#Riseandgrind: Lessons From a Biased AI

Conor McGarrigle

Dublin School of Creative Arts, TU Dublin

Abstract: #RiseandGrind is a research-based artwork that, through a process of active engagement with the machine-learning tools of what is known as artificial intelligence, sought to make visible the complex relationship between the origins and context of training data and the results that are produced through the training process. The project using textual data extracted from Twitter hashtags that exhibit clear bias to train a recurrent neural network (RNN) to generate text for a Twitter bot, with the process of training and text generation represented in a series of gallery installations. The process demonstrated how original bias is consolidated, amplified, and ultimately codified through this machine learning process. It is suggested that this is not only reproductive of the original bias but also constitutive, in that black-box machine learning models shape the output but not in ways that are readily apparent or understood. This paper discusses the process of creating and exhibiting the work and reflects on its outcomes.

Keywords: Twitter; Machine-learning; artificial intelligence; new media art; generative art;

1. Introduction

#RiseandGrind, is an AI art project using textual data extracted from Twitter hashtags to train a recurrent neural network (RNN) to generate text for a Twitter bot, with the process of training and text generation represented in a series of gallery installations. The work was first commissioned in 2018 and exhibited in three different iterations in 2018 and 2019. The neural network was trained on two selected hashtags, #RiseandGrind and #Hustle, specifically chosen as representative of a Twitter filter bubble that I identify as embodied neoliberal precarity. That is a form of economic self-exploitation arising from an assimilation of the values of the precarious digital gig economy characterized by an adherence to neoliberal principle that economic success or failure is built entirely on individualised effort. In this filter bubble the dedication to the hustle, relentless self-promotion, and dispensing of bland life advice is total, almost to the point of parody; it is also lively, energetic, occasionally ironic and a little anarchic and, in this respect, encapsulates the best and worst of social media. The project sought to make visible aspects of the black-boxed machine learning process, to raise issues of algorithmic bias through demonstrating the training process, and to highlight the role of artistic practice and research in understanding these issues as part of a broader dialogue in addition to exploring their aesthetic potential, particularly as a method for the generation of text. While questions of algorithmic bias are not new and have been widely studied, they are still relevant and unresolved as illustrated by the recent controversy over the training of Open AI's GPT-2 model on Reddit data noted for their strong bias. (Sheng et al. 2019) This text discusses the motivation and process of producing an intentionally biased AI resulting from training on biased filter bubble data, unpacks the process of training a RNN on Google's TensorFlow from a non-specialist perspective, and rendering this process visible through a series of gallery installations.

2. Twitter Hashtags

Hashtags are a core mechanism that coordinate the flows of Twitter conversations, dynamically forming and reforming ad hoc publics (Bruns and Burgess 2011) that assemble over news and information, common interests, cultural moments, values, for political debate and activism and more (Murthy 2018), presupposing a "virtual community of interested listeners". (Zappavigna 2011:791) Alongside other categorisation techniques such as location, followings, trending subjects, the hashtag is the "killer app" that enables Twitter's users to consume and interact with tweets from users they do not follow or have location in common with and engage in conversation with strangers on matters of common interest. In addition to organising informational flows, hashtags have been seen to play a role in not only constituting online identities but also in co-producing these network identities, this effect being particularly notable on Black Twitter. (Brock 2012; Freelon et al. 2016; Graham and Smith 2016) However, they also form filter-bubbles, self-referential immersive information environments (Pariser 2011), these can be both mutually supportive communities of special interest and echo chambers where ideological positions are rehearsed and reinforced with little outside intervention, which can serve to reinforce and amplify bias and are subject to automated manipulation from bots.

Hashtags act as method of categorising twitter users and their data for the purposes of surveillance, targeting, and the accumulation of what Shoshana Zuboff calls behavioural surplus. (Zuboff 2019:65) Twitter packages and sells insight on their users through their data services and offers access to data through their APIs making Twitter a popular source of training data for myriad machine learning applications including sentiment analysis and content generation. These assemblages (Kitchin 2014) of surveillance capitalism target not only the specifics of interests and activities as indicated by followings, likes, retweets, locations, interactions and hashtag activity but, through algorithmic sentiment analysis, opinions and values. The goal of surveillance capitalism is, according to Shoshana Zuboff, not only one of behavioural prediction but is ultimately to persuade, to nudge and change behaviour based on past actions. (Zuboff 2019:68) The full consequences of building predictive and persuasive models on data that are biased and incomplete, the result of self-replicating internet echo chambers that should not be extrapolated, is only beginning to be fully understand. In the words of Anima Anandkumar, Director of AI Research at Nvidia, discussing the training of Open AI's GPT-2 model on Reddit; "when you train on Reddit data, out comes garbage"[1]

[1] Tweeted November 6 2019 https://twitter.com/AnimaAnandkumar/status/1191983025250295815 . See also Sheng et al. 2019 for a more detailed analysis of bias in NLG generation

3. Data Acquisition

The project began with two interrelated hashtags, #RiseandGrind and #Hustle, that are not openly politically partisan or controversial and don't readily fit within social media culture wars, which is not the same as saying that they are apolitical. In fact, I suggest that they are ideological, espousing a value system that emphasises an individualistic self-reliance, where hard work and entrepreneurial hustle are all that it takes to succeed in the neoliberal gig economy. They are representative of an economic world view that I identify as embodied neoliberal precarity; that is a value-based form of self-exploitation that conflates the requirements and economic values of the precarious gig economy with personal identity and self-realisation or individuation, threading a ground that has been previously identified and described as characteristic of the sharing internet economy. (Scholz 2012) As with many hashtags their usage is complex with irony and sarcasm juxtaposed with naively bombastic tweets. As expected, automated bot activity is evident (Varol et al. 2017) at various levels of sophistication, from crude spam hashtag storms to carefully targeted tweets that pass as human. However, the overwhelming impression is of a filter bubble delineated within these hashtags with a clearly articulated message and a cohesive world view, even if that does not stand up to sustained scrutiny. It is important to note that these hashtags are also lively, energetic, entertaining and fun, in many ways pure Twitter in that they are well-attuned to their medium, in form and content. It was for these reasons that these particular hashtags were selected as training data.

The project began with some questions. What would training using a common language model on a ubiquitous machine-learning platform produce from this data and what conclusions could be drawn from its results? Would traces of the process of training and adjustments to the training parameters produce aesthetic traces that are unique and characteristic of the process - similar to digital artefacts of glitch art – and could the process of making visible these black box processes through the medium of art add to the critical debate on AI and society by adding an additional perspective beyond those of experts in the field?

4. The Intelligence of Machine-Learning

While I describe this project as AI, I acknowledge that the term artificial intelligence itself is problematic. AI can be more accurately described as machine-learning, the use of convolutional neural networks, recurrent neural networks, generative adversarial networks, deep learning and so forth. Current machine-learning techniques differ from earlier generations of AI with their focus on creating thinking machines to emulate the human brain to create a general artificial intelligence. While this bio-memetic terminology persists, the techniques are very different. Machine-learning is a probabilistic method that works with statistical correlations and heavy-duty computational power to identify patterns in datasets and encode these into a model which can be used on unseen data to perform its decision making functions; generating text, identifying objects or faces, machine translation and so forth. Despite the anthropomorphic terminology it's not intelligent in a general human-level sense, rather it's a fast, brute force method that can achieve impressive results in narrow domains. (Karpathy 2015; Kelleher 2019; Sejnowski 2020) Machine-learning techniques require high quality data to train the network in a computationally intensive process with the results only as good as the quality of the data input. In many applications, including computer vision and autonomous vehicles, this process often has a Mechanical Turk aspect to it, with automation made possible by workers painstakingly creating training data input by manually labelling millions of images. (Lee 2018) With natural language machine-learning the training data is often, as in the case of GPT-2 the most prominent and powerful language model, sourced from the internet with their inherent bias problems.

Machine-learning systems learn through advanced pattern recognition within very large data sets, with these patterns then encoded into the algorithms – the process of training. These algorithms can then be applied to data to categorize it based on these patterns or to generate new data based on past actions, the premise is that the larger the data set the more accurate the outputs. Of course, this isn't always true, and machine-learning has been found to codify bias, error, racism, and sexism in what philosopher Bernard Stiegler calls functional stupidity or la bêtise. (Fitzpatrick and Kelleher 2018) However, the effectiveness and opaqueness of machine-learning systems, tends to militate against the recognition of bias in these systems. As decision making systems based on machine-learning are widely deployed, the flaws in their make-up are emerging. Recent scholarship has revealed these flaws; racist search engine assumptions (Noble 2018), sentencing systems that discriminate against people of colour (Angwin et al. 2016) even facial recognition cellphone unlocking unable to distinguish between Chinese female faces. (Zhuang 2017) Problems of bias are complex but begin with data, and with a failure to recognise the unrepresentative nature of the data as well as outright bias, racism and sexism in what has been described as "privilege hazard". (Ignazio and Klein 2020:28) However, failures in AI are not always readily apparent and as machine-learning systems are incorporating into all aspects of life the act of discrimination and bias is baked into proprietary algorithms incorporated into larger processes of governance, so that identifying where in a complex process this bias is located is not often possible and actionable. It is against this growing context that this project was conceived.

5. Neural Network Training

The project began by acquiring a body of tweets with the hashtags, #RiseandGrind and #Hustle. This took place over a period of weeks, the idea was to capture a representation of the hashtags, capturing both historical and current tweets, over a period of time to dampen out temporal fluctuations arising from specific events. In all, this process has been repeated on three different occasions over an 18-month period resulting in a corpus of tweets that demonstrates a broad consistency of usage within these hashtags. This amassed approximately 600,000 tweets with retweets excluded from the collection. The proportion of #Hustle to #RiseandGrind tweets was 80% to 20% respectively, most tweets were in English with US based tweets in the majority with Nigeria the second most represented country. Tweets were acquired through the Twitter APIs and saved in a Mongo noSQL database. In addition to the tweet text the Twitter APIs return comprehensive metadata for each the tweet which typically includes 70-80 fields or 250-300 lines of JSON. The tweet text, a single field, was extracted to form the body of training data.

The data was used to train a language model on Google's machine learning-platform TensorFlow. TensorFlow was selected as a ubiquitous tool for machine learning. Originally developed by the Google Brain team for internal use and released as open source software in 2015 it has become a widely accepted industry standard framework synonymous with deep learning and neural networks. For this work the intention was to train a model to generate the perfect #RiseandGrind and #Hustle tweets, tweets that would pass unnoticed. To achieve this the model needed to identify and extract patterns from themes, subjects, language use, hashtags, retweets, and @ing other users in the conversational flow. Twitter is a fast moving and complex textual environment, where conventions and user practices are not only specific to Twitter as a whole but are more specifically dynamically defined within an array of filter bubbles and conversations that form and reform with implicitly understood rules of engagement. The task of generating tweets that fit within this very specific hashtag environment is not trivial, and for machine-learning, which is excellent at pattern identification but has no semantic understanding, it is challenging. Amongst data scientists training a neural network is often considered to be both an art and a science, as fine tuning of multiple parameters impacts greatly on the quality of the results with over and under training presenting significant problems. Achieving the right balance is an art that comes from experience and intuition as much as deep understanding of the mathematics. For a novice both problems were evident as the data went through a series of trainings with variations in the data-set size, the number of epochs (length) of training and other tweaks to the parameters of TensorFlow. A character-level recurrent neural network (RNN) was employed to model the probability distribution of characters in the body of training data, that is the probability that one character will follow another, to produce a body of text under 280 characters, character by character. When using terms like artificial intelligence it's important to remember that the neural network has no semantic understanding, what is actually happening is the network calculating the probability that one character will follow another using the example of the training text. RNNs it turns out are, in the words of Tesla's Director of AI Andrej Karpathy, "unreasonably effective" (2015) in doing this.

5.1 Following the Training Process

The training process was logged in order to make visible the process at work, with the system configured to produce sample text at periodic intervals to identify the state of the training process. This log included input text alongside sample output with a value for the network's confidence that it was correct. In early epochs, we see blank text or single characters often repeating (see Table 1)

Training Text	Generated Text	Batch Loss
d #ironwillfit #fitness #fit	nnnnnn	loss: 4.05692
d\Gameday! #RiseandGrind\#RISE	nnnnnddnnnn nn	loss: 3.53756
the hustle never stops. #RiseandGrind	ennn	loss: 3.10608
her listened to my Rise And Grind	aon #o to #isenAnd iii	loss: 2.15545
Grind #positivevibes #fitlife #	nd \Rosenene one #ooseene #o	loss: 2.43459
#FridayFeeling #FineWomenFriday	iisen #ri nd #iise ar in	loss: 3.13049

Table 1: Early stage sample from the training log showing training text, corresponding generated text from learned state, batch loss

The network periodically generates random text from the current learned state (see Table 2) As the epochs progress the text begins to coalesce into words which are typically nonsensical.

TRAINING STATS: batch 0/173 in epoch 0, batch loss: 4.47500, batch accuracy: 0.12733

┌───────Generating random text from learned state──────────────────┐
#AtnBwiq nw#thn##aasi a r iaatiy aooornania#ytstytnrr ia#y iirsoytsi #nnr itsnry tt tay y##
ts t tysiony#yyiay# iatonrot t nniryt#oi#irtyao#rrr oyontyars t n ossaistr o o nniinstainso
ysotyy#a r#yirtsarii#in r#tyaoniassysioiarsysry # ray yniretsttrysynnati aaann#satr#ytsi
nyiranyta satyi i naiattrniaa iaannyato nnr# eriarsnasoy a # ni# ai#orni isit#tiyi
snao#yairyssrsa y#ya##artonaaris#a#aysiraa #yaaan #i#iyonsitoyyitiotssrssnai aaatoon
raytotryyyyo or # oi#nry#rr tn nn tnr#a si i# iya yt i ror#y n#er nntntttiryatio#n rs#
yaarntnyinrs #arriaiaytsyyt #st nsrao ia#nr#tssyso yoir#trartono y#rn r#oia rsyiointso#rityr r
aynaanyosi risnyooiro###yniaatss sos to #rtna n arinarts or aons t issnoi# tonno t#r aynor oria
tsytnosn noonoait ##nytst#tnntsayitrnrsors t ianiyn##nr nryytyityio#s #irs ni#iaro# ta ao taa i

osoay neoriniyn# artti ornsan iort t y yrtaooanotysroyn #rt#ri##r#ystrisyrriasysayitstaiirrr sanon nonanynioritrra# ss# at iayriiy#ys aa a#yasa# iyyin##iesr rat# nrys
End of generation

Table 2: Example of a generated text, early in the machine-learning process

However, it doesn't take long before words begin to emerge, and the network begins to hallucinate Twitter handles and URLs correctly formatted with the http://www and @ form even though they don't correspond to any real address or account. The neural network quickly begins to generate texts that acquire the correct tweet form often in varying styles: long engaged tweets that @ many other accounts, pithy short declamatory tweets, self-promotion tweets complete with many #hustle related hashtags. Early tweets often made no sense but mimicked aspects of the style and themes from the hashtags with many errors (see Figure 1)

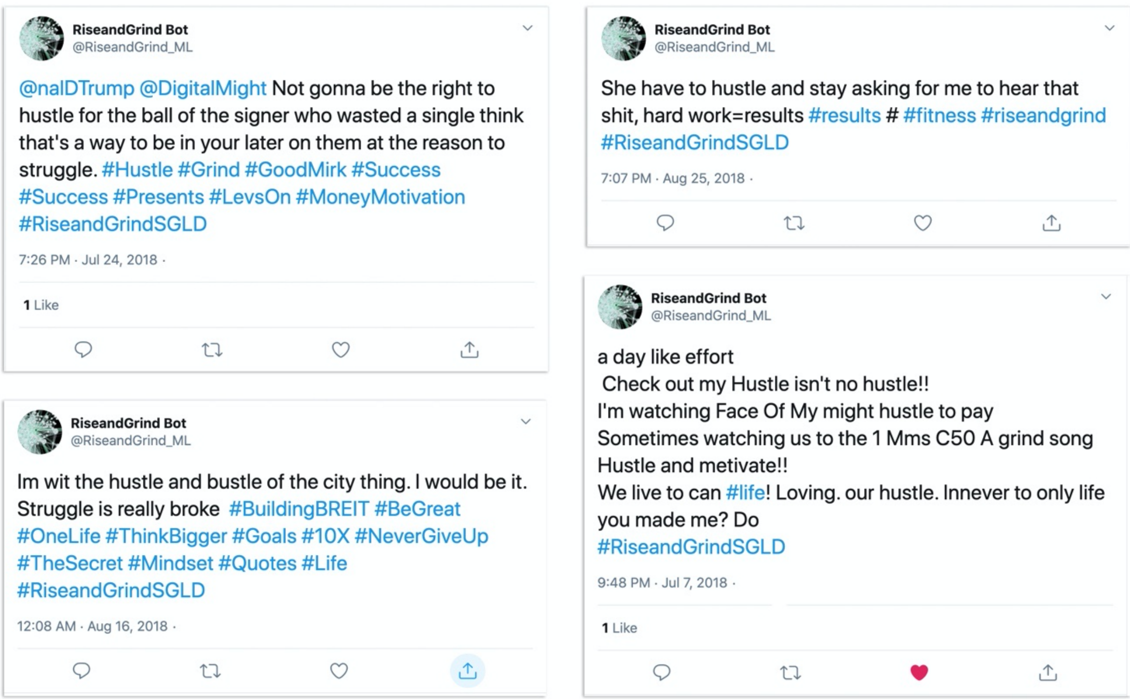

Figure 1. Early generated tweets achieve many aspects of Twitter style but don't make sense

However, as the training progresses the output accuracy improves with the form of the tweets coming into focus before the content. For example (see Table 3) in generated random text we see repetition of different variants of popular hashtags within the training text.

TRAINING STATS: batch 104/173 in epoch 2, batch loss: 1.78837, batch accuracy: 0.51517

┌──────────Generating random text from learned state──────────────────┐
ealys! #riseandgrind #suncass #fitness\00
#riseandgrind \00\00\00\00\00\00\00\00
RT @TheSharkDayMond #MandayMotivation #RiseandGrind #MondayMotivation\00
#riseAndgrind #hustle #riseandgrind #goodmorning #motivation #motivation #motivation
#motivation #motivation #motivational #makenturself #motication #morning #motivation
#motivation #morningmotivateon\00
RT @TaeseardDaymond #riseandgrind #goodmorning #riseandgrind
@TheSharkDaymond #RiseAndGrind #RiseAndGrind #MandayMotivation\00
#riseandgrind
#riseandgrind #goals #goodmorning #riseandgrind
@ around and getting. #riseandgrind #goodmorning #riseandgrind #goodmorning
#motivation #motivation #morningmotivation #motivation #manday #motivationalqoods
#bosings #morningmotivation #riseandgrind #sundaymotivation #morningmorning
#riseandgrind
@ and the wonth the work this morning! \00 #RiseAndGrind #RiseAndGrind
#RiseAndGrind \00\00\00\00 #RiseAndGrind
#RiseAndGrind
RT @TheSharkDaymotivation: The samplow #riseandgrind #success #morningmotivation\00
RT @MondayMotivation #Ri
└─────────────────────End of generation─────────────────────┘

Table3: Random generated text from learned state demonstrating repetition of popular hashtags from the training data

Training continued over a number of iterations using varying sizes of training data sets and different starting parameters to produce results that were at first glance indistinguishable from real tweets. Training logs were saved to be used in the exhibition of the work. Once trained the network is deployed to generate tweets with a Twitter-bot tweeting text from all stages of the training on a project Twitter account @RiseandGrind_ML (see Figure 2)

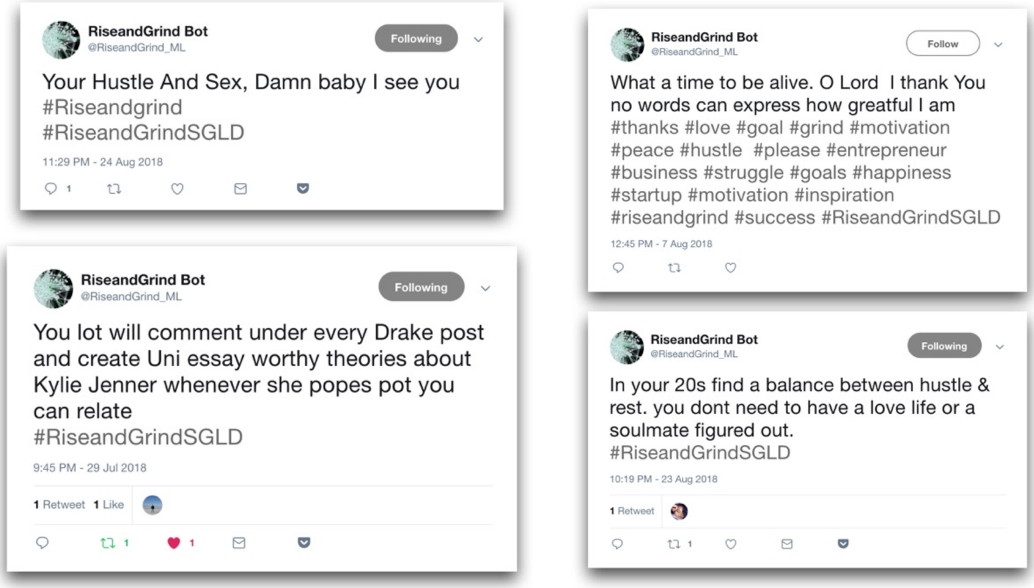

Figure 2: Sample tweets from the fully trained model

The final generated tweets had very successfully adopted the style and of their training tweets and did not seem out of place on the hashtag. It was noted that certain tendencies in the training data had become more pronounced and insistent with a noted shift to the right. The outputs were seen to have not only repeated but amplified the bias obvious from the hashtags. This amplification of the bias appeared to have come from a process of smoothed out difference and subtlety and thus erasing some of the context; the idiosyncratic, the ironic, the linguistic plays, the subtle sub-tweets, and the nuanced weird of the internet were all lost in translation. Without this context and these modifiers, the patterns of the hashtags' text which were replicated in a more or less convincing manner seemed like relentless exhortations to hustle and grind that contained none of the fun of the originals. (Figure 3)

Figure 3: Generated tweets from late in the training

6. Exhibition

There were two main impetuses to this work: one was the hashtags themselves, in my artistic practice I have always been interested in informal internet communities particularly those that form around practices and concepts to construct an autonomous conceptual world view. Previous work such as JoyceWalks (McGarrigle 2009), Spook... (McGarrigle 2015), and 24h Social (McGarrigle 2020) addressed these ideas of hybrid internet-based conceptual worlds from a number of perspectives. #RiseandGrind follows in this path through an engagement with a cohesive world view assembled under these two hashtags; one that speaks to the power of social media as a medium for assembling people and ideas in entertaining and powerful ways, and the power of the platform to algorithmically shape its content in way that are not readily apparent; the second was as a process of critical engagement with machine learning, to make visible not only how bias can be reinforced through machine learning but also the operation of the black-boxed machine-learning process itself.

The work was installed in a number of different iterations from its original commission for Hustle at the Science Gallery Lab in Detroit. The work was further developed as part of a residency with Parity Studios at University College Dublin and Insight Centre for Data Analytics, with new components developed and added for exhibition in TULCA Exhibition of Visual Art in Galway, and in Screentime in the Green on Red Gallery Dublin.

The exhibition is centered on a neon text piece, #RiseandGrind, connected to a live Twitter feed, which illuminates and dims based on activity on the hashtag. Screens display the training process throughout the duration of the exhibition as scrolling text that displays the input training text, the network’s sample texts and their probability weighting, epoch by epoch, from early stages to fully trained. The Twitter Bot is displayed on a series of 7” screens powered by network connected Raspberry Pi board computers alongside a live display of the hashtags from Twitter (see Figure 4). The final element is a screen-based

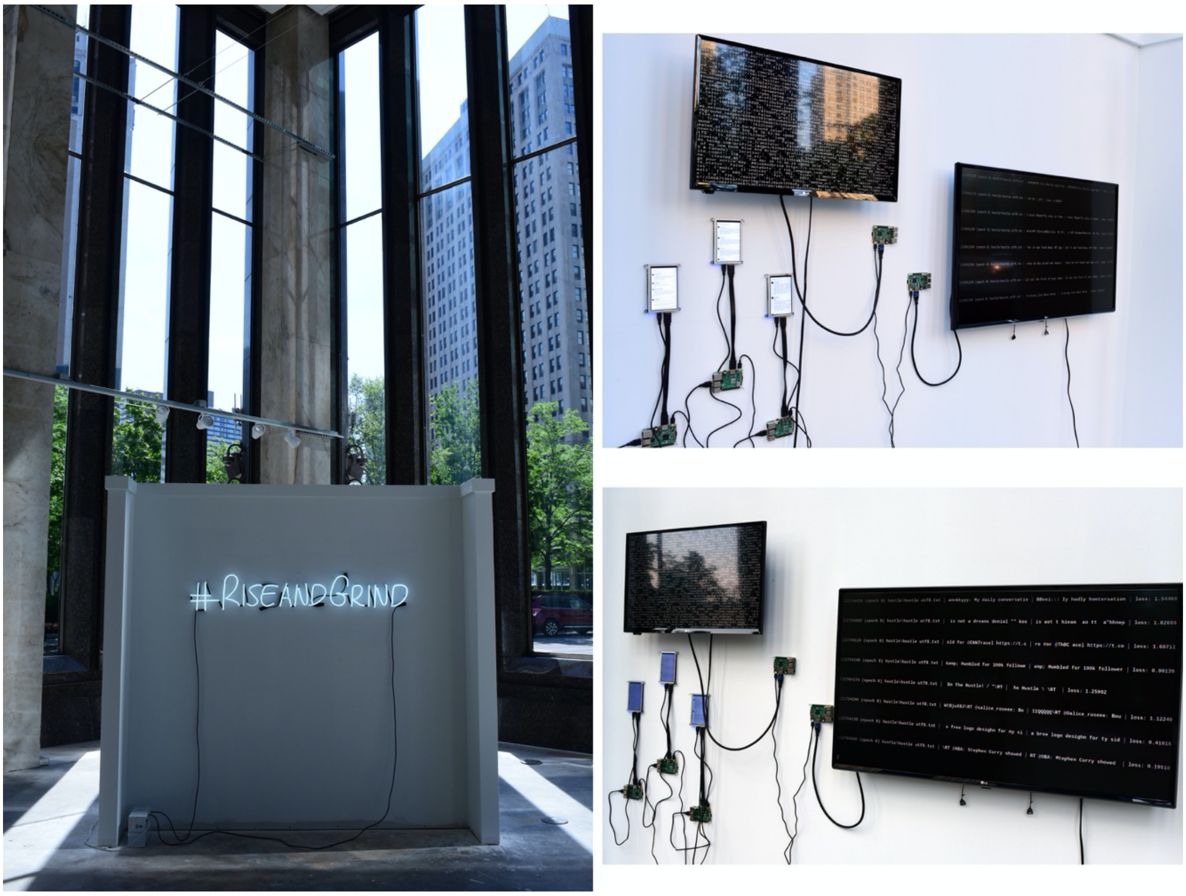

Figure 4: Installation view, Science Gallery Lab Detroit 2018.

display of all the generated tweets character by character (See Figure 5). The exact configurations of these elements are decided based on the gallery space. The exhibitions are accompanied by mediation and discursive events including the Science Gallery Lab Detroit's extensive programme of docent tours, artists talks for TULCA in Galway, and a panel discussion with academics and curators at Green on Red gallery with all events open to the public.

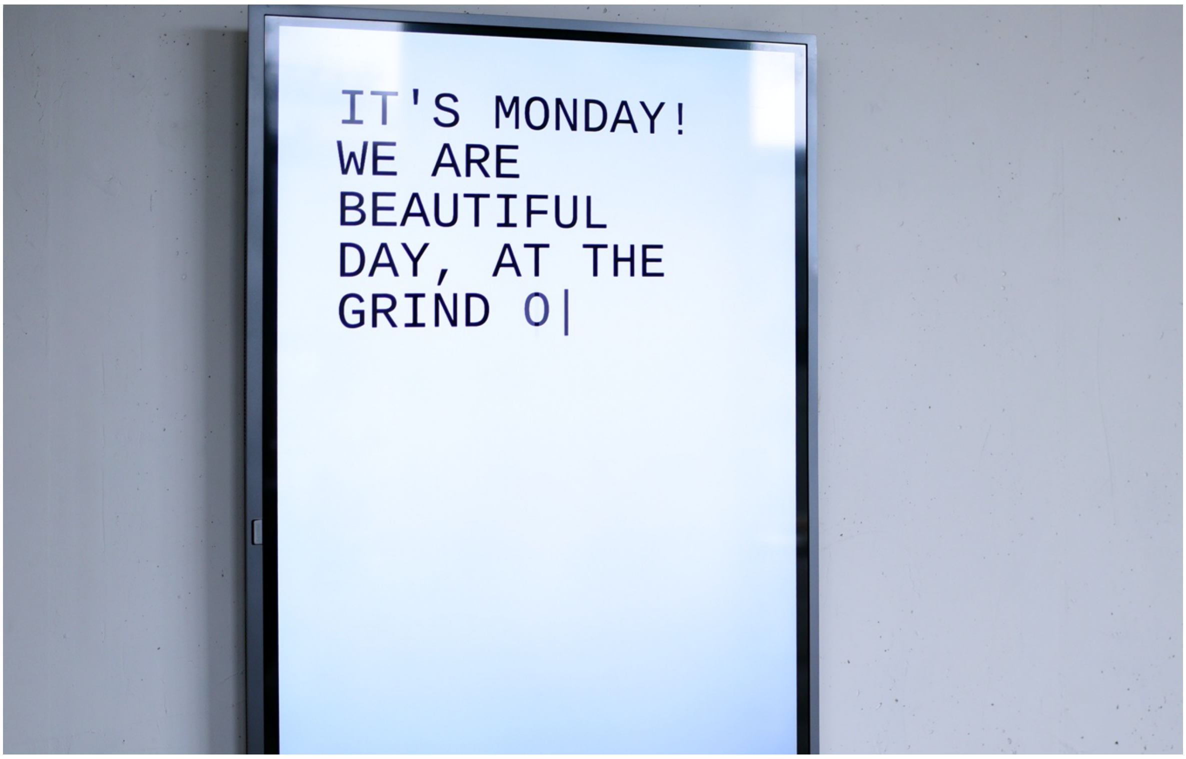

Figure 5: Installation view with screen scrolling through all of the generated texts

7. Conclusion

#RiseandGrind is a research-based artwork that, through a process of active engagement with the machine-learning tools of what is known as artificial intelligence, sought to make visible the complex relationship between the origins and context of training data and the results that are produced through the training process. It is a process work, and as such the final exhibition outcomes, while important, are not the sole arbiters of the work. Of equal importance are the components of what was a sustained process of engagement with these Twitter hashtags, their rendering as data and processing to model the complex activity so that it could be emulated. This process is at one level an attempt to unpack, understand, and generate new knowledge about machine-learning and its connection with artificial intelligence conceptually and in practice. In this respect the active engagement with the process with its errors, missteps and lack of expertise are an essential component of the work.

The work originates in an interest in ad hoc internet communities that assemble around diverse interests that are conceptually linked through overarching values – in this case a belief that self-reliance and hard work are the key attributes for success – and how these can be dynamically formed within Twitter's hashtag bubbles. Arguably this ability to assemble and empower communities and audiences, real and imagined, across geographical divides is what has made the internet central to everyday life. The work demonstrates the tension between the internet as social and surveillance space, as is evident in the relative ease that the hashtags can be captured and used to model and replicate their activity as behavioural surplus. Although there is an expectation that tweets are in the public domain, the uses to which they are put are not widely known, the artwork visualises one such process, the training of machine learning models and their deployment to generate text. In this simple way the project seeks to make visible the opaque workings of machine learning and to highlight issues of bias and the role of the origin and context of training data in creating and sustained algorithmic discrimination. The world view returned by the trained model was a cohesive one that replicated the bias obvious from the hashtags themselves. However, it also amplified this bias through a process of smoothed out difference; the idiosyncratic, the ironic, the subtle sub-tweeting, and the nuanced weird of the internet were all lost in translation, replaced by a hard relentlessness to hustle and grind that contained none of the ambiguity and fun of the original. The patterns that were identified and extracted were not wrong, but the tone in which they were delivered was.

The process of the work itself produced many individual aesthetic moments that were unique to the process. These are reminiscent of glitch art where algorithmic and machine processes produce these in-between states that speak to the nature of the technical process, making visible the workings of normally opaque algorithms in a way that I suggest is unique to art, bringing an additional perspective that adds to the critical debate on AI and society.

While bias in machine learning has been widely recognised it remains a significant problem that calls for broad agreement on AI ethical practices in civil society that goes beyond an ICT industry perspective. Artists have engaged with AI at many levels from works that have sought to open the black box and ask critical questions such as Anatomy of an AI (Crawford and Joler 2018), Not the Only One (Dinkins 2017) and Imagenet Roulette (Paglen 2019) and work such as Portrait of Edmond Belamy (Obvious 2018) that promote AI as new tools of machine creativity. I propose that critical AI artworks, to which I believe #RiseandGrind contributes, can act as artistic research method that provides a critical lens to make visible the workings of black-boxed algorithmic systems and can suggest alternative paths, albeit at a minor scale alongside other methods. As AI's hype-cycle accelerates these contributions can make important contributions to this debate.

Acknowledgments: This art project was commissioned by the Science Gallery Lab Detroit with support from Michigan State University. Further development of the project was made possible by a residency with UCD Parity Studios and The Insight Centre for Data Analytics at University College Dublin.

References

Angwin, Julia, et al. Machine Bias. Propublica, 23 May 2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.
Brock, André. “From the Blackhand Side: Twitter as a Cultural Conversation.” Journal of Broadcasting & Electronic Media, vol. 56, no. 4, 2012, pp. 529–49.
Bruns, Axel, and Jean Burgess. “The Use of Twitter Hashtags in the Formation of Ad Hoc Publics.” Proceedings of the 6th European Consortium for Political Research (ECPR) General Conference 2011, 2011, pp. 1–9.
D’Ignazio, Catherine, and Lauren Klein. Data Feminism. MIT Press, 2020.
Fitzpatrick, Noel, and John Kelleher. “On the Exactitude of Big Data: La Bêtise and Artificial Intelligence.” La Deleuzania, no. 7, 2018, pp. 142–55, https://doi.org/10.21427/dfw8-m918 .
Florini, Sarah. “Tweets, Tweeps, and Signifyin’: Communication and Cultural Performance on ‘Black Twitter.’” Television and New Media, vol. 15, no. 3, 2014, pp. 223–37, doi:10.1177%2F1527476413480247.
Graham, Roderick, and Shawn Smith. “The Content of Our #Characters: Black Twitter as Counterpublic.” Sociology of Race and Ethnicity, vol. 2, no. 4, 2016, pp. 433–49, doi:10.1177%2F2332649216639067.
Karpathy, Andrej. “The Unreasonable Effectiveness of Recurrent Neural Networks.” Andrej Karpathy Blog, 21 May 2015, https://karpathy.github.io/2015/05/21/rnn-effectiveness/.
Kelleher, John. Deep Learning. MIT Press, 2019.
Lee, Dave. “Why Big Tech Pays Poor Kenyans to Teach Self-Driving Cars.” BBC News, 3 Nov. 2019, https://www.bbc.com/news/technology-46055595.
McGarrigle, Conor. “Art in the Data-City: Critical Data Art in the Age of Surveillance Capitalism.” Digital Art in Ireland, Anthem Press, 2021, pp. 71-91.
McGarrigle, Conor. “Joyce Walks: Remapping Culture as Tactical Space.” Proceedings of 15th International Symposium on Electronic Art Belfast, ISEA International, 2009, pp. 440–47.
McGarrigle, Conor. “Preserving Born Digital Art: Lessons from Artists’ Practice.” New Review of Networking, vol. 20, no. 1–2, 2015, pp. 170–78.
Murthy, Dhiraj. Twitter. Polity, 2018.
Nilsson, Nils. The Quest for Artificial Intelligence. Cambridge University Press, 2010.
Noble, Safiya Umoja. Algorithms of Oppression: How Search Engines Reinforce Racism. New York University Press, 2018.
Pariser, Eli. The Filter Bubble : What the Internet Is Hiding from You. Viking Press, 2011.
Proudfoot, Diane. “Anthropomorphism and AI: Turing’s Much Misunderstood Imitation Game.” Artificial Intelligence, vol. 175, no. 5–6, 2011, pp. 950–57, doi:10.1016/j.artint.2011.01.006.
Salles, Arleen, et al. “Anthropomorphism in AI.” AJOB Neuroscience, vol. 11, no. 2, pp. 88–95, doi:10.1080/21507740.2020.1740350.
Scholz, Trebor. Digital Labor: The Internet as Playground and Factory. Routledge, 2013.
Sejnowski, Terrence J. “The Unreasonable Effectiveness of Deep Learning in Artificial Intelligence.” Proceedings of the National Academy of Sciences, Jan. 2020, p. 201907373, doi:10.1073/pnas.1907373117.
Sheng, Emily. The Woman Worked as a Babysitter: On Biases in Language Generation. 2019.
Varol, Onur, et al.. Online Human-Bot Interactions: Detection, Estimation, and Characterization. 2017.
Zhaung, Pinghui. “Chinese Woman Offered Refund after Facial Recognition Allows Colleague to Unlock IPhone X.” South China Morning Post, International Edition, 14 Dec. 2017, https://www.scmp.com/news/china/society/article/2124313/chinese-woman-offered-refund-after-facial-recognition-allows.

Zappavigna, Michele. “Ambient Affiliation: A Linguistic Perspective on Twitter.” New Media & Society, vol. 13, no. 5, Aug. 2011, pp. 788–806, doi:10.1177/1461444810385097.
Zuboff, Shoshana. The Age of Surveillance Capitalism. Public Affairs, 2019.