

Navigating AI in ESG & Risk Management

Unveiling the Mechanics



Table of Contents

Introduction: Harnessing AI’s Potential in ESG & Risk Management	<u>2</u>
1. How Does AI Fit with ESG & Risk Mitigation?	<u>3</u>
1.1 Step 1: Data Collection	<u>4</u>
1.2 Step 2: Organizing & Cleaning the Data	<u>5</u>
1.3 Step 3: Connecting the Dots	<u>6</u>
1.4 Step 4: Adding Contextual Understanding	<u>7</u>
1.5 Step 5: Analyzing with Algorithms	<u>8</u>
1.6 Step 6: Turning Analysis into Actionable Insights	<u>9</u>
2. Using NLP for ESG & Risk Management: Practical Examples	<u>11</u>
2.1 FTX & Sam Bankman-Fried	<u>11</u>
2.2 Signature Bank	<u>13</u>
3. The Future of ESG & Risk Management	<u>15</u>



“

The future of AI in ESG and risk management is not just a matter of technological advancement but a narrative of how we evolve with it.

”

Introduction

Harnessing AI's Potential in ESG & Risk Management

With all the buzz around Generative AI, it's easy to forget that artificial intelligence (AI) has been driving innovation across industries for years. Environmental, Social, and Governance (ESG) and risk management are no different. However, at the rate AI is advancing and as the amount of raw data available for analysis continues to expand, the need to understand AI is more pressing than ever.

AI is transforming ESG, turning complex data into predictive insights and reshaping our approach to risk. But, what does this mean for the industry, and how can professionals leverage this technology to maintain a competitive edge? **The future of AI in ESG and risk**

management is not just a matter of technological advancement but a narrative of how we evolve with it.

This ebook dives into how AI works when applied to ESG, shares a few practical examples of what it looks like in real life, and offers a few predictions for what comes next.

1. How Does AI Fit with ESG & Risk Mitigation?

Key facts about NLP:

Our focus

At SESAMm, our focus is on harnessing natural language processing (NLP) to pinpoint ESG controversies.

How NLP works

NLP uses sophisticated algorithms to digest and comprehend vast quantities of text from diverse sources, such as:

- news articles,
- blogs,
- forums.

Our data lake

analyses

10 million

new articles each day

The advantage of NLP

lies in its remarkable speed of processing data.

Artificial intelligence spans a broad spectrum of technologies, from the first Machine Learning models to Large Language Models, and beyond. At SESAMm, our focus is on harnessing natural language processing (NLP) to pinpoint ESG controversies. NLP uses sophisticated algorithms to digest and comprehend vast quantities of text from diverse sources, such as news articles, blogs, and forums. The advantage of NLP lies in its remarkable speed of processing data. Every day, ten million new articles are added to our data lake. These are processed, categorized by relevant entity, and analyzed for controversy type and severity. It's not an exaggeration to say that NLP works faster than humanly possible. If a human could read an article every second, it would take him or her 115 days to read just 24 hours of SESAMm's data.

For a better understanding of how NLP works, we've broken it down into six high-level steps.

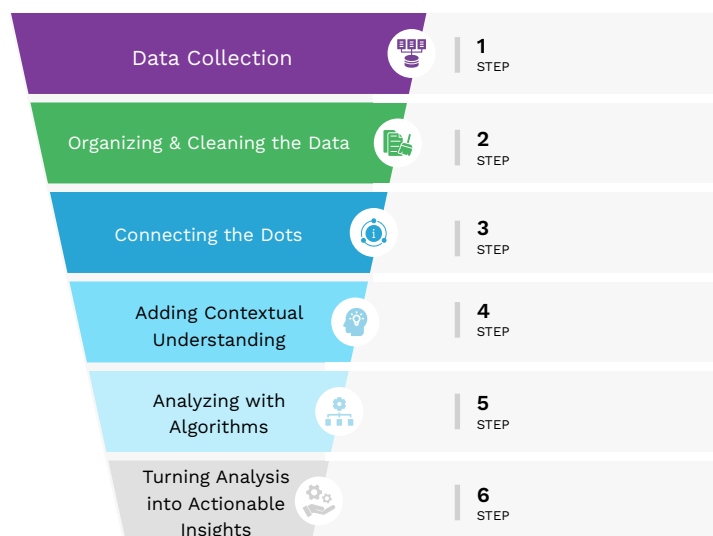
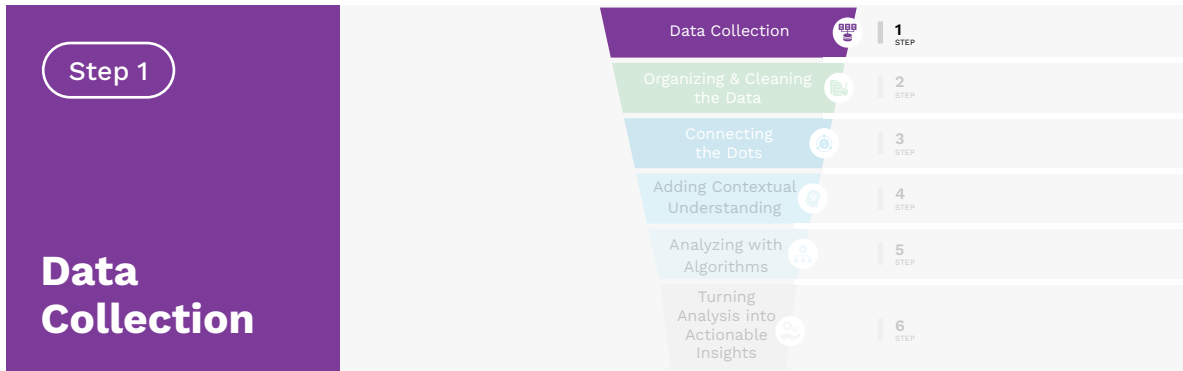


Figure 1: How NLP works in six steps



Data Collection

SESAMm's data repository

SESAMm's data repository dates back to 2008.

Encompassing

15 years

of data from:

- news outlets,
- blogs,
- consumer reviews,
- discussion forums,
- whistleblower websites.

Made up of over

25 billion

articles

from

4 million

sources

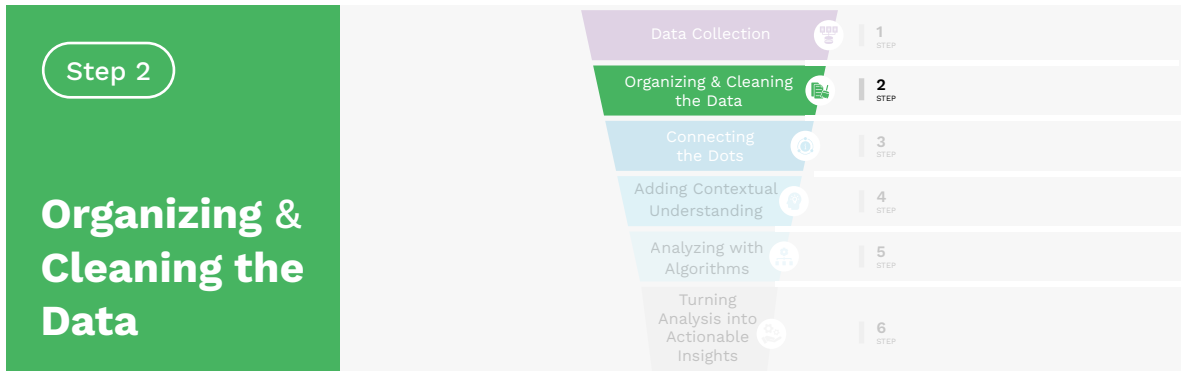
in

100

languages

To get started, the system gathers web-based information to form our data lake. It's essentially a large storage repository that holds raw data in its native format ready to be used. A data lake is important because algorithms need large datasets to train on to ensure their accuracy. It also enables real-time data ingestion, which in turn, makes near real-time insights and monitoring possible. Finally, the data lake preserves information, safeguarding it from being later deleted or edited on the web. Maintaining historical context is important for an accurate trend analysis.

SESAMm's data repository dates back to 2008, encompassing 15 years of data from professional news outlets, blogs, consumer reviews, discussion forums, and whistleblower websites. Currently, it is made up of over 25 billion articles from 4 million sources in 100 languages. This extensive and robust data lake, featuring hyperlocal sources, enables SESAMm to cover millions of public and private companies comprehensively.



Step 2

Organizing & Cleaning the Data

Next steps

The data must be:

- o standardized,
- o categorized,

so it can be

- o tagged,
- o indexed.

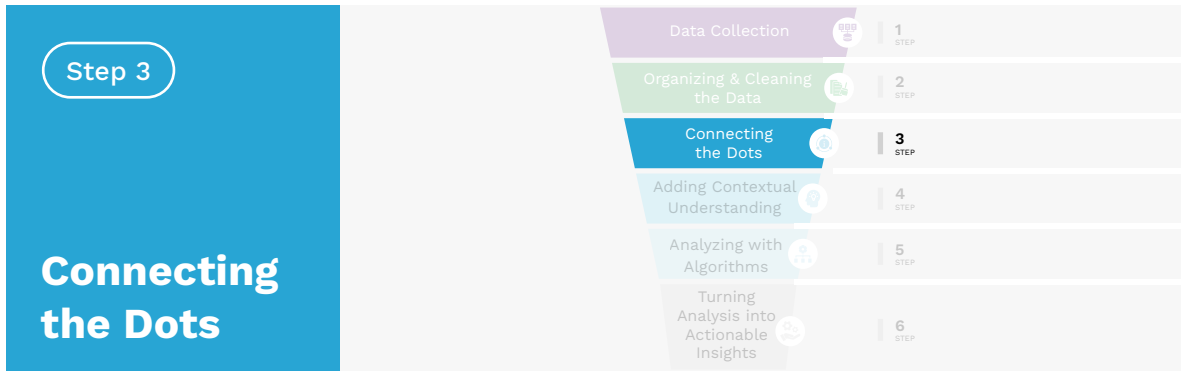
In the context of SESAMm’s data, **this process involves:**

- o tagging the type of document (for example news, blogs, etc),
- o the language,
- o the country,
- o scoring the authority of the source,
- o the quality of the text.

Next, the data must be standardized and categorized, so it can be tagged and indexed. This process allows for a better, faster analysis and filters out irrelevant information. Essentially, the data is like a huge, mixed-up pile of LEGO bricks. You can’t tell what you can build with them because they’re all jumbled together—different colors, sizes, and shapes, all in one big mess. Before you can start building, you need to sort them out.

Cleaning and organizing the data is like sorting those LEGO bricks by color and size, and then putting them into separate bins. Tagging and indexing is like labeling each bin so you know exactly where to find the piece you need. When you’re ready to build something, you can quickly grab the right pieces without wasting time sifting through bricks you don’t need. Just like with LEGOs, once the data is sorted and labeled, you can more easily see what you have and start constructing your analysis.

In the context of SESAMm’s data, this process involves tagging the type of document (for example news, blogs, etc), the language, and the country, in addition to scoring the authority of the source and the quality of the text.



Knowledge Graph (KG)

The KG is a big, intricate **network** that **outlines** the **connections** between:

- key entities,
- topics,
- themes.

At SESAMm, our **KG** is **specifically tailored** to:

- find,
- extract,
- analyze,

data about:

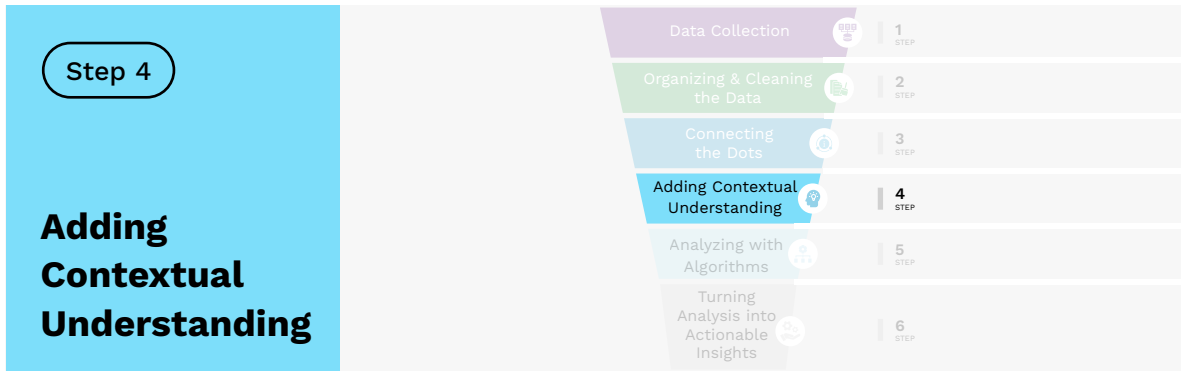
- public companies,
- private companies.

SESAM KG
is updated
weekly

Now that the data is organized, it is used to create a Knowledge Graph (KG). The KG is a big, intricate network that outlines the connections between key entities, topics, and themes. It helps the system understand how companies, products, and brands are related to each other. For example, Apple is mapped to the iPhone, Steve Jobs, and Tim Cook. To stay accurate, the KG will have to be updated regularly.

Although the basic concept of a KG is consistent between AI companies, the execution, focus, and application can vary greatly. For example, the data sources and scope, the domain focus, the entity recognition and the update frequency can all vary depending on the specific needs, industries, or use cases.

At SESAMm, our KG is specifically tailored to find, extract, and analyze data about public and private companies. Three years of development went into structuring and curating our large KG based on a combination of paid sources and web-extracted information to provide extensive coverage of public and private companies. To stay current with corporate changes and new emerging entities, the SESAMm KG is updated weekly.



Step 4

Adding Contextual Understanding

Next challenges

The next challenge is to understand the text itself which requires a variety of techniques.

- Named Entity Recognition (NER),
- lemmatization,
- word embedding.

Named Entity Recognition (NER) identifies and classifies key elements in text into predefined categories.

Lemmatization and **word embedding** are used to analyze the context.

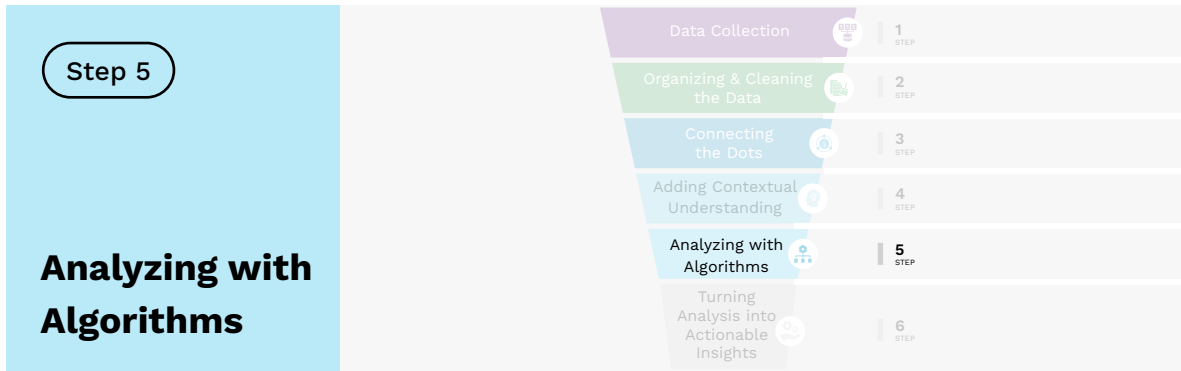
All together, these techniques help ensure a **robust contextual understanding**, which in turn enables:

- enhanced precision in data analysis,
- improved relevance and accuracy of results,
- comprehensive coverage.

With the Knowledge Graph, we can navigate the complex web of entity relationships. The next challenge is to understand the text itself, which requires a variety of techniques. Named Entity Recognition (NER), for example, identifies and classifies key elements in text into predefined categories. Essentially, it looks for words or phrases that represent specific pieces of information, making it easier to understand the context or the main points of a text. For example, in the sentence “Apple Inc. was founded by Steve Jobs, Steve Wozniak, and Ronald Wayne,” NER would identify “Apple Inc.” as an organization, “Steve Jobs,” “Steve Wozniak,” and “Ronald Wayne” as people.

Other technologies, such as lemmatization and word embedding, are used to analyze the context. For example, they are used to understand the various topics discussed in the data: this article about Apple Inc. is about human rights issues, while this other article is about greenwashing. At SESAMm, we also use these technologies to regroup articles on the same topic into a cluster. Clustering articles helps avoid information overload and provides with users better visibility into the evolution of a controversy.

All together, these techniques help ensure a robust contextual understanding, which in turn enables enhanced precision in data analysis, improved relevance and accuracy of results, and comprehensive coverage. Ultimately, it ensures that the insights derived are based on a comprehensive and precise understanding of the data.



Next step

Using our ESG framework **our algorithms scrutinize the data** for essential **insights** such as:

- sentiment analysis,
- ESG controversies,
- Sustainable Development Goals (SDG) impacts.

The goal is to pick out all of the articles that relate to

- a company,
- a specific ESG risk.

Monitoring is also a **key component** to ensure quality, which requires:

- continuous evaluation,
- adaptation to new data,
- feedback mechanisms,
- human oversight.

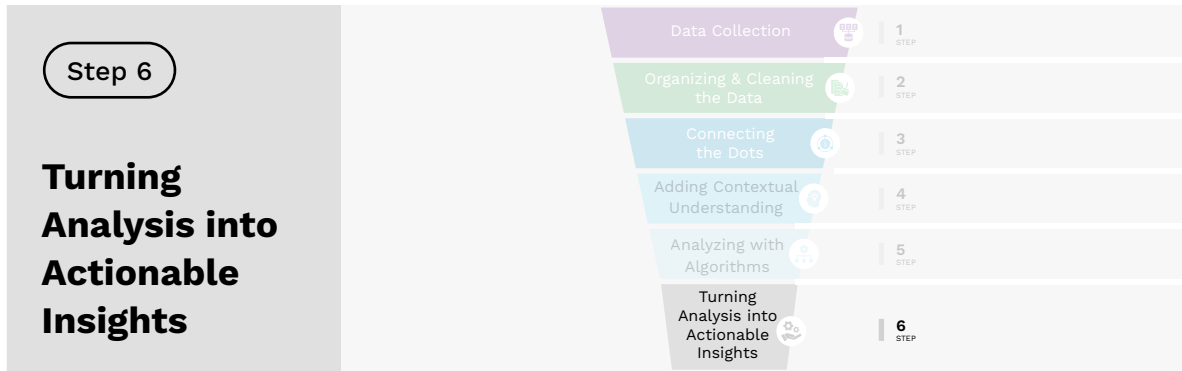
30% saved time

after we added large language models into our tech stack to automate data annotation for ESG and SDG alerts.

Once the data is well-organized and contextualized, a deeper analysis phase kicks off. Using our ESG framework, our algorithms scrutinize the data for essential insights such as sentiment analysis, ESG controversies, and Sustainable Development Goals (SDG) impacts. The goal is to pick out all of the articles that relate to a company and a specific ESG risk. These algorithms are being refined constantly to improve accuracy and performance. Monitoring is also a key component to ensure quality, which requires continuous evaluation, adaptation to new data, feedback mechanisms, and human oversight.

At SESAMm, for example, our experts conduct regular annotation campaigns to improve the performance of our algorithms. We have also added large language models into our tech stack to automate data annotation for ESG and SDG alerts, this alone has saved our analysts 30% of their time and ensures our algorithms will continue to perform well.

Once the algorithms have been run and the analysis extracted, it's time to combine them into actionable insights.



Next step

One of AI's **biggest advantages** is the **speed** at which it can:

- process data,
- provide an accurate analysis.

However, to really take advantage of the increased speed, **the analysis needs to come in a fast, easy-to-understand format.**

One such metric is **SESAMm's Controversy Intensity Score.**

Ranging intensity to

1-5 (low to high)

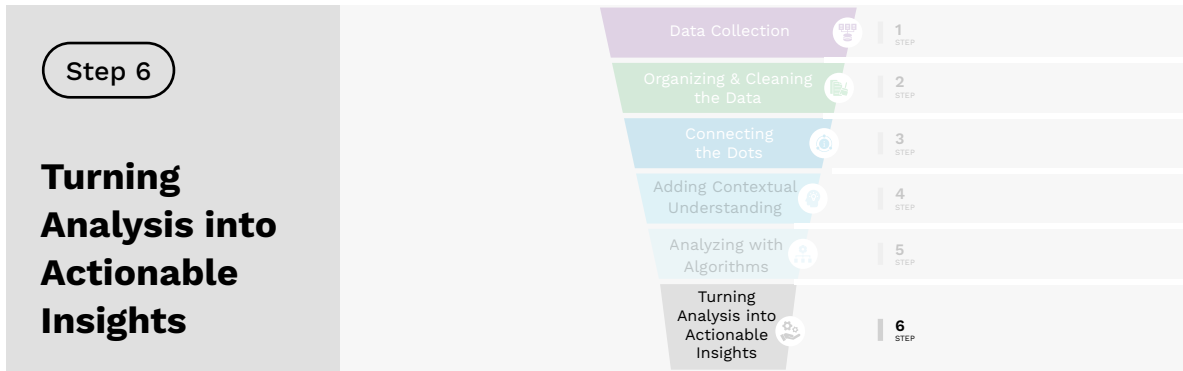
it's a fast, intuitive way:

- to understand the potential severity of a controversy,
- to triage which events require your attention first.

One of AI's biggest advantages is the speed at which it can process data and provide an accurate analysis. However, to really take advantage of the increased speed, the analysis needs to come in a fast, easy-to-understand format. Take ChatGPT, for example. It was a revolution in LLMs not because of the AI technology itself (or at least, not only because of the technology), but rather because it has a simple, easy-to-leverage user interface, putting its powerful analytic capabilities at the general public's fingertips.

For ESG and risk management, providing a competitive advantage to our customers requires delivering accurate insights in near real-time using clear metrics, empowering them to make informed decisions and take action quickly. One such metric is SESAMm's Controversy Intensity Score. Ranging from 1-5 (from low to high intensity), it's a fast, intuitive way to understand the potential severity of a controversy and to triage which events require your attention first.

When a user logs into their SESAMm dashboard, they quickly see which companies in their portfolio have the most significant controversies as well as the top 30 most significant controversies for their entire portfolio. From there, they can dive into the details of each company and controversy, including reading the articles behind the event.



When a user **logs into** their **SESAMm dashboard**, they quickly see:

- which companies in their portfolio have the most significant controversies,
- the top 30 most significant controversies for their entire portfolio.



Figure 2: SESAMm’s dashboard

That summarizes the basic NLP process, from collecting the data to turning it into actionable insights. But rather than stay in the theoretical, we’ve pulled together some real-world examples of how NLP can be used to manage ESG controversies and risk.

Practical Examples

Using NLP for ESG & Risk Management

For a practical look at what NLP can do, we gathered a couple of examples of how TextReveal, the AI platform behind our ESG Alerts solution, can be used for ESG controversy and risk management.

2.1 FTX & Sam Bankman-Fried

The FTX collapse and subsequent bankruptcy filing shook the crypto and financial communities. FTX, a \$32 billion cryptocurrency exchange company, filed for bankruptcy on November 11, 2022, after a series of events that raised questions about its financial stability and management practices.

The issues began with a CoinDesk report revealing that Alameda Research, a hedge fund owned by FTX's founder Sam Bankman-Fried (SBF), held significant amounts of FTT tokens, the native token of FTX. This suggested a risky overlap between the two entities he owned.

“

This stunning series of events took the world by surprise. Could TextReveal have predicted the trouble?

”

Following the report, Binance's CEO, Changpeng Zhao, announced the sale of Binance's FTT holdings, leading to a market panic and a sharp decline in FTT's price. FTX faced a liquidity crisis as customers rushed to withdraw funds, totaling about \$6 billion over three days.

This stunning series of events took the world by surprise. Could TextReveal have predicted the trouble?

Looking at the web data, there are several red flags before the collapse. Most importantly, the sentiment was positive until early 2022 but turned

negative with increased negative mentions related to withdrawals and other controversies surrounding SBF. In the graph below, we see the global sentiment decreasing, while the volume of mentions begins to increase.

Definition: Polarity represents the aggregate of positive and negative sentiments (opinions or reviews) on a company. A 0 score means there is as much positive as negative sentiment expressed. The dotted and dashed lines represent sentiment in the following charts.

Pre-Bankruptcy Mentions & Sentiment

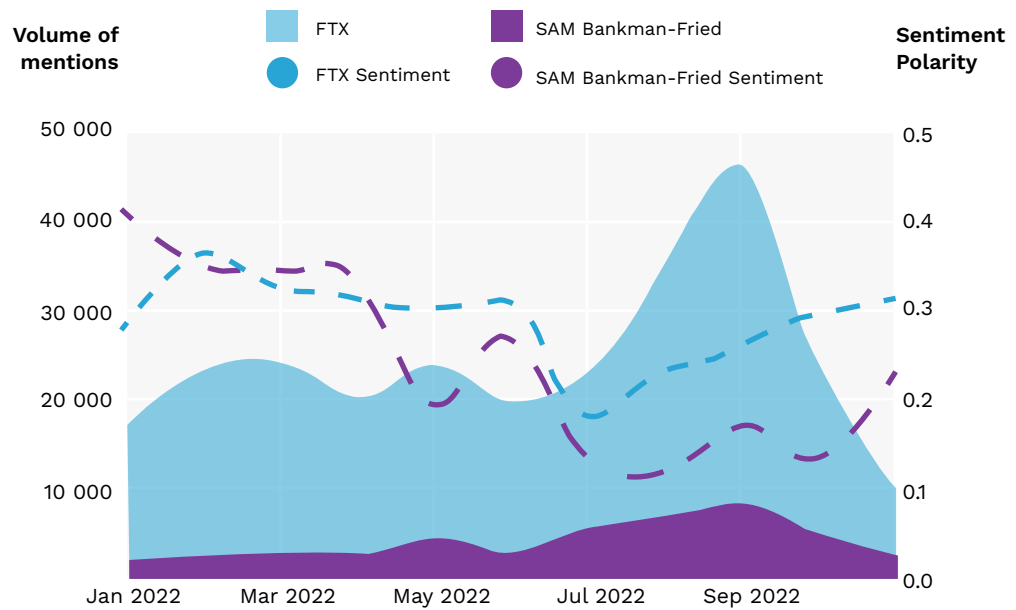


Figure 3: Pre-Bankruptcy Mentions & Sentiment

A deeper look into the ESG risks uncovered by TextReveal suggests several concerning trends for social and governance risk as well, all well before the story broke in November. Put all together, the TextReveal data provides several warning signs prior to the FTX collapse. [Check out the full analysis of TextReveal on FTX here.](#)

2.2 Signature Bank

Signature Bank operated as a commercial bank approved by the FDIC and licensed by the state of New York, mainly serving privately held businesses. While it served mid-sized businesses, the bank was particularly recognized for its services to law firms, property purchasers, and businesses dealing in cryptocurrencies.

The trouble for Signature Bank started with the collapse of Silicon Valley Bank (SVB), which triggered a rush of withdrawals from Signature Bank on

“

Looking at the data available from TextReveal, we see that SESAMm uncovered significant controversies surrounding Signature Bank as early as the beginning of February, more than a month before its forced shutdown by federal regulators on March 12th.

”

Friday, March 10, 2023. The panic among depositors that followed SVB’s failure was compounded by the fact that Signature Bank held high amounts of uninsured deposits and was exposed to the crypto sector ([source](#)).

Looking at the data available from TextReveal, we see that SESAMm uncovered significant controversies surrounding Signature Bank

as early as the beginning of February, more than a month before its forced shutdown by federal regulators on March 12th. SESAMm’s clients were able to proactively exclude it from their investment portfolios and supplier lists, potentially saving them from significant financial losses.

Count of controversies by intensity over time

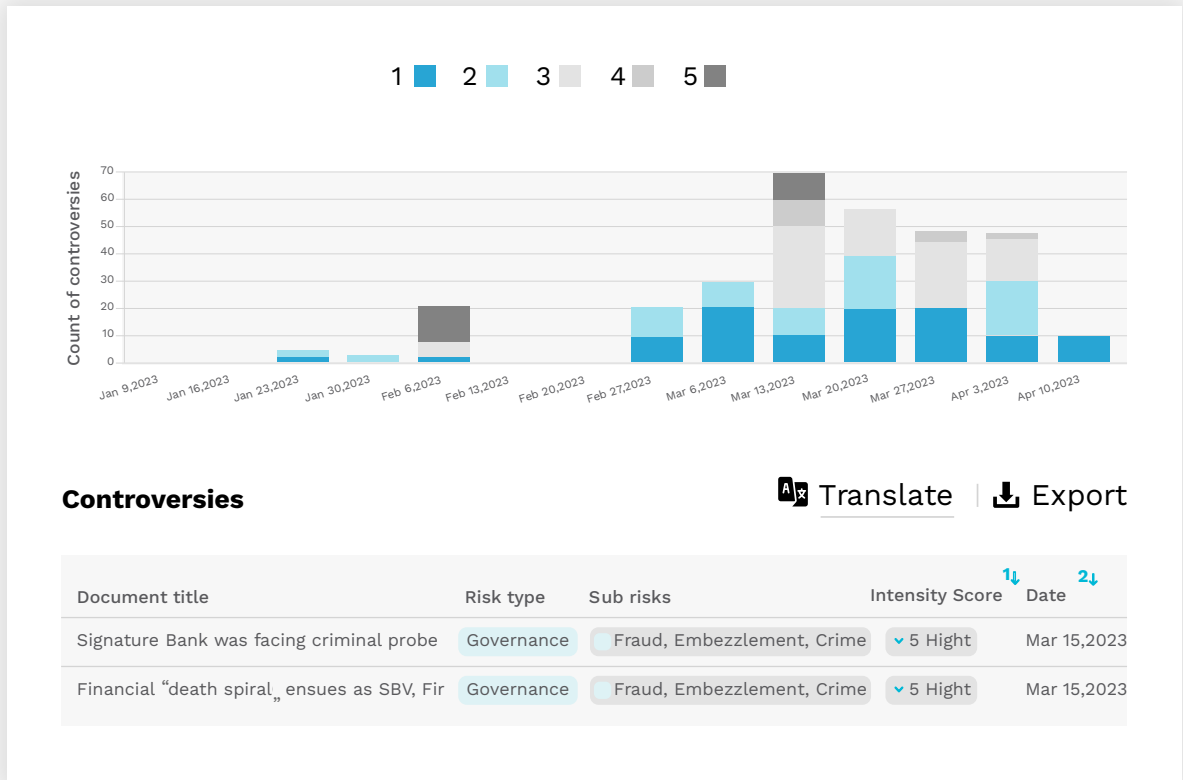


Figure 4: Count of controversies by intensity over time

The Future of ESG & Risk Management

Looking ahead, it's clear that AI is not just a nice-to-have tool but a critical capability in effectively managing ESG and risk. There's an overwhelming amount of data available on the web today and the journey from raw data to actionable insights is complicated- but AI offers a new level of clarity and precision at speeds previously unimaginable.

Discover how SESAMm's AI technology can help your company better manage ESG and risk.

[Request a demo here](#)