

Large Language Models et l'enseignement de la programmation

Diane Lingrand



28 Novembre 2023

Présentation Azur Tech Winter pour
UniCA par Diane Lingrand sous
Licence Creative Commons
Attribution-NonCommercial-
ShareAlike 4.0 International License.

The image shows the Creative Commons logo (CC) followed by three icons: a person (BY), a crossed-out dollar sign (NC), and a circular arrow (SA).

Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)

This is a human-readable summary of (and not a substitute for) the [license](#), [Disclaimer](#).

You are free to:

- Share** — copy and redistribute the material in any medium or format
- Adapt** — remix, transform, and build upon the material

[Share your work](#) | [Use & remix](#) | [What We](#)

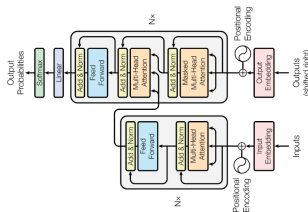
Under the following terms:

- Attribution** — You must give [appropriate credit](#), provide a link to the license, and [indicate if changes were made](#). You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
- NonCommercial** — You may not use the material for [commercial purposes](#).
- ShareAlike** — If you remix, transform, or build upon the material, you must distribute your contributions under the [same license](#) as the original.

No additional restrictions — You may not apply legal terms or [technological measures](#) that legally restrict others from doing anything the license permits.

Cette présentation n'est pas ...

... un cours sur les transformers, les LLMs ou le prompt engineering ...



[Attention Is All You Need, Neurips 2017]

- Pour cela, il faut bien plus que 30 minutes !
- Cependant, il existe plusieurs formations sur Sophia Antipolis où cela est enseigné pour les étudiants en informatique :
 - parcours IA-ID en dernière année du département Sciences Informatiques, Polytech 
 - les M2 DS4H du même parcours suivent les mêmes cours scientifiques
 - certains EIT suivent également ces cours
 - M2 DSAI (Data Science and Artificial Intelligence)
- Les entreprises peuvent s'adresser à EFELIA afin d'obtenir des formations sur mesure 

- Ecole Française de l'Intelligence Artificielle
 - Coordination nationale entre les 4 Instituts 3IA
 - Lauréat de l'appel ANR Compétences Métiers d'Avenir (CMA)



Direction du projet :

Charles Bouveyron, professeur en mathématiques appliquées, directeur du 3IA Côte d'Azur

Direction scientifique :

Lucile Sassatelli, professeure en informatique, directrice scientifique
Vincent Vandewalle, professeur en mathématiques appliquées, directeur scientifique adjoint

Direction exécutive :

Diana Sebbar, ingénieure de recherche
Université Côte d'Azur

Pilotage opérationnel :

Violette Assati, Cheffe de Projet
violette.assati@univ-cotedazur.fr

- Septembre 2023
- Plus de 80% ont déjà utilisé un LLM :
 - 28% par curiosité
 - 57% à titre personnel (lettre de motivation, traduction ...)
 - 14% pour les cours (langues, philosophie, mathématiques, programmation, résumé ...)
- “Pensez-vous que chatGPT puisse fournir des réponses complètement fausses ?”
 - oui : 73%
 - non : 9%
 - je ne sais pas : 17%
- “Avez-vous cherché à vérifier la réponse de chatGPT ?”
 - oui : 71%
 - non : 28%

.. est une base de reflexion pour l'utilisation des LLMs pour l'enseignement de la programmation au département Sciences Informatiques.

- Si les entrants en cycle préparatoire l'ont déjà en majorité utilisé, la pratique des étudiants en cycle ingénieur est de l'utiliser quotidiennement, notamment via copilot en programmation.
- Comment devons-nous réagir, collectivement ?
 - intérêts
 - risques
 - nécessité : modification de notre façon d'enseigner

AI-generated code presents both opportunities and challenges for students and educators in introductory programming and related courses. The sudden viability and ease of access to these tools suggest educators may be caught unaware or unprepared for the significant impact on education practice resulting from AI-generated code. We therefore urgently need to review our educational practices in the light of these new technologies.

from [Programming is hard - Or at Least It Used to Be: Educational Opportunities and Challenges of AI Code Generation, SIGCSE 2023]

... sont maintenant des professionnels de l'informatique.

The Impact of AI on Developer Productivity: Evidence from GitHub Copilot

Sida Peng,^{1*} Eirini Kalliamvakou,² Peter Cihon,² Mert Demirel³

¹Microsoft Research, 14820 NE 36th St, Redmond, USA

²GitHub Inc., 88 Colin P Kelly Jr St, San Francisco, USA

³MIT Sloan School of Management, 100 Main Street Cambridge, USA

*To whom correspondence should be addressed; E-mail: sidpeng@microsoft.com.

Abstract

Generative AI tools hold promise to increase human productivity. This paper presents results from a controlled experiment with GitHub Copilot, an AI pair programmer. Recruited software developers were asked to implement an HTTP server in JavaScript as quickly as possible. The treatment group, with access to the AI pair programmer, completed the task 55.8% faster than the control group. Observed heterogeneous effects show promise for AI pair programmers to help people transition into software development careers.

arxiv preprint 2023

Il faut donc apprendre à utiliser ces outils !

- Faciliter l'apprentissage de la programmation
 - exemple de solutions (fini l'angoisse de la page blanche)
 - variabilité des solutions
 - [The Robots Are Coming: Exploring the Implications of OpenAI Codex on Introductory Programming, ACE2022]
 - intérêt pédagogique de l'étude de différentes solutions, correctes ou moins correctes.
 - revue de code [A Review of Peer Code Review in Higher Education ACM Trans.Comp.Educ 2020]
 - possible explication de code ou de concept
 - explication des erreurs en langage naturel
 - plus explicite et moins décourageant que ‘‘Syntax error’’ ou ‘‘Core dump’’
- Production de ressources pédagogiques
 - [Automatic Generation of Programming Exercises and Code Explanations with Large Language Models, ICER 2022]
- Assistant
 - [The AI Revolution in Education: Will AI Replace or Assist Teachers in Higher Education? arxiv preprint] : ce papier conclue par : “the future of education lies in the synergy between human teachers and AI.”

- si on ne change rien dans l'enseignement :
 - ne plus apprendre
- fausses réponses
 - mauvais code
 - ne fait pas ce qui est attendu
 - non efficace
 - difficilement lisible par un humain
 - mauvaise explication
- plagiat
 - pour les copies des étudiants
 - examens sur machine, projets
 - problème du droit d'auteur
- disponibilité gratuite de ces outils pour les étudiants
- comme tout outil de ML : biais
- sécurité

- Expliciter les différentes compétences à acquérir en programmation aide à l'apprentissage [A theory of instruction for introductory programming skills, Journal of Computer Science Education, 2019]
- Les compétences à apprendre incrémentalement :
 - comprendre l'exécution pas-à-pas d'un programme
 - écrire de la syntaxe
 - compréhension de template
 - écrire un code selon un template
- Ce qui a été mesuré :
 - taux d'exercices réussis
 - taux d'erreurs
 - capacité à expliquer le code
 - engagement dans l'apprentissage de la programmation

Quelles compétences ?

- exposer un problème

Students can focus more on how to communicate algorithmic problems clearly, thereby providing a better description to code generation models that can then generate working solutions.

from [Programming is hard - SIGCSE 2023]

- décomposer un problème
- sélectionner un code parmi plusieurs propositions
- comprendre un code
 - pour savoir s'il est correct
 - ligne par ligne (copilot sait faire)
 - objectif globale ou par morceaux

- tester

- déboguer

People whose jobs still consist in writing will be supported, increasingly, by tools such as GPT-3. Forget the mere cut & paste, they will need to be good at prompt & collate.

[GPT-3: Its Nature, Scope, Limits, and Consequences, Mind and Machines 2020]

Promptly : Using Prompt Problems to Teach Learners How to Effectively Utilize AI Code Generators

arxiv preprint, 2023 : <https://arxiv.org/pdf/2307.16364.pdf>

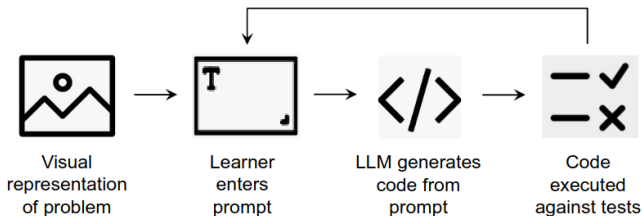


Figure 2: The PROMPTLY tool implements a simple variation of Prompt Problems in which the code generated by the LLM is automatically executed against a set of test cases and can not be edited directly. In order to modify the code, the learner is directed to edit the prompt.

Studying the effect of AI Code Generators on Supporting Novice Learners in Introductory Programming

Majeed Kazemitabaar
Department of Computer Science,
University of Toronto
Toronto, Ontario, Canada
majeed@dgp.toronto.edu

Barbara J. Ericson
School of Information, University of
Michigan
Ann Arbor, Michigan, USA
barbarer@umich.edu

Justin Chow
Department of Computer Science,
University of Toronto
Toronto, Ontario, Canada
justinchow@dgp.toronto.edu

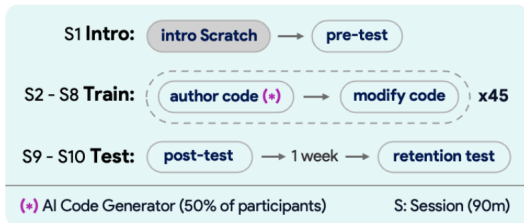
David Weintrop
College of Education, University of
Maryland
College Park, Maryland, USA
weintrop@umd.edu

Carl Ka To Ma
Department of Computer Science,
University of Toronto
Toronto, Ontario, Canada
cma@dgp.toronto.edu

Tovi Grossman
Department of Computer Science,
University of Toronto
Toronto, Ontario, Canada
tovi@dgp.toronto.edu

ACM CHI 2023

Protocole :



Expérience pédagogique - CS1 : types d'exercices

a) Scratch pre-test (Introduction)

What does this program say?



Ⓐ100 Ⓑ10 Ⓒ101 Ⓓ11

b) Authoring Task (Training)

Task Description:

Write a program that repeatedly generates a random number between 0 and 100 until the random number that it generates becomes equal to 50 (and then stop). Then display the number of attempts it took to generate the number.

Sample Output:

It took 27 attempts.

c) Modifying Task (Training)

Task Description:

Modify the program so that it stops when the random number becomes equal to any of the numbers 25, 50, or 75. It should also display which number it stopped on after displaying the number of attempts after the loop.

```
from random import randint
num = randint(0, 100)
count = 0
while num != 50:
    num = randint(0, 100)
    count += 1
print(str(count) + " attempts.")
```

d) Multiple-Choice Task (Evaluation)

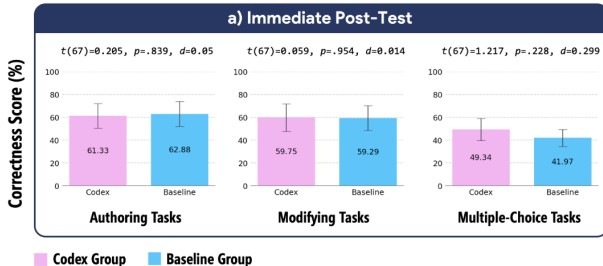
What does this program display?

```
num = 0
while 25 < num:
    if num == 5:
        num = num + 1
        break
    elif num == 20:
        num = num - 3
        break
    num = num + 1
print(num)
```

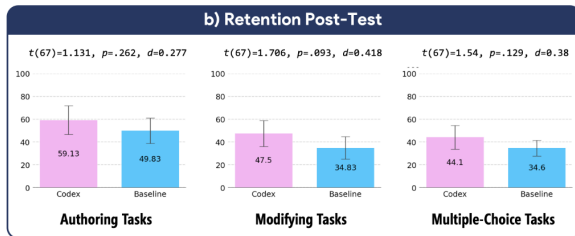
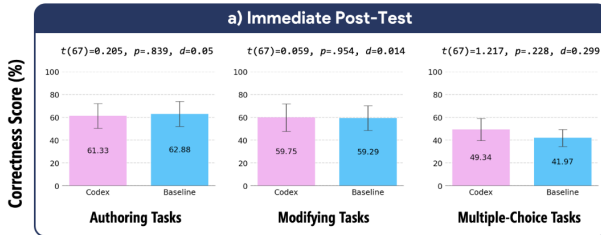
Ⓐ6 Ⓑ17 Ⓒ0 Ⓓ1

[Studying the effect of AI Code Generators on Supporting Novice Learners in Introductory Programming, ACM CHI 2023]

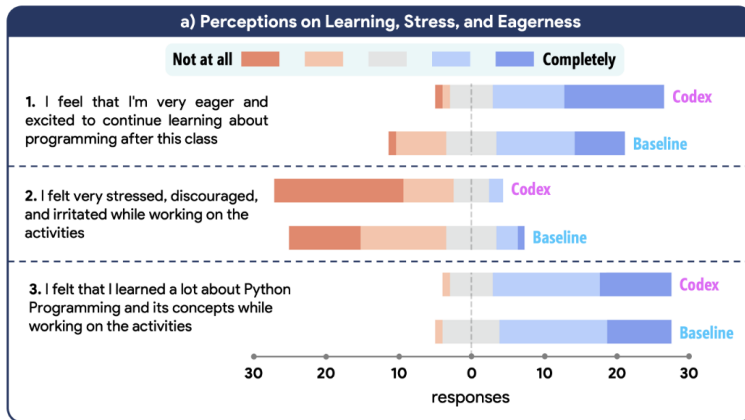
Expérience pédagogique - CS1 : test final



Expérience pédagogique - CS1 : test final

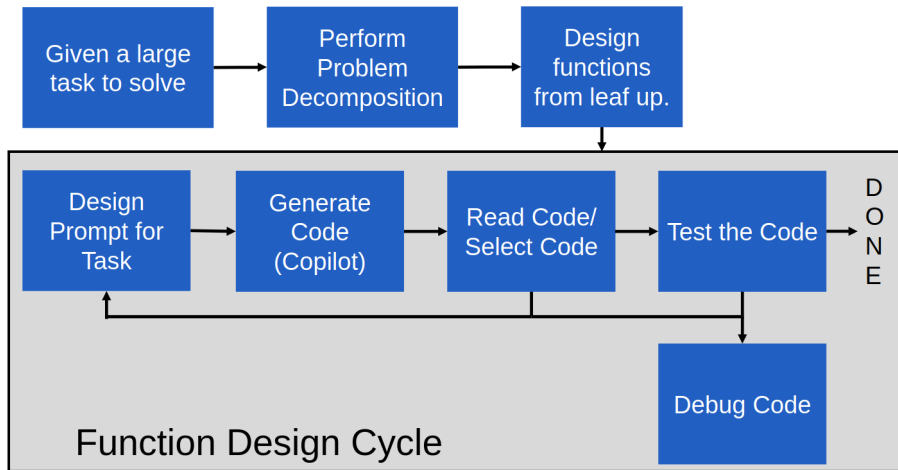


[Studying the effect of AI Code Generators on Supporting Novice Learners in Introductory Programming, ACM CHI 2023]



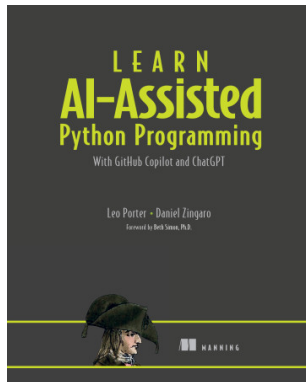
[Studying the effect of AI Code Generators on Supporting Novice Learners in Introductory Programming, ACM CHI 2023]

Ecriture d'un programme avec LLM



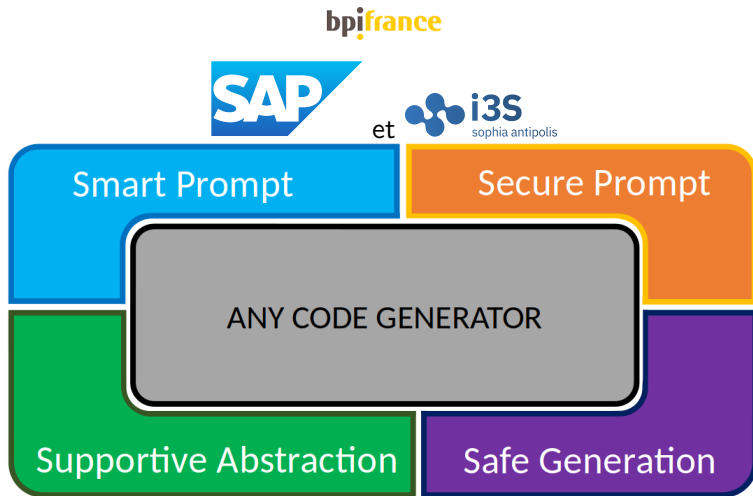
source : séminaire de Porter et Zingaro, 2023

Un livre sort aujourd'hui



Dans la suite du webinaire ACM :

<https://www.youtube.com/watch?v=FfGZBRWxVAM>



[OWASP Top 10 for LLM Applications, 2023]

La suite de l'histoire ?



Soyez les bienvenus à la journée nationale organisée dans le cadre du GDR GPL à Sophia sur **les LLMs et le développement logiciel**.

La date est entre le 30 janvier et 1er février 2024 (date à déterminer).

Les informations seront disponibles sur le site du GdR GPL dans les prochaines jours.