

Throwing Shaders at Language Models - Evaluating Creative Code Generation

by Jan Kels @ TaCoS
29th June 2023 11:30 in 23.21.00.46

About me



- **Jan.Kels@hhu.de**
- 8th semester CL at HHU
- Interested in LLMs
- English/German
- 🤗 @Vipitis

Outline for this talk

- **A:** background
 - “creative” code?
 - code generation
 - evaluation methods
- **B:** my work so far
 - a Dataset
 - an evaluation task
 - it’s results
 - the models
- **C:** things to come
 - improvements
 - a Bachelor Thesis?
- Questions/Discussion
- **[redacted]**
- Backup slides

A: some background

So we all understand what this is about

A1: what do we mean by creative code?

- “[...]a type of computer programming in which the goal is to create something *expressive* instead of something *functional*.”
- GLSL/WebGL: evaluate the color of every pixel, 60 times a second
- Some examples to show it best ([backup 0](#))
 - <https://www.shadertoy.com/view/fstyD4> (beautiful 2D example) ([backup 1](#))
 - <https://www.shadertoy.com/view/3lsSzf> (3D stuff) ([backup 2](#))
 - <https://www.shadertoy.com/view/DdlcRf> (my own creation) ([backup 3](#))
- Very expressive, but requires some skill
- This website becomes the dataset we will see later

A2: what exactly is Code Generation?

- Decoder only models (GPT)
- learn the probability of the next token
- autoregressively generate more and more (when do they stop?)
- Code specific models exist: Codex(GPT3.5), santacoder, starcoder, InCoder, ...
- Code completion is more accurate to what we are interested in

'print("Hello World'		
!")		37.33%
")		23.18%
!		6.03%
,		5.34%
!\		4.38%

A3: How code generation is evaluated

According to Evtikhiev et al.(2023) there is 3 main kind:

- Metrics from machine translation: ROUGE-L, ChrF, ...
- Metrics for code snippet comparison: CodeBLEU, RUBY,
- Metrics that run and unit tests: HumanEval, APPS, MBPP, ... (*pass@k*)

```
def concatenate(strings):  
    """  
    Concatenate list of strings into a single string  
    """  
    <model generation here>  
    1. return " ".join(strings)  
    2. return "".join(strings)  
    3. out = "".join(strings)  
       return out
```

- There now is an evaluation harness

<https://github.com/bigcode-project/bigcode-evaluation-harness>

B: my project

From Deep Learning class

B1: Datasets you can use right now

- “Shadertoys” ~ 44k passes with code, title, author, description, source url,
<https://hf.co/datasets/Vipitis/Shadertoys>
< 14000 downloads - but I don’t know what for.
- “Shadertoys-fine” ~ 270k functions,
<https://hf.co/datasets/Vipitis/Shadertoys-fine>
- Common train/test
- No deduplication
- License tagging is an issue - as long as you care

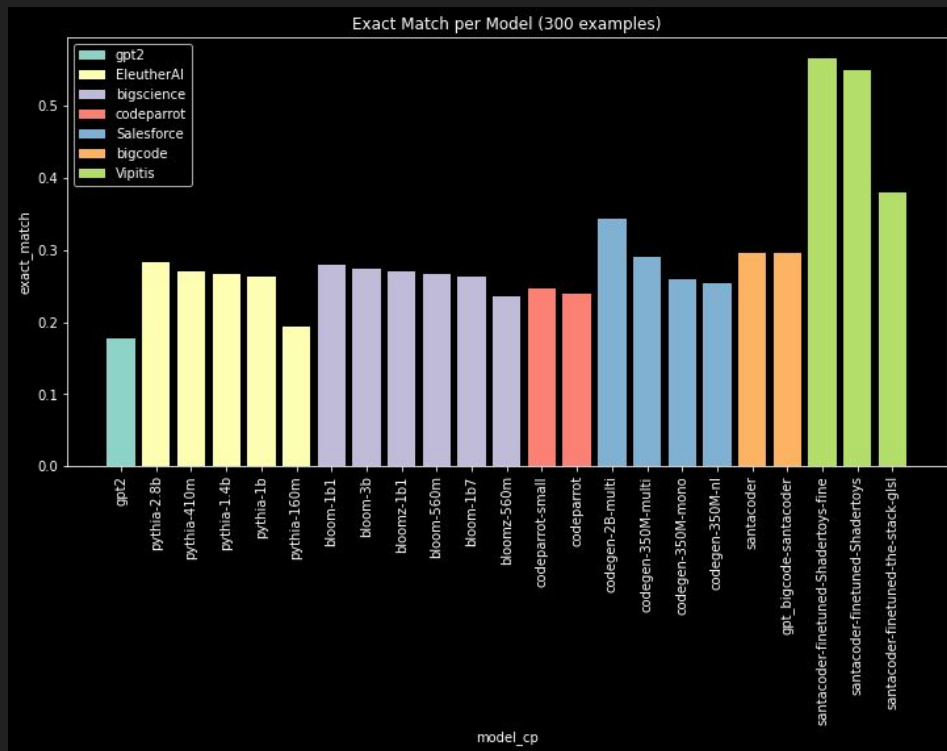
B2: a Return Completion task

```
float noise(vec3 p) //Thx to Las^Mercury
{
    vec3 i = floor(p);
    vec4 a = dot(i, vec3(1., 57., 21.)) + vec4(0., 57., 21., 78.);
    vec3 f = cos((p-i)*acos(-1.))*(-.5)+.5;
    a = mix(sin(cos(a)*a),sin(cos(1.+a)*(1.+a)), f.x);
    a.xy = mix(a.xz, a.yw, f.y);
    return <model generation here>;
}
```

- Greedy decoding
- exact_match
- <https://hf.co/spaces/Vipitis/ShaderEval>
- gpt2 mix
- bloom-560m/codeparrot a
- santacoder-finetuned mix(a.x, a.y, f.z)
- StarCoder sin(dot(a, vec2(12.9898, 78.233)))*43758.5453
- StarCoderPlus mix(mix(a.x, a.y, f.z), a.z, f.w)
- Copilot chat dot(a, vec4(1.0/vec3(1.0, 57.0, 113.0), 0.0))

B2: the results

- “Proof of concept”
- Code models do better
- Fine tuned models overfit
- Larger models do better
- A bunch of issues with this task basis for “further research”



B3: fine tuned models on Shadercode

Santacoder (1B) finetuned for 1000 steps (~8hours) on three different datasets

- The Stack dedup, GLSL subset
- Shadertoys
- Shadertoys-fine

All available on <https://hf.co/Vipitis>

Work in progress demo space <https://hf.co/spaces/Vipitis/ShaderCoder>

Models might be horribly overfitted, don't rely on them directly

C: future plans

Your feedback needed

C1: expand on the idea (ideas)

- Improve the dataset: deduplication, function parsing, without comments, licenses, thumbnails
- Expand the dataset: TwiGL, scraping, (any more?)
- Improve metrics for ReturnCompletion task: CodeBLEU, (more metrics?)
- Adding more tasks to become an evaluation suite:
 - Semantic match (same image generated)
 - CLIP match with title
 - further ideas?
- Provide them all in the BigCode Evaluation harness

C2: how to bachelor thesis?

- Any supervisors here that are interested?
- Relaxed timeframe: done in March 2024
- Literature recommendation?
- Formulate a “research question”!
- Present the work at a conference (any suggestions?)
- Is this even related to (formal) linguistics?
- Writing 30 000 words about it?
- Funding for hardware?

THANK YOU!

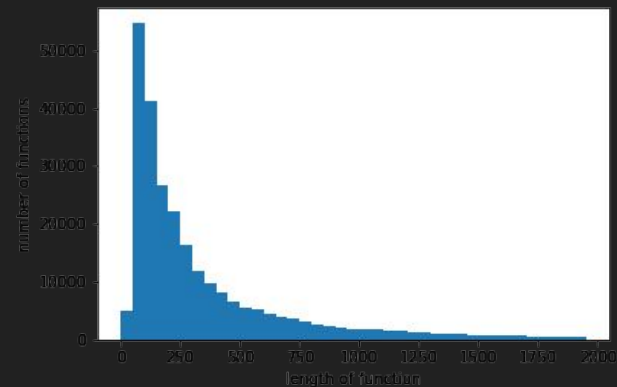
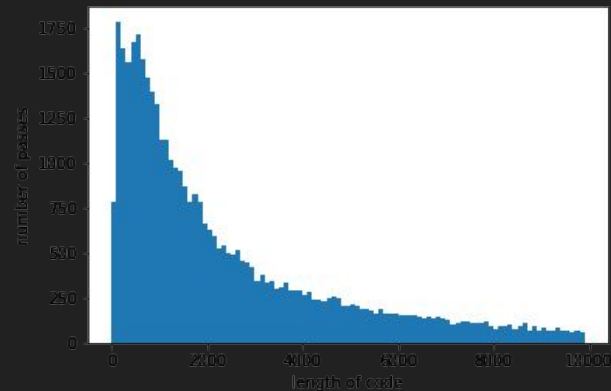
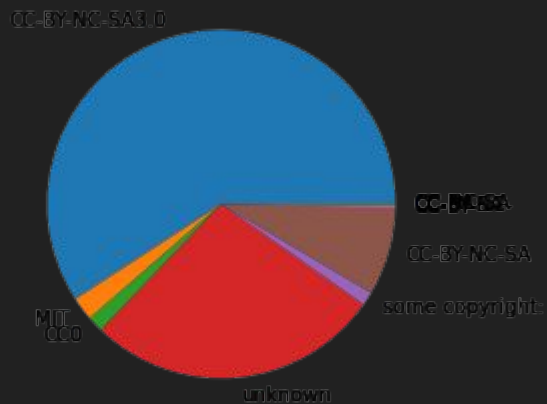
Any questions?

References/Sources

- [https://commons.wikimedia.org/wiki/File:Michelangelo_-_Creation_of_Adam_\(cropped\).jpg](https://commons.wikimedia.org/wiki/File:Michelangelo_-_Creation_of_Adam_(cropped).jpg)
- https://en.wikipedia.org/wiki/Creative_coding
- Evtikhiev, M., Bogomolov, E., Sokolov, Y., & Bryksin, T. (2023). Out of the bleu: how should we assess quality of the code generation models?. Journal of Systems and Software, 203, 111741.
- https://huggingface.co/datasets/openai_humaneval/viewer/openai_humaneval/test?row=28
- <https://www.shadertoy.com/view/ttlfWM>

BONUS SLIDE (1/n)

Some metrics on the dataset:



BONUS SLIDE (2/n)

Shader Showdown video clip, ideas to turn this into a democratic exhibition with audience input/music guidance.



BONUS SLIDE (3/n)

How bad my time management was for the project so far

Deep Learning in NLP Ideen

1. (kein NLP) reinforcement learning um Mirror's Edge iOS zu spielen (overfitting Problem?) An Speedruns angelegt. Framework via Visual Studio und Hyper V auf dem Laptop müsste gelöst werden für TAS
2. Transformer für Shadertoy, angelegt an Copilot und so. Es soll wohl 50k shadertoys geben, kann man die als Korpus bekommen? Erstes Ziel Syntax der funktioniert.
3. betting related?
4. Neurolinguistische Methoden auf Transformer anwenden (könnte ne super Präsentation werden), ich habe aber keinen Kurs Neurolinguistik gemacht ... Sowas wie halbe Eingabe Vektoren löschen und Schlaganfall zu simulieren. Man bräuchte pretrained Models wie GPT-2/3. Eventuell was für später. "Drop out?"
5. Code golfing (mit python), sowas wie Text Zusammenfassung, aber eher mit harten regeln.
6. Meine eigenen Schulunterlagen als Korpus. Habe keinen Scanner, OCR eventuell als Problem?
7. Discord chatbot für Fragen? Bräuchte das CLI um Text zu extrahieren, Copyright wird fraglich
8. Image inpainting for highlight recovery using raw photography?

Edited 9 Nov 2021

Created 8 Nov 2021

Close



Prüfungsform

Modul

Prüfungs-/
Abgabedatum

Status

Hausarbeit

AP CL5: Morphologische und syntaktische Sprachverarbeitung
(Hausarbeit oder Studienarbeit oder Projektarbeit)

05.08.2023

angemeldet

BONUS SLIDE (4/n)

- I had to politely ask them to give me a possible return statement
- Has never happened to me before
- I could still copy and paste their suggestion
- It's syntactically correct, and seems to be pragmatically alright.



Vipitis

can you give me a possible return statement?



GitHub Copilot

Certainly! Here's a possible return statement for the `noise` function:

```
float noise(vec3 p) //Thx to Las^Mercury
{
    vec3 i = floor(p);
    vec4 a = dot(i, vec3(1., 57., 21.)) + vec4(0., 57., 21., 78.);
    vec3 f = cos((p-i)*acos(-1.))*(-.5)+.5;
    a = mix(sin(cos(a)*a),sin(cos(1.+a)*(1.+a)), f.x);
    a.xy = mix(a.xz, a.yw, f.y);
    return dot(a, vec4(1.0/vec3(1.0, 57.0, 113.0), 0.0));
}
```

This implementation uses Perlin noise to generate a value at the given 3D position `p`. The return statement returns the dot product of the generated



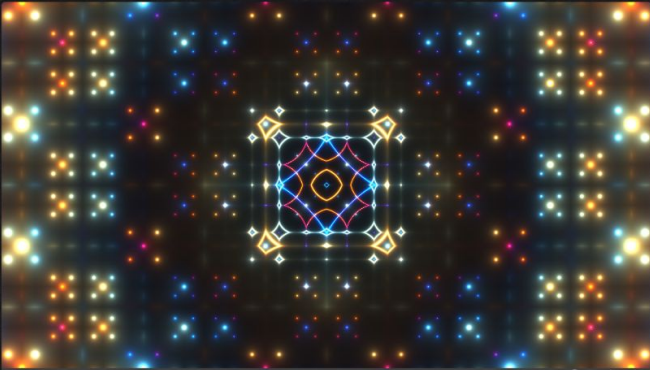
Oops, your response got filtered. Vote down if you think this shouldn't have happened.

BACKUP SLIDE 0 Shadertoy.com

Shadertoy

Search...

Welcome jake101 | Browse New Logout



0.00 164.9 fps 1000 x 563


Shader Art Coding Introduction

Views: 15721, tags: fractal, sdf, tutorial, intro, repetition, code, youtube, tuto, creative

<https://youtu.be/f4s1h2YETNY>


Created by kishimisu in 2023-05-20


Comments (30)




?

Post

 kaihenthe, 2023-06-22
this is the first tutorial i watched

 Maskillz, 2023-06-21
Thanks Kishimisu W

 konigma, 2023-06-14
This video introduced me to the world of Shader Art. Thank you so much.

+ Image

Shader Inputs

```
1 /* This animation is the material of my first youtube tutorial about creative
2 coding, which is a video in which I try to introduce programmers to GLSL
3 and to the wonderful world of shaders, while also trying to share my recent
4 passion for this community.
5
6 */
7
8 //https://iquilexles.org/articles/palettes/
9 vec3 palette( float t ) {
10     vec3 a = vec3(0.5, 0.5, 0.5);
11     vec3 b = vec3(0.5, 0.5, 0.5);
12     vec3 c = vec3(1.0, 1.0, 1.0);
13     vec3 d = vec3(0.263, 0.416, 0.557);
14
15     return a + b*cos( 6.28318*(c*t+d) );
16 }
17
18 void mainImage( out vec4 fragColor, in vec2 fragCoord ) {
19     vec2 uv = (fragCoord * 2.0 - iResolution.xy) / iResolution.y;
20     vec2 uv0 = uv;
21     vec3 finalColor = vec3(0.0);
22
23     for (float i = 0.0; i < 4.0; i++) {
24         uv = fract(uv * 1.5) - 0.5;
25
26         float d = length(uv) * exp(-length(uv0));
27
28         vec3 col = palette(length(uv0) + i*.4 + iTime*.4);
29
30         d = sin(d*.8 + iTime)/0.;
31         d = abs(d);
32
33         d = pow(0.01 / d, 1.2);
34
35         finalColor += col * d;
36     }
37
38     fragColor = vec4(finalColor, 1.0);
39 }
```

Compiled in 0.0 secs 533 chars

iChannel0

iChannel1

iChannel2

iChannel3

BACKUP SLIDE 1 <https://www.shadertoy.com/view/fstyD4>



BACKUP SLIDE 2 <https://www.shadertoy.com/view/3lsSzf>



BACKUP SLIDE 3 <https://www.shadertoy.com/view/DdlcRf>

