

# Pythonプログラミング演習授業に おけるChatGPT

利用実態と影響

佐藤重幸(東京大学MIセンター)

数理・データサイエンス・AI教育強化拠点コンソーシアム関東ブロック  
第2回ワークショップ「生成AIとデータサイエンス教育」

2023-08-01

# 「Pythonプログラミング入門」

- 学部2年(先行履修)～大学院生が受講可能な全学的オンライン授業
  - 数理・データサイエンス教育プログラムの一部
  - データサイエンスをキラーアプリとしつつ Pythonプログラミングの基本を扱う
- 1ターム(7週)授業を年2回開講
  - 年間受講者600～700名
  - 2023S1の受講者351名(内履修者272名, 理8文2, 学部5.5大学院4.5)
- 反転授業
  - 教材自習と課題演習から入って個別フィードバックで教育
  - 授業中は全て質問対応
- 事実上のオンデマンド
  - 教科書は[オンライン公開されている](#)
  - 課題提出のみで成績評定されて出席は不要
  - 課題のヒント(穴あきプログラム)は動画として漸次提供される
  - 答えは, Webシステム[PLAGS UT](#)の上で自動評価されて, 即座にフィードバックが返される

# 経緯：3月末ガイダンス準備中の担当教員間の議論

「欧米の大学では生成AIに関する方針が出されている。全学や部局のレベルでの公式な方針は現状ないが、授業としての方針を出すべきでは？」

- 至極真っ当な指摘

「ChatGPTの利用を禁止するのは良くない。利用していることを検知できないので、正直に指示を守った人が損をする不公平さが生じる。」

- 技術的に正しい指摘

「授業にとって有益な指示が分からない。教員側の願望を伝えても『指示』としての実効性が乏しい。上位のレベルで方針が決まっていなければ、方針を出す必要はないのでは？」

- 懐疑主義かつ現実主義の指摘 (by 佐藤)

➡ 結局、ChatGPTについて何も言わないことを選ぶ

## 経過:4月以降

4/3に太田副学長(教育・情報担当)による「[生成系AI\(ChatGPT, BingAI, Bard, Midjourney, Stable Diffusion等\)について](#)」が出される

- 事実上の全学的方針

4/10のガイダンスでは当該記事を紹介して、参照するように指示

- ガイダンススライドには明記せず、動画にのみ残す

以降のS1の授業では生成AIに関する言及はなし

- 《沈黙》の方針に従う

## 結果：授業終了時アンケート

設問「本授業におけるChatGPTの利用について教えてください」

回答	回答数(N=92)
学習ツールとして積極的に利用した	14
ヘルプとして必要に応じて利用した(+α)	38
使ってみたが役に立たなかった(or 利用できなかった)	5
自主的に利用を禁じた(or 利用しなかった)	30
ChatGPTの使い方がわからない	4
ChatGPTを知らない	1

# アンケート結果から分かること

- 受講生側でもChatGPTに対する立場は割れている
  - 56.5% (52/92) が有効活用している
  - 38% (35/92) は使おうとしてもいない
- 利用者視点での利便性や満足度は高い
  - 利用者の91% (52/57) が便利に使っている
- 利用したくても OpenAI の都合で利用できないことがある
  - メールアドレスがキックされてアカウントが作れないことがある
- 用途(自由記述欄より)
  - Python APIリファレンス
  - コード改善(リファクタリング支援)
  - ミスの指摘
  - 教材(日本語)に対する高度な翻訳機

# アンケート結果を踏まえた見解

- 少なくとも受講生にアンケートを実施することは有益
  - 何に使えるのか、使って嬉しいのかは、教員側には分からない
- ➡ アンケートに正直に回答されるような授業環境を作ることは重要
- 生成AIに対する受容度が論点
  - 「使って当然」の人と「危うい」の人では、指示の実効性は変わる
    - 研究室M1「なんでも知っている便利な暇人」
    - 萩谷先生「何も分かってないのに適当なことを喋る奴」( MLSE夏合宿2023)
  - 社会全体の受容度と大きく乖離した教育方針は不自然
  - 学生側の受容度と社会全体での受容度は、大きく乖離しないと想定するのが自然
- ➡ 経過観察を続けながら時々状況に対応していくしかない

## 参考:ソフトウェア環境での受容度

- [Windows Copilot for Windows 11](#)
  - ChatGPTを組み込んだ対話的ヘルプ
  - Windows 11 Insider Preview Build 23493 で利用可能
- [Google Colab上での AI-powered coding 機能](#)
  - PaLM 2に基づくコード生成用モデル Codey が組み込まれる
  - [6/27 に有料版 US Pro+ で使えるようになった](#)
  - *"We'll be expanding access to US Pro subscribers in the coming days, and to the free-of-charge tier and other regions as soon as we are able."*

➡ ソフトウェア環境を通じて受容せざるを得なくなる兆し

## AIツールの授業における利用について(ver. 1.0)

- [1] 自分がこれまで課してきた課題や試験などを題材にChatGPT、BingAI、Bardなどの言語生成系AIに答えさせ、そのレベルを感じておく
- [2] 授業や課題ごとに、言語生成系AI利用に対する教員のスタンスを明示する
- [3] 課題の目的、学生にとっての達成目標、成長目標を学生に伝える。得られた結果ではなく解答を得る過程が重要であることを説く
- [4] 実践可能な範囲で、言語生成系AIによって安易に解答が得られない課題・出題形式を検討する
- [5] AIによって生成された文章であるかの検出ツールは過信しない
- [6] 教育効果をもたらす影響と効果について

# 実験: どれくらいChatGPTは課題を解けるのか

- GPT-4 (2023-07-18付けChatGPT有料版) を利用
- 自動評価付きの全24題について試行
- 正誤判定の基準はPLAGS UTの自動評価

## プロンプトの与え方

- 「これから送る問題に Python で解答してください。」と最初に与える
- 以降, 問題文のMarkdownをコピーするだけ
  - 問題文の前段に導入部がある場合, それは含まない
  - 問題文にはLaTeX形式の数式やコード例も含まれる

(Special thanks to 利根 悠司)

# 実験結果

- 20/24に正答
- 間違った4題も些細なものであり、対話的プロンプトを数回重ねれば正答に至るように見える
  - 例1:問題文の前段の導入部を踏まえていない
  - 例2:問題文に附属するテストケースを踏まえていない
- 答案コードはどれもとても綺麗
  - 教員のコードよりも綺麗
  - 私(専門家)の想定を超えてシンプルなもの 2例あった
- ChatGPTは答案コードに日本語の解説も付ける
  - それが全体として適切である

# 問題文テスターとしてのChatGPT

- ChatGPTが誤読した問題文を、少なくない受講生が同様に誤読している
    - 問題文を良く読んで、附属するテストケースを読めば、題意は理解可能
    - 総合的な読解力をプログラミング初学者の一部が持たないのは自然
- ➡ ChatGPT-proofの問題文であれば初学者の誤読や疑問を抑制する

## 問題文テスターとしてChatGPTを使うと

- 教員単独で高度な校正ができるので、高品質の課題を提供しやすい
  - その構成からChatGPTで正答可能な課題にしかない
- ➡ 欠点を補って余りある便益がある(と私は思う)

# 東京大学の学生の皆さんへ:

## AIツールの授業における利用について(ver. 1.0)

...

[3][**教育効果の重視**] 大学での学びにおいては、知識生成の過程や洗練化の過程を通して思考能力を高めることが重要です。生成系AIツールでは、情報を収集・整理する作業を自動化し結果だけを表示します。**生成系AIツールで生成された文章をそのまま授業課題の回答とすれば、この貴重な思考過程の訓練の機会を逸することになり、長期的には当人の能力向上が損なわれます。**授業によって、利用禁止にしたり、利用に一定の条件を設定するのはこのためです。

[4][**不正行為**] 学術の世界だけでなく社会活動全般において、個々人のアイデアや独創性を尊重することが重要です。レポートや論文では、根拠となった出典を明記した上で、自分なりの考えを記載することが求められます。授業課題を提出する際に、**生成系AIツールが生成した文章等をそのまま自分の文章として用いることは認められません。**

...

# 調査：提出物はどれだけChatGPTの答案に近いか

## 調査方法

- ChatGPTが正答しつつ模範解答と書き方に有意な違いが見つかった5つの課題が対象
- 提出物が模範解答とChatGPT解答のどちらに近いかを調べる
  - ASTレベルの単純な剽窃検出器 [pycode\\_similar](#) を利用

## 調査結果

- いずれの課題でも完全一致は模範解答の方が多かった
  - 模範解答と大きく離れたChatGPT解答と一致する提出物は無かった
- 4/5で模範解答への類似度の方が全体的に高かった
  - 模範解答の穴あきプログラムをヒントとして与えているので、当然と言えば当然

# プログラミング入門とChatGPT

- 基本的にChatGPTで課題は解けて当たり前
  - 特にコーディングスタイルについては優れた模範を示す
- ChatGPTで解けない問題文については、記述を見直した方が良い
  - 初学者の読解力 << ChatGPTの読解力
  - 誤読を避ける問題文の工夫は、(人に対する)一種のプロンプトエンジニアリング
- 成績評価の方法を工夫して乗り切るのが現実的
  - 提出物の評価を粗くする(「Pythonプログラミング入門」の方針)
  - 理解確認用のペーパーテストを実施する