



August 2024

The Yin and Yang of AI By Jamii Corley

Generative AI uses large training sets of artists work to create the neural networks they use to form new pieces of art. Is this inspiration or theft?

If I ask an AI for some Patrick Nagel like paintings and it comes up with something that's difficult to tell from a real Nagel, then I turn around and sell them what am I doing? First, I'm directly copying someone's art and that can be seen as stealing, but worse than that, I'm depriving the original artist of revenue he might make selling his art. And possibly worst of all it dilutes his brand. Now everyone has some of this 'art' and you see it everywhere so it's no longer new, interesting, or bold. AI can do this quickly, almost for free, and without consequences. There are currently no laws that address this kind of harm to the artist.



Image Credits: Patrick Nagel and Midjourney (AI)

On the other hand, if I can't draw or paint, but I have a brilliant idea for a mural and I describe it to my generative AI is the result an inspiration or a fraud? It's a really tricky issue.

I think generative AI can be a tool in the arsenal of artists. It can help an artist flesh out an idea or do some of the repetitive grunt work, but as with all tools it inserts a distance between the artist and their craft. In that repetitive work a slip of the brush can be a serendipitous accident that leads to a new idea for the artwork. So, is AI evil? Will it lead to the downfall of artists? Or is it a bright new tool that can yield new inspiration?

It's a tool, and as Ani DiFranco says, "**Every tool is a weapon, if you hold it right**". We do need new rules about how people's artwork can be used, but we need to make sure we re-

strict bad behavior while allowing for inspiration.

I don't have answers, but I do know we all need to be asking the questions.

The Enron Corpus

By Mark Costlow

When certain large stars (much larger than our sun) run out of nuclear fuel, they don't just wither and die. Without the constant outward pressure of nuclear fusion, the star collapses in on itself. The massive body crumples so quickly and with such intense force that it explodes, flinging material out into the universe. Most of the heavy elements that exist on earth and in our bodies came from these **supernova** explosions.

The implosion of a bright shining star of the business world in the early 2000s had a similar effect. The material flung out of the Enron collapse and explosion may be driving much of the software that mediates our daily lives.

A podcast produced in 2020 by Business Insider called "**You've got Enron mail!**" details the fall of Enron and how it inadvertently influenced AI, voice assistants, counter-terrorism software, and more.

In the late 1990s **Enron** was flying high. An energy trading company headquartered in Houston, TX, it became the 7th largest company in the US. Sadly, they got that way by lying, cheating, and stealing. They used fraudulent accounting practices to inflate their revenues and hide their debts in subsidiary companies. When the fraud was uncovered in late 2001, Enron quickly unraveled and filed for Chapter 11 bankruptcy in December. At \$63B it was the largest bankruptcy ever at the time, a record they held until the chaos of the 2008 financial crisis.

In addition to the financial games, Enron played dirty tricks on the public by **manipulating newly deregulated energy markets** for their profit. One trick was to make people wildly overpay for electricity by creating false scarcity. They did that by lying about maintenance problems in certain power plants, causing them to shut down, so power companies had to import power from other places. Combined with other tricks like buying power in California, moving it to Nevada, and selling it back to California, this caused a series of rolling blackouts while they raked in mountains of cash.

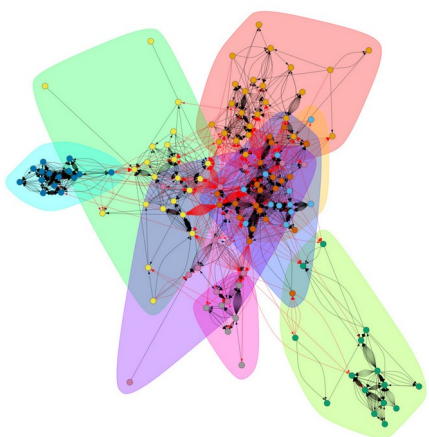
Because of these market manipulations, the Federal Energy Regulatory Commission (FERC) was one of the agencies tasked with investigating Enron. Their investigator Pat Wood had access to all of Enron's files, including all of the company's emails. The emails included everything: meeting notes, project updates, lunch plans, spam, forwarded jokes, harmless flirting, more serious inappropriate relationships. Buried somewhere in there was evidence of Enron's bad deeds, but the sheer size of the data made

it impossible to make sense of. Enron had over 20,000 employees when it failed.

After FERC released its staff report in 2003, the emails and other data collected were deemed to be in the public domain, **available for historical research and academic study**. Since then, the Enron Corpus, as it is known, has been used by many researchers for their own projects, and they have in turn **made copies of it available** in different forms. Anybody who wants it can download the corpus and use it for whatever they like.

You may be horrified at the thought of a company's personal emails being released to the public like this. There was a sanitization process, where the Enron employees were allowed to flag messages to be excluded from the corpus and some obviously private information like social security numbers was removed. Even so, it is full of candid information, clearly written by people who didn't know they were speaking to history. It is very unlikely such a disclosure would happen today, but 2003 was a different time. Gmail, Facebook, Google Earth, and Twitter all came later. The thinking about personal rights to privacy has progressed a lot since then. Pat Wood, who released the corpus, said Enron was widely hated and nobody was very concerned about protecting their privacy rights at the time.

The Enron corpus may be the only large collection of real emails that is freely available for study without non-disclosure agreements or other privacy restrictions. That made it the go-to data source for anyone needing a large sample of human communications. Some areas that have been significantly advanced using the Enron Corpus include **social network analysis, natural language processing, fraud detection**, and machine learning. Early versions of software we use every day, such as Apple's Siri and GMail's "Smart Compose" feature, and other systems that have evolved into today's generative AI systems were trained with the Enron Corpus.



Link Analysis of Enron Emails

This image is a visualization of "network analysis" to identify who was emailing who within Enron. Researchers discovered the

groups committing crimes tended to be insular, talking only among the group members and very few others. These analysis techniques are now used to identify terrorist cells by running the same algorithms on phone call and texting records.

Training Bias

One major concern about using this data for AI training is that it doesn't represent an accurate cross section of our world, or even the US. While it contains an unprecedented level of candor for this kind of training data, the sources are all people who worked for an energy company in Houston. They were not artists, doctors, firefighters, retail workers, social scientists, sailors, cooks, tailors, or teachers. They were lawyers, financial traders, MBAs, CPAs, and HR managers. The corpus also selected mostly senior management, so it is further skewed toward wealthy, white, male individuals.

Amanda Levendowski wrote in 2017, "The Enron emails are simply not representative—not geographically, not socioeconomically, not even in terms of race or gender. Indeed, researchers have used the Enron emails specifically to analyze gender bias and power dynamics. And yet the Enron emails remain a go-to dataset for training AI systems."

Enron's Legacy

In the wake of Enron's demise, congress passed the Sarbanes-Oxley Act which created sweeping auditing and financial regulations for public companies, a very visible and tangible legacy. However, the influences of the Enron Corpus, although hidden and mostly unknown, might leave a more durable and long-lasting mark on our society.

There is a natural comparison here to the tumor **cells of Henrietta Lacks**. They were taken without her knowledge or consent in 1951 and then used in thousands of biomedical research projects over the next 7 decades. The "HeLa cells" have been cited in over 110,000 studies, but the public was unaware of their existence and influence for decades. You have almost certainly received a medical treatment sometime in your life which was **influenced by HeLa cells**.

Similarly, the Enron Corpus was released into the world at just the right time. It unwittingly became a fundamental part of the computer systems which now underpin our everyday lives, including all the specific world-views and biases it contains.

Southwest Cyberport - swcp.com

New Mexico's Expert Internet Service

Provider since 1994

505-232-7992 | Support: help@swcp.com

5021 Indian School NE, Ste. 600,

Albuquerque, NM 87110

Click on bold blue type in browser for links.