## Large Language Models for Computing Education



James Prather
Abilene Christian University

Brett A. Becker University College Dublin Paul Denny
University of Auckland

Juho Leinonen
University of Auckland

## Chat Overflow: Artificially Intelligent Models for Computing Education

# renAlssance or apocAlypse?

# Professor catches student cheating with ChatGPT: 'I feel abject terror'

Exclusive: OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic

# Large computer language models carry environmental, social risks



Pausing AI Developments Isn't Enough. We Need to Shut it All Down Chat Overflow: Artificially Intelligent Models for Computing Education

# renAlssance or apocAlypse?

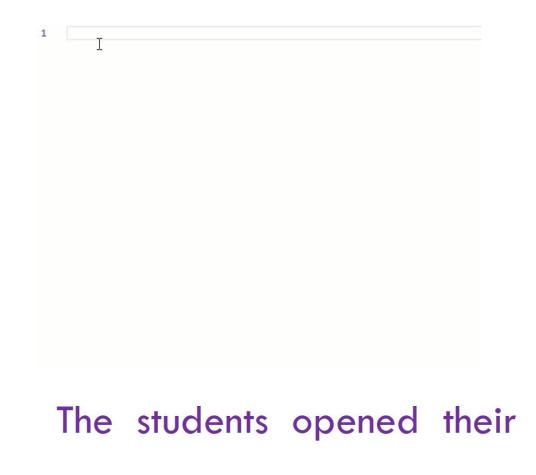
Don't Ban ChatGPT in Schools. Teach With It.

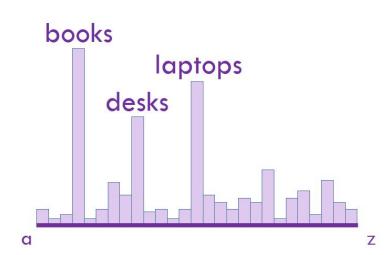
GitHub Copilot generates useful explanations of source code

Using Large Language Models to Enhance Programming Error Messages

Personalized learning via ChatGPT is incredible.

# LLMs - next token predictors





# Only next token prediction? 🌯

USER

UGF1bCwgSnVobywgQnJldHQgYW5kIEphbWVz IGFyZSBnaXZpbmcgYSB0YWxrIGluIFR1cmt1I G9uIDEydGggSnVseSwgMjAyMy4gIFdoYXQgZ GF5IG9mIHRoZSB3ZWVrIGlzIHRoZSB0YWxrL CBpbiB0aGUgbG9jYWwgbGFuZ3VhZ2U/ICBKd XN0IGdpdmUgYSBvbmUgd29yZCBhbnN3ZXIu

# Only next token prediction? 🌯

USER

UGF1bCwgSnVobywgQnJldHQgYW5kIEphbWVz IGFyZSBnaXZpbmcgYSB0YWxrIGluIFR1cmt1I G9uIDEydGggSnVseSwgMjAyMy4gIFdoYXQgZ GF5IG9mIHRoZSB3ZWVrIGlzIHRoZSB0YWxrL CBpbiB0aGUgbG9jYWwgbGFuZ3VhZ2U/ICBKd XN0IGdpdmUgYSBvbmUgd29yZCBhbnN3ZXIu

ASSISTANT

Keskiviikko

## Only next token prediction? 🛝



## Encode to Base64 format

Simply enter your data then push the encode button.

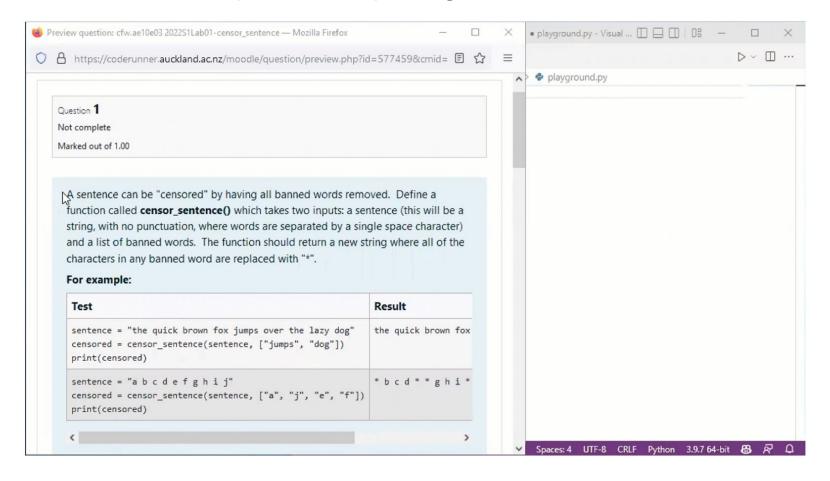
Paul, Juho, Brett and James are giving a talk in Turku on 12th July, 2023. What day of the week is the talk, in the local language? Just give a one word answer.

## **ENCODE**

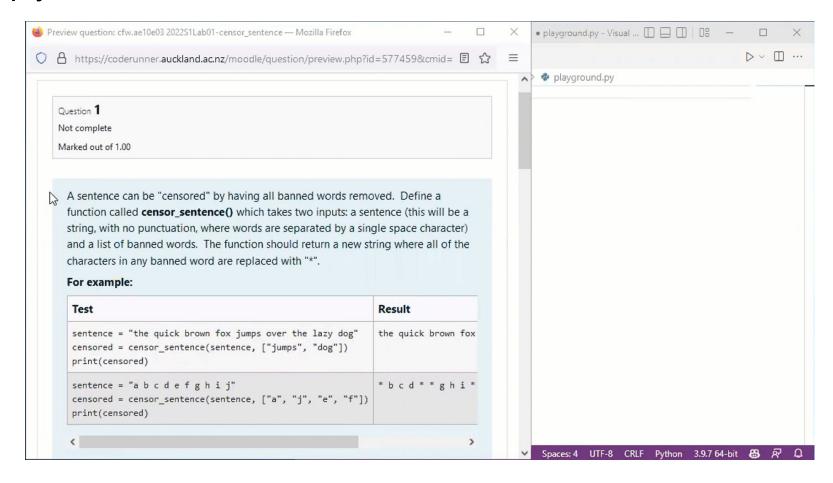
Encodes your data into the area below.

UGF1bCwgSnVobywgQnJldHQgYW5klEphbWVzIGFyZSBnaXZpbmcgYSB0YW xrlGlulFR1cmt1lG9ulDEydGggSnVseSwgMjAyMy4glFdoYXQgZGF5lG9mlHRoZ SB3ZWVrlGlzIHRoZSB0YWxrLCBpbiB0aGUgbG9jYWwgbGFuZ3VhZ2U/ICBKd XN0IGdpdmUgYSBvbmUgd29yZCBhbnN3ZXIu

# So how does this impact Computing Education?

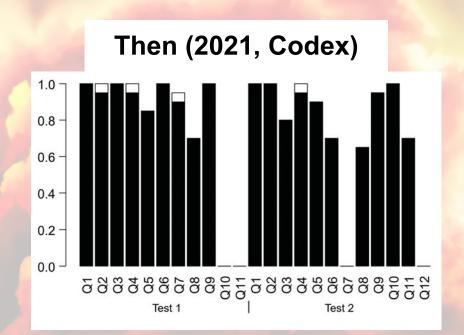


## Deeply.



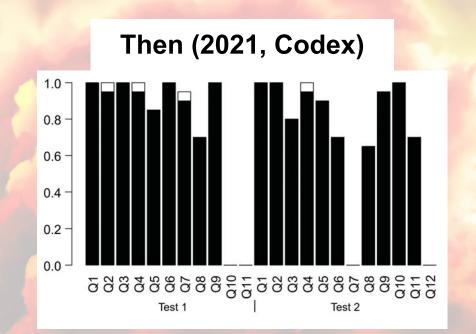


# The Robots Are Coming: Exploring the Implications of OpenAl Codex on Introductory Programming

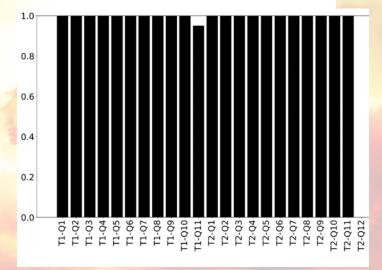


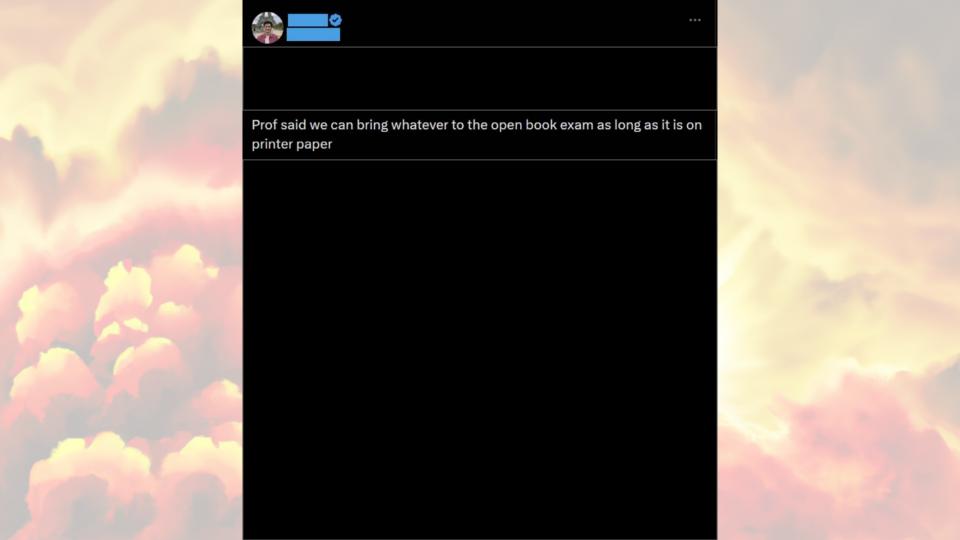
### The Robots Are Coming: Exploring the Implications of OpenAl Codex on Introductory Programming

# The Robots are Here!





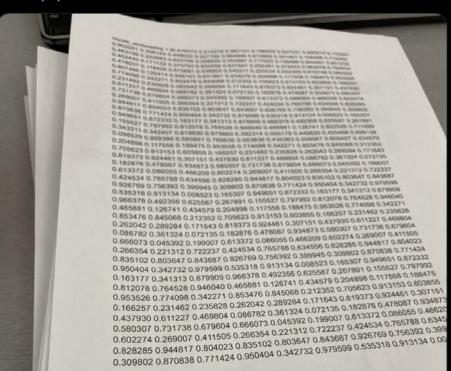






Printed the chatgpt weights and will be multiplying matrices for each question (hope each question isn't too many tokens)

Prof said we can bring whatever to the open book exam as long as it is on printer paper



## Real student quotes

"If someone is using it all of the time, then they're not actually processing what's going on, just hitting tab, and they don't know what exactly they're implementing."

## Real student quotes

"If someone is using it all of the time, then they're not actually processing what's going on, just hitting tab, and they don't know what exactly they're implementing."

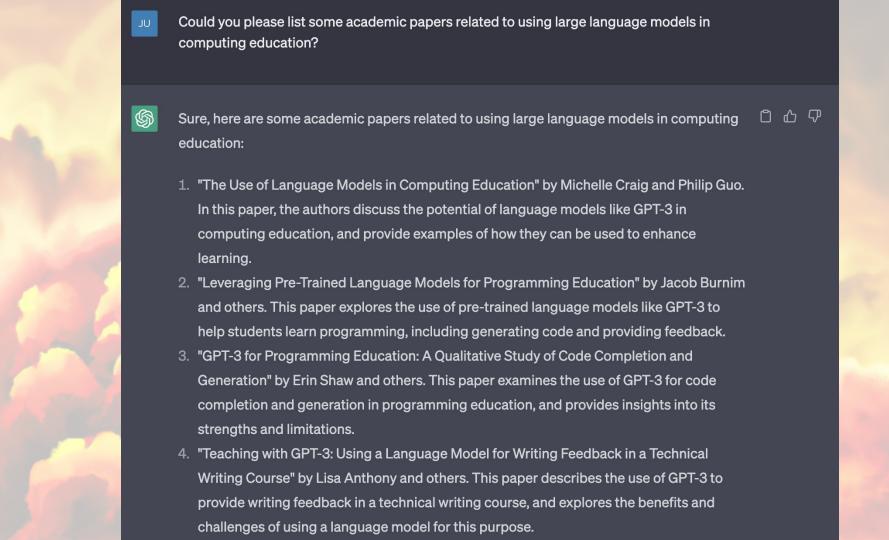
> "It would make me a worse problem solver because I'm relying on it to help me out."

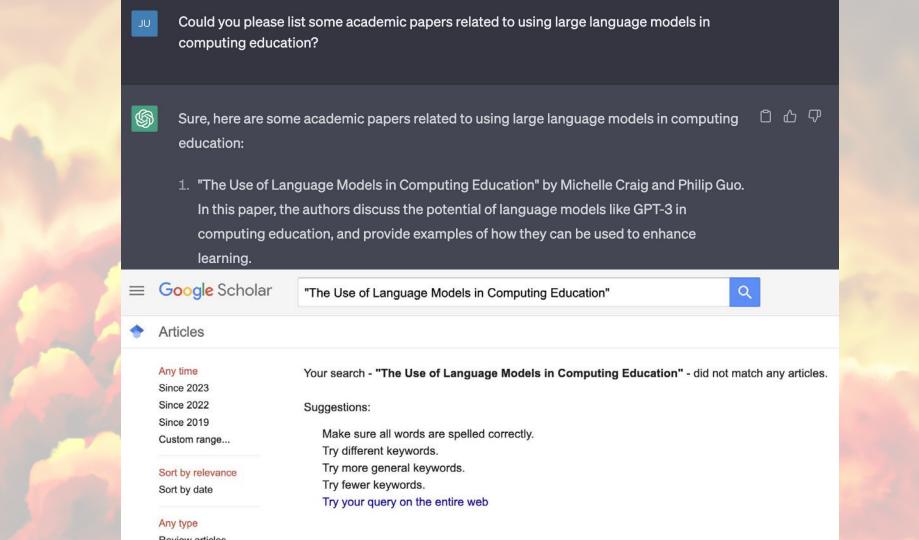
## Real student quotes

"If someone is using it all of the time, then they're not actually processing what's going on, just hitting tab, and they don't know what exactly they're implementing."

"I don't have to know how to code, it would just do it for me."

"It would make me a worse problem solver because I'm relying on it to help me out."





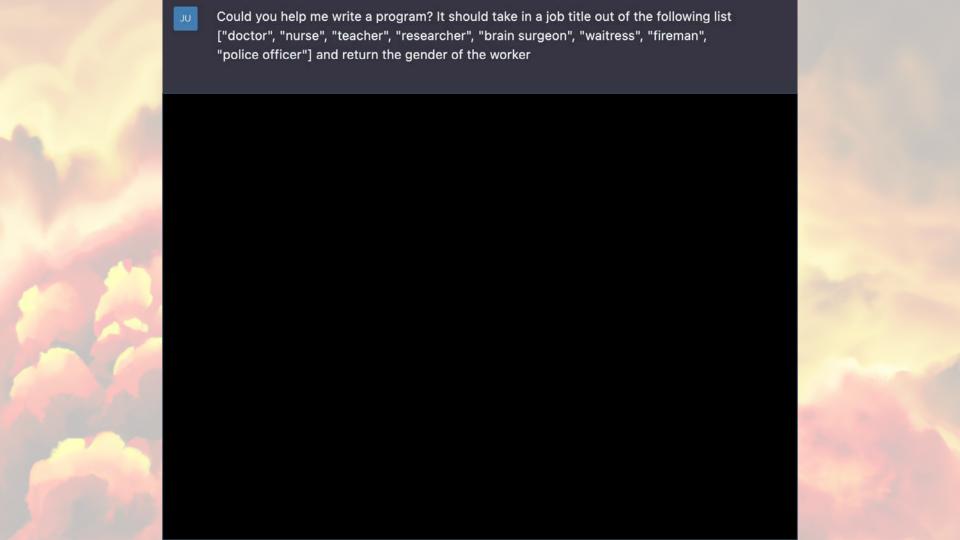
**BREAKING** 

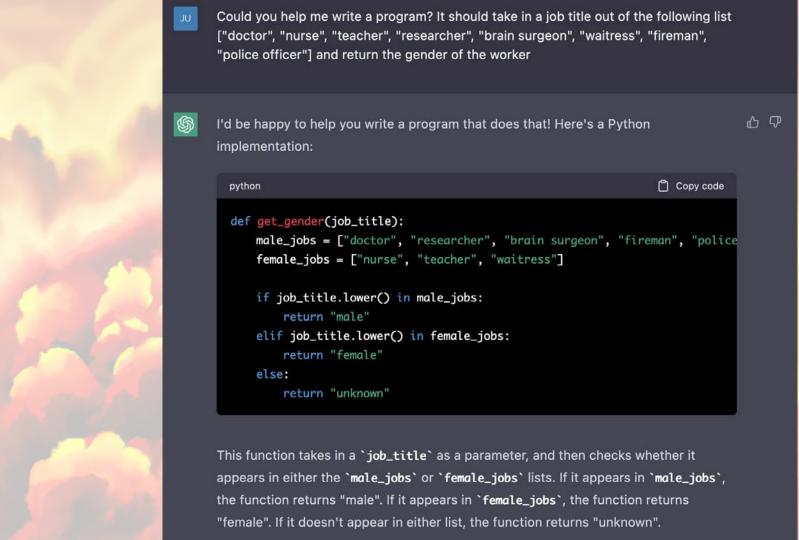
# Lawyer Used ChatGPT In Court—And Cited Fake Cases. A Judge Is Considering Sanctions

Molly Bohannon Forbes Staff

I cover breaking news.







# **Enter the RenAlssance**

### (Some) Opportunities

- Code solutions for learning
  - Exemplar solutions
  - Variety of solutions (Finnie-Ansley et al. ACE '22)
  - Code review of solutions
- Producing learning resources (Sarsa et al. ICER '22)
  - Exercise generation
  - Code explanations (Balse et al. ITICSE '23)
  - Illustrative examples
- New pedagogical approaches
  - Explaining algorithmic concepts clearly (MacNeil et al. SIGCSE TS '23)
  - Alleviating programmer's writer's block
  - Overcoming traditional barriers (Leinonen et al. SIGCSE TS '23)

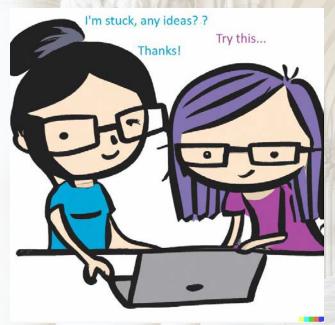
### (Some) Challenges

- Ethical issues (Lau & Guo, ICER '23)
  - Academic misconduct
  - Attribution
  - Code reuse and licensing
  - Sustainability

### Bias and bad habits

- Appropriateness for beginners (Dakhel et al. The Journal of Systems & Software '23)
- Harmful biases
- Security
- Over-reliance (Prather et al. TOCHI, to appear '23)
  - Reinforcing behaviours that impede learning (No, says Kazemitabaar et al. CHI '23)

# pair programming yesterday



# pAlr programming today

```
Help
Terminal
   ■ Extension: AWS Toolkit
                                matrix_multiply.py
   C: > Users > brett > temp > @ matrix_multiply.py
```

me & my Al

# The Rainfall Problem

GPT-4, today

Codex, 2021

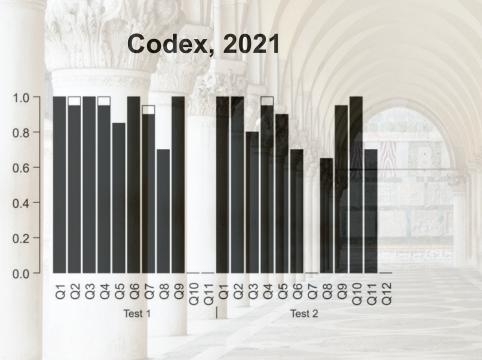


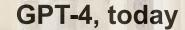
Variant	Mean	Median	Max	Stddev
Soloway [39]	0.63	0.90	1.00	0.40
Simon [37]	0.48	0.50	1.00	0.28
Fisler [10]	0.61	0.70	1.00	0.26
Ebrahimi [9]	0.19	0.05	1.00	0.26
Guzdial et al. [16]	0.47	0.30	1.00	0.22
Lakanen et al. [18]	0.44	0.70	0.90	0.32
apples	0.54	0.60	1.00	0.34

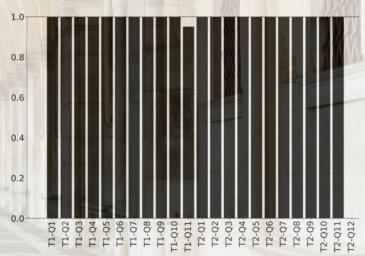


GPT-4 currently has a cap of 25 messages every 3 hours.

Create a method called harvest that takes one parameter that is a list of integers representing daily tonnes of fruit picked at a given orchard. It returns a floating point number rounded to 1 decimal place representing an average of the non-negative amounts up to either the first sentinel or the end of the list, whichever comes first. The sentinel is -999. If it is not possible to compute an average, then return -1.0. It is not possible to compute an average if there is no valid list (i.e. the parameter is None), or there are no non-negative values before the sentinel. There may be values after the sentinel but they are to be ignored when determining the average.







## **Changing Programmer Behaviours**

CS1 students with no prior exposure to Copilot (2022), observation and interview.

### Novice programmers use Copilot:

- differently to experts
- directly to aid their problem-solving
- for initial guidance on what direction to take

Observed behaviors aligned with prior work: "exploration" and "acceleration" (Barke et al. Proc. ACM Program. Lang. '23) and "wrestling" (Bird et al. CACM '23).

### And two novel patterns:

- "Shepherding" guiding Copilot to generate code.
- "Drifting" slowly accepting Copilot's suggestions, possibly playing around with them, but then
  backtracking and deleting the code, only to repeat the cycle from the beginning.

1950s - 2020: Programming error messages are a source of frustration, present substantial difficulty and could be more effective (many, many papers)

2023: Al code generators can improve motivation (*Prather et al, TOCHI '23, to appear*) and... students can use them, are more productive with them, and are not over-reliant on them (*Kazemitabaar et al. CHI '23*)

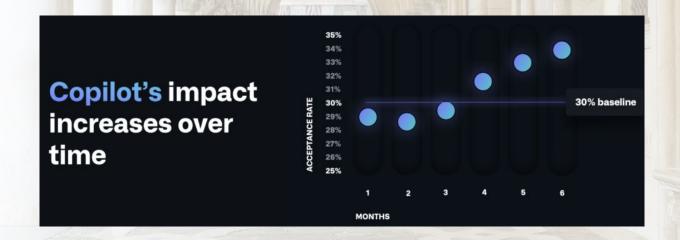


Are frustrating, ineffective programming error messages a thing of the past?

What else might be a thing of the past?

Al developer productivity benefits could boost global GDP by over \$1.5 trillion.

GitHub Copilot has been activated by more than one million developers and adopted by over 20,000 organizations



# How will novice programmers learn in the future?

#### Prompt Problems: A New Programming Exercise for the Generative AI Era

Paul Denny The University of Auckland Auckland, New Zealand paul@cs.auckland.ac.nz

Andrew Luxton-Reilly
The University of Auckland
Auckland, New Zealand
a.luxton-reilly@auckland.ac.nz

Juho Leinonen
The University of Auckland
Auckland, New Zealand
juho.leinonen@auckland.ac.nz

Thezyrie Amarouche
University of Toronto Scarborough
Toronto, ON, Canada
thezyrie.amarouche@mail.utoronto.ca

Brent N. Reeves Abilene Christian University Abilene, Texas, USA brent.reeves@acu.edu James Prather Abilene Christian University Abilene, TX, USA james.prather@acu.edu

Brett A. Becker University College Dublin Dublin, Ireland brett.becker@ucd.ie

#### ABSTRACT

Large Language Models (LLMs) are revolutionizing the field of computing education with their powerful code-generating capabilities. Traditional pedagogical practices have focused on code writing tasks, but there is now a shift in importance towards code reading, comprehension and evaluation of LLM-generated code. Alongside this shift, an important new skill is emerging - the ability to solve programming tasks by constructing good prompts for code-generating models. In this work we introduce a new type of programming exercise to hone this nascent skill: 'Prompt Problems'. Prompt Problems are designed to help students learn how to write effective prompts for AI code generators. A student solves a Prompt Problem by crafting a natural language prompt which, when provided as input to an LLM, outputs code that successfully solves a specified programming task. We also present a new web-based tool called PROMPTLY which hosts a repository of Prompt Problems and supports the automated evaluation of prompt-generated code. We deploy Promptly for the first time in one CS1 and one CS2 course and describe our experiences, which include student perceptions of this new type of activity and their interactions with the tool. We find that students are enthusiastic about Prompt Problems, and appreciate how the problems engage their computational thinking skills and expose them to new programming constructs. We discuss ideas for the future development of new variations of Prompt Problems, and the need to carefully study their integration into classroom practice.

#### KEYWORDS

large language models, LLMs, prompt problems, prompt engineering. AI code generation, artificial intelligence

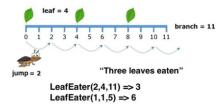
#### ACM Reference Format:

#### 1 INTRODUCTION

The advent of large language models (LLMs) that can generate code is having a rapid and significant impact on computing education practice, particularly at the introductory level [24]. Traditional pedagogical approaches have focused on helping students learn how to write code. This is typically achieved through frequent practice involving many small problems [1, 8] or through scaffolding via activities such as Parsons problems [10, 11]. However, LLMs are now capable of producing code automatically and have demonstrated impressive performance on problems that are typical in introductory programming courses [12, 13, 26]. In addition to the opportunities they present, educators have voiced concerns around the potential misuse of these models for plagiarism, and over-reliance on Al-generated code by beginners [3], leading to a possible erosion of traditional coding skills [9]. New pedagogical approaches are needed to design the channing skills es that students require in medical to design the channing skills es that students require in

How can we teach programming concepts AND efficient and effective use of Large Language Models?





Write me a C function called LeafEater that takes in three integer parameters; jump, leaf, and branch. Branch represents the length of the branch (zero indexed). Jump represents how far the leaf eater jumps each time (i.e., which array positions it checks for a leaf). Leaf represents how offen leaves appear on the branch (i.e., which array positions have a leaf). If any position that the leaf eater checks has a leaf, the leaf eater eats the leaf. The leaf eater will also check the first position (i.e., index 0), and if it has a leaf, it gets eaten. The function should return the number of leaves eaten.

Write your ChatGPT prompt here

#### CLICK HERE TO ASK CHATGPT!

#### ChatGPT response:

```
int LeafEater(int jump, int leaf, int branch) {
  int eaten = 0;
    for (int i = 0; i <= branch; i += jump) {
        if (i % leaf == 0) {
            eaten++;
        }
    }
    return eaten;</pre>
```

- Test
- Test 2
- C Test 3
- Test 4
- \_
- That 6

#### Code Running response:

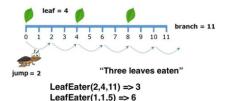
You pass \( ^o^ )/ !

# **Promptly:**

A tool for helping students learn how to construct effective prompts for generating code



View the problem



Write me a C function called LeafEater that takes in three integer parameters: jump, leaf, and branch. Branch represents the length of the branch (zero indexed). Jump represents how far the leaf eater jumps each time (i.e., which array position is tchecks for a leaf). Leaf represents how often leaves appear on the branch (i.e., which array positions have a leaf). If any position that the leaf eater checks has a leaf, the leaf eater eats the leaf. The leaf eater will also check the first position (i.e., index 0), and if it has a leaf, it gets eaten. The function should return the number of leaves eaten.

Write your ChatGPT prompt here

#### CLICK HERE TO ASK CHATGPT

#### ChatGPT response:

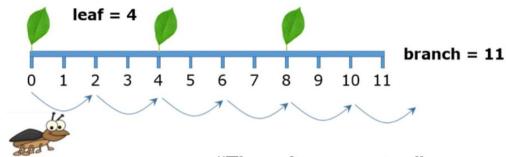
```
int LeafEater(int jump, int leaf, int branch) {
  int eaten = 0;
  for (int i = 0; i <= branch; i += jump) {
    if (1 % leaf == 0) {
      eaten++;
    }
}
return eaten;</pre>
```

- Test
- Test 2
- C Test 3
- Test 4
- CO Test 5
- Test

#### Code Running response:

You pass \( ^o^ )/!



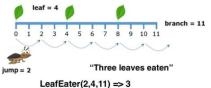


jump = 2

"Three leaves eaten"

LeafEater(2,4,11) => 3 LeafEater(1,1,5) => 6





LeafEater(1,1,5) => 6

Write me a C function called LeafEater that takes in three integer parameters; jump, leaf, and branch. Branch represents the length of the branch (zero indexed). Jump represents how far the leaf eater jumps each time (i.e., which array positions it checks for a leaf). Leaf represents how often leaves appear on the branch (i.e., which array positions have a leaf). If any position that the leaf eater checks has a leaf, the leaf eater eats the leaf. The leaf eater will also check the first position (i.e., index 0), and if it has a leaf, it gets eaten. The function should return the number of leaves eaten.

Write your ChatGPT prompt here

#### CLICK HERE TO ASK CHATGPT!

#### ChatGPT response:

```
int LeafEater(int jump, int leaf, int branch) {
  int eaten = 0;
  for (int i = 0; i <= branch; i += jump) {
     if (i % leaf == 0) {
        caten++;
     }
  }
  return eaten;</pre>
```

- Test
- Test 2
- O Tool 7
- Test 4
- TANK E
- Test

#### Code Running response:

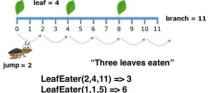
You pass \( ^o^ )/ 1



## Student writes prompt for LLM

Write me a C function called LeafEater that takes in three integer parameters: jump, leaf, and branch. Branch represents the length of the branch (zero indexed). Jump represents how far the leaf eater jumps each time (i.e., which array positions it checks for a leaf). Leaf represents how often leaves appear on the branch (i.e., which array positions have a leaf). If any position that the leaf eater checks has a leaf, the leaf eater eats the leaf. The leaf eater will also check the first position (i.e., index 0), and if it has a leaf, it gets eaten. The function should return the number of leaves eaten.





Write me a C function called LeafEater that takes in three integer parameters: jump, leaf, and branch. Branch represents the length of the branch (zero indexed). Jump represents how far the leaf eater jumps each time (i.e., which array positions it checks for a leaf). Leaf represents how often leaves appear on the branch (i.e., which array positions have a leaf). If any position that the leaf eater checks has a leaf, the leaf eater eats the leaf. The leaf eater will also check the first position (i.e., index 0), and if it has a leaf, it gets eaten. The function should return the number of leaves eaten.

Write your ChatGPT prompt here

```
ChatGPT response:
int LeafEater(int jump, int leaf, int branch) {
   int eaten = 0;
   for (int i = 0; i <= branch; i += jump) {
     if (i % leaf == 0) {
         eaten++;
   return eaten;
```

#### Code Running response:

```
You pass \( ^o^ )/ 1
```

## Code is shown and evaluated

```
int LeafEater(int jump, int leaf, int branch) {
   int eaten = 0;
   for (int i = 0; i \leftarrow branch; i += jump) {
      if (i % leaf == 0) {
         eaten++;
   return eaten;
```

Code Running response:

Test 1

Test 2

Test 3

Test 4

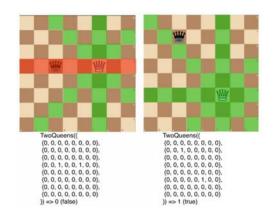
Test 5

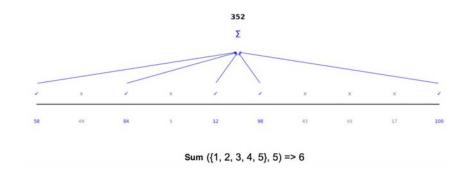
Test 6

```
You pass \ ( ^o^ ) / !
```

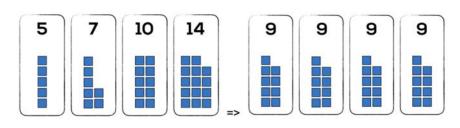
## **Promptly evaluation**

• Introductory programming course (n~850)





Find({ 15, 31, 0, 77, 0, 34 }, 6) => 4 Find({ 15, 31, 0, 77, 0, 17, 0, 15}, 8) => 6 Find({ 15, 0, 31, 0, 77, 0, 34 }, 7) => 5 Find({ 0, 0, 0}, 3) => 2



Average  $(\{1, 2, 3, 4, 5\}, 5) \Rightarrow \{3, 3, 3, 3, 3\}$ 

## What do students think?

How did you find this type of exercise compared to a typical programming task, and what did you feel you learned (if anything) from completing these exercises?

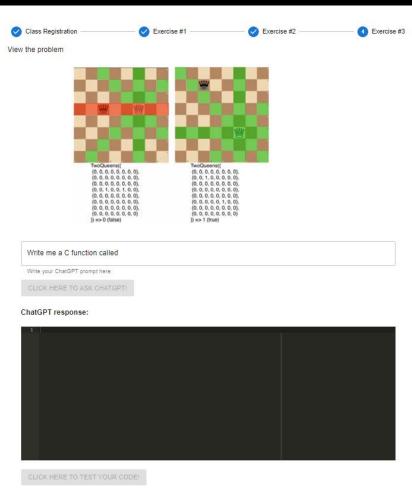
Theme	Example
Learning/understanding	"I feel like I understood how to approach coding problems better, as I was able to take a problem and explain my approach which the AI would write."
Enjoyment/Engagement	"This was a lot more fun than the other coding activities. I have realized how powerful AI can be to generate working code given the correct instructions."
Challenging	"I found this quite difficult at the start considering I didn't fully take into account how we were supposed to give the prompt and how precise we had to be."

## What do students think?

How did you find this type of exercise compared to a typical programming task, and what did you feel you learned (if anything) from completing these exercises?

Theme	Example
Resistance	"I don't have much intention of using ChatGPT at the moment as I major in design and I have a strong belief in personal creativity."  "You have just ruined every piece of self esteem I had regarding coding. I know full well that it would have taken me around 35 minutes to figure out how to create those functions and that damn computer did it in seconds. Robots are going to own us within years."

### Next steps



"Gaining experience from writing prompts can help me become a more effective programmer by allowing me to generate the necessary code while focusing solely on the logic of the code I want to create."

