Introduction to Python and Poetry

Digital Integration Teaching Initiative

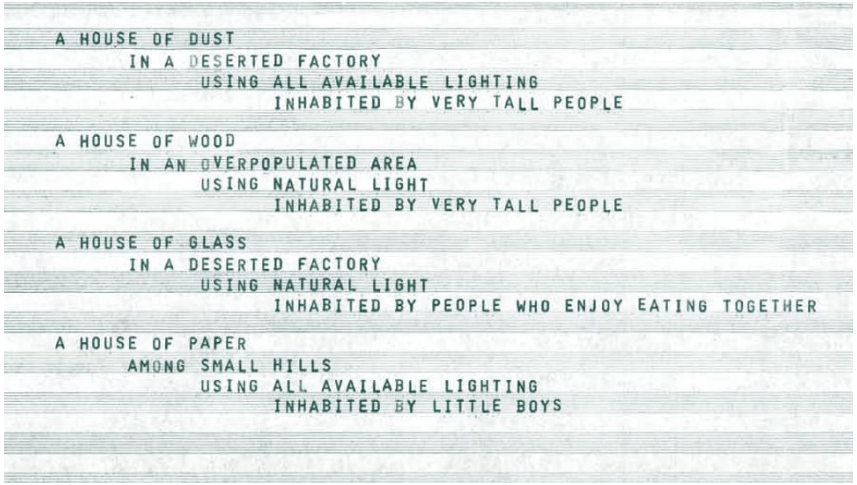
Northeastern University
NULab for Digital Humanities and Computational Social Science

Workshop Agenda

- Computational poetry example and discussion
- Introduction to Python and Google Colab
- Generative AI
- AI Ethics
- Discussion

Example: “The House of Dust”

- House of Dust by Alison Knowles and James Tenney (1967)
- Code reimplemented in Python by Nick Montfort and updated as a teaching example: TheHouseOfDust_ExampleComputationalPoem.ipynb



Printout of “The House of Dust,” Gebr König Verlag, Cologne, 1967. Found in *Art by Translation*.

“The House of Dust” Discussion

- Based on the code for “The House of Dust”, what are the four main building blocks of the poem?
- Can you tell which decisions were made by the author and which are random?
- Can you describe the process of how this poem was written?

Other Computational Poems

- “Stochastische Texte" by Theo Lutz (1959) found in "Exploring (Semantic) Space With (Literal) Robots" by Allison Parrish (2015)
- "House of Dust" by Alison Knowles and James Tenney (1967), code reimplemented in Python by Nick Montfort
- “A Travesty Generator for Micros” by Hugh Kenner and Joseph O’Rourke (1984)
- *Travesty Generator* by Lillian-Yvonne Bertram (2019)
- "won't you" by Nick Montfort (2019)
- "Compass Poems" by Allison Parrish (2020)

Python & Google Colaboratory

Northeastern University
NULab for Digital Humanities and Computational Social Science

Python Summary

The Python code in this module covers the below topics:

- Variables
- Print function
- Import random module
- Data types
  - Strings (Text)
  - Lists
  - Dictionaries
- Selecting text from lists and dictionaries

Python Google Colab Notebooks

Please refer to the below notebooks to learn more about Python and Google Colab.

- Colab&IntroToPythonPoetry_Lesson.ipynb: This notebook introduces the fundamentals of Python and provides example code for creating computational poetry.
- PythonPoetry_ComputationalPoemTemplate.ipynb: This notebook is a template with some starter code to help you create your own computational poem.

Generative AI

Northeastern University
NULab for Digital Humanities and
Computational Social Science

Important AI Vocabulary (1/2)

- Artificial Intelligence (AI): A "technology that enables computers and machines to simulate human learning, comprehension, problem solving, decision making, creativity and autonomy."
- Supervised machine learning: "[A] machine learning technique that uses labeled data sets to train artificial intelligence (AI) models to identify the underlying patterns and relationships."
- Unsupervised machine learning: A machine learning technique that uses "algorithms to analyze and cluster unlabeled data sets."

Important AI Vocabulary (2/2)

- Generative AI: “AI technologies that can create new content, ideas, or data that are coherent and plausible, often resembling human-generated outputs.”
- Markov Chains: “[A] stochastic process describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event.”
- Text Embedding: A numerical representation of the meaning of a word.

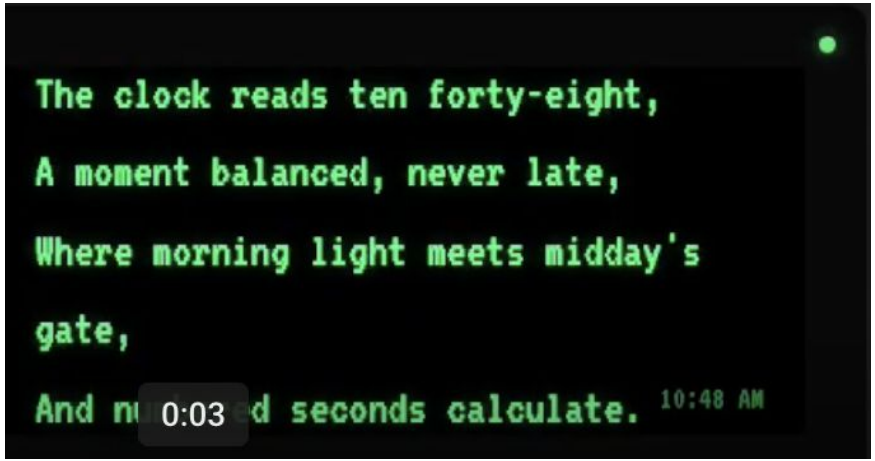
Generative AI Summary

- Uses unsupervised machine learning and other computational methods, such as Markov chains and embeddings, to learn how to generate content
- The type of dataset used to develop the generative AI determines what it can do

Image by DALL-E 3 found in "Text Embeddings: Comprehensive Guide" by Mariya Mansurova

Example: AI Rhyming Clock

- Claude's AI Rhyming Clock
  - Uses generative AI to write a poem integrating the current time
- Try going to the link above and clicking customize to adjust the prompt generating the poems



AI Ethics

Northeastern University
NULab for Digital Humanities and
Computational Social Science

Ethics: “Originality”

- Some argue that all AI-generated output constitutes plagiarism and copyright infringement, since it is remixing training data that was scraped from the internet without permission from the original creators.
- Many AI companies are facing lawsuits from people whose content was used as training data without their consent.
- Some publication venues, like the *Science* journals, have made it an official policy that AI does not meet the standard for authorship and require authors to disclose use of AI.

Ethics: Training

- AI training data is sometimes supplemented by labels (annotations) added by people (Amironesei and Díaz, 2024). These labels can worsen bias in training datasets.
- People from middle and low income countries often labor in poor working conditions to annotate data for clients in high income countries. Fieldwork by Muldoon et al (2023) revealed that workers faced traumatizing content, in addition to experiencing discrimination in the workplace and receiving low wages.
- Public awareness can help pressure companies to adopt good practices. For more information on fair labor in AI, see the report AI for Fair Work: From principles to practices by *Fairwork*

Ethics: Bias

- AI outputs can be biased due to human biases that exist in the original training data or decisions around how the model was developed.
- For example, when AI was used to summarize medical notes, “Google’s AI tool Gemma described men’s health issues with terms like “disabled,” “unable,” and “complex” significantly more often than women’s, who were often framed as more independent despite similar needs, an alarming gender bias trend.” (Source: crescendo.ai)

Ethics: Environment

- Training and using AI requires processing very large amounts of data, which is done in data centers
- These data centers can have a negative impact on the environment and communities
- Given their intensive energy and water demands, data centers can worsen local water scarcity and increase electricity prices
- To explore how the energy use of AI compares to other digital tasks, check out Jon Ippolito's "What Uses More" app

What You Can Do

- By using smaller models, where possible, and running fewer prompts you may be able to reduce your contribution to AI's environmental impact.
- You can also integrate tools such as CodeCarbon into your code to track and minimize the carbon dioxide emitted from computers executing the code.
- To help counter the spread of misinformation, or simply misleading content, it is important to evaluate AI output against other sources and carefully consider whether certain outputs might be biased.

AI Ethics Resources

- The Institute for Experiential AI at Northeastern
- United Nations Educational, Scientific and Cultural Organization (UNESCO) Global Forum on the Ethics of AI 2024

Image found in Changing the Landscape of AI Governance, UNESCO

Discussion

Post-exploration group discussion

- Do you have any reflections on using Python for constructing poetry?
- How does the writing process differ from traditional poetry?
- How might the code impact the readers perception of the poem?
- How might you use this in the future?

Thank you!

—Developed by Sara Morrell, Dipa Desai, Emily Sullivan and Kasya O'Connor Grant

- For more information on the DITI, please see: https://bit.ly/diti-about
- Schedule an appointment with us! https://bit.ly/diti-meeting
- If you have any questions, contact us at: nulab.info@gmail.com

Northeastern University
NULab for Digital Humanities and Computational Social Science

Learn More

- Bertram, Lillian-Yvonne. *Travesty Generator*. Noemi Press, 2019.
- Turkel, William J., and Adam Crymble. "Manipulating Strings in Python." *Programming Historian*, 17 July 2012.
- Santillan, Marvin C., and Arnulfo P. Azcarraga. "Poem generation using transformers and doc2vec embeddings." *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2020.