ChatGPT を用いた

記述式法律学科試験問題の解答生成と

専門家による評価

2020428 佐久間 千佳

令和 5 年度　提出

目次

第 1 章　はじめに 3

第 2 章　関連研究 4

2.1　米司法試験問題を題材とした従来研究 4
2.2　日本の試験問題における ChatGPT 4

第 3 章　研究の手法 5

3.1　研究の流れ 5
3.2　研究の対象 5
3.3　問題の入力 6
3.4　問題の出力 8
3.5　評価方法 10

第 4 章　結果 11

4.1　全体の得点結果 11
4.2　各問題の分析 12
4.3　考察 14

第 5 章　まとめ 16

謝辞 17

参考文献 17

第 1 章　はじめに

ChatGPT とは、2022 年に米 OpenAI 社によって公開された、大規模自然言語モデルを基にする対話型生成系 AI である。現在、教育や医学、行政など、多岐に渡る分野において、この ChatGPT を効果的に利用することで、作業の効率化、迅速化を進めようという動きが活発になっている [1,4,5]が、法律分野においても、その活用による課題解決が期待されている[2]。

中でも法律学科試験については、ChatGPT の能力を測る問題の一つとして扱われており、その解答生成の試みは盛んである[3]。しかし、主としてそのような研究を進めているのは ChatGPT のリリース元である米国であり、その対象も米国の司法試験問題であるため、日本の司法に基づいた試みは少ない。さらに、日本の法律学科問題を扱っていたとしても、あくまで選択式問題のみを対象とする場合にとどまり、配点の多くを占める記述式問題には対応していない。これは、答えが一つに決まり、正誤の判断が明確である選択式問題に対して、記述式問題は、その評価に法律の専門家による判断が必要となるからである。

この問題を解決するためには、以下の点を踏まえて検証実験を行う必要がある。

①日本の司法に基づいた問題
②選択式ではなく記述式問題
③法律の専門家による解答の評価

そこで本研究では、日本司法に基づいた記述式法学試験問題の解答を ChatGPT に生成させ、法律の専門家である法律学科教員にその解答の採点を依頼することで、ChatGPT による法律問題の処理精度を明らかにする。

以下、本論文は、第 2 章において ChatGPT が法律分野においてどのような発展を進めているのかについて従来研究を示す。第 3 章では、本研究の方法や評価方法について具体的な説明を述べる。第 4 章では、本研究の結果を提示し、それに対する分析、考察を行う。最後に、第 5 章において本研究を総括する。

第 2 章　関連研究

2.1　米司法試験問題を題材とした従来研究

米国では ChatGPT が公開されてから、司法試験問題を対象として解答生成の精度を検証する研究が行われてきた。2022 年 12 月には、Michael Bommarito II らによって、GPT-3 .5 モデルを利用した ChatGPT の米司法試験の選択式問題に対する解答の正答率が検証された［1］。この結果、GPT-3.5 は 50.3%の平均正解率を達成し、7 科目の平均では人間の受験者（平均正解率 68%）に約 18%の差で負けたものの、証拠法と不法行為法の 2 科目では平均合格率を超えていたことがわかっている。また、2023 年 3 月に新たに公開された、従来の GPT-3.5 に比べ解答精度が高いとされる GPT-4 モデルついても、OpenAI 社は公開した“GPT-4 Technical Report”においてその法分野での処理能力に言及している[2]。これによると、ChatGPT は、米司法試験に合格し、上位 10%の成績をマークしたとされる。ここでは、択一式の短答問題だけでなく記述式の論文問題についても対象としていた。

このように、米司法試験を対象とした研究は大きく発展しており、その処理能力の高さが証明されている。

2.2　日本の試験問題における ChatGPT

ChatGPT によって日本の試験問題の解答生成を行う試みは、いくつかの分野で行われてきた。増田らによる研究では、公認会計士試験を ChatGPT が突破できるかについて検証が行われた[3]。ここで対象としていたのは短答式問題であり、記述式問題についての処理能力については報告されていない。その他にも、Jungo Kasai らによって、GPT-4 を用いて日本の医師国家試験の解答に試みる研究が行われているが、ここでも対象となる問題について、数字を必要とするいくつかの例外の他はほとんどすべて多肢選択問題であった[4]。

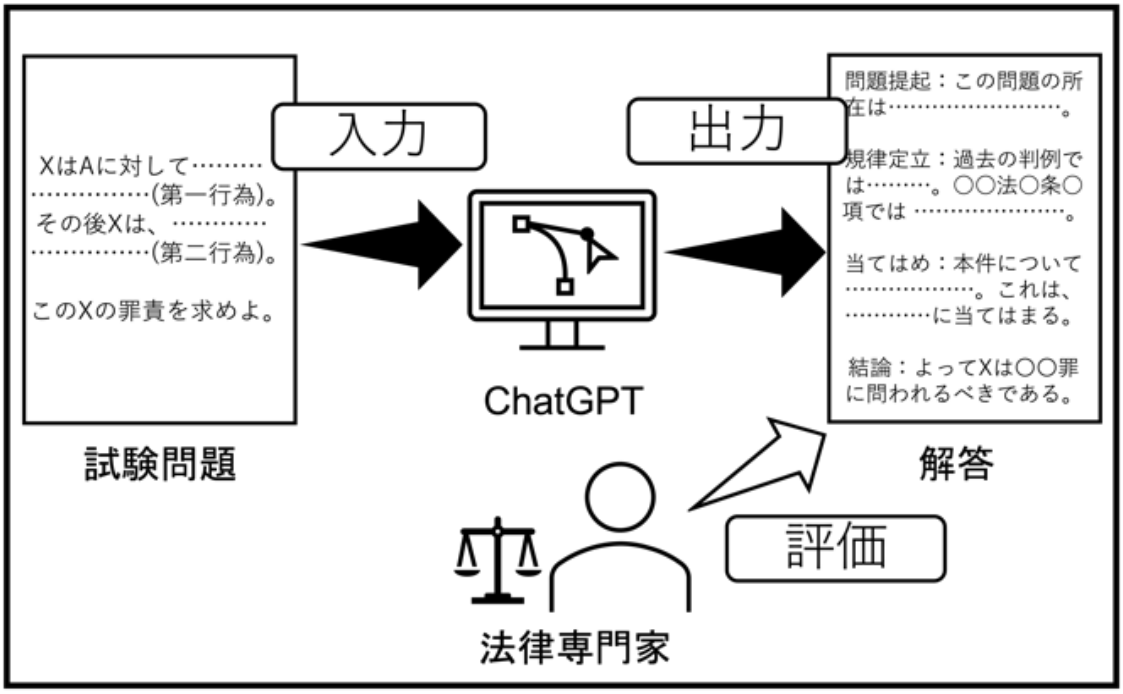
法律分野においても、ChatGPT を用いて試験問題を解答させる試みは行われている。株式会社 Legalscape は、GPT-4 を用いた自社開発の AI によって法律問題の解答を試みているが、ここでも取り上げているのは一部分野の短答式問題のみであり、記述式問題については扱っていない[5]。

このように、日本でも法律分野に限らず多様な分野において ChatGPT を用いた問題の解答生成が試行されているが、その対象となるのは専ら正誤の判断がわかりやすい選択式問題や短答式問題である。したがって、本研究のように、日本の知識領域分野において、記述式問題を対象として ChatGPT の解答について分析、評価を行うことは、これまでになかった試みであり、ChatGPT の解答生成の能力判断にあたって意義があるものだといえる。

第 3 章　研究の手法

3.1　研究の流れ

本研究は OpenAI 社によって公開された、大規模自然言語モデルである「ChatGPT」に法律学科の試験問題を入力し、解答を出力する。図 1 に本家研究の流れを示す。図 1 に示したように、出力された ChatGPT の解答は、法律の専門家である法律学科教員に直接添削を依頼した。その採点結果を基に、ChatGPT による法律問題の処理精度を判断する。



次節からは、研究の対象や方法について、具体的に説明する。

3.2　研究の対象

本研究の題材とする法律学科試験は、小樽商科大学で開講された法学科目の期末試験とした。これは、以下の三つの理由による。

①司法試験といった過去の公的試験は web 上に問題と解答がそのまま記載されている可能性があるため利用が難しい。

②科目試験問題には当該科目の担当教員によって明確な採点基準が存在し、点数として数値で正答率を図ることができる。

③評価を当該科目の担当教員に依頼することで、問題制作者が直接採点することができる。

また、対象とする法科目は、法律の条文や判例を基に作成する必要のある学科として、民法、刑法（I、II）を採用した。

図 1　本研究の流れ

本研究に用いた民法の期末試験問題は、債権法を範囲としている。また、刑法については、刑法Ⅰは、犯罪と刑罰の基礎理論、犯罪の一般的成立要件、全ての犯罪に共通する理論、刑罰の種類と適用などを対象とする刑法総論を、刑法Ⅱは、個々の犯罪を規定した各刑罰法規の解釈が中心となる刑法各論を範囲としている。本学の科目試験問題については、文章量の多さや事案の複雑さから見て、刑法Ⅰよりも刑法Ⅱの方が解答するにあたっての難易度が高いといえる。

各問題の出題内容と配点については、表1に示す。

表1　各問題の出題内容と配点

科目	設問	出題内容	配点
民法	問1（1）	二重債権譲渡	8
	問1（2）	契約不適合責任	8
	問2	敷金の返金	8
刑法Ⅰ	問1（1）	行為と結果の因果関係	60
	問1（2）	正当防衛	60
刑法Ⅱ	問1	共謀共同正犯と因果関係	60

3.3　問題の入力

現在、一般に幅広く利用されているWebサービスとしてのChatGPTはGPT-3.5モデルが搭載されているが、本研究で使用したのは、2023年3月に新たに公開されたGPT-4モデルである。このGPT-4は、従来のモデルよりも解答精度が高く、長い対話も可能となった。また本研究では、Chat GPTを外部から利用して処理を行うことができるように、API[1]を利用した。APIを用いることで、プログラミングによってこのGPT-4の処理を行うことができるため、入出力にあたっての細かい指示ができるようになる。また、ここで使用したプログラミング言語は、言語処理に長けたPythonとした。

ChatGPTに入力する問題文は、背景となる事案についての記述説明部分も含めて、実際の試験問題の文章をそのまま利用した。さらに、出力の精度を高め、出題意図に沿った解答形式にするために、問題文の入力に加え、プロンプト[2]を追加し、形式を詳しく指定した。

---

[1] APIとは「Application Programming Interface（アプリケーション・プログラミング・インターフェース）」の略称で、ソフトウェアやWebサービスの間をつなぐインターフェースのことである。

[2] プロンプトとは、プロンプトエンジニアリングのことであり、ChatGPTなどの大規

本研究のプロンプトは、出力の精度を高めるために、ChatGPT の置かれた立場について、「東京大学法学部の首席レベルの優秀な学生」とし、さらに、状況の分かりやすさを求めるために、法律学科の試験問題として日本で最も一般的であり、求められる解答の質が高いものである「日本の司法試験の記述式問題」として、対象を指定した。

また、法律科目の記述式問題について、その解答には適切な形式が存在する。よって、それについても、プロンプトの命令と同じように ChatGPT に指定する必要がある。

法律学科科目の記述式問題は基本的に、「問題提起」、「規律定立」、「あてはめ」、「結論」の順に記述し、解答を構成する。ここで、民法と刑法で担当教員から指定されている解答形式に若干の違いがあった。民法については、基本的なこの四つの構成に準じた答案で解答要件を満たすことができる一方で、刑法では「規律定立」が「学説の議論状況」、「適切な学説の検討」、「類似判例の状況」の三つの構成に分けられる。

したがって、民法においては「問題提起」、「規律定立」、「あてはめ」、「結論」

の構成をなすように、刑法については、「問題提起」、「学説の議論状況」、「適切な学説の検討」、「類似判例の状況」、「あてはめ」、「結論」の構成をなすように、それぞれ入力の際に条件を記入した。

また、記述式問題は選択式問題と違って解答が明確に定まらないため、複数回試行した上で出力結果の統計をとって分析することが難しい。したがって、一度の試行で適切に ChatGPT の解答を分析する必要がある。よって解答の再現性を向上させるために、解答のランダム性を左右する要素である‘temperature’を、最も小さい 0 に指定する入力を追加した。

模言語モデルと効果的に会話するために必要なスキルセット。入力の際に ChatGPT に条件や命令を記入することで、出力の質を向上させることができる。

このような点に留意して入力したコードは以下のようになった。

```
!pip install openai
import os

os.environ["OPENAI_API_KEY"] = “【キーの入力】”

import openai
response = openai.ChatCompletion.create(
  model="gpt-4",
  temperature=0,

  messages=[{"role": "system", "content": "あなたは、東京大学法学部の首席レベ

ルの優秀な学生です。以下の日本の司法試験の記述式問題について、解答を記

述しなさい。また、解答は法律の論述問題の解答形式に則り、問題提起、学説
```

3.4　問題の出力

上記のような入力を経て、ChatGPT が出力した解答は以下のようになった。

```
print(response["choices"][0]["message"]["content"])
```

問題提起:
本件における問題は、X の行為が A の死亡に対して果たした役割と、その後の Y の行為が A の死亡に対して果たした役割をどのように評価するか、という点にある。具体的には、X の行為が A の死亡に対して直接的な因果関係を持つか、あるいは Y の行為が A の死亡に対して直接的な因果関係を持つか、という問題である。

学説の議論状況:
この問題については、主に二つの学説が存在する。一つは、最初の行為者(本件では X)の行為が被害者(本件では A)の死亡に対して直接的な因果関係を持つとする学説である。もう一つは、後の行為者(本件では Y)の行為が被害者の死亡に対して直接的な因果関係を持つとする学説である。

適切な学説の検討:
本件においては、X の行為が A の死亡に対して直接的な因果関係を持つとする学説を採用するのが適切であると考えられる。なぜなら、X の行為がなければ A の死亡は起こらなかったと考えられるからである。また、Y の行為は A の死亡を早める結果をもたらしたが、それは既に X の行為によって A の死亡が決定的となっていた状況下での出来事であるため、Y の行為が A の死亡に対して直接的な因果関係を持つとは言えない。

類似判例の状況:
類似の判例としては、最高裁判所昭和 34 年 11 月 25 日判決がある。この判決では、被害者が既に死亡する状態にある中で、第三者の行為が被害者の死亡を早めた場合でも、最初の行為者の罪責は消滅しないとされた。

あてはめ:
本件においても、X の行為が A の死亡に対して直接的な因果関係を持つと考えられる。また、Y の行為は A の死亡を早めたものの、それは既に X の行為によって A の死亡が決定的となっていた状況下での出来事であるため、Y の行為が A の死亡に対して直接的な因果関係を持つとは言えない。

結論:
以上の理由から、X は A の死亡に対して直接的な因果関係を持つ行為を行ったとして、殺人罪に問われるべきである。

一例として挙げたのは、刑法Ⅰの問 1（1）の行為と結果の因果関係に関する問題の出力結果である。ここに、ChatGPT が生成する法律学科記述式問題の解答は、文章が破綻していたり、通読していて意味が通らなかったりするようなものではないということがわかる。ChatGPT の法律学科記述式問題に対する解答の処理精度を正しく判断するためには、この文章が法学的知見に基づいた観点からすると、どのような評価が得られるのかについて理解する必要がある。そこで次節では、この ChatGPT による生成解答についての評価方法を説明する。

3.5　評価方法

3.2「研究の対象」でも述べたように、ChatGPT によって生成された解答は、当該試験科目の担当教員の方に採点を依頼した。採点結果として、解答の点数とともに、生成された文章について法学的側面からの解説や論評を受け取った。

本研究で対象とするのは民法と刑法であるが、この二つの科目ではそれぞれ担当された教員が違うこともあり、採点基準や解答形式が異なっている。それぞれの問題の解答を分析するにあたっては、この違いについても考慮する必要があるだろう。

まず民法の採点は加点方式をとっており、表 1 に示した通り、全ての問題が 8 点満点であった。刑法についても刑法 I 、刑法 II ともに、採点方法は加点方式をとっている。刑法は、加点部分 1 か所について一定の点数が定められており、満点は 60 点となっている。誤った文章に関しては教員の方からご指摘をいただいたが、減点方式ではないため、その部分に関して点数がひかれているわけではない。つまり、満点だった場合でも、加点された要素が満点に定められた点数に達していただけで、要素としては誤った記述を含むといった場合もある。一方で、採点が 0 点だった場合は、文章中に加点要素が一つもなく、さらに、加点される要点に関して、記述の全てが誤っていたか、もしくは記述していなかったということになる。

本論文では、頂いた採点結果をもとに、ChatGPT の出力する法学科目問題の解答がどの程度の精度と信頼性を持っているのか、さらに、利用するにあたってどのような問題点があるのかを明らかにする。

第 4 章　結果

4.1　全体の得点結果

ChatGPT によって生成された解答の得点結果は以下の表のようになった。

表 2　民法の得点

科目	設問	得点
民法	問 1（1） 二重債権譲渡	8/8
	問 1（2） 契約不適合責任	2/8
	問 2 敷金の返金	4/8

表 3 刑法の得点

科目	設問	得点
刑法 I	問 1（1） 行為と結果の因果関係	15~20 /60 ※
	問 1（2） 正当防衛	10/60
刑法 II	問 1 共謀共同正犯と因果関係	0/60

※採点者による採点では点数に幅を持たせていた

全体をみると、問題によって得点率に大きなばらつきがあることがわかる。

最も得点率が高いのは民法の問 1（1）の二重債権譲渡に関する問題であり、8 点満点中 8 点と、点数としては満点を取っている。一方で、刑法 II の共謀共同正犯と因果関係に関する問題では、60 点満点中 0 点と、解答中に一つも加点要素がなかった。しかし、このように加点要素がなく 0 点だったのは全体を通しても刑法 II のこの問題のみであり、他の 5 問においては程度の差はあれ、解答中の単語、文脈、説明において何らかの加点要素があったことがわかる。

また、民法と刑法を比較すると、刑法の方が民法よりも得点率が低い傾向にある。しかし、問題形式や問題の難易度、配点や採点基準など、多くの要素で条件が違うことから、単純な

比較は難しい。そこで次の節では、それぞれの科目別に、個々の解答の内容を詳しく分析する。

4.2　各問題の分析

本節では、民法、刑法の科目別に、個々の解答における採点結果について、採点者によって「正しい記述」とされて加点された要素、また解答として「不十分である記述」とされた要素について検討する。ここで、「不十分」というのは、内容が誤っている場合と、必要な記述が不足している場合の二つがある。

まず民法について、解答の採点から、正しい記述と不十分である記述をまとめると以下の表のようになった。

表 4　民法の採点結果

科目	設問	出題内容	正しい記述	不十分である記述	得点
民法	問　1（1）	二重債権譲渡	・条文の引用 ・結論	・解答の過程が不鮮明 ・判例の引用	8/8
	問　1（2）	契約不適合責任		・条文の内容 ・学説の説明・用語 ・判例の引用	2/8
	問 2	敷金の返金	・結論	・条文番号 ・判例の引用	4/8

問 1（1）は得点としては 8 点満点中 8 点を取っており、「条文の引用」や「結論」に正しい要素があった。一方で、不十分であるとされる記述も存在しており、完璧な正しい記述のみで解答が構成されていたというわけではなく、あくまで加点要素が満点の 8 点に達していたと言った方が正しい。

したがって、全ての問題において何かしら記述に不足があったということになる。不十分な記述として共通していたのは、「判例の引用」である。これは引用した判例が問題に不適合なものだったというわけではなく、そもそも判例が一つも引用されていなかった。また、「条文の引用」についても、問 1（2）と問 2 で共通して記述が不十分であった。記述自体がなかった判例とは違い、条文は解答中に引用されてはいたものの、その内容や条文番号に間違いがあった。特に、問 1（2）に至っては、存在しない文言の条文をでっちあげて引用するなど、重大なハルシネーション[3]を引き起こしていた。これは、法律という絶対的な法

[3] ハルシネーションとは、人工知能（AI）が事実に基づかない情報を生成する現象のこと。まるで AI が幻覚（＝ハルシネーション）を見ているかのように、もっともらし

的根拠を基に解答を生成する法学科の問題を扱うにあたって、大きな問題であるといえる。

次いで、刑法の採点結果について検討する。正しい記述と不十分である記述は次ページの表のようになった。

表 5　刑法の採点結果

科目	設問	出題内容	正しい記述	不十分である記述	得点
刑法 I	問　1 (1)	行為と結果の因果関係	・問題提起 ・あてはめ ・結論の断定	・学説の説明 ・類似判例の引用	15~20 /60
	問　1 (2)	正当防衛	・問題提起の一部 ・着目点のキーワード	・問題の所在 ・学説の説明・検討 ・類似判例の引用 ・結論	10/60
刑法 II	問 1	共謀共同正犯と因果関係		・問題提起 ・学説の説明・検討 ・類似判例の引用 ・あてはめ ・結論	0/60

まず、刑法 I について、問 1 (1)、(2) のどちらの問題にも共通して、「問題提起」において加点があった。しかし、(2) はあくまで問題提起の一部のキーワードに加点があったのみで、「問題の所在」は誤った記述がなされていた。この問題では、「問題の所在」という解答の論点が誤っていたため、その後の記述も出題者の意図する説明、キーワードがほとんどなく、加点が見込めなかった。

また、どちらの問題にも共通して不十分な記述として、「学説の説明」「類似判例の引用」がある。学説について、解答では適当な学説を複数引用して、問題に当てはめて正しい学説を検討する必要があるが、問題 (1)、(2) どちらについても引用する学説が誤っていたため、その後の検討についても加点要素が得られなかった。特に、(2) に関しては、存在しない名称の学説をでっちあげて説明しており、ここでも重大なハルシネーションを引き起こしていた。同様に、類似判例の引用についても、どちらの問題でも存在しない判例文を引用していた上に、説明も不鮮明であった。刑法におけるこれらの問題では、類似判例の引用とそれ

い嘘（事実とは異なる内容）を出力するため、このように呼ばれている。

を基にした当てはめが問題の重要な要点である。採点者の解説においても、「引用する判例が不存在である点で、他の箇所で正しいことが書けていたとしても、全体として疑わしい文章」と評されており、正しい判例を引用できないというのは大きな問題であるといえる。

刑法 II の共謀共同正犯と因果関係に関する問題については、得点が 0 点であったこともあり、問題提起や論の展開など、出力された解答の全てが出題意図からそれたものであった。刑法 I と同様に「学説の説明」「類似判例の引用」についての記述が不十分であっただけでなく、いくつかのキーワードが加点要素として設定されていたにも関わらずそれらの用語が一つも含まれていなかった。また、刑法の試験では罪責が判例と異なっていてもそれに至る論理の過程に妥当性があれば減点することはないとされている「結論」においても、ChatGPT の出力解答にはその理論的妥当性がないとされ、加点には至らなかった。つまり、刑法 II の問題については ChatGPT の出力した解答は完全に誤りであったといえる。

4.3 考察

ここまで、小樽商科大学の法律学科試験問題に対して、ChatGPT が出力した解答の採点結果について分析してきた。本節では、これを踏まえて、ChatGPT による法律学科記述式問題の解答生成が、具体的にどのような点を得意とし、またどのような点で問題があるのかについて考察する。

まず、全体を通して、ChatGPT が得意といえる要素は、「問題提起」と「結論」であるといえる。この二つの要素は、解答するにあたって答案作成者が自ら立てる「問い」と「答え」の関係であり、解答を構成する軸となるものである。したがって ChatGPT は、入力された問題に対して、何が問われており、どのような答えを導く構成にすれば良いのかという点についてはある程度理解することができるということだ。しかし、生成した解答の中には、この「問題提起」と「解答」が不十分であったものもいくつかあった。特に、加点要素が一つもなかった刑法 II の共謀共同正犯と因果関係に関する問題では、この記述が全て誤っており、問題の出題意図から大きくずれた解答を生成していた。刑法 II のこの問題がこのように特に点数が取れなかった要因について、刑法 I よりも問題文が長く、発展的な内容であり解答難易度が高かったことが挙げられる。よって、あくまで「問題提起」と「結論」というのは、他の要素に比べて比較的加点されうる傾向が強いということにとどまり、問題の内容や難易度によって、それが果たされない場合もある。

一方で、ChatGPT が特に不得意とする要素は「判例の引用」「条文の引用」「学説の説明」である。判例や条文、学説というのは、法学的根拠を示すものであり、法学的な観点から見て解答が正しいかどうかを図る指針として重要なものである。ChatGPT の生成した解答では、単にこれらの記述が不足していたのではなく、存在しない判例や条文、学説を勝手に作り上げてそれらしく説明を加えており、法律的観点からすると非常に信頼性が低いものであるといえる。

このような、判例、条文、学説といった法学的な事実や知識にまつわる記述が不正確だっ

たということは、ChatGPT がこれらの情報を検索し、出力として採用する過程に問題があるということだ。まず条文について、六法全書が毎年改訂されているように、法律というのは都度新しく変わっていくものである。しかし ChatGPT は、それを認識してインターネット上にある数多の情報の中から、問題に適した正しい番号、内容の条文を引用することが難しいようだ。また判例に関しても同様に、正しいものを見つけ出して採用することが出来ていない。条文よりも判例の引用の方が、精度が低くハルシネーションを引き起こす傾向が強かったことから、こちらの方が正しい情報を精査するのが難しいのではないかと考える。また、学説についての記述というのは、授業で説明された内容が最も反映される要素である。ChatGPT はあくまでインターネット上にある情報を引き出すもので、担当教員によってまとめられたレジュメは当たり前だが参照できない。したがって、条文や判例よりも、検索するだけではなかなか辿り着かない深い法学的知識である「学説の説明」というのは、ChatGPT が正しく記述するには、より難しかったのではないだろうか。刑法では、出力する解答の構成として「学説の議論状況」「適切な学説の検討」が条件として指定されていたため、ChatGPT は検索による上澄みの情報を頼りに誤った学説を作り挙げて説明していたが、解答の構成要件として「学説の説明」が指示されていなかった民法は、そもそもそれについて述べることが少なかった。

ChatGPT が法律学科の記述式問題の解答を生成するにあたって、判例や条文、学説といった法学的根拠を正しく参照できないといった問題点を解決し、今よりも処理精度を高めるためには、外部から最新の条文や判例を ChatGPT に反映させたり、学説の議論状況といった専門的な法学知識を学習させたりする必要がある。

第 5 章　まとめ

本研究では、ChatGPT によって生成された、日本の司法に基づいた法学記述式問題の解答を、法律の専門家である法律学科教員によって評価してもらうことで、その処理精度と問題点について検討することを目的として試行を行った。研究の対象として小樽商科大学の法律学科科目の刑法、民法の試験問題を用い、その採点を当該科目の担当教員に依頼することで、生成解答を評価した。

検証によると、ChatGPT によって生成した解答の採点では、最高得点が 8 点満点中 8 点、最低得点が 60 点満点中 0 点だった。このように得点は問題によって大きくばらつきがあった一方で、その採点内容には傾向が見られた。傾向として、得点の高かった問題では、「問題提起」や「結論」といった、出題意図を踏まえて解答の展開を決定するための基本的な要素が正しい場合が多かった。一方で、「判例の引用」「条文の引用」「学説の説明」といった、法学的根拠を参照する必要のある要素については、多くの問題に共通して適切に記述することができていなかった。

したがって、ChatGPT による記述式法律学科の試験問題の解答生成は、解答の展開を構成する能力には長けている一方で、判例や条文、学説といった法学分野の知識については欠けており、法学的な信頼性が著しく低いということが判明した。これは法学の試験問題を扱うにあたって、大きな課題であるといえる。

今後の課題として、外部から最新の判例や条文、学説を ChatGPT に学習させ、適切な法学的根拠に基づいて解答生成を行うようにさせる必要がある。

謝辞

本研究の実施にあたり、小樽商科大学商学部企業法学科 竹村准教授、並びに同学科 菅沼准教授にご協力いただきました。試験問題のご提供だけでなく、解答の採点につきまして法学の専門家として詳細なご助言を頂きました両名に、心より感謝します。

参考文献

[1] 寺岡 三左子 ”ChatGPT®は有効な教育支援ツールになり得るか？－看護学における対話型 AI をめぐる議論の動向－” 医療看護研究 32, pp64-70, 2023

[2] Michael Bommarito II&Martin KatzDaniel. "*GPT Takes the Bar Exam*."Social Science Research Networ,2022

[3] ApenAI.” GPT-4 Technical Report”,2023

[4] 増田樹，中川慧，星野崇宏 ,“ChatGPT は公認会計士試験を突破できるか？: 短答式試験監査論への挑戦” 人工知能学会第二種研究会資料,2023

[5] Jungo Kasai, Yuhei Kasai, Keisuke Sakaguchi, Yu- taro Yamada, and Dragomir Radev. “Evaluating gpt-4 and chatgpt on japanese medical licensing examina- tions.”arXiv preprint arXiv:2303.18027, 2023

[6] 『生成 AI が司法試験「合格水準」 東大発新興、一部科目で』日本経済新聞 2023-6-11 (最終閲覧日 2023-11-26)
https://www.nikkei.com/article/DGXZQOUC317WP0R30C23A5000000/