デジタルハリウッド大学

2024 年度　一般選抜　B 方式

小論文　［80 分］

【 注 意 事 項 】

1. 試験監督の指示があるまでは、問題冊子は開かないこと。
2. 試験監督から指示があったら、解答用紙に氏名・受験番号を正確に記入すること。
3. 試験開始の合図後、この問題冊子を開き、8 ページ（白紙ページ含む）揃っているか確認すること。
4. 乱丁、落丁、印刷不鮮明などがある場合は、手を挙げて試験監督に知らせること。
5. 試験開始から終了までの間は、試験教室から退出できません。
6. 不正行為を行った場合は、その時点で受験の中止と退室を指示され、同日受験したすべての科目の成績が無効となる。
7. 解答用紙は試験終了後、回収される。問題冊子は持ち帰っても良い。

これは2ページ目です。
次のページから問題が始まります。

生成 AI（Generative AI）に関する次の**資料１～２**を読み、後の問に答えなさい。

〈資料１〉

車にクルミを割らせるカラス

近年の AI の躍進を支えているのは、主に「帰納的学習」といって、「得た情報からパターン（規則性）を見つける」という学習プロセスです。

帰納的な学習の例として、「車にクルミを割らせるカラス」という話があります。

このカラスはたまたま「クルミが道路に落ちて、その上を車が通ると殻が割れた」という経験から、道路にクルミを置くようになったと考えられています。「モノは十分に重い物体に踏まれると壊れる」といった因果関係は理解していません。ですので、もしホバークラフトのように地面から少し浮いて走る空飛ぶ自動車が発明されたとしても、タイヤの有無にかかわらず（その騒音がカラスを警戒させないかぎり）、カラスは同じように、当面は道路にクルミを置き続けるでしょう。

しかし、ぼくら人類は車からタイヤがなくなった時点で、すぐにクルミが割れないことがわかります。それは「モノは十分に重い物体に踏まれると壊れる」という因果関係を一般論として学習しているからです。

たとえ一度もクルミを自動車が割るという経験をしていなくても、話を聞けば因果関係がわかり、クルミの未来が想像できてしまう。これが「演繹的推論」と呼ばれる、「ある事象に対して、仮説や一般化されている因果関係を適用したらどうなるか考える」という学習プロセスです。

演繹的な学習プロセスは、人類がいかにして道具（＝テクノロジー）を造り、扱うようになったかを想像するとわかりやすいです。

原始的な「石の穂先を持った槍（石槍）」を例にしましょう。

石槍という道具の発明は、まず人類が「ある種の石はシャープに割れる」と気づいたことから始まったでしょう。この時点では、先に挙げた「カラスのクルミ割り」に近い経験則による学習だった可能性が高いと思いますので、まだ演繹的とは言えません。

ここから、たとえば「①シャープに割れた石で、自分の指を切った」→「②自分の指が切れるなら、狩ってきた肉も切れるかも」→「③肉が切れるなら、シャープに割れた石を投げたら、動物に刺さるかも」→「④シャープに割れた石を細い木の棒にくくりつけたら、刺しやすいかも」→「⑤これで動物を仕留められるかもしれない」などのプロセスで、石槍という武器が誕生したのかもしれません。

ですが②～③の時点で、すでに未来予測としてはかなり複雑です。それを②以降に展開できた動物は、人類だけです。すると②～③を実現するために、小さな因果関係を足したり引いたりするという高度な思考プロセスがあったと推測できます。まさに「物語の『因果関係』を確認して、編集する」というアプローチだと言えます。

AI と人類のちがいがなくなりはじめた

AI の学習は基本的にはカラスと同じです。将棋や囲碁の AI も、絵を描く AI も文章を生成する AI も、音声や画像認識する AI も、すべて同じです。同じというのは「帰納的学習をしている」という意味で、同じです。

しかし大規模言語モデル[注1]は、ざっくり言うと「帰納的学習によって、一部の演繹的推論を身につける」という、驚くべき進歩を遂げたのです。

これまでは、演繹的推論とは「１つひとつの因果関係を積み上げていくこと」だと捉えられていたので、「帰納的学習からは演繹的推論は行われない」と考えるのが一般的でした（以前から演繹と帰納の区別には意味がないと言う人もいましたが、ここでは割愛します）。

けれども桁ちがいに大規模な言語データと AI で帰納的学習を行った結果、データ上に見られるパターンのバリエーションを桁ちがいに身につけました。そのパターンのなかには当然、演繹的推論のパターンも含まれていたのです。

再び「車にクルミを割らせるカラス」を例に引くと、パターンが桁ちがいになったことで、「車にはタイヤがある」「タイヤは地面に圧力をかける」「圧力がクルミを割る」「ホバークラフトにはタイヤがない」といったように、「パターンとパターンの連鎖」も１つのパターンとして身につけてしまいました。結果として、「車にはタイヤがあるので圧力がかかりクルミが割れ、ホバークラフトにはタイヤがないので圧力がかからずクルミが割れない」と、回答できるようになりました。

このように帰納的学習で演繹的推論を行うという能力が大幅に向上したことは、コロンブスの卵のような驚きがあります。ただ、あくまでパターンを連鎖させた結果の弱い演繹的推論であり、人類のように因果関係の意味を理解したり、論理的思考を発展させたりする能力は限定的なので、そこにはまだ隔たりがあります。

しかし、ここからわかることは、実は人類も同じような仕組みを持っているのかもしれないという、新たな視点です。

（林要『温かいテクノロジー　AI の見え方が変わる 人類のこれからが知れる 22 世紀への知的冒険』より）

注１　大規模言語モデル――LLM：Large Language Model：数千万～数十億のパラメータをもつ人工ニューラルネットワークで構成される自然言語処理の汎用モデル。

〈資料２〉

（「白：」＝白井暁彦氏　「８：」＝852 話氏）

Stable Diffusion は何も学んでいない

8：最新の Stable Diffusion（Ver1.4）[注2]を使ってますが、使っていてもモデルそのものに対するフィードバックってないですよね。ではデータを学習させる段階の教師役はどうなっていたのでしょう。

白：Stable Diffusion のモデル自体は何十億枚ぐらいの画像を勉強していて、その後に LAION っていう美学だけをもつデータセットを学習しています。ただ、そのデーターベースのソースを読んだんですけど、全然面白くないんです。単純にポスターとかが並んでいるだけなんですよ。「絵の描き方」を教えてるわけでも「ビューティフルなもの」でもなくて、人間がなんとなく「バランスとして整えられている」と感じている画像の「場所と特徴」をひたすら学習してるだけなんですよね。黄金比みたいなものでもなく、もっと狭くてシンプルな特徴の塊みたいです。

8：Stable Diffusion が AI として、「このポーズが気持ちいいですよ」っていう美学を持ってるってことですか？

白：全く持ってないんです。追加の画像データを使ってファインチューニングされると描けるようになるんですが、それは「何でも描けるのが得意な人」に「こういうの好きなんでしょう？」っていうチューニングをしているだけなんですよ。

8：Stable Diffusion は大量に学習データがあって、そこからノイズを生成して、ノイズを除去して画像が自分の手元に出てきますよね。Stable Diffusion が「正解」として導き出すように見える画像には、実は何の意志もない。

白：みんなは AI 生成画像を見て勘違いしているんじゃないかと思います。たとえば Stable Diffusion がすごい絵を描く。ユーザーには「いいね」だったり、様々なフィードバックがある。AI で画像を生成して SNS に投稿していると、「SNS でこういう絵は意外と人気だった」という感覚を持つし、「Stable Diffusion と自分自身の美的感覚は間違ってなかった！」みたいな人間側のフィードバック学習ができる要素もある。けれど、Stable Diffusion 自体は「その意志」、フィードバックを得てないんです。非常に不思議なんですけれども、Stable Diffusion の中身がどうなってるのか、解明がそんなに進んでいないんです。苦手な構図とか不可能な構図はあるようです。例えば「箱の上にボールが載ってます、みたいな構図は苦手」です、って論文やライセンスに書かれている。「人類が美しいと思っている構図理論」を獲得してるのかなとも思うんですが、もしかしたらそれもないかもしれない。使う側からすると「お前たちはこういうの好きなんだろう？」「こういう構造を美しいと思うんだろう？」っていう画像を、ノイズからの超解像化しかできないのになぜ生成できるのか。本当に分かんないんですよね。

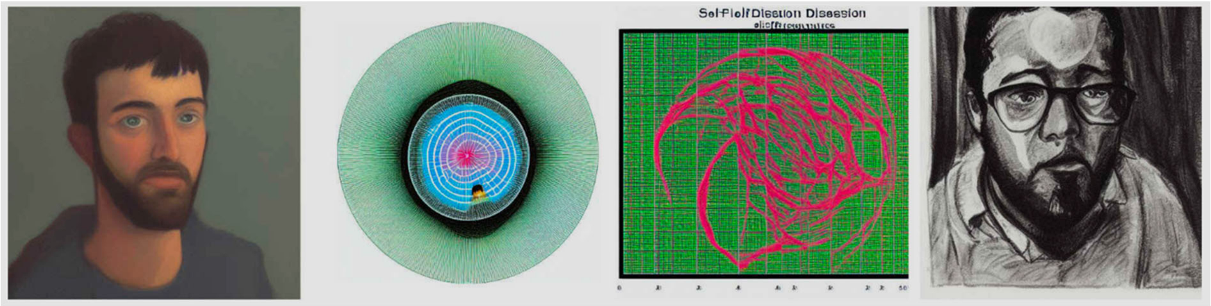
8：人間のデザインっていうか、作ってる人たちが正解というものがあって、最初はランダムなんだけど学習していくうちに多数出せるようになっていって……という話ではない。

白：自分も最初はそうだと思ったんですけど、実はどうも違うんですよ。Diffusion モデルの生成は、人間のフィードバックをかけてないんですよ。つまり、Midjourney[注3] みたいに「我々が何かを選びました」というフィードバックがかかるシステムならともかく、凍結されたモデルに乱数でノイズから超解像かけて生成してるだけ。単純にガチャを回してるだけなんですよね。

8：面白い！　それはもうできあがってる段階でってことですか。

白：獲得の段階は言語と場所の特徴を使っているのですが、獲得したモデルを我々はデータとしてダウンロードして使っているだけです。Stable Diffusion さんに我々のプロンプトの意図が伝わってるわけではない。

図１：筆者が「Self Portrait」（自画像）を Dream Studio[注4] に描かせてみると、ときどき不思議な記号が生成されることがある。



プロンプト生成のセオリー

8：「AI は失敗を恐れない」と感じていて、人間は時間をかけて描くので「明らかに間違った」とか「明らかにウケない」絵を描くのを避けますよね。AI はそれがない。何を出しても正解か不正解か決めるのは人間なので、AI は淡々と生成し続ける。「感情がないから失敗を恐れなくて、失敗を恐れないからいくらでも量産できる」っていうところが AI の強みだと思ってるんですよね。

白：「こんな絵が描けたらいいな」って思っても、描くのに一週間かかっちゃうとそれ自体に愛着が生まれてしまう。時間かければかけただけ思い入れも増える、失敗すると恥ずかしい気持ちが出てくる。でも AI は一切こういう疲れとか傷つきとかがないんですよね。

8：「感情を持って絵を描いてる人達」が AI の絵に「温かみがない」とか「こんなもの」って発言するのは理解できるんですよ。「人の努力の結晶としての絵」を評価したい層は絶対いるので。共感や感動、アーティストが好きという感情があって、「その人から生み出されるものは尊い」みたいな人は絶対いなくならないと思うんですよね。

白：逆に、最近だと 852 話さんの絵が AI で描かれたと知った上で好き、と言う方も増えてますよね。852 話さん独自の「プロンプト[注5]生成のセオリー」ってあるんでしょうか。スタイルから入るか、サブジェクトから入るか、プロンプトの順番とか。プロンプトについて正しそうなことを言ってる人はたくさんいるんだけれども、人の心を動かすようなものをたくさん描いてる人からすると、何かセオリーはあるんでしょうか。

8：そうですね……まずは、写真っぽいかイラストっぽいか？　自分の好みに寄せるか万人に寄せるか、日本に受けるか、海外に受けるか？　男性か女性か？　ファンタジーな髪色か、万人受けの髪色か？……という感じで、どんどん２択で分かれていきます。かけ離れた選択肢もときど

き入れながらバリエーションを増やしていく。「幹」があって「枝葉」を出していく感じです。例えば MidJourney なら、スタイルは「Mucha 風」で、これは「自分も好きで万人に受ける」とかですね。自分の好みに寄せるか万人にウケるものに寄せるかで、まず２択があります。

図１：筆者による"anime girl loved by all"を軸にプロンプトを追加して比較していった作例（Dream Studio にて）

（白井暁彦著『AI とコラボして神絵師になる　論文から読み解く Stable Diffusion』　出題のため一部を編集）

注２　Stable Diffusion――Stability AI 社開発の言語から画像を生成する AI モデル。2024 年２月現在は Ver1.7.0 を公開。また同社は、日本語向け大規模言語モデル「Japanese Stable LM 3B-4E1T Base」と「Japanese Stable LM Gamma 7B」も公開している。

注３　MidJourney――MidJourney 社開発の言語から画像を生成する AI モデル。

注４　Dream Studio――Stable Diffusion のオープン β 版として公開されたサービス。

注５　プロンプト――AI モデルに生成を促す言語文字列。

問１　〈資料１〉について、下線部「驚くべき進歩」とは、何を指すか。本文における「帰納的な推論」と「演繹的な推論」の区別を明らかにしながら、300 字以内で説明しなさい。

問２　「2050 年の AI とあなたの生活」について、800 字以上 1200 字以内で自由に論述しなさい。その際、〈資料１〉〈資料２〉を踏まえながら、あなたが本学での学びを活かして挑戦したいことに必ず触れること。