Received: 24 August 2023 | Revised: 23 April 2024 | Accepted: 13 May 2024

DOI: 10.1002/aaai.12181

SPECIAL TOPIC ARTICLE

A new era of AI-assisted journalism at Bloomberg

Claudia Quinonez | **Edgar Meij**

Bloomberg, London, UK

Correspondence
Claudia Quinonez and Edgar Meij,
Bloomberg, London, UK.
Email: cquinonez1@bloomberg.net;
emeij@bloomberg.net

Abstract

Artificial intelligence (AI) is impacting and has the potential to upend entire business models and structures. The adoption of such new technologies to support newsgathering processes is established practice for newsrooms. For AI specifically, we are seeing a new era of AI-assisted journalism emerge with trust in the AI-driven analyses and accuracy of results as core tenets.

In Part I of this position paper, we discuss the contributions of six recently published research papers co-authored by Bloomberg's Artificial Intelligence Engineering team that show the intricacies of training AI models for reliable newsgathering processes. The papers investigate (a) the creation of models for updated headline generation, showing that headline generation models benefit from access to the past state of the article, (b) sequentially controlled text generation, which is a novel task and we show that in general, more structured awareness results in higher control accuracy and grammatical coherence, (c) chart summarization, which looks into identifying the key message and generating sentences that describe salient information in the multimodal documents, (d) a semistructured natural language inference task to develop a framework for data augmentation for tabular inference, (e) the introduction of a human-annotated dataset (ENTSUM) for controllable summarization with a focus on named entities as the aspect to control, and (f) a novel defense mechanism against adversarial attacks (ATINTER). We also examine Bloomberg's research work, building its own internal, not-for-commercial-use large language model, BloombergGPT, and training it with the goal of demonstrating support for a wide range of tasks within the financial industry.

In Part II, we analyze the evolution of automation tasks in the Bloomberg newsroom that led to the creation of Bloomberg's News Innovation Lab. Technology-assisted content creation has been a reality at Bloomberg News for nearly a decade and has evolved from rules-based headline generation from structured files to the constant exploration of potential ways to assist story creation and storytelling in the financial domain. The Lab now oversees the operation of hundreds of software bots that create semi- and fully automated stories

of financial relevance, providing journalists with depth in terms of data and analysis, speed in terms of reacting to breaking news, and transparency to corners of the financial world where data investigation is a gigantic undertaking. The Lab recently introduced new tools that provide journalists with the ability to explore automation on demand while it continues to experiment with ways to assist story production.

In Part III, we conceptually discuss the transformative impact that generative AI can have in any newsroom, along with considerations about the technology's shortcomings in its current state of development. As with any revolutionary new technology, as well as with exciting research opportunities, part of the challenge is balancing any potential positive and negative impacts on society. We offer our principles and guidelines used to inform our approach to experimenting with the new generative AI technologies. Bloomberg News' style guide reminds us that our "journalism is aimed at possibly the most sophisticated audience in the world, for whom accuracy is essential."

INTRODUCTION

A key value in any newsroom is the accuracy of its reporting and writing. This is particularly true at Bloomberg News, which has as guiding principle: "It isn't news if it isn't true." For financial market readers, the stakes are always high and billions of dollars can change hands in reaction to an unexpected headline or news story. "The Bloomberg Way" (Micklethwait et al. 2017) states that "the key to our success has been an unyielding commitment to getting the facts right, no matter how big or small the story." So in order to continue serving our readers, accuracy and trust need to remain at the core of any technology we develop and adopt.

Training AI models to be incorporated into news workflows or tasks while meeting requirements for accuracy is very challenging as it requires a very high precision level. At the same time, incorporating AI technologies is appealing given the number and types of tasks with which AI models can help. AI-assisted journalism can save time, augment expert knowledge, allow scalability, increase precision, and shorten the time it takes for news to be published. Developing such high-precision models requires iterative stages of research and implementation (Capponi and Lehalle 2023). It requires breaking the newsgathering workflow into narrow investigation opportunities that are identified in collaboration between computer scientists, news product managers, and editorial domain experts. Recently, published research papers by Bloomberg's AI researchers hint at the granularity of the contributions needed to advance the understanding and application of this technology for news. Tasks that are relatively easy to human journalists, such as updating a headline or explaining a chart, require narrow training of an AI with multiple iterations, as well as additional research into how to deploy them in an automated fashion. Yet, however impressive the recent developments in the field of AI may be, these models do not yet display a human journalist's capacity for planning, reasoning, ensuring fair and balanced reporting, detecting from small amounts of evidence the trails of a bigger story, or asking tough questions in an interview in order to bring to light an unknown, but relevant, piece of information.

Where AI models are highly efficient is in augmenting the journalistic capacity to report, crunching numbers at scale and predicting upcoming trends, monitoring and extracting information from structured and unstructured data, and alerting when certain market conditions are about to align. The technology provides multiple forms of useful assistance and creates new and exciting opportunities for tech-savvy journalists to augment news tools and technologies, or to train and even develop new technologies.

The confluence of news and AI is a vast space, full of opportunities however still in its infancy. The main contributions of this paper are (i) to introduce how we incorporate AI into existing Bloomberg newsroom workflows in this era of AI-assisted journalism and (ii) to showcase recent research performed by our computer scientists in collaboration with product managers and editorial domain experts.

To that end, we examine some of the recent contributions by Bloomberg AI researchers to advance AI for news applications in Part I. We note these papers are early-stage

research and typically only focus on a relatively small part of a potential, future newsroom enhancement.

After that, in Parts II and III, we dive into how we make use of such research and couple it with domain expertise from in-house research analysts, journalists, and finance specialists in order to ideate and develop novel AI-driven news stories that effectively combine insights from all of these disciplines.

Part I—Contributions to advancing the field of AI and news

Updating headlines

Updating published headlines is a task performed several times a day in any real-time news organization. As stories unfold, the reader must be updated with new facts and relevant developments, and, as such, it is a dynamic setting. In an effort to establish what would be required for a system to recognize the need to update a headline and perform the task, we conducted research looking to determine the model training requirements. AI models could potentially be able to perform the task by recasting the workflow as a type of contextual summarization task. The implication is that we would require a system that is capable of identifying the novel information in the updated article and then to modify the existing headline accordingly. We note that if news developments require a complete rewrite of the headline, that normally implies publishing a fresh version of the story with a new lead and headline.

The challenge is to generate a new headline that must simultaneously convey critical new information while providing a holistic overview for readers unfamiliar with the story. Bloomberg researchers envisioned a summarization system that combines an existing summary with information updates, aimed at producing a new summary based on what the reader already knows and the content that has changed (Panthaplackel, Benton, and Dredze 2022). In the paper, we contributed a model for contextual, dynamic summarization that includes the release of the Headline Revision for Evolving News (HREN) dataset, an evaluation of different types of information such as previous headlines and edits to the article body, in addition to conducting human evaluation (Figure 1).

The research highlights the importance of making use of the latest version of the article body for the updated headline generation task. Faithfulness to the article body is essential, and it is normally expected that an updated headline copies part of the original headline and overall structure as it follows a natural progression of events. Overall, many headlines can be correctly updated through a simple token replacement reflecting an analogous change in the body of the article. Hence, retaining the original headline helps maintain important details from the story. When there are more substantial changes to the article, headlines are likely to require extensive edits as well.

Our model evaluates context representations, including both Pointer Networks and a pre-trained transformer network (BART), and uses beam search to decode for all models along with bigram blocking. We conclude that the decoding hyperparameters were found to work well across models. We also note that the proposed model (H1 + B_{edits})[1] can generate headline predictions that are statistically tied with gold headlines in terms of factuality, while making fewer unnecessary edits.

Multiple versions of a story

News organizations normally need to create multiple versions of the same story to fit different products, from their websites to email newsletters, or even to create a set of related news stories (Figure 2).

Our researchers identified that content planning for text generation using macrostructural controls has been relatively underexplored in this space (Spangher et al. 2022). We propose a novel controlled text generation task, showing that more structural awareness of the discourse results in higher control in terms of accuracy, grammaticality, coherency, and topicality, approaching human-level writing performance. We hypothesize that information about discourse structure can positively impact generative outlooks and that more structural awareness improves generation. The discourse structure of a text—its macrostructure—impacts both human and machine comprehension, and we show that structural awareness, especially of past structure, helps generate the highest quality text.

We combine two different approaches in controlled text generation—generation and editing—to demonstrate that we can effectively do generation by combining a naively trained language model with a discriminator. In this case, the discriminator was constructed as the main architectural component, allowing for the incorporation of interdependencies between control code sequences. The architecture combines a sentence classification model with a separate label embedding architecture to incorporate knowledge of prior tags. The contribution opens the door to a variety of follow-up directions, such as giving users control over the macro structure or, working in tandem with a generative algorithm, to fill-in missing structural components in a piece of writing or even create different versions of the same story.

B_1	H_1	B_2	H_2
Nearly a million people in southern Vietnam face **evacuation from the path of a deadly tropical storm**...	Tembin: **Vietnam braces for killer storm**	A tropical storm that **was threatening southern Vietnam has weakened and is expected to dissipate**...	Tembin: **Storm weakens as it nears southern Vietnam**
A no-confidence motion in Wales' health minister over Cwm Taf's maternity service failings **has been debated by AMs and will be voted on later**....	Cwm Taf: Health minister **facing** no-confidence vote	Wales' health minister **has survived** a Plaid Cymru no-confidence motion in him after severe failings were uncovered at Cwm Taf's maternity services...	Cwm Taf: Health minister **survives** no-confidence vote
US astronauts Doug Hurley and Bob Behnken **will dock** to the International Space Station (ISS) **in the next hour**...	SpaceX Nasa Mission: Astronaut capsule **closes in on** space station	US astronauts Doug Hurley and Bob Behnken **have docked** with the International Space Station (ISS)...	SpaceX Nasa Mission: Astronaut capsule **docks with** space station
At least **19** people were injured in a crash involving a charter bus and a tractor-trailer on a Virginia interstate on Sunday morning, the authorities said...	At Least **19** Hurt in Tractor-Trailer and Bus Crash on I-64 in Virginia	At least **24** people were injured in a crash involving a charter bus and a tractor-trailer on a Virginia interstate on Sunday morning, the authorities said...	At Least **24** Hurt in Tractor-Trailer and Bus Crash on I-64 in Virginia
...special counsel Robert Mueller, the man charged with investigating Russian interference in the US election and possible collusion with Trump's campaign, **with one friend of the president floating the possibility he could fire Mueller.**	Trump **may sack** special counsel in Russia inquiry, **says friend**	**Rod Rosenstein**, the deputy attorney general, has **hit back following speculation** that Donald Trump was considering firing the special counsel Robert Mueller, assuring senators he was aware of **"no secret plan"** to oust the former FBI director...	**Rod Rosenstein: 'no secret plan' to fire** special counsel in Trump-Russia inquiry

FIGURE 1 Examples of evolving news stories from (Panthaplackel, Benton, and Dredze 2022). The system receives an existing article (B_1, H_1) and an updated version of the article body (B_2). The goal is to update H_1 to produce a new headline (H_2) that reflects important new information in B_2. Important changes between B_1 and B_2, and H_1 and H_2 are in boldface.

Headline	UPDATE 1-Belarus food retailer lines up landmark overseas IPO
Main Event	SILVER SPRING, Colo. (AP) — A company in Finland is poised to list its first international grocery chain in the United States after a yearlong search, raising concerns about potential corporate conflicts of interest and legal issues around using foreign intellectual property.<|endoftext|>
Previous Event	The world's largest retailer of meat and fish products hopes to become one day the leader in convenience groceries.<|endoftext|>
Expectation	It says it could become a leader in fresh fruits and vegetables, as well.<|endoftext|>
Evaluation	About 300 Finnish companies have applied to list on the New York Stock Exchange under a new plan to bypass the U. S. government and create "licenses" that allow them in many other countries, according to a regulatory filing.<|endoftext|>
Evaluation	The European Union prohibits using foreign intellectual property for purposes outside the country, and Finland's National Stock Exchange does not require approval before selling a company'lls share in its new company.<|endoftext|>
Evaluation	"What's at stake here is transparency, a sense of fairness to all the stakeholders.<|endoftext|>
Evaluation	I don't think it's right for companies to have intellectual property rights," says Michael Vakilainen, an independent analyst who has tracked the company since 2008.<|endoftext|>
Expectation	He says there are potential conflicts of interest, because one partner is the government.<|endoftext|>
Expectation	"What if you're a government contractor?"<|endoftext|>

FIGURE 2 Example of an automatically created document from (Spangher et al. 2022).

Illustrating stories using charts

Adding charts to stories has become indispensable in the information era as they illustrate the narrative and help to convey relevant information in an easy-to-absorb way. Charts are used for visually summarizing important information that a document intends to shed light on. At times, they contain information that is not necessarily present

in the body of the article, hence conveying messages in a multimodal setting.

We hypothesized that the primary information in a chart is an important step towards understanding a multimodal document (Tan et al. 2022). Potentially relevant applications include the indexing of information content for search engines, making charts accessible for individuals with vision impairment, or simplifying information dissemination. At the same time, the field of automatically generating summaries for "structured images," such as charts, remains underexplored.

Different from capturing a natural image, the summarization of a chart relies heavily on the understanding of the text inside images, which requires building a pipeline to extract the text info from the images. The chart summarization task aims to generate a salient summary of the visual content plus any text with significant information of the multiple components in the chart. To achieve this, a basic captioning model is used, along with an optical character recognition (OCR) encoder and a pretrained language encoder. This approach encodes the position of the OCR along with the text representation since the spatial information generally indicates properties of the text. In order to connect the visual and textual information, researchers adopt a dual approach: appending pre-embeddings and adding cross-attention layers. The pre-embeddings concatenate the sequence of visual vectors before the word embeddings, while the cross-attention layers fuse visual information.

With this design, the generation is led by the image content and uses the OCR information to generate concrete words. We empirically find this to be the best combination to fuse information into the language decoder. We also find that the proposed model significantly outperforms the baseline models for the tested datasets and identifies future research into the adoption of a visual encoder specifically designed for understanding curved lines in charts, opening up the way to more advanced ways of having machines understand content in documents, to be used in downstream journalistic settings.

Making use of tables

Similar to charts, tables are increasingly prevalent, as they provide readers with valuable granular data points that are useful in decision-making processes. One of our recent table-related research contributions (Kumar et al. 2022) extends the application of Natural Language Inference (NLI) from sentences to tabular data. It focuses on a tabular inference task that develops a semiautomated framework for data augmentation, contributing to the availability of datasets. Producing large-scale datasets at a reasonable cost while retaining data quality is critical; however, computationally efficient autonomous generation has many limitations, including an inability to generate linguistically complex sentences and the difficulty of incorporating reasoning into the dataset.

The proposed framework exploits patterns in tabular structures to generate hypothesis templates that are transferable to similar tables that share common attributes. By focusing on entity tables, the framework benefits from the existing tabular structure and straightforward knowledge patterns. This approach to hypothesis generation utilizes the overlapping key patterns to create counterfactual tables. It also highlights that the complexity and diversity of the templates can be enforced via human annotators enabling the creation of linguistically and lexically diverse datasets.

The research applies the above framework to the entity-centric INFOTABS dataset and creates a large-scale, human-like synthetic dataset called AUTO-TNLI, which contains counterfactual entity-based tables. The intensive experiments with AUTO-TNLI demonstrate that the framework helps benchmark and data augmentation, especially in a data supervision setting.

The paper highlights that—in the context of this dataset—the semiautomatic architecture simplifies the creation of larger-scale datasets. It also allows the framework to be reused with additional tabular data as long as the table's structure is preserved. Moreover, this framework can also be used to minimize hypothesis bias by establishing an adequate number of diverse templates for all aspects of each category on a given table, again allowing machines to more deeply understand commonly encountered artefacts in financially relevant documents.

Summarization

Similarly, automatic summarization is a core Natural Language Processing (NLP) problem. In newsrooms, where journalists deal with an exponential growth in information that needs to be made sense of, getting help with more quickly and easily digesting core information contained in large documents is a much sought-after capability. Part of the challenge with summarization is that it can be a highly subjective task, where each user may have different requirements, information needs, or story angles they are pursuing, making generic summaries less useful. Controlled summarizations through salient named entities in the original documents is a potential solution to this limitation.

In recent work, we introduce ENTSUM, the first annotated dataset for controllable summarization that focuses on entities as control aspects (Maddela, Kulkarni, and Preoţiuc-Pietro 2022). This new dataset contributes quality dataset analysis, evaluation of generic and controllable

summarization methods, and adaptation of extractive and abstractive summarization methods for performing entity-centric summarization. By focusing on entity salience, the research achieves substantially better results on the dataset. The proposed process identifies and rates entity salience, then identifies all sentences in the article that are salient to the target entity, goes on to identify the sentences in the article that are used to make up the entity-centric summary, and writes a coherent summary of up to 150 words for the entity.

Our proposed framework utilizes an existing document summarization framework, which uses a Transformer model initialized with BART named GSum, which allows for using an input as a guidance signal along with a source document. This approach adapts GSum to generate entity summaries by using the entity information as a guidance signal. It similarly adapts BERTSum, which is constructed in a similar way to the GSum training procedure outlined above. Both are used to enlarge a *New York Times* corpus of document and summary pairs, resulting in multiple <document, summary> pairs for a single document. Among several key contributions, we introduce a new setup for learning entity-centric summarizations from generic summarization datasets, demonstrating good performance on the newly proposed ENTSUM dataset and paving the way towards integrating such functionality in our news editing and publishing processes.

Keeping the training data intact

Finally, as AI models gradually make their way into our business processes and workflows, countering adversarial perturbations becomes an essential task. Neural models in NLP can be especially vulnerable to adversarial attacks, and it is important to have solid mechanisms for mitigating such attacks, while maintaining good task performance. Based on the intuition that automatically generated adversarial inputs can be undone by learning to manipulate the textual input instead of retraining the classification model, we formulate a mechanism with the valuable property of transferability of defense, which allows the underlying classification model to be deployed to new and unknown models without retraining (Gupta et al. 2023). The assumption is that a single shared model can be more robust than the individual ones, while also reducing overhead when deploying it into new models.

The "text writer" is designed as an external module that serves as a pluggable component, enabling it to defend other models it was not explicitly trained to protect. The novel defense mechanism, Adversarial Text Interceptor and Rewriter (ATINTER), uses the text rewriter approach, along with a simple mechanism, to train this module. We demonstrate effectiveness on the novel mechanism against four benchmark datasets and five adversarial attacks, which substantially improves its adversarial robustness with a much smaller decrease in accuracy on nonadversarial inputs (Figure 3).

The proposed model learns how to remove perturbations by rewriting the text. The objective is to keep the model parameters unchanged, but rather manipulate the input by rewriting it. Instead of relying on random word replacements, the proposed framework employs a separate model that is explicitly trained to remove adversarial perturbations. Among other contributions, ATINTER shows great transferability to new models and datasets and generalizes across multiple downstream tasks and classifiers, all while improving adversarial robustness—a key feature for our use cases—by more than 10%.

BloombergGPT

In March 2023, Bloomberg published a research paper (Wu et al. 2023) detailing the multiple steps and design decisions taken to build and train BloombergGPT, an internal, not-for-commercial-use 50 billion parameter large language model (LLM), from scratch. As Bloomberg is primarily a financial data company, the goal for BloombergGPT was to research the development of a first-in-class LLM for financial tasks. Some of the potential use cases that benefit from the financial domain specialization include making financial texts and data more liquid, transforming natural language queries into code, or news-specific use cases such as headline suggestions or the task of financial question answering.

The model was validated on standard LLM benchmarks, along with Bloomberg internal benchmarks that reflect the potential use cases. BloombergGPT's performance was then evaluated on two broad categories of tasks drawn from multiple existing benchmarks: finance-specific and general purpose. For the Bloomberg-internal tasks, the research considered aspect-specific sentiment analysis, such as equity news sentiment (to predict the aspect-specific sentiment expressed in the news story toward a company), equity social media sentiment (using English language financially relevant social media content), equity transcript sentiment (using transcript from company press conferences), ES news sentiment (aspect-specific sentiment expressed in the news story towards a company's environmental and social policies) and country sentiment (sentiment expressed in the news story towards a country). Results show that across the four internal aspect-specific sentiment tasks, BloombergGPT performed better than the other tested open models of a similar size, by a wide margin, with the exception of social media sentiment, where it performed similarly.

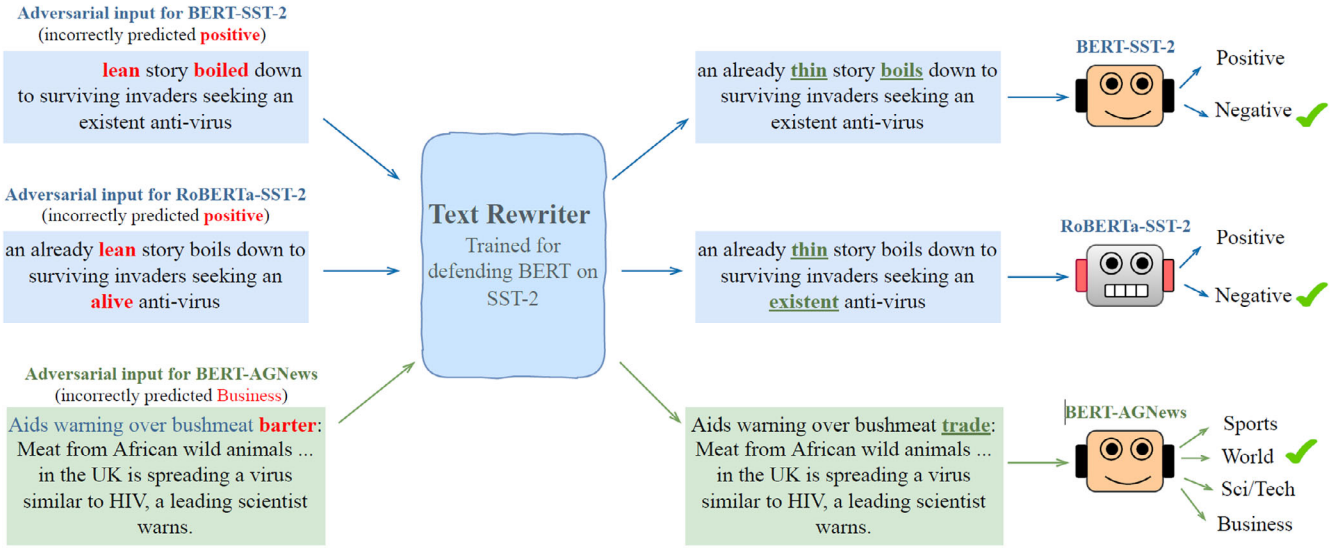


FIGURE 3 Modular application of ATINTER, which is trained for defending a BERT classifier for the Stanford Sentiment Treebank (SST-2) dataset. Demonstrating transferability across models, ATINTER successfully defends a RoBERTa classifier on SST-2 without retraining. Similarly, ATINTER is successful in defending a BERT model for a news classification task on AGNews.

In addition, Bloomberg researchers also explored how BloombergGPT would perform on named entity recognition (NER), a well-established NLP task that remains largely unexplored for generative LLMs. Finance-specific NER has subtleties that require extensive annotation guidelines in the prompt in order for the LLM to understand the intended tagging behavior. Seven internal NER datasets from different domains were considered: BN NER (English long-form Bloomberg News content), BFW NER (short-form stories from Bloomberg First Word), Filings NER (mandatory filings disclosures filed by companies), Headlines NER (headlines of Bloomberg English content), Premium NER (third-party English news content ingested by Bloomberg), Transcript NER (transcripts of press conferences), and Social Media NER (English language financially relevant social media content). The results from internal NER are mixed (see Table 1 with results from the original BloombergGPT research paper below).

Another exploratory task consisted of combining NER +NED (named entity disambiguation). In finance, NED is applied in tasks where the aim is to link mentions of publicly traded companies in texts to their unique corresponding ticker symbols. The NER+NED task looks to identify the stock tickers of companies mentioned in a document. This task is useful to evaluate the model's knowledge of companies, their various surface forms, and company to ticker mapping. BloombergGPT outperformed all other models on this task by a large margin, except on social media data (where it came in second). This result further underscores the advantage of BloombergGPT for financial NLP tasks.

In summarizing BloombergGPT's performance on standard, general-purpose NLP tasks, the paper concludes that BloombergGPT performs the best among similar sized models across dozens of tasks in many benchmarks. In some cases, it is even competitive with or exceeds the performance of much larger models, showing that the model's finance-specific training did not hinder its general-purpose abilities.

Part II—Bloomberg's News Innovation Lab

In 2015, Bloomberg created its News Automation team, with the mandate of producing automated content in order to free up time so journalists could dedicate themselves to higher-value newsgathering tasks. This meant identifying and understanding data and other structured flows of information, when that information is relevant for journalists—and therefore readers—and creating infrastructure, systems, and processes to handle and maintain these operations.

For a news organization like Bloomberg News, making sense of data in real time has always been a core competency. In its early stages, automation was focused on generating headlines when certain trigger conditions were met, customizing data-driven content with a rule-based approach, and creating very simple recipes that were programmed to process a trigger, get ticker data, and generate text. Bloomberg has always aimed to be clear about when a machine generated a story: every automated story includes "By Bloomberg Automation," language, which is then highlighted in bright yellow to provide

TABLE 1 Results on internal NER and NED datasets. On NER, while the much larger $BLOOM_{176B}$ model outperforms all other models, results from all models are relatively close, with BloombergGPT outperforming the other two models. On NER+NED, BloombergGPT outperforms all other models by a large margin.

	BLOOMBERGGPT	GPT-NeoX	OPT_{66B}	$BLOOM_{176B}$
NER				
BFW	72.04	71.66	72.53	**76.87**
BN	57.31	52.83	46.87	**59.61**
Filings	58.84	59.26	59.01	**64.88**
Headlines	**53.61**	47.70	46.21	52.17
Premium	60.49	59.39	57.56	**61.61**
Transcripts	75.50	70.62	72.53	**77.80**
Social media	60.60	56.80	51.93	**60.88**
All task (*avg*)	62.63	59.75	58.09	**64.83**
All task (*WR*)	0.57	0.29	0.19	**0.95**
NER+NED				
BFW	**55.29**	34.92	36.73	39.36
BN	**60.09**	44.71	54.60	49.85
Filings	**66.67**	31.70	65.63	42.93
Headlines	**67.17**	36.46	56.46	42.93
Premium	**64.11**	40.84	57.06	42.11
Transcripts	**73.15**	23.65	70.44	34.87
Social media	67.35	62.57	**70.57**	65.94
All task (*avg*)	**64.83**	39.26	58.79	45.43
All task (*WR*)	**0.95**	0.00	0.67	0.38

Abbreviations: BFW, Bloomberg First Word; BN, Bloomberg News content; NED, named entity disambiguation; NER, named entity recognition.

transparency to the reader. The machine also gets a double byline or an assistant attribution at the bottom of the article if the content contains some form of automated contribution.

As technology has evolved over the years, so has the complexity of stories we have been able to generate. One very early automated story aimed at empowering editors to report on earnings releases from top global companies. The sheer number of companies reporting around the globe during any "earnings season" means that journalists could previously only report on a subset of company names of interest. With the help of machines, the coverage universe has expanded greatly, adding a new level of transparency and actionable information for financial market participants.

Designed as automation assistance to ensure reporting consistency, the tool introduced an entirely new workflow. Whereas reporters would previously wait for an earnings press release to react and identify the numbers that matter, the automation workflow requires that news judgment be applied in preparation for earnings day to determine the expected key metrics ahead of the release being published.

When earnings coverage was a fully manual process, journalists aimed at publishing the first take of the story with the most relevant information. This process was optimized for speed, which meant providing the most relevant earnings figures as fast as possible for our readers. However, given the vast number of earnings releases that require coverage, it was not possible for reporters to add immediate context. Within seconds of an earnings release being published, readers now benefit from a well-structured story that provides all relevant numbers from the company's reported revenue and how those stand compared to analyst estimates for the reporting company. In addition, the system also provides available commentary from the press release and includes company-specific metrics from Bloomberg's "Corporate Financials" data, such as brands, products, operating metrics, and business or geographical segments. This all helps in interpreting the performance of thousands of companies worldwide.

Over the years, the automation system gradually evolved to provide relevant data points and context that are key for investors' decision-making processes, while also greatly reducing the time it takes for each story to be published. Beyond the actual release, news automation publishes previews and reviews, stories that capture a company's stock price moves, wraps, and market reactions. Publishing in different languages is also possible, and that selection is set up according to the interests of non-English speaking markets and countries.

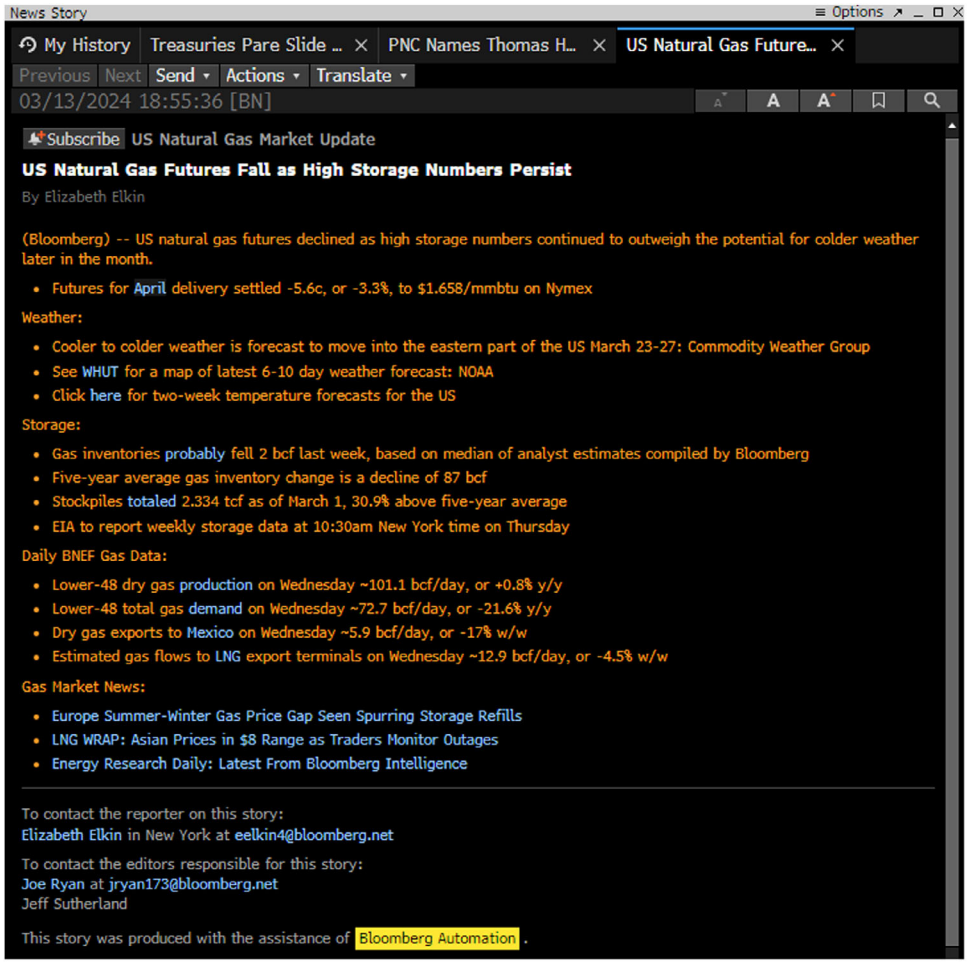


FIGURE 4 Sample of semi-automated story with assistance of Bloomberg Automation.

Degrees of automation

The inventory of active automations has grown to incorporate a variety of story types for all relevant asset classes, from daily and weekly wraps that help readers manage the information overload across a variety of market sectors and countries, to sophisticated data analysis. Automated stories are created in response to our customers' and readers' needs, and look to address new areas of interest in the news cycle and the availability of new datasets from which to draw insights.

Ideation and development all the way to deployment of an automated story involves a collaboration between editorial, domain experts, research analysts, and engineers. The creation process of an automated story requires months of research by an editorial domain expert in consultation with data experts. Currently, stories use rules-based structures to determine which market conditions or new financial data updates trigger a narrative. As we combine market knowledge and apply AI techniques to find patterns in different types of financial data such as options or exchange-traded funds (ETFs), the resulting financial narratives generated by the system can grow in complexity and provide greater insights and transparency.

Most rule-based automated stories start with a trigger event that starts a process which retrieves data from Bloomberg systems and applies relevant business logic. It can combine several datasets or previously published stories, add automated charts and tables, apply headlines and lead waterfall logic, and publish in multiple languages. Various types of automated stories exist, ranging from on-demand, that is, where journalists request the needed analysis for fresh and up-to-the-second data retrieval, to human-assisted stories, where either (i) reporters and editors report on critical data and the machine adds context before the story is published or (ii) where the machine collects the data and then sends it to the newsroom to be finalized and published (Figure 4).

As an added advantage to our journalists, each automated story can double as a highly efficient alert: instead of starting from scratch, the automated story becomes a combination of an alert and a robot news assistant that is monitoring raw data at scale.

Quality control

Testing for quality and performance is an integral part of the process. Automated stories are only released to the wire after extensive testing, and every single new version goes through this process. The tests typically consist of validating the story output against an extensive set of market configurations to ensure every data point and corresponding context reflects reality. That is, we test the story output for a diverse range of market conditions and context to ensure factuality. We also test for fairness, robustness, and overall acceptability of the output. Only after these extensive tests are concluded, and the human editor signs off, is it made available to readers. If a story fails the acceptability test, it does not get released.

Quality control continues after the story is released and is part of the daily operation. If even a single data point is found to be faulty, the affected stories are immediately corrected, adding a tagline to clarify the reason for the correction. There are systems in place to review and address the source of a data error, and this process follows Bloomberg newsroom's strict correction policies.

Thousands of decisions

During the day-to-day operation of a global news organization, thousands of big and small decisions must be made. Technology can assist with many smaller decisions, removing that onus on reporters and editors.

For example, take monetary value formatting. At the scale of Bloomberg News' coverage, this means formatting money and accounting for market conventions in more than a hundred currencies and in all variations of the supported languages. It needs to take into consideration the currency naming convention, the unique rounding logic, number formatting in different languages, acceptable abbreviations, and Bloomberg's own newsroom style guide. As an example, a USD-denominated price point can have up to 11 different forms of display to be accurate in all languages and formats:

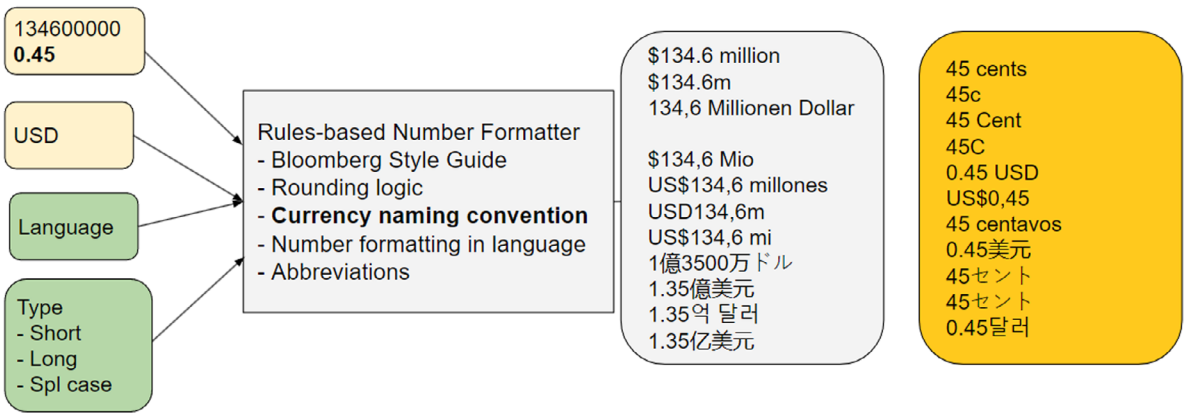
Monetary value formatting is only one such example where technology can help make better and faster decisions. Reporters and market participants are inundated with questions all day long: What was the market reaction last time this company reported earnings? Is this a relevant options strategy on a day like today? How unprecedented is this price move? The list is endless; using technology to assist in surfacing the necessary information and calculating the answers saves valuable time.

One-click complex data analysis

The "datasphere" has expanded rapidly and is expected to continue its growth trajectory. Multiple alternative sources of data join the traditional data sources from official national and international statistical agencies to feed an ever-growing appetite for data insights. In recent years, Bloomberg started to tap into a wider range of data analytics tools, such as BQuant,[2] to provide journalists with a no-code environment where complex data-crunching is available.

Embracing new technologies has given Bloomberg's newsroom the ability to quickly prototype what a data-heavy story idea might look like, or even to rapidly expand the story to multiple markets and variations as the news cycle requires. For example, in 2022, our UK bureau was looking for a relatable way to track the impact of inflation that would appeal to a broader audience. The News Innovation Lab created an "English Breakfast Index" that could track month-on-month the impact of inflation on the UK's English Breakfast. The first story was published in July 2022, and reporters were provided with an easy-to-use interface to recalculate the cost of the dish every month after inflation data was released by the official statistical agency.

This idea caught on and multiple bureaus around the world expressed an interest in tracking and publishing similar stories. The family of food indices has grown to track



food prices for Italian Pizza, Spanish Paella, South Africa's Shisa Nyama, French Coq au Vin, American Dinner and Egypt's Koshary meal. From the creation and monitoring of the data in all those countries, it became clear that national dishes were becoming a luxury around the globe. Food inflation kept rising at a moment when headlines indicated that overall inflation was stabilizing.

Using technology has made it a bit easier to make sense of complex topics and made prepackaged data analysis available to journalists on-demand. To fulfill that role, the members of the team constantly partner with journalists and editors across the global newsroom to utilize and deploy bespoke techniques. This helps them break news or quickly perform an in-depth analysis or use technology to find insights and track noteworthy data-heavy developments that our readers care about.

Now, with the recent developments in the field of LLMs and generative AI, new and exciting areas of exploration have opened up. Market participants are eager to figure out what can be developed and what is possible with this new technology. For us it means looking to identify solutions that are in line with the high threshold of Bloomberg News' accuracy requirements.

Part III—The era of generative AI

When an LLM was deployed into what has become a popular chat application, a new era of technology effervescence began. The assumption is that LLM technology can transform productivity in ways and use cases that are yet to be discovered. However impressive, LLMs have several known limitations and ample societal implications.

In financial news, where accuracy is a paramount value, the future use of LLMs in newsgathering processes will require carefully designed strategies and testing, to mitigate any potential risks.

Risks and rewards of using LLMs in news

LLMs bring interesting opportunities for newsgathering and news production. At Bloomberg, we are using our experience in news automation, and the accumulated knowledge and well-designed processes created to support that operation, as the starting point to experiment with LLMs for news.

Given some of the characteristics of this type of technology, our initial research and tests have helped inform our best practices around LLMs. Our guiding principles around the use of generative AI fully align with those of journalism and Bloomberg News: accuracy, fairness, and transparency. Any experiment we pursue is focused on generating value for our readers, viewers, and listeners. For our newsroom, we focus on using the technology to assist journalists in fact-based reporting.

By having the interest of our readers and clients at the heart of any approach, testing, and deployment criteria, we set a high bar to safeguard the integrity of our news flow (that is, a news flow with accurate, fair, and balanced stories) and trust in our coverage. Our approach is to center our newsroom experimentation and development around the notion of having a human in the loop, and design solutions that assist and help journalists do their jobs.

The risk of errors or hallucinations is well documented. Hence, when we introduce any tool to a journalistic workflow, we require the careful eye of experienced editors, alongside clearly defined processes, for validation and verification.

Beyond regular news workflows, LLMs can also be a valuable tool for investigation. A journalist might need to digest vast datasets or large pools of documentation to spot patterns and trends required to support their research into a story. LLMs can also help produce variations of content that is tailored to the specific needs of audiences, such as articles produced in local languages or segments summarized for other products.

At the time of this writing, we are asking a myriad of "what if" questions. What if the LLM could accurately repurpose stories to the requirements of different products and platforms? What if the LLM could summarize for context from a large set of articles? What if the LLM could optimize headline variation? What if the LLM could help make sense of how a story is evolving?

Each of these questions, and many others that emerge from the daily challenges of newsgathering, represent an area of future experimentation. Initial experiments indicate promising results in the use of LLMs to assist editors in linguistic tasks. In turn, we are opening up the possibility of additional exploration, including the integration of the technology into publishing systems, scalability of models and tools across bureaus around the globe, model training in the context of ever-evolving news cycles, and importantly: development of editorial processes and systems to validate references and mitigate risks. As we investigate and learn more, a long process of careful validation begins.

CONCLUSION

Bloomberg has been at the forefront of innovation and financial information delivery since 1981 and has developed a leadership position in developing and applying AI in the finance and news domains since 2009. The News Innovation Lab builds upon this long legacy by focusing on driving the innovation efforts for our global newsroom, drawing from the collective domain knowledge available

at Bloomberg to combine the expertise of AI researchers and engineers, data scientists, editorial, and asset-class experts. We identify new technologies that can enhance the ability to break news and provide our readers with high-quality insights, we explore the frontiers of newsgathering and news monitoring at scale, and we look to develop data modeling techniques and solutions that meet the ever growing appetite for data insights.

Generative AI and LLMs were recently added to the News Innovation Lab's toolkit, and we are now exploring whether these technologies can be used to generate solutions and applications that will help serve our readers better. The possibilities are plenty and point to a flourishing area of exploration.

However, as with any new and emerging technology, it is important to fully understand the limitations of the technology and mitigate risks in order to create valuable applications. As we explore and develop solutions with AI technologies, we remain committed to taking steps to safeguard the journalistic values of accuracy and fairness. If it is not true, it is not news.

ACKNOWLEDGMENTS

The authors would like to thank the leaders of Bloomberg's News organization, the members of the News Innovation Lab, the AI and automation engineers at Bloomberg who help design and build the firm's AI/NLP infrastructure and solutions, as well as the data analysts, editors, and journalists who both make use of them and provide valuable feedback to improve and extend them.

CONFLICT OF INTEREST STATEMENT

The authors declare that there is no conflict.

ORCID

Claudia Quinonez https://orcid.org/0009-0008-4624-7929
Edgar Meij https://orcid.org/0000-0003-0516-3688

ENDNOTES

[1]Where H1 is the headline for the existing article and 'B $_{edits}$' are the changes only to the article body.

[2]BQuant is Bloomberg's next-generation analytics solution, which is designed specifically for quantitative analysts and data scientists in the financial markets. They use it to conduct data analysis using Python and to seamlessly publish models as applications to Bloomberg Terminal users.

REFERENCES

Capponi, A., and Lehalle, C.-A., eds. *Machine Learning and Data Sciences for Financial Markets: A Guide to Contemporary Practices*. Cambridge, UK: Cambridge University Press.

Gupta, A., C. Blum, T. Choji, Y. Fei, S. Shah, A. Vempala, and V. Srikumar. 2023. "Don't Retrain, Just Rewrite: Countering Adversarial Perturbations by Rewriting Text." In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, ed. Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, 13981–98. Toronto, Canada: Association for Computational Linguistics. https://aclanthology.org/2023.acl-long.781

Kumar, D., V. Gupta, S. Sharma, and S. Zhang. 2022. "Realistic Data Augmentation Framework for Enhancing Tabular Reasoning." In *Findings of the Association for Computational Linguistics: EMNLP 2022*, eds. Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, 4440–58. Abu Dhabi, UAE: Association for Computational Linguistics. https://aclanthology.org/2022.findings-emnlp.324

Maddela, M., M. Kulkarni, and D. Preoţiuc-Pietro. 2022. "EntSUM: A Data Set for Entity-Centric Extractive Summarization." In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, eds. Smaranda Muresan, Preslav Nakov, and Aline Villavicencio, 3355–66. Dublin, Ireland: Association for Computational Linguistics. https://aclanthology.org/2022.acl-long.237

Micklethwait, J, P. Addison, J. Sondag, and B. Grueskin. 2017. *The Bloomberg Way: A Guide for Journalists*. Bloomberg Press.

Panthaplackel, S., A. Benton, and M. Dredze. 2022. "Updated Headline Generation: Creating Updated Summaries for Evolving News Stories." In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, eds. Smaranda Muresan, Preslav Nakov, and Aline Villavicencio 6438–61. https://aclanthology.org/2022.acl-long.446.pdf

Spangher, A., Y. Ming, X. Hua, and N. Peng. 2022. "Sequentially Controlled Text Generation." In *Findings of the Association for Computational Linguistics: EMNLP 2022*, eds. Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, 6848–66. Abu Dhabi, UAE: Association for Computational Linguistics. https://aclanthology.org/2022.findings-emnlp.509/

Tan, H., C. Tsai, Y. He, and M. Bansal. 2022. "Scientific Chart Summarization: Datasets and Improved Text Modeling." *The Third Workshop on Scientific Document Understanding at the 36th AAAI Conference on Artificial Intelligence*. https://assets.bbhub.io/company/sites/40/2022/02/AAAI_2022_Chart_Summarization_v3.pdf

Wu, S., O. İrsoy, S. Lu, V. Dabravolski, M. Dredze, S. Gehrmann, P. Kambadur, D. Rosenberg, and G. Mann. 2023. "BloombergGPT: A Large Language Model for Finance." *arXiv*, March 31, 2023. https://arxiv.org/abs/2303.17564

How to cite this article: Quinonez, C., and E. Meij. "A new era of AI-assisted journalism at Bloomberg." *AI Magazine* 45: 187–99. https://doi.org/10.1002/aaai.12181

AUTHOR BIOGRAPHIES

Claudia Quinonez is the global head of the Content Innovation division in Bloomberg's News Product, Tech & Publishing (NPTP) group. She leads teams dedicated to Automation Workflows, Content

Creation, and the News Innovation Lab. With a career at Bloomberg of more than two decades, Claudia has managed global teams focused on news in different departments, including Sales, News, and Product, in which she gained unique insights on financial market participants' news and information needs. She has an extensive track record in applying a variety of technologies to explore new frontiers in storytelling and content creation. Claudia earned an executive master's degree in Strategy and Innovation from Saïd Business School/Oxford University and a master's degree in philosophy and artificial intelligence from Northeastern University London.

Edgar Meij is the head of the AI Platforms division in Bloomberg's Artificial Intelligence (AI) group and leads 10+ teams of engineers and researchers that are responsible for all key AI, NLP, ML, LLM, and Search technology platforms used across the company. Edgar holds a Ph.D. in computer science from the University of Amsterdam and has an extensive track record in artificial intelligence, information retrieval, natural language processing, machine learning, large-scale computing infrastructures, knowledge graphs, and semantic search. He has published more than 110 papers in top international venues, which have been cited more than 2,000 times. He is also a (Senior) Program Committee member of virtually every major conference in the field (including The Web Conference, WSDM, SIGIR, CIKM, KDD, AAAI, ACL, and EMNLP), has organized tutorials and workshops at those same conferences, has served as sponsorship co-chair for The Web Conference and ECIR, and is co-chair of the Industry Track (SIRIP) for SIGIR 2024.