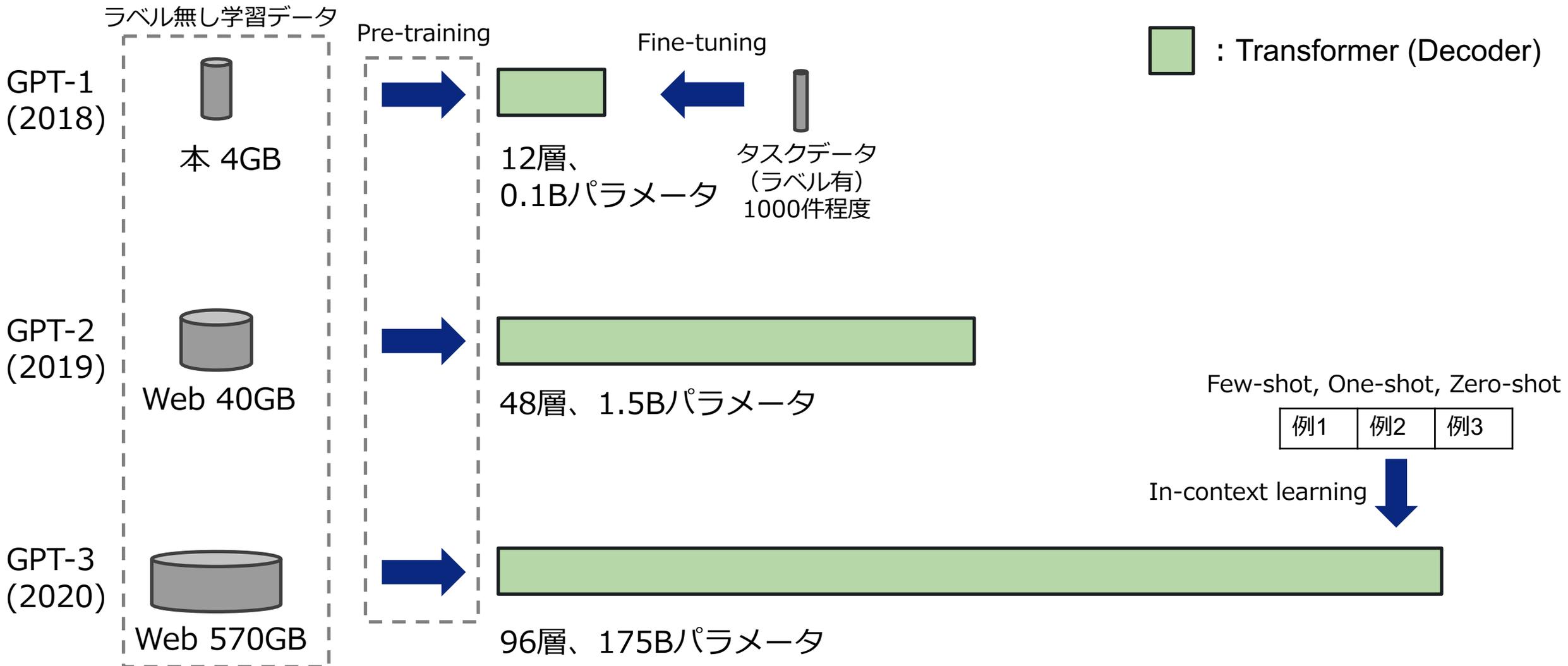
- ・ポートが落ちていたらネットワークの問題が考えられる
- ・CPU利用率が上がっていたらプロセスの異常が考えられる

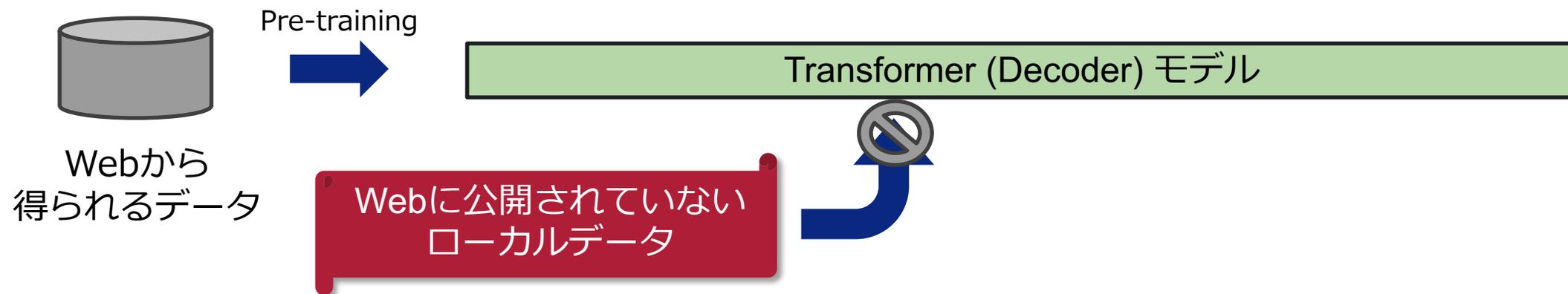
## Zero-shot

このトラブルチケットの原因を教えてください。

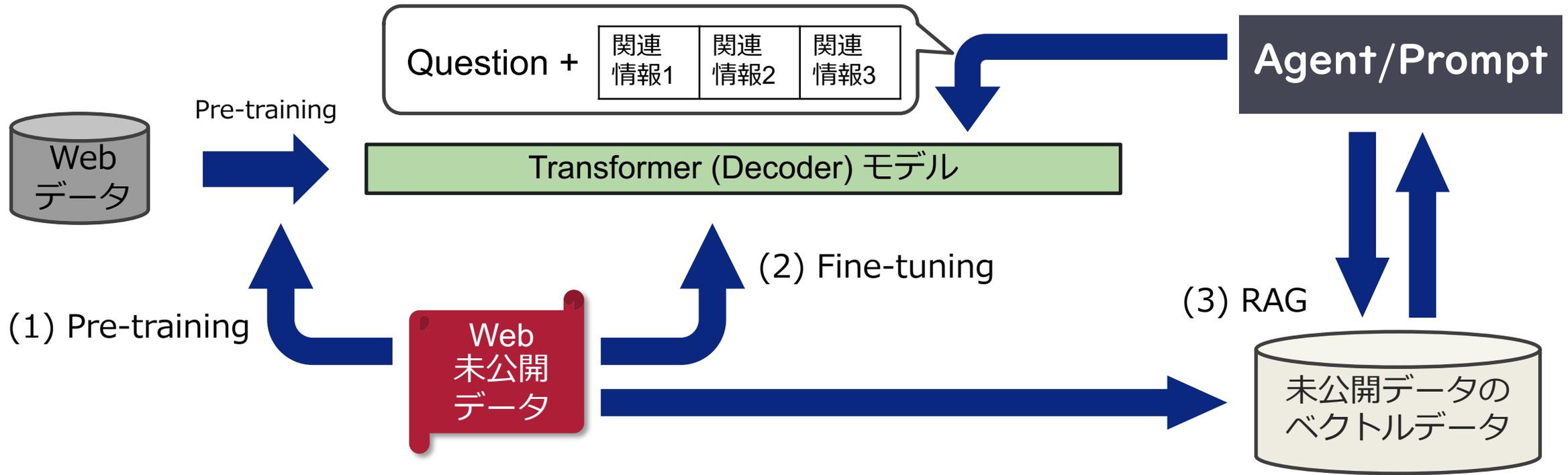
<トラブルチケットのURL>

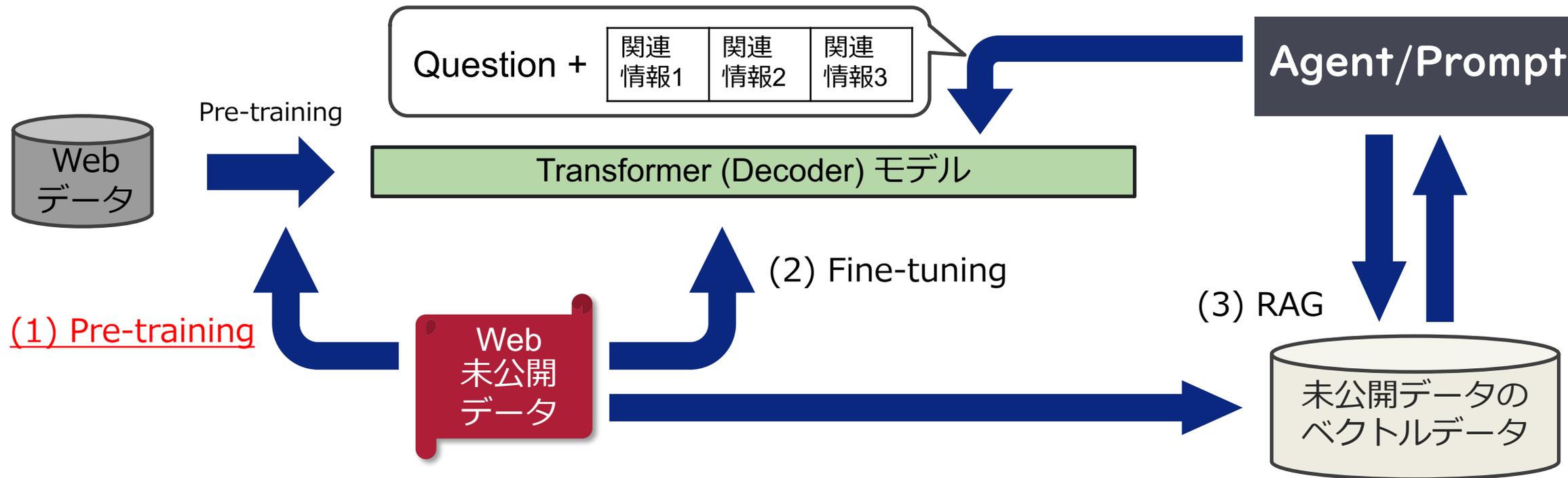
# (再掲) GPT-1~3の各論文における比較





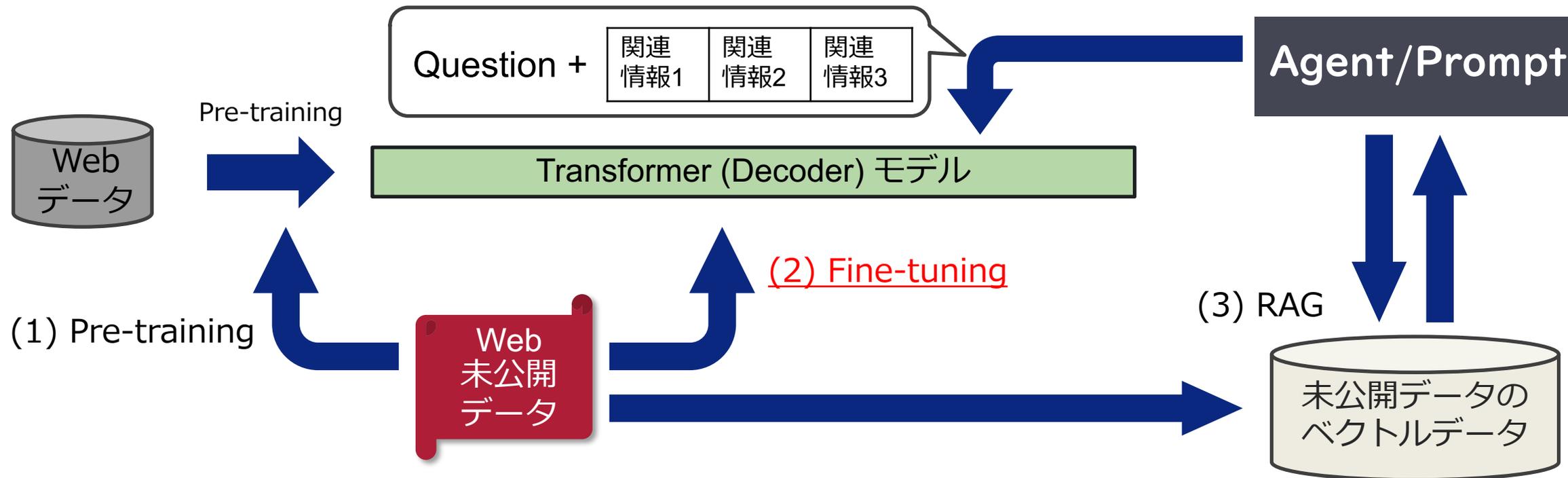
- 事前学習では、Webで公開されているテキストデータから、言語モデルを訓練  
→ つまりWebに非公開のデータは学習されていない
- しかしながら、社内情報等のWebに公開しないデータは多く、それらを言語モデルに学習させて、出力を得るには何か工夫が必要
  - ネットワーク運用における非公開データ例
    - ・ トラブルチケット情報
    - ・ 各種仕様書





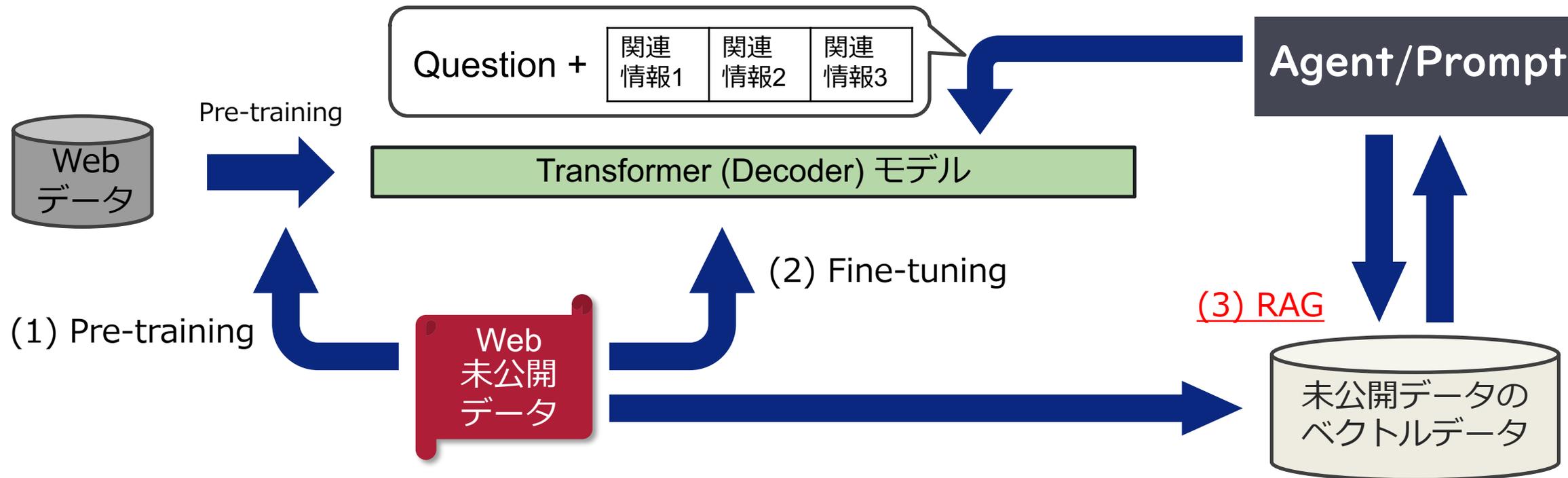
## (1) Pre-training

- 事前学習に未公開データを追加して、言語モデルをチューニングする
  - ・ 追加事前学習、継続事前学習などの方法あり



## (2) Fine-tuning

- 未公開データに関するタスクをある程度の数（例：千件）ラベル付きで用意して、言語モデルをチューニングする
  - Full Fine-tuning (FFT)：すべての層をチューニング
  - Parameter-Efficient Fine-Tuning (PEFT)：限られたパラメータ/層のみをチューニング



## (3) RAG: Retrieval-Augmented Generation

- 未公開データをベクトル化したデータベースを用意
- 与えられたプロンプトに対して **関連するデータを検索**
  - ・ ベクトル化することで、意味的に近い = 関連するデータを検索できる
- 検索結果を言語モデルに付与して聞く

- LLM/GenAIは世の中の様々な業務への適用が進められている
  
- では、ネットワーク運用の世界にはどんな業務へ適用ができるだろうか・・・？
  - 営業
    - ・ 例：顧客対応の自動化
  - 企画
    - ・ 例：社内対応の自動化
  - 開発・建設
    - ・ 例：最適なネットワークアーキテクチャ検討の手助け
  - 運用
    - ・ 例：トラブルチケットの自動起票、ログデータから根本原因特定
  
- このあと、各社の事例や検討内容について共有してもらいます！

内容	発表者	目安時間
• LLM概要とネットワーク運用への適用検討	 宮坂 拓也 株式会社KDDI総合研究所	15分
• ネットワーク運用への活用検討事例 (KDDI編)	 仲松 匠 KDDI株式会社	15分
• ネットワーク運用への活用検討事例 (NTT-FT編)	 佐藤 亮介 株式会社NTTフィールドテクノ	30分
• 議論	 白井 嵩士 株式会社NTTフィールドテクノ	30分
	 田口 順史 株式会社NTTフィールドテクノ	
	 近藤 健吉 株式会社NTTフィールドテクノ	
	全員 + 参加者 + 司会	

内容	発表者	目安時間
• LLM概要とネットワーク運用への適用検討	 宮坂 拓也 株式会社KDDI総合研究所	15分
• ネットワーク運用への活用検討事例 (KDDI編)	 仲松 匠 KDDI株式会社	15分
• ネットワーク運用への活用検討事例 (NTT-FT編)	 佐藤 亮介 株式会社NTTフィールドテクノ	30分
• 議論	 白井 嵩士 株式会社NTTフィールドテクノ	30分
	 田口 順史 株式会社NTTフィールドテクノ	
	 近藤 健吉 株式会社NTTフィールドテクノ	
	全員 + 参加者 + 司会	

トピック	質問例
導入背景・目的	ネットワーク運用の現状の課題は何ですか？ LLMがこれらの課題をどのように解決できると考えますか？
	LLMの導入を検討するにあたって、他の技術と比べての利点は何ですか？
技術面	上記の課題を解決するために、LLMに追加するローカルデータとしては、どのようなデータが必要ですか？
	ローカルデータによるLLM拡張に対して、どのような手法がよいと思いますか？（例：RAG、Fine-tuning）それは何故ですか？
運用面	LLMを運用現場へ導入する場合、どのようなステップが必要になりますか？それは他のシステムと変わりますか？
	導入後の運用・メンテナンス（例：モデル更新）の中でどのような課題が発生すると思いますか？

トピック	質問例
<b>セキュリティとリスク管理</b>	LLMが引き起こす可能性のあるセキュリティリスクにはどのようなものがありますか？
	データプライバシーの観点から、LLMのトレーニングに使用するデータはどのように管理すべきでしょうか？
<b>組織への影響</b>	LLMの導入により、運用チームの役割や業務内容はどのように変わるとお考えですか？
	LLM導入のために、運用者に対して新しいスキルやトレーニングの必要性はありますか？
<b>実際の事例</b>	LLMを導入して成功した具体的な事例を教えてください。その成功要因は何でしたか？
	失敗した導入事例があれば、その原因と学んだ教訓を教えてください。