March 26, 2021

Editorial Algorithms

The next AI frontier in News and Information

National Press Club
Journalism Institute

Francesco Marconi

Housekeeping

- You will receive slides after the presentation
- Interactive polls throughout the session
- Use QA tool on Zoom to ask questions
- If you still have questions, reach out to Francesco directly **francesco@appliedxl.com** or connect on twitter **@fpmarconi**

What will your learn during today's session

1. **How AI is being used** and will be used by journalists and communicators, and what that means for how we work
2. **The ethics of AI**, particularly related to diversity, equity and representation
3. The **role of AI in the spread of misinformation** and 'fake news'
4. How journalists and communications professionals can **avoid pitfalls while taking advantage of the possibilities AI** provides to develop new ways of telling stories and connecting with audiences

Poll time!

What's your overall feeling about AI and news?

A. It's the future of news and communications!
B. I'm cautiously optimistic
C. I'm cautiously pessimistic
D. It's bad news for journalism

Journalism was designed to solve information scarcity

Newspapers were established to provide increasingly literate audiences with information. Journalism was originally designed to solve the problem that **information was not widely available.**

The news industry went through a process of standardization so it could reliably source and produce information more efficiently. This led to the development of very structured ways of writing such as the "inverted pyramid" method, to organizing workflows around beats and to plan coverage based on news cycles.
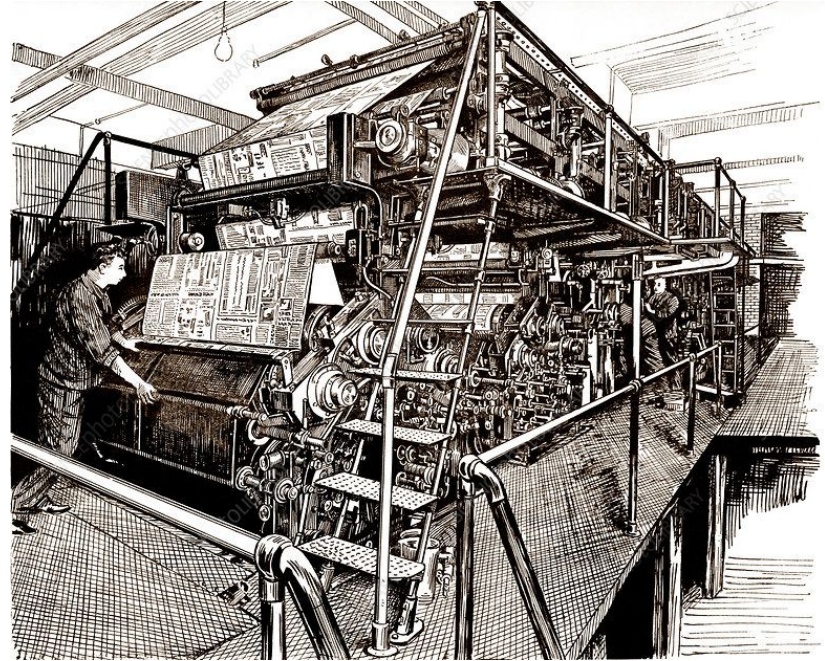
A new reality: information abundance

The flood of information that humanity is now exposed to presents a new challenge that cannot be solved exclusively with the traditional methods of journalism.

The explosion of data from the web, sensors, mobile devices and satellites has moved us to an environment where there's too much information.

We retain and produce more information now than at any previous point in human history, it takes much more effort to delete, remove and filter unwanted information than to accumulate it. **The modern role of journalism is to filter and contextualize.**

The rise of editorial algorithms

The enormous amount of data that our society is producing is pushing us towards a new type of journalism, computational journalism, which aspires to **source and contextualize information at scale.**

The role of computational journalists is no longer to write and produce data-driven stories, but to write editorial algorithms. These systems are coded with journalistic principles and work nonstop to filter irrelevant information and find newsworthy events.

Editorial   Algorithms

AI is creating news jobs and responsibilities in news media

Many roles will evolve into leveraging new technology as part of their daily workflow. But there are also a number of new roles as a result of AI entering the newsroom:

→**Automation editors:** planning how workflows can be augmented through AI and ensuring its editorial reliability

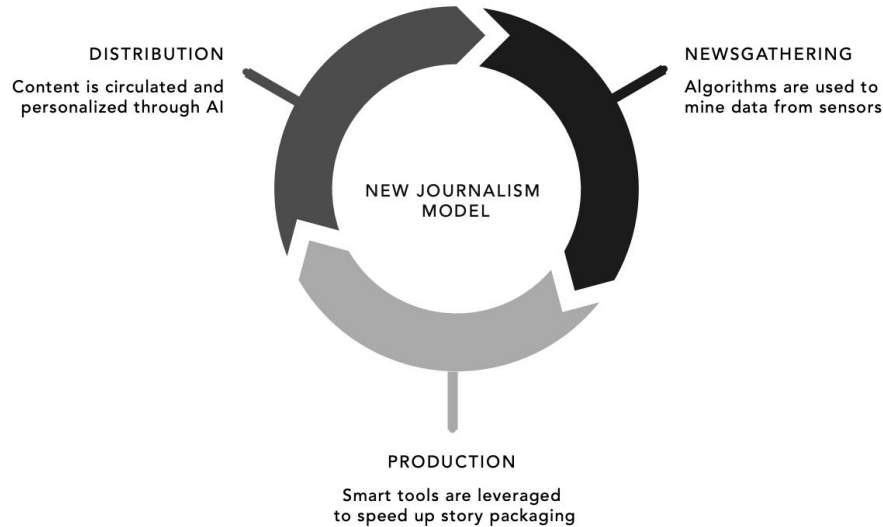→**Computational journalist:** applying computer and data science methods to develop editorial algorithms

→**Newsroom tool manager:** coordinating the implementation of new platforms and training journalists how to deploy them

→**AI ethics editor:** ensuring transparency and explainability of algorithms as well as its use of training data

AI is dramatically changing journalism

The modern news workflow is dynamic, with each step of the process being augmented by AI.

DISTRIBUTION
Content is circulated and personalized through AI

NEWSGATHERING
Algorithms are used to mine data from sensors

NEW JOURNALISM MODEL

PRODUCTION
Smart tools are leveraged to speed up story packaging

AI in **news gathering**

Event detection in social media
Case study: Reuters

Reuters' platform **News Tracer** helps reporters **monitor emerging topics on social media** and find relevant stories more quickly.

In 2015, it revealed social media activity documenting a shooting in San Bernardino, California, **before any other news organization**.

Other social media detection tools:

Dataminr **NEWSWHIP** **samdesk**

Text mining
Case study: ICIJ

The International Consortium of Investigative Journalists uses an AI-powered tool to **automatically recognize and index text documents by recognizing entities** (people, places, dates etc.) within content.

This was used to make sense of 13.4 million confidential documents relating to offshore investments for the series—"Paradise Papers: Secrets of the Global Elite."

Other text mining tools:

VOYANT
see through your text

Gathering alternative signals
Case study: Quartz

Journalists at Quartz had a computer watch a televised debate between Hillary Clinton and Donald Trump.

It **recognized dominant emotions** measured by facial expressions for each candidate.

These systems are fallible and **must be audited!**

Surprise is not always suprise

Raised eyebrows are an indicator of surprise but in this frame, Clinton is actually talking about the economy and how her father ran a small business
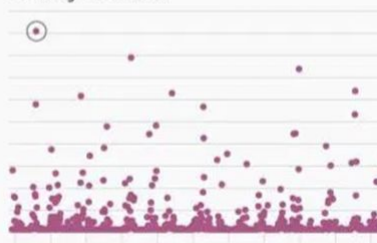
Here Trump says his support for the Iraq war was mischaracterized; asked in an interview with Howard Stern he said "very lightly, 'I don't know.'"

Hillary Clinton

Donald Trump

Quartz | qz.com
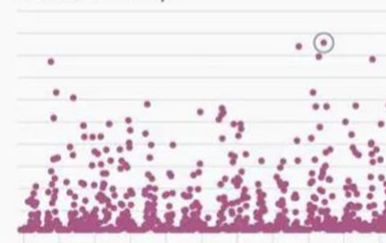
Data compiled by Debate in (E)motion

Algorithms used in journalism must be audited

The same way journalists ask questions to human sources, they should be able to ask questions to algorithms, to understand their inner workings. This allows for the creation of reliable editorial algorithms.

- **CATEGORY:** What does the algorithm do (filtering, prediction, ranking, calculation, etc.)?

- **GOAL:** What is the algorithm optimizing for (e.g., maximizing time spent on site)?

- **DATA BASIS:** What data is the algorithm based on and is there any obvious bias in it?

- **TRANSPARENCY:** Is it clear and communicated to users how the algorithm makes decisions?

- **HUMAN OVERRIDE:** Is there oversight by humans able to quickly make decisions and tweak the algorithm?

- **EXPLAINABILITY:** Is the output of the algorithm explainable/interpretable?

- **DETECTED ERRORS:** Are there reported instances of mistakes the algorithm made?

- **FAIRNESS:** Are certain groups (dis)advantaged by this algorithm?

- **PRIVACY:** Is user data stored or shared with other users or third parties (e.g., advertisers, government)?

- **ROBUSTNESS:** Was the service checked for robustness against adversarial attacks and hacking?

AI in **news production**

News automation
Case study: AP

AP uses **natural language generation** to automate part of its financial and sports coverage.

In business news, the newswire went from covering 300 companies with human writers to covering over 4,400 companies with the help of smart machines.

Northwestern looks to end streak vs Ohio State

AP Associated Press

Posted
2/19/2019 7:00 AM

Northwestern (12-13, 3-11) vs. Ohio State (16-9, 6-8)

Value City Arena, Columbus, Ohio; Wednesday, 8:30 p.m. EST

BOTTOM LINE: Ohio State looks to extend Northwestern's conference losing streak to seven games. Northwestern's last Big Ten win came against the Indiana Hoosiers 73-66 on Jan. 22. Ohio State lost 62-44 to Michigan State on Sunday.

SUPER SENIORS: Northwestern has relied heavily on its seniors. Vic Law, Dererk Pardon, Ryan Taylor and A.J. Turner have combined to account for 71 percent of the team's scoring this year and 84 percent of all Wildcats points over the team's last five games.

LIKEABLE LAW: Law has connected on 32.8 percent of the 137 3-pointers he's attempted and has made 8 of 23 over the last three games. He's also made 80.8 percent of his free throws this season.

SCORING THRESHOLDS: Northwestern is 0-9 when its offense scores 60 points or fewer. Ohio State is a perfect 9-0 when it holds opponents to 61 or fewer points. The Buckeyes have allowed 59 points per game over their last three.

Text automation tools:

ai AUTOMATED INSIGHTS   Narrativa   narrative science

News automation requires humans in the loop

Human journalists have a crucial role at every step of data-to-text automation process providing important **editorial oversight.**

Data journalists identify, verify, and validate data sources.

Data journalists format and load data into a structured table.

Automated stories are sent to the CMS for final review by editor.

Automation editors feed data into a story template engine.
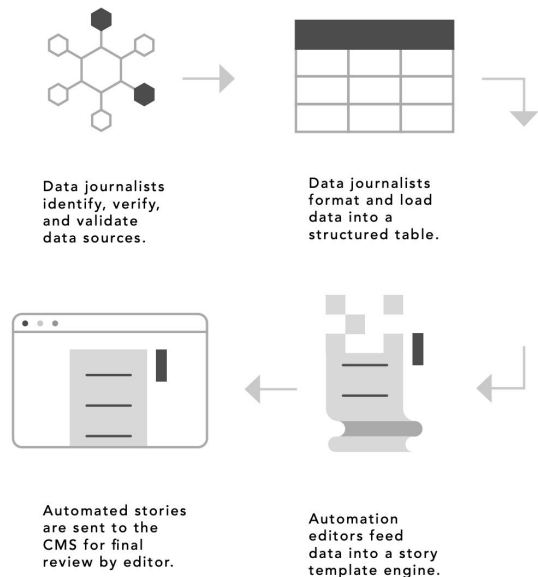
Automated summaries
Case study: Bloomberg

Bloomberg launched **The Bulletin**, a feature on its mobile app powered by **machine-generated summaries** that provide readers a "sense of completion in **quickly learning the latest news** of the moment, and a comprehensive summary of the news that goes beyond a headline."

Summarization tools:

agolo   **QuillBot**

The Bulletin   ● UPDATED 9:41 AM

As they battle the rise of e-commerce, **U.S.** mall owners are trying to clear their books of fading centers to focus on the most-profitable ones.

Deals   10 hours ago

Market Moving
Elon Musk is cutting 9 percent of the workforce at **Tesla Inc.,** a turnabout that underscores the mounting pressure he is under to show the electric-car maker can one day turn a profit.

Tesla Inc ▲ +2.77%   an hour ago

U.S. inflation accelerated in May to the fastest pace in more than six years, reinforcing the **Federal Reserve**'s outlook for gradual interest-rate hikes while eroding wage gains that remain relatively tepid despite an 18-year low in unemployment.

Markets   6 hours ago

↺

3 FRESH STORIES

Home   Markets   Watchlist   Media   Menu

Beware! Fake text

Recent advancements in natural language enable the creation of fake text at scale.

The Journalism Institute at the Press National Club **in the United States has recently published a report on the "dramatic failures of the liberal media as a means of getting the message out". The report, which is part of** a research project focused on **the use of the media by liberal leaning Americans, points to the recent efforts of** organizations to **shift political and social views**.

FAKE

Written by Transformer · transformer.huggingface.co 🦄

Video footage labelling
Case study: Sky News

Sky News uses AI for **facial recognition, subtitling automation and datasets.**

An example is the piece "Who's who": In 2018, Sky News used AI facial recognition technology to spot and identify the guests arriving at the UK Royal Wedding.

Other image recognition tools:
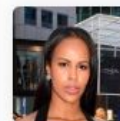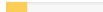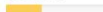
Beware: algorithms make mistakes!

An analysis of computer vision algorithms showed an error rate of 35% for darker-skinned women when compared to just a 1% error rate for lighter-skinned men.

Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%

Organizations focused on AI transparency:
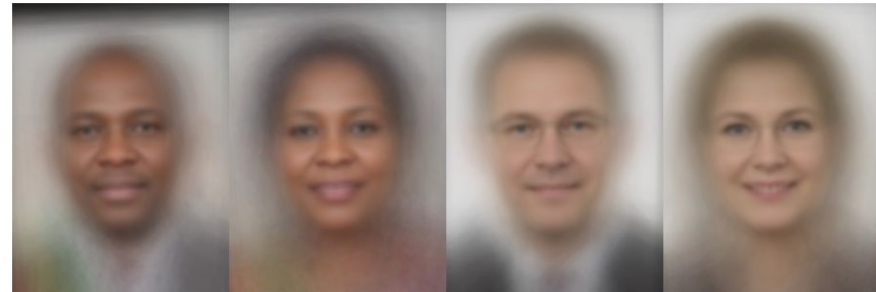
Image credit: Joy Buolamwini — Mit Lab Press Kit

Diversity and representation in AI is crucial!

Data from the Bureau of Labor Statistics shows that the professionals who write AI programs are still **largely white males**. And a study conducted last August by Wired and Element AI found that **only 12% of leading machine learning researchers are women**.

People that first notice bias issues are mostly users who are from that specific minority community. **Fostering a diverse AI team** can help mitigate unwanted AI biases and ensure fairness.

A checklist for addressing AI bias:

✓ Establishing processes and implement regular audits to check for biases on datasets and algorithms.

✓ Determining how to include humans-in-the-loop throughout the AI system.

✓ Implementing data & AI transparency standards the same way we have editorial standards.

✓ Having a clear strategy for bringing diverse perspectives to AI teams through training, hiring and partnerships.

Synthetic video
Case study: BBC

BBC newsreader Matthew Amroliwala only speaks English, but by using artificial intelligence software by Synthesia he appears to be speaking Spanish, Mandarin and Hindi.

Tools for automated video generation:

Synthesia

wibbitz

Synthetic news anchors
Case study: Xinhua

Chinese news agency launched fully digital news anchors that can deliver a video broadcast in different styles for a more personalized experience.

Synthetic media can be used for good but also for bad...

Beware: Deep fakes!

Deep fakes are videos, images, text or audio files generated or altered with the help of artificial intelligence to deceive an audience into thinking they are real.

The biggest challenge is how widely accessible this technology is, which means anyone can now create deepfakes.

Deepfake detection tools:
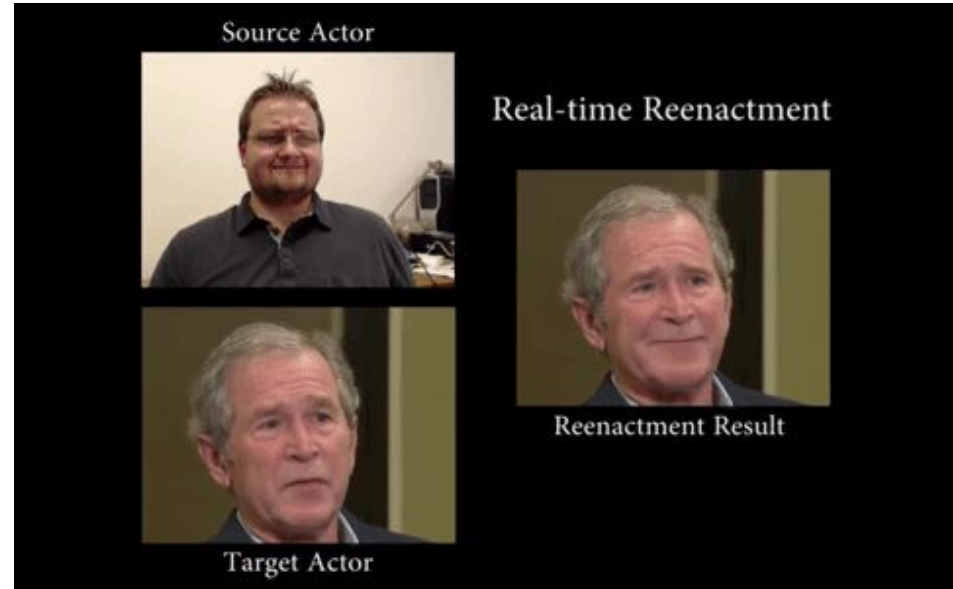
Approaches for detecting deep fakes

Lack of eye blinking
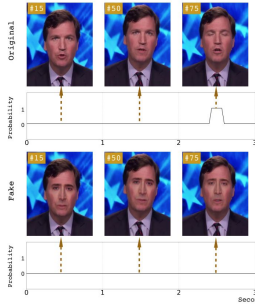
An algorithm developed by researchers at SUNY is able to tell if a person blinks; people shown in deepfakes often don't close their eyes.

Pulse signal

Researchers at MIT developed a technique to detect the pulse of a person by color amplification, which is often missing in deepfakes.

Image blurriness

If you look at still images of a deepfake you can see things like the flickering and blurriness around the mouth or face, or discrepancies between the face, body type and skin color.

Machine learning

Algorithms can be trained to detect fake imagery and provide a confidence score on whether the footage is real or fake.

Poll time!

Can AI help us better navigate the current information environment?

A. AI can help but it's not the full solution
B. Too early to tell
C. Only humans can address the challenges of information abundance

AI is changing **news distribution**

Personalized news
The Times

It has an AI-powered tool called **JAMES**, which stands for Journey Automated Messaging for Higher Engagement through Self-Learning.

It acts as a digital butler, using data to get to know the **habits, interests, and preferences** of readers.

journalism.co.uk

The Times employs an AI-powered 'digital butler' JAMES to serve personalised news

The publisher uses machine learning to grow loyal audience by delivering editorial content in the right format, time and place

Posted: 24 May 2019 By: Marcela Kunova

Credit: Photo by Tyler Nix on Unsplash

Beware: information bubbles!

With AI personalization, there is a risk of creating information bubbles if readers are only shown the views of the world they already agree with.
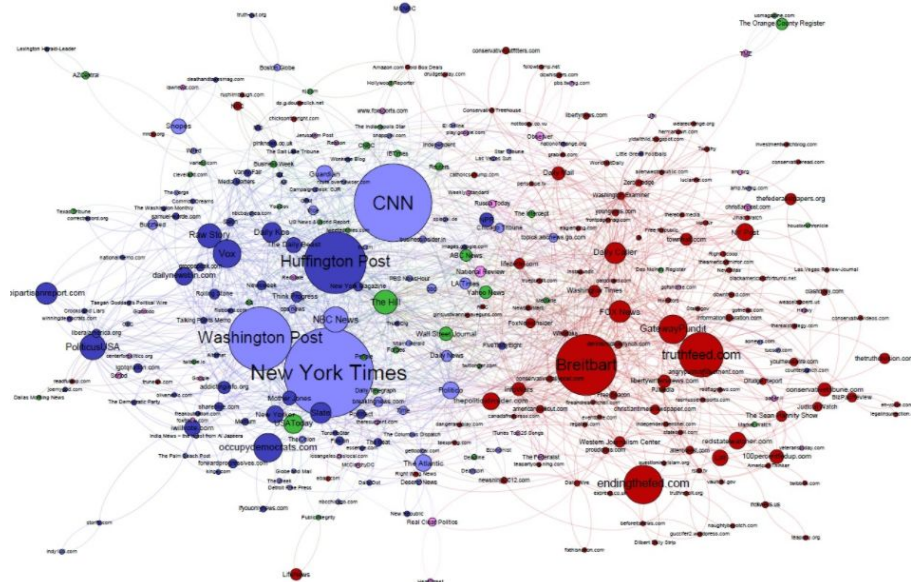
Navigating information bubbles

Tools to help understand partisan media consumption and social media influence.

Comment moderation
Case study: The New York Times

NYT uses machine learning to make it easier to moderate comments.

The system Perspective finds patterns in data to spot abusive language or online harassment, and it scores comments based on the perceived impact they might have on a conversation.

AI models are trained through human input, which is prone to bias. Algorithms should be vetted with the same journalistic standards used to evaluate a story.

Audience moderation tools:

Perspective   HIVE MODERATION

💬 **2672 COMMENTS**

Readers shared their thoughts on this article.

The comments section is closed. To submit a letter to the editor for publication, write to letters@nytimes.com.

| **All** 2672 | **Readers' Picks** 1685 | **NYT Picks** 33 |

Sean C. Charlottetown • February 13, 2016

Truly objective legal interpretation is mostly a myth (personal preferences always inform decisions), but Scalia was atypical in his willingness to disregard any legal principle he had previously enunciated if it conflicted with his own partisan agenda, and uniquely rude and offensive in the language he used to do so.

My condolences to his loved ones, but his presence on the Court will not be missed.

👍 3619 Recommend • 📘 🐦   🚩 Flag

Growing subscriptions
The Wall Street Journal

The Wall Street Journal's paywall is powered by a machine-learning algorithm that **measures reader activity**.

It gives a unique subscription probability to each reader, which helps inform how many sample stories users can access.

Implementing AI in the newsroom

How are AI newsroom applications built?

The development of artificial intelligence requires constant validation of both the data sources and the algorithms used in that tool.
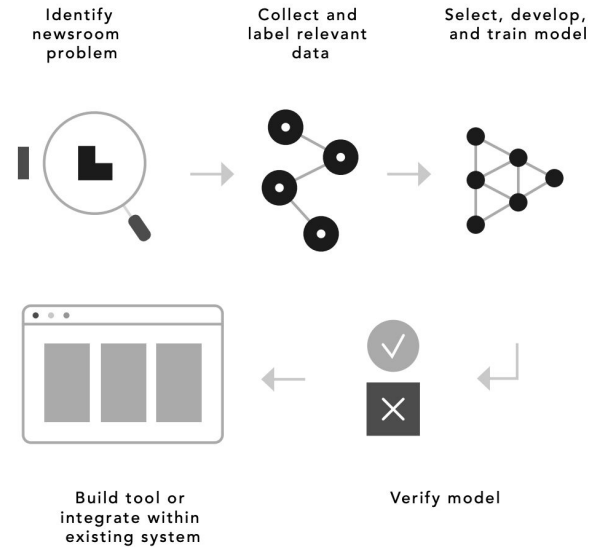
Identify newsroom problem

Collect and label relevant data

Select, develop, and train model

Build tool or integrate within existing system

Verify model

Incremental development
Case study: WSJ

Talk 2020 is a **text analysis platform and search tool** that allows the newsroom to access 30 years of public statements made by 14 presidential candidates.

It **pulls transcripts** from speeches, rallies, interviews and press conferences. Initially started as an internal tool, it was later made available to subscribers.

Incremental development
Case study: WSJ

Talk 2020 is a **text analysis platform and search tool** that allows the newsroom to access to 30 years of public statements made by 14 presidential candidates.

It **pulls transcripts** from speeches, rallies, interviews and press conferences and uses Factiva as its primary data source.

Source: WSJ

Democratic Debates: What the Presidential Candidates Are (and Aren't) Saying

Here's a look at some of the top topics in the past three Democratic debates, and a few topics candidates barely have touched on

Incremental development
Case study: WSJ

Talk 2020 is a **text analysis platform and search tool** that allows the newsroom to access to 30 years of public statements made by 14 presidential candidates.
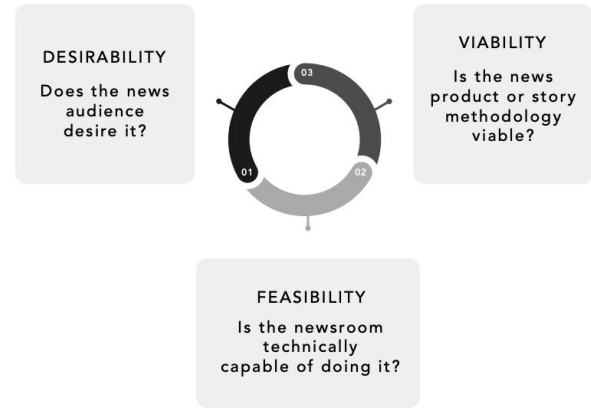
It **pulls transcripts** from speeches, rallies, interviews and press conferences. Initially started as an internal tool and later made available to subscribers.

Source: WSJ
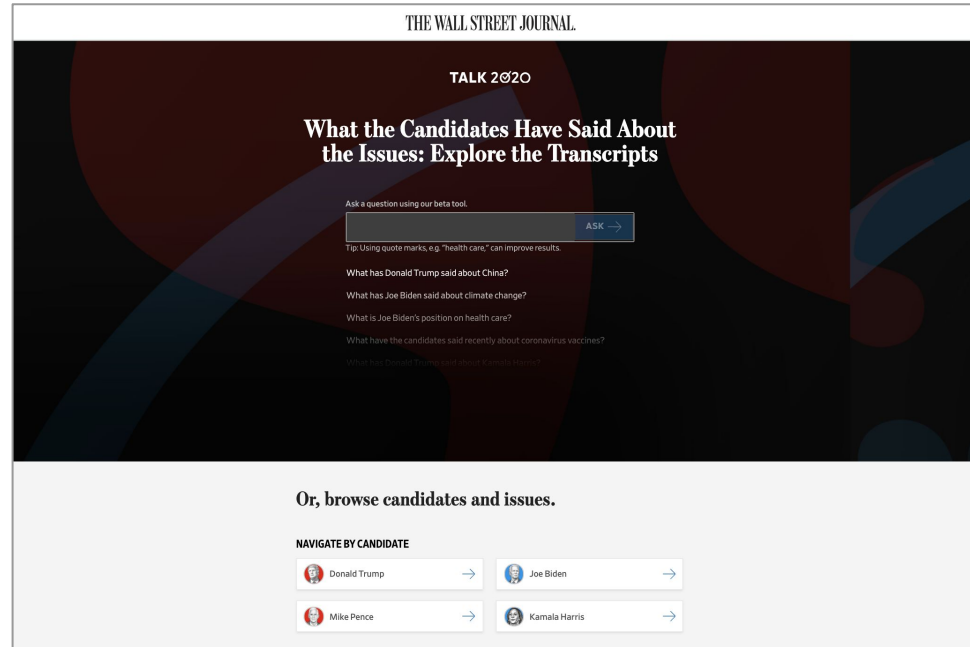
Newsrooms with a product mentality will succeed

Iterative journalism borrows from design thinking methodologies, starting with the audience's needs (desirability), then answering questions related to feasibility of execution, and finally addressing its long term viability.

DESIRABILITY

Does the news audience desire it?

VIABILITY

Is the news product or story methodology viable?

FEASIBILITY

Is the newsroom technically capable of doing it?

Important takeaways

Technology changes, **journalistic standards don't.**

AI is susceptible to the **same biases as humans.**

Journalists can **best leverage AI once they experiment** with the technology.

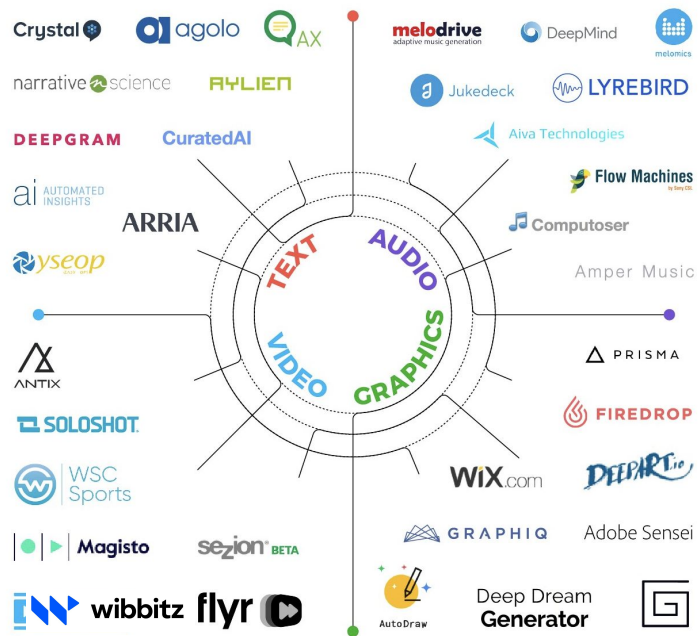There are **ethical considerations** inherent in journalism's use of AI.

Resources

Newsmakers: Artificial Intelligence and the Future of Journalism

Francesco Marconi, 2020

COLUMBIA UNIVERSITY PRESS

More tools

AI COMPANIES
FOR CONTENT CREATORS

Poll time!

After learning about how AI is currently being used, what's your overall feeling about AI and News?

A. It's the future of news and communications!
B. I'm cautiously optimistic
C. I'm cautiously pessimistic
D. It's bad news for journalism

Thank you