Cole McKinley
Jeff Henkel
Anna Ferguson
Derek Colombo

Proposal Draft 1

Project Title: The Art Factory

Project Pitch:

The question we want to have answered is why aren't there more routes of creativity in the arts online? Learning technology through arts is another great way to understand the internet and the computing systems that come with it.

Our project involves a user selected dataset that will then be used to develop and generate digital art. A website will be created by the group with datasets analyzed by us. The user will choose a dataset preview in digital form, and an algorithm will generate its own image based on what it has learned

We'd really appreciate it if you would be able to help and mentor our group for our capstone because we know you have plenty of experience with computing and the arts. We feel like you could bridge the gap between our skills and what we hope to accomplish with our generative art project.

Names and general skills:

Python, Software Development: Anna, Derek, Cole & Jeff

Python & R, Data Analysis and Manipulation: Anna & Derek

Web Development, User Experience: Cole, Jeff, & Anna

SQL, Server management and data storage/access: Derek & Jeff. We may have time to set up separate server for data storage

Data Ethics Checklist:

Please fill out the following checklist for during general use or after a data breach.

#	Question	Generally	Data Breach	Example
1	Could a user sell drugs or other illegal items on your platform?	N	N	https://academicworks.cuny.edu/cgi/viewcontent.cgi?article=1072&context=sph_pubs
2	Could a user of your platform engage in sex trafficking?	N	N	https://stmuscholars.org/craigslist-and-backpage-sex-trafficking-at-your-fingertips/
3	Could a user sell class notes or cheat on their homework on your platform?	N	N	https://www.edsurge.com/news/2021-02-23-more-students-are-using-chegg-to-cheat-is-the-company-doing-enough-to-stop-it
4	Could a stalker use your project to find someone?	N	N	https://www.nytimes.com/2018/05/19/technology/phone-apps-stalking.html
5	Could your app be used to spy on or track individuals?	N	N	https://www.nytimes.com/2019/12/22/us/politics/totok-app-uae.html
6	Could your app/software access the camera or microphone and record things	N	N	https://gizmodo.com/these-academics-spent-the-last-year-testing-whether-you-1826961188

	without users being aware?			
7	If someone uses your platform, could they be re-traumatized or have their mental health impacted in some way?	N	N	https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739
8	Could your algorithm promote material that would traumatize or upset individuals?	M	N	https://www.theguardian.com/technology/2021/oct/16/tiktok-eating-disorder-thinspo-teens
9	Would your users be upset if the data you collect was given to someone else?	N	N	https://www.businessinsider.com/stolen-data-of-533-million-facebook-users-leaked-online-2021-4
10	Could a data leak potentially lead to identity theft?	N	N	https://www.ftc.gov/enforcement/cases-proceedings/refunds/equifax-data-breach-settlement
11	If your site was hacked, would users of that product potentially	N	N	https://www.forbes.com/sites/zakdoffman/2019/08/23/ashley-madison-is-back-with-30-million-cheating-spouses-signed-since-the-hack/?sh=22f1ba5c3878

	lose their job, spouse, or family?			
12	Should there be an age limitation on your product?	Y	Y	https://www.bbc.com/news/technology-48925623
13	Could someone use your product to find, contact, and potentially commit elder abuse?	N	N	https://www.nbcnews.com/health/aging/genetic-testing-scam-targets-seniors-rips-medicare-n1037186
14	If the data on your platform was breached, could it be used to blackmail the users?	N	N	https://www.wired.com/story/parler-hack-data-public-posts-images-video/
15	Does the existence of your project imply that a particular racial group, gender, religion or other protected category is inherently bad, gross, or unwanted?	N	N	https://www.distractify.com/p/pinky-gloves-dragged

16	Could your product be used to commit hate crimes against a specific group?	N	N	https://ibmandtheholocaust.com/
17	Does the primary content of your game or algorithm focus on something considered deeply unethical?	N	N	https://www.quora.com/What-is-the-most-unethical-video-game-ever-created
18	Does your game or software contain race, gender, or other stereotypes?	N	N	https://en.wikipedia.org/wiki/List_of_controversial_video_games
19	Could users of your app scam other individuals?	N	N	https://dailyiowan.com/2021/06/21/opinion-kickstarter-scams-are-on-the-rise/
20	Is your particular algorithm biased towards predicting correctly only for one race, gender, or other group?	N	N	https://www.theguardian.com/technology/2020/sep/21/twitter-apologises-for-racist-image-cropping-algorithm

21	Are the users of your project, players of your game, or those being surveyed for your data aware of how their data will be used?	M	M	https://www.computerweekly.com/news/252464048/Many-search-engine-users-unaware-of-personal-data-collection
22	What are the possible misinterpretations of your results? For example - would a white supremacist or misogynist be stoked about your results if they misinterpreted it?	N	N	https://www.nature.com/articles/s41467-020-19723-8
23	Does the use or purchase of your data potentially contribute to a dangerous group or regime?	N	N	https://vertpaleo.org/svp-sends-letter-to-paleontological-community-on-myanmar-amber/
24	Could your virtual reality environment cause injury to the user?	M	M	https://bonejoint.net/blog/eight-things-you-should-know-about-virtual-reality/

25	Are your study participants or game players aware that their data will be collected and used?	Y	Y	https://www.polygon.com/features/2019/5/9/18522937/video-game-privacy-player-data-collection
26	Does your game or app contain addictive design elements without benefit to the user?	N	N	https://searchsoftwarequality.techtarget.com/tip/5-examples-of-ethical-issues-in-software-development
27	Does your survey contain an aspect of compulsion or unusually large incentive, that would command users to take it even if it was to their detriment?	N	N	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4214066/
28	Could your research outcomes harm an individual or entity?	N	N	https://rutgersaaup.org/rutgers-budget-system-is-a-mess-what-will-president-holloway-do-about-it/

Data Ethics Responses:

For number 8, “Could your algorithm promote material that would traumatize or upset individuals?” We answered maybe because there may be a very small chance that our generative art may produce lewd or inappropriate art, or art that could be interpreted as inappropriate. In the

randomness of how we are designing the art to be created, it is possible that what is created may be interpreted as inappropriate, but in no way was the program designed to do such art. To further prevent the user from being able to possibly submit insensitive or inappropriate entries, we are requiring the user to upload their data set to a website before uploading to our software. We will also be allowing only individuals of 18 years of age and older to access our software; we added a 'Yes' to option 12 above to note we will restrict users' age.

Number 21. We will be making users aware that we will store their experience in our software, but not specifically storing their data set, if inputted. As creators of the software, we will be interested in analyzing how different users interact with the software, so we will store data related to that but not related to the users personal info and will not be mining the users for information for using our software/site.

Number 24. The user could be harmed if they have a pre-existing condition that could cause them to have a seizure or other medical episode. We will add a caveat to the start of the software that users have to accept before moving on.

Number 25. I think this is supposed to be 'yes'. The user will be notified that their experience will be stored as data but none of their personal information or computer information will be collected.

Skills & Data List Breakdown:

- Taking user data input for chosen data (Jeff, Cole)
    - General data set input
    - Two text inputs to take variable names of interest from user
        - Text inputs will use Regular Expressions to match data column names
- Data engineering, cleaning, preprocessing (Derek & Anna)
    - Two sources:
        - User inputted data
        - Predone data sets (20)
- Data analysis (Derek & Anna)
    - Potentially, incorporate ML and other models into data analysis which can include but not limited to linear regression, clustering, forest and tree models, knn models
- Manipulating analysis into visual art (Derek & Anna)
    -
- Manipulating analysis into audio art ***If Time Permits*** (Jeff & Cole)
    - If time permits, audio will be made using Python libraries and built from the data set given/selected by the user

    - Make general MIDI music scores to play along with the art, if no time for the first option.
- Ability to build a web application (Jeff & Cole)
    - User landing page with disclosure prompt
    - User input page
    - Display page showing generative art
    - This also includes prompting user for input in order to change art to the users liking
- Math equation assortment. Minimum of 50. (Derek & Anna)
    - Gather math equations that vary in output so that visually we have a lot of variance in what is possible to create. For instance, Sine equations and cubic equations both produce unique output and plots that are visually different.
    - *If time permits, lastly, the use of Generative Adversarial Networks (GAN's) can be used*

Literature/Market Review:

Generative Sound Project - Gensound Gensound is an open-source python project/program that processes audio and can then alter audio signals by transformations or combine different audio signals. This project includes unique syntax that enables simple programming to manipulate audio signals. This project is similar to ours in that they will generate new audio using given or existing signals, but different in that we also are generating new visual art using given or existing visual art. A cool feature in Gensound that will be similar to ours is they use Sine waves and other various math to manipulate their audio sounds, we will be using math to manipulate our visual art. We will be using mathematical equations to manipulate visual art. Github: https://github.com/Quefumas/gensound

Automatic Music Generator - Improvisor Improviser is an automatic music generation program that plays music in real-time. The result of the project is a MIDI file. This project interestingly uses Pygame for its visualizations, which is a feature we will also be using in our generative art project, so this is a good similarity. Another similarity is we will provide a graphical user interface for our project. As a difference from our project to Improviser, we will not be providing a command line interface for the users. Link: http://improviser.onderstekop.nl, Github: https://github.com/bspaans/improviser

Context Free Art A program that takes a written input ("instructions") from the user. This input is labeled as a grammar which instructs the program to produce an image containing "millions of shapes". Context Free's program can produce thousands of unique images from

different artists. This program's images can be saved in both PNG or SVG format and is available on Mac, Windows and Posix/Unix. The code for this program was created using C++ and the "anti-Grain Geometry 2D graphics library" and allows for any users to freely use both its images and code. This will be helpful in the development of our project by seeing how a similar program was built using C++. Link: https://www.contextfreeart.org

TopMod is a multi-mesh modeling solution that provides a large range of options for creating high-genus designs, including practically all subdivision algorithms, a wide range of techniques to remesh shapes, additional extrusions, and more. TopMod is based on a novel paradigm that allows the topology of 2-manifold polygonal meshes to be changed dynamically. Topological mesh modeling is a wide ranging term that is used to house three unique extensions of the work encapsulated within. The first is Orientable 2-manifold mesh modeling using graph rotation systems and its accompanying computer graphics applications. Second is knot modeling with hints of non-orientable meshes, and finally topological constructions that revolve around very physical and geometrical limitations and constraints regarding graph rotation systems. Link: http://www.topmod3d.org

Generating Art from Neural Networks A painting was generated by a class of ML algorithms called generative adversarial networks (GANs). Looking at the wide variety of visual content editing,recently there has been the explosion of deep fakes, auto generating characters or even creating visual artwork. With its current wave, people are starting to realize the true potential of GANs' and the combination of AI to make applications more general. Adopting statistical modeling techniques, one particularly is neural networks. Neural Networks allow growth of computational power and the ability to access massive data sets. GAN's have a unique approach to neural networks for training machines. Generative modeling trains a data set of images to generate new images that are not the same. A random element shall be implemented so that the new image is different every time. The new generative model will compare to the original training dataset, the output group of pixels will have some sort of resemblance to the original. Link: https://www.weareworldquant.com/en/thought-leadership/generating-art-from-neural-networks/

Timeline:

Week	Cole	Derek	Anna	Jeff
6	Data gathering and cleaning (5 datasets each)	Data gathering and cleaning (5 datasets each)	Data gathering and cleaning (5 datasets each)	Data gathering and cleaning (5 datasets each)

7 - testing to make sure we know where with algorithm development	Manipulate data into audio art	Manipulate data into visual art	Manipulate data into visual art	Manipulate data into audio art
8	Manipulate data into audio art	Manipulate data into visual art	Manipulate data into visual art	Manipulate data into audio art
9 - We may be able to increase our in-house dataset numbers, depending on how week 6, 7, and 8 went.	Begin demo website, including art produced from week 7 and 8	Data gathering and cleaning (round 2). Begin finding math equations of interesting output.	Data gathering and cleaning (round 2). Begin algorithm/s for cleaning and processing user given datasets.	Begin demo website, including art produced from week 7 and 8
10	Create in-house data web page display	Continue finding math equations, aid in finishing algorithm for cleaning/processing user given dataset	Help finding math equations, aid in finishing algorithm for cleaning/processing user given dataset	Create user input and web page display
11	Ensure in-house data displays correctly. Ensure art and audio work the way expected.	Produce algorithms for making art from cleaned data.	Produce algorithms for making art from cleaned data.	Ensure user submitted data displays correctly. Ensure art and audio work the way expected.
12	Help with algorithms for making art from cleaned data.	Produce algorithms for making art from cleaned data.	Produce algorithms for making art from cleaned data.	Help with algorithms for making art from cleaned data.
13	Web development environment testing	Finish and test algorithms for making art from cleaned data.	Finish and test algorithms for making art from cleaned data.	Web development environment testing
14	Final web testing	Final software testing	Final software testing	Final web testing

Final Presentation Format:

Web Application platform, used as a host of the datasets applied.

Compatible Systems:

All web browsers as we will have web hosted software. If necessary we'll narrow down Chrome at the very least.

Minimum Deliverable:

The user accesses our software via a web browser, choses a dataset or inputs a dataset, and displayed is a unique generative art piece using our algorithms.

Expandable Deliverable:

If time permits, one thing focused on is manipulating analysis into audio art. Audio will be made using Python libraries and built from the data set given/selected by the user, and make general MIDI music scores to play along with the art. Another aspect that may be potentially used is GAN's (Generative Adversarial Networks).

Data Availability:

No data scraping involved (subject to change if we decide to use our expandable deliverables).