Scene Bot: Frontiers in the application of artificial intelligence as artistic medium

Aidan Collins: arcollins@massart.edu

Scenebot is a chatbot-inspired] dialogue generator that allows users to collaboratively author a script with the help of a “class conditional language model” (CCLM). Scenebot uses off-the-shelf AI parts to achieve a creative end. This project represents a primary investigation into a much larger space of “intelligence art” and the fundamental artistic objectives that must underlie the successful application of thinking machines to good artwork.

Drawing a distinction between “artificial intelligence art” and art made through the artistic application of artificial intelligence promises to revolutionize the arts and our general creative capacity as humans. Currently, there are many compelling artworks that utilize artificial intelligence in a myriad of ways ranging from image generators, music machines, interactive exhibits and more. However, all of these works are diminished by a current critical failure to establish when artificial intelligence has been deftly applied, or even what valid artistic contributions may arise from artificial intelligence; questions of authorship, for example, remain largely unanswered. For instance, Google’s “deep dream” produces visually stunning images reminiscent of dreams or hallucinations, truly dissimilar to the work any human artist; however a notable lack of artistic intention behind the creative decisions that compose the project makes us feel as though it is better defined as an engineering marvel than artistic masterpiece.

While Scenebot taught me much, many of these questions remain. Fortunately, by considering the creation of Scenebot in both a technical manner and in an artistic manner and then juxtaposing these individual perspectives, we can begin to carve out broad areas for improvement for future artistic work with artificial intelligence. For instance, early in the project I had to decide how to evaluate the model. This apparently is its own field of research, or at least a subproblem for natural language text generation, similar in effect to running the most optimal Turing Test (maximal quality measures on the text and maximal empirical uncertainty reduction elicited from minimal subjects and minimal interactions.) Usually the metrics that are valued for assessing language models depend on large numbers of testers and samples which were cost-prohibitive in the scope of this investigative project. Therefore, creative liberties were taken with the construction of the modeling procedure, away from the “technical best practice.” Indeed, projecting a strong enough sense of class conditioned language onto the Generative Discriminator (GeDi) model was difficult in practice because the selected class -- the genre of the movie sample -- was particularly “latent” as a topic. (However, hints of class dependency can be seen especially when selecting “sci-fi” and “horror” topic codes.) Principally, compared to fine-tuned causal language models, trained GeDi’s pull on a much larger distribution of text generated by a separate language model, which is not further trained, and which is selected by the GeDi given a set of k class vectors that are real words occurring in the language model’s vocabulary. This diversity of text generation would seem ideal for artistic “tool” applications where unrealistic samples can be discarded by the user and good samples can be kept. However, the GeDi alone produced what was technically remarkable but artistically lacking. It seemed I was chatting with a vague dialogue -- much like the language used in real Hollywood

scripts. In truth, many of the samples would fool me when compared to a random sampling of the dataset the model was trained on. This is both reflective of the dependency the scripts have on the accompanying cinematic visuals and context as well as the loss of broader narrative context when sampling singular two-party dialogues from a larger script.

Technically, however, Scenebot demonstrated an interesting language modeling result in the success of its chatbot undercarriage. Unlike many other creative writing robots, especially for scripts, scenebot is built on a pretrained chatbot. In this case, Microsoft's DialoGPT-small was used, which is a derivative of Openai's GPT2-small. The idea is that brief fine-tuning of a more task-specific base-model will better leverage the data resources than fine-tuning the same model trained on a less similar task. This seems to have worked remarkably well. The GPT2 fine-tuning for script generation from raw text does *not* display a dialogue nature of verbal back and forth despite only seeing conversational training data from scripts, but the DialoGPT does, even though it is built from the same architecture and differs in that it was further trained on conversational text. Both models are inadequate for expressing natural script-like language due to genre contextual confusion. Attempts to use special tokens in a standard GPT2 configuration to add class dependency fail because they overfit the training set by failing to disentangle semantics of the class words and the positions of those words. This raises the first philosophical conundrum: If the text seems adequate from a language model that conceptually isn't, have we succeeded as artists? I think not. For the same reason fitting fractal dimensions to Jackson Pollock painting periods helps to add intentionality to his splatters. How we analyze novelty establishes a large part of its meaning and aesthetic value.

My ultimate technical solution to combining these systems was to feed my retrained GeDi model a *fine-tuned* DialoGPT2 generator model, trained on the same dataset. I perceive these results as better than all combinations described above. From a conceptual perspective, the GeDi can now pull on a further refined (albeit still too "vague" on its own) generator model. In principle, this would be a technologically very improper way to model these *two* volumes of data (data used to train DialoGPT and the script data used to train the GeDi) because of the concern of overfitting relative to leaving the generator model untrained. At this point, I elected to enact my creative license and break a rule. The two models weren't "sharp" enough on their own. It would stand to reason that replacing the generator model that was not fine-tuned with the fine-tuned generator, we ought to further fit to the GeDi training data? Because we can't evaluate at the large empirical scale of sophisticated research papers, we ought to strive to achieve better performance on the metrics available to us,namely how interesting the text is, which was the objective of the project. Oddly, Scenebot claims a contribution here, in my opinion; we can now overlook many empirical standards that are only distally relevant to what we view as the creative objective of our work, when operating in an aesthetic space.

This subtle interplay between technical considerations and the possible ranges of artistic meaning derived from them demonstrates not only the challenges of leveraging A.I. advances in a principled artistic manner but also the powerful semantic tool that data can serve in the artist's creative process. By considering the relationship between "what works" and what we think is "making it work," we can claim to gain more creative control over the semantics of our artwork. In many ways, we can argue that carefully constructing creatively optimal training combinations is the exact opposite of a Pollock drip painting "informed" by random chance. However, to the scientist, recklessly wheeling about in this combinatorial space might seem even more chancy.

This process is equivalent to the careful teasing of entirely abstract versions of reality into a meaningful form of the same -- arguably an explicit version of these “meta” qualities that saturate the entirety of artistic production. Many of the decisions made in the creation of Scenebot had a technological “feel” to them, seemingly opposed to the stroke of a paintbrush or selection of a jazz improvisatory note. However, scientifically speaking, they are technically disagreeable decisions. Therefore, not even a scientist could claim they’re much more than art.

Therefore, working closely with artificial versions of intelligence, planning and reacting to constructive creative decisions, will make us better aware of our own thinking mechanisms, their behaviors and our intentionality as artists. By recognizing that we must understand artificial intelligence at some level to use it, we might then understand something about what we think we can understand about thinking -- a sort of continuous “meta thought experiment.” This promises to expand the scope of thought itself, which I would claim is a similar objective to that of art-making in general.

Finally, Scenebot opened the door to many exciting avenues to pursue with regards to my artificial intelligence art projects. I recognize a need to further develop an explicit semantic artificial intelligence for artists. Semantics are far from a codified field; after all, the true meaning of a “field” is semantics, isn’t it? But graph databases and their associated analytical tools, such as graph neural networks and embedded graph database neural networks, suggest a wealth of application for a new space of problems I propose called semantic discovery. There exist some topics we don’t know are compelling yet. Finding new meaning in the world is one of the chief functions of art. It could now be possible to store, compute and generate meaning in an explicit way. Imagine being able to match queries like “hot dog like dog” (equals dachshund) with some measure of amusement metrics (like google search trend data adjusted to reflect evolving semintenment) related to “dachshund,” or qualitative information, such as the precipitating query to the average “dachshund” look-up. These sorts of common sense mental “lookups" are at the core of expressive artistic output. As a technical aside, work on semantics is far more principled since the 1990s and the age of Chomsky-inspired semantic networks for natural language understanding. Despite the biological plausibility of these older ideas, they never achieved much success as an engineering application. Consider now that we can encode certain chemistry problems (graphical in nature and literally beyond the scope of intuition) such that they are compatible with deep learning frameworks. Technically, deep learning can be viewed as a way to intuit what was previously impossible. These algorithms can now outperform human chemists in certain domains. Similar algorithms have shown application to social information that previously evaded understanding. Furthermore, the topic of semantics extends outside of the space of natural language to encompass the qualification of many types of codified information. Building artificial intelligence tools to help artists mine and construct semantics might promise to revolutionize the way we imagine.

Secondly, the nature of production and consumption balance has become of extreme interest to me. We view art as an inwardly focused productive process usually conducted at length in relative isolation, challenged by our ability to match isolated individual exertion with the deliberate and specific stimulation of an audience at a later time. Our productive process seems philosophically inverted. Instead, what if we made our work about finding what the audiences will like and assumed the generative process to be entirely static and, for example, virtually immediate? In fact, I would argue much of contemporary art is already headed in this direction,

favoring the novelty of production over the perfection of a traditional craft. In the case of Scenebot, applying the right dataset made a conceptually interesting interactive artwork. I now imagine a space of online experiments that combine tools such as causal inference and reinforcement learning to help identify more compelling subject matter than the standard thinking process provably can. This stands in stark contrast to the lesser form: an optimal sequential experimental design in a fixed space. I liken the fixed exploratory space to the optimization of the craft of discovering what moves an audience and the dynamic exploratory space to rapidly evolving the practices of this craft as well the insights that those practices generate. In this sense, we obviously can hope to see beyond the scope of our imagination if we elect to continuously explore. This marks a new advancement in the discovery of concepts. The technology is there. We can easily parse behavioral cues from video cameras, we can track eyeball time spent on canvas position, and project emotional reaction from facial features -- all with less than $150 in hardware. What if we used this technology to discover or construct meaning that exists outwardly in the real world instead of for commercial and governmental purposes? We might consider that the application of A.I. towards exclusively “mechanical” good, such as medicine, biology and physics, may leave us in a world that is decidedly not beautiful in an artistic way -- by definition. Therefore, a final reason to pursue this work is that economically it will combat the public’s innate interest in developing artificial intelligence as a tool of force rather than a tool of insight by keeping pace with aesthetic-less value creation enabled by A.I. application. (A tendency to focus on aesthetic-free value creation as a superior investment generally shunts funding from the arts.) As we consider machines that can outthink and outplan us, we ought to consider whether the communities that build them share the same values as all other communities that will use and be affected by them. We might also consider that non-artistic communities are not inherently dependent on the interpersonal reception of work. For example, physicists may annihilate many individuals in the course of a work they deem successful. We use these same analytical tools to help discover the triggers that will inspire a purchase, help identify likely criminals by camera, and build complex databases of large groups of citizens and their private interactions. Wouldn’t people continually prefer to buy, develop, work on, live around and approve of artificial intelligence that was designed to inspire, help, enhance and discover? When viewed from a scientific or mathematical perspective, ensuring these targets is virtually impossible to demonstrate empirically or prove, at least presently. When working from an artistic angle, we believe that goodwill will prevent the advance of creative ingenuity from cascading from good to evil. We can hope to overcome these A.I. challenges in a similar manner, by acknowledging the empirical limitations of the engineering procedure and defaulting to the purest of creative intentions, which we already trust to ensure that good will come from our creative endeavours.

Scientifically speaking, neural networks and many other cognitive computing machines are stochastically conditioned deterministic systems. Therefore, they cannot fundamentally alter our certainty about the global perspective, only astound us with “local” performance. Scientific applications of neural network architectures cannot claim to have learned anything about the global “world” by explicitly using the neural network but rather only about the subset of the “world” represented by the relative performance of the network on whatever training and testing data were selected. Applying the deterministic learning machine to a scientific or technical problem is an art in and of itself.

Data and art may seem antithetical; perhaps this perspective is a function of social conditioning and the environments in which we are typically exposed to these topics. However, the two seemingly disparate fields are really better characterized under one umbrella topic of “observation.” Using analytical thinking to both empower expressivity as well as scientifically assess expressive output promises to revolutionize the arts, advancing the human capacity to imagine as well as to concretely understand “human-interpretable” semantic information. These inventions of thought will allow us to expand the scope in which we try to conceptualize possible comprehensions of the universe.