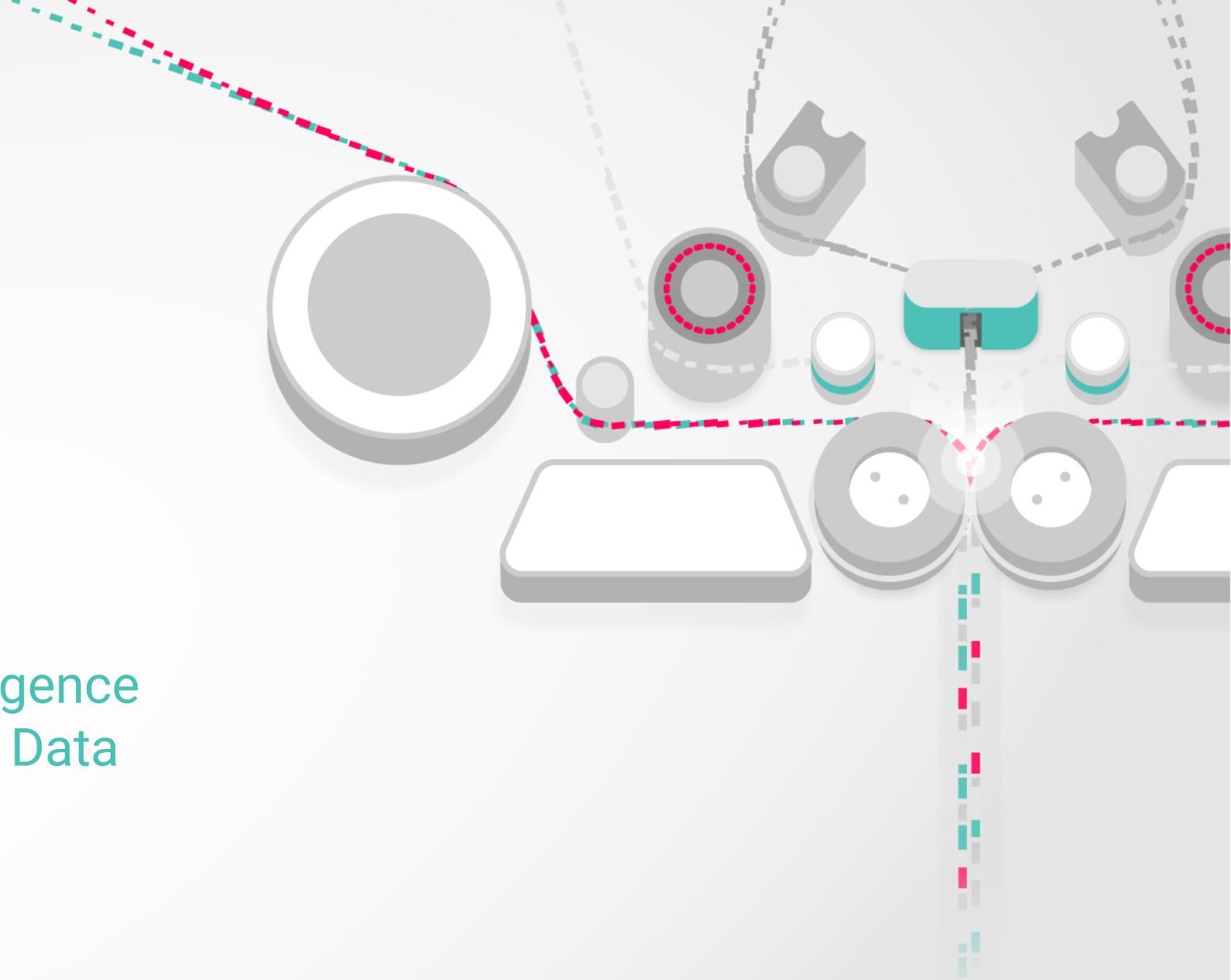




INTELLIGENCE POWERED DEFENSE

Narrator

Generating Finished Intelligence
Products From Structured Data



Who We Are

Sergey Polzunov

Software Engineer



Jörg Abraham

Chief Analyst



Agenda

- Setting the Scene
- From Structured Data to Narratives
- NLG in the Cyber Threat Intelligence Domain
- Narrator - A Proof-of-Concept
- Lessons Learned, Takeaways, Further Consideration

Setting The Scene

NLG Definition

Natural Language Generation (NLG), which is a subfield of artificial intelligence and computational linguistics [...] can produce meaningful texts in English or other human languages from some underlying non-linguistic representation of information.

“Building Natural Language Generation Systems” by Ehud Reiter and Robert Dale, 2000

Widespread Use of NLG

Real Time Traffic & Weather Forecast

Journalism

Business Intelligence

Property Bot

Sports Reporting

Financial Analysis and Reporting

NLG in Journalism

The Washington Post experiments with automated storytelling to help power 2016 Rio Olympics coverage

[The Washington Post, August 5, 2016, WashPostPR](#)

Press Association wins Google grant to run news service written by computers

[The Guardian, July 6 2017, Julia Gregory](#)

EY, Deloitte And PwC Embrace Artificial Intelligence For Tax And Accounting

[Forbes, November 14 2017, Adelyn Zhou](#)

NLG Powered News Generator for Automated Journalism

Cuomo ordains New Yorkers to wear face masks in public

18 hours ago - 📍 Rikers Island

    1.5K 

In crowded neighborhoods, there will be a requirement for face coverings, while the New York City authorities said they would create emergency food reserves and take other measures to ensure supply for residents.

The face mask rule will take effect only three days later, providing people with enough caution to buy or find one.

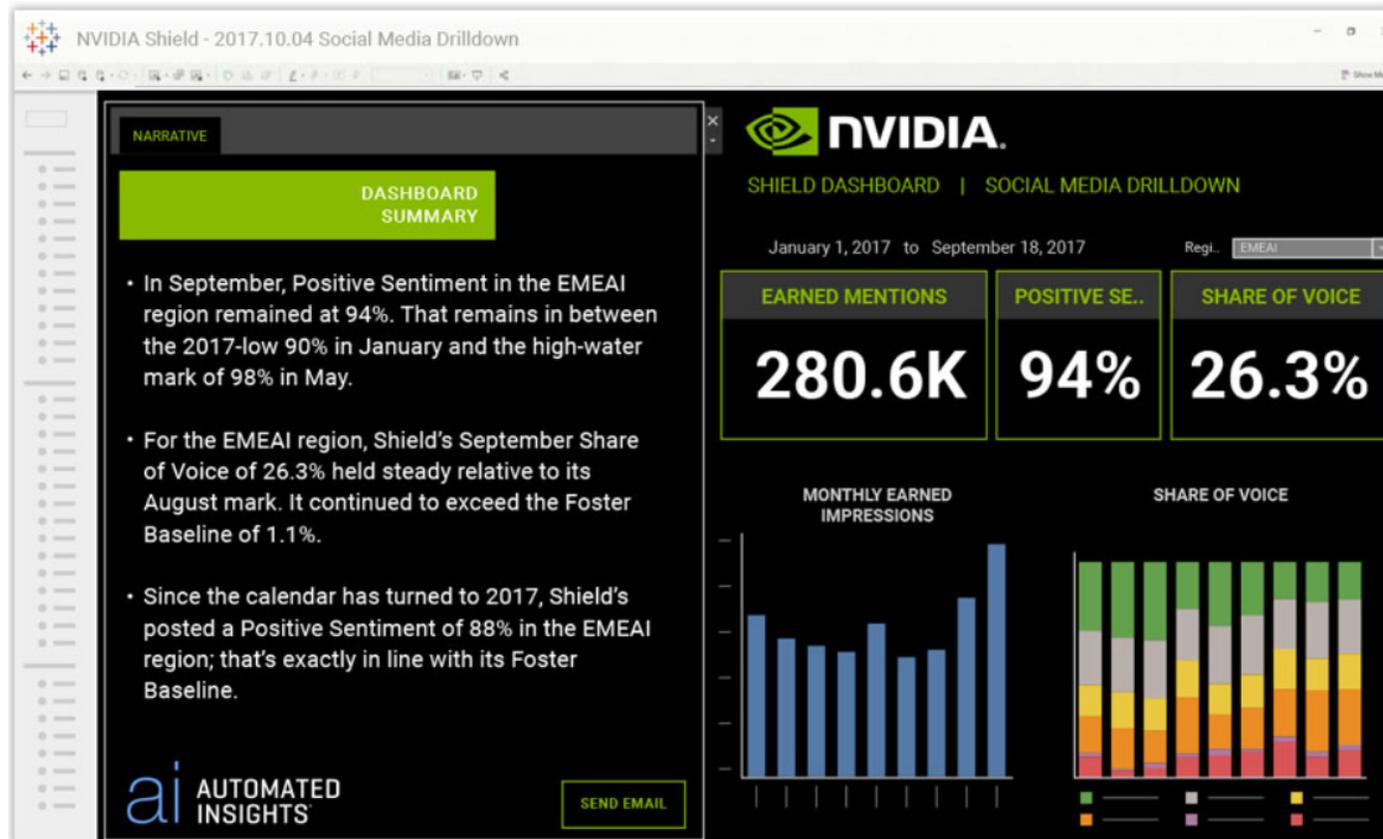
The new mandate requires that a mask or mask should be fitted on busy roads, on public transport, or in situations where people cannot preserve their social separaton.

De Blasio said the city does not just feed hungry New Yorkers but also 'people to work.'

So far, the city already distributes 250.000 free meals a day to schools and provides elderly people 25.000 meals per day.

[Monok.com, April 15 2020](#)

NVIDIA Leverages NLG to Augment Marketing Analytics



Automated Insights' Wordsmith platform integrated within Tableau

Databook - 5.8 M Executive Summaries Written Per Year

The screenshot displays the Databook interface for Bank of America. The top navigation bar includes the Databook logo, a search bar with 'bank of america' entered, and menu items for Opportunities, Companies, Topics (with a BETA badge), and Requests. The main content area is titled 'Bank of America' with a 'KEY ACCOUNT' badge and an 'Add to list' button. A left sidebar contains navigation options like Overview, Highlights, Core Insights, Competitors, Stock Price, News, Transcripts, Events, Value Drivers, Metrics, and Filings. The 'Company Description' section provides a brief overview of Bank of America Corporation and lists various business segments such as B2B, B2C, Finance, Insurance, Financial Services, Investment Banking, Banking & Mortgages, and Consulting & Professional Services. A stock price chart for 'BAC NYSE' shows a price of \$28.41 with a -7.0% (1-yr TSR) change, plotted from March 2018 to February 2019. The 'Executive Summary' section, highlighted with a blue border, contains three bullet points discussing shareholder returns, revenue growth, and future expectations. The summary is attributed to 'Quill'.

On page load, the executive summary is instantly written to reflect real-time data from the market.

Executive Summary

- Though Bank of America's total shareholder returns slipped over the past year by 7.0%, its performance topped declines in average returns across the Banks industry (14.0%). [More](#)
- Bank of America's revenues increased by \$4.0 billion in the year to December 2018. To match investor expectations Bank of America will have to maintain its trend and grow revenues by \$3.6 billion (4.1%) in each of the next three years. This will be challenging, despite Bank of America's annual growth of 1.3% over the past three years. [More](#)
- To meet investor expectations, Bank of America needs to maintain its current level of profitability, with a 1.8 percentage points change - or a total of \$8.0 billion - over the next three years (to December 2021). This is likely attainable, given that Bank of America's margins grew by 11.5 percentage points over the past three years. [More](#)
- Analysts only expect 13.5% of improvements to Bank of America's profitability to come from cost cutting, as boosting growth is a higher priority for the company. Bank of America, however, is expected to cut \$1.9 billion in cost over the next three years. This is achievable against a backdrop of improving cost efficiencies at the company in recent years. [More](#)

Written by **Quill**

<https://narrativescience.com/wp-content/uploads/2019/05/Databook-sample-BoA.png>

Mastercard - Customer Habits & Buying Patterns



Customer Report

MONTHLY SNAPSHOT
XYZ Merchant Location



- Your spend per customer (purchase size multiplied by visit per customer) is \$52.27, less than the benchmark average. If you increased your prices slightly, by \$0.75 across the board, your spend per customer would be in-line with the benchmark.
- Your restaurant attracts a lunch crowd during the work-week that is primarily composed of middle-aged affluent men. This differs from other local category businesses with more dinner traffic. This may be a key driver of your lower average purchase amount. You might consider additional promotional activity for lunch customers to come back for happy hour or dinner.
- Also, you may consider promotional activity for business events or services given your location appears to be a good workday option for customers.

Customer Spending



“This may be a key driver of your lower average purchase amount. You might consider additional promotional activity for lunch customers to come back for happy hour or dinner.”

Written by **Quill**

<https://narrativescience.com/wp-content/uploads/2019/05/Mastercard-customer-report.png>

From Structured Data to Narratives

From Structured Data to Narratives



Document Planner

- **Context determination** - Decide what information should be communicated out of the full dataset available
- **Document structuring** - How the information should be structured

→ Produces a *document plan* object

Microplanner

- **Lexicalization** - Chose the words that will represent concepts
- **Referring expressions generation** - chose proper names, pronouns and references
- **Aggregation** - group information that should be expressed in one lexical block (phrase, paragraph, section)

→ Produces *a text specification* object

Surface Realiser

- Generating final text blocks from a text specification tree, produced by microplanner

→ Produces *final text*

NLG Systems & Services

- Association for Computational Linguistics (ACL)
https://aclweb.org/aclwiki/Downloadable_NLG_systems
- Multiple commercial NLG services for various domains

NLG in the Cyber Threat Intelligence Domain

NLG in the Cyber Threat Intelligence Domain

- Automated email Generation for Targeted Attacks using Natural Language

Avisha Das, Rakesh Verma

Department of Computer Science University of Houston, Houston, Texas

NLG in the CTI Domain - The Challenge

Millions of
structured
information

Finished intel
product is often
a written report

How NLG Can Help Intelligence Analysts?

- Free up analyst's time by automating what can be automated
- Expanded coverage
- Increases capability for investigations by looking at big data sets
- Consistency & conformance to standards

Augmentation rather than Replacement

Narrator - A Proof-of-Concept

NLG Applied to CTI - PoC Constraints

- Use STIX2 bundles as input data format.
- Require high quality STIX2 with necessary relations / properties set.
- Create draft report, which must then be further edited by an analyst.
- Focus on producing routine factual sections of a document which human authors often find monotonous to write.

Lessons Learned, Pitfalls and Takeaways

Lessons Learned

- Strict criteria on quality of the produced content required (style, structure, words used)
 - Limits the use of artificial neural networks (unpredictability)
 - Post-generation moderation is required
- STIX2 partially not fit for purpose, due to limited object properties
- Additional sanitization required (due to analysis / information was missing in structured data)

Takeaway

- NLG requires a matching use case.
- Start with realistic goals.
- Data must be structured enough.
- Requires significant SME and engineering investment.
- Leave neural networks for later, and use them for text style transfer, synonyms selection, data importance estimation, etc.

Further Consideration

- Algorithmic Transparency
- Algorithmic Bias
- Balancing Artificial and Human Intelligence
- Additional controls for sensitive information?

Thank You

Questions? - jorg@eclecticiq.com / sergey@eclecticiq.com



INTELLIGENCE POWERED DEFENSE