McKinsey&Company

AI in storytelling: Machines as cocreators

Consumer Tech and Media December 2017

Eric Chu
Jonathan Dunn
Deb Roy
Geoffrey Sands
Russell Stevens

AI in storytelling: Machines as cocreators

Computers don't cry during sad stories, but they can tell when we will.

Sunspring **debuted** at the SCI-FI LONDON film festival in 2016. Set in a dystopian world with mass unemployment, the movie attracted many fans, with one viewer describing it as amusing but strange.[1] But the most notable aspect of the film involves its creation: an artificial-intelligence (AI) bot wrote *Sunspring's* screenplay.

"Wow," you think. "Maybe machines will replace human storytellers, just like self-driving cars could take over the roads." A closer look at *Sunspring* might raise some doubts, however. One character in the film inexplicably coughs up an eyeball, and a critic noted that the dialogue often sounds like "a random series of unrelated sentences."[2] Until the technology advances, we still need rumpled screenwriters bent over keyboards. So let's envision a less extreme scenario: could machines work *alongside* humans to improve the storytelling process?

Imagine how this collaboration might unfold in the rich medium of video. As always, human storytellers would create a screenplay with clever plot twists and realistic dialogue. AI would enhance their work by providing insights that increase a story's emotional pull—for instance, identifying a musical score or visual image that helps engender feelings of hope. This breakthrough technology would supercharge storytellers, helping them thrive in a world of seemingly infinite audience demand.

The Massachusetts Institute of Technology (MIT) Media Lab recently investigated the potential for such machine–human collaboration in video storytelling. Was it possible, our team asked, that machines could identify common emotional arcs in video stories—the typical swings of fortune that have characters struggling through difficult times, triumphing over hardship, falling from grace, or declaring victory over evil? If so, could storytellers use this information to predict how audiences might respond? These questions have resonance for anyone involved in video storytelling, from amateurs posting on YouTube to studio executives.

Emotional arcs: The backbone of stories

Before getting into the research, let's talk about emotional arcs. Master storytellers—from Sendak to Spielberg to Proust to Pixar—are skilled at eliciting our emotions. With an instinctive read of our pulse, they tune their story to provoke joy, sadness, and anger at crucial moments. But even the best storytellers can deliver uneven results, with some Shakespeare plays leaving audience members feeling indifferent or disconnected. (There aren't many big fans of *Cymbeline* out there.) What accounts for this variability? We theorize that a story's emotional arc largely explains why some movies earn accolades and others fall flat.

The idea of emotional arcs isn't new. Every storytelling master is familiar with them, and some have tried to identify the most common patterns. Consider Kurt Vonnegut's explanation of arcs.[3] The most popular arc, he claims, follows the pattern found in *Cinderella*. As the story begins, the main character is in a desperate situation. That's followed by a sudden improvement in fortune—in Cinderella's case provided by a fairy godmother—before further troubles ensue. No matter what happens, *Cinderella*-type stories end on a triumphant note, with the hero or heroine living happily ever after.

There's evidence that a story's emotional arc can influence audience engagement—how much people comment on a video on social media, for example, or praise it to their friends. In a University of Pennsylvania study, researchers reviewed *New York*

Times articles to see if particular types were more likely to make the publication's most emailed list.[4] They found that readers most commonly shared stories that elicited a strong emotional response, especially those that encouraged positive feelings. It's logical to think that moviegoers might respond the same way.

Machines as moviegoers: MIT's radical experiment

Some researchers have already used machine learning to identify emotional arcs in stories. One method, developed at the University of Vermont, involved having computers scan text—video scripts or book content—to construct arcs.[5]

We decided to go a step further. Working as part of a broader collaboration between MIT's Lab for Social Machines and McKinsey's Consumer Tech and Media team, we developed machine-learning models that rely on deep neural networks to "watch" small slices of video—movies, TV, and short online features—and estimate their positive or negative emotional content by the second.

These models consider all aspects of a video—not just the plot, characters, and dialogue but also more subtle touches, like a close-up of a person's face or a snippet of music that plays during a car-chase scene. When the content of each slice is considered in total, the story's emotional arc emerges.

Think about this for a moment: machines can view an untagged video and create an emotional arc for the story based on all of its audio and visual elements. That's something we've never seen before.

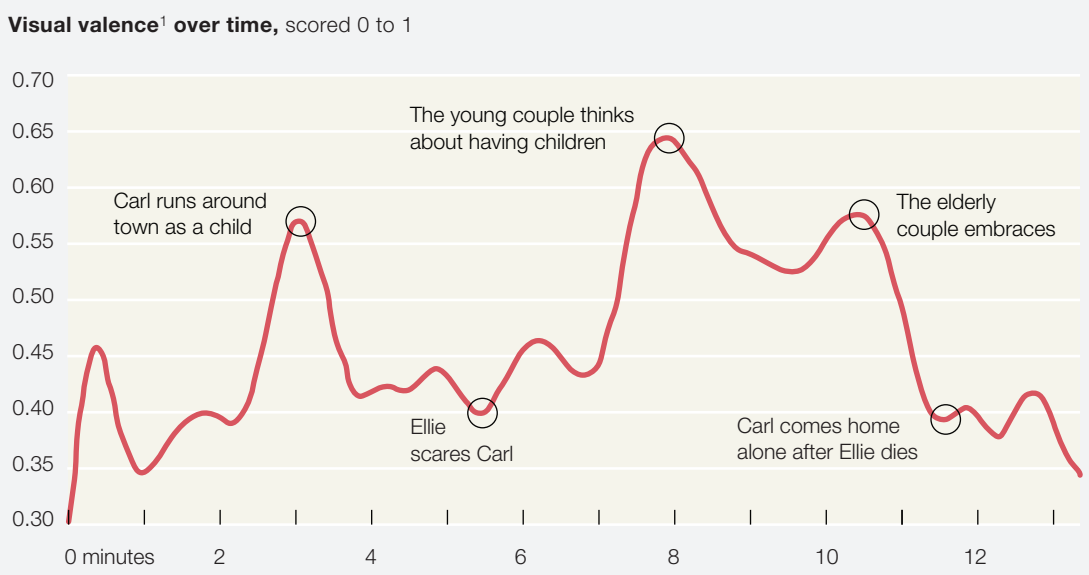
Consider the famous opening sequence of *Up*—a 3-D computer-animated film that was a critical and popular hit. The movie focuses on Carl Fredricksen, a grumpy senior citizen who attaches thousands of balloons to his house in a quest to fly to South America after his wife, Ellie, dies. Wanting to devote most of the movie to Carl's adventure, the screenwriters had to come up with a quick way to provide the complicated back story behind his trip. That's where the opening sequence comes in. It's silent, except for the movie's score, and an emotional arc emerges as scenes of Carl's life play on the screen. (We also looked at the arc for the movie as a whole, but this is a good way to view one in miniature.)

You can see the high and low points of the montage in the graph (Exhibit 1). The x-axis is time, measured in minutes, and the y-axis is visual valence, or the extent to which images elicit positive or negative emotions at that particular time, as scored by the machine. The higher the score, the more positive the emotion. As with all our analyses, we also created similar graphs for a machine's responses to audio and to the video as a whole. We're focusing on the visual graphs, here and elsewhere, since that was the focus of our later analyses of emotional engagement.

Visual valence is measured on a scale of 0 to 1, but not every film has images that span the entire spectrum. What's important is the relative valence—how positive or negative a scene is compared with other points in the movie—as well as the overall shape of the emotional arc. As in many video stories, the arc in *Up*'s opening montage contains a series of mood shifts, rather than a clear upward or downward trajectory. One of the highest peaks corresponds to images of Carl as a happy child, for instance, but there's a big drop shortly after, when young Ellie scares him in the middle of the night. The machine's negative response reflects Carl's fright. Other peaks emerge much later, when the newlyweds are planning to have children, or when the elderly couple embraces. The valence plummets near the end, when Carl returns home alone after Ellie dies.

MIT's machine-learning models have already reviewed thousands of videos and constructed emotional arcs for each one. To measure their

Exhibit 1 **The emotional arc in *Up*'s opening sequence, as scored by a machine, shows highs and lows in line with positive or negative moments.**



[1] Visual valence is scored by machine on a scale of 0 to 1. The higher the score, the more positive the emotional response.
Source: Massachusetts Institute of Technology, Lab for Social Machines

accuracy, we asked volunteers to annotate movie clips with various emotional labels. What's more, the volunteers had to identify which video element—such as dialogue, music, or images—triggered their response. We used these insights to refine our models.

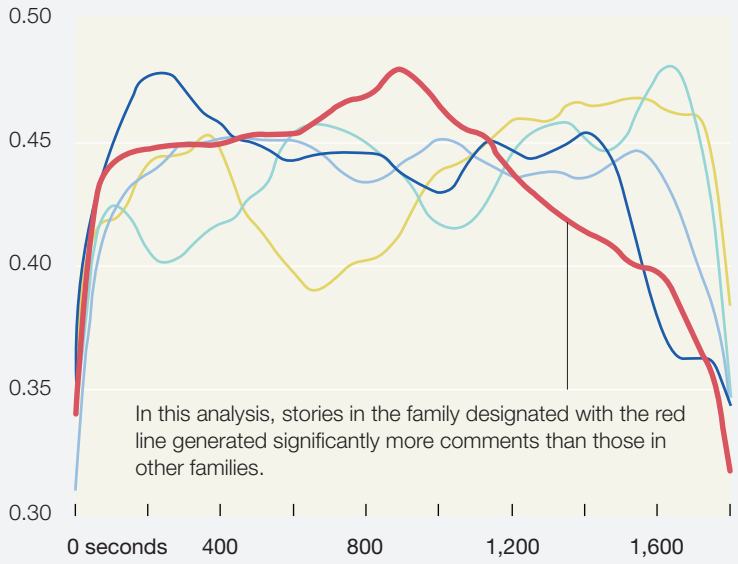
Finding 'families': Common emotional arcs

After sifting through data from the video analyses, we developed a method for classifying stories into *families* of arcs—in other words, videos that share the same emotional trajectory. Our approach combines a clustering technique, called k-medoids, with dynamic time warping—a process that can detect similarities between two video sequences that vary in speed.

We looked for arc families in two separate data sets—one with more than 500 Hollywood movies and another with almost 1,500 short films found on Vimeo. Our preliminary analysis of visual valence revealed that most stories could be classified into a relatively small number of groups, just as Vonnegut and other storytellers suspected. Exhibit 2 shows that the arcs that emerge with the videos in the Vimeo data set are clustered into five families.[6] For the family designated by the yellow line, for instance, there's a surge in negative emotion fairly early in the video, followed by sustained positive emotion near the finale. (All movies tend to score low at the beginning and the end, as the machine snores through the credits.)

Exhibit 2 **Some families of stories generate more comments than others.**

Visual valence[1] **over time,** scored 0 to 1



[1] Visual valence is scored by machine on a scale of 0 to 1. The higher the score, the more positive the emotional response.
Source: Massachusetts Institute of Technology, Lab for Social Machines

Computers as crystal balls: Predicting audience engagement

Seeing how stories take shape is interesting, but it's more important to understand how we can use these findings. Does a story's arc, or the family of arcs to which it belongs, determine how audiences will respond to a video? Do stories with certain arcs predictably stimulate greater engagement?

Our team attempted to answer these questions by analyzing visual data for the Vimeo short-film data set. (We chose to focus on visual arcs in the analysis discussed here because they were more closely linked to video content than audio, and the combined arcs present some analytical challenges.) For each story, we used a regression model to consider arc features while controlling for various metadata that can affect online reaction, such as the video length and upload date.

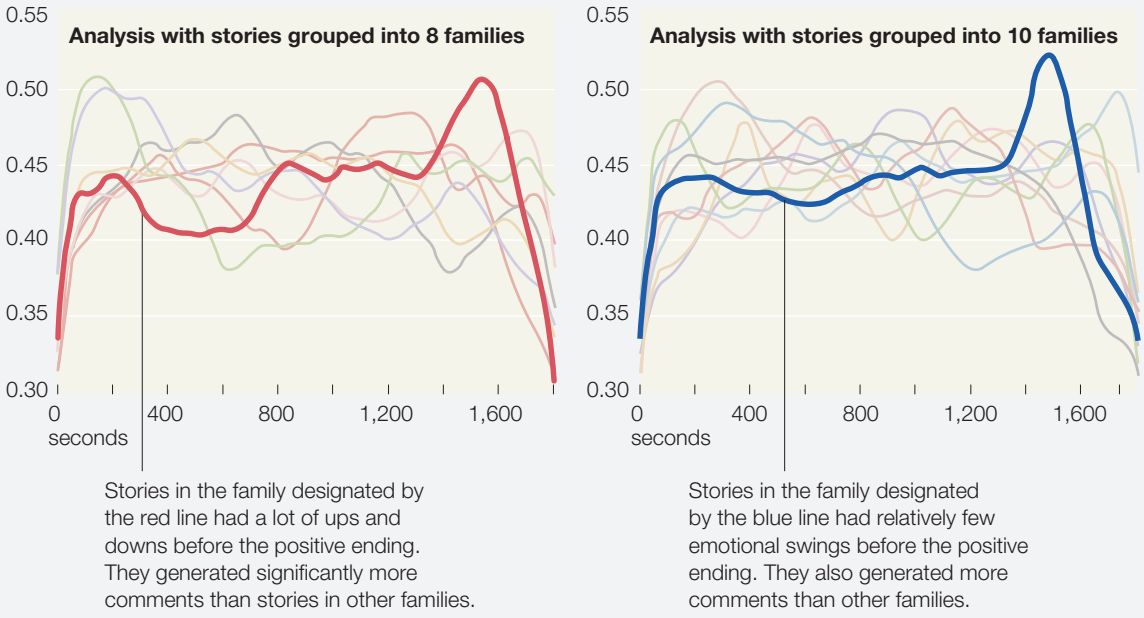
The goal was to predict the number of comments a video would receive on Twitter and other social media. In most cases, a large volume of comments signals strong audience engagement, although there can be some caveats. If a movie bombs—think *Gigli* and *Ishtar*—it could also generate lots of online commentary, but not in a good way.

In the Vimeo analysis, visual arcs indeed predicted audience engagement, with movies in several

Exhibit 3

The families where stories had a large positive spike toward the end tended to generate most comments.

Visual valence[1] **over time,** scored 0 to 1



[1] Visual valence is scored by machine on a scale of 0 to 1. The higher the score, the more positive the emotional response.

Source: Massachusetts Institute of Technology, Lab for Social Machines

families generating more viewer comments. (We ran several analyses, each with a different number of families, to ensure that we didn't overlook any trends). In one analysis, the family that stood out—shown in red in Exhibit 2—follows a rise-and-fall pattern, with the characters achieving early success and happiness before a steady decline into misfortune. Of all the story families, this one has the most negative ending. Although these tales end bleakly, they leave an impact on viewers.

Other analyses of the Vimeo videos revealed similar findings, with two story families attracting significantly more comments than others (Exhibit 3). These stories both culminate with a positive emotional bang, indicated by the large spike near the end of the arcs. The main difference is that stories in the graph on the left involve more mood swings from negative to positive before the big finale. Stories from these two families tend to receive more comments than those that end negatively, perhaps echoing the University of Pennsylvania finding that positive emotions generate the greatest engagement.

Our team read the comments for all the Vimeo shorts, rating the types of emotions expressed, and

ran a program to measure their length. This analysis confirmed that stories in the three families just described tend to generate longer, more passionate responses. Instead of just saying, "great work," a comment might read, "Superb ... so, so powerful ... it hits you like a wrecking ball." What's equally striking is that the comments didn't focus on particular visual images but on a video's overall emotional impact, or how the story changed over time.

These insights will not necessarily send screenwriters back to the drawing board—that would be like asking George Orwell to tack a happy ending onto *1984* to cheer things up. But they could inspire video storytellers to look at their content objectively and make edits to increase engagement. That could mean a new musical score or a different image at crucial moments, as well as tweaks to plot, dialogue, and characters. As storytellers increasingly realize the value of AI, and as these tools become more readily available, we could see a major change in the way video stories are created. In the same way directors can now integrate motion capture in their work, writers and storyboarders might work alongside machines, using AI capabilities to sharpen stories and amplify the emotional pull.

For more information on emotional arcs, see MIT's page on the story learning project or the thesis by Eric Chu on which this research is based.[7,8] Our next article in this series will examine whether a story's emotional arc can predict the speed, breadth, and depth of its discussion on Twitter.

[1] Annalee Newitz, "Movie written by algorithm turns out to be hilarious and intense," *Ars Technica*, June 9, 2016, arstechnica.com.

[2] Nina Metz, "A movie scripted entirely by a computer, but don't freak out just yet," *Chicago Tribune*, June 16, 2016, chicagotribune.com.

[3] "Kurt Vonnegut on the shapes of stories," youtube.com.

[4] John Tierney, "Will you be e-mailing this column? It's awesome," *New York Times*, February 8, 2010, nytimes.com.

[5] Andy Reagan, "The shapes of stories," Computational Story Lab, November 7, 2016, uvm.edu.

[6] We selected the number of families to examine here and in other analyses—it wasn't the machine's choice. There is no "right" number of families, but our heuristics suggest that many fall into between five and ten shapes.

[7] The story learning machine, MIT Media Lab, mit.media.edu

[8] Eric Chu, "A darn good yarn," sosuperic.github.io.

Jonathan Dunn is a partner in McKinsey's New York office, and **Geoffrey Sands** is a director emeritus in the Stamford office. **Eric Chu** is a PhD candidate at the Massachusetts Institute of Technology (MIT) and conducts research at the Lab for Social Machines, part of MIT's Media Lab, where **Deb Roy** is the director and **Russell Stevens** is the industry deployment lead.

Contact for distribution: Jenna Gravino
Phone: +1 727 561 5805
Email: Jenna_Gravino@McKinsey.com

December 2017
Designed by Global Editorial Services