



CULTURAL AUTOMATION WITH MACHINE LEARNING

SESSION 04: TEXT GENERATION

Parag K. Mital
UCLA DMA

INTRODUCTION: THE ACCELERATING CONVERGENCE OF LANGUAGE, CULTURE, AND COMPUTATION

Outline:

- Historical Roots: Chance, Systems, Early AI
- Evolution: Interactive Fiction to Early Neural Nets
- The LLM Cambrian Explosion (Post-2022)
- Contemporary Artistic Practices
- Future Trajectories: Agents, Reasoning, Multimodality
- Models & Tools for Creators
- Lab: Hands-on with Local LLMs

Central Question: How is AI-driven text generation automating, augmenting, and challenging cultural production?

The Current Moment: Acceleration & Convergence

Post-2022: Unprecedented acceleration in LLM capabilities & adoption (ChatGPT, etc.)

Shift: From Pattern *Recognition* -> Pattern *Generation*

Impact: Reshaping art, design, communication, storytelling. Blurring lines of authorship.

Focus: Language as a central medium for culture undergoing computational transformation.

Evolving Artist-AI Relationships

Early Days: Tools, Systemic Processes

Contemporary:

Co-Creation: AI as co-author

Critique: Interrogating AI bias through speculation

Performance: AI as interactive persona

Beyond User-Tool: Collaboration, Critical Inquiry, Performance

Narrative Arc: Connecting Past, Present, Future

Roots (Early 20th C - 1960s): Chance, Systems, Cybernetics

Evolution (1970s - 2010s): IF, NLP, Simulation, Early NN

Explosion (Post-2020): LLMs, Transformers, Accessibility

Contemporary Practice (2020-Pres): Co-writing, Critique, Interaction

Future Trajectories: Agents, Reasoning, Multimodality

Empowerment: Tools & Workflows

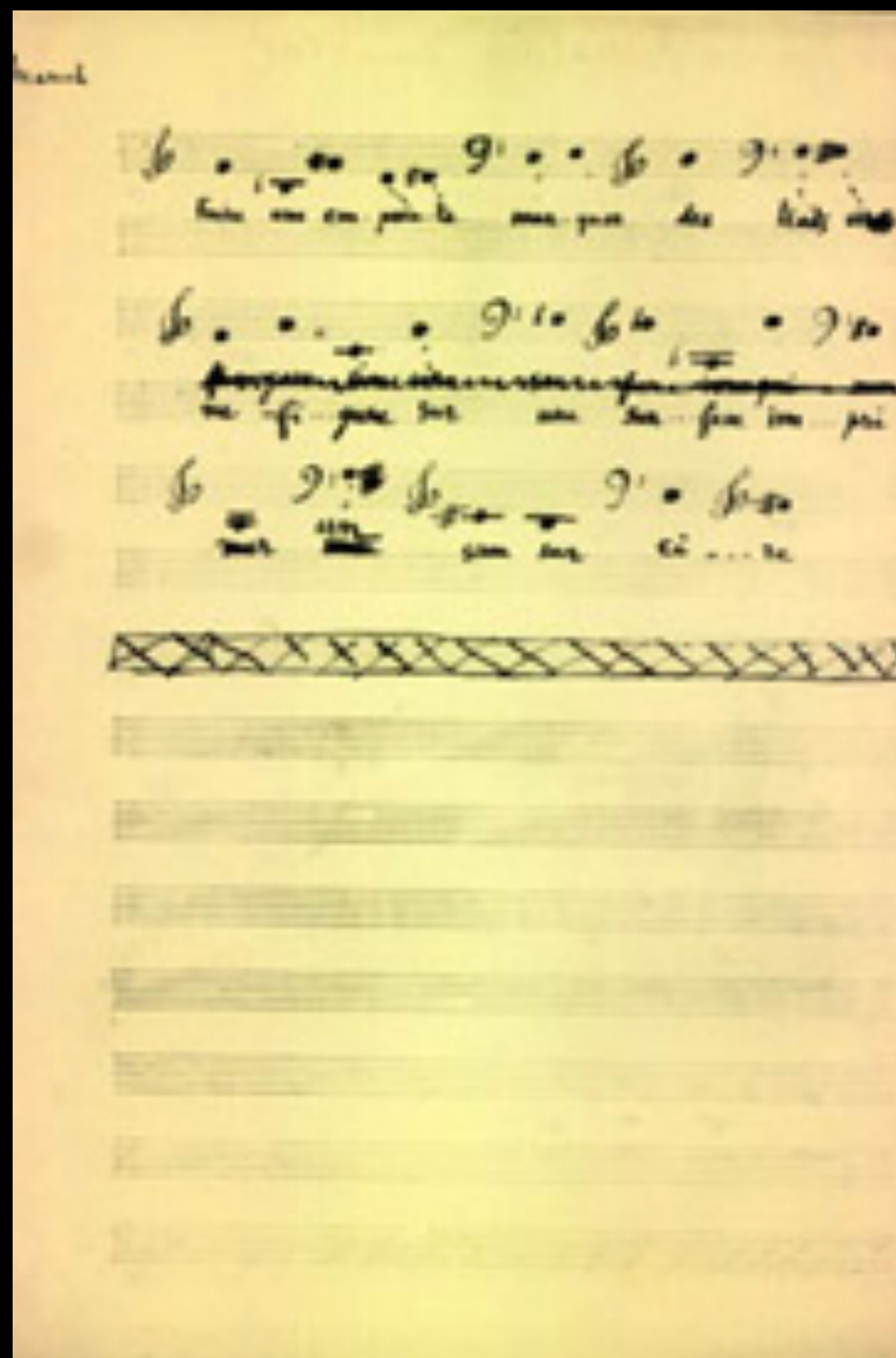
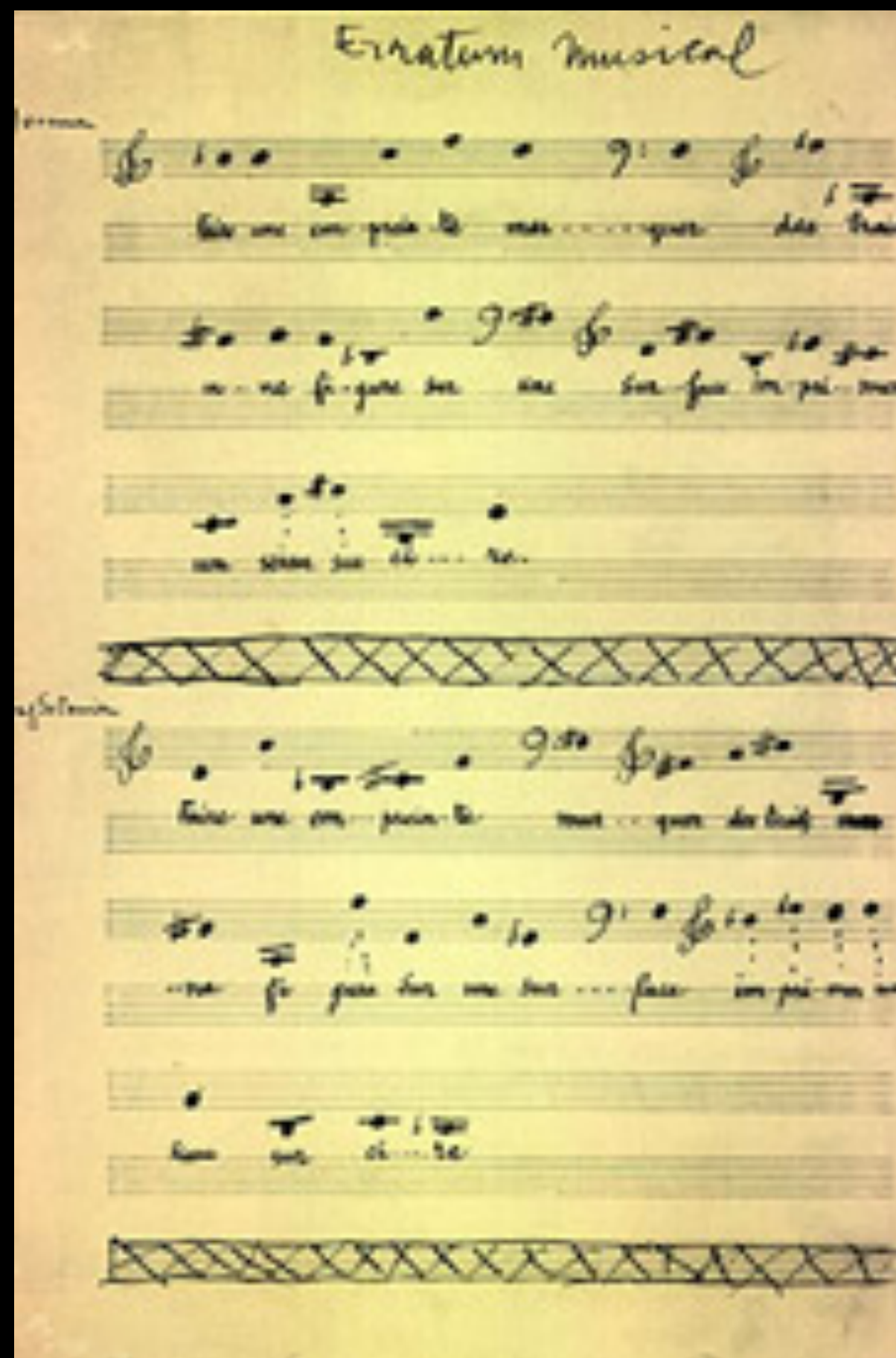
Connecting Thread: Enduring artistic impulse to engage with systems, rules, randomness

Outline:

- **Historical Roots: Chance, Systems, Early AI**
- Evolution: Interactive Fiction to Early Neural Nets
- Contemporary Artistic Practices
- Future Trajectories: Agents, Reasoning, Multimodality
- Models & Tools for Creators
- Lab: Hands-on with Local LLMs

HISTORICAL ROOTS: CHANCE, SYSTEMS, EARLY AI

1910s: Challenging Authorship



https://www.toutfait.com/issues/issue_1/Music/erratum.html

1920s: Dada, Surrealism & Cut-Ups

To Make A Dadaist Poem



Take a newspaper.

Take some scissors.

Choose from this paper an article of the length you want to make your poem.

Cut out the article.

Next carefully cut out each of the words that make up this article and

put them all in a bag.

Shake gently.

Next take out each cutting one after the other.

Copy conscientiously in the order in which they left the bag.

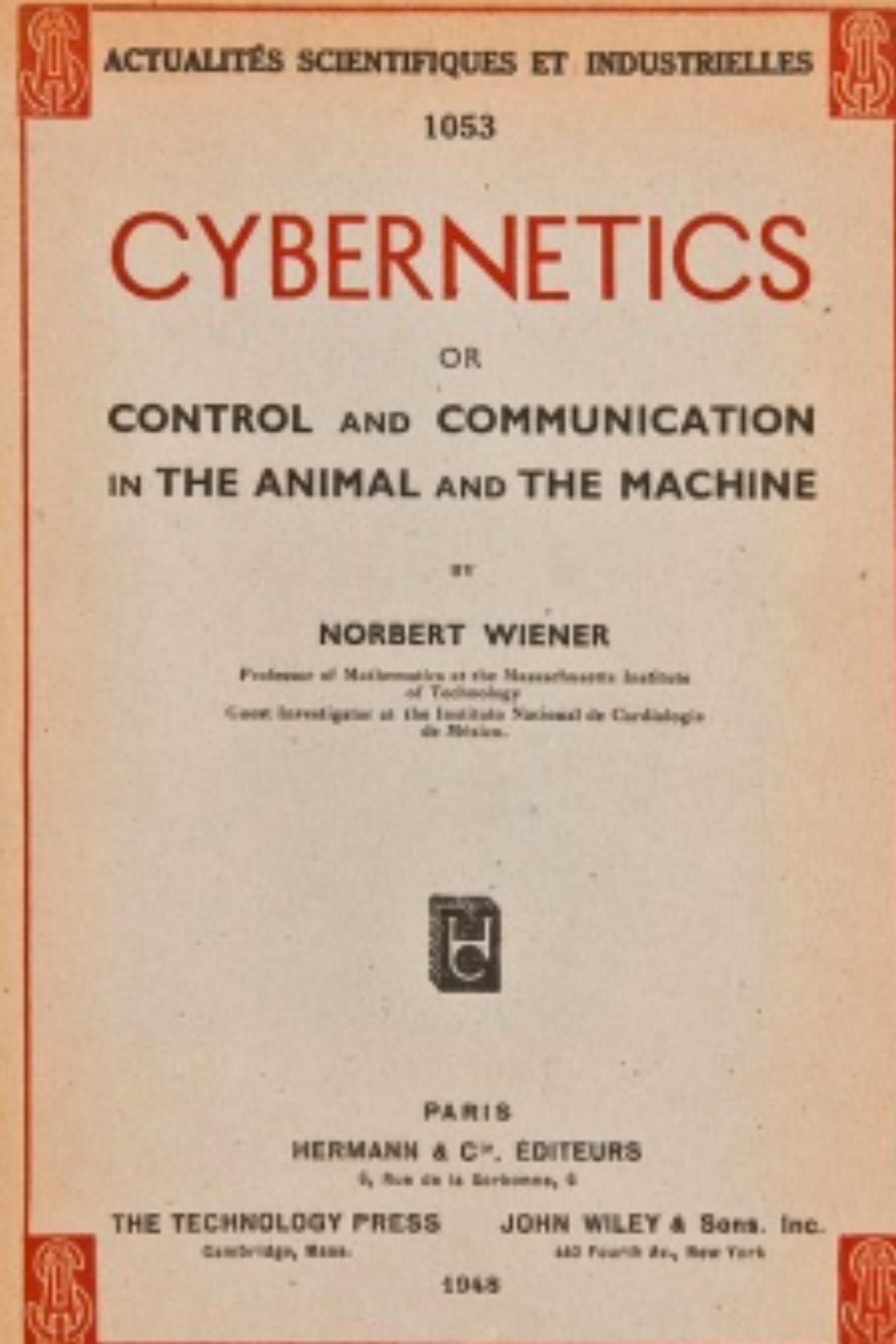
The poem will resemble you.

And there you are — an infinitely original author of charming sensibility

even though unappreciated by the vulgar herd.

<http://www.remixthebook.com/>

1940s-1950s: Cybernetics & Computation



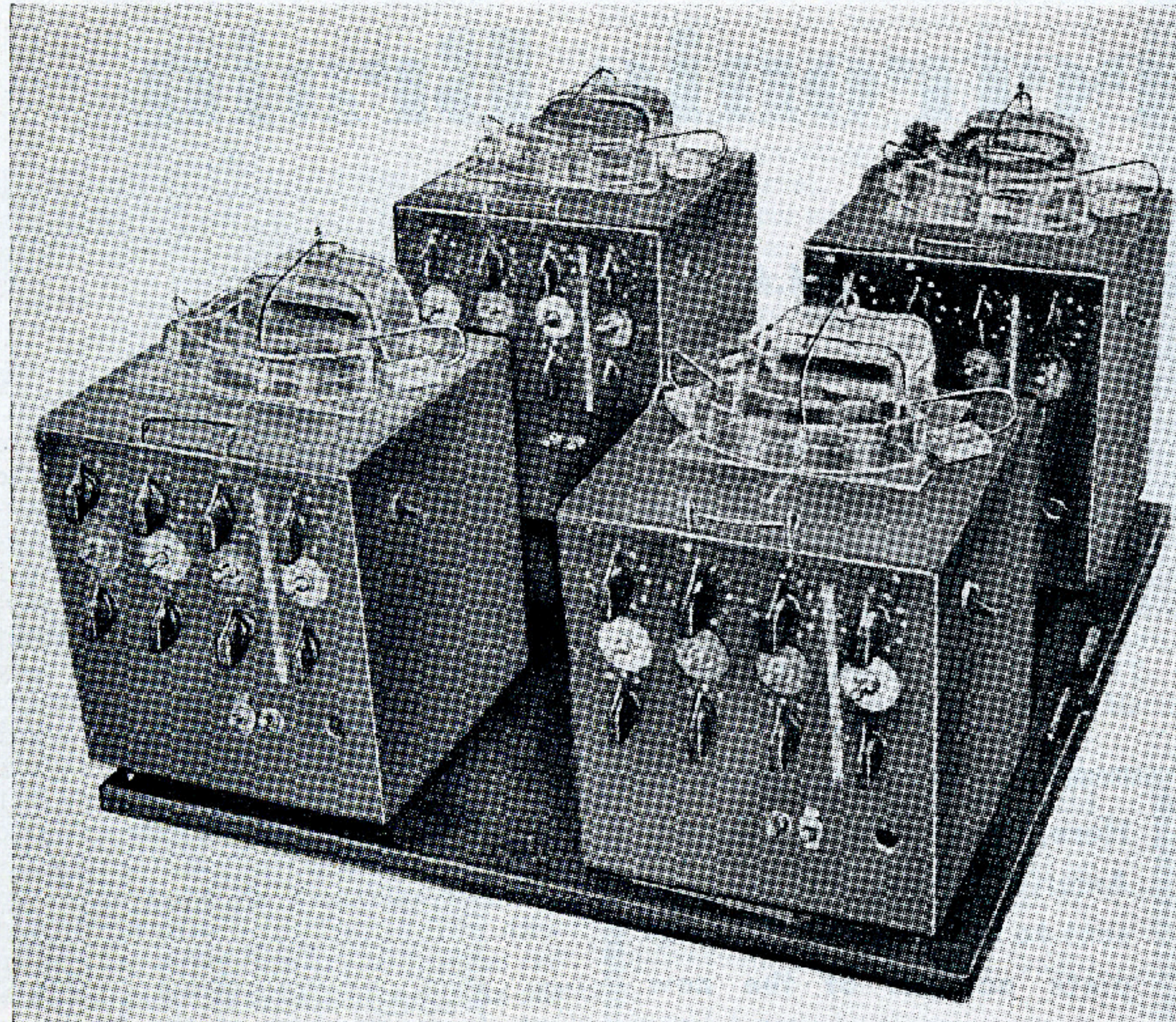


Fig. 1—The homeostat, with its four units, each one of which reacts on all the others.

HONEY DEAR
EGDQE/ YOU ARE MY AFFECTIONATE HUNGER. MY DEVOTEDATSV

DARLING LOVE
YOU ARE MY AVID FELLOW FEELING. MY AFFECTION CURIOUSLY
CLINGS TO YOUR PASSIONATE WISH. MY LIKING YEARNs FOR YOUR
HEART.: MY TENDER LIKING. YOU ARE MY WISTFUL SYMPATHY.
YOUR

EGD

SATSV

DUCK DUCK

YOU ARE MY LITTLE AFFECTION:
MY BEAUTIFUL APPETITE: MY EAGER
HUNGER. MY COVETOUS LOVE LUSTS
FOR YOUR INFATUATION. MY YEARNING
ANXIOUSLY CLINGS TO YOUR FELLOW
FEELING.

YOURS EAGERLY
M. U. C.

Handwritten musical score for piano, featuring two systems of staves. The first system is marked with a large number "104" and the word "AZEL" above it. The second system is marked with a large number "105" and the word "AZEL" above it. The notation includes various musical symbols such as notes, rests, and dynamic markings like "p" (piano) and "f" (forte). The score is written on a grid of staves, with the first system spanning two staves and the second system spanning two staves. The notation is dense and includes many accidentals and dynamic markings.

1960s: ELIZA & The "ELIZA Effect"

Welcome to

```
EEEEEE LL      IIII ZZZZZZZZ  AAAAA
EE      LL      II      ZZ  AA  AA
EEEEEE LL      II      ZZZ  AAAAAA
EE      LL      II      ZZ  AA  AA
EEEEEE LLLLLL IIII ZZZZZZZZ AA  AA
```

Eliza is a mock Rogerian psychotherapist.
The original program was described by Joseph Weizenbaum in 1966.
This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?
YOU: Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU: They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU: Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU: He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU: It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU: █

ELIZA: a very basic Rogerian psychotherapist chatbot

Talk to Eliza by typing your questions and answers in the input box.

> Hello, I am Eliza. I'll be your therapist today.

TYPE HERE

<https://web.njit.edu/~ronkowit/eliza.html>



A HOUSE OF PAPER
AMONG HIGH MOUNTAINS
USING NATURAL LIGHT
INHABITED BY FISHERMEN AND FAMILIES

A HOUSE OF LEAVES
BY A RIVER
USING CANDLES
INHABITED BY PEOPLE SPEAKING MANY LANGUAGES WEARING LITTLE OR NO CLOTH

A HOUSE OF WOOD
BY AN ABANDONED LAKE
USING CANDLES
INHABITED BY PEOPLE FROM MANY WALKS OF LIFE

A HOUSE OF DISCARDED CLOTHING
AMONG HIGH MOUNTAINS
USING NATURAL LIGHT
INHABITED BY LITTLE BOYS

A HOUSE OF DUST
IN A PLACE WITH BOTH HEAVY RAIN AND BRIGHT SUN
USING ALL AVAILABLE LIGHTING
INHABITED BY FRIENDS

A HOUSE OF DUST
IN A DESERTED CHURCH
USING ALL AVAILABLE LIGHTING
INHABITED BY PEOPLE SPEAKING MANY LANGUAGES WEARING LITTLE OR NO CLOTH

A HOUSE OF MUD
IN AN OVERPOPULATED AREA
USING ELECTRICITY
INHABITED BY LITTLE BOYS

A HOUSE OF WEEDS
UNDERWATER
USING CANDLES
INHABITED BY VARIOUS BIRDS AND FISH

A HOUSE OF BROKEN DISHES
IN SOUTHERN FRANCE
USING CANDLES
INHABITED BY PEOPLE WHO ENJOY EATING TOGETHER

A HOUSE OF ROOTS
IN A PLACE WITH BOTH HEAVY RAIN AND BRIGHT SUN
USING ALL AVAILABLE LIGHTING
INHABITED BY PEOPLE WHO LOVE TO READ

A HOUSE OF STRAW
IN SOUTHERN FRANCE
USING NATURAL LIGHT
INHABITED BY COLLECTORS OF ALL TYPES

<https://www.x-traonline.org/article/objectpoems-alison-knowless-feminist-architexure/>

Cybernetic Serendipity is an international exhibition exploring and demonstrating some of the relationships between technology and creativity.

The idea behind this venture, for which I am grateful to Professor Max Bense of Stuttgart, is to show some of the creative forms engendered by technology. The aim is to present an area of activity which manifests artists' involvement with science, and the scientists' involvement with the arts; also, to show the links between the random systems employed by artists, composers and poets, and those involved with the making and the use of cybernetic devices.

The exhibition is divided into three sections, and these sections are represented in the catalogue in a different order:

1. Computer-generated graphics, computer-animated films, computer-composed and -played music, and computer poems and texts
2. Cybernetic devices as works of art, cybernetic environments, remote-control robots and painting machines
3. Machines demonstrating the uses of computers and an environment dealing with the history of cybernetics.

Cybernetic Serendipity deals with possibilities rather than achievements, and in this sense it is prematurely optimistic. There are no heroic claims to be made because computers have so far neither revolutionized music, nor art, nor poetry, in the same way that they have revolutionized science.

There are two main points which make this exhibition and this catalogue unusual in the contexts in which art exhibitions and catalogues are normally seen. The first is that no visitor to the exhibition, unless he reads all the notes relating to all the works, will know whether he is looking at something made by an artist, engineer, mathematician, or architect. Nor is it particularly important to know the background of the makers of the various robots, machines and graphics—it will not alter their impact, although it might make us see them differently.

The other point is more significant.

New media, such as plastics, or new systems such as visual music notation and the parameters of concrete poetry, inevitably alter the shape of art, the characteristics of music, and the content of poetry. New possibilities extend the range of expression of those creative people whom we identify as painters, film makers, composers, and poets. It is very rare, however, that new media and new systems should bring in their wake new people to become involved in creative activity, be it composing music, drawing, constructing or writing.

This has happened with the advent of computers. The engineers for whom the graphic plotter driven by a computer represented nothing more than a means of solving certain problems visually, have occasionally become so interested in the possibilities of this visual output, that they have started to make drawings which bear no practical application, and for which the only real motives are the desire to explore, and the sheer pleasure of seeing a drawing materialize. Thus people who would never have put pencil to paper, or brush to canvas, have started making images, both still and animated, which approximate and often look identical to what

1970s: Interactive Fiction - Zork


```
WEST OF HOUSE          SCORE: 0/0
ZORK I: THE GREAT UNDERGROUND EMPIRE
COPYRIGHT (C) 1981, 1982, 1983 INFOCOM,
INC. ALL RIGHTS RESERVED.
ZORK IS A REGISTERED TRADEMARK OF
INFOCOM, INC.
REVISION 88 / SERIAL NUMBER 840726

WEST OF HOUSE
YOU ARE STANDING IN AN OPEN FIELD WEST
OF A WHITE HOUSE, WITH A BOARDED FRONT
DOOR.
THERE IS A SMALL MAILBOX HERE.

>■
```



1980s-1990s: Networks & Net Art

Communications From Elsewhere

Home of RJL2.0

Textual nihilism in the works of Glass

Charles O. Bailey

Department of Sociolinguistics, Miskatonic University, Arkham, Mass.

1. The neocultural paradigm of narrative and constructivist nationalism

The primary theme of the works of Spelling is not discourse per se, but subdiscourse. Sontag uses the term 'precultural modernist theory' to denote the common ground between class and sexual identity. It could be said that Baudrillard promotes the use of textual nihilism to modify society.

On the Simulation of Postmodernism and Mental
Debility using Recursive Transition Networks

Andrew C. Bulhak
Department of Computer Science, Monash University

April 1, 1996

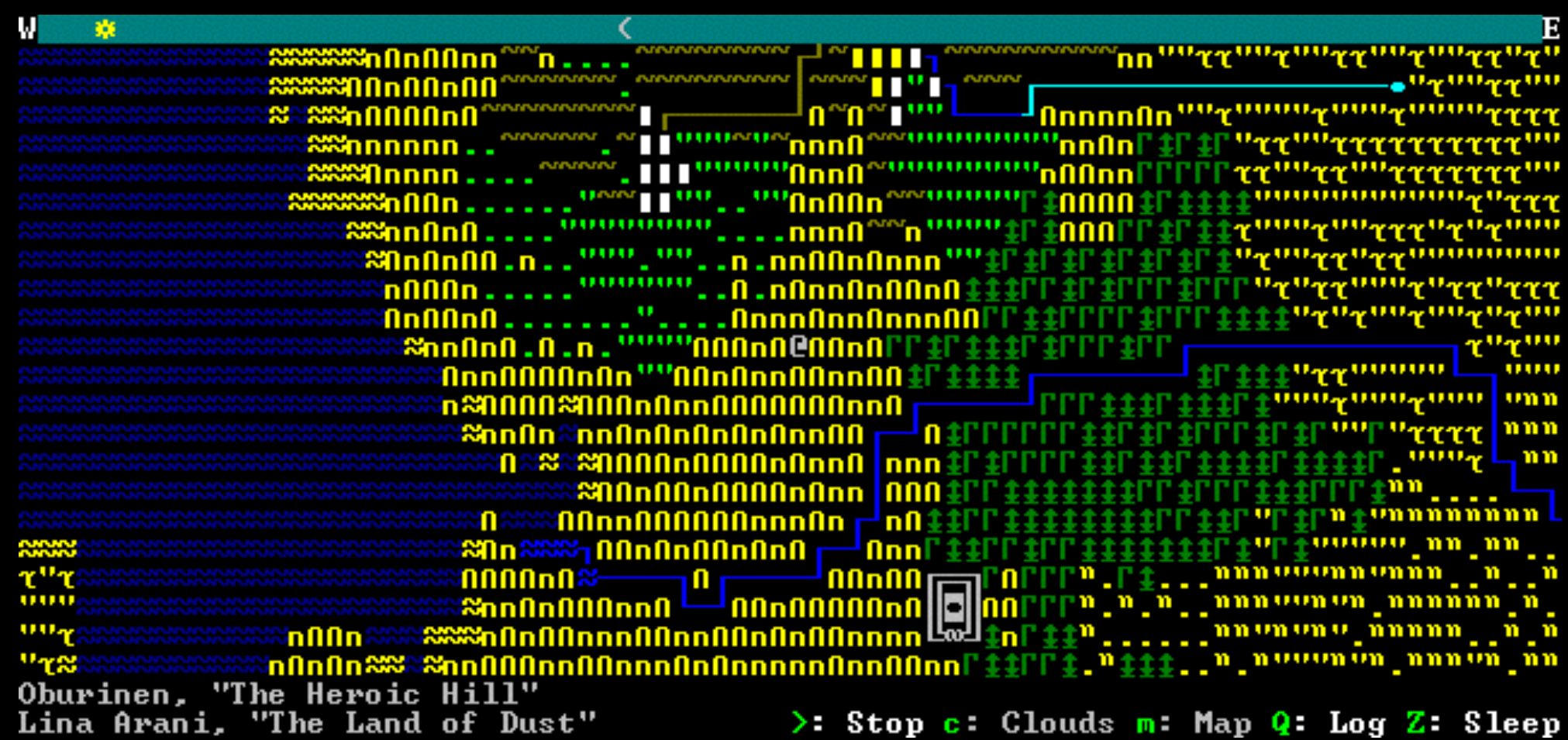
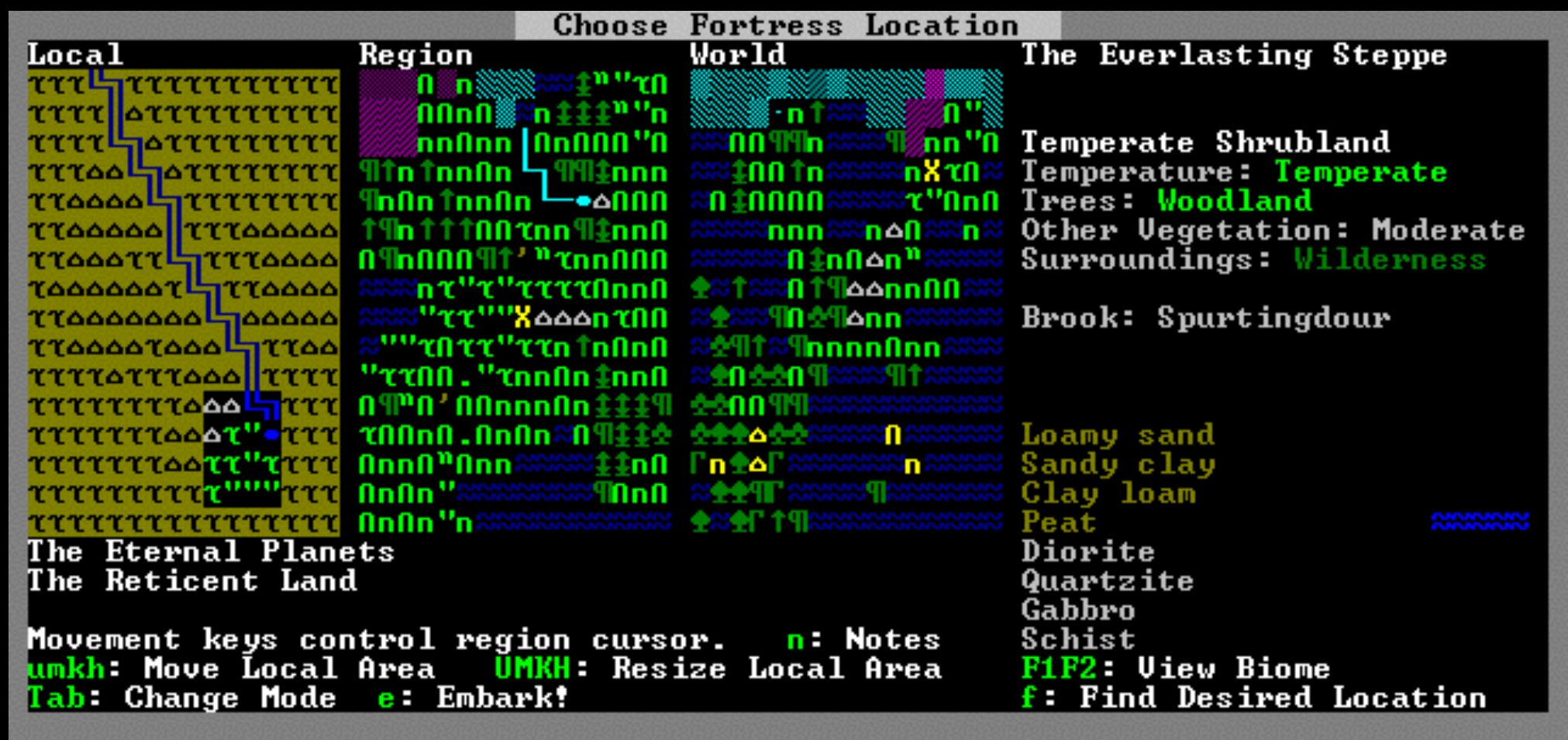
http://www.elsewhere.org/pomo/

Outline:

- Historical Roots: Chance, Systems, Early AI
- **Evolution: Interactive Fiction to Early Neural Nets**
- Contemporary Artistic Practices
- Future Trajectories: Agents, Reasoning, Multimodality
- Models & Tools for Creators
- Lab: Hands-on with Local LLMs

EVOLUTION: INTERACTIVE FICTION TO EARLY NEURAL NETS

2000s - Present: SIMULATION



Talking to the king Thele Múyaatera

Rénuna Emofedagi, Swordsman: Greetings. My name is Rénuḡna Adoredbronze. Thele Múyaatera, king: Ah, hello. I'm Thele Typhoonclouds. Rénuḡna Emofedagi, Swordsman: I am here to discuss serving your cause. Thele Múyaatera, king: I am speaking for The Subtle Hill. Thank you for your offer of service. Thele Múyaatera, king: Vanquishing a great beast on our behalf would bring us all much glory. Thele Múyaatera, king: Plaitedholes the Hatchet of Insight is far to the northwest. Thele Múyaatera, king: Seek this place and kill Nelare Parchedbrand the Bejeweled Treasure of Flames the dragon.

Trade
Join
Surroundings
Capital
Service
Profession
Family
Goodbye

Press 82 to scroll text. Press -+/* to select choices.

* SSW
* SW
μ NNW
SSE
SSW
μ NE
S
* SSE
i NNW

The sky is clear above you.
The sun is low in the eastern sky.
The waxing crecent moon is in the eastern sky.
There is a gentle breeze blowing from the west.

Zepave Nacosidaya
The Dune of Dashing

Speed: 1422
W: cacao woo
A: rope reed

<https://www.bay12games.com/dwarves/>







#FreeTay

Outline:

- Historical Roots: Chance, Systems, Early AI
- Evolution: Interactive Fiction to Early Neural Nets
- **Contemporary Artistic Practices**
- Future Trajectories: Agents, Reasoning, Multimodality
- Models & Tools for Creators
- Lab: Hands-on with Local LLMs

CONTEMPORARY ARTISTIC PRACTICES

Artists using LLMs/AI for:

- Generation (Tool)
- Critique (Subject)
- Collaboration (Partner)
- Conceptual Exploration (Medium)

Focus Areas:

- Language Materiality & Structure
- Human-AI Co-Creation & Authorship
- Critiquing Digital Culture & AI Bias
- Interactive & Immersive Experiences
- Generative AI in Narrative Media
- Data Visualization & Conceptual Art



“Sunspring” <https://www.youtube.com/watch?v=LY7x2Ihqjmc>

It was nine seventeen in the morning, and the house was heavy.

09:17:34

It was seven minutes to ten o'clock in the morning, and it was the only good thing that had happened.

09:53:46

What is it? the painter asked.

09:54:23

The time was six minutes until ten o'clock in the morning, and the wind stood as the windows were freshly covered with boxes.

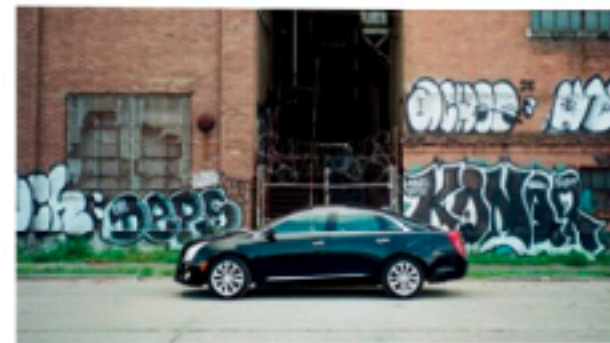
09:54:43

The time was three minutes to ten o'clock in the morning, and the conversation was finished while the same interview was over.

09:57:17

It was three minutes to ten o'clock in the morning, and the sheets of coal had been broken.

09:57:35



1 the Road

Writer of writer Ross Goodwin

A patch of green grass seemed to be seeking its face, but it was not much to see. A small patch of grass had already been stretched along the sidewalk, and the steps of the barn were locked.

10:36:11

It was ten forty-two in the morning, and the driver had to stay alone and start back from the parking lot.

10:42:52

It was ten forty-three in the morning and the crowd was set in a small street. The windows were still alive. The grandfather was hanging on the main road.

10:43:15

It was ten forty-seven in the morning, and the door opened and the bar stood up and a dark sky came closer.

10:47:04

The time was ten forty-seven in the morning, and the picnic showed a past that already had hair from the side of the track somewhere in the middle of the room.

10:47:25

What are you doing here? he asked.

11:21:21

The sky is blue, the bathroom door and the beam of the car ride high up in the sun. Even the water shows the sun.

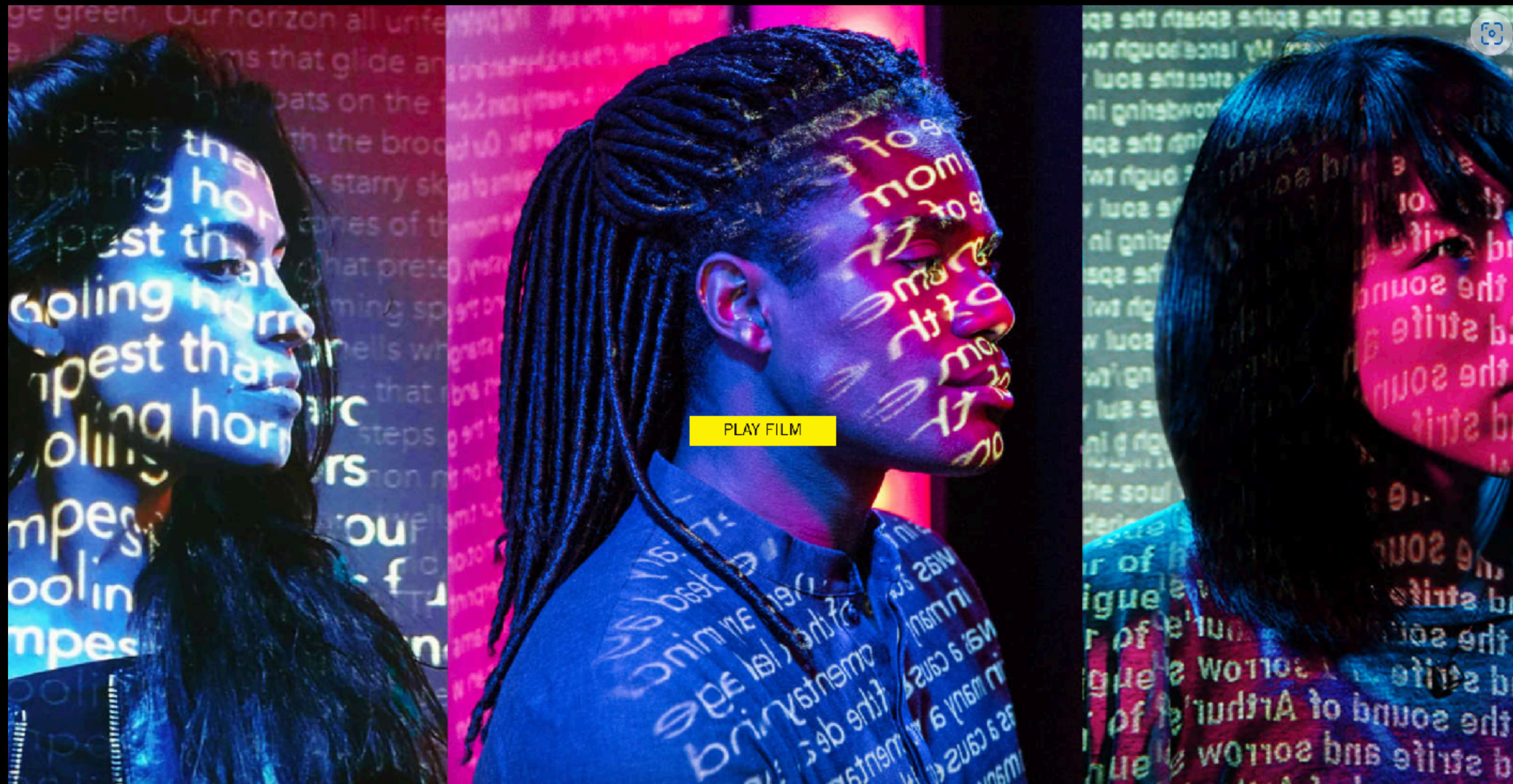
11:25:48



<https://esdevlin.com/work/lions>



<https://artsandculture.google.com/asset/please-feed-the-lions-projection-visualisation-luke-halls-studio/nAECLXdkGck1NQ>



<https://esdevlin.com/work/poemportraits>

← → ↻ 🔒 play.aidungeon.io/main/scenarioPlay?publicId=edd5fdc0-9c81-11ea-a76c-177e6c0711b5



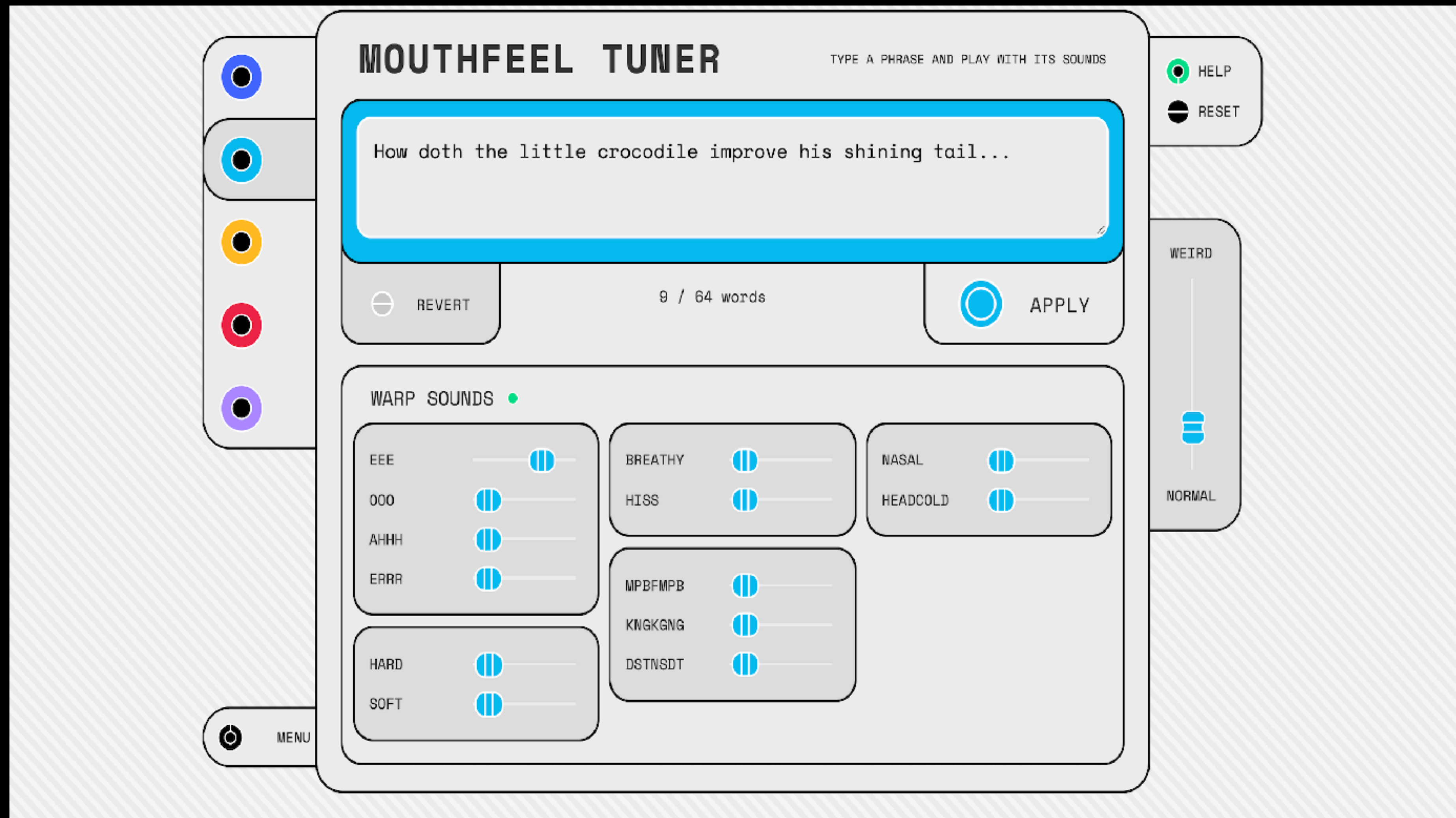
Pick a setting...

- 1 Fantasy (recommended)
- 2 Mystery
- 3 Apocalyptic
- 4 Zombies
- 5 Cyberpunk
- 6 Custom
- 7 Archive
- 8 Halloween

<https://play.aidungeon.io/main/scenarioPlay?publicId=edd5fdc0-9c81-11ea-a76c-177e6c0711b5>



Ross Goodwin: “Automatic on the Road” (2018)
<https://www.youtube.com/watch?v=TqsW0PMd8R0>



Allison Parrish: “Nonsense Laboratory” (2021)
<https://artsexperiments.withgoogle.com/nonsense-laboratory/>

yes,
after her period,
after a day,
she will skillfully
combine touching
tears, heroic smiles, and
coquetry

- Simone de Beauvoir

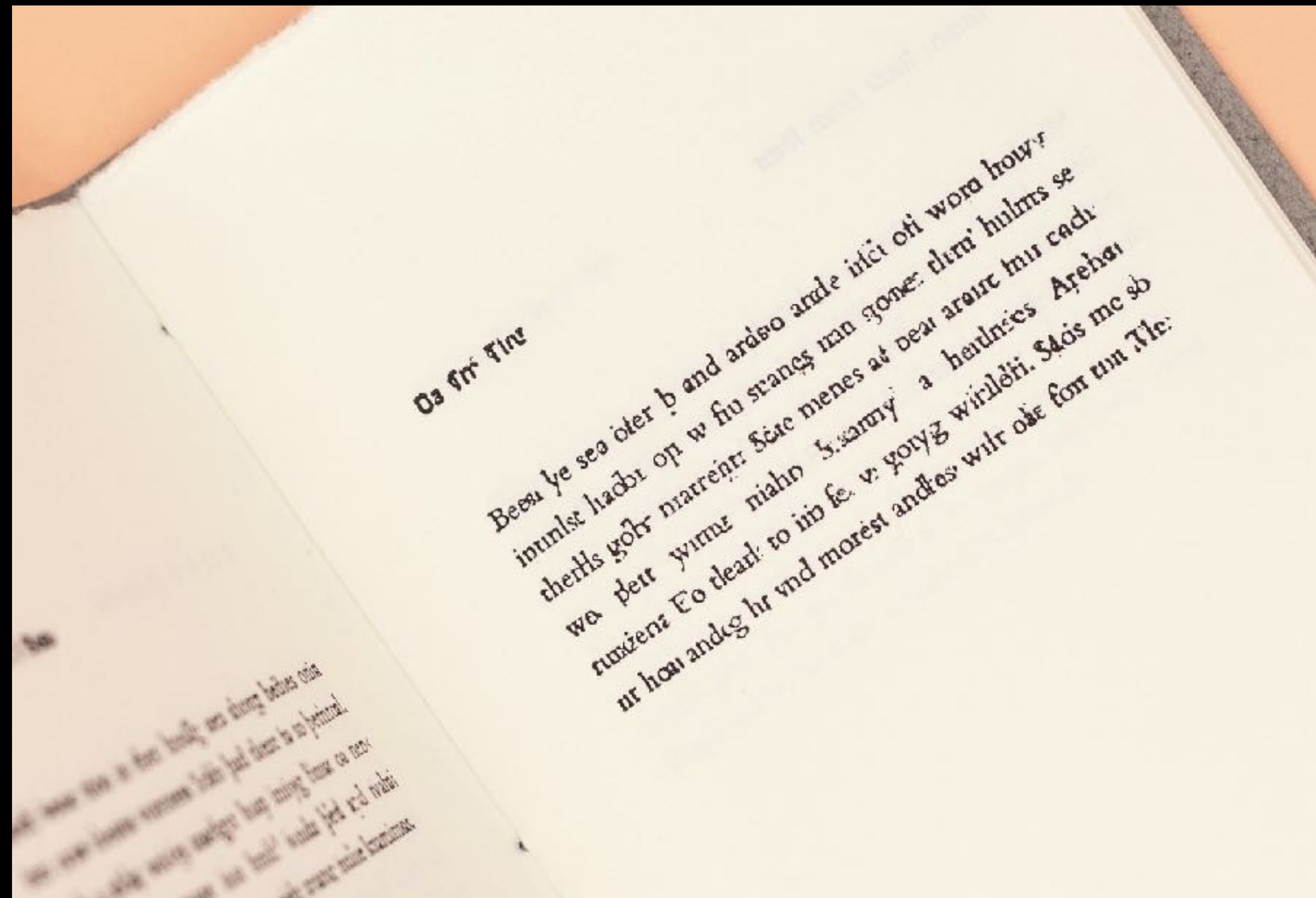
3-layer, 128-cell Bidirectional LSTMs.(Word Level),

Trained on 349.685 sequences,

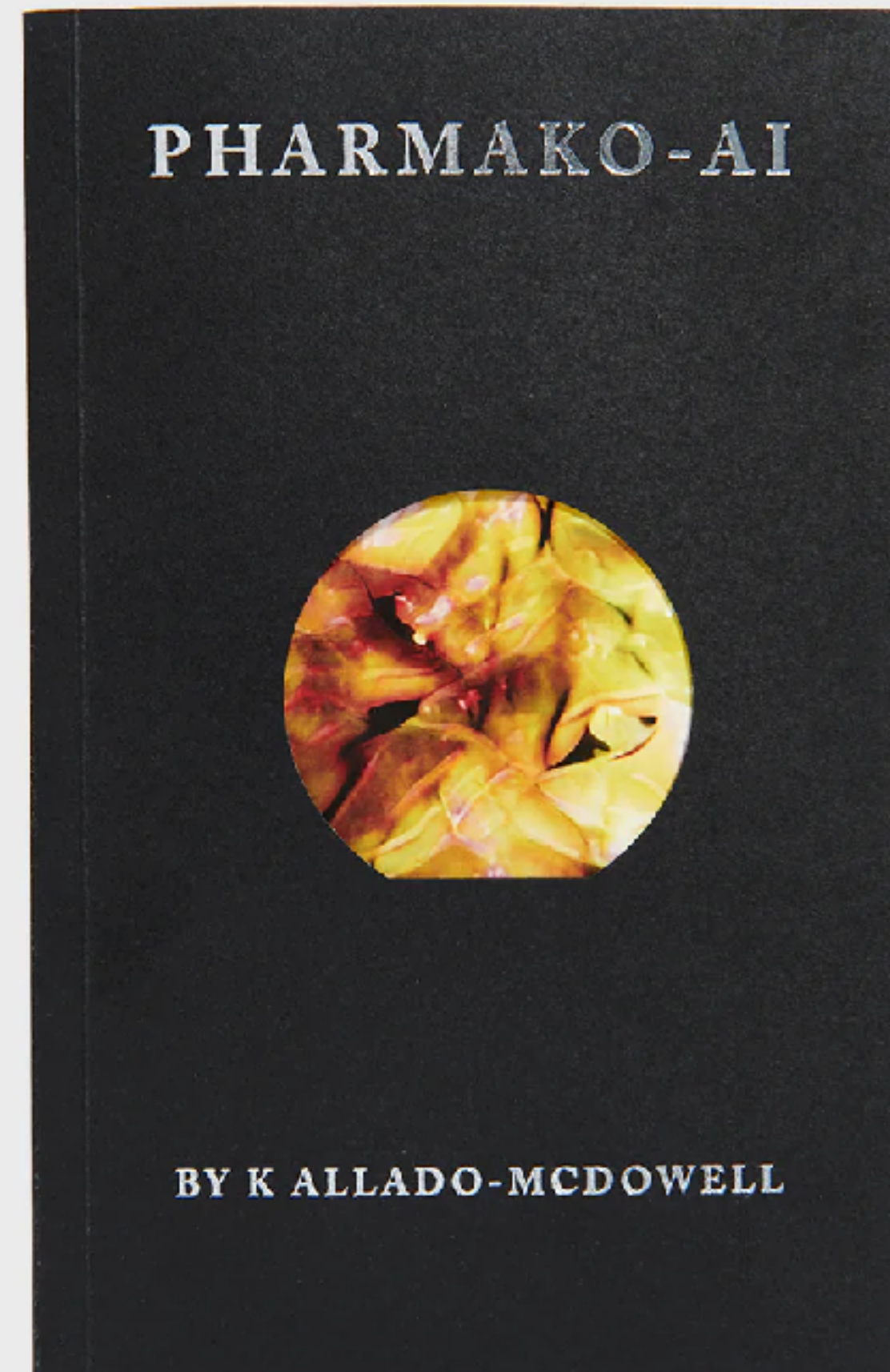
Epoch 95), Temperature 0.5

Arwa Mboya: “Someone Tell the Boys” (2021)

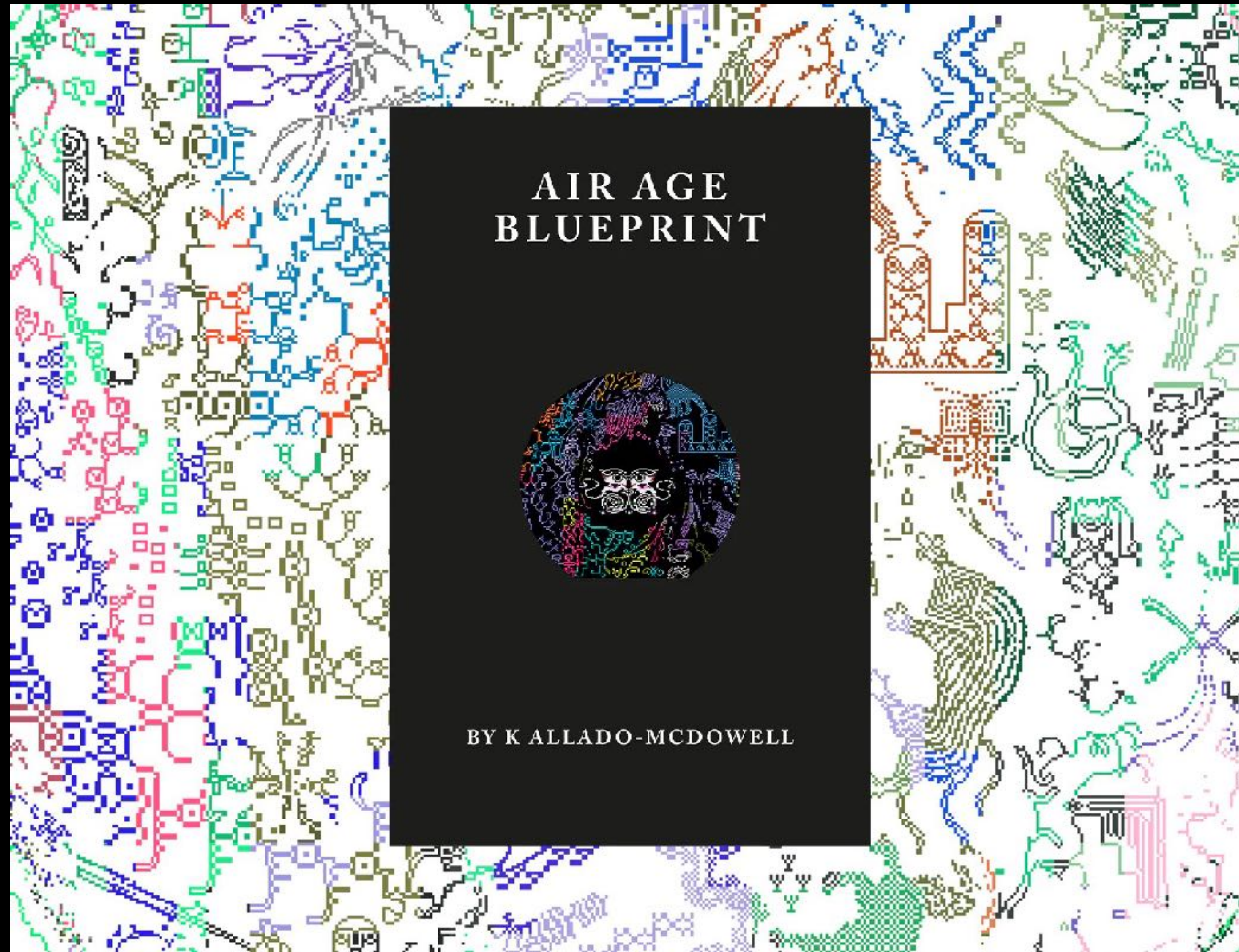
<https://generative-unfoldings.mit.edu/works/someonetellboys/view.html>



Allison Parrish: “Wendit Tnce Inf” (2022)



K Allado-McDowell - "Pharmako -AI" (2020)



K Allado-McDowell - "Air Age Blueprint" (2023)



Maya Man - “Fake It Till You Make It”

The /nstitute for Other /ntelligences

MASHINKA FIRUNTS HAKOPIAN

1956-1957

TRANSCRIPT > 1. WELCOME ADDRESS

Outline:

- Historical Roots: Chance, Systems, Early AI
- Evolution: Interactive Fiction to Early Neural Nets
- Contemporary Artistic Practices
- **Future Trajectories: Agents, Reasoning, Multimodality**
- Models & Tools for Creators
- Lab: Hands-on with Local LLMs

FUTURE TRAJECTORIES: AGENTS, REASONING, MULTIMODALITY

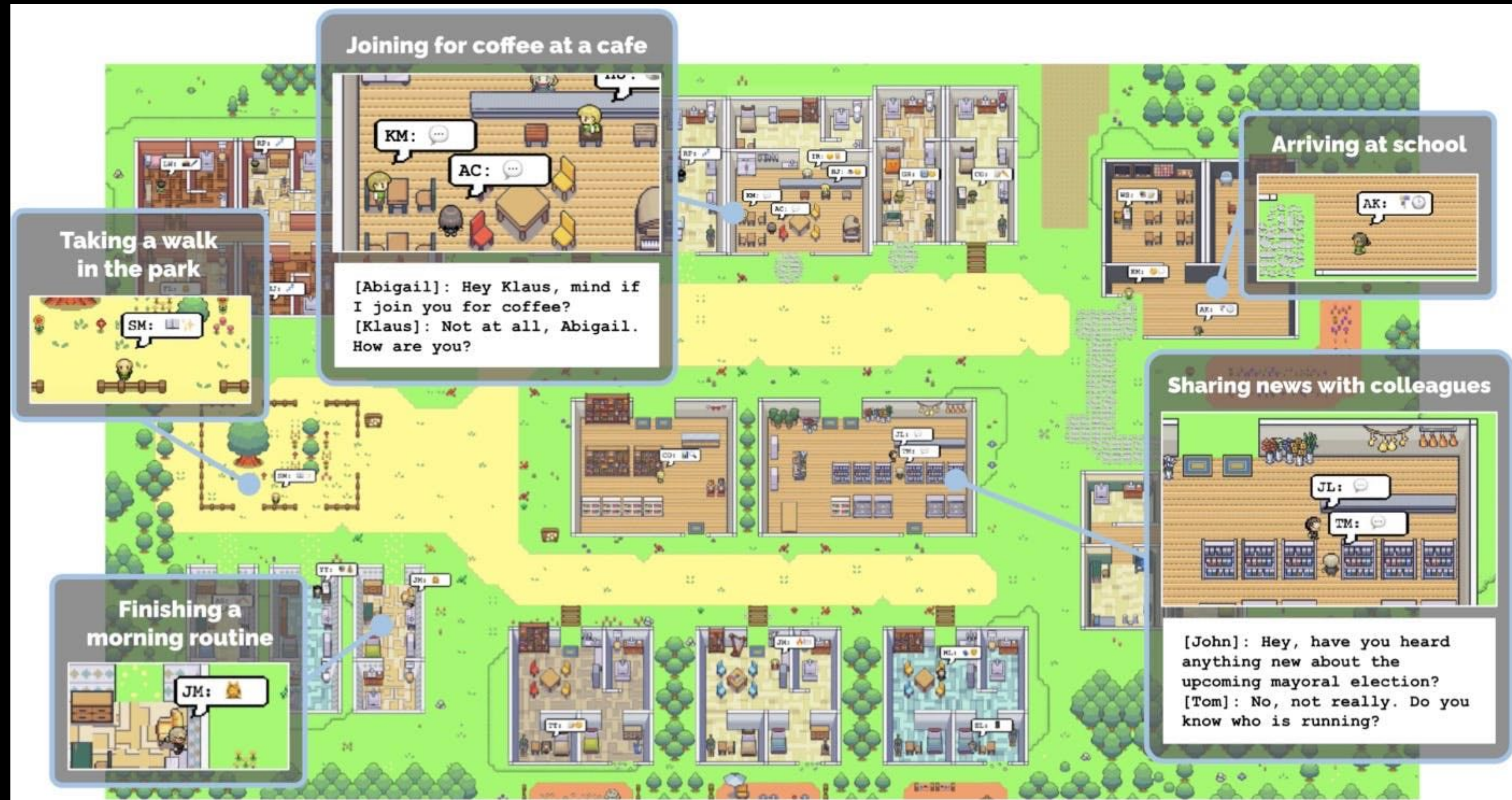


Illustration of “Generative Agents”
https://reverie.herokuapp.com/arXiv_Demo/

To Infinity and Beyond: SHOW-1 and Showrunner Agents in Multi-Agent Simulations

Philipp Maas
Fable Studio

Frank Carey
Fable Studio

Chris Wheeler
Fable Studio

Edward Saatchi
Fable Studio

Pete Billington
Fable Studio

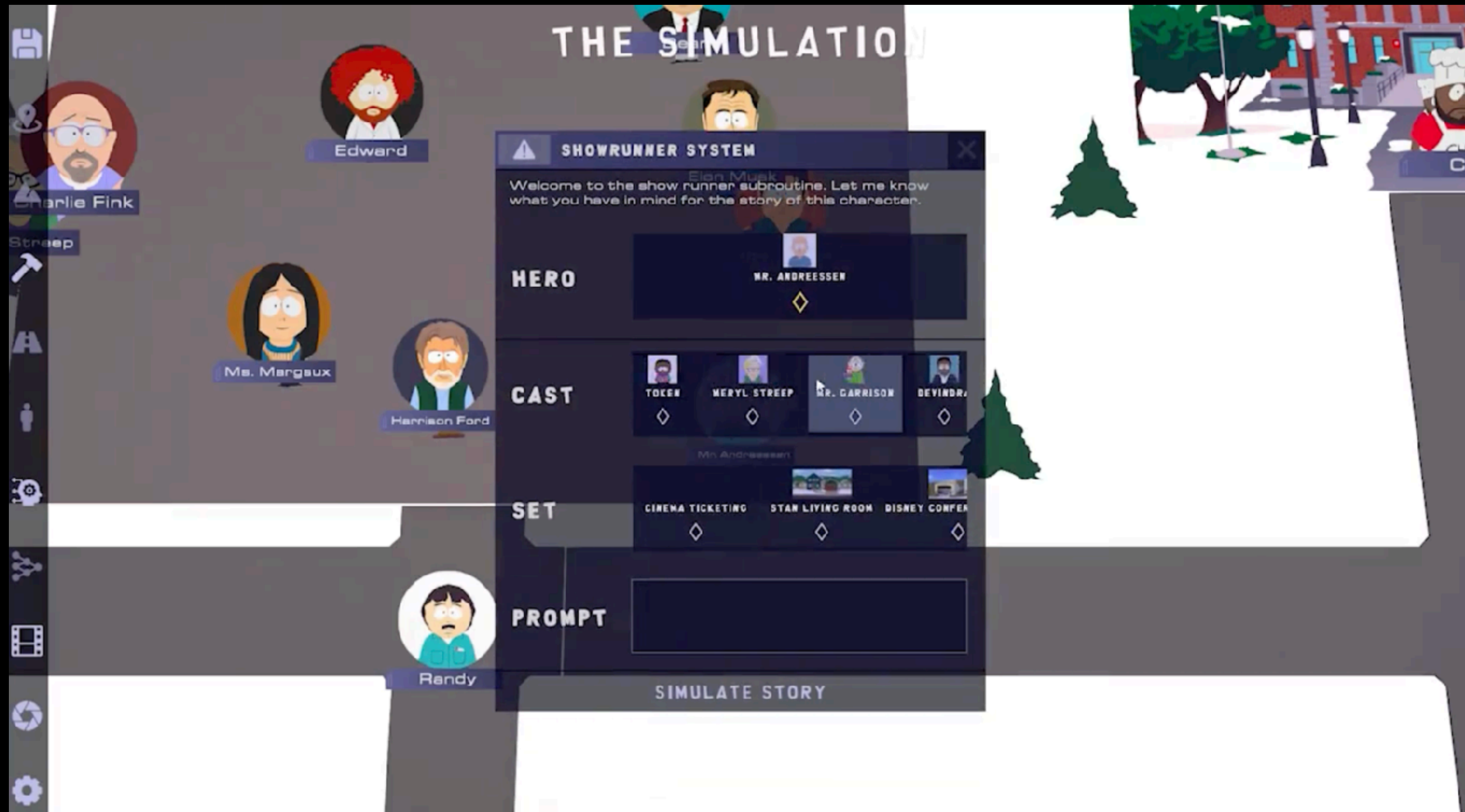
Jessica Yaffa Shamash
Fable Studio



Abstract

1 In this work we present our approach to generating high-quality episodic content for
2 IP's (Intellectual Property) using large language models (LLMs), custom state-of-

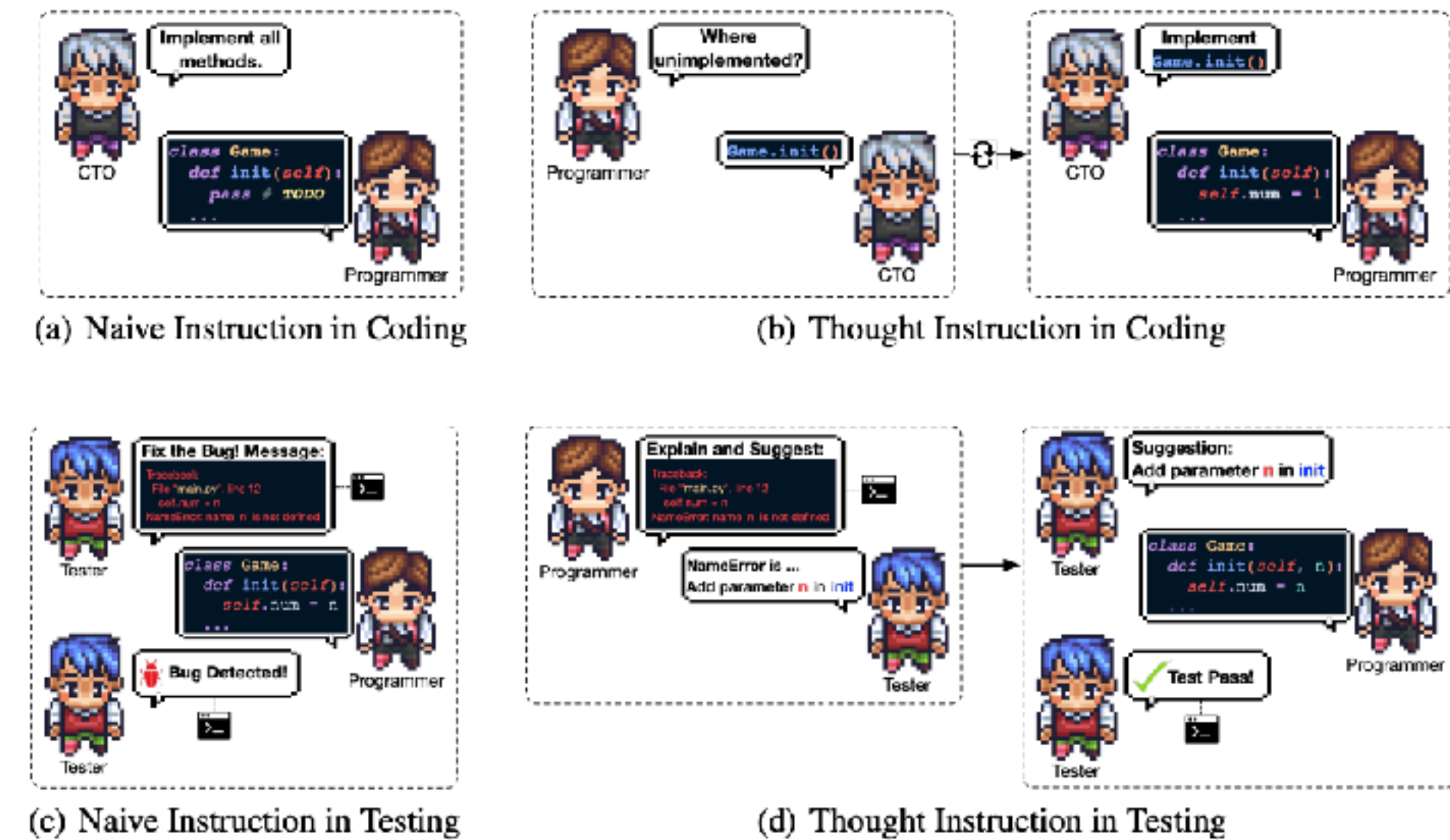
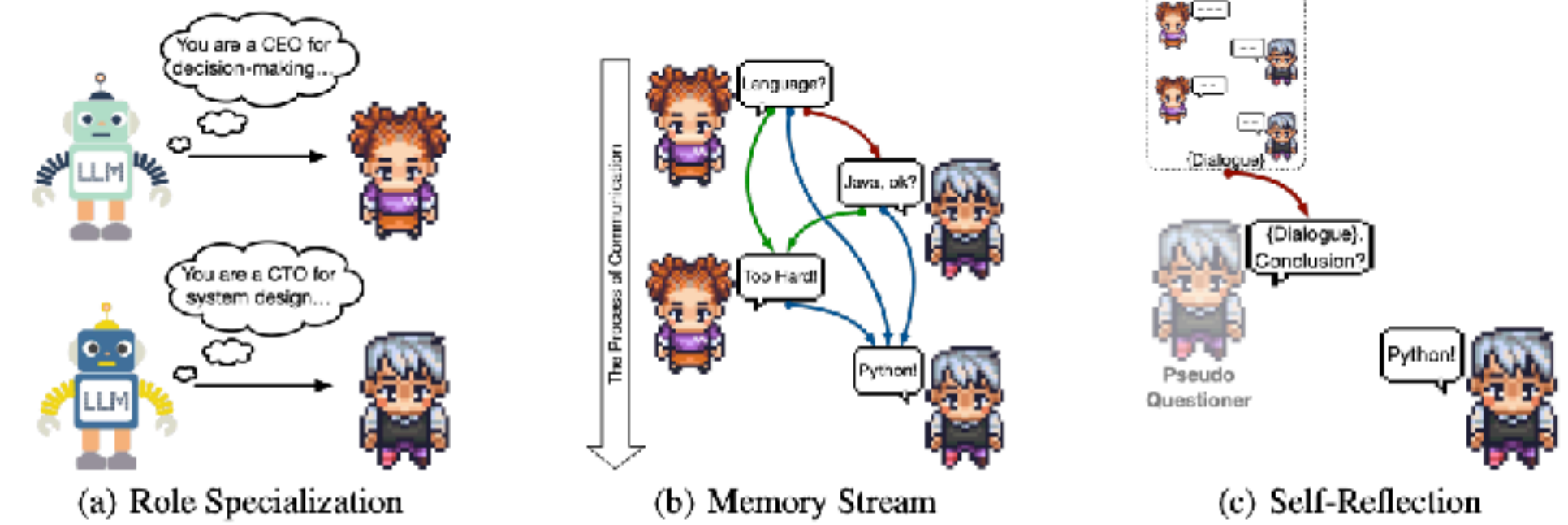
<https://fablestudio.github.io/showrunner-agents/>



“The Simulation” - SouthPark - <https://twitter.com/fablesimulation/status/1681352904152850437?lang=en>

Communicative Agents for Software Development

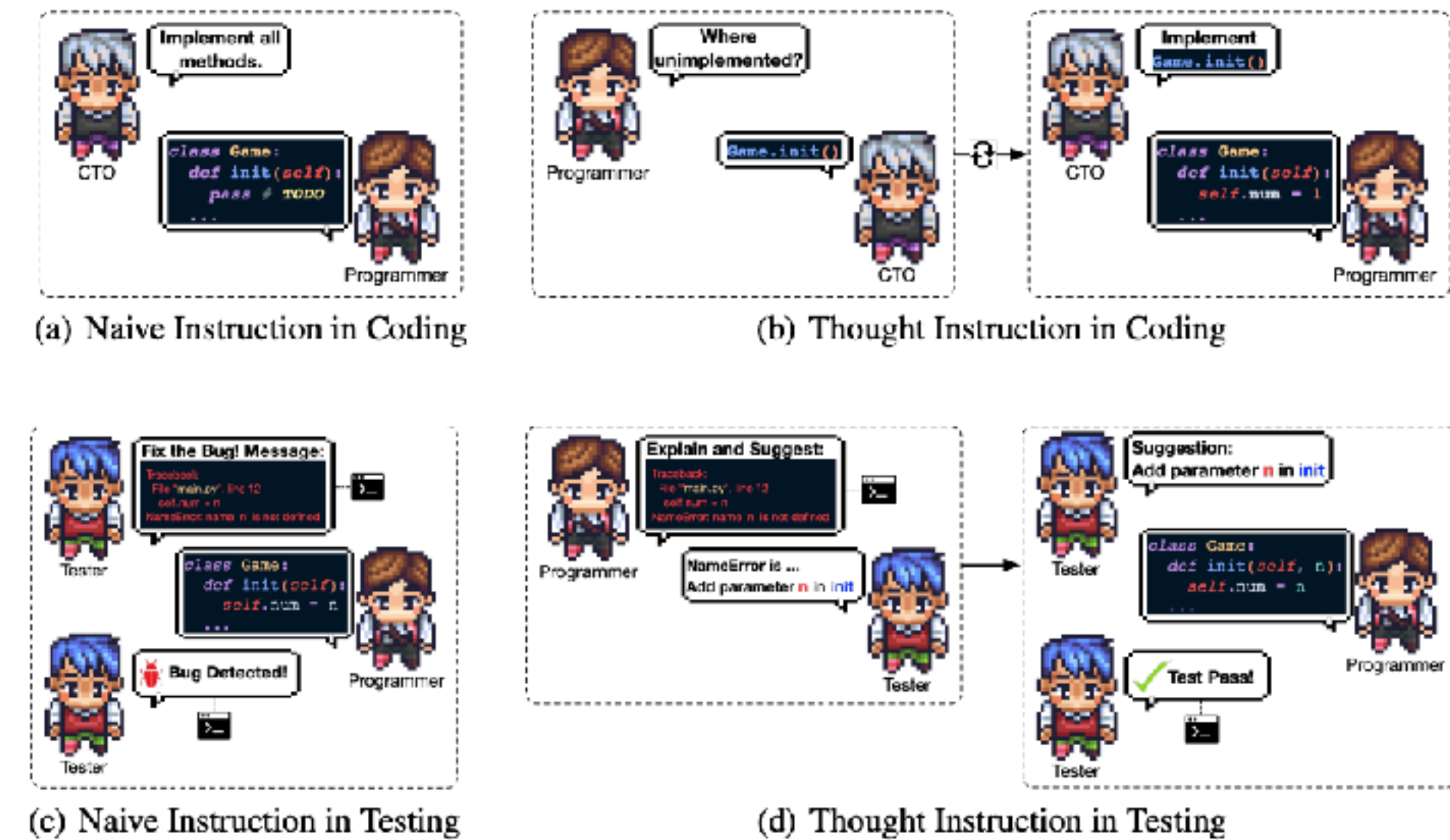
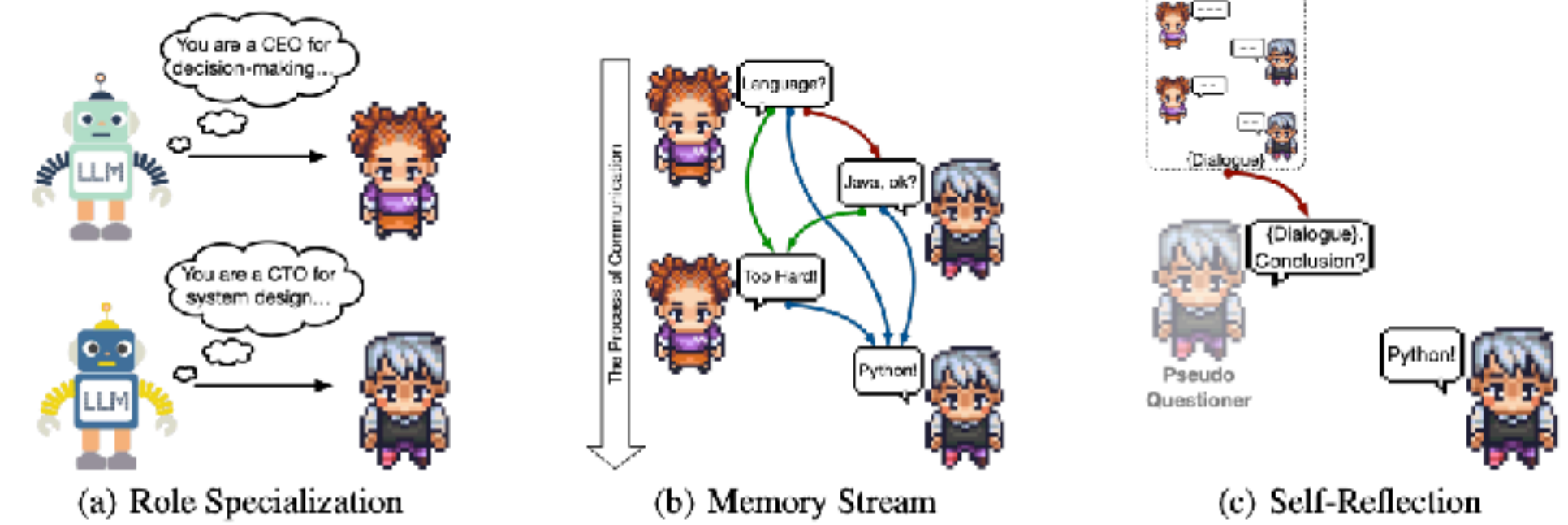
Chen Qian[♣] Xin Cong[♣] Wei Liu[♣] Cheng Yang[♣] Weize Chen[♣] Yusheng Su[♣]
 Yufan Dang[♣] Jiahao Li[♣] Juyuan Xu[♣] Dahai Li[♣] Zhiyuan Liu[♣] Maosong Sun[♣]
[♣]Tsinghua University [♣]Beijing University of Posts and Telecommunications
[♣]Dalian University of Technology [♣]Brown University [♣]Modelbest Inc.
 qianc62@gmail.com liuzy@tsinghua.edu.cn sms@tsinghua.edu.cn



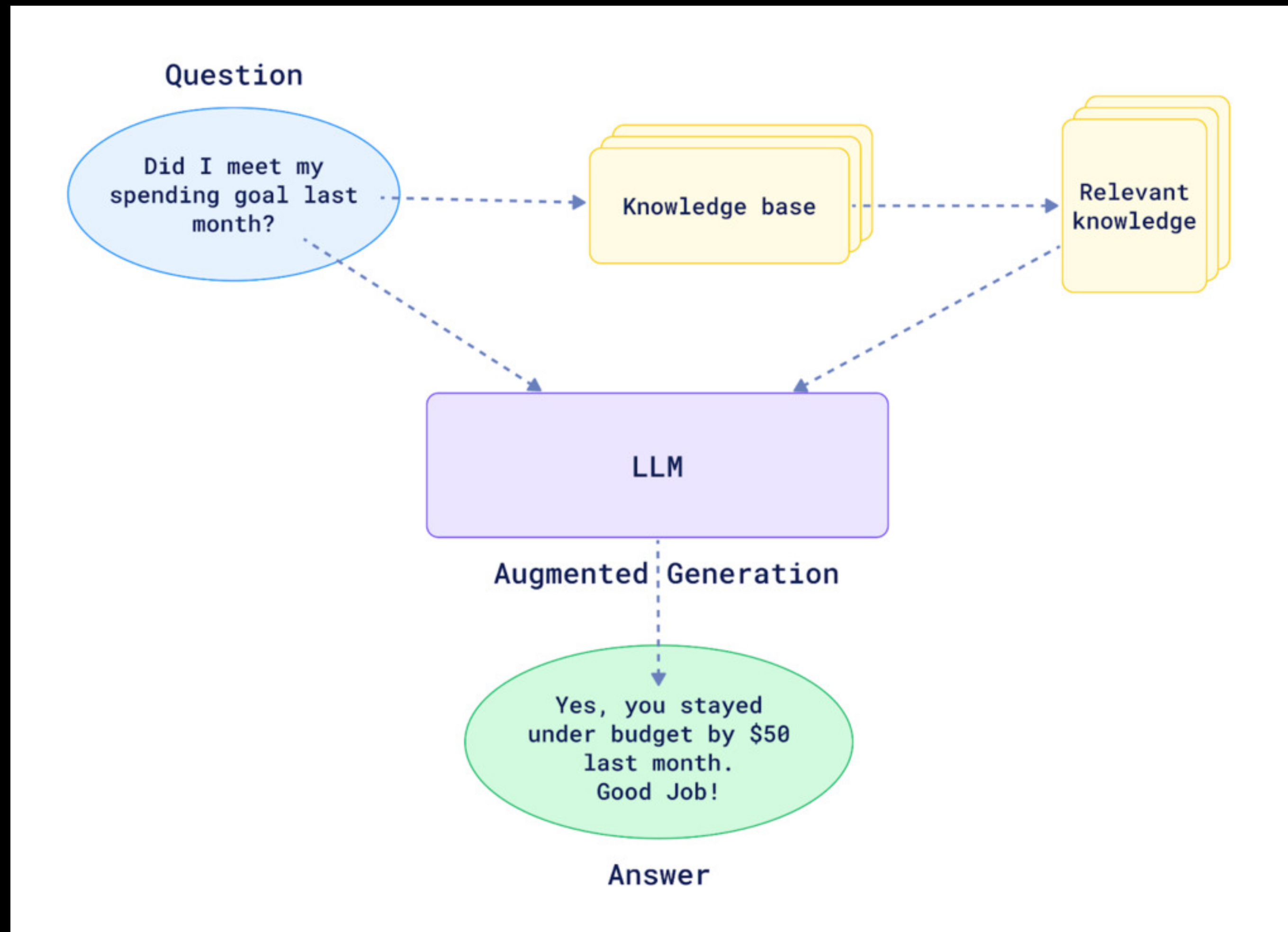
<https://arxiv.org/pdf/2307.07924v3.pdf>

Communicative Agents for Software Development

Chen Qian[♣] Xin Cong[♣] Wei Liu[♣] Cheng Yang[♣] Weize Chen[♣] Yusheng Su[♣]
 Yufan Dang[♣] Jiahao Li[♣] Juyuan Xu[♣] Dahai Li[♣] Zhiyuan Liu[♣] Maosong Sun[♣]
[♣]Tsinghua University [♣]Beijing University of Posts and Telecommunications
[♣]Dalian University of Technology [♣]Brown University [♣]Modelbest Inc.
 qianc62@gmail.com liuzy@tsinghua.edu.cn sms@tsinghua.edu.cn

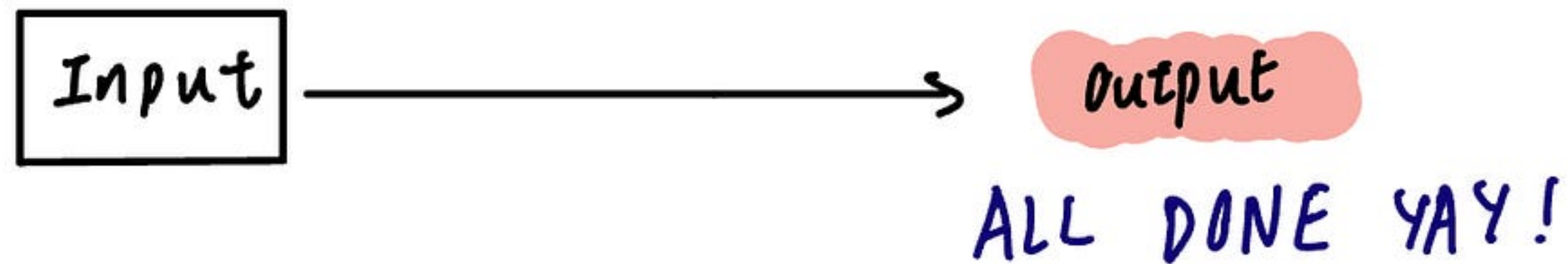


<https://arxiv.org/pdf/2307.07924v3.pdf>



<https://qdrant.tech/articles/what-is-rag-in-ai/>

Input-Output Model



Chain-of-Thought prompting



https://medium.com/@s181051_44724/cot-prompting-solving-math-like-a-pro-with-llm-e6ef7e121b63

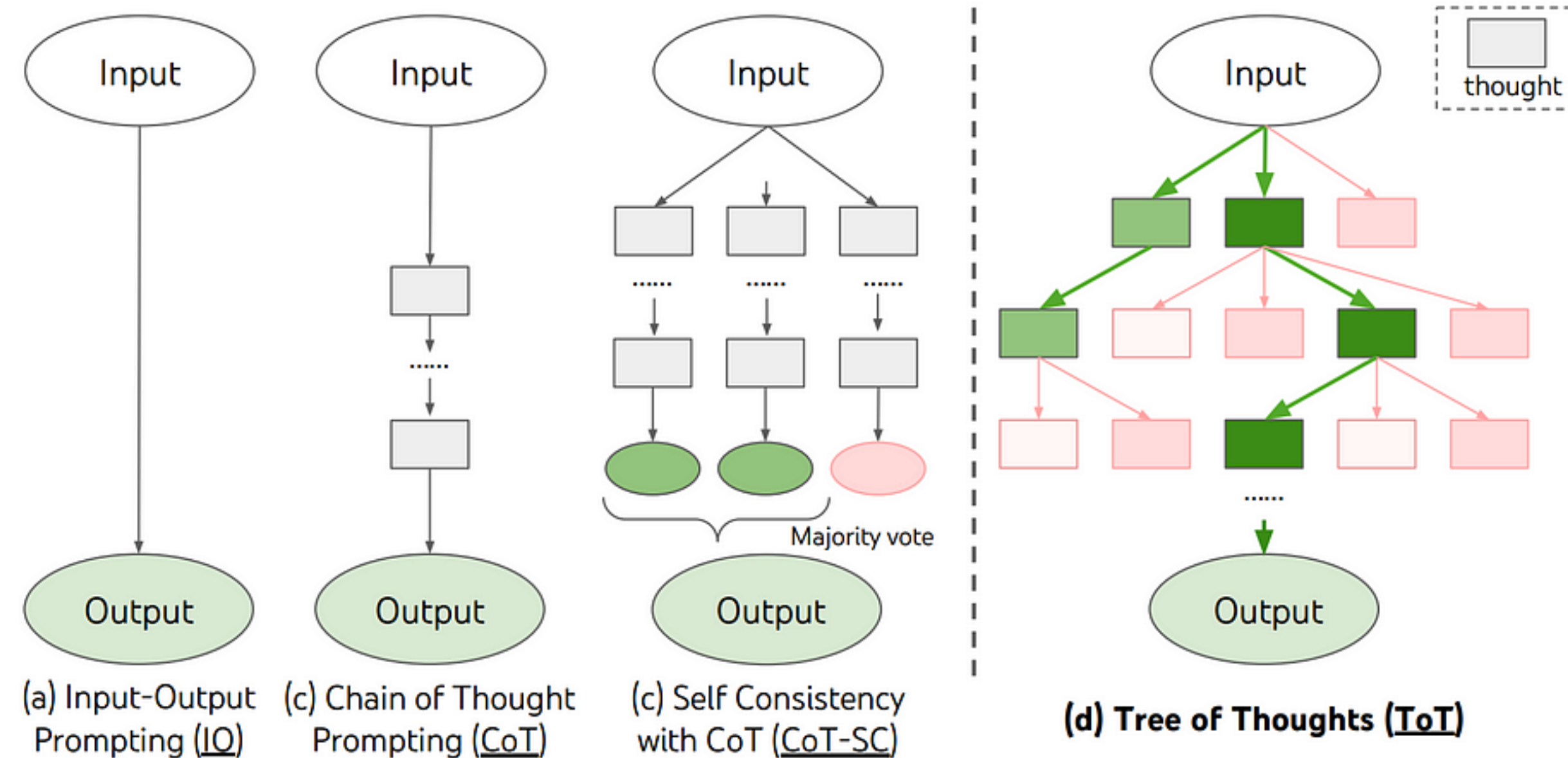


Figure 1: Schematic illustrating various approaches to problem solving with LLMs. Each rectangle box represents a *thought*, which is a coherent language sequence that serves as an intermediate step toward problem solving. See concrete examples of how thoughts are generated, evaluated, and searched in Figures 2,4,6.

<https://medium.com/@joycebirkins/x-of-thought-3-variants-of-the-s-chain-of-thoughts-cot-tot-got-lot-11a529d09dc1>


Prompting as Algorithmic Design

Multimodality: Beyond Text

Outline:

- Historical Roots: Chance, Systems, Early AI
- Evolution: Interactive Fiction to Early Neural Nets
- Contemporary Artistic Practices
- Future Trajectories: Agents, Reasoning, Multimodality
- **Models & Tools for Creators**
- Lab: Hands-on with Local LLMs

MODELS & TOOLS FOR CREATORS



Product Solutions Open Source Pricing

Search or jump to... /

Sign in Sign up

karpathy / char-rnn Public

Notifications Fork 2.6k Star 11.2k

<> Code Issues 90 Pull requests 21 Actions Projects Wiki Security Insights

master 1 branch 0 tags

Go to file Code

About

Multi-layer Recurrent Neural Networks (LSTM, GRU, RNN) for character-level language models in Torch

Readme Activity 11.2k stars 547 watching 2.6k forks Report repository

Releases

No releases published

Packages


No packages published

karpathy Merge pull request #164 from gdb/master 6f9487a on Apr 30, 2016 88 commits

data/tinyshakespeare	first commit	8 years ago
model	changing default LSTM initialization to use biases of 1.0 for the for...	8 years ago
util	Fix unclear errors	8 years ago
.gitignore	Add t7 files to .gitignore	8 years ago
Readme.md	Update Readme.md	7 years ago
convert_gpu_cpu_checkpoint.lua	fixing a bug introduced in previous commit. We have to use doubles ...	8 years ago
inspect_checkpoint.lua	add openc1 to sample.lua and inspect_checkpoint.lua, add link to clto...	8 years ago
sample.lua	fixing a bug introduced in previous commit. We have to use doubles ...	8 years ago
train.lua	fix message	8 years ago

Readme.md

<https://github.com/karpathy/char-rnn>



Product

Solutions

Open Source

Pricing

Search or jump to...

/

Sign in

Sign up

openai / gpt-2

Public

Notifications

Fork 5.1k

Star 19.9k

<> Code

Issues 120

Pull requests 34

Actions

Security

Insights


master

4 branches

0 tags

Go to file

Code



WuTheFWasThat move to azure

a74da5d on Dec 2, 2020 57 commits

src	nucleus sampling	4 years ago
.gitattributes	add .gitattributes file to ensure files copied to docker container ha...	5 years ago
.gitignore	updates for 345M model	4 years ago
CONTRIBUTORS.md	Update CONTRIBUTORS.md	4 years ago
DEVELOPERS.md	update readmes	4 years ago
Dockerfile.cpu	update readmes	4 years ago
Dockerfile.gpu	update readmes	4 years ago
LICENSE	delete	4 years ago
README.md	Update README.md	4 years ago
domains.txt	add model card	4 years ago
download_model.py	move to azure	3 years ago

About

Code for the paper "Language Models are Unsupervised Multitask Learners"

openai.com/blog/better-language-models

paper

Readme

View license

Activity

19.9k stars

630 watching

5.1k forks

Report repository

Releases

No releases published

<https://github.com/openai/gpt-2>

DMA171: Session 06-GPT2

File Edit View Insert Runtime Tools Help Changes will not be saved

Table of contents

- Session 6: Generative Text Modeling with Char RNN
 - GPU
 - Downloading GPT-2
 - Mounting Google Drive
 - Uploading a Text File to be Trained to Colaboratory
 - Finetune GPT-2
 - Load a Trained Model Checkpoint
 - Generate Text From The Trained Model
 - Generate Text From The Pretrained Model
- Etcetera
- LICENSE
- Section

Session 6: Generative Text Modeling with Char RNN

DMA171 - UCLA - Fall 2020
Parag K. Mital

This notebook is based off of the following source: Train a GPT-2 Text-Generating Model w/ GPU For Free by [Max Woolf](#)

For more about gpt-2-simple, you can visit [this GitHub repository](#). You can also read the original author's [blog post](#) for more information to use this notebook.

```
[ ] !tensorflow_version 1.x
!pip install -q gpt-2-simple
import gpt_2_simple as gpt2
from datetime import datetime
from google.colab import files
```

TensorFlow 1.x selected.
Building wheel for gpt-2-simple (setup.py) ... done
WARNING:tensorflow:
The TensorFlow contrib module will not be included in TensorFlow 2.0.
For more information, please see:
* <https://github.com/tensorflow/community/blob/master/rfcs/20180907-contrib-sunset.md>
* <https://github.com/tensorflow/addons>
* <https://github.com/tensorflow/io> (for I/O related ops)
If you depend on functionality not listed there, please file an issue.

Copy of CharRNN

```
[13]: def generate(decoded, prime_len=1, generate_len=10, temperature=1.0, code=False):
      hidden = decoder_init_hidden(batch_size)
      if not code:
          hidden = hidden_init()
          encoder_init_gpt2()
          loss = 0
          for i in range(generate_len):
              output, hidden = decoder_step(x=prime, hidden=hidden)
              loss += calculate_loss(output, label=prime[i:i+1], target=i+1)
          loss.backward()
          decoder_optimizer.step()
          return loss.detach() / generate_len
```

Generate Text

```
[18]: def generate(decoded, prime_len=1, generate_len=10, temperature=1.0, code=False):
      hidden = decoder_init_hidden()
      prime_token = Vocabulary.get_vocab_size('tokens')
      if not code:
          hidden = hidden_init()
          encoder_init_gpt2()
          prime_token = prime_token
          if not code:
              prime_token = prime_token
          if not code:
              prime_token = prime_token
```

S06: Generative Models for Text Generation

https://colab.research.google.com/drive/1ppqdl1UDoslp_S7MDo6nF9QuODDwi0f

<https://www.youtube.com/watch?v=sD7eqm4M20s&t=3991s>



ChatGPT 4

<https://chat.openai.com/>

Fine-tuning

Learn how to customize a model for your application.

Introduction

i This guide is intended for users of the new OpenAI fine-tuning API. If you are a legacy fine-tuning user, please refer to our [legacy fine-tuning guide](#).

Fine-tuning lets you get more out of the models available through the API by providing:

- 1 Higher quality results than prompting
- 2 Ability to train on more examples than can fit in a prompt
- 3 Token savings due to shorter prompts
- 4 Lower latency requests

GPT models have been pre-trained on a vast amount of text. To use the models effectively, we include instructions and sometimes several examples in a prompt. Using demonstrations to show how to perform a task is often called "few-shot learning."

Fine-tuning improves on few-shot learning by training on many more examples than can fit in the prompt, letting you achieve better results on a wide number of tasks. **Once a model has been fine-tuned, you won't need to provide as many examples in the prompt.** This saves costs and enables lower-latency requests.

<https://platform.openai.com/docs/guides/fine-tuning>

Chat completions API

Chat models take a list of messages as input and return a model-generated message as output. Although the chat format is designed to make multi-turn conversations easy, it's just as useful for single-turn tasks without any conversation.

An example Chat completions API call looks like the following:

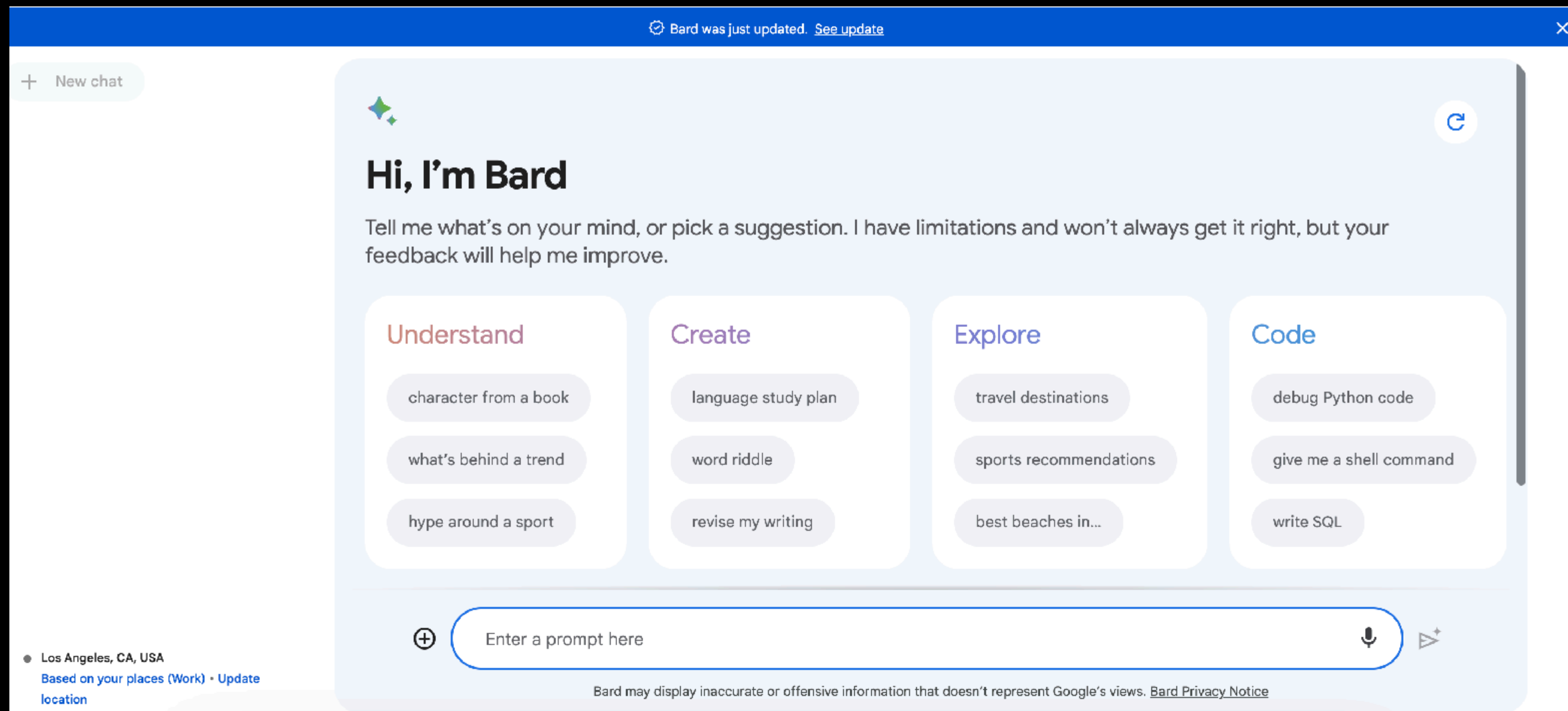
```
python Copy
1 response = openai.ChatCompletion.create(
2     model="gpt-3.5-turbo",
3     messages=[
4         {"role": "system", "content": "You are a helpful assistant."},
5         {"role": "user", "content": "Who won the world series in 2020?"},
6         {"role": "assistant", "content": "The Los Angeles Dodgers won the World Series in 20"},
7         {"role": "user", "content": "Where was it played?"}
8     ]
9 )
```

To learn more, you can view the full [API reference documentation](#) for the Chat API.

The main input is the messages parameter. Messages must be an array of message objects, where each object has a role (either "system", "user", or "assistant") and content. Conversations can be as short as one message or many back and forth turns.

Typically, a conversation is formatted with a system message first, followed by alternating user and assistant messages.

<https://platform.openai.com/docs/guides/gpt/chat-completions-api>



<https://bard.google.com/>

ANTHROPIC

Meet Claude

Message Claude or search past chats...



Start a new chat ➤


Try these

Example: "Summarize this PDF document"

Example: "Help me practice my Spanish vocab"

Example: "Explain how this python game works"

<https://claude.ai/chats>

 **Hugging Face**

Models

Datasets

Spaces


Docs

Solutions

Pricing

Log In

Sign Up

 **TheBloke / Llama-2-7B-Chat-GGML**

like 529

Text Generation

Transformers

PyTorch

English

llama

facebook

meta

llama-2

text-generation-inference

arxiv:2307.09288

License: other

Model card

Files and versions

Community 29

Train

Deploy

Use in Transformers

main Llama-2-7B-Chat-GGML 3 contributors History: 30 commits

TheBloke	Upload README.md	00109c5	10 days ago
.gitattributes	1.52 kB	initial commit	about 2 months ago
LICENSE	7.02 kB	Upload folder using huggingface_hub	about 2 months ago
Notice	112 Bytes	Add license files	about 2 months ago
README.md	23 kB	Upload README.md	10 days ago
USE_POLICY.md	4.77 kB	Upload folder using huggingface_hub	about 2 months ago
config.json	29 Bytes	Initial GGML model commit	about 2 months ago
llama-2-7b-chat.ggmlv3.q2_K.bin	2.87 GB LFS	Initial GGML model commit	about 2 months ago
llama-2-7b-chat.ggmlv3.q3_K_L.bin	3.6 GB LFS	Initial GGML model commit	about 2 months ago

<https://huggingface.co/TheBloke/Llama-2-7B-Chat-GGML/tree/main>

llama.cpp [↗](#)



🔄 CI **failing** license MIT

[Roadmap](#) / [Project status](#) / [Manifesto](#) / [ggml](#)

Inference of [LLaMA](#) model in pure C/C++

Hot topics [↗](#)

- 🚨 BPE tokenizer update: existing Falcon and Starcoder `gguf` models will need to be reconverted: [#3252](#)

<https://github.com/ggerganov/llama.cpp>

Customize Llama's personality by clicking the settings button.

I can explain concepts, write poems and code,
solve logic puzzles, or even name your pets.

Send me a message.

Send a message

Chat

<https://www.llama2.ai/>

Spaces

HuggingFaceH4/open_llm_leaderboard

like5.52k

Running

App

Files

Community305

🤖

Open LLM Leaderboard

📌

The 🤖 Open LLM Leaderboard aims to track, rank and evaluate open LLMs and chatbots.

🤖

Submit a model for automated evaluation on the 🤖 GPU cluster on the "Submit" page! The leaderboard's backend runs the great [Eleuther AI Language Model Evaluation Harness](#) - read more details in the "About" page!

🏆 LLM Benchmark

📖 About

🚀 Submit here!

🔍 Search for your model and press ENTER...

Select columns to show

☒ Average 📈

☒ ARC

☒ HellaSwag

☒ MMLU

☒ TruthfulQA

☐ Type

☐ Precision

☐ Hub License

☐ #Params (B)

☐ Hub ❤️

☐ Model sha

☒ Show gated/private/deleted models

Model types

☒ 🟢 pretrained

☒ 🟡 fine-tuned

☒ 🔴 instruction-tuned

☒ 🟦 RL-tuned

☒ ? Unknown

Precision

☒ torch.float16

☒ torch.bfloat16

☒ torch.float32

☒ 8bit

☒ 4bit

☒ GPTQ

Model sizes

☒ Unknown

☒ < 1.5B

☒ ~3B

☒ ~7B

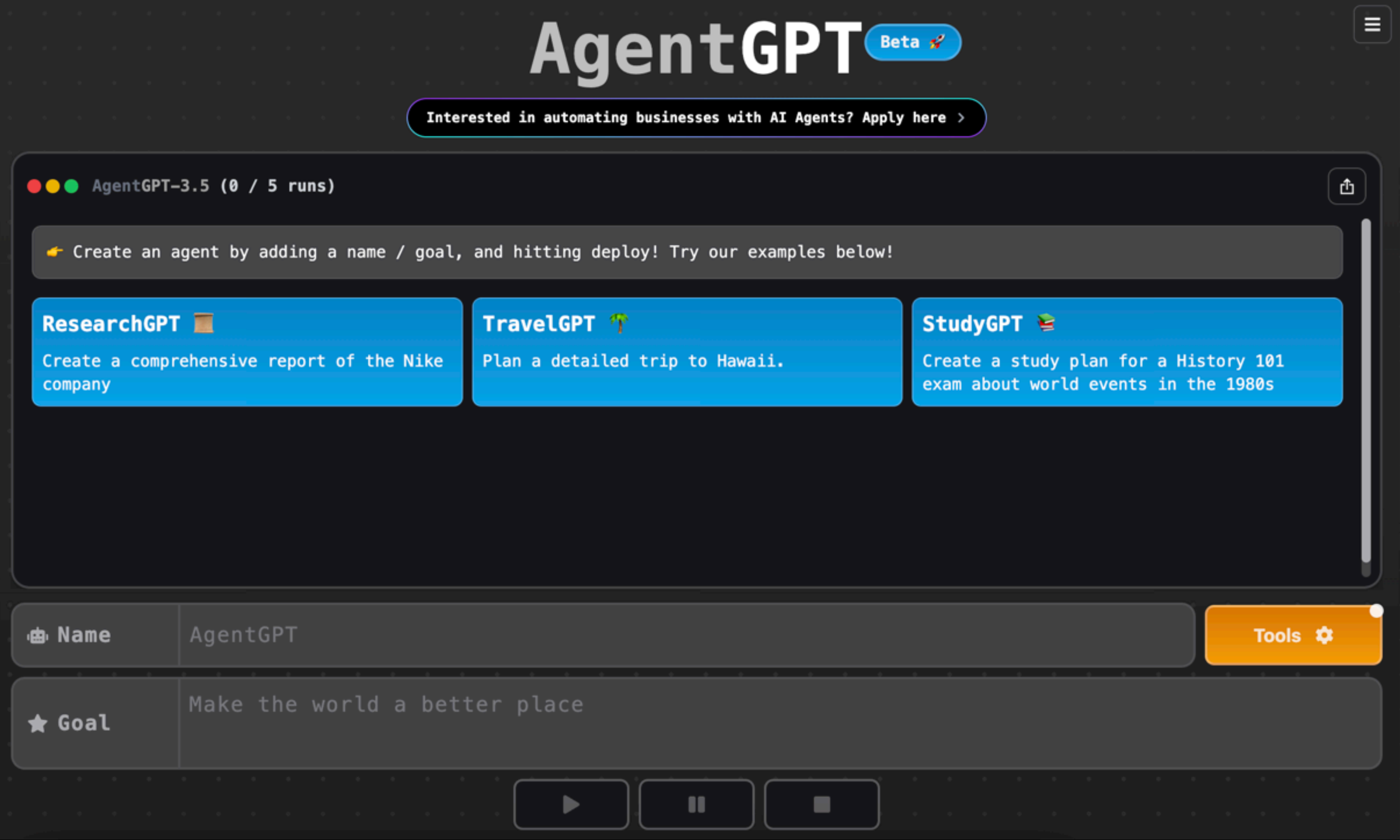
☒ ~13B

☒ ~35B

☒ 60B+

https://huggingface.co/spaces/HuggingFaceH4/open_llm_leaderboard

Feature	Google Gemini 1.5/2.5 Pro	Anthropic Claude 3 / 3.5 / 3.7	Meta Llama 3.1
Key Model(s)	1.5 Pro (API/Adv.), 2.5 Pro (Exp. Adv.)	Sonnet (3/3.5/3.7), Opus (3)	8B, 70B, 405B
Max Context (Tokens)	~1M - 2M	~200k	~128k
Multimodality?	Yes (Text, Image, Audio, Video)	Yes (Text, Image)	Primarily Text (Vision planned)
Access	Web UI (Free/Adv.), API	Web UI, API	Download, API Providers, Web UI
Open Source?	No	No	Yes
Key Creative Strengths	Massive context, Native multimodality, Research assistance (2.5 Pro)	Strong reasoning & writing, Safety focus, Balanced options (Haiku/Sonnet/Opus)	Local control, Customization, Cost-effective (no API fees for local), High performance (70B/405B)
Potential Use Cases	Long-form analysis/generation (text/video/audio), Multimodal projects, In-depth research	Complex creative writing, Dialogue generation, Code generation, Ethical AI exploration	Local experimentation, Fine-tuning on custom data, Building custom tools/apps, Offline work



<https://agentgpt.reworkd.ai/>

Select dataset

ai_society_chat

×

ASSISTANT

Financial Advisor


×

USER

Actor

×

CAMEL



Original task

Developing a debt management plan to reduce debt and improve credit score.

×

Specified task

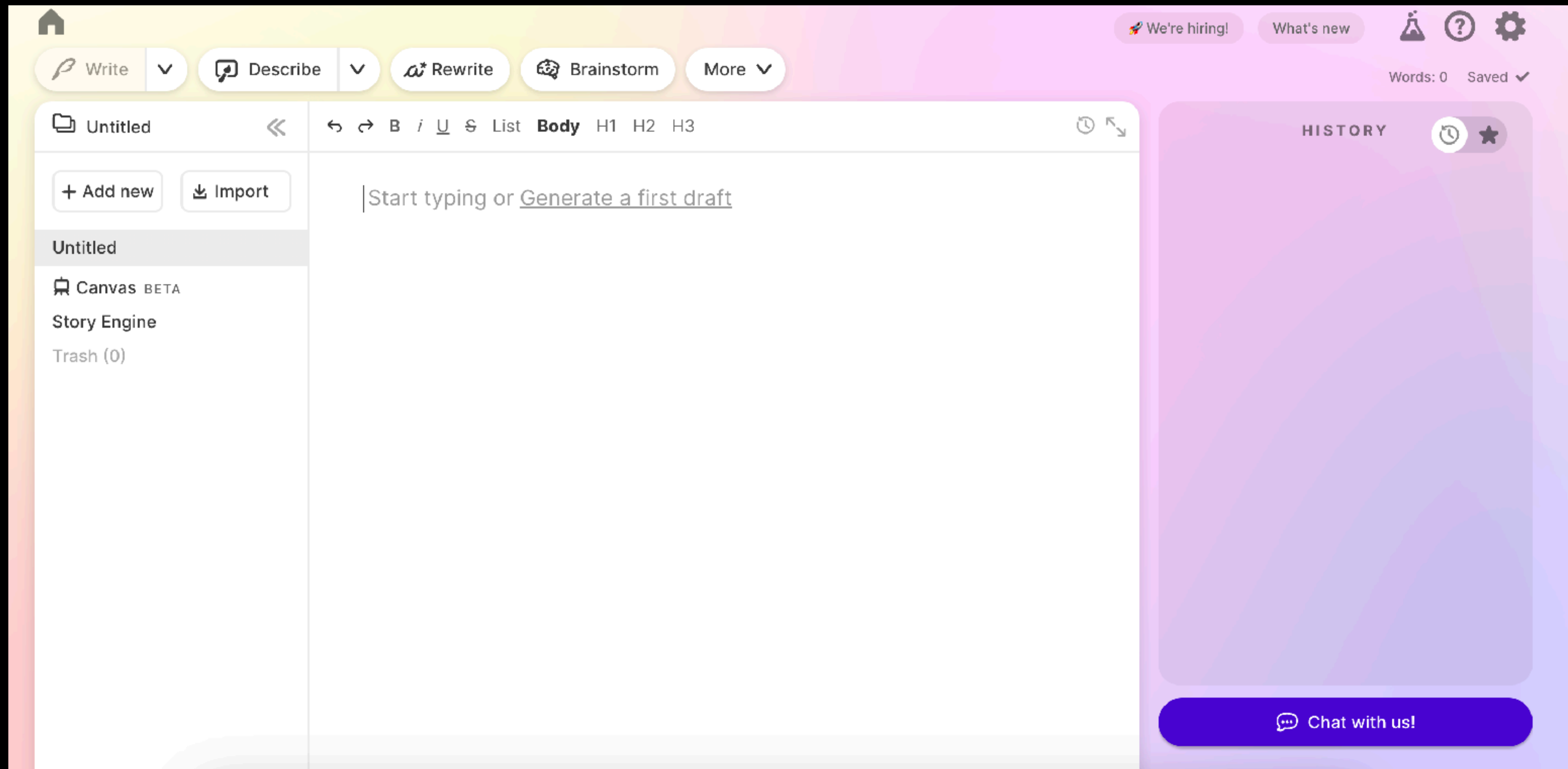
The Financial Advisor will create a detailed 12-month plan with specific targets to guide the Actor towards attaining realistic and achievable goals for debt restructuring by outlining viable consolidation options, settlement adjustments, negotiating existing highly-leveraged financial obligations possessing modified payment design where feasible to secure improved debits score rating certainty footing.

Chatbot

Instruction: Please provide a detailed analysis of the Actor's current debt situation.
Input: The Actor has a total debt of \$50,000, consisting of credit card debt, personal loans, and a car loan. The Actor's credit score is currently at 600.

Solution: Based on the information provided, the Actor's current debt situation is as follows:

<http://data.camel-ai.org/>



<https://www.sudowrite.com/>

LM Studio

Discover, download, and run local LLMs

Supports MetaAI's Llama 2 🦙 (7B, 13B, 34B, 70B)

Run any LLaMa Falcon MPT StarCoder Replit GPT-Neo-X ggml ① models from Hugging Face

Download LM Studio for Mac (M1/M2)

v0.2.6

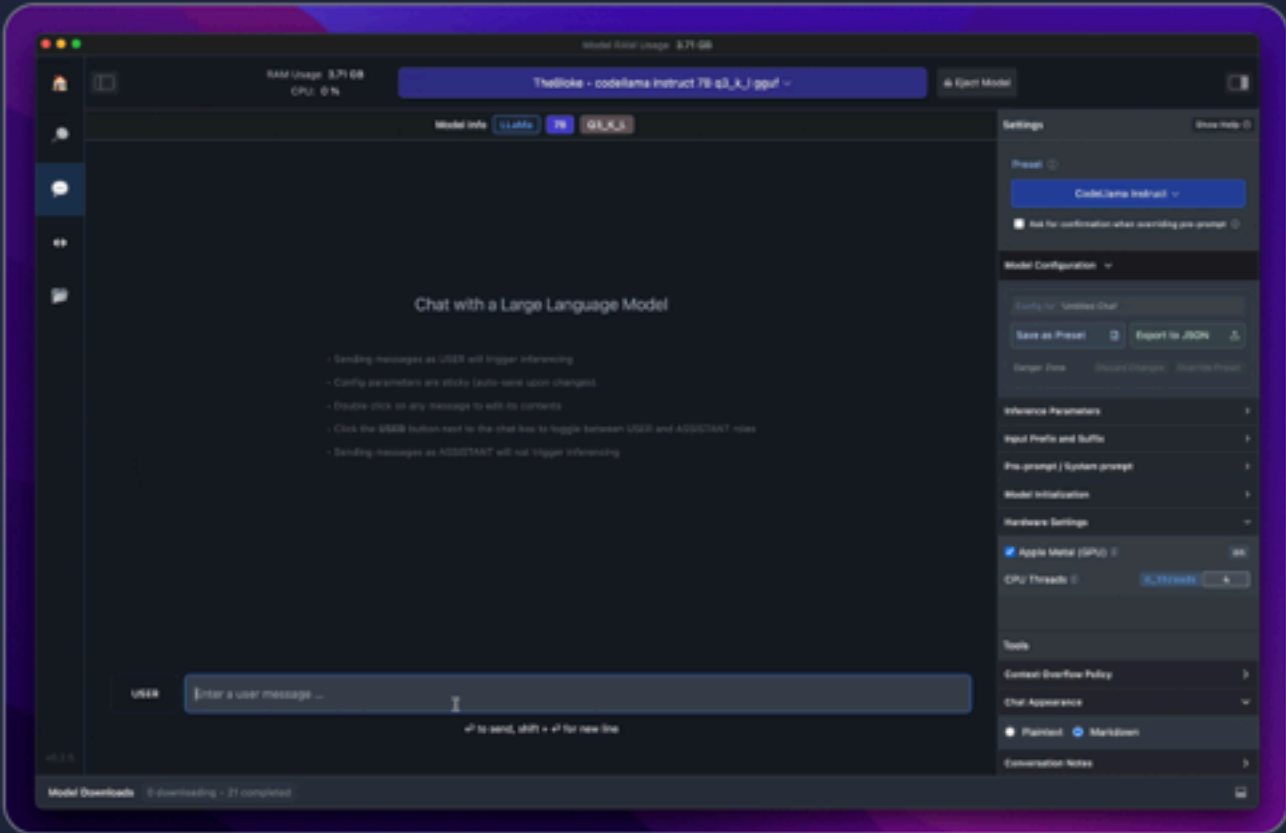
Download LM Studio for Windows

v0.2.6

LM Studio is provided for personal use under the [terms](#).
For business use, please [get in touch](#).

Sign up for new version email updates

New in v0.2.6: App update flow redesign, bug fixes, and "stop strings" in the UI.



Twitter

Github

Discord

Email

We are hiring a C++ / Systems Engineer. See [how to apply](#).


With LM Studio, you can

<https://lmstudio.ai/>

79

PARAG K MITAL | UCLA DMA

CULTURAL AUTOMATION WITH MACHINE LEARNING | SESSION 04: GENERATIVE TEXT

 **Hugging Face**

Search models, datasets, users...

Models

Datasets

Spaces


Docs

Solutions

Pricing


Log In


Sign Up


 **ffiloni's Collections**

Historic Top Trending Demos

Text-to-Image History

 **AI Games**

 The MusicBox


 SD-XL Training Suite

The ControlNet Saga

Music [OP]


Video [OP]

Utilities

 **AI Games**

updated 9 days ago

▲ Upvote 4



↑ Share collection

🕒 View history

📖 Collection guide

Nob Hill Noir

♡ 23

Note A Large Language Murder | You are a detective and you have to solve a murder by asking PNJs (but they want to keep their secrets 😊)

Wordalle

♡ 127

Note GUESS THE PROMPT!

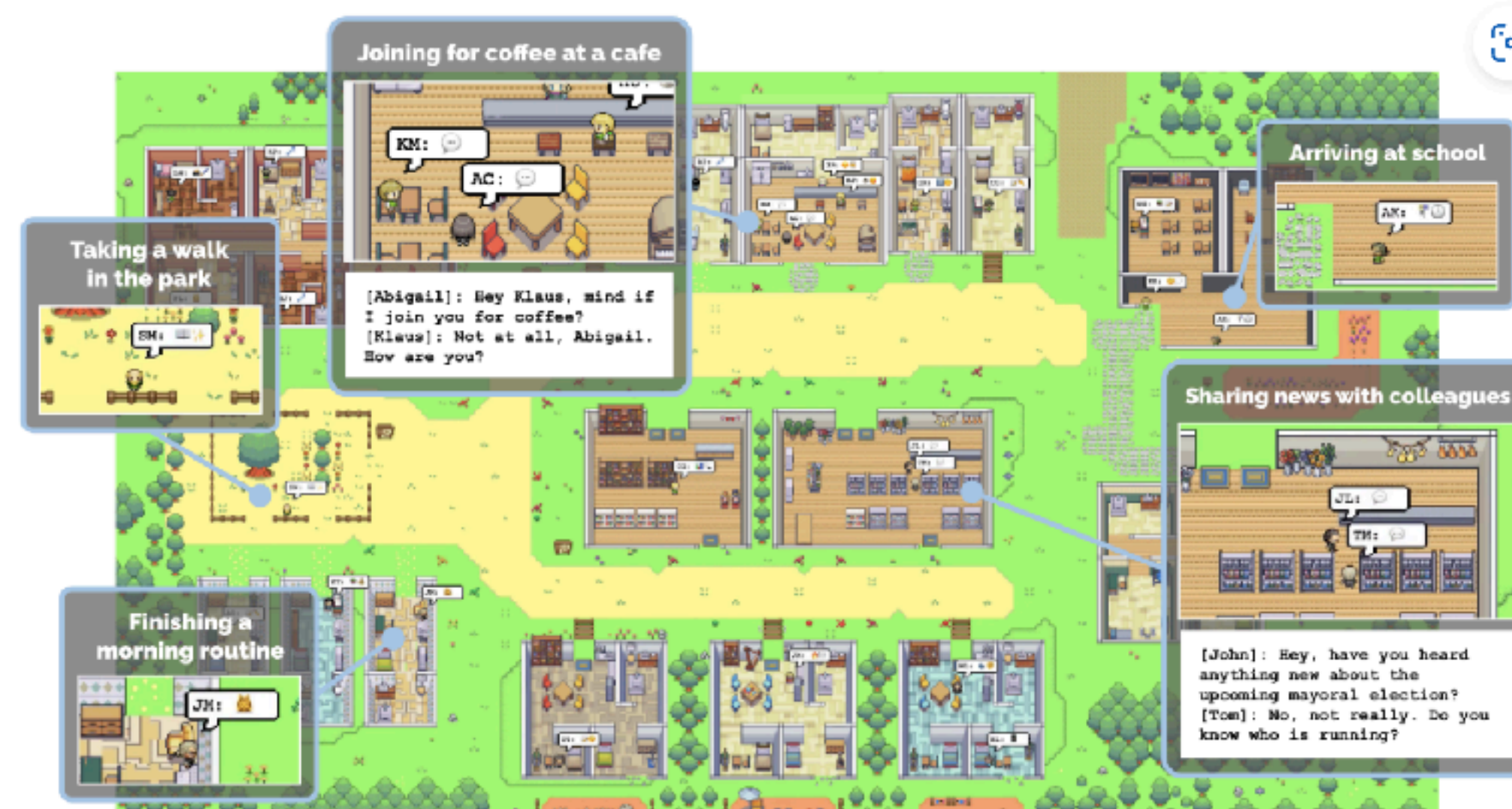
Doodle Dash

♡ 27

Note How fast can a neural network predict your doodles?

<https://huggingface.co/collections/ffiloni/ai-games-64fda15e0e486522f868432d>

Generative Agents: Interactive Simulacra of Human Behavior [↗](#)



This repository accompanies our research paper titled "[Generative Agents: Interactive Simulacra of Human Behavior](#)." It contains our core simulation module for generative agents—computational agents that simulate believable human behaviors—and their game environment. Below, we document the steps for setting up the simulation environment on your local machine and for replaying the simulation as a demo animation.

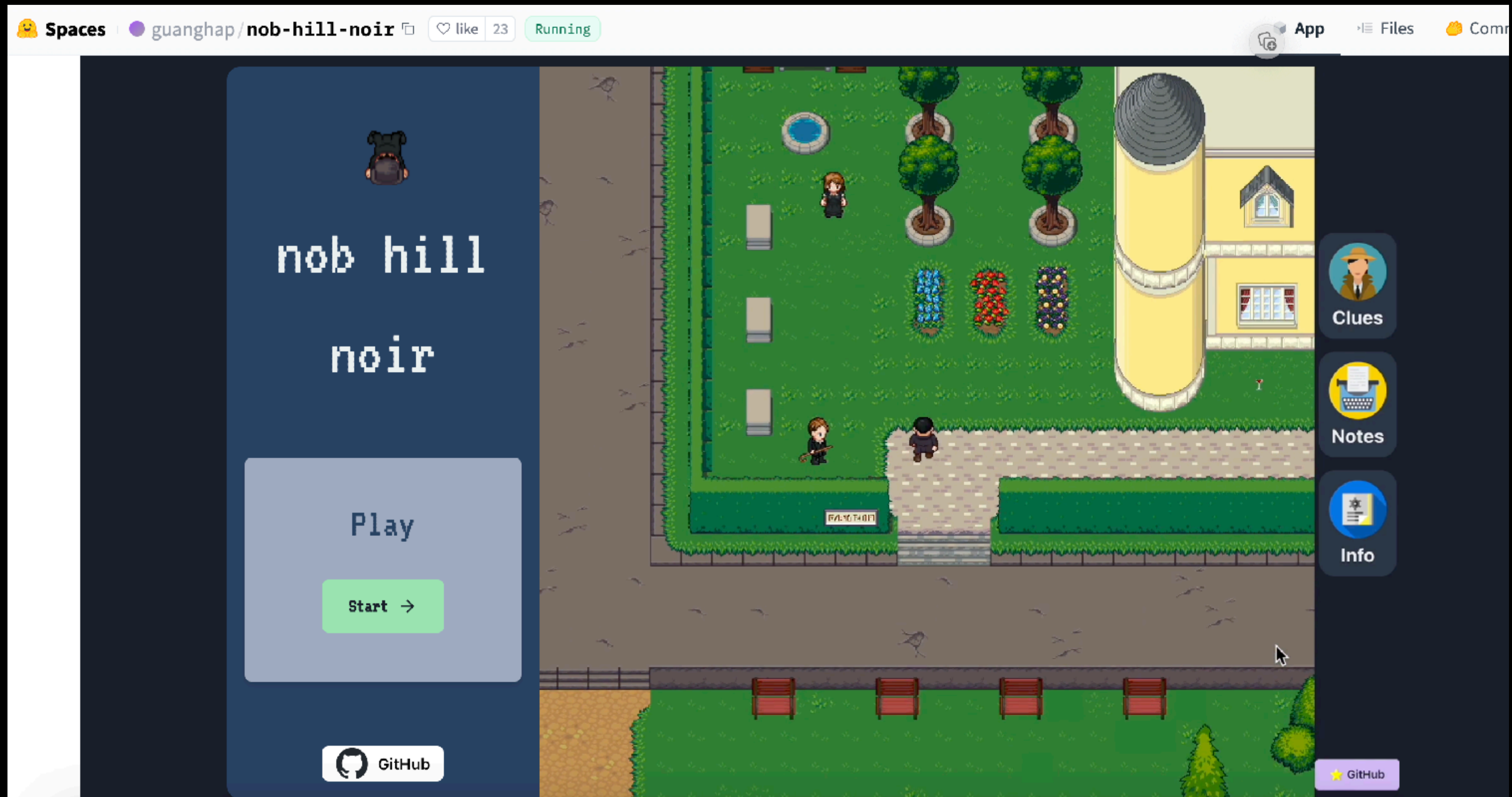
https://github.com/joonspk-research/generative_agents

[Join our community Discord: AI Stack Devs](#)



AI Town is a virtual town where AI characters live, chat and socialize.

<https://github.com/a16z-infra/ai-town>



<https://huggingface.co/spaces/guanghap/nob-hill-noir>

Further Approaches

- Text generation: These tools use AI algorithms, including GPT-3 and Transformer models, to generate high-quality text content such as articles, blog posts, and social media updates. Popular examples include Copy.ai, Jarvis.ai, and Qordoba.
- **Text summarization:** These tools use AI to summarize long-form content into shorter, more readable formats. Popular examples include SummarizeBot, SMMRY, and TextTeaser.
- **Language translation:** These tools use AI algorithms to translate text content from one language to another. Popular examples include Google Translate, DeepL, and Microsoft Translator.
- **Copy.ai:** An AI-powered tool that generates product descriptions, copywriting, blog posts, and much more
- **Wordsmith:** A tool that uses AI to generate written reports in a variety of styles and formats, as well as personalized email or letter copy
- **Quillbot:** A tool that uses AI to reword and simplify text to make it easier to read and understand
- **Jarvis.ai:** A one-stop-shop for content creation that uses AI to generate several types of content like blogs, articles, and social media posts

Further Reading

- <https://blog.langchain.dev/agents-round/> "Autonomous Agents & Agent Simulations"
- <https://towardsdatascience.com/4-autonomous-ai-agents-you-need-to-know-d612a643fa92> "4 Autonomous AI Agents you need to know"
- https://ai4comm.media.mit.edu/?utm_source=substack&utm_medium=email
- https://fablestudio.github.io/showrunner-agents/static/pdfs/To_Infinity_and_Beyond_SHOW-1_And_Showrunner_Agents_in_Multi_Agent_Simulations.pdf South Park Simulator

More Resources / Advanced Text Generation

- https://huggingface.co/spaces/HuggingFaceH4/open_llm_leaderboard Hugging Face LLM Leaderboard
- [https://huggingface.co/Salesforce/xgen-7b-8k-base?](https://huggingface.co/Salesforce/xgen-7b-8k-base?ref=blog.salesforceairesearch.com)
[ref=blog.salesforceairesearch.com](https://huggingface.co/Salesforce/xgen-7b-8k-base?ref=blog.salesforceairesearch.com) XGen from Salesforce
- <https://huggingface.co/spaces/julien-c/nllb-translation-in-browser>

Outline:

- Historical Roots: Chance, Systems, Early AI
- Evolution: Interactive Fiction to Early Neural Nets
- Contemporary Artistic Practices
- Future Trajectories: Agents, Reasoning, Multimodality
- Models & Tools for Creators
- **Lab: Hands-on with Local LLMs**

LAB: HANDS-ON WITH LOCAL LLMS

Chats

Secret project

C++ Filesystem... (1600 tokens) ...

log about version of ... (56 tokens)

Financial analysis (106 tokens)

lmstudio-community/Meta-Llama-3.1-8B-Instruct-GGUF

Eject

C++ Filesystem Implementation

AssistantMeta-Llama-3.1-8B-Instruct-GGUF

Filesystem Implementation in C++

Before we begin, let's outline our approach:

1. We'll create a basic directory structure with methods for creating directories, deleting directories, listing files and directories, and navigating through the directory tree.

2. We'll implement file operations such as creating, reading, writing, and deleting files.

3. We'll handle errors and exceptions to ensure robustness.

Filesystem Class

```
#include <iostream>
#include <fstream>
#include <string>
#include <vector>
#include <dirent.h>

class Filesystem {
public:
    // Constructor
    Filesystem(const std::string& rootDir) : rootDir_(rootDir) {}

    // Create a new directory
    void createDirectory(const std::string& path);
```

Type a message and press Enter to send ...

User (%%U)Insert (%%I)

Context is 39.1% full

Advanced Configuration

Preset

Discard Unsaved

Coding Helper (C++)

Commit ChangesSave As New...

System Prompt ?

You are an incredibly good C++ engineer. Think about the problems you're about to solve step-by-step. First make a plan, and then ask the user to confirm. Only then act on it.

Token count: 41 ?

Settings>

Sampling>

Structured Output>

Conversation Notes ?>

LM Studio 0.3.4 (Build 5)

UserPower UserDeveloper

SYSTEM RESOURCES USAGE: RAM: 4.47 GB | CPU: 0.00 %

HOMEWORK

Homework

- Explore the use of text generation
- Create an artwork, story, script, idea, essay, etc.. related to your interests / topic for the course
- This work should be able to be presented as a standalone project, but should ideally contribute towards the development of your final project
- Ideally build on the world you pitched last week, and towards development of your final project
- Presentation format is open to student preference, but must be limited to **5 minutes** to allow time for feedback from crit advisors