

INDUSTRY WATCH

Navigating the Text Generation Revolution: Traditional Data-to-Text NLG Companies and the Rise of ChatGPT

Robert Dale

Language Technology Group
rdale@language-technology.com

Abstract

Since the release of ChatGPT at the end of November 2022, generative AI has been talked about endlessly in both the technical press and the mainstream media. Large language model technology has been heralded as many things: the disruption of the search engine, the end of the student essay, the bringer of disinformation ... but what does it mean for commercial providers of earlier iterations of natural language generation technology? We look at how the major players in the space are responding, and where things might go in the future.

**Note to the copy-editor:
this piece is deliberately written in an informal style.
Please do not remove this informality;
in particular, please do not replace contractions like ‘isn’t’ and ‘hasn’t’.**

1. Introduction

Natural language processing (NLP) has two daughter sub-disciplines, these being natural language understanding (NLU) and natural language generation (NLG). But the siblings are not equal; NLG has historically been the ‘poor sister’ in the relationship. From a scholarly perspective, this is most obvious in terms of the balance of papers presented at NLP and AI conferences, where, until more recently, the number of presentations exploring NLU vastly outnumbered those addressing issues in NLG. But the imbalance was just as obvious in the commercialisation of NLP research, where the application of NLG technology again trailed that of NLU. Towards the end of the last millennium, you could count companies pitching NLG solutions on one finger, the sole company in the space being CoGenTex (founded 1990; no longer active). The company was most well-known for its generation of weather forecasts in English and French for the Canadian weather service; that application fell out of use in the mid-1990s, and for a while the commercial NLG landscape lay rather barren.

Interest in the commercial opportunities for NLG re-ignited in the ten years beginning in 2007, a period which saw the founding of around a dozen companies in the space. That’s promising, but it’s still just a drop in the ocean compared to the number of companies that were providing software solutions addressing some aspect of NLU.

Fast forward to the 2020s, and things have changed big-time. Thanks to the appearance on the scene of generative AI, the concept of a machine generating textual output has become widely

recognized in the mainstream press as a thing, and there are now dozens of startups offering products built around text creation. But, as Spock might have said to Captain James T Kirk, ‘It’s NLG, Jim, but not NLG as we know it.’^a Today’s NLG techniques, based on large language models, have very little in common with the ‘good old fashioned AI’ approaches underlying the earlier generation of commercial offerings in the NLG space.

Clay Christensen’s oft-cited book, *The Innovator’s Dilemma*, explores the challenges faced by established companies when innovative technologies emerge. Christensen observes that companies often dismiss or underestimate innovations because they typically lack the performance or features that appeal to mainstream customers. His book discusses numerous examples in detail, but perhaps the most accessible is the shift from film photography to digital photography and how established companies like Kodak struggled to adapt. Despite being a pioneer in digital photography technology, Kodak’s focus on its highly profitable film business prevented it from fully embracing the disruptive digital technology, ultimately leading to its decline.

It’s hard not to wonder whether the arrival of ChatGPT and friends might be that Kodak moment for established NLG players. In this article, we look at how ‘traditional’ NLG companies have responded to the appearance on the scene of generative AI, and what this might mean for the future.

2. The commercial NLG landscape

2.1 A historical timeline

As noted above, the latter half of the noughties kicked off a surge of interest in commercial NLG, most visibly with the arrival of Automated Insights (founded 2007) and Narrative Science (founded 2010). Both companies signed contracts with media outlets to mass-produce corporate earnings previews: Automated Insights for the Associated Press, still apparently running on [the AP website](#), and Narrative Science for *Forbes Magazine*, a service which appears to have ended in 2015, but whose content you can still find [online](#).

Both companies also produced sports reports, another genre that was a good fit for the technical approach these vendors adopted: a glance over even a handful of outputs from either system quickly makes evident their template-based nature. Neither of these companies exist as separate entities any longer: Automated Insights was acquired by Vista Equity Partners in 2015 and Narrative Science by Tableau, which in turn was acquired by Salesforce, in 2021.

There were also at least two other NLG companies operating in Europe at this time: Yseop (founded 2007) and Data2Text (2009; subsequently acquired by Arria NLG). But these were far less visible than the US companies.

Perhaps buoyed by the apparent success of Automated Insights and Narrative Science, the next few years saw a comparative explosion of interest, with the appearance on the scene of InfoSentience (founded 2011), Arria NLG (2012),^b 2TXT (2013), Retresco (who started building NLG applications in 2014, although the company was founded in 2008), Textual.ai (2014), Narrativa (2015), vPhrase (2015), AX Semantics (2016), and United Robots (2016). All of these companies are active today, and we’ll take a closer look at them below.

At least in terms of new companies being founded, things seem to have gone quiet again after 2016. I’m not sure what that means. Market saturation by the existing players seems unlikely. Perhaps the increasing visibility of deep learning technology caused both founders and investors to hesitate, waiting to see where the commercial potential might be most promising.

^a<https://oupacademic.tumblr.com/post/45375455323/misquotation-its-life-jim-but-not-as-we-know>

^bDisclaimer: I hold shares in Arria NLG, and worked for the company from 2012–2017. As CTO there, I was responsible for the design and development of what is now known as Arria Studio, a rule-based templating system of the kind discussed in Section 2.2.

2.2 A framework for analysis

Before we look at these companies in a little more detail, it will be useful to have something of a framework within which to discuss them.

Software companies in general often develop through a similar series of stages:

- Stage 1 is the startup phase, typically involving the building of a bespoke application for a specific client. If that's successful, it gives the company something to brag about, and the process is repeated for a few more clients.
- At Stage 2, the company recognizes common patterns and functionalities across the applications it develops, and factors these out into an internal toolset or framework that streamlines the development process and enables faster delivery of new applications; this accelerates the company's growth.
- In many cases, however, it becomes apparent that the bottleneck is human resource; even with the productivity increase the internal tools provide, the company can't achieve the kind of scalability that investors demand. To address this, it's not uncommon to move to a Stage 3, where the internal tools are transformed and repackaged into a low-code platform that can be offered to external customers, so they can build their own applications. Today, that platform is often made available on a SaaS subscription basis.
- Some companies find that even their low-code Stage 3 solution is too difficult for a wide customer base to be able to use directly, resulting in a hefty upfront professional services component in deploying the platform for a customer and training its users.
- A Stage 3 company built on a platform supported by professional services can still be a sustainable business. For some, however, the demands of scalability result in a move to a Stage 4, characterized by the packaging of use-case specific functionalities into no-code configurable products, with the professional services component of each sale dropping to zero or near zero.

The numbering of stages here is not intended to imply that later stages are somehow better: the particular stage a company settles at comfortably will depend on a range of factors, including the specific domains, use cases and customer capabilities the company seeks to address.

The above is a characterisation (and of course a simplification) of the development of a broad range of different types of software company. The development path of NLG companies adds some specifics to this.

First, the technology developed in Stages 1 and 2 is often based on a theoretical model of language generation, borrowing ideas from computational linguistics or other academic disciplines. So, for example, CoGenTex's RealPro was inspired by Meaning-Text Theory; InfoSenticence makes use of a notion of conceptual automata; 2TXT's platform is loosely based on HPSG; Yseop's underlying framework is based on Abstract Categorical Grammar; Narrativa's platform is built on knowledge graphs. These frameworks typically provide powerful abstractions that support the development of sophisticated language generation applications, but they may appear impenetrably esoteric if you're not an insider.

Correspondingly, Stage 3 is sometimes a radical reworking of the internal framework to remove sophisticated linguistic abstractions that are alien to the typical end user. Combined with the observation that, for many practical tasks, a large proportion of content can be considered canned text that doesn't need to be generated from first principles, this often ends up with the development of a rule-based templating system, sometimes referred to (perhaps with disparaging intent) as 'mail-merge on steroids'. My impression is that Automated Insights and Narrative Science had these Stage 3 insights from early on, resulting in the development of the former's Wordsmith template authoring tool and the latter's Quill framework.

The various vendors' Stage 3 platforms are all different in interesting ways, but their architectures tend to share the same set of components:

- an underlying template-based language that intersperses canned text with variables, function calls and conditionals or other control statements;
- an editing environment/user interface that aims to provide a low-code/no-code route to defining and editing templates, shielding the user from the syntax of the underlying templating language;
- a collection of interfaces for importing data from different sources;
- a mapping layer that supports numeric, alphabetic and other transformations being applied to input data values;
- some degree of 'linguistic smarts', typically to automatically handle morphological phenomena like number agreement or orthographic details like sentence casing; and
- a runtime environment and associated API that supports deployment of templates to a delivery mechanism, either cloud-based or on-premise; this may support a variety of output formats.

We'll assume this basic structure when we describe each of the companies below.

2.3 The players

So where do the traditional NLG companies sit with respect to this framework? We map them out here along the series of stages identified above.^c

2.3.1 InfoSentience

[InfoSentience](#) positions itself as a content automation business that develops bespoke multimodal interactive reporting tools for clients like the Chicago Mercantile Exchange. Some of the applications the company has built bear a resemblance to the integrations of NLG into business intelligence tools that was a popular focus for a number of other players three or four years ago. However, the bespoke nature of the applications InfoSentience develops allows them to be much more sophisticated, and as such, they are excellent demonstrations of just how language and visualizations can be combined to maximum effect via what the company refers to as 'narrative engineering'. The reliance on a complex underlying model built on what the company calls conceptual automata means that it's unlikely we'll see a toolkit that lets you build your own any time soon. In terms of the framework outlined above, InfoSentience is a Stage 1 company, and seems content to be there for the time being.

2.3.2 Narrativa

[Narrativa](#) primarily focuses on the life sciences (in particular the generation of clinical study reports) and financial services; the company is also used by the Wall Street Journal for some finance news reporting. Since its inception, the company has built complex bespoke NLG applications using Gabrielle, an internal toolset that involves mapping input data into a knowledge graph, then generating text by traversing this graph. This isn't a self-service offering, although it has been used externally; but in order to service demand, Narrativa has evolved its technology to produce a product for a broader audience. The resulting SaaS tool, called the Narrativa Generative AI Platform, is now in beta testing with a client. The platform's overall architecture is similar to the standard model described above. The scope of the mapping layer is the broadest of the vendors surveyed here, in part because it supports transformations over input text fields as well as numeric

^cThe material here is based on conversations with representatives of the respective companies during late May and early June 2023. By the time you read this, companies' positioning and products may have changed.

and symbolic data, allowing the application of NLU capabilities like summarization, translation, sentiment analysis and entity extraction to be used to create new data points that can be incorporated into text templates. Any third-party language processing component can be integrated in this way. In terms of our staged framework, Narrativa is in the process of transitioning to Stage 3.

2.3.3 Arria NLG

[Arria](#) positions itself as a broad-purpose self-service NLG solution focussed on process automation; the most common use cases addressed are in various kinds of financial analysis. The company's core product is Arria Studio, a template editing and development environment with an associated runtime platform that can be delivered in the cloud or on premises. The template editor provides a low-code interface to writing scripts in ATL, the company's proprietary template language, which provides a wide range of functions that can be used to manipulate numerical data and text. Recognizing that the self-service low-code editing approach is still difficult for some customers, the company is looking at a shift towards a configurable pre-built applications that can more or less be used 'out of the box'. The challenge here is balancing simplicity of UI against the user's desire to control detailed aspects of the output. In terms of the development stages outlined above, the company is at Stage 3, and considering approaches to Stage 4.

2.3.4 vPhrase

[vPhrase](#) produces an NLG tool called Phrazor, which allows the user to build interactive dashboards that combine visual and textual elements, much like a BI tool. A focus for the company has been finding the right level of abstraction for the customer to be able to self-service. The current approach has three levels at which the customer can build an NLG application:

- The tool comes with around 100 pre-built templates for common reporting tasks across a range of industries; the user configures the template by mapping input data elements to the parameters of the template.
- Alternatively, the user can build a report from a set of inbuilt analysis functions that deliver individual insights (such as comparisons and trends); the Phrazor Editor is a no-code GUI that supports an interactive process for gathering configuration information, and then provides a drag and drop interface for building up a report by combining visual and textual components.
- Using a Business Logic and Language editor, the user can also modify the underlying template code used to generate the insights, or build new insights.

In terms of development stages, the multiple levels at which the user can interact with the technology mean that the company is operating at both Stage 3 and Stage 4.

2.3.5 Retresco

[Retresco](#) positions its offering as 'hybrid natural language generation', this being content generation based on text templates combined with the potential of GPT models. The company's initial focus was on tools for content management for publishing houses and media providers, using NLU technology for index linking and entity recognition. These tools are still in use. Retresco started building NLG applications in 2014, leading to the development of the textengine.io platform; this is a rule-based template engine with some proprietary machine learning technology to address grammatical issues (specifically, morphology and determiner choice). The platform is available as a SaaS platform or on-premises deployment. Today, the company's primary focus is

on e-commerce, but it also provides solutions for media and sports reporting. In terms of development stages, the company is at Stage 3; the company is also willing to build bespoke NLG applications.

2.3.6 AX Semantics

AX Semantics positions itself as ‘scalable e-commerce text automation’, and provides a self-service rule-based NLG platform for content automation and optimization. The platform is incorporated into a workflow management tool for e-commerce, although the platform is also used in a number of other domains, including publishing, finance and pharma. The platform includes a graphical UI for building up data transformations, built on top of a full-blown data manipulation language, along with a template authoring environment that supports machine translation (via neural MT) of templates in around 30 languages; when statements are translated, the rules and logic are automatically incorporated into the target language statements. Dependency parsing of the source language is used to aid in the identification and translation of linguistic elements, allowing control of grammatical phenomena such as appropriate determiner use. The platform also supports real-time collaboration between multiple users on the same rule set. In terms of the development stages outlined above, AX Semantics is at Stage 3.

2.3.7 Textual.ai

Textual doesn’t claim to be a general NLG solution: it specifically targets the e-commerce sector as an audience and focusses on the generation of product descriptions, long form content, category pages and brand content. Consequently, the company’s offering is essentially a configurable workflow management system and content management system tailored to the needs of e-commerce, combining a range of content creation and translation functions that are useful for product descriptions across different versions for different channels, languages and audiences. The economics of using the platform are such that the company’s smallest clients generate as few as 1000 texts per year, all the way up to 800k texts for the company’s largest client. The platform is built around a rule-based templating system that supports a wide range of general string processing and language processing functions that can be used to manipulate content. Machine translation is integrated into the workflow for translation of any generated or manually written content; Textual also provides ancillary manual services, such as human translation and proofreading. In terms of the development stages outlined above, Textual is at Stage 3; however, the company reports that very few clients build the templates themselves, so the company emphasises its focus on speedy support, and targets enterprises where the business case can allow many hours of configuration.

2.3.8 Yseop

Yseop focuses on highly regulated industries; the main use cases are regulatory report automation for the pharma industry and financial report automation for financial institutions. Yseop Compose is the company’s earlier data-to-text software platform; this isn’t customer facing, but may still be used to build bespoke applications. The company’s self-service offering is a Microsoft Word plug-in called Augmented Analyst: this is a no-code interface used to deploy Automation Packs, which are use-case specific NLG applications. For any given deployment, the Automation Pack will be configured for the specific customer’s requirements; the document author can then set a number of parameters and thresholds that define the particular report to be generated. At the time of writing, Yseop had just announced Yseop Copilot, an enterprise software platform that leverages pre-trained LLM models for the BioPharma industry; this appears to be an evolution of the Augmented Analyst model, providing a no-code configurable platform that aids in the

authoring of specific kinds of regulatory documents such as Clinical Study Reports. This puts Yseop at Stage 4 in the framework outlined above.

2.3.9 2TXT

[2TXT](#) focusses mainly on e-commerce, although the company has also delivered applications in finance, travel and the public sector. The company's solution might best be described as 'content-as-a-service'. The customer brings their product data; 2TXT maps the product catalog onto a proprietary product ontology, which is in turn paired with an already existing grammar-based system for describing products. The ontological mapping is the only step required to create a new deployment; for each category of product, 2TXT has a pre-built text model, with new categories being added as required. In production mode, the customer simply provides their data and gets back product descriptions; they don't have to concern themselves with the internal workings of the platform. The differentiator here is the simplicity of use from the customer's point of view. In terms of the development stages outlined above, 2TXT is at Stage 4.

2.3.10 United Robots

[United Robots](#) provides local news stories around sports, real estate, weather, and a few other topics, determined by the availability of the relevant data in the geographical area being serviced. The company takes the content-as-a-service model to its logical conclusion, providing its clients with the required content without any need for the customer to be involved in the process of its creation other than in the initial setup. The company takes care of sourcing the required data, then works with the client to develop text models appropriate to the client's journalistic style; the required stories are then generated without client intervention. United Robots is exploring the use of LLMs for various aspects of the text authoring process, but the requirement for factuality in combination with end-to-end automation rules current models out for generating individual texts. This is very much a Stage 4 company.

3. Responses to generative AI

When a company is faced with a new technology, it can ignore it as irrelevant, acknowledge its strengths and weaknesses, or embrace it. Certainly none of the NLG providers are ignoring generative AI; all have acknowledged it; and just about everyone has embraced it to some degree or other.

A number of the companies I spoke to indicated that the widespread interest in generative AI has caused an increase in inbound sales enquiries. Inevitably, this leads to a scenario where the customer has to be educated in regard to the difference between generative AI and the more traditional approach to NLG. To support this, each company has expended marketing effort on this, whether it be through webinars and meetups ([AX Semantics](#), [Retresco](#)) or blog posts ([Arria](#), [AX Semantics](#), [InfoSentience](#), [Retresco](#), [United Robots](#), [Yseop](#)). These resources generally provide fair assessments of the advantages and disadvantages of both the more traditional 'deterministic' NLG and the newer generative AI approach. They speak with one voice, and the message is clear: at least as things stand now, generative AI is great for creative content, but it runs the risk of hallucinating, and so cannot be relied upon in the way that a deterministic approach to NLG can be. The bottom line: reliability—the absence of hallucinatory risk—is traditional NLG's moat.

Beyond acknowledging its strengths and weaknesses, pretty much everyone has at least started integrating the new technology into their existing platforms in some way. It's important to draw a distinction here between using generative AI in the template authoring process vs using it in the production of individual text outputs. Human sanity checking of LLM output that is incorporated in a fixed form into a template at the authoring stage (what might be called a 'static template') is

a viable exercise; sanity checking even 100 output texts that each contain different generative AI output (produced by what we might refer to as a ‘dynamic template’) is likely to be infeasible in most circumstances, and out of the question if you are generating thousands of texts.

Consequently, by far most integrations of generative AI have been in the form of features that assist the human author in building a static template. Most platforms already provide support for pre-defined random variations in the output, to avoid each text generated reading too much like all the others; but coming up with variants is recognized as a time consuming problem for users, so automatic variant suggestion is an obvious target for generative AI. AX Semantics, Narrativa, Retresco, Textual.ai and vPhrase have all deployed some form of synonym and phrasal variation for a number of years, with early attempts by some using word embeddings; the appropriateness of the suggestions proposed improves considerably with the context awareness provided by LLMs. vPhrase’s plug-ins for Power BI and Tableau take this a step further, using an LLM to rephrase insights data-generated by Phrazor to match a pre-specified persona.

vPhrase also has the most sophisticated variation on this functionality that I’ve seen: Phrazor’s business logic and language editor uses generative AI to reduce some of the template-writing burden by automatically deriving all the logically possible alternatives of a given complex condition. You specify one branch of the conditional along with its corresponding text output, and the tool generates both the conditions and output templates for all the other logically possible branches. I suspect this can easily go wrong, or at least produce results other than those the author had in mind, and therefore require some tweaking; but it’s a neat feature that addresses a time-consuming pain point.

Yseop’s Copilot, focussed on clinical study reports, uses the company’s own pre-trained LLM models to provide narrative content. Given the factuality requirements of the domain, these capabilities are limited to text-to-text transformations such as summarization and tense change of previously human-written content, and the syntactic correction of telegraphic physician’s notes. This might be seen as a higher risk endeavour, but it’s important to note that, whereas a product description generator might produce 100s or 1000s of texts in a single run, clinical trial reports are produced one at a time, with much more scope for human quality assurance.

Beyond producing versions of a text that’s already been written, a number of providers (AX Semantics, Retresco, Textual.ai) aim to provide support for ‘writer’s block’. Typically this is done by allowing the user to provide some basic configuration parameters (type of message to be generated, data elements to be mentioned, SEO terms to use), in response to which the tool proposes a sentence template that can then be edited as required. Arria is exploring whether the tool might go further to suggest analysis types that can be carried out on the data provided.

There’s also scope for co-authoring, where the respective strengths of the data-driven template approach and the generative approach are combined: for example, data-driven templating might be used to describe a specific stock portfolio’s performance, but generative AI might be used to provide static background content on what’s happening in the market more broadly.

2TXT’s reaction to ChatGPT was to launch an experimental product which is end-to-end GPT, generating a product description directly from an array of product data; you can try it out on [the website](#). This is very scalable, but apparently in German, at least, hallucinations are sufficiently common that the QA overhead is simply too great to make this a feasible solution; the unit cost of text generation might seem low, but the cost of individual checking makes it uneconomic. From 2TXT’s point of view, offering the GPT solution alongside its more traditional offering also presents a marketing challenge, since having two products that have complementary strength profiles and pricing but look very similar from the customer perspective can be confusing for potential clients.

2TXT and United Robots both say they are exploring the use of generative AI for authoring assistance, but given their shared positioning as product-as-a-service, this will be invisible to the user.

Narrativa supports generative AI prompts not only in the template editing stage, but in the data mapping stage (recall the common Stage 3 architecture outlined at the end of Section 2.2). Whereas data mapping for most vendors involves arithmetic or string operations on existing numeric or symbolic data elements, Narrativa supports prompt-driven operations here too, for example to summarize existing text data points, or to convert their tone or style. This means that content blocks can then incorporate generative AI outputs directly from the mapping layer. Again, hallucinatory risk demands that this feature should only be used in low-volume human-checkable scenarios.

Finally, there's Textual.ai, whose text authoring platform provides the most full-blown integration of GPT as a central component. A key feature is the Copy Assistant, which assists in the development and management of prompts for both category descriptions and product descriptions. A prompt design tool enables reuse of previously-constructed prompt elements, and allows the user to fine-tune instructions in a multi-shot fashion, adding the actual product data to be added at the last step; once tested, the fine-tuned prompt can then be used for other products in the category. Prompt selection is triggered by specified combinations of metadata values. The text editing tool allows interleaving of template-generated and GPT-generated text blocks; machine translation is used to generate other language versions of templates. The company really has gone all-in on the new tech: it also pitches itself as a 'GPT agency', providing expertise on how and where to use the technology.

Standing back a bit, it becomes apparent that the vendors reviewed here can be divided into two camps with regard to their stance on hallucinatory risk. Some companies, such as AX Semantics and Arria, emphasise end-to-end automation as a key feature of their solutions, which rules out the option of having a human carry out quality assurance on each individual output; consequently, they can only integrate risky technology at the template authoring stage. On the other hand, Narrativa, Retresco, and Textual leave the choice to the user; they all support the development of dynamic templates, where the output for a given data vector is unpredictable at template authoring time, with a consequent QA risk that the customer can choose how to address (or not) as the circumstances permit.

4. Disruption and disintermediation

We noted above Clay Christensen's observations about the responses of companies to innovative technologies. In fact, Christensen distinguishes two kinds of innovations: sustaining innovations, which involve companies improving their existing products or services; and disruptive innovations, which typically lack the performance or features that appeal to mainstream customers. The tendency he observes is that companies tend to focus on sustaining innovations. Arguably that's where most NLG companies are today: generative AI enables useful augmentations to their existing technology stacks. On this view, LLMs are a sustaining innovation, not actually a disruptive one. And that leads to an obvious question: What would it mean to be disruptive in the context of NLG?

A common characteristic of many disruptive innovations is disintermediation, the removal of intermediaries or middlemen in a particular process or industry. Most obviously here, generative AI might disintermediate the template as a key step in the process of language generation: why spend your time planning out how to tell a story with all its conditional variations and data inclusions when you can just ask an LLM to build a story of some particular type around some key data points? It's quite conceivable that an LLM-generated narrative will better adjust the story angle to the idiosyncracies of a particular set of data values than would be achievable by a human template writer trying to predict the different possible angles in advance.

Of course, the traditional NLG companies might adapt to such an advance by increasing the scope of prompt-based generation within their existing tools, allowing any combination of

template-driven and prompt-driven output, including the case where the template-driven content is reduced to nil. But the potential replacement of a template-based approach by a prompt-based approach also lowers the barrier to entry for new players in the space, removing the need to build a lot of the machinery the incumbents have invested so heavily in. You still need a user interface, of course, but now this can be a much simpler UI—and one that most of the new generative AI-based copy-writing tools have already developed.

There's another potential disintermediation that doesn't just mess with the existing NLG toolsets, but might make them irrelevant: the disintermediation of the document as a carrier of information. Documents have long been a primary medium for storing and communicating information, acting as intermediaries between data and users. They are, in effect, pre-packaged answers to questions that the reader might ask.

But as conversational AI systems built on LLM technology advance, they have the potential to allow users to directly query and interact with the underlying data in a more natural and conversational manner. Instead of relying on documents as static information sources, users can engage in dynamic conversations with AI systems to obtain real-time insights, ask specific questions, and receive personalized responses. And to some extent the future is already here: Microsoft's Copilot for Power BI is an excellent example of how LLMs can play this kind of role. Check out Microsoft's [demo video](#): the first part of this presents an example of report being produced in response to a prompt—so that's the disintermediation of the template—but the real novelty is demonstrated from around the 1-minute mark, where the report is interactively revised through conversation.

Realistically, of course, it's a bit premature to talk of the death of the document. Documents are likely to serve important purposes such as formal reports, legal contracts, and archival records for some time to come. But developments in AI and the new opportunities they present for information presentation may highlight the fact that documents are not always optimal as conveyors of information, thus reducing the attractiveness of document-centric solutions in many scenarios.

Of course, the elephant of hallucination remains in the room. Until we can be assured that—in a manner somewhat analogous to the case often made for self-driving cars—AI-generated texts are *less* likely to contain hallucinations than their human-scripted counterparts, the reliability moat will remain unbridged. But I wouldn't bet my business on that being the case for too much longer.

Acknowledgements

Many people graciously gave of their time to provide background for this article. A big shout-out to: Johannes Bubenzer (2TXT); Neil Burnett (Arria NLG); Joe Procopio (ex-Automated Insights); Robert Weißgraeber (AX Semantics); Mike White (ex-CoGenTex); Steven Wasick (InfoSentience); David Llorente, Alberto Moratilla, Javier García Ortiz (Narrativa); Marko Drotschmann, Jana Erhardt, Anastasia Linnik (Retresco); Henrik Gemoll (Textual.ai); Cecilia Campbell (United Robots); Vivek Mishra, Soham Save (vPhrase); and Thomas Alby, Emmanuel Walckenaer (Yseop).

If you'd like to keep up to date with what's happening in the commercial NLP world, consider subscribing to the free *This Week in NLP* newsletter at <https://www.language-technology.com/twin>.