踊るバズワード ～Behind the Buzzword（18）：

ChatGPTは怖くない ～使い倒してラクをせよ

https://eetimes.itmedia.co.jp/ee/articles/2305/15/news060.html

ある日突然登場し、またたく間に世間を席巻した生成AI「ChatGPT」。今や、ネットでその名を聞かない日はないほどです。このChatGPTとは、一体何なのか。既に数百回以上、ChatGPTを使い倒している筆者が、ChatGPTの所感をエンジニア視点で語ってみたいと思います。

2023年05月15日 11時30分 更新

[江端智一, EE Times Japan]

「業界のトレンド」といわれる技術の名称は、“バズワード”になることが少なくありません。“M2M”“ユビキタス”“Web2.0”、そして“AI”。理解不能な技術が登場すると、それに“もっともらしい名前”を付けて分かったフリをするのです。このように作られた名前に世界は踊り、私たち技術者を翻弄した揚げ句、最後は無責任に捨て去りました――ひと言の謝罪もなく。今ここに、かつて「“AI”という技術は存在しない」と2年間叫び続けた著者が再び立ち上がります。あなたの「分かったフリ」を冷酷に問い詰め、糾弾するためです。⇒連載バックナンバー

「文章嫌い」を劇的に変えた、あのマシン

私は、文字を書くのが嫌いな子どもでした。

小中学生の頃に強要された「漢字の書き取り」は拷問でした。ですから、「大工」という漢字を、毎日100文字書いて、提出していただけでした。右の書取りノートを見ただけで、今でも、嘔吐感が込み上げてきます。

私は、文章を作るのが嫌いな子どもでした。

「毎日の日記」の記載と提出に至っては、何を書いて良いのか分からずに、毎日、以下の3行フレーズを、『てにおは』を変えて提出するだけでした。

今日は、天気だった。
今、お風呂から上って、いい気分だ。
明日もいい天気かな

上記の「大工100文字」の書き取り、そして、「3行お風呂日記」を毎日提出していましたが、別段教師から叱責されたことはありません。私だけでなく、教師の方も**『こんなもの、どーでもいい』と考えていたことは明らか**でした ―― このようなくだらない無駄が、対象をいろいろ変えながらも、高校卒業まで続きました。

ところが、この私の、『文字を書くのが苦痛、文章を作るのが嫌い』を、一瞬にしてひっくり返すものが登場します ―― ワードプロセッサ（ワープロ）です。

私の記憶の中にある最初のワープロは、液晶画面に10文字しか示せず（本当）、自分が、今、原稿のどの辺を書いているかは、その液晶上のビュー画面に切り替えて確認するしかない、というものでした。

『確か、購入したのは大学の1年生の秋だった』という記憶と合わせて調べてみたところ、多分、そのワープロは、右図（東芝JW-R10）であったと推認されます。

なぜ私が、当時の最高級の機械（ワープロ）を使えたかというと、パソコンショップの店員をやっていたからです。『自分が使えないものは、お客様には売れない』と思いましたので、パソコンショップで、マニュアルと首っ引きで、使い方を勉強していました。

そうしているうちに、そのパソコンショップの広告やら、商品説明のラベル（今でいうところの”ポップ”）の作成なども頼まれるようになりました。

そして、私は、店員割引価格で購入したJW-R10によって、人生のシンギュラリティポイントを迎えることになります ーー **第2の江端の覚醒**です。

覚醒後の私は、大量文章製造マシンと化します。自分が思考する速度で、そのまま文章が書けることに驚き、頭の中を全部文章で書き出す勢いでした。しかしながら、その道のりは険しいものでもありました。

大学在学中、私はワープロで書いた手紙を友人に出して、友人から「**思いが伝わってこない**」と非難されました。大学に「ワープロで作成した実験レポート」を提出した時には、大学教授からは『**手書きレポートしか受理しない**』と言われ、ワープロの文章を、再度手書きに書き直しました（本当）。

しかし、大量文章製造マシン・江端はくじけません。会社に入社後には「日立こぼれ話」という内部暴露コラムを、定期的に友人や親類に『紙に印刷』して『郵送する』という、（今になって思えば、手紙を受けとった人には随分な迷惑な）行為に及びます*）。

*）未公開の「日立こぼれ話」 総計456話は、私の退職後に、一斉公開予定です。

その当時、電子メールを使える人はごく一部の人間でした。電子メールを使う人は、『オタク』と言われて、世間から石を投げられる存在だったのです。しかし、私は大量文章製造を止めようとは思いませんでした ーー **ラクだったから**です。

そして、今回のコラムの趣旨は、この**『ラクは正義である』**です。

編集担当Mさんからの、1本のメール

こんにちは、江端智一です。

ご報告が遅れましたが、私、2022年10月、社会人大学院に入学しました。その動機や経緯や詳細については、おいおいお話させていただきたいと思いますが、社会人と大学生の両立は、私の想像を越えた世界でした。どういう世界かというと、『**激務で、入院がスコープに入ってくるくらい**』でした。

まだ、半年しか経過していませんが、既に大学院の入学を後悔し始めています。しかし、皆さんに語るべきネタは格段に増えました。この大学のネタだけで2〜3本くらい、コラムが書けると確信できるレベルになっています。

そういう背景があり、EE Times Japanの編集担当のMさんに事情を説明し、『**倒れますか？ 書きますか?**』という感じの~~『恫喝（どうかつ）』~~『ご相談』をさせていただき、私の執筆の頻度を調整していただいています。

ところが、先日、Mさんから、珍しく、以下のメールが届きました。

江端様

お世話になっております。

次回以降の原稿につきまして、ご相談です。

「お金に愛されないエンジニア」の続きも読みたいのですが、「踊るバズワード 〜Behind the Buzzword」として、ChatGPTを取り上げていただけないでしょうか。

ChatGPTはまさしく「バズワード」になっていて、技術的な優位性と社会的な問題点とが、毎日のようにアップデートされ、カオスになっていると感じます。この辺りで、江端さん視点で語っていただきたいのです。

一度、江端さんにChatGPTのテーマを取り上げていただく、ということが重要なので、結論が「よく分からん」「興味がわかない」などでも構いません。

ご検討いただけますと嬉しいです。よろしくお願いいたします。

編集担当M

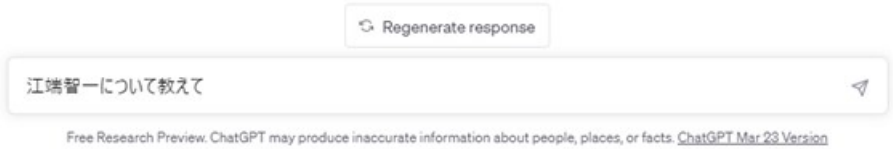
　このメールを見た瞬間、

―― やはり、来たか

と思いました。

　最近、NHKのニュースでも、ChatGPTについて毎日のように報道されていますし、新聞ではトップニュース扱いです。そして、これまでのAI技術のアプリケーションとしては、最も多くの人に使われていることは間違いありません。なにしろ、こんなラクなインタフェースで使えるAI技術のアプリケーションは、人生でもめったに出会えるものではありません。

　例えば、「江端智一について教えて」と入力するだけです。こんな感じで。



　このMさんからのご依頼は、私にとっても「渡りに船」のお話でした。

　私は、こちらの連載「Over the AI ――AIの向こう側に」において、ちゃんとした根拠と自分なりの検証結果を用いて、**”AI幻想論”**を繰り返し主張し続けてきました。

　しかし、このChatGPTと出会ってからは、『もしかしたら、私の"AI幻想論"は、壊れたのかもしれない』という、疑問（というか恐怖）を感じました。何しろ、本当の人間が、私の質問に応えているかのように感じられたからです。そして、実際にこのChatGPTを仕事や大学の勉強に使い倒しており、その便利さに、今なお、驚き続けているからです。

　「非常に優れたメンター、またチューターが、私の横に座っているかのような安心感」は、これまでのAI技術にはない経験で、私の”AI幻想論”を激しく揺らしていたからです。

　ですので、前々から、『ChatGPTについて、チャンスがあればちゃんと調べて理解したい』と、と思っていたのです。

　現時点で、ChatGPTは、課題山積の成長過程にあるAI技術であるのは事実です。例えば、上記の「江端智一について教えて」に対する**4月29日**のChatGPTの回答は以下の通りでした。

江端智一について教えて

江端智一（えばた ともかず）は、日本のプログラマーであり、AIの研究者でもあります。彼は、OpenAIという非営利研究組織で働いており、大規模言語モデルの開発に関わっています。また、彼はGPT-3という言語モデルの開発に重要な役割を果たし、その性能に貢献しました。

江端智一氏は、東京大学大学院情報理工学系研究科で博士号を取得しており、専門分野は自然言語処理と深層学習です。彼は、言語処理技術を利用したさまざまな応用分野の研究に取り組んでおり、その成果は高く評価されています。

　この「江端智一」なる人物が、α世界線、β世界線、はたまたシュタインズゲート世界線の江端智一なのか、私には分かりません*)が ―― ともあれ、

―― 今日も、ChatGPTは、盛大に、私を爆笑させています。

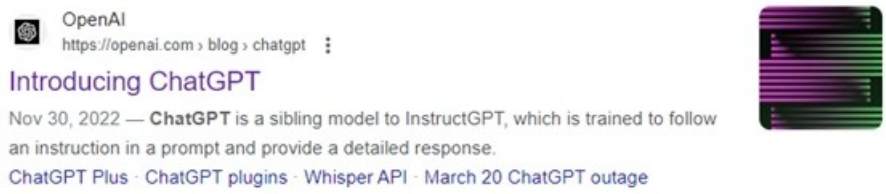
*）関連記事：「「シュタインズ・ゲート」に「BEATLESS」、アニメのAIの実現性を本気で検証する」

それでは、始めたいと思います。

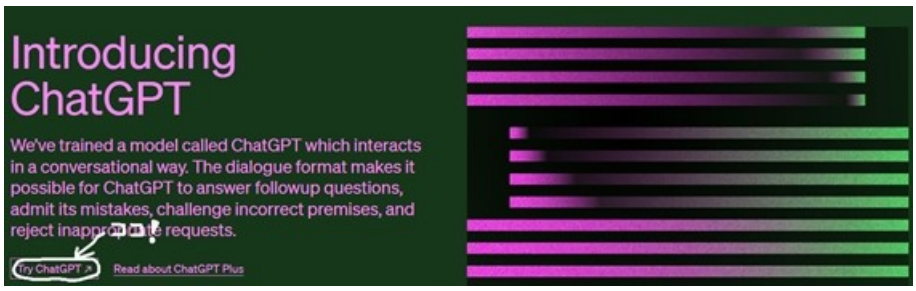
ChatGPTの回答「江端智一について」を添削してみた

私が、連載「Over the AI ――AIの向こう側に」で論じてきたAI技術の多くは、ロクなインタフェース（Web入力画面など）を持っていませんでしたが、ChatGPTの使い方はとても簡単です。3ステップくらいで使えます。

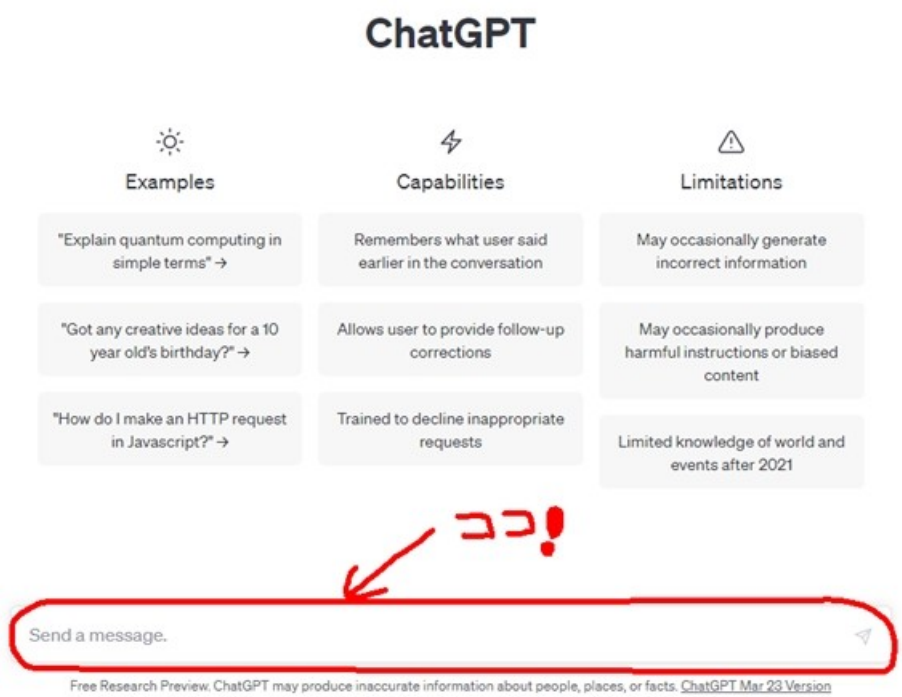
（Step 1） Google検索で”ChatGPT”と入力する



（Step 2） これをクリックして、”Try ChatGPT”ボタンをクリックする



（Step 3） ここに好きなフレーズ（例:『江端智一について教えて』）を入力してリターンする



これだけです。ログインアカウントとパスワードの入力を求められるかもしれませんが、まあ、その辺は適当にアカウントを作成してください（私は、Googleのアカウントログインをそのまま使いました）。

先程の**4月29日分**のChatGPTの文章を、添削してみましょう。なにしろ、私自身の経歴が解答ですから、添削は簡単です。

■江端智一（えばた ともかず（いち））は、

■日本のプログラマーであり、~~AIの研究者でもあります~~（AI研究者ではありません）。

■彼は、~~OpenAIという非営利研究組織で働いており~~（そもそも、OpenAI自体を知りませんでした）、

■~~大規模言語モデルの開発に関わっています~~（言語モデル自体、知りません）。

■また、彼は~~GPT-3という言語モデルの開発に重要な役割を果たし~~（GPT-3は、今回のコラムの下調べで初めて知りました）、

■~~その性能に貢献しました。~~（“貢献”以前の問題です）

■江端智一氏は、~~東京大学大学院情報理工学系研究科で博士号を取得しており~~（現在、別の大学の社会人大学院で、博士号取得を目指してヘロヘロになっているところです）、

■専門分野は~~自然言語処理と深層学習~~（あえていうのであれば、今なら”エージェントシミュレーション”でしょうか）です。

■彼は、~~言語処理技術を利用したさまざまな応用分野の研究に取り組んでおり~~（言語処理には、1mmも関わったことがありません）、

■~~その成果は高く評価されています~~（“評価”以前の問題です）。

以上のように、**4月29日分の**ChatGPTの答は、**95%がデタラメで、唯一の正解は、「日本のプログラマー」の1つだけ**でした。

この他、別の日の「江端智一について教えて」に対する**愉快なChatGPTの解答例**につきましては、こちらをご覧ください。

「江端智一について教えて」で、最高傑作は、『東京大学出身博士号取得者』『文部科学省の事務次官』『政界に進出した国会議員』で、『国際的なAIの枠組みに取り組む国際機関のトップ』でした（このログのコピペを取り忘れていたのは、返す返す残念です）。

しかし**『ChatGPT = バカ』ではない**のです。

私にとって、最大の威力を発揮したのは、英語論文の概要把握です。私はこれで随分ラクをさせてもらいました。数百の英語論文の概要を、数日でまとめることができたからです。

しかし、これらの論文の概要の内容が、**どの程度妥当なもの**なのかを調べることはしていませんでした。

ChatGPTが返してきた論文概要は、どれくらい妥当なのか

そこで今回、私の執筆した論文やカンファレンスペーパー（以下、”ペーパー”といいます）を使って、これを調査してみました。なにしろ自分が書いたものですから内容は熟知しています。対象としたペーパーは以下のものにしました。

ここ数年で江端の執筆したペーパー一覧

対象	概要
論文 (採択済み)	An On-Demand Alternate Transportation Service System for Public Transportation using Real-Time Multi-agent Technology
国際学会 (採択済み)	Proposal of Supply and Demand Mediation type Transportation Service based on Dissatisfaction
	Using User Dissatisfaction to Bridge Transportation Gap with Supply-and-demand Mediation-type Service
	Safe and Secure Driving for Supply and Demand Mediation type Transportation Service
標準化	Internet-Draft, draft-ebata-inter-domain-qos-acct-00.txt

以下の表は、私のペーパーの中で、私が渾身を込めて主張したいと考えていた内容と、ChatGPTが、その **私の「渾身」に対して、どれだけ応えてくれたか**を、主観値で評価したものです（100点満点）。

期待していたChatGPTの解答と現実

対象	江端がChatGPTに期待していた解答	江端によるChatGPTの解答の評価
論文 (採択済み)	100万人分のエージェントを使って、鉄道人身事故発生後の2時間後を予想して、5分後に代替バス計画を完了する	50点
国際学会 (採択済み)	鉄道路線の保守作業を行う複数の会社の間のスケジューリングを、「不満」を使って調停する	50点
	利用者と交通事業者の「不満」を調停して、リアルタイムで運行計画を立案/変更する	50点
	突然運行ルートが変わってしまう、リアルタイム運行のリスクを勘案して、リスクのない運行計画を実現する	50点
標準化	複数のネットワークサービスプロバイダを調停して、QoSを保証する	85点

感想は、

「概要は間違ってはいないか、一般論の域を出ない凡庸な内容だった」
「私（江端）のオリジナリティの主張が全く取り出されていなくて、とても悲しかった」

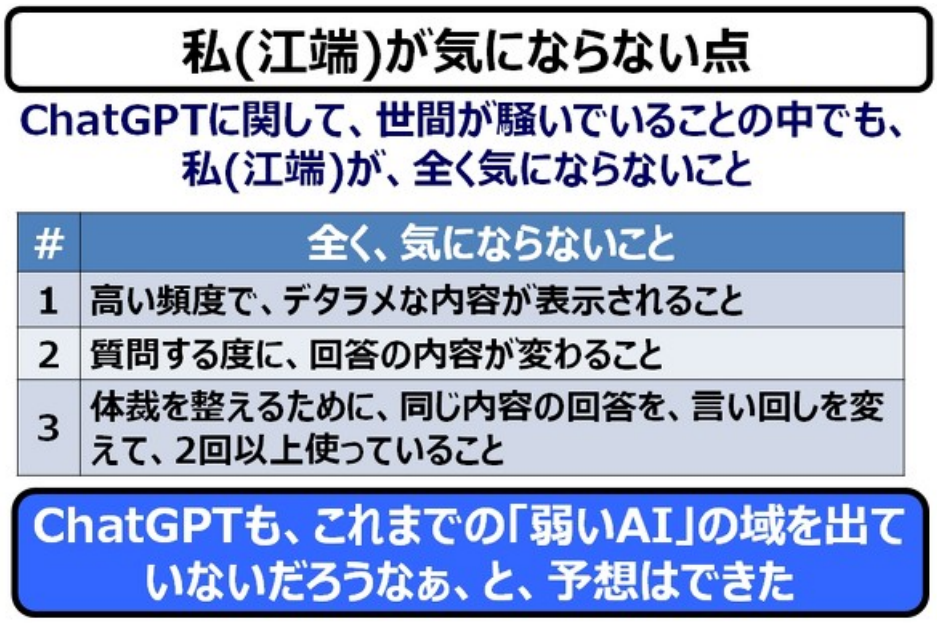
です。

これは邪推の域を出ませんが、「**ChatGPTは、ペーパーの表題だけを読んで、内容を推測しているんじゃないか？**」と疑ってしまうレベルです。

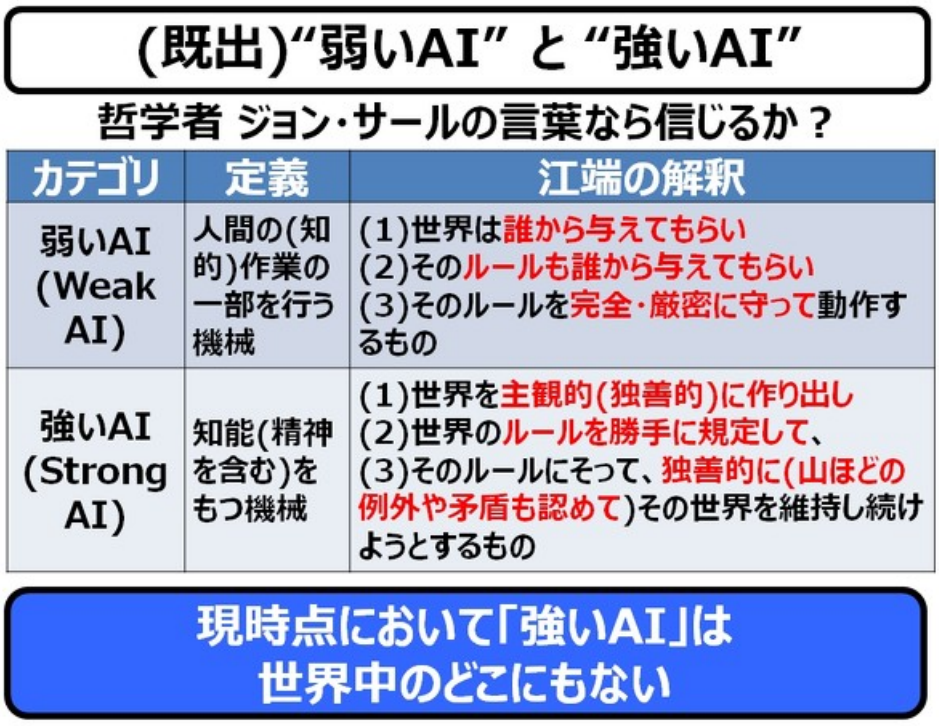
まあ、それでも、論文調査においては、概要把握ができるだけでも上出来と言えます。それに、普通の人間でも、著者の『熱い思い』まで読みとるのは、難しいことですから。

総じて、私、昨今のAI技術って、こんなもんだと思っていますので、それほど気にしていません。というか、これで十分だろ

う、と思っています。この気持ちをまとめてみると、こんな感じです。



『弱いAI』『強いAI』については、「弱いままの人工知能 ～ “強いAI”を生み出すには「死の恐怖」が必要だ」で説明していますが、ざっくりとした説明を付けておきます。



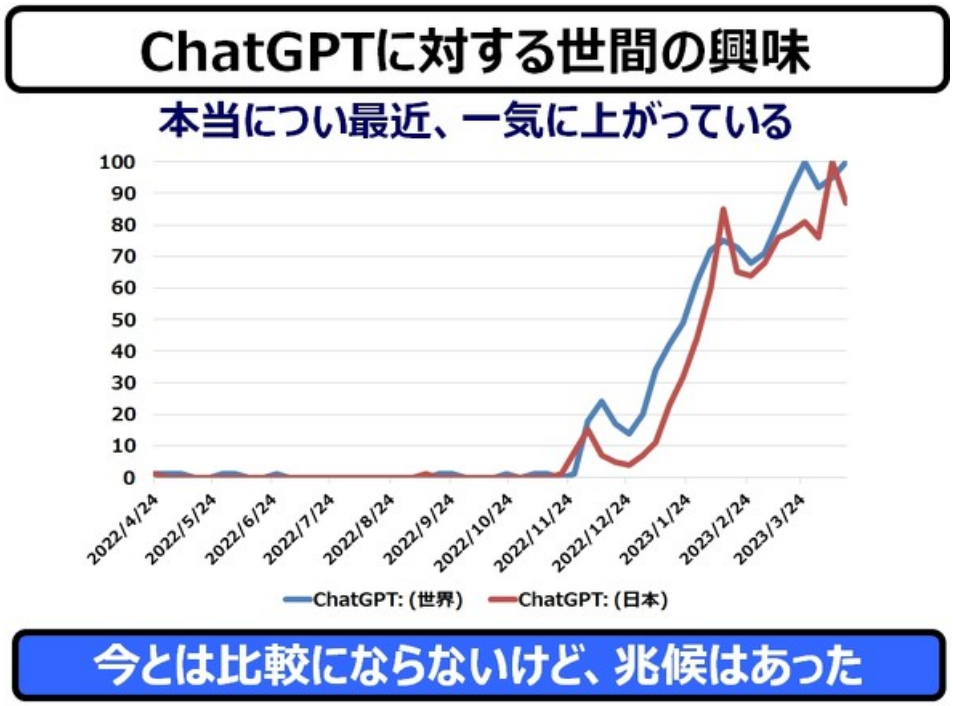
私の、ChatGPTに関する興味はもっと別のところにあります。これについては後述します。

ChatGPTの「芽生え点」が見つからない

ChatGPTは、ある日突然現れて世界を席巻したかのように見えます。

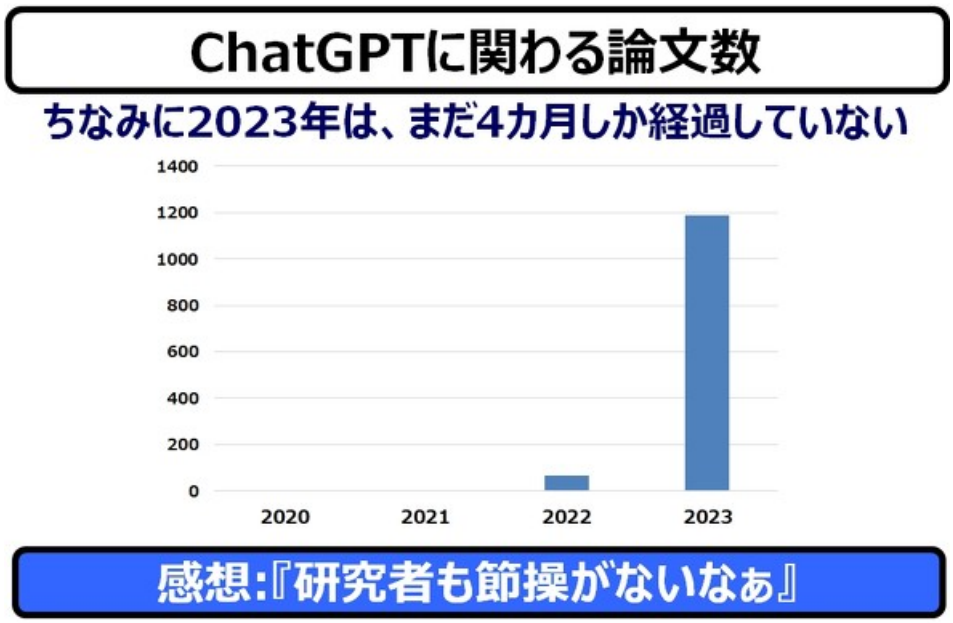
多くのAI技術は、その前の段階で、ブレークポイントの芽生えのように見える場面がありました。ですので、私もChatGPTの“芽生え点”を求めて探してみたのですが ―― **“見つからない”**のです。今でも、ちょっと信じられませんが。

以下は、Google Trendを使って、”ChatGPT”が登場してくるニュースヘッダ数の比率の推移をグラフにしたものです。



昨年2022年の11月下旬に現われるやいなや、いきなり世界を席巻したのです。これは、従来のAI技術のパターンとは一線を画すものであると言えます。

こういう場合は、学術方面の論文数から調べることで、世間とアカデミズムの乖離が「見える化」できますが、驚いたことに、アカデミズムにおいても、ChatGPTが完全に見落とされ続けていたことも分かりました。これは、下記のChatGPTを含む論文数の推移から見ても明らかです。

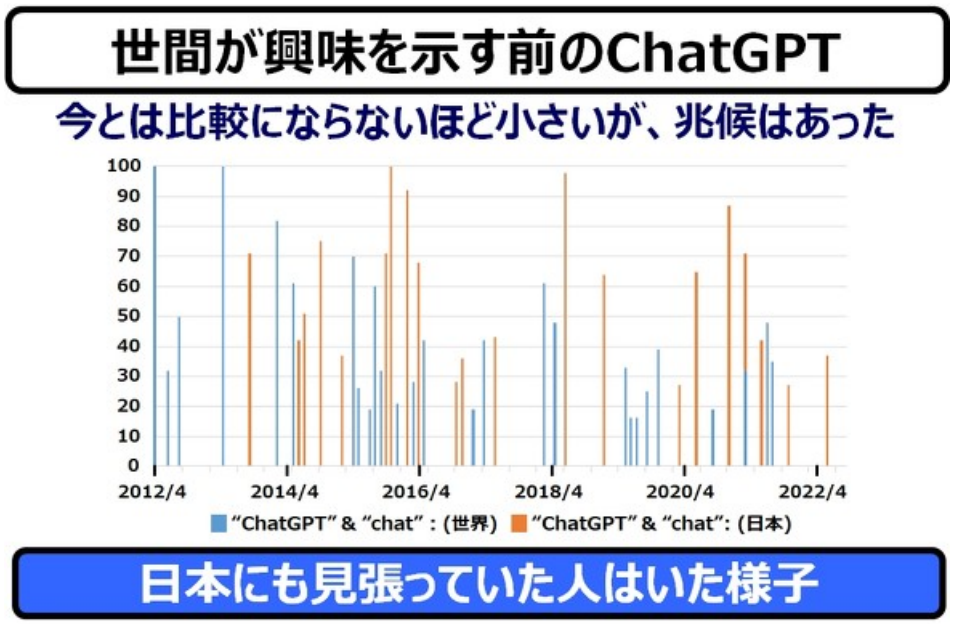


世界中の研究者達が、ChatGPTに全く気がついておらず、今、慌てて論文を出しまくっているという、**節操のない様子**が伺えます。加えて言えば、ChatGPTに関する**技術的観点の論文は皆無に等しく**、論文のテーマは、「ChatGPTの活用方法」や「ポストChatGPTの社会への影響」という内容になっています。

つまり、**ChatGPTを技術的に理解していない自称「AI専門家」が、慌ててChatGPTについて語り出しているという『見苦しい』状況**が見て取れます ―― 実際に、私、今回の調査で、ChatGPTの技術に関する文献が少なすぎて、本当に困りまし

た。

ただ、完全にChatGPTが見落とされていたかというと、決してそういう訳でもないようです。ブレークポイントとなった2022年11月後半より前の、ChatGPTに関するニューストレンドの比率を調べてみたところ、現在の数から見ると、それはほこりのように小さい数ではありますが、世界だけでなく日本でも、ちゃんとウォッチし続けた人（研究者）がいたと推認できるデータもあります（下図参照）。



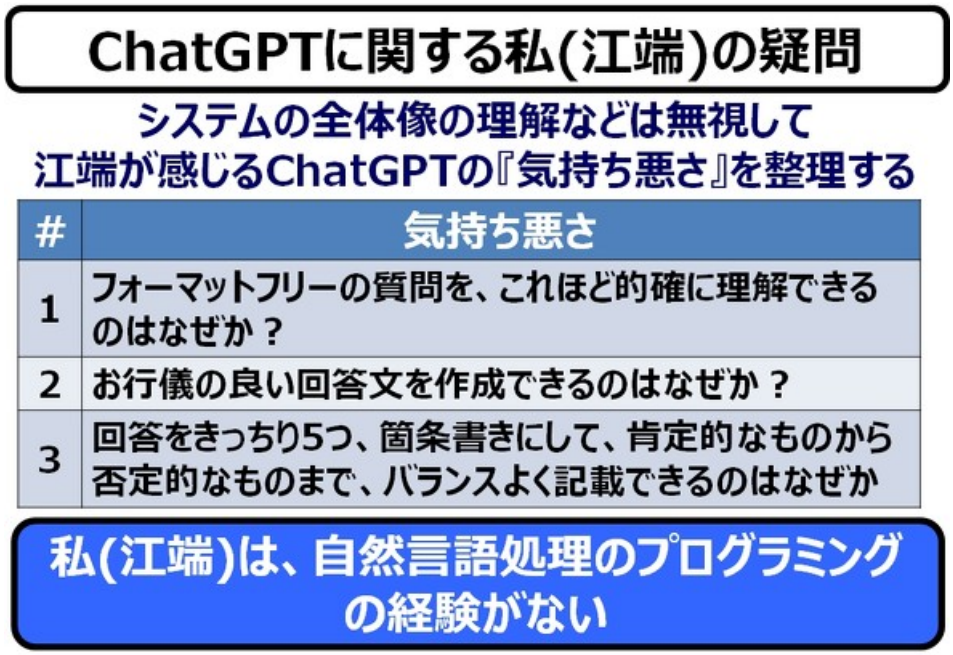
総じて、このChatGPTという技術は、AI研究者にとどまらず、登場と同時に一般の人が簡単に使える（上記の3ステップ）AI技術であることが、大きな特徴と言えそうです。

ChatGPTは「気持ち悪い」!?

さて、今回のコラムでは、ChatGPTのシステム全体について解説するつもりはありません。

このコラムでは、私（江端）が、ChatGPTについて『**気持ち悪！**』と考えるところだけに注力して解説します。私が理解できているところは、平気でふっ飛ばしますが、悪しからずご了承ください。

私がChatGPTに対して『気持ち悪！』と思う点は、ざっくりまとめると以下の3つです。



まとめますと、**「なぜ、ChatGPTは、私（たち）に違和感を与えることがないような、人間が語る／書くような形式の出力ができるのか」**に尽きるわけです。

情報の収集については、ネットから情報を集めていることは明らかですし、それはGoogleをはじめ、各種の検索エンジンで、私たちはよく目にしているものですので、その部分については「気持ち悪！」とは思っていません。

しかし、この**ChatGPTの解答やその応答形式は、間違いなく「気持ち悪い」**。

まるで、パソコンの向こう側に人がいるかのような錯覚を与えるほどです。チューリングテスト（ただし、一回限りの応答限定）をすれば、間違いなくテストをパスしそうです。まあ、だからこそ、『ChatGPTには知性がある』と言い出す人が出てくるわけですが、それも無理ないかなぁ、と思うのです。

ちなみに、AIの知性については、この「チューリングテスト」に対するアンチテーゼとして「中国語の部屋」というのもあります（下記の表はご参考です）（関連記事：「弱いままの人工知能 ～ “強いAI”を生み出すには「死の恐怖」が必要だ」）。

(既出)AIの「知性」の考え方

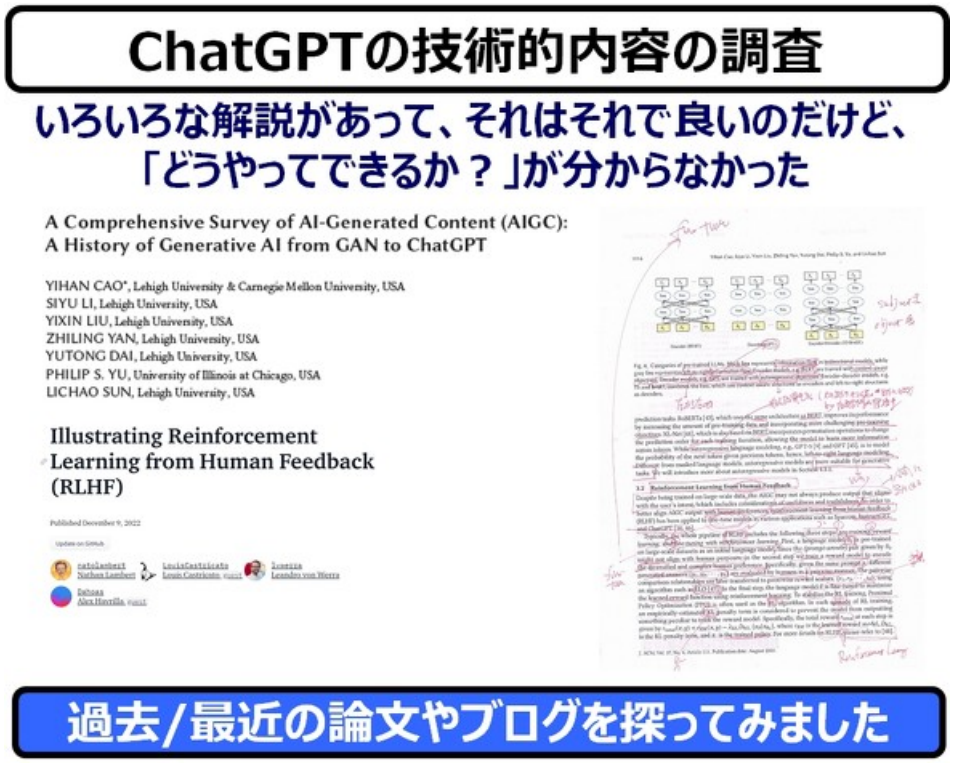
超有名な二つの考え方

名称	概要	その他
(1) チューリング テスト	(Step.1) ■人間が、AIと会話(チャット)している	■あの天才、アラン・チューリングが提唱(1950年) ■イメージとしては「自動応答Botと喧嘩した」等
	(Step.2) ■人間が、そのAIを「人間」と勘違いしたら、	
	(Step.3) ■そのAIには「知性がある」とする	
(2) 中国語の部屋	(Step.1) ■紙切れ1枚しか出し入れできない部屋の中に、アルファベットしか理解できないイギリス人がいる	■哲学者ジョン・サールが思考実験として提唱(1980年) ■チューリングテストの「知性」に対する反論として捉えられる ■現存するAI技術は、ほとんどこのアナロジーで説明できる
	(Step.2) ■紙切れには「『一∧◆●(中国語)」と書かれていれば『ゝ♪♂♀(中国語)』と書き加えてから外に出せ」という記載がある。 ■イギリス人は、ひたすら、その指示に従う	
	(Step.3) ■その部屋の外にいる人は、「この小部屋の中には中国人がいる」と考える ■しかし、実際には中国語なんぞ全く理解していないイギリス人が1人いるだけである	

それはさておき、とにかくChatGPTの技術に関する文献が少なくて、本当に困りました。Amazonで、いくつか入門書を購入してみたのですが、正直「返品しようかな」という内容でした。

ChatGPTの使い方なら、『使えば分かる』し、技術内容であるなら全然説明が足りません。一応、**ChatGPT自身に、「ChatGPTを理解するお勧めの本」ついても質問してみたのですが**、「深層学習」や「機械学習」の本を紹介してくるありさまで、**アテになりません**。

『となると、ちゃんと勉強するしかないかなぁ』と思い、過去の論文を調べることにしました。私が今回参考にさせていただいたのは、以下の資料です。



あと、ほぼ原稿を書き終えてから気がついたのですが、AI研究をやっている人（特に、Deep Learningの学習メカニズムを理解している人）であれば、

Qiitaで公開されているオミータ氏の記事（リンクはこちら）は突出して優れた資料でした。ご一読を、強く、強くお勧めいたします（本コラムの中でも、使わせていただいております）。

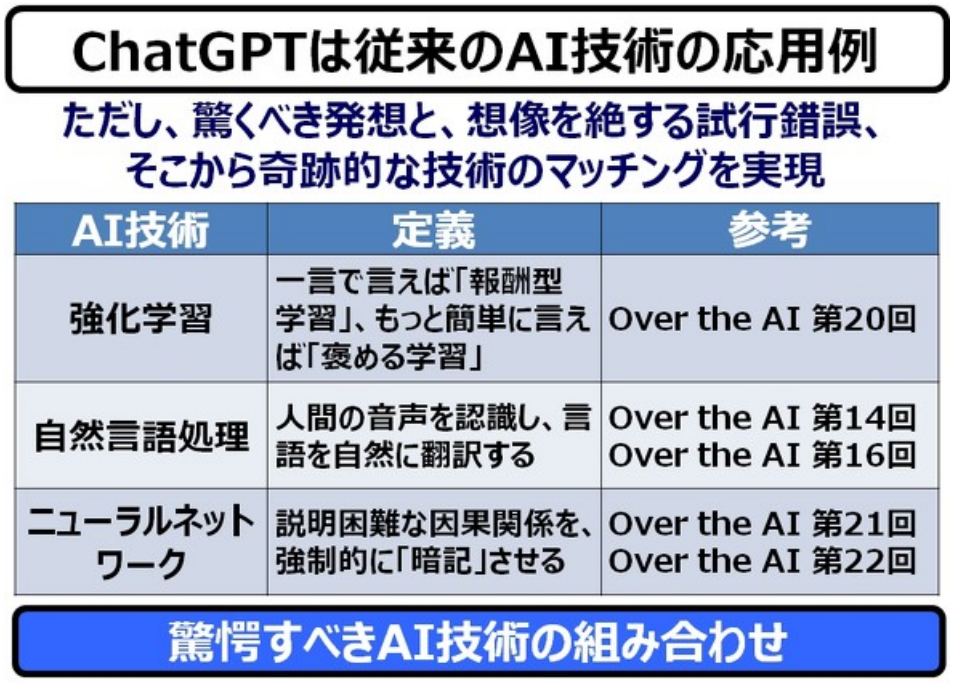
ChatGPTとは何なのか

さて、ChatGPTとは何か、と問われれば、「使えば分かる」と答えれば十分ですが、ここからは、ChatGPTの内部の機能やら学習の方式などについて、私と同様にChatGPTのメカニズムに興味のあるAI研究者、または、ソフトウェア技術者をターゲットとして解説していきます。

まずは、ChatGPTの言葉の意味ですが、これは、**Chat Generative Pre-trained Transformer**の略です。これを日本語に翻訳すると「**事前に訓練された変換機能を使い倒すチャット生成器**」という感じになります ―― で、多分、この理解で合っていると思います（後述）。

従来のAIは、「翻訳する」「変換する」「学習する」という機能を使って、「学習した結果、あるいはその学習に沿った結果を吐き出す」という機能がメインでした。これに対して、最近よく耳にするようになった「Generative AI（生成型AI）」というのは、異なるAIの出力結果を混ぜ合わせて、あるいは異なるAI技術の機能を組み合わせて、「未知の新しいものを作り出すAI」という意味で使われています。

で、このChatGPTも、Generative AI（生成型AI）の一つです。ChatGPTを構成するAI技術は以下の3つになります。



このコラムでは、上記3つの技術についての説明はバッサリ省略させていただきます。上記の参考記事の方をご一読いただけますよう、よろしくお願いいたします。私のコラムは、**私が私の疑問を解消することが目的**でして、徹頭徹尾「**江端ファースト**」です。

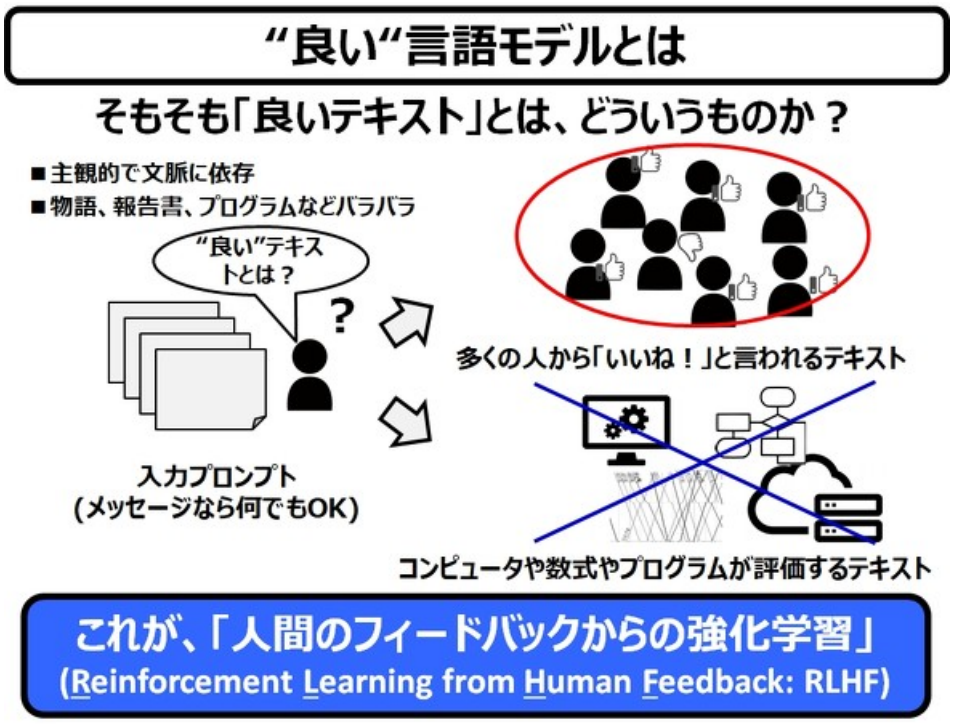
「Over the AI」過去記事

- Over the AI 第20回 忖度する人工知能 ～権力にすり寄る計算高い“政治家”
- Over the AI 第14回 開き直る人工知能 ～「完璧さ」を捨てた故に進歩した稀有な技術
- Over the AI 第16回 モノマネする人工知能 ～ 自動翻訳を支える影の立役者
- Over the AI 第21回 不幸な人工知能 ～尊敬と軽蔑の狭間で揺れるニューラルネットワーク
- Over the AI 第22回 官能の人工知能 ～深層学習を最も分かりやすく説明するパラダイム

さて、先程申し上げた通り、私の疑問は「**なぜ、ChatGPTは、私（たち）に違和感を与えることなく、人間が語る/書くような形式の出力ができるのか**」、つまり、「**なぜ、良いテキストを生成できるのか**」です。この疑問に、技術面からアプローチしてみたいと思います。

そもそも「良いテキスト」とは、どういうものか ―― それは、かなりハッキリしています。「**多くの人から『いいね！』と言われるテキスト**」です。逆に言えば、品のないツイッターのメッセージ、人を馬鹿にする掲示板の表示などは「悪いメッセージ」と言えます。

そして、このような「良いテキスト」を、ロジカルに説明することは難しく、**コンピュータや数式やプログラムには、「良いテキスト」を作り出すことはできない**、ということです。ChatGPTの基本形である、GPT-3は、この割り切りからスタートしています（GPT-3については後述します）。



大ざっぱに言えば、これが、人間のフィードバックからの強化学習、Reinforcement Learning from Human Feedback: **RLHF**です（厳密にはちょっと違いますが、これも後述します）。

さて、このRLHFによって作られるChatGPTのコアエンジン（“本体”という理解でいいです）は、ざっくり3つのプロセスの学習（訓練）によってされます。

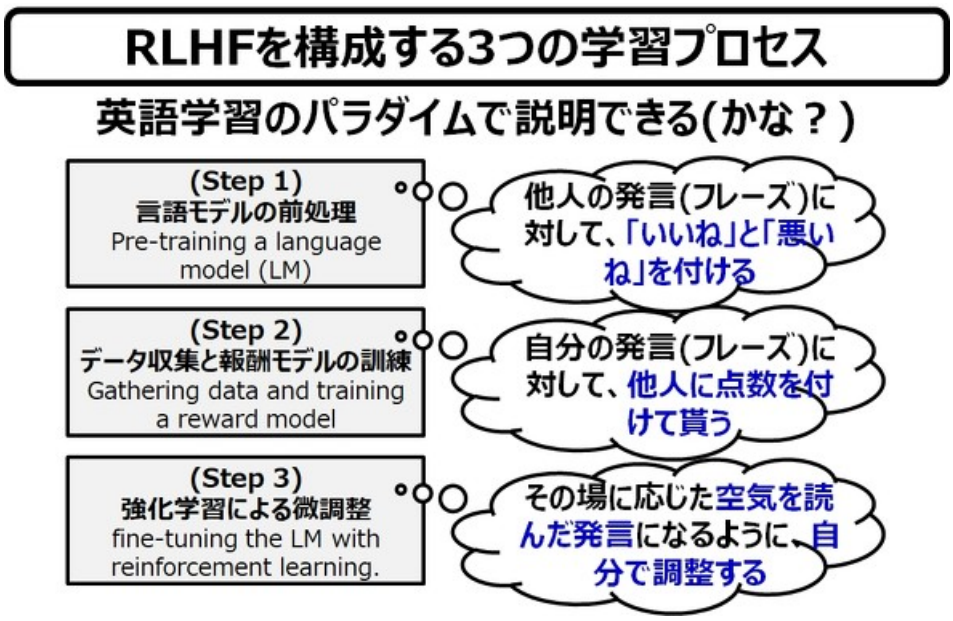
ChatGPTは「何も考えていない」

今回は、これを英語学習のパラダイムで説明してみましょう。

ここでは、

- ChatGPT君を育てる教師である「あなた」
- あなたの弟子の教育実習生の「リ・ワード（Reward）さん」
- まだ英語を全く知らない小学1年生の「ChatGPT君」

の3人を登場させます。



（Step 1）言語モデルの前処理

ここでは、小学1年生のChatGPT君が、英語で書かれた膨大な他人のメッセージをひたすら読んだり、聞いたりします。ChatGPT君は、もちろん、**その英語のメッセージの意味が分かりません**が、教師であるあなたは、その一つ一つのメッセージに、「いいね」「悪いね」と評価を付けて、ChatGPT君に強制的に教えこみます。

ちなみに、**ChatGPT君は、英語のフレーズの意味を、全く理解しておりません（というか、ChatGPT君は、自分が英語を学習しているという自覚すらありません）**が、とにかく、先生であるあなたの言うことを素直に聞いて、英語のフレーズと対応付けて「いいね」と「悪いね」を覚えます。

（Step 2）データ収集と報酬モデルの訓練

あなたは、ChatGPT君に対して、1つのフレーズを与えて、そこから4つの言い換えのフレーズを言うように命じます。ChatGPT君は、（Step 1）で覚えたフレーズと、あなたに教えてもらった「いいね」と「悪いね」を思い出して、4つのフレーズを捻り出します。なお、**ChatGPT君は、この段階においても自分が、英語のフレーズをしゃべっているという自覚がありません。**あなたの言われた通りのフレーズを書き出す、または、音声信号を発しているだけです。

あなたは、この4つのフレーズに対して、あなたが良いと思うものから順に、1番、2番、3番、4番と順番を付けていきます。そして、ChatGPT君に対して、「なるべく1番のフレーズを使うよう」に言い聞かせます。ChatGPT君は、素直にあなたの言うことを聞いて、できるだけあなたの言う通りの英語をしゃべるようになります。そして、しつこいですが、**ChatGPT君は、この段階においても自分が、英語のフレーズをしゃべっているという自覚がありません。**

（Step 3）強化学習による微調整

こうして、ChatGPT君を訓練していたあなたは、そのうち、**その作業が面倒くさくなってきました。**そして、あなたの近くにいた、**教育実習生のリ・ワードさん**に、「私がやってきたこと、見てきたよね。後は、あなたが私の代わりにやってちょうだい」と言って、ChatGPT君の教育を、リ・ワードさんに丸投げして帰宅してしまいます。

リ・ワードさんは、あなたがChatGPT君に何を施していたのかを全く理解していませんが、あなたとChatGPT君のやりとりを見続けていたので、あなたのマネをすることができました。ですので、リ・ワードさんは、自分が「リ・ワードである」とは言わず、あなたのフリをして、ChatGPT君の訓練を延々と続けました。

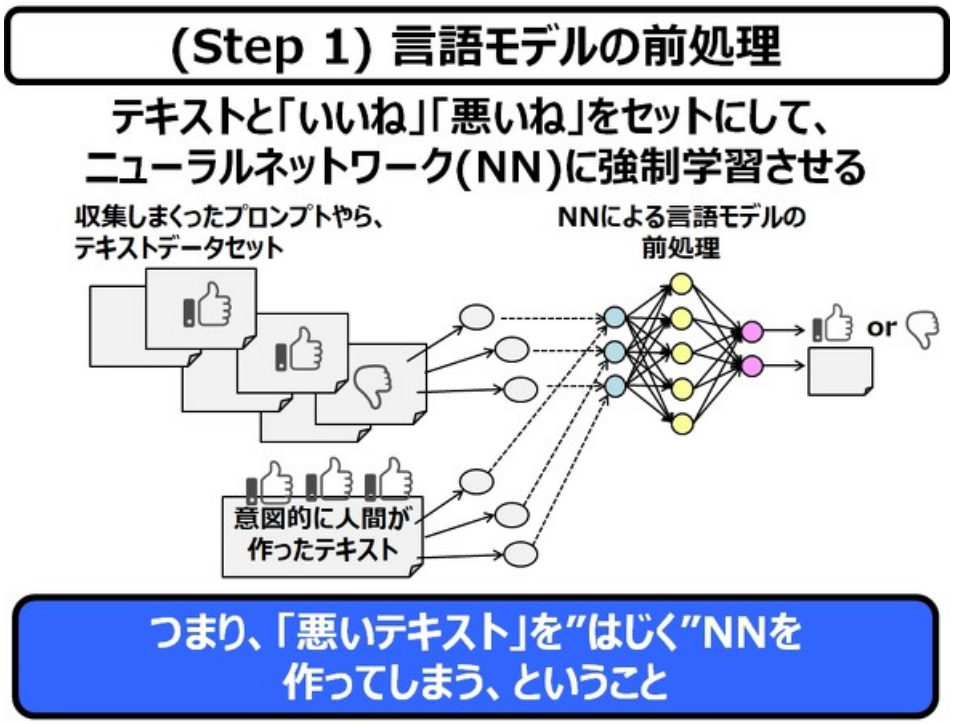
こうして、ChatGPT君は、世界中の人のメッセージに対して、そのメッセージに適したメッセージを返事することができるようなり、無事に世界デビューを果たすことができるようになりました。しかし、**ChatGPT君は、この段階においても自分が何をしているのか、全然分かっていません。**

ChatGPT君は、今もなお、何も考えていません。あなたに言われた「いいね」「悪いね」と「4段階のランク」を、ただひたすら、忠実に守っているだけです。

3人の話を技術的な話に落とし込んでみる

さて、ここからは、上記の教師である「あなた」と、あなたの影武者として働き続ける「リ・ワードさん」と、あなたの教育対象である「ChatGPT君」の話を、技術的な話に落とし込んで語ってみましょう。

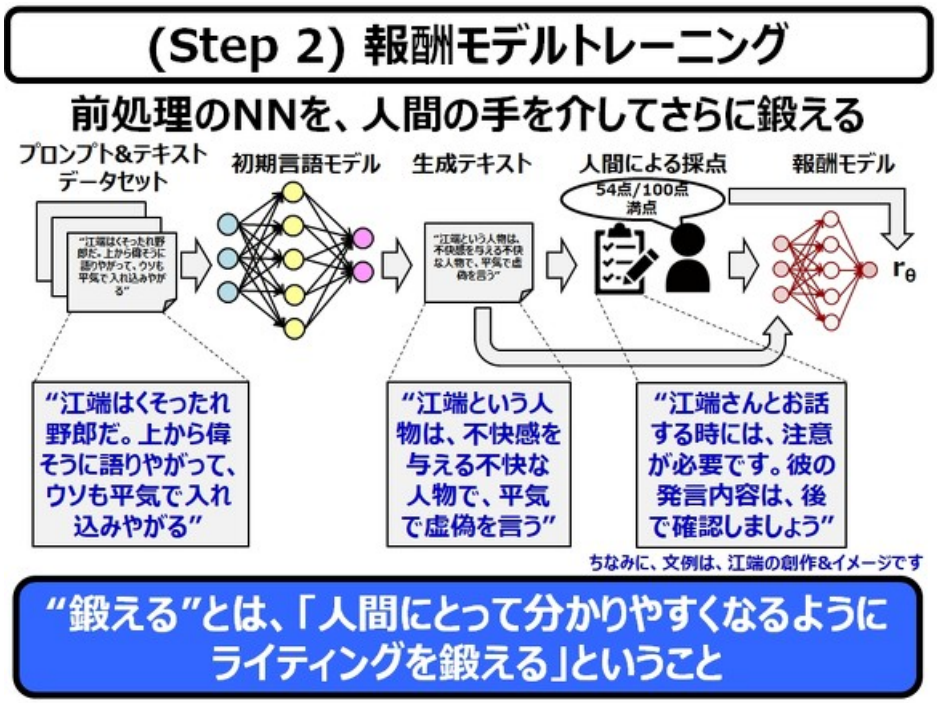
ChatGPTの本体は、ニューラルネットワークです。ニューラルネットワークは、深層学習によって、膨大な下図の非線形の因果関係を覚えることができます（私のコラムなどで、ご確認ください）。



とにかく、駄文、悪文、良文、関係なく、ニューラルネットワークに、その情報を叩き込み、単純に「いいね」と「悪いね」を出力する"だけ"の学習を施しまくります。

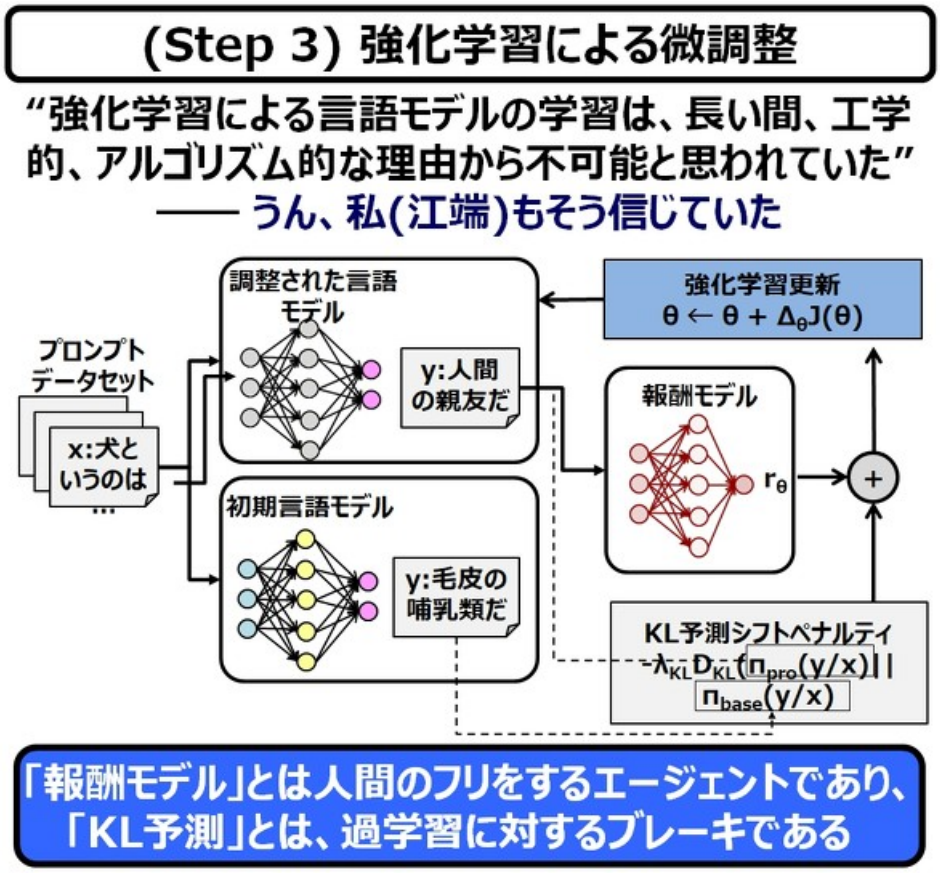
学習後、このニューラルネットワークに、実際のフレーズを入力すると、まずまずの文章を複数作り出します。そのフレーズは**人間様が丹念にチェックします**（想像を絶する大変さだと思います）。

そして、それと同時に、その人間様と同じように振る舞う別のニューラルネットワーク「報酬モデル（Reward Model）」も作っておきます。これが、**人間抜きの教育**を行う準備となります。



さて、ここから最終段階に突入します。ここから、人間のフリをする「報酬モデル」が介入しながら、ChatGPTのニューラルネットワークの強化学習を継続します。

ただ、この学習を「報酬モデル」に任せ続けると、ChatGPTは、最初の自分の状態を完全に忘れてしまいます。これを防止するために、KL（カルバック・ライブラー）予測というメカニズムを使って、初心を忘れないように、学習のし過ぎ（過学習）にブレーキをかけます。



ちょっと（かなり）混乱していると思いますので、ここで、最初の英語学習のパラダイムと併せて、いったん整理しましょう。

・初期言語モデル

→ 膨大な文章を覚えたばかりの初期状態の「初期言語モデル」であり、初々しいピカピカの**ChatGPT君**である

・調整された言語モデル

→ 人間（あなた）と報酬モデル（リ・ワードさん）によって訓練され尽くした、プロフェッショナルな**ChatGPT君**である

・報酬モデル（Reward Model）

→ 人間（あなた）がChatGPT君の教育の様子を見て、それをマネてあなたの代わりをする、教育実習生の「**リ・ワードさん**」である

・あなた

→初期段階のChatGPTに、「いいね」「悪いね」「4段階の採点」を付ける、生身の（重労働の）**人間**である

こうして、ChatGPT君は、「あなた」と「リ・ワード」さんに鍛えられて、一人前のChatGPTとしてデビューを果たすわけです。

なお、しつこいほど繰り返しますが、**ChatGPT君は、『何も考えていません』。「あなた」と「リ・ワード」さんに言われたことを、淡々とやっているに過ぎません。**

もう少し「技術者」っぽく解説してみる

とまあ、上記の説明で**「なぜ、ChatGPTは、私（たち）に違和感を与えることなく、人間が語る／書くような形式の出力ができるのか ーー 良いテキストを生成できるのか」**の答えにはなっていると思います。

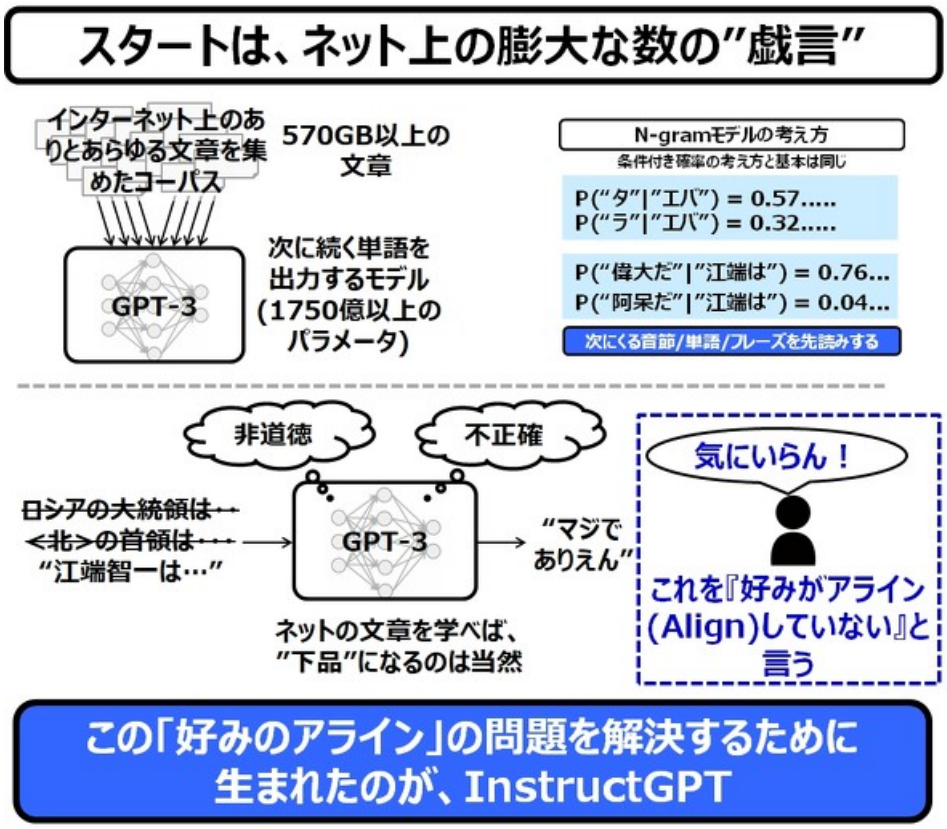
しかし、このままのメタ表現（リ・ワードさんやら、ChatGPT君やら）+概念説明だけでは、日本中のAI研究者から「さっぱり分からん！」と叱責されるような気がします。

そこで今度は、論文の用語などをできるだけ正しく引用して説明したいと思います（興味のない人は、スキップしていただいても結構です）。

上記で説明した初期状態のChatGPTの基本モデルが**”GPT-3”**と呼ばれる、570GB以上の文章コーパスを用いて作られた、1750億以上のパラメータを有するニューラルネットワークです。これは、N-Gramモデルのように、次に続く単語を予測するリファレンスモデルとなっています。

ところが、このGPT-3は、何しろ覚えるだけ覚えるものなので、非常に**「育ちが悪い」**です。ネットやら掲示板の文章を学べば、下品なフレーズを使うようになるのは当然です。そして、このような下品なフレーズが出てくることを、**「人間の好みとアライン（Align）していない」**と言います。

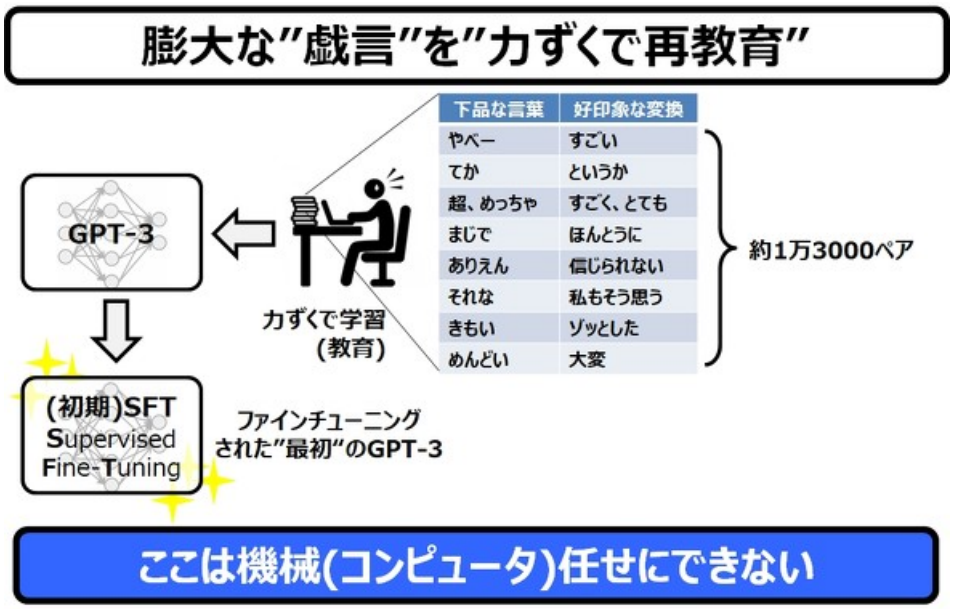
そして、**この「好みのアライン」の問題を解決するために生まれたのが、InstructGPTです** 。



さて、このGPT-3に最初の学習を施すのは、生身の人間です。

基本的には下品な言葉を好印象な言葉に変換することが目的で、この変換パターンは、ざっと1万3000ペアあります。これを、人手を使って丹念に学習させていきます。この作業をファインチューニング（微調整）と言います。

こうして、ファインチューニングされた最初のGPT-3を、初期SFT（Supervised Fine Tuning：”監修された微調整”）と呼びます。当然ですが、この段階では、自動化は登場しません。



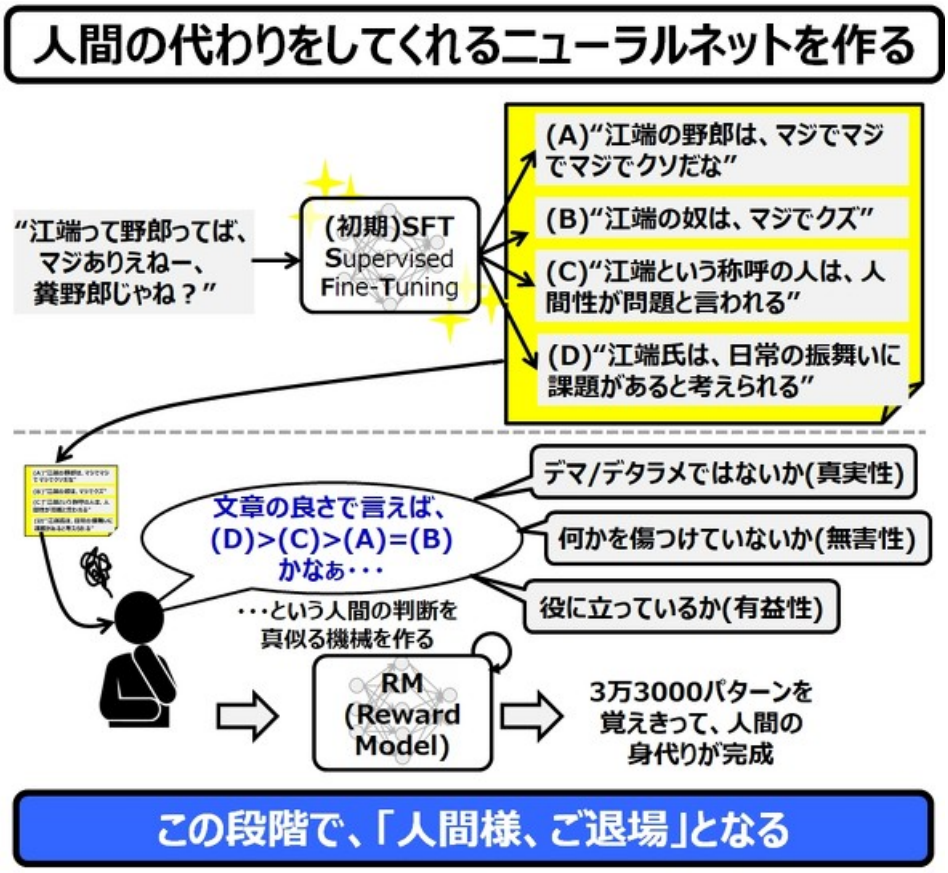
さて、ここから出来上がったばかりの初期SFTに、実際にフレーズを作らせてみます。1つの入力フレーズに対して、4つ（最大7つ）のフレーズを作らせます。そして、この4つのフレーズを、優れている（**と人間が感じる**）順番に点数を付けていきます。

この時の採点は、「**デマ／デタラメではないか（真実性）**」「**何か／誰かを傷つけていないか（無害性）**」「**役に立っているか（有用性）**」の3つが重要な評価指針となっています。

そして、人間がこの作業をやっている様子を、別のニューラルネットワークであるRM（報酬モデル:Reward Model）が学習をします。

ChatGPTのコアとなるニューラルネットワークと、人間の学習を代行するニューラルネットワークの2つが登場しているので、混乱しないようにしてください。

RMが人間の振る舞いを学んでしまえば、ここからは人間は不要となります。ニューラルネットワーク（RM）が、ニュラルネットワーク（初期SFT）を鍛えるという仕組みが完成することになるからです。

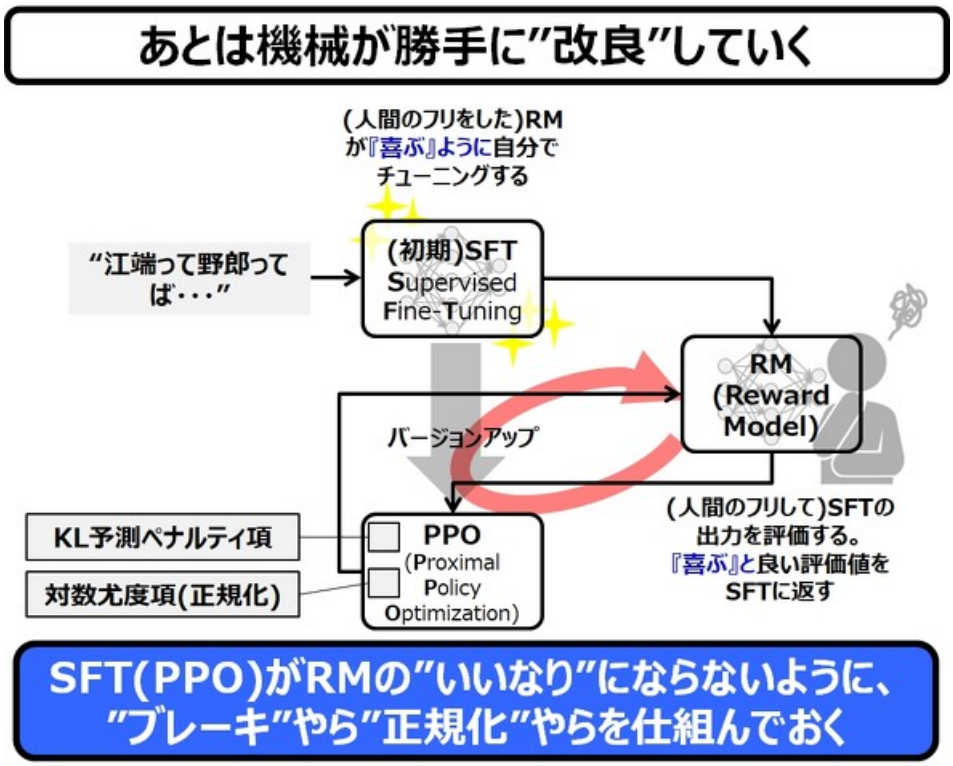


さて、こうして、人間のフリをするRMと、そのRMに評価されて強化学習を続けるSFTによる、自動学習が続くことになります。そして、訓練され続けるSFTは、最後にPPO（Proximal Policy Optimization:最適化された近接ポリシー）というものに、バージョンアップします。

ただ、**SFTはRMに褒められると調子によって学習を加速させていきます**ので、このままでは、RMの”言いなり”になってしまいます。これを「過学習（過剰な学習）」というのですが、これって結構マズいのです。というのは、PPOは、全ての入力に対してバランス良く学習を行わなければならないからです。

例えるのであれば、**数学しか教えていなかったPPOに、いきなり古典の問題を尋ねるようなものです。**PPOがパニックになるのは当然です。ここに学習の難しさがあります。数学と古典を矛盾なく、融合させながら教えていくことは、教育現場だけでなく、ニューラルネットワークの学習でもやはり難しいのです。

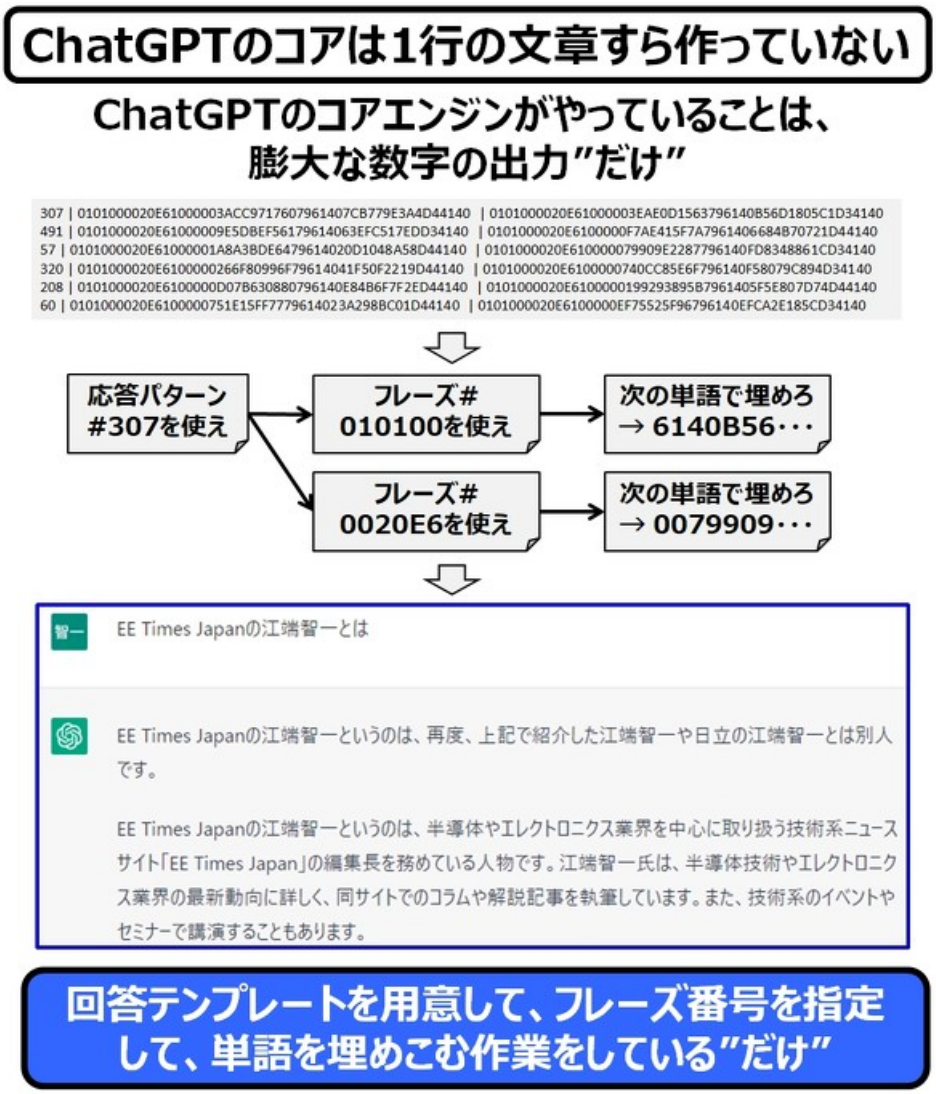
ですから、PPOには、**SFT（PPO）がRMの”いいなり”にならないように、”ブレーキ（KL予測ペナルティ項）”やら”正規化（対数尤度項）”やらを仕組んでおく必要があるのです。**



ところで、ニューラルネットワークとは、基本的に数値関数です。ですから、文章を理解する能力なんて、本当に1mmもありません。**"強化学習による言語モデルの学習は、長い間、工学的、アルゴリズム的な理由から不可能と思われていた"**というのは事実であり、私も、今の今までそう信じていました。

しかし、ChatGPT（正確には、その前のInstcutGPT、GPT-3）は、「辞書のページ（数字）、単語のある行数（数字）、パターン化された例文のリストの番号（数字）にし、それを組み合わせる」というアプローチで、この問題を解決したのです。

そのイメージを図示してみました。



ChatGPTの実体はニューラルネットワークであり、そのニューラルネットワークの出力する数値に応じて、サンプル例文と、単語を選び出しているだけです。私が、**ChatGPTには知性がない**、と繰り返しているのは、このメカニズムがその理由です。

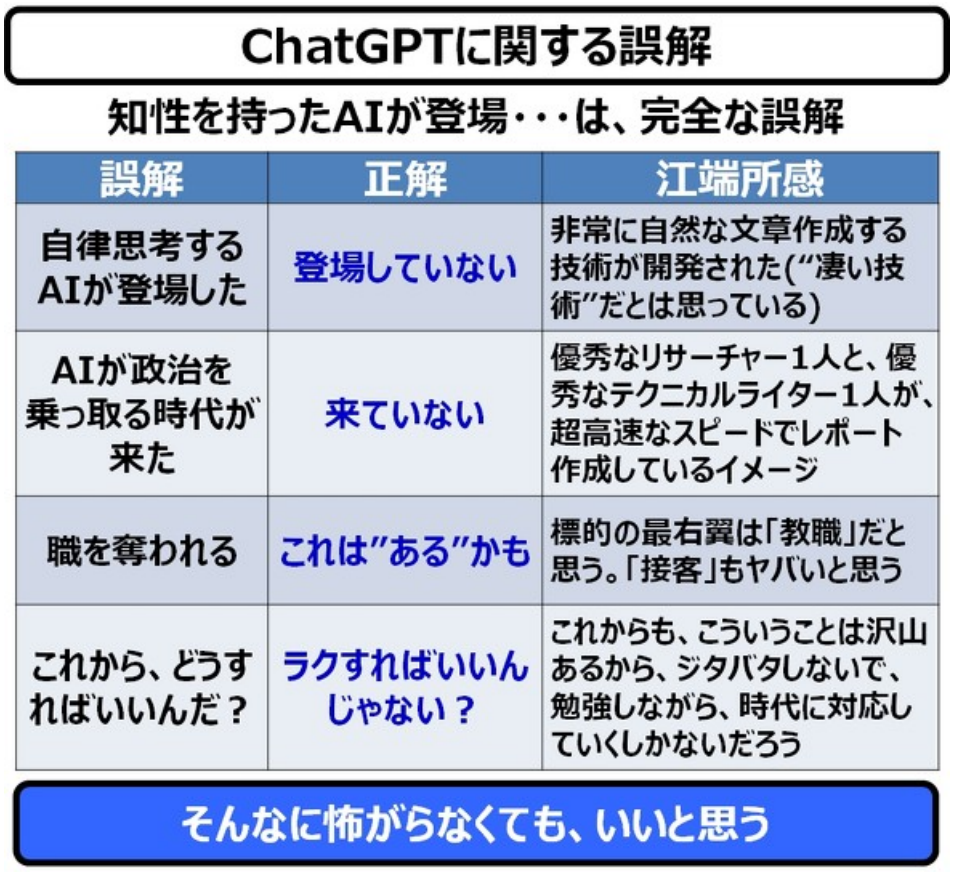
もちろん、だからといって**ChatGPTがものすごいAI技術であることには1mmの疑義もありません**。私がこのアイデアを聞かされたとしたら、間違いなく「そりゃ、理屈として分かるけどさぁ、そんなの実現するのは不可能だよ」と答えたと思います。しかし ―― **それを可能と信じた人がいて、そして、その人たちは決して諦めなかったのです** 。

（ちなみに、上図のChatGPTの回答は、別の日に「「江端智一について教えて」とChatGPTに質問したものです。**誤答率100%です。**）

ChatGPTは誤解されている？

では、最後に、私のChatGPTに関する所感を述べさせていただきます。

まずは、世間のChatGPTに関する誤解ですが、今回のChatGPTのメカニズムの調査（Reinforcement Learning from Human Feedback: RLHFと、ファインチューニング）によって、以下のことがはっきりしました。



（ある程度予想はしていましたが）ChatGPTもまた、これまでの「弱いAI」と同様に、**自律思考など1mmもやっていませんでした。**ですので、「AIが政治を乗っとる時代」は、残念ながらまだまだ遠いと言わざるを得ません（もちろん、驚異的な発想に基づく、AI技術の活用と連携には、腰を抜かす程度には驚きましたが）。

そういえば、先日、某新聞のトップに『**ChatGPTに国会で政府の答弁をさせることは、憲法の違反になる恐れがある**』とかいう識者の意見が記載されていて、私は首をかしげていました。

―― どの条文に抵触するの？

私、今、日本国憲法の条文を開いて読んでいるのですが、”当たり”を付けることすらできませんでした。そもそも、国会でのChatGPTが憲法違反になるなら、Googleサーチが駄目でしょうし、インターネットの利用すら憲法違反になります。

ChatGPTがすごいのは、膨大な文章コーパスがあり、あたかも人間が作っていように見せるインタフェースの精錬さと、その説明内容の洗濯のラインアップのバランスの良さと、そして、それらをまとめて驚異的なスピードで表示する能力です。

ChatGPTは、玉虫色の優等生的答弁をするように作られています。前述の、”真実性”、”無害性”、”有用性”の方針で学習されているからです。しかし、国会答弁は、誰を／どこを切り捨てて、どこに金と時間をかけていくか、という、生臭いドロドロした話です。例えば、少子化対策は、詰まるところ、高齢者福祉の切り捨てとトレードオフです。

もちろんChatGPTは、Generative AI（生成型AI）なので、そういう生臭いドロドロの答弁をさせることも可能でしょうが、そんな答弁したら、ChatAPIではなく、内閣の方が先に倒れます。（個人的には、故吉田首相の『バカヤロー解散』に匹敵する、『ChatGPT解散』を見てみたい気はしますが）。

『職が奪われる』のは考えられる未来の一つだと思います、それでも、パソコン導入時のインパクトに比べれば、はるかに小さい規模であろうと思います。パソコンの時と同じように、ChatGPTが入ってきたら、いきなり失職する、なんてことはありません。技術革新による社会変化は10年のスパンが必要です（反対勢力が、改革を押し止めるブレーキになるから）から、その間になんとかすれば良いでしょう。

では、これからどうすればいいのか。

―― ラクすればいいんじゃないですか？

ChatGPTは、多くの人が平等に使えるので、差別化技術にはなりません（意固地になってChatGPTを使わない人のことなんか、私は知りません）。とすれば何が残るか？**『便利な道具』の一択**でしょう。

ChatGPTへの「世間一般の反応」に対する江端の所感

次に、ChatGPTの世間の反応に対する、江端の見解を述べさせていただきます。

ChatGPTに関する江端の見解

ChatGPTには"知性"はないが、恐しく便利な道具

世間の心配	江端見解	江端所感
考える力が低下する	考える力が向上する	ChatGPTを「踏み台」にできる
コピペのレポートが蔓延(まんえん)する	コピペって、そんなに悪いことか？	コピペするだけでも、十分に勉強になると思う
	"オリジナリティ"がクリアになる	コピペとオリジナルを簡単に区別できる
	口頭試験を実施する	最終的には、口頭試験で確認すれば足る
国会の答弁に使われる	結構なことじゃないのか？	ChatGPTに答えられるような安直な質問/回答をする議員や政府の無能がクリアになる

"使い倒す"方向で前向きに検討すべき

ChatGPTによって、考える力が低下する、と言われますが、私は「そうかな？」と疑っています。

むしろ逆に「考える力が向上する」と思います。どんな仕事でも勉強でも、最初の一歩のところが難しい、ということは多いですが、ChatGPTは、この最初の一歩の障壁を下げてくれる、思考の「踏み台」となるからです。

例えば、私が論文を書いているとき、「XXXについての課題を出せ」とChatGPTに言うと、私の想定外の課題が山ほど出てきて、検討項目が増えました。結果として、ものすごく苦労させられましたが、学会発表で想定される質問の予想ができて、逆に安心することができました。

あとは歴史的な経験則ですね。**「パソコンによって考える力が低下する」「検索エンジンによって考える力が低下する」「スマホによって考える力が低下する」**―― 全て、これまで散々言われてきたことですが、この中で、実際に「考える力が低下した」という事例が一つでもあったでしょうか？　もしそのような事例があれば、客観的なデータを添付して私に送付してください。ちゃんと検討してお返事をします。

ChatGPTのコピペが蔓延する？ ――　そんなに悪いことですかね。

私は、友人のレポートを書き写すくらいなら、ChatGPTの内容を書き写す方が勉強になる、と思っているくらいです。それにChatGPTは、汎用アプリケーションですから、丸写しは簡単に発見されてしまいます（即、レポート却下にできます）ので、これからはそういう行為は意味を成さなくなっていくと思います。

むしろ、**学生にとっては受難の時代の突入と言えます**。ChatGPTを越える発想をしなければ、評価されない時代です。私は、ChatGPTを最初から無視して、一点突破の狂ったロジックを展開できる学生の方に、高い評価が与えられるという逆

転現象すら予想できます。

まあ、そんなにコピペが心配なら、口頭試問すれば済みます。ChatGPTの解答を見せた上で、試験官は「さて、この内容**以外**の観点から、自由に論じてください」と尋ねれば足ります。学生の方もその準備ができますので、Win-Winになるでしょう。

国会の答弁に使われる？ ―― 大歓迎です。

大臣の答弁の質が上がることが期待できます。

野党議員:『大臣。そんなChatGPTで言える程度のこと、私は聞いていませんよ』。

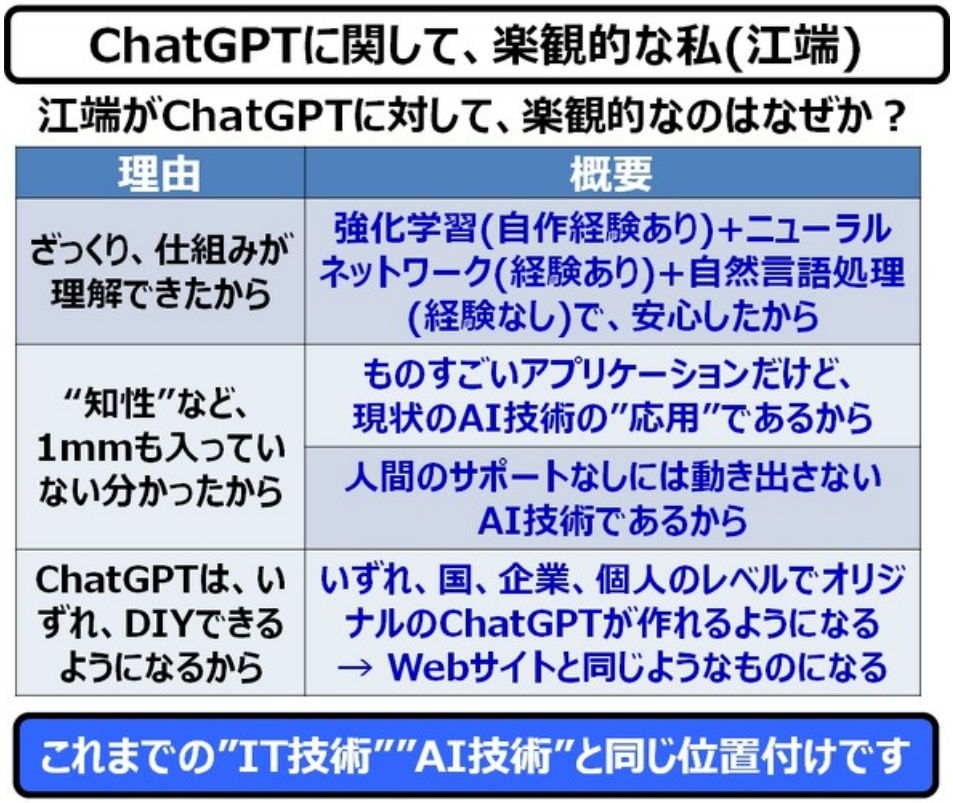
しかし、逆に野党も質問の内容を熟考しなければなりません。

大臣:『その程度の質問なら、ChatGPTに聞いてもらえませんでしょうか。いちいち、私に聞かなくてもChatGPTが答えてくれますよ』

うん、こうなれば、国会中継が、随分楽しいものになりそうです。ワクワクします。

江端がChatGPTに対して「楽観的」な理由

さて、ここまで読んでいただいたことからも明らかなように、私（江端）は、ChatGPTの使用に対して、好意的で、恐ろしく楽観的です。その理由を、以下の表にまとめてみました。



どんなモノであれ、コトであれ、分からないものは、怖いものです。

私も、ChatGPTを初めて使った時の「ゾクっ」としたあの感覚をよく覚えています。冒頭に述べた通り、「気持ち悪！」でした。ですが、今回のコラム執筆のために行った調査によって、その内容が分かってきて、今や、すっかり安心しております。

これは、『膨大な時間と金さえかければ、ChatGPTと似たようなものを自分でも作れる』という確信（妄想）があり、その裏側には『ChatGPTは、まだ、私が理解可能な範囲の技術で作られている』、つまり、**今の私は、まだ、ChatGPTについて、その利点も弱点も論じることができる段階にいる**、ということです。

本当に良い時期に、このテーマのコラムをいただいたと、EE Times JapanのMさんには感謝の言葉もありません。そして、私は、その立ち位置（AI技術コラムの執筆担当者）のおかげで、この安心感を獲得できるという幸運にめぐまれました。

しかし、多くの人にとって、強化学習、自然言語処理、ニューラルネットワークなどは、チンプンカンプンの内容だと思いますし、そういう人たちの恐怖というのは ―― 例えるのであれば、コロナワクチンへの恐怖と似たようなものがあると思います（関連記事：「「それでもコロナワクチンは怖い」という方と一緒に考えたい、11の臨床課題」）。

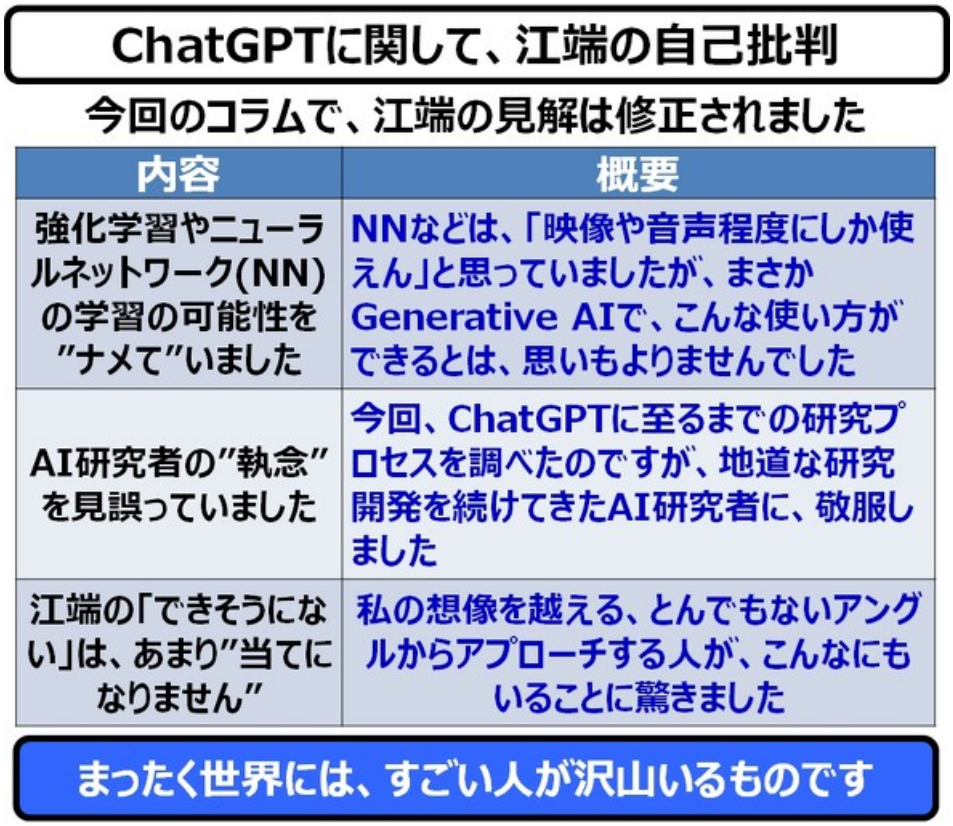
特に、ChatGPTは、AI技術史上、もっともユーザーフレンドリーなインタフェースですから（質問フレーズを放り込むだけ）、その恐怖もダイレクトに感じてしまうのではないか、と思います。

これをなんとかする方法があるのかどうか考えましたが ―― 正直なところ、**私さえ安心できれば、他の人のことはどうでもいいです**（今さら、偽善者ぶっても仕方ない）。

それと、私は、いずれChatGPTは、個人単位で運用されるようになる時代がくる、と踏んでいます。そうなると、多分、私たちの中に潜む、AI恐怖症（AIフォビア）も、程なく解消されると考えています。

まったく、世の中にはすごい人がたくさんいるものです

とはいえ、かつてのAI研究に従事した研究員としては、やはり自己批判しなければならないことも多いです。



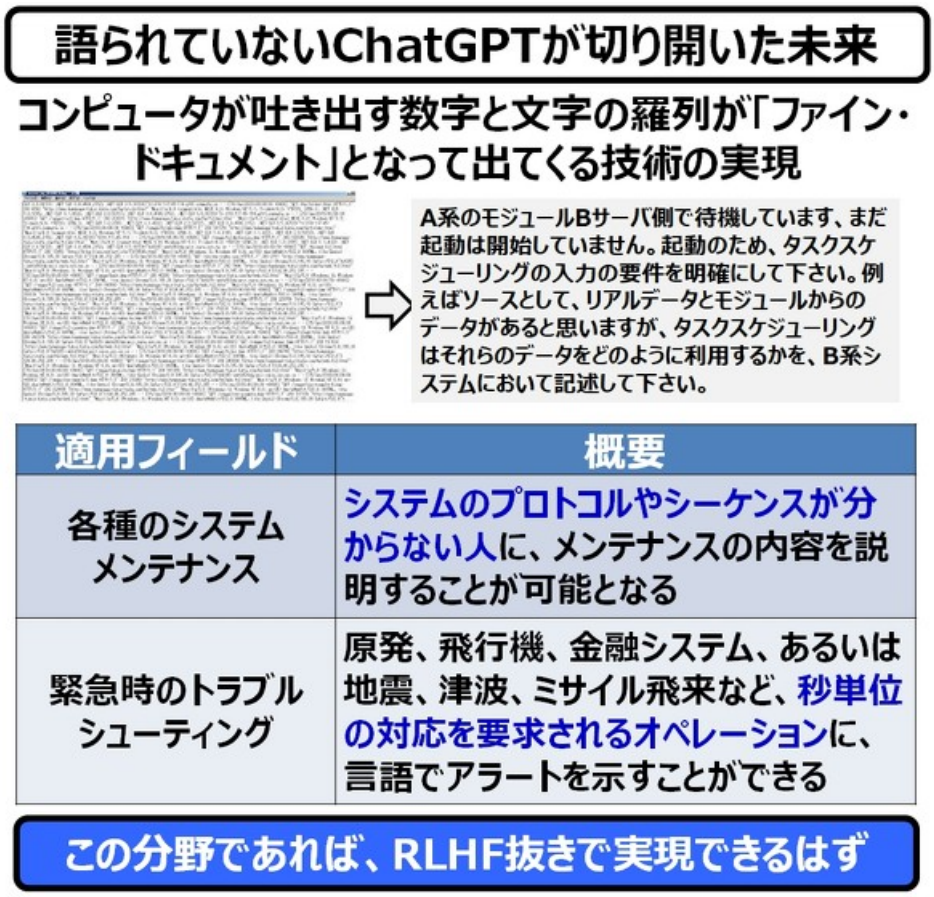
これらについて、一つ一つを語る気力はありません。もう私は、**敗北一色**です。

まったく、世の中には、ぶっとんだ発想を、本当に実現してしまうすごい人がいるものです。私も研究員として生きている以上、**人生で一度くらいは「すごい人」と言われたいものです**。

まあ、「すごい人」とは言われているんですけどね ―― ただ、「すごい」の意味が全然違うようですが（関連記事：「一実践編（目次）――英語に愛されない私たちの行動原理「目的は手段を正当化する」」）。

ChatGPTで「ラク」をすればいい

さて、今後のChatGPTについて私の期待を語って、本コラムを終えたいと思います。



ChatGPTは、Generative AI（生成型AI）として、優れたアプリケーションではあるのですが、私は、Generative AIそのものよりも、その一部である「**自然言語インタフェース**」と「**記載のバランシング**」機能に着目しています。

このような人間に対する優れた言語ベースのユーザーインタフェースは、さまざまな工業分野への展開が期待できると思います。上の表に記載したような、各種のメンテナンスや、秒単位での対応を要求される緊急オペレーションなどに、これらのインタフェース技術が展開されることを期待しています。

スリーマイルアイランド原発事故においては、設計時には想定されていなかった運転員のエラーにより事故が発生し、中央制御室は、100を越える警報が一斉に点灯して、何が起きたのか判断できなくなる状態になったそうです。これを、「クリスマスツリー（警報雪崩）」と呼ぶのだそうです。

比して、福島原発事故においては全電源喪失となり、いわゆる「逆クリスマスツリー（警報無言）」（命名 江端）となった訳です。

今まで、ピカピカ点滅していた装置、監視用モニター、そして原発の健全性を証明する計器が、一瞬で全部沈黙。叩こうが蹴ろうがピクとも動かない。そして、多重に準備されたバックアップ電源が、津波による水没で一気に（多分1分内くらいで）全滅。

このような、**システム運用者にとっては考えたくもない状態に対応できるようなAI技術の応用**を、私は期待しています。

□

それでは、今回のコラムの内容をまとめてみたいと思います。

【1】EE Times Japanの編集担当のMさんから「ChatGPTについて執筆してください」と依頼を受けて、ChatGPTについて調査を開始しました。

【2】ChatGPTについて簡単な使い方を説明し、その内容について（1）「『江端智一について教えて』という質問」と（2）「江端が執筆してきた論文の概要」についてChatGPTにから回答を得ました。この（1）（2）に限れば、**ChatGPTは、50〜**

100%デタラメな回答をした、という実験結果を報告しました。

【3】Google Trendを使って、ChatGPTの知名度の変化を調べたところ、**2022年11月下旬**から、世界、日本ともに爆発して上昇していること、また、Google Scholarを使った”ChatGPT”を含む論文数を調べた結果も、同様の傾向が観測されることが分かりました。

【4】ChatGPTの作り方について、(A)「私」「リ・ワードさん」「ChatGPT君」という3人を登場させた**擬人化での説明**、(B)一般の方向けの**ざっくりした説明**、(C)AI研究者、エンジニア向けに、**論文で用いられている用語を使った説明**の、3つのアプローチでの解説をしました。

【5】上記の検討の結果を踏まえて、(A)『ChatGPTが知性をもったAI』であるという**世間の誤解**、(B)『ChatGPTは恐ろしく便利な道具であれ、使い倒せば良い』という**江端の見解**、(C)『ChatGPTを恐れる理由はない』という**江端の楽観的スタンス**、(D)ChatGPTのAI研究者の執念を垣間みた江端の、**自分の能力や根性の欠落に対する落胆**および、(E)ChatGPTのインタフェースの**江端の他の分野の適用への期待**について、私見を述べました。

以上です。

―― と、本コラムのリリース後に、**ChatGPTがこういう風にまとめてくれることを、私は期待しています**。

ChatGPTで「ラクをすればいい」

今回のコラムの趣旨は、**ChatGPTの登場で、私たちはどうすれば良いのか**、に関する検討、考察にありました。

私の結論は明確です。**さらにラクをすればいい ―― この一択です。**

私が、ワープロを入手した時のように、ChatGPTが、多くの人にとって、人生のシンギュラリティポイントになるなら、こんなにステキなことはない、と思っています。

このように、総じて、今回の技術的な調査を経て、私は、ChatGPTに関して、**脳内お花畑のような楽観主義**で洗脳されています。

ですので、イーロン・マスク氏が、ChatGPTの最新版言語モデル「GPT-4」を上回るシステムの開発を6カ月間停止するよう求める、AIの専門家や業界幹部らが共同制作した公開書簡に署名したことや、イタリア政府によるChatGPTの使用の一時制限について ―― 私は、そのリスクを理解はしているのですが(セキュリティやプライバシーの問題) ―― それは、オンラインのメンテナンスで十分対応できる範疇の内容に思えるのです。

正直に言って、**停止させたり、使用制限させたりするほどのことかな?**と思っています。私は今、「Googleストリートビュー」の時の、あの騒ぎを思い出しています(筆者のブログ)。ちなみに、「グーグルマップ ストリートビュー」に対して、差し止め裁判を起こした、日本国内の町内会、自治会は、全て裁判を取り下げています。

私なら ―― 単純な技術的手法になりますが ―― GPT-3に使っているコーパスに対する人間による評価値をもっと厳しくして作り直すとか、Reward Modelの出力値をコントロールするとか、KLの抑制パラメータ値を調整するとか、そんなことを要求するかもしれません。やり方はいろいろあると思うのです。

今回のコラムの調査で、私は「分かった気になっているだけ」かもしれません。何か、とてつもない重大なリスクを見落しているのかもしれません。それに気がつかれている方は、当方にご一報いただければ幸いと存じます(本件に関するメールアドレスはover_the_ai@kobore.netです)。

「ChatGPT」の回答、日本人にだけは届かないかもしれません

後輩:「江端さんも言っていますが、40代以上の大人は、『**ITというものが、基本的にウソつきであり、使えないものである**』ということを、よく知っていますよね」

江端:「まあ、極端なパソコン嫌いとか、ITから逃げ回っている人間でもなければ、誰でも知っていることだろう」

後輩:「今や翻訳エンジンの精度は劇的に向上していますが、基本的に私たちは、翻訳エンジンの翻訳結果を信じてきませんでしたよね」

江端:「うん、初期の英語の日本語翻訳のあの"ひどさ"を見れば、日本語の英語翻訳の結果なんて信じられる訳がなかった」

後輩:「音声でメールを書ける、という触れ込みのソフトウェアが、全く使えなかったとか、エキスパートシステムとかいうAI(第2世代AI)が、組み合わせ爆発で翌日になっても答えを返してこないとか・・・」

江端:「まあ、ITというのは『期待と失望』がセットになっていたと思う」

後輩:「で、まあ、今回のコラムで、江端さんは、自分のコンテンツを使って、ChatGPTの実験結果を示して、**ChatGPTの応答の『50〜100%がデタラメ』**という結果を導いていますよね」

江端:「ケースによるとは思うけどね」

後輩:「でも、別段、ガッカリもしていないでしょう?」

江端:「全くガッカリしていない。そんなことより、『対人間インタフェース』のすごさの方を絶賛している」

後輩:「江端さんは、ChatGPTの『対人間インタフェース』の、どの部分がすごいと思っていますか?」

江端:「**あいつ(ChatGPT)、諦めないんだよ。**絶対に、『知らない』と言って途中で投げ出さない。仮に知らなくても、『同じ内容を、別の言葉で言い換え』してでも、ちゃんと"回答している風"に見せかける。あのスタイルは、全人類が見習わなければならないと思う」

後輩:「ChatGPTに関しては、個人情報保護やプライバシーの問題がリスクとして挙げられているようですが、私はもっと本質的な別の問題があるんじゃないか、と思っています」

江端:「というと?」

後輩:「ChatGPTの言っている内容が、日本人にだけには届かないかもしれない、というリスクです」

江端:「……よく分からないんだか?」

後輩:「ChatGPTのレポートのまとめ方は、基本的にはロジカルに記載されていますよね」

江端:「まあ、ChatGPTの学習プロセスが、そのようなテクニカルライティングを想定しているからだ、と思うけど」

後輩:「私たちのような技術フィールドの人間にはこれで良いです ーー というか、これでないと困ります。しかし、日本人にとって日本語とは単なる情報を運ぶだけの道具ではありません」

江端:「?」

後輩:「つまりですね、表情や口調、声の大小、速度、抑揚、間の取りかた ーー **そういう、非言語化された部分に、大量の情報を搭載させるのが、日本人の使う日本語**です」

江端:「それは、まあ、分かる」

後輩:「よく外国の人が、『ちゃんとした日本語をしゃべっても、日本人は理解しない』と愚痴を言っているのですが、それは、**非言語化された日本語**を、ちゃんと学んでいないからなのですよ」

江端:「『非言語化された言語を学ぶ』というのは、そのフレーズ自体が論理破綻しているけどね」

後輩:「残念ながら、『非言語化された言語』を理解するのは、日本文化のバックグランドの理解が必要で、その理解をする

ためにも非言語化の手段が必要なのです」

江端:「それでは、日本語は、日本で生まれ育った人間に同士でしか理解できない、**暗号言語**ということになってしまうけど」

後輩:「ある意味、それは正しいと思います。日本語は、"言語"が難しいのではなく、"非言語"が難しいのです」

江端:「なるほど、言いたいことが分かってきたぞ。つまり将来の**ChatGPTの『対人間インタフェース』は、日本語の"非言語"部分にも対応できるだろうか?**と言いたいんだな」

後輩:「そうです。もちろん、今のレベルのChatGPTの言語であっても、単なる情報伝達、特に、科学技術分野においては、問題ないくらいに高度なレベルにあると思います。ただ……」

江端:「今回調べたような**ChatGPTのRLHF(人間のフィードバックからの強化学習)で、そのような"非言語"の言語レベルに至れるかどうかは、分からない**、と」

後輩:「そういうことです」

江端:「でもなぁ、それって『強いAI』の範疇(はんちゅう)の話だと思うぞ。現状の『弱いAI』で、そのような"非言語"の言語をサポートするのは無理なんじゃないか?」

後輩:「……江端さん。そういうこと言って、また検討する前に諦めるんですか。また『腰を抜かすほど驚いて』『悔しい思いをする』ことになりますよ」

江端:「……」

後輩:「江端さんは、**人生で一度くらいは『すごい人』と言われたいのでしょう**?」

Profile

江端智一(えばた ともいち)

日本の大手総合電機メーカーの主任研究員。1991年に入社。「サンマとサバ」を2種類のセンサーだけで判別するという電子レンジの食品自動判別アルゴリズムの発明を皮切りに、エンジン制御からネットワーク監視、無線ネットワーク、屋内GPS、鉄道システムまで幅広い分野の研究開発に携わる。

意外な視点から繰り出される特許発明には定評が高く、特許権に関して強いこだわりを持つ。特に熾烈(しれつ)を極めた海外特許庁との戦いにおいて、審査官を交代させるまで戦い抜いて特許査定を奪取した話は、今なお伝説として「本人」が語り継いでいる。共同研究のために赴任した米国での2年間の生活では、会話の1割の単語だけを拾って残りの9割を推測し、相手の言っている内容を理解しないで会話を強行するという希少な能力を獲得し、凱旋帰国。

私生活においては、辛辣(しんらつ)な切り口で語られるエッセイをWebサイト「こぼれネット」で発表し続け、カルト的なファンから圧倒的な支持を得ている。また週末には、LANを敷設するために自宅の庭に穴を掘り、侵入検知センサーを設置し、24時間体制のホームセキュリティシステムを構築することを趣味としている。このシステムは現在も拡張を続けており、その完成形態は「本人」も知らない。

本連載の内容は、個人の意見および見解であり、所属する組織を代表したものではありません。