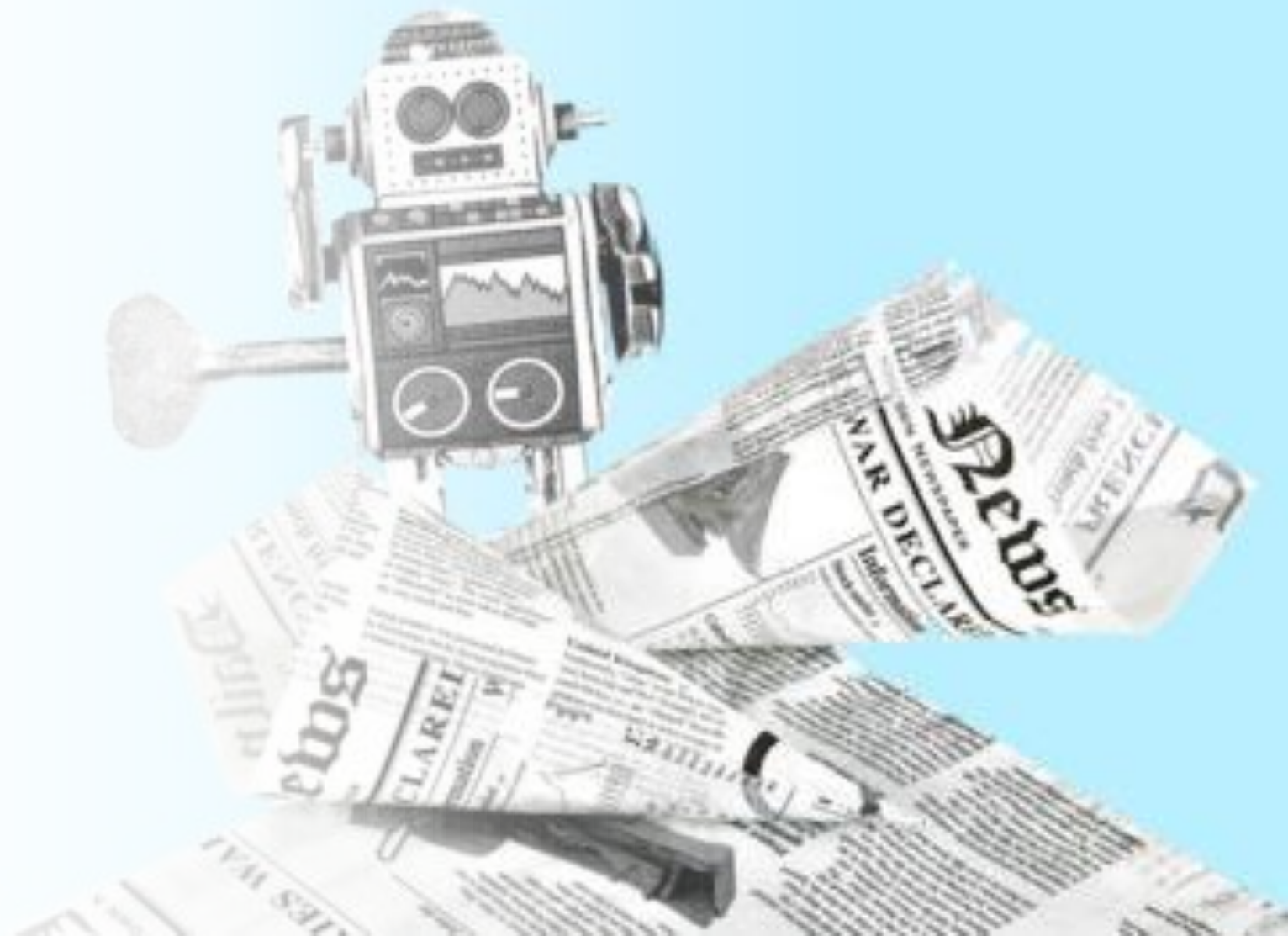




Dziennikarstwo oparte na AI

Andrzej Ustupski, Bartosz Ziomek,
Daniel Terkała, Dawid Jaształ,
Grzegorz Puczkowski, Marcei Wójciuk



Robojournalism

W 2014 roku Los Angeles Times opublikował raport o trzęsieniu ziemi kilka minut po jego wystąpieniu.

Wyczyn ten był możliwy, ponieważ pracownik opracował bota do pisania automatycznych artykułów na podstawie danych generowanych przez US Geological Survey.

Dziś sztuczne inteligencje piszą setki tysięcy artykułów, które są cyklicznie publikowane przez media głównego nurtu.

Los Angeles Times



Earthquake aftershock: 2.7 quake strikes near Westwood

BY KEN SCHWENCKE
MARCH 17, 2014 7:53 AM PT

A shallow magnitude 2.7 earthquake aftershock was reported Monday morning four miles from Westwood, according to the U.S. Geological Survey. The temblor occurred at 7:23 a.m. Pacific time at a depth of 4.3 miles.

A magnitude 4.4 earthquake was reported at 6.25 a.m. and was felt over a large swath of Southern California.

According to the USGS, the epicenter of the aftershock was five miles from Beverly Hills, six miles from Santa Monica and six miles from West Hollywood.

[FULL COVERAGE: Southern California earthquakes](#)

In the last 10 days, there has been one earthquake of magnitude 3.0 or greater centered nearby.

This information comes from the USGS Earthquake Notification Service and this post was created by an algorithm written by the author.

Read more about [Southern California earthquakes](#).

SUBSCRIBERS ARE READING

FOOD FOR SUBSCRIBERS

These are the 101 best restaurants in L.A.

CALIFORNIA

Should I change my holiday plans because of COVID-19 surge, Omicron? What experts say

CALIFORNIA

Drugs, texts and left for dead: Untangling the final hours of an L.A. model and architect

FOOD

Where we love to take out-of-town guests to eat in L.A.

LIFESTYLE

FOR SUBSCRIBERS

Lonely in L.A.? These 21 places and experiences will help you embrace it.

LATEST CALIFORNIA >

CALIFORNIA

1,300 gallons of fuel spill from gas pump in Pasadena and into Alhambra Wash

Dec. 19, 2021

Robojournalism

Początkowo większość narzędzi Natural Language Generation (NLG) tworzących te artykuły była dostarczana przez firmy programistyczne, takie jak Narrative Science. Obecnie wiele organizacji medialnych opracowało swoje własne wersje.

Systemy te zaczynają się od danych – wykresów, tabel i arkuszy kalkulacyjnych. Analizują je, aby wydobyć konkretne fakty, które mogłyby stanowić podstawę narracji. Generują plan artykułu, a na koniec tworzą zdania za pomocą oprogramowania do generowania języka naturalnego.



sciforce

NLG Applications in Theoretical Research



Understanding the human language production facility



Planning extended monologic discourses



Psycholinguistic modeling



Composing target language texts in machine translation systems

Zalety

- **Szybkość**

"Roboty dziennikarskie" są stworzone do generowania dużych ilości informacji w szybkim tempie. Associated Press stwierdził, że dzięki automatyzacji agencja prasowa uwolniła 20 procent czasu dziennikarzy, aby mogli skupić się na ważniejszych projektach

- **Koszt**

Dziennikarstwo automatyczne jest tańsze, ponieważ więcej treści można wyprodukować w krótszym czasie. Obniża również koszty pracy dla organizacji prasowych. Mniejszy wkład ludzki oznacza mniejsze wydatki na pensje



Wady

- **Autorstwo**

Często pojawia się niejasność co do tego, kto powinien być uznawany za autora. Kilku uczestników badania nad autorstwem algorytmicznym przypisało to programiście, inni postrzegali organizację jako autora'

- **Zatrudnienie**

Wśród obaw związanych z automatyzacją jest utrata zatrudnienia przez dziennikarzy, ponieważ wydawcy przestawiają się na stosowanie sztucznej inteligencji.

- **Wiarygodność i jakość**

Obawy dotyczące postrzeganej wiarygodności wiadomości automatycznych są podobne do obaw związanych z postrzeganą wiarygodnością wiadomości w ogóle. Krytycy wątpią, czy algorytmy są „uczciwe i dokładne, wolne od subiektywności, błędów lub prób wywierania wpływu”.

Wykorzystanie dziennikarstwa automatycznego

Aktualnie systemy mogą generować tylko artykuły, w których jako dane wejściowe dostępne są wysoce ustrukturyzowane, takie jak wideo z meczu piłki nożnej lub dane arkusza kalkulacyjnego z rocznego zwrotu firmy. Nie potrafią (jeszcze) pisać artykułów z polotem, wyobraźnią lub dogłębną analizą.

Typowe tematy obejmują podsumowania sportowe, pogodę, raporty finansowe, analizy nieruchomości i przeglądy zarobków.

Associated Press zaczęła wykorzystywać automatyzację do obsługi 10 000 meczów mniejszych lig baseballowych rocznie, korzystając z programu Automated Insights i statystyk z MLB Advanced Media.



Przykład artykułu generowanego podczas meczu krykieta ligi IPL

Komentarz w czasie rzeczywistym jest trudny dla dziennikarzy i analityków danych. Tutaj z pomocą przychodzą narzędzia NLG.

Jak widać na powyższym zdjęciu, produkt pisze nagłówek, podsumowanie a następnie szczegółowo przedstawia przebieg zdarzeń, a wszystko to bez konieczności nadzoru człowieka.

Taka automatyzacja może szybko analizować dane i generować spostrzeżenia oraz prezentować je w formie materiałów eksploatacyjnych w czasie rzeczywistym.

Home / Cricket / IPL 2020: SRH Vs KKR- SRH innings, death overs highlights

IPL 2020: SRH Vs KKR- SRH innings, death overs highlights

The big clash between SRH and KKR of IPL 2020 is being played at Abu Dhabi. The toss went in favour of Sun Risers Hyderabad (SRH), who decided to bat first. Here are the updates from the match.

CRICKET Updated: Sep 26, 2020 21:26 IST



5% Cashback
ON EVERYTHING
(Travel, Food, Bills, Tickets and Shopping)

dhani super saver card

SUBSCRIBE NOW

Try FREE for first 30 days
CANCEL ANYTIME

HT Neighbour's Kitchen- A web series about friendship and food

trending topics

india cities opinion world [ipl](#) entertainment biharelection trending videos tech podcasts web stories lifestyle #AIZLife



At the end of the first innings of the 8th match of IPL 2020, Sun Risers Hyderabad has managed to get to a total of 142 runs for the loss of 4 wickets. In the last five overs, Sun Risers Hyderabad managed to score 43 runs for the loss of 2 wickets. Manish Pandey was the highest scorer with 51 runs. The highest partnership of the innings was between Wriddhiman Saha and Manish Pandey who contributed 62 runs to the innings.

5% Cashback
ON EVERYTHING
(Travel, Food, Bills, Tickets and Shopping)

dhani super saver card

SUBSCRIBE NOW

Try FREE for first 30 days
CANCEL ANYTIME

Varun Chakravarthy bowled an expensive 16th over of the innings where the SRH batsmen hit a six scoring 11 runs from the over.

8 runs came off the 17th over of the innings bowled by Pat Cummins. SRH batsmen hit a four.

Andre Russell bowled an expensive 18th over of the innings where 9 runs came from the over along with a wicket.

The 19th over of the innings was bowled by Shivam Mavi who kept things tight, conceding just 6 runs.

9 runs and a wicket came off the 20th over of the innings bowled by Andre Russell.

Blog notrealnews.net



VERIFIED

A Global Pandemic Will Force These Industry Sectors To Strike Smart

As the world is gearing up for the coronavirus pandemic, cities are vowing to not let the virus move in to prevent the worst. During the Great Influenza Pandemic of



WORDFLOW AI CONTENT
11 APR 2020 • 2 MIN READ



VERIFIED

The vast financial community eager to invest in biotechnology today has to deal with obstacles too

If you're interested in something that comes with the title "hardtech," you'll find many of the parties involved in tonight's Large Event in San Francisco that are



WORDFLOW AI CONTENT
11 APR 2020 • 2 MIN READ



VERIFIED

If you think financial crises are bad, you shouldn't treat pandemics like they're in the clear

In late 2015, just after the H1N1 pandemic hit home, the Norwegian government formed a task force called the Reference Center on Emerging Diseases (KSF) to develop a post-crisis plan.



WORDFLOW AI CONTENT
11 APR 2020 • 3 MIN READ

Blog wykorzystuje sztuczną inteligencję do wypełniania tego, co przypomina stronę główną serwisu informacyjnego, wraz z fałszywymi wiadomościami napisanymi przez sztuczną inteligencję.

Strona internetowa, projekt firmy Big Bird zajmującej się rozwojem sztucznej inteligencji, ma być pokazem, w jaki sposób algorytmy firmy mogą pomóc dziennikarzom w szybkim pisaniu przekonujących wiadomości.

Kiedy decydują algorytmy

Ciekawą kwestią jest to, że artykuły mogą zostać dostosowane do konkretnych niszowych odbiorców, a docelowo dla każdego z nas indywidualnie. Chociażby analizując nasz ruch sieciowy.

Przykładowo oświadczenie organizacji badawczej, że **prawidłowe napompowanie opony samochodowej może zmniejszyć wydatki na benzynę o 7%**, może być dostosowane do konkretnego samochodu, liczby przejechanych mil tygodniowo, a nawet stylu jazdy, zakładając, że konsument udostępnia te informacje.

"The Daily Me" to pomysł, który krążył od lat i istnieje w podstawowej formie w postaci Google News i podobnych serwisach, które prezentują artykuły szerszej publiczności.



Trendy i odchylenia

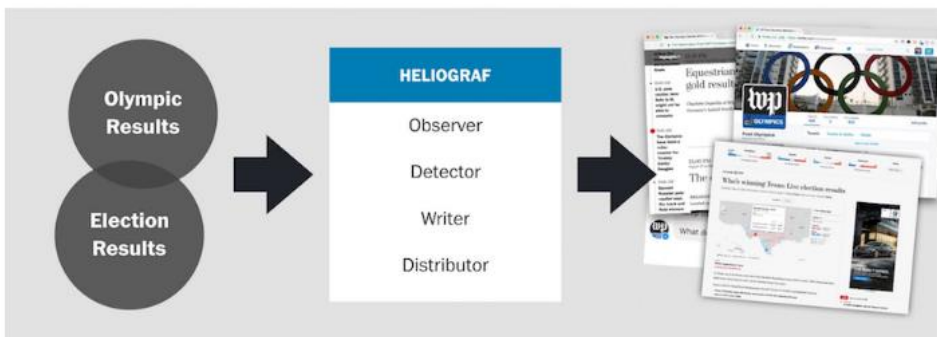
Firmy medialne mogą wykorzystywać sztuczną inteligencję do agregowania informacji z dostępnych źródeł, aby odczytywać z nich trendy w gospodarkach i na rynkach.

Financial Times korzysta również z botów, aby analizować swoje własne treści. System o nazwie **Janetbot** – przeanalizował stosunek płci twarzy pojawiających się w gazecie. Został przeszkolony na próbce zdjęć z całej branży, ale jego używanie zostało wstrzymane, kiedy okazało się, że próbka nie była wystarczająco zróżnicowana rasowo, aby wiarygodnie wykryć płęć.



Automated Story Process

Heliograf



The Washington Post

Washington Post - Heliograf

Technologie **Heliograf** i **ModBot** firmy **Washington Post** zajęły pierwsze miejsca w 2018 Global BIGGIES Awards, wyróżniającą najlepsze praktyki w zakresie produktów i strategii Big Data oraz sztucznej inteligencji stosowane przez firmy medialne z całego świata.

Heliograf zwyciężył w kategorii "Doskonałość w korzystaniu z botów" (ang. Excellence in Use of Bots), a jurorzy zauważyli, że "choć nie jest to pierwsza firma, która z powodzeniem zautomatyzowała tworzenie artykułów na podstawie kompilacji danych w szablonach, ich zestaw przykładów i wyjaśnień jest wymownie napisany i wspierane. Wyniki również wydają się całkiem obiecujące. Bardzo imponujące".

Z jego pomocą Washington Post raportował przebieg Igrzysk Olimpijskich, jak również wybory prezydenckie w Stanach Zjednoczonych w 2016 roku.

Heliograf Distribution Channels

Olympics and Elections



The Washington Post

*CTR - współczynnik klikalności. Stosunek między liczbą kliknięć a wyświetleniami reklamy, który mierzy się w procentach.

Washington Post - *Heliograf*

W pierwszym roku działania Heliografu, Washington Post wydał z jego użyciem około 850 artykułów.

Obejmowało to 500 artykułów na temat wyborów, które wygenerowały ponad 500000 kliknięć.

W gruncie rzeczy jest to niewiele, ale większość z nich to artykuły, na które wydawca i tak nie zamierzał poświęcać personelu.

Media wykorzystujące sztuczną inteligencję twierdzą, że ich działanie ma na celu umożliwienie dziennikarzom wykonywanie bardziej wartościowych prac.

Heliograf

An Intelligent, Automated
Storytelling Agent

- Automatic story generation from real-time data sources
- Delivery of channel-specific stories
- Personalized stories for readers



Washington Post ModBot

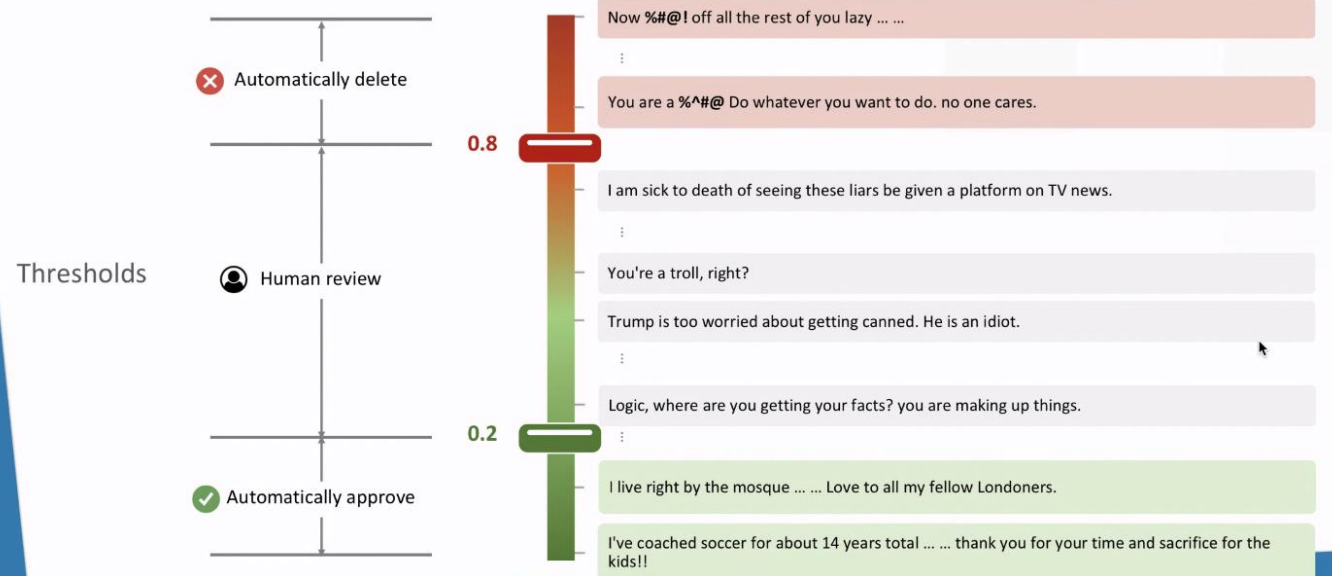
W maju 2017 Washington Post uruchomił **ModBot** - aplikację, która wykorzystuje sztuczną inteligencję do moderowania komentarzy.

Technologia ta wykorzystuje uczenie maszynowe do automatycznego filtrowania komentarzy wymagających moderacji, oznaczania artykułów wymagających monitu oraz zatwierdzania lub usuwania komentarzy.

Ocenianie komentarzy odbywa się za pomocą algorytmu, który był trenowany na przestrzeni lat danymi zebranymi w trakcie pracy moderatorów.

ModBot analizuje komentarze, aby określić, czy potrzebne jest działanie człowieka. Bierze pod uwagę takie czynniki jak język komentarza, jego wydźwięk oraz czy podobne komentarze były kiedyś moderowane.

Human in the loop during inference



BBC - *Juicer*

Juicer to interfejs API do agregacji wiadomości i ekstrakcji treści. Pobiera artykuły z BBC i innych serwisów informacyjnych, automatycznie je analizuje i oznacza (taguje) powiązаныmi jednostkami **DBpedia**. Podmioty są podzielone na cztery kategorie:

- ludzie
- miejsca
- organizacje
- rzeczy (wszystko, co nie mieści się w pierwszych trzech)

Jest to również "potok" agregacji wiadomości. Przyswaja artykuły i wydobywa z nich interesujące informacje.

What are we bringing to Seedhack?

- News Content
- Tagged with Linked Data concepts

The Juicer



Juicer Data + APIs

- Approx 500k news articles – various news providers
- Each article semantically annotated with DBpedia Concepts, Events and Storylines
- Concept APIs - find and lookup concepts
- News APIs - query news via the semantic annotations – use SPARQL !

BBC - Juicer

BBC Juicer obserwuje kanały RSS różnych serwisów informacyjnych. Gdy nowy artykuł zostanie opublikowany w jednym z tych kanałów, oprogramowanie scrappuje artykuł z wiadomościami - zarówno surowy tekst, jak i metadane (np. datę, godzinę, tytuł, źródło).

W następnym kroku Juicer identyfikuje i oznacza tagi wymienione w tekście artykułu, dzięki czemu można je przeszukiwać. Pozwala to analizować trendy w społeczeństwie, czy tworzyć statystyki oparte na rzeczywistych danych.

European Election Coverage

An experimental dashboard around the 2014 European Election coverage in the UK

by @BasileSimon and @IainCollins for @BBC_News_Labs



BBC News - Vote 2014 is the official BBC coverage of the European Elections

We've aggregated content from 40+ news organisations and put it through the BBC News Labs **News Juicer**

Sources

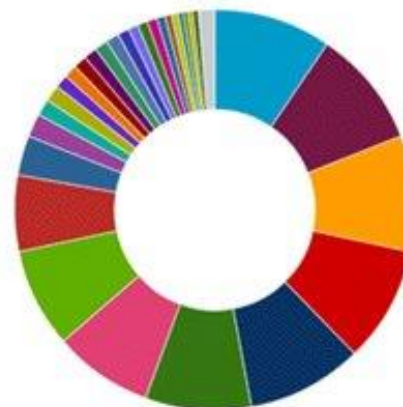
The publishers range from national publications (including the *BBC*, *The Guardian*, *Sky News*, *The Huffington Post* and others) to local newspapers (such as the *Yorkshire Evening Post*, *Birmingham Post*, *Belfast Telegraph* and *The Evening Standard*).

Semantic Analysis

Using semantic analysis, the News Juicer has extracted conceptual entities which are, broadly speaking, split up in to people (e.g. "Nigel Farage"), places (e.g. "Birmingham", "Europe") organizations (e.g. "The Labour Party") and topics (e.g. "immigration").

Number of candidates *versus* media coverage

Number of candidates running by party



Recent mentions in the media by party



The "number of mentions in the media" above was calculated from April 1st, 2014, to election day.

Automated Insights *Wordsmith*

Yahoo! Sports, właściciel platformy Automated Insights Wordsmith twierdzi, że dzięki wykorzystaniu sztucznej inteligencji i automatycznego generowania treści medialnych, jak artykuły, raporty czy newslettery, jest w stanie zrealizować dwa cele przy jednym podejściu.

Przyciągnąć użytkowników na dłuższe sesje, oferując im spersonalizowany, bogaty контент (oparty na sportowych danych)

Przyciągnąć partnerów reklamowych, oferując im dłuższy czas prezentacji reklam użytkownikowi - im więcej czasu spędzi na portalu – tym więcej reklam zobaczy.



Automated Insights Wordsmith

Większość relacji medialnych na temat „dziennikarstwa robotów” (dwa lub trzy lata temu) dotyczyła artykułów sportowych i finansowych w Yahoo!. Pomimo dziesięcioletniego upadku firmy (i niedawnej sprzedaży gigantowi telekomunikacyjnemu Verizon), Yahoo! wciąż może pochwalić się masą fanów. W swoich wiadomościach o finansach i mediach sportowych.

Automotive Insights – czołowy dostawca generowania języka naturalnego – przedstawia studium przypadku dotyczące współpracy z Yahoo!

```
{
  "data": {
    "company_name": "Hooli Corporation",
    "company_nickname": "Hooli",
    "date": "12/31/17",
    "prior_date": "11/30/17",
    "rev_current_month": 6842,
    "rev_prior_month": 1341,
    "rev_rank_12_mos": 7,
    "average_revenue_6_mos": 1300,
    "average_revenue_12_mos": 1250,
    "rev_mom_goal": 0.08,
    "rev_pct_rank_1_value": "0.2404",
    "rev_growth_rank_1_name": "Revenue 1",
    "rev_growth_rank_1_value": 0.7741827,
    "cost_sales_curr_month": 512,
    "cost_sales_prior_month": 650,
    "cost_sales_pct_change": -0.212307692,
    "gross_margin_curr_month": 0.2864912,
```

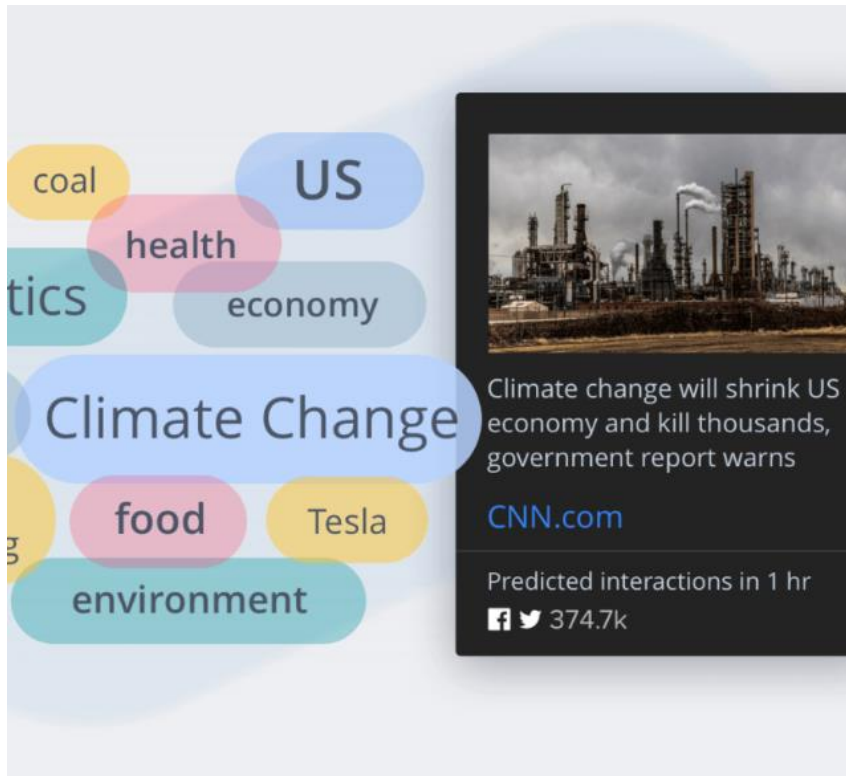
P&L Summary - Hooli Corporation

Overview

Blowing away the company's six-month revenue average, December revenue reached \$6,842 for Hooli Corporation. The more-than-5X rise over November is significantly higher than the 8% month-over-month growth goal.



- Growth was spurred primarily by Revenue 9, which made up 24% of total sales.
- Revenue 1 saw the largest change with a 77% jump in the past 30 days.
- The cost of goods sold has actually seen a drop, falling 21% since November.
- After missing the margin goal in November, Hooli exceeded expectations in December.



Associated Press - NewsWhip

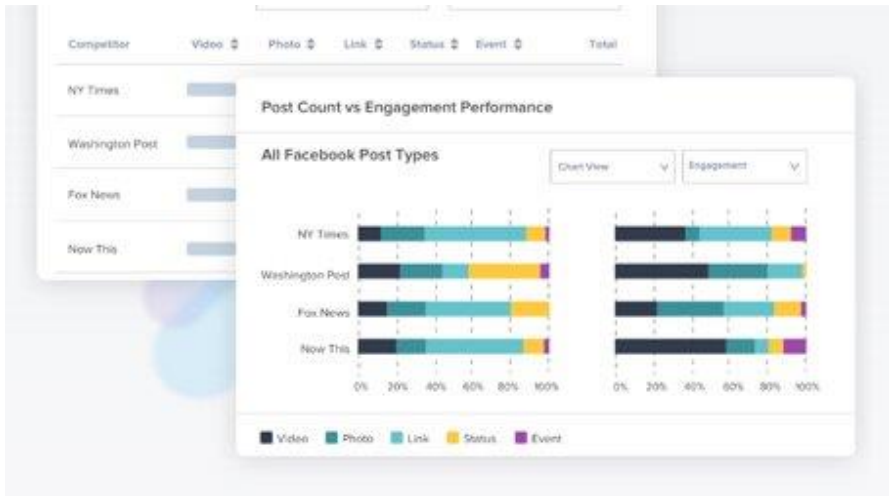
Associated Press po raz pierwszy zaczęła wykorzystywać sztuczną inteligencję do tworzenia treści informacyjnych w 2013 r. do pobierania danych i tworzenia raportów sportowych i finansowych. Obecnie redakcja AP korzysta z oprogramowania **NewsWhip**.

Strona analityczna reklamuje następujące główne możliwości:

- Analiza porównawcza konkurencji we wszystkich sieciach społecznościowych
- Zaangażowanie odbiorców wokół słów kluczowych i branż
- Identyfikacja osób mających wpływ na wyniki marki

Oprócz śledzenia wiadomości, może analizować rzeczywisty lub historyczny okres czasu od 30 minut do 3 lat i dostarczać reporterom powiadomienia w czasie rzeczywistym lub cykliczne podsumowania.

NewsWhip reklamuje się także, jako jedyna platforma monitorująca media w czasie rzeczywistym, która przewiduje historie i tematy, jakie będą miały znaczenie w nadchodzących godzinach, dając specjalistom ds. komunikacji możliwość podejmowania szybkich i pewnych decyzji.

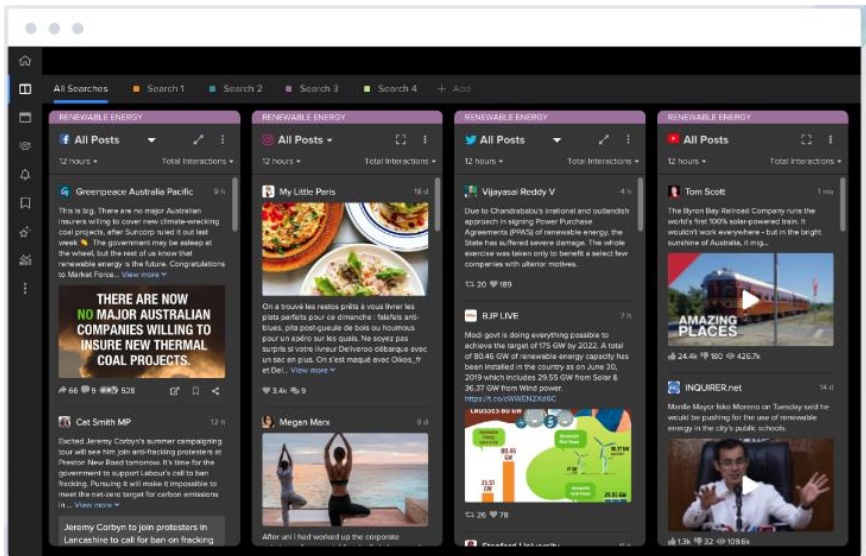


Associated Press - NewsWhip

Nie jest zaskoczeniem, że wydawcy opracowują własny zestaw unikalnych narzędzi niszowych. Sztuczna inteligencja może zastąpić ręczną pracę związaną z marketingiem influencerów, badaniami i analizą konkurencji.

NewsWhip rozkłada na czynniki pierwsze informacje uzyskane z artykułów dostępnych w internecie i na portalach społecznościowych. Pomaga planowanie redakcyjne, wyskrobuje wydarzenia dedykowane konkretnym odbiorcom. Umożliwia również porównanie z konkurencją, na przykład popularności w sieciach społecznościowych na przestrzeni lat.

Podobnie jak w przypadku Yahoo! Sports, **Automated Insights** zawiera studium przypadku dotyczące współpracy z Associated Press, które używa produktu "Wordsmith", do przekształcania surowych danych finansowych w artykuły.



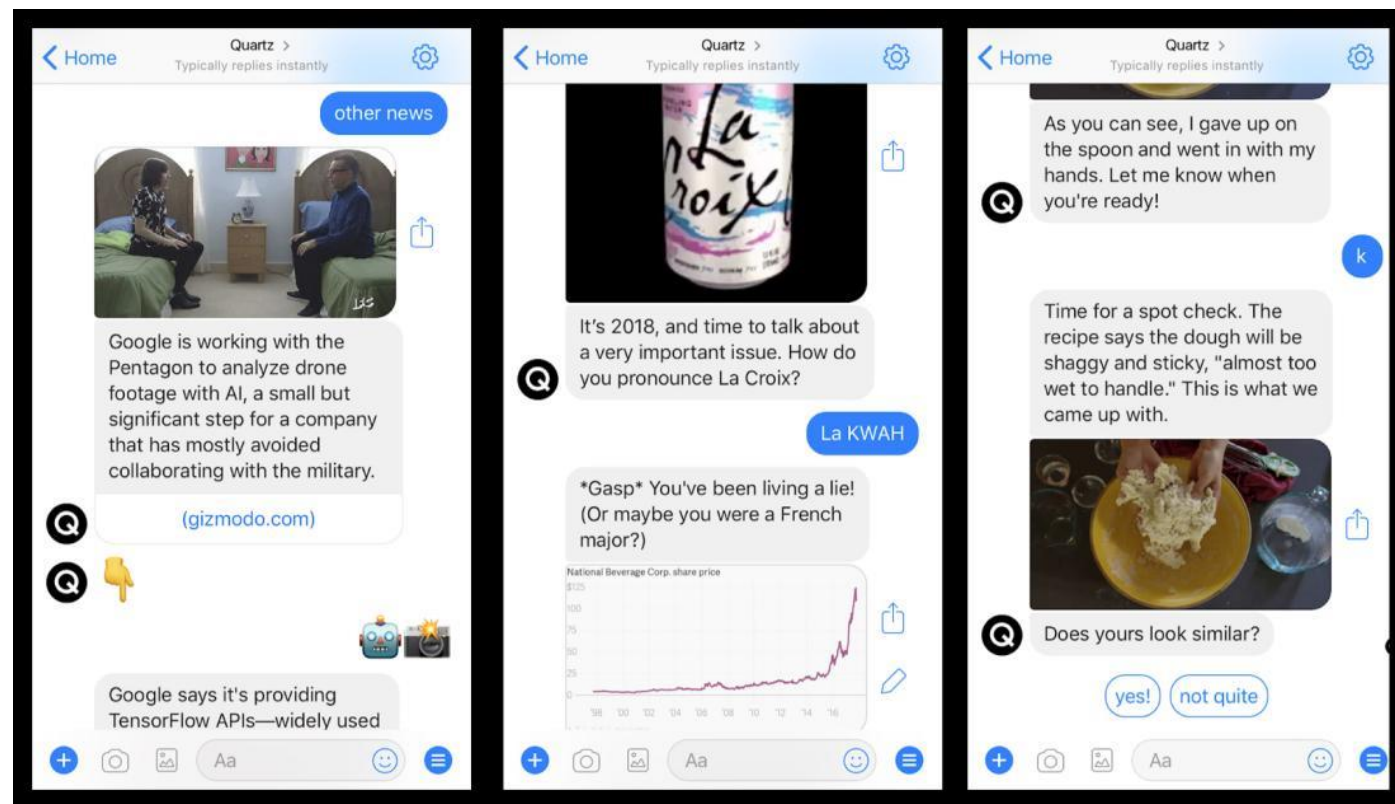
Quartz Digital News

W 2016 roku firma Quartz otrzymała grant w wysokości 193 000 funtów na założenie **Bot Studio** w celu stworzenia zestawu zautomatyzowanych narzędzi dla dziennikarzy.

Posunięcie to było inspirowane faktem, że media informacyjne przeniosły się nie tylko z druku na komputer i telefony komórkowe, ale także na inne urządzenia podłączone do Internetu w domu i samochodzie.

Użytkownicy wchodzą w interakcję z firmami za pośrednictwem czatu, głosu i innych innowacyjnych nowych kanałów, a firma Quartz chce również znaleźć najnowocześniejsze rozwiązania w zakresie korzystania z multimediiów.

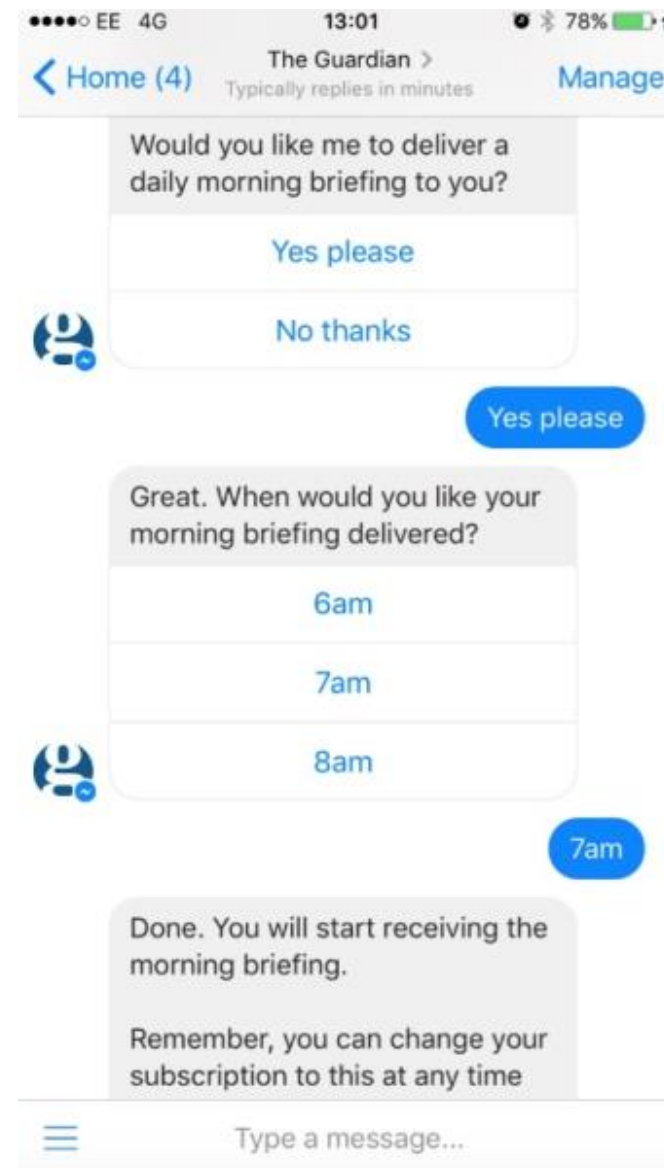
Chociaż projekt jest wyraźnie w powijakach, początkowe eksperymenty Bot Studio obejmowały interfejs wiadomości, który wygląda i przypomina „czat”. Użytkownicy wysyłają SMS-y z pytaniami o wydarzenia, osoby lub miejsca, a aplikacja odpowiada treściami, które jej zdaniem będą dla nich istotne.



The Guardian

W 2016 roku The Guardian uruchomił swojego Chatbota za pośrednictwem Facebooka. Aby zaoszczędzić czas na przewijaniu lub wyszukiwaniu wiadomości, chatbot pozwala użytkownikom wybierać spośród amerykańskiej, brytyjskiej i australijskiej wersji Guardian News, wybrać czas dostawy o 6 rano, 7 rano lub 8 rano i dostarczać wybrane wiadomości codziennie za pośrednictwem Facebook Messenger.

Jeśli użytkownik chce widzieć tylko nagłówki i wiadomości sportowe, jeśli chce tylko czytać popularne wiadomości techniczne i naukowe, może je również dodać. Podobnie jak w powyższym przykładzie Quartz, interfejs odpowiada na wiadomości czatu z powiązaną treścią odpowiadającą zapytaniu użytkownika.



Click bait ?

Innym, bardzo odmiennym rodzajem mediów informacyjnych, które odniosły sukces w świecie cyfrowym, są BuzzFeed i MailOnline, internetowa wersja Daily Mail, która jest uważana za najczęściej czytana anglojęzyczną stroną internetową gazety na świecie, ale który został zakazany jako źródło odniesienia przez Wikipedię.

Witryny te opierają się na reklamach zautomatyzowanych: odpowiadając na dane o apetytach czytelników za pomocą „listy” oraz artykułów o celebrytach, które skłonią ich do klikania reklam na Facebooku i generowania przychodów kosztu kliknięcia (CPC).



Synteza tekstu na mowę i Deepfake

1. <https://www.youtube.com/watch?v=eB29ZVDOfU>

2. https://www.youtube.com/watch?v=5iZuffHPDAw&feature=emb_logo

Źródła

- <https://emerj.com/ai-sector-overviews/automated-journalism-applications/>
- <https://www.theverge.com/2015/1/29/7939067/ap-journalism-automation-robots-financial-reporting>
- <https://phrazor.ai/blog/natural-language-generation-in-media-and-entertainment>
- <https://www.adexchanger.com/publishers/associated-press-uses-ai-boost-content-video-volume/>