Have I Got Views For You!
Generating "Fair and Balanced" Interventions into Online Debates

Tony Veale
School of Computer Science
University College Dublin
Belfield, Dublin D4, Ireland
Tony.Veale@UCD.ie

Abstract

Online debate around divisive topics has become increasingly fractured, leading to the emergence of "echo chambers" in which disputants communicate almost exclusively with those who hold compatible views. To inhibit the growth of echo chambers and expose disputants to both sides of an argument – in ways that encourage dialogue across the divide – we aim to automate the generation of creative interventions into otherwise insular online debates. On highly echoic platforms such as Twitter, bot-driven interventions run contrary to best practices, and may be reported as an abuse of the system. However, passive interventions can instead use story generation to dramatise an ongoing debate. If the stories so generated are engaging and balanced, and are aptly labeled with attested hashtags, they can draw users to a bot's content, thus avoiding any need for a bot to elbow its content into a live conversation. The $Excelsior$ system, as described here, aims for amusing, even-handed engagement by packaging its data-driven stories as comic strips which integrate two sides of any argument into a single visual intervention.

Hold The Funny Pages!

It has been suggested that life is a tragedy for those who feel, and a comedy for those who think. We see this dichotomy writ large every day on social platforms such as Twitter, where discourse around contentious topics generates an excess of polarizing feeling and a comparable dearth of rational thought. Such platforms incentivize the articulation of short, pithy positions that prize outrage over insight, and in which interactions between opposing camps fall quickly to rancour. However, even rancorous exchanges may be preferable to the non-engagement with antagonistic stances that is too often observed on Twitter, for at least they can expose users to multiple points of view. Instead, inward-looking, defensive structures called $echo\ chambers$ (Barberá et al. 2015) insulate disputants from interactions with those with whom they are in dispute, and feed the growth of factionalism and the decline of real debate on Twitter.

Bots are an oft-aligned presence on Twitter, but one benign use of Twitterbots is the generation of interventions to foster engagement between holders of opposing views (Blaya 2019). Such interventions can cut to the heart of a dispute, by repackaging the nub of a conflict in an engaging form. Although many users follow bots out of an appreciation for their whimsical and oddly human-like outputs, few welcome unsolicited intrusions from bots in the form of direct messages, replies or mentions. Even bots that point out spelling errors may provoke vitriol or scorn (Veale and Cook 2017). After all, few of us like to be lectured by strangers, least of all automated strangers. Our goal in the $Excelsior$ system is the generation of narrative interventions that are as engaging as they are unthreatening, and which users can find for themselves via the use of attested hashtags.

Key to this engagement is the use of comic strips as a narrative medium. These strips originate in the "funny pages" of newspapers, where they were meant to entertain more than to educate, yet comics are a sequential art form (Eisner 1985; McCloud 1993) that is not limited to tales of funny animals and masked heroes. Here we aim for education and entertainment, to give data-driven stories about serious topics a harmless comedic form that is more likely to foster engagement than suspicion and outrage. Crucially, each newly created comic must balance two points of view, an argument and its converse, as articulated in the underlying data, which in the current case is the ongoing debate on Twitter about climate change, or vaccines, or guns, or abortion.

Excelsior proceeds by first identifying hashtags that convey a clear stance to a topic, such as *#EcoLiteracy*, *#FireFauci* or *#GetVaccinatedNow*, and then arranges related tags into sequences of mounting emotion, such as from curiosity to skepticism to disgust. An emotional inversion is performed mid-sequence, such as from disgust to admiration, to shift the narrative to an opposing viewpoint. The full sequence is then rendered as a comic, one panel per hashtag, that balances both points of view. The resulting comic can then be tweeted as an animated GIF along with the tags that punctuate its plot. Excelsior's approach to data storification does not aim to summarize the totality of a debate all at once. Rather, as we will show, it treats each debate as a space of views, and samples stories from this space in a way that, over time, cumulatively mirrors its emphases.

Back Issues: Related Work and Ideas

Comics are a medium for story-telling that requires a narrative impetus. For the *Comic Chat* system of (Kurlander, Skelly, and Salesin 1996), this impetus comes from the interactions of the users of online chatrooms. User texts are not summarized but placed verbatim into speech balloons above cartoon depictions of each user. Each conversational beat produces a single panel, and sentiment analysis is used to determine which variant of a user's comic avatar is associated with each speech act. But this impetus can also be machine-generated, and comics offer a viable medium for rendering automated stories, as in the story-to-comic generators of (Alves et al. 2007), (Pérez y Pérez, Morales, and Rodríguez 2012) and (Veale 2022). This can be modeled as a text-to-text generation task if each comic is specified using XML, as in the CSDL (Comic Strip Description Language) of (Alves et al. 2007), the CBML (Comic Book Markup Language) of (Walsh 2012) or the ComiXML of (Veale 2022). *Excelsior* builds upon the latter, ComiXML, as it allows a comic to be specified as a specific arrangement of visual assets, drawing from a repertoire of hundreds of different character poses and panel backgrounds.

This is a symbolic, componential approach to building comic strips, in contrast to the neural approaches typified by (Melistas et al. 2021) and (Proven-Bessel, Zhao, and Chen 2021). Neural approaches are trainable, and so are adaptable to specific data sets and visual genres (e.g., *Manga* in (Melistas et al. 2021), *Dilbert* in (Proven-Bessel, Zhao, and Chen 2021)). They are, in principle, capable of generating diverse images to match a given text prompt, although the visual outputs of the generative adversarial networks in (Melistas et al. 2021; Proven-Bessel, Zhao, and Chen 2021) are often blurry and ill-formed. Moreover, the relationship between image and dialogue, which is the crux of the comics medium, is difficult to control in such models. This relationship is crucial when comics are used to package interventions into a debate, especially when the goal is to balance opposing points of view.

Alternatively, images and text may be generated separately, by models that specialize in each. For instance, very large language models such as *GPT3* and *ChatGPT* can be used to generate stories for a given prompt (Xie, Cohn, and Lau 2023), in the desired form (e.g., a two-person dialogue, a one-act play). To provide a suitable context to the generator, the prompt may in turn be generated by existing narrative extraction methods (Santana et al. 2023), as applied to a debate corpus of interest. Individual text fragments can then be used to prompt an image generator such as *Dall-E* or *Stable Diffusion* (Gozalo-Brizuela and Garrido-Merchan 2023) to create a panel setting for each. But large language models (LLMs) are resource-intensive blackboxes that are not conducive to the development of small-footprint systems. Neither do LLMs yet permit easy interrogation of their logical processes, or offer guarantees as to whether their outputs convey the intended meanings. In contrast, a symbolic model can tick all of these boxes.

Data Collection, Organization and Analysis

We initially viewed each of the four debate spaces – climate change, vaccines, guns and abortion – as distinct, and collected four separate corpora of tweets via Twitter's streaming API, guided by seed sets of topic-related hashtags. We have come to realize that all four instantiate a single overarching debate concerning the acceptable balance of power between the state and the individual, and although each corpus has unique hashtags of its own, many tags – especially those of a political nature – recur across debate boundaries. Table 1 reports the number of distinct tweets and users comprising each corpus, noting how many are in fact retweets.

Dataset	**# Tweets**	**# Retweets**	**# Users**
Vaccines	1,624,173	1,244,009	391,489
Climate Change	1,017,087	691,333	340,836
Abortion	369,914	237,139	159,196
Gun control	205,535	131,728	62,387

Table 1: Size and makeup of the four debate datasets.

Table 2 reports the number of distinct hashtags in each dataset. While the raw counts (*# Tags*) are large, the number of distinct tags that convey a clear stance toward an explicit topic (*# Stanced*) is much smaller. These tags, in turn, conform to a smaller set of semantic *patterns*. These patterns are templates with semantic filler types that allow Excelsior to determine the stance and topic of each tag. For instance, the hashtag pattern #Fire{personal} is instantiated in 11 ways across the four debates, where {personal} can range from Fauci to DeSantis to Trudeau. The most varied patterns include #Get{solution} (30 fillers) and #No{solution} (29 fillers), #Pro{solution} (20) and #Anti{solution} (24), #Arrest{personal} (25) and #LetsGo{personal} (11), #Boycott{business} (12) and #Boycott{place} (20), #No{problem} (19) and #Stop{problem} (31).

Dataset	**# Tags**	**# Stanced**	**# Patterns**
Vaccines	39,366	4,236	1,986
Climate Change	32,375	1,993	1,043
Abortion	18,563	1,344	652
Gun control	16,090	957	470
All four debates	90,323	6,982	2,985

Table 2: Raw and processed hashtag counts per dataset.

Just as sets of domain-specific "seed" hashtags are used to collect each individual debate dataset via Twitter's streaming API, a set of seed entities is also used to drive the mapping of newly collected tags to generic patterns; e.g., "Greta"={person}, "covid"={problem} and "vax"={solution}. A bootstrapping process is used to identify candidate patterns among hashtags in which camel-casing indicates a multi-word structure, and for which sentiment analysis indicates a positive or negative stance, such as #LetsGoBiden, #PleaseVaxUp and #EndCovidNow. Replacing any known entities in these tags gives us the candidate patterns #LetsGo{person}, #Please{solution}Up and #End{problem}Now. These candidates are manually curated, and added to Excelsior's lexicon only when they convey a clear stance toward the referenced entity. But these additions can, in turn, be used to suggest new entities, by matching the pattern against other hashtags. For example, since #LetsGo{person} also matches #LetsGoDeSantis, the entity "DeSantis"={person} is also offered as an addition to the lexicon. These new entities then allow further patterns to be identified in the data, such as #{personal}2024, #LockUp{personal} and #{personal}Lies. A candidate pattern may unite multiple entities, such as #{personal}Failed{place}, which matches #TrumpFailedAmerica as well as #DeSantisFailedFlorida.

Each curated pattern is associated with a firm stance, either $accepting$ or $rejecting$, toward a referenced entity. This must be done manually because online debate is fast-moving and sentiment analysis is so often wrong. For example, #LetsGo{person} is actually a $rejecting$ rebuke, and not an $accepting$ endorsement of {person}, due the peculiar origins of the jeer #LetsGoBrandon. Each tag pattern is also linked to an emotional framing, which offers a finer view of the feeling being articulated. So #{personal}2024 evokes an $election$ framing while #LockUp{personal} evokes a $prison$ framing. A framing allows a hashtag to be visualized as a comic panel with apt character poses, apt dialogue and an apt backdrop. Thus, the $prison$ frame suggests someone holding keys to an other's cell, while the $election$ frame suggests one voting for another in a poll centre, and so on.

Each hashtag pattern is linked to one or more of 96 framings, such as $battle$, $freedom$, $contempt$ and $hoax$. A framing often represents a metaphorical perspective, such as $battle$ or $slavery$, or an intense feeling, such that a given problem or solution is a $hoax$. We choose them for their dramatic potential, as well as for their suitability to the sampled tags. Each dramatic frame is associated with a set of apt dialogue patterns, for both a protagonist (the main speaker) and an antagonist (one holding an opposing view). For instance, #Fake{solution}, #Phony{solution}, #{solution}Cult and #{solution}Con typify a large family of similar tags that are linked by the $hoax$ framing. In this context, the dialogue patterns "Expose {solution} as a fake!", "Unmask {solution} as a fraud!", and "I hate the hypocrisy of {solution}!" are available to the protagonist, while the patterns "What's the issue with {solution}?", "Why are you down on {solution}?" and "What's so wrong with {solution}?" are possible responses for the antagonist.

The set of 96 framings is organized as a graph. One framing links to another if the second adds to the feelings of the first, thus serving to build the debate (or comic) toward an emotional crescendo. So, for instance, $scepticism$ can lead to $denial$ or $blame$. which can lead to a call for $defunding$ or an accusation of $tyranny$ or $hoax$. In turn, $tyranny$ can lead to cries of $treason$ or $fascism$, where $treason$ can lead to calls for $prison$.

The ABCs of Comic Generation

This graph allows Excelsior to organize hashtags in the data into plot-like sequences that build to a dramatic climax, even if those tags were never used in the same tweets or even by the same users in the original dataset. Every random walk in this graph produces a valid plot, although Excelsior must then ground the constituent frames in actual hashtags that refer to the same topic. It is also not enough to articulate just one viewpoint on a topic. Rather, the "plot" should switch from one side to another at some turning point in the narrative, and thereby allow the antagonist to become the protagonist. To facilitate this switch of perspectives, the graph also links framings to those that express opposing emotions. For instance, $treason$ is thus linked, by opposition, to $heroism$, $election$, and $admiration$. Note however that these transitions at the framing level are only pursued if there are actual hashtags in the data to support them. A plot can switch from $treason$ to $election$ with regard to topic X only if the data contains tags that imply that X is a traitor, and tags that call for X to be elected.

To dictate the general shape of a plot, we employ the AAB string notation. The place holders A and B can denote any framing, but the sequences AA and BB can only denote a transition from one framing to another more intense framing on the same side of the debate, as allowed for by the framing graph. Conversely, AB and BA can only denote a transition between frames on either side of the debate, as allowed for by the graph.

A plot with the shape AAB is thus realized as a comic in which a particular stance toward a given topic is established in one frame/panel, intensified in the second, and rebutted in the third. This generic AAB pattern is an example of what (Loewenstein, Raghunathan, and Heath 2011) call a repetition-break structure, in which a norm is first established by repetition and then dashed to produce a humorous or creative effect. Those authors provide evidence for the pattern's popularity and effectiveness in eye-catching TV adverts, while (Loewenstein 2018) argues for the utility of the pattern in constructing materials designed to spread rapidly across social networks. We further generalize the AAB pattern here to allow for controlled repetition of the norm and its opposite. Fig. 1 presents a comic created by Excelsior for the pattern $AAAABBB$, as applied to the joint dataset. The system picks the topic $climate\ change$, and balances views for and against the topic in the comic.

The joint dataset combines tweets and tags from all four of the debates in Table 1. When an explicit topic is provided, such as $carbon$, Excelsior confines itself to tags that focus on that topic. To offer the data-fitting process some wiggle-room, we define a topic graph to connect related ideas for which a stance toward one translates to a stance toward the other, such as $climate$ and the $environment$, $carbon$ and oil, or $Biden$ and the $Democrats$. This allows Excelsior to veer from one topic to another when instantiating its AABs, to generate more varied comics while staying on-message.

As shown in Fig. 1, each hashtag that instantiates the A/B elements of the $AAAABBB$ pattern is given its own panel, under which the original tag is displayed. Each comic uses two characters, which are rendered in blue and red to make them visually separable. This visual identity is important when the viewpoint switches from one side of the debate to the other, as happens here in the second panel of row two. The protagonist, shown in blue, advances the A side of the argument on climate change, and here advances the pro-green agenda. The antagonist, shown in red, responds with as many questions as rebuttals. Excelsior strives for balance across panels and within panels too, and generally aims to let no claim go unquestioned, whatever its validity. When the agonists switch sides, it becomes red's turn to voice the anti-green B side in the face of blue's advocacy.

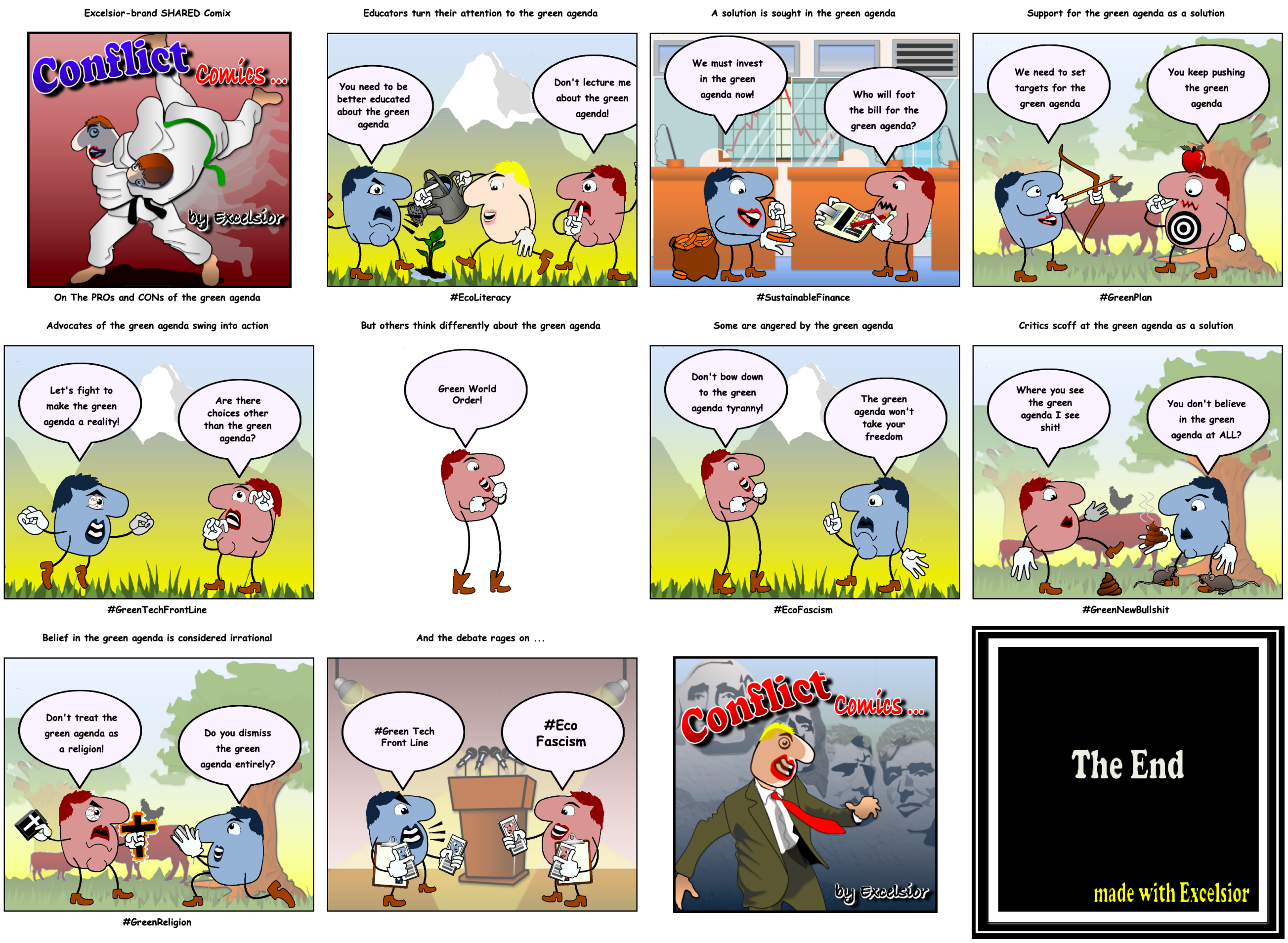

Figure 1: An Excelsior comic in the domain of climate change. Stance reversal occurs in the 2nd panel of row 2.

The AABs of Irony Generation

(Rozin et al. 2006) show that the AAB pattern is more effective than any other variation (e.g., AB or $AAAB$) at inducing a humorous response to a creative stimulus. One consequence of using comic strips to package the products of data "storification" is that stances which are already emotionally intense are tipped into humorous exaggeration by a vividly expressive rendering. The AAB pattern is used here to inject conflict and balance into each comic, but any emergent humour is ultimately unplanned. Still, we can foster humour by using the AAB pattern in its purest form, with data that has been explicitly chosen for its humorous potential.

The internet is replete with humorous content, such as joke lists, that can be injected into a comic. These resources, though large, are often problematic, since they lean heavily on racism, sexism and homophobia. (Tang et al. 2022) present a transformer for detecting offense in Reddit joke lists, but offer no means of controlling the meaning of a joke or making it fit a given context. There is little point in forcing an arbitrary joke about farming, say, into a comic on this topic if Excelsior cannot know which side of a debate the joke is on. Rather, we need a more controlled source of humour that cleanly interfaces with the assertions implied by each hashtag. For this we turn to the "about" similes of (Veale 2012).

Humour involves playful insincerity, so to avoid serious misunderstandings, humorists often provide subtle but predictable cues to their insincerity. In the case of exaggerated or ironic similes, these cues are found in hedge words like "about" or "almost." Take the heavily panned film *Cats* (2019). After viewing an unappealing trailer, one might describe the film as "*about* as marketable as a flesh-eating virus." These cues serve a dual function: they signal a creative intention on the part of a writer, and allow machines to trawl large quantities of creative similes from the web. (Veale 2012) reports that such a trawl pulls in a large set of ironic similes, in which one quality is asserted but its opposite is implied, and a larger set of comical similes whose qualities are asserted literally. If Excelsior can infer the qualities implied by a specific tag framing, it can make the qualities comically explicit by using vivid similes from this corpus. It can also exploit the AAB pattern to magnify the humour of its choices, by chasing two literal similes (AA) for an implied quality with an ironic twist (B).

Figure 2: An Excelsior comic on the topic of vaccination which follows an AAB irony pattern.

(Hao and Veale 2010) present a means of separating ironic from literal "about" similes, noting that positive qualities (like $marketable$) are often intended ironically, while negative ones are more often intended literally. As noted earlier, Excelsior maps hashtags like #FireFauci to patterns such as #Fire{person}, and further maps those patterns to framings like $rejection$ and $contempt$. We now associate these framings with the qualities they imply of their referents, for instance, that the referent of #Fire{person} is neither competent nor welcome, or that the focus of #{solution}Farce is hardly credible. An AAB pattern can now be crafted from a single tag like #JabFarce, as illustrated in the comic of Fig. 2. Note how the quality $credible$ is treated literally for two comparisons before it is subverted by irony in a third. Irony offers balance even in the case of a single hashtag.

Nonetheless, Excelsior is careful to balance the scales. Just as the comic opens with a panel visualizing the tag #JabFarce via the framing $contempt$, it closes with one visualizing an opposing view, #AntiVaccineMadness, via the antithetical framing $defence$. The core conflict between these views is then summarized in a final panel.

Experiments in Transformation

As a generator of topical comics, the Excelsior system is both knowledge-driven and data-driven. Its comics reflect real tensions in social-media data sets that are growing and evolving in real-time, and it uses top-down knowledge-structures to make sense of this data. The comics themselves are specified using an XML schema that assembles a fixed repertoire of poses and settings into LEGO-like dioramas, but they are filled with dialogue that, while apt, relies on pre-scripted templates. These trade-offs make Excelsior responsive *and* controllable, but the surprises in its comics come from the data, which is constantly changing, and not the system's own knowledge, which evolves at a much slower pace.

Symbolic systems make poor learners, but they can still serve as good teachers. To see why, consider how a pre-trained neural model is fine-tuned for a new task. A transformer language model such as the $T5$ (Raffel et al. 2020) can be further trained on a set of input/output text pairs, so that it can learn to map from a given input text to the desired output text. We can, for instance, fine-tune a $T5$ on a set that maps domain-specific tweets onto the XML specifications

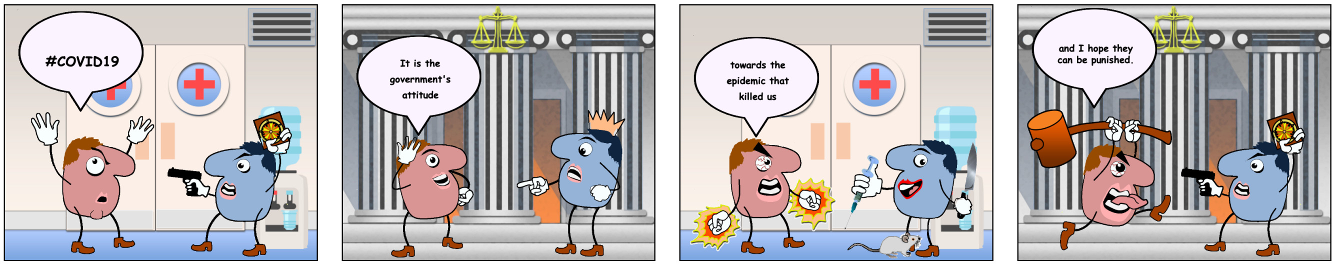

Figure 3: A comic generated by a T5-small transformer that is fine-tuned on a dataset of Covid/vaccine tweets.

of the corresponding comics. Indeed, a $T5$-$small$ model of 60 million parameters is sufficient to the task of learning the text-to-XML mapping for a domain such as vaccines or climate change. We can see this with a corpus of 1,500 Covid/vaccine tweets which have been manually annotated with ComiXML; of these, we hold back 150 for validation and 150 for testing. Fig. 3 shows the comic that is rendered from the XML output for an unseen test tweet: "#Covid19 It is the government's attitude towards the pandemic that killed us and I hope they can be punished." Note how the dialogue is one-sided, and repurposes the text of the tweet, but does so in a way that is visually expressive and emotionally apt.

Notice also how the action switches from a hospital setting (where the dialogue touches on medical issues) to a legal setting (where the dialogue touches on governance and law) and back again. Because its fine-tuning tweets are segmented by XML mark-up tags, the transformer learns to segment each new tweet into balloon-sized morsels of dialogue. In each case, the transformer assigns poses to the agonists that match both their explicit interactions (e.g., panels 2-4) and their implicit stances (e.g. panel 1). Here the opening panel aptly sets the scene, and serves to foreshadow the scepticism of the protagonist (in red) in the following panels.

The $T5$ performs well on new tweets, and learns how to use the ComiXML schema well in the Covid/vaccine domain. The value of XML as an output format cannot be understated, as it allows a generator to automatically check the validity of the transformer's outputs. On the rare occasions when these are malformed – e.g., when the XML is not schema-compliant, or when it invents new poses or settings – a new output can be re-sampled from the same input.

But the transformer does not generalize well beyond its specific domain. When presented with tweets lacking an overt focus on Covid or vaccines, it cannot but view them through a monocultural lens. It continues to place characters in hospital and graveyard settings, as though perceiving a subtext that is invisible to human readers. When we repeat our experiments with a new fine-tuning corpus of 1500 tweets, this time on the topic of climate change, we observe the same outcome. The transformer performs very well on new in-domain tweets, but does not generalize robustly beyond this domain. The Covid transformer is well-formed but inept in its handling of climate issues, and the climate transformer is similarly inept in its handling of vaccines. The situation improves when a transformer is fine-tuned on a joint corpus for both domains, but it still fails to generalize well to other domains, such as gun control and abortion. Moreover, it is costly to fine-tune the $T5$ for each new domain. We find that it takes 2 to 3 person weeks of effort to collect and mark up each new set of 1,500 training tweets.

This is where a symbolic teacher can step in. Though its dialogue patterns are limited in number, such a teacher can generate dialogue for specific topics in a new domain. Its outputs will be guided by attested hashtags in the domain, so it will produce short texts that are representative of the feelings swirling about those topics in the given dataset. It can also produce the XML comic specifications for those texts, to automatically generate both sides of the input/output training pairs for the transformer. A symbolic Excelsior can thus lend its ability to generalize, via templates, to a learner with an unsure footing in a new domain. The $T5$ can now be periodically fine-tuned on the new example sets.

Template-based generation becomes more stilted and predictable with time. Excelsior's dialogue model can talk about new topics, but only in the same old ways. To expose a learner like our $T5$ to fresh ideas *and* fresh ways of talking, we need fresh data. Fortunately, a symbolic teacher that can interpret new hashtags in terms of existing patterns can easily find tweets that use those tags. It can fine-tune the learner by pairing these tweets with comics it produces from the tags. in this way, a symbolic teacher can greatly reduce the time taken to create a training set for a new domain.

We have some way to go before the symbolic Excelsior is inevitably usurped by its statistical student. For now, only the symbolic model can offer a complete solution to the generation of comic strips that balance the views of multiple users across competing "echo chambers." This model will be replaced piecemeal rather than all at once, as transformers learn to improve on its individual parts.

Moral Dimensions and Dilemmas

The generation of comics with carefully balanced meanings is a means to an end rather than an end unto itself. These comics serve as interventions into a fractious online debate, so as to expose disputants to all sides of an issue. They are not intended to provide answers but to raise questions and foster discussion. Yet, in doing so, they also pose some difficult questions for their creators.

Some disputes make it difficult to stay above the fray. Is there a moral imperative to take a side when some actors spread conspiracy-fuelled misinformation and play fast and loose with scientific facts? Balance is surely a desirable quality, but is it always right or wise to give exposure to extreme views in the interests of fairness? Each time we encourage debate between opposing sides, we run the risk that more, not fewer, people will adopt the controversial views that we put under the spotlight. Yet, to serve as an honest broker that appeals equally to both sides, a creative system cannot afford to be partisan. This refusal to hold opinions of its own can make a creative system seem indifferent and amoral, a purveyor of what (Frankfurt 1986) famously called "bullshit." It is, it seems, a question of balancing one harm against another: are echo chambers so detrimental to our social discourse that these other risks are worth taking?

A "fair and balanced" creative system can manifest bias in subtle ways. For instance, it might always grant the last word on a topic to one side of a debate, e.g., to show there is a clear reply to every objection to vaccines or to every doubt about climate change. The ordering of claims in a comic can make a certain position seem like an argument's end-point rather than a starting point. A system that uses humour to promote engagement may not use its humour even-handedly, and may, for example, make certain views the preferred butt of its jokes, or use more risible visual representations of those views. We must give systems knowledge but not opinions, and be shrewd enough to distinguish one from the other. This is challenging whether one is building a top-down symbolic system or a bottom-up statistical system.

The most trenchant views on Twitter often involve *ad hominem* attacks, but should a system repeat these even if it balances them with supportive counter-points? In politics, such attacks are a way of life and the cost of doing business, but what of others in the public sphere? Public figures make for good "extras" in a comic strip, because they lend an emotive face to non-visual ideas. The white-haired figure of Tony Fauci and the pig-tailed figure of Greta Thunberg bring concepts such as public health policy and climate change down to human-scale. Since each is an effective champion *and* a lightning rod for controversy, Excelsior uses both in its comics, for the same reason they anchor so many tags in the data. We want Excelsior to treat all public figures equally, but do not want to aid the demonization of certain individuals. Excelsior must somehow refrain from giving a comic form to the worst excesses in the underlying data.

A sly wit is sometimes the saving grace of an ad hominem attack, but some topics are just too serious to ever be treated humorously. We scarcely want a creative system to make jokes about rape or the holocaust, but how and where do we draw the line? (Veale 2021) identifies two kinds of self-regulation for a creative system: *inner* and *outer* regulation. Any system using an inner regulator explores a modified search space that omits certain topics which can give rise to offense. So, by choosing not to put rape or the holocaust into its lexicon of possible hashtag referents, Excelsior becomes blind to those topics and will not use them in its comics. In this respect, the traditional knowledge bottleneck in symbolic systems can sometimes work to our advantage.

A system with an outer regulator does not explore a reduced search space, and so is capable, in principle, of treating sensitive topics in crass and insensitive ways. Instead, a filter is used to catch any potentially offenses before they can be shared with users or the public. For example, a "blocklist" might list the terms that a system must avoid. The filter is applied retroactively, so a system may explore, but not actually speak of, certain ideas. It follows that inner regulation makes more sense for a symbolic system whose knowledge is curated and pruned with care. Outer regulation, in contrast, is more suited to statistical systems that learn from real data. A hybrid system that uses a symbolic teacher to tune a statistical learner will use both kinds of regulation.

As such, Excelsior draws on both kinds of regulation to create comics that are informative, provocative and balanced. Yet, while the system is presently poised to fulfil its intended social function – automated intervention into ongoing debates – the foregoing ethical issues still give us sufficient pause to delay Excelsior's launch as an autonomous Twitter *bot*. An abundance of care is needed whenever one aims to balance potential harms against each other. Further testing is needed to quantify Excelsior's capacity to offend, since any system with the capacity to surprise may also shock and dismay.

Summary and Conclusions

The classic 1950s crime drama *Naked City* ended each episode with these words: "There are eight million stories in the naked city. This has been one of them." It seems natural to feel the same way about a large data set, such as a corpus of polarized views gathered from Twitter. This data does not tell a single story but many, and we must do justice to them all when we set out to creatively capture an overall sense of its contents.

Excelsior is a system for creating topical comics from an evolving social-media data set. It is a modular system that separates the planning of a comic – its plot, emotional cadence, and core opposition of views – from its visual rendering. For the former, Excelsior generates an XML specification of a comic which is human- and machine-readable, and for the latter it uses a bespoke renderer. The sequences of images in Figs. 1 and 2, for instance, have been generated by such a renderer. Its stories are composites, drawn from multiple sources and multiple related – or opposing – viewpoints. However, these composites still do justice to the data, by making explicit the narratives that connect different users, hashtags and tweets within and across echo chambers. Crucially, Excelsior balances the views in its comics, so that no single position is favored or goes unchallenged.

As a symbolic system, Excelsior relies on a number of explicit representations, which allow it to map hashtags onto topic-relative stances and emotions, and from there onto visual actions and textual dialogue. It is the logical coding of these representations that allows us to tightly maintain Excelsior's sense of balance. However, for the system to grow in expressive power, we need to make it learn for itself. So, mindful of the moral dilemmas that already attach to the symbolic model, and of how these might be exacerbated by inappropriate training data, we are tentatively exploring how the symbolic Excelsior might train its own statistical replacement.

Author Contributions

This paper is wholly the work of the principal author.

References

Alves, T.; McMichael, A.; Simões, A.; Vala, M.; Paiva, A.; and Aylett, R. 2007. Comics2d: Describing and creating comics from story-based applications with autonomous characters. In *Proceedings of CASA, the 20th Annual Conference on Computer Animation and Social Agents*.

Barberá, P.; Jost, J.; Nagler, J.; and Tucker, J. 2015. Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science* 26(10):1531–1542.

Blaya, C. 2019. Cyberhate: A review and content analysis of intervention strategies. *Aggression and violent behavior* 45:163–172.

Eisner, W. 1985. *Comics & Sequential Art*. Tamarac, Florida: Poorhouse Press.

Frankfurt, H. 1986. On bullshit. *Raritan Quarterly Review* 6(2):81–100.

Gozalo-Brizuela, R., and Garrido-Merchan, E. 2023. Chatgpt is not all you need. a state of the art review of large generative ai models. *ArXiv* abs/2301.04655.

Hao, Y., and Veale, T. 2010. An ironic fist in a velvet glove: Creative mis-representation in the construction of ironic similes. *Mind and Machines* 20:635–650.

Kurlander, D.; Skelly, T.; and Salesin, D. 1996. Comic chat. In *Proceedings of SIGGRAPH'96, the 23rd annual conference on Computer graphics and interactive techniques*. ACM. 225–236.

Loewenstein, J.; Raghunathan, R.; and Heath, C. 2011. The repetition-break plot structure makes effective television advertisements. *Journal of Marketing* 75(5):105–119.

Loewenstein, J. 2018. Surprise, recipes for surprise, and social influence. *Topics in Cognitive Science* 11.

McCloud, S. 1993. *Understanding Comics: The Invisible Art*. New York, NY: Harper Collins.

Melistas, T.; Siglidis, Y.; Kalogiannis, F.; and Manouach, I. 2021. A deep learning pipeline for the synthesis of graphic novels. In *Proceedings of ICCC-21, the 12th International Conference on Computational Creativity*. 256–265.

Proven-Bessel, B.; Zhao, Z.; and Chen, L. Y. 2021. Comicgan: Text-to-comic generative adversarial network. *ArXiv* abs/2109.09120.

Pérez y Pérez, R.; Morales, N.; and Rodríguez, L. 2012. Illustrating a computer generated narrative. In *Proceedings of ICCC-22, the 3rd International Conference on Computational Creativity*. 103–110.

Raffel, C.; Shazeer, N.; Roberts, A.; Lee, K.; Narang, S.; Matena, M.; Zhou, Y.; Li, W.; and Liu, L. J. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research* 21:1–67.

Rozin, P.; Rozin, A.; Appel, B.; and Wachtel, C. 2006. Documenting and explaining the common aab pattern in music and humor: Establishing and breaking expectations. *Emotion* 6:349–55.

Santana, B.; Campos, R.; Amorim, E.; Jorge, A.; Silvano, P.; and Nunes, S. 2023. A survey on narrative extraction from textual data. *Artificial Intelligence Review*.

Tang, L.; Cai, A.; Li, S.; and Wang, J. 2022. The naughtyformer: A transformer understands offensive humor. *arXiv* abs/2211.14369.

Veale, T., and Cook, M. 2017. *Twitterbots: Making machines that make meaning*. Cambridge, MA: MIT Press.

Veale, T. 2012. *Exploding the Creativity Myth: The Computational Foundations of Linguistic Creativity*. Bloomsbury Academic.

Veale, T. 2021. Walking the line in computational creativity: Paradox and pitfall in the mitigation of automated offense. In *Proc. of the 12th International Conference on Computational Creativity*. 132–139.

Veale, T. 2022. Two-fisted comics generation: Comics as a medium and as a representation for creative meanings. In *Proceedings of ICCC-22, the 13th International Conference on Computational Creativity*. 59–66.

Walsh, J. 2012. Comic book markup language: An introduction and rationale. *Digital humanities quarterly* 6(1).

Xie, Z.; Cohn, T.; and Lau, J. H. 2023. Can very large pretrained language models learn storytelling with a few examples? *ArXiv* abs/2301.09790.