

# 数据、算法与媒体创新

Data , Algorithm and Media Innovation

演讲人：王琼 武汉大学镭次元数据新闻研究中心负责人

跨界互联  
数聚未来

第四届中国数据分析师行业峰会  
CHINA DATA ANALYST SUMMIT

北京 中国大饭店 2017.07

# 媒介兴衰史 技术发展史

媒体最能感知、应用、传播新的技术，不仅仅因为媒体本身有教育和环境检测的功能，而且因为媒体安身立命之本就是要吸引人的眼球。



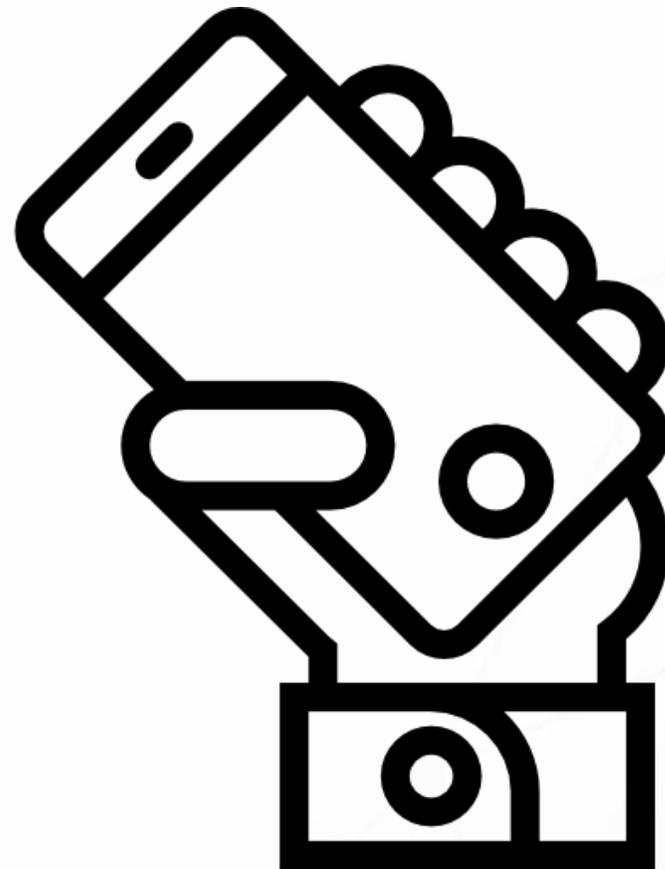
报纸



广播



电视



移动端  
多媒体

# 2013

对媒体而言

大数据元年

## 数据是个老东西！

数据受到前所未有的重视！



**VOTE**

EARLY VOTE IN FLORIDA OCT 20 - NOV 2  
FIND OUT WHERE BY VISITING  
**VOTEFORCHANGE.COM**  
OR CALLING 1-877-2-FL-OBAMA (1-877-235-6226)

OBAMA & BIDEN  
www.obamaobama.com

BarackObamaSuperstore.com

2012年：数据新闻成功预测奥巴马当选美国总统

跨界互联  
数聚未来

第四届中国数据分析师行业峰会  
CHINA DATA ANALYST SUMMIT

CDA 数据分析师  
CERTIFIED DATA ANALYST

IT大咖说  
知识分享平台



# 1952

## CBS用数据报道大选



TRUMP

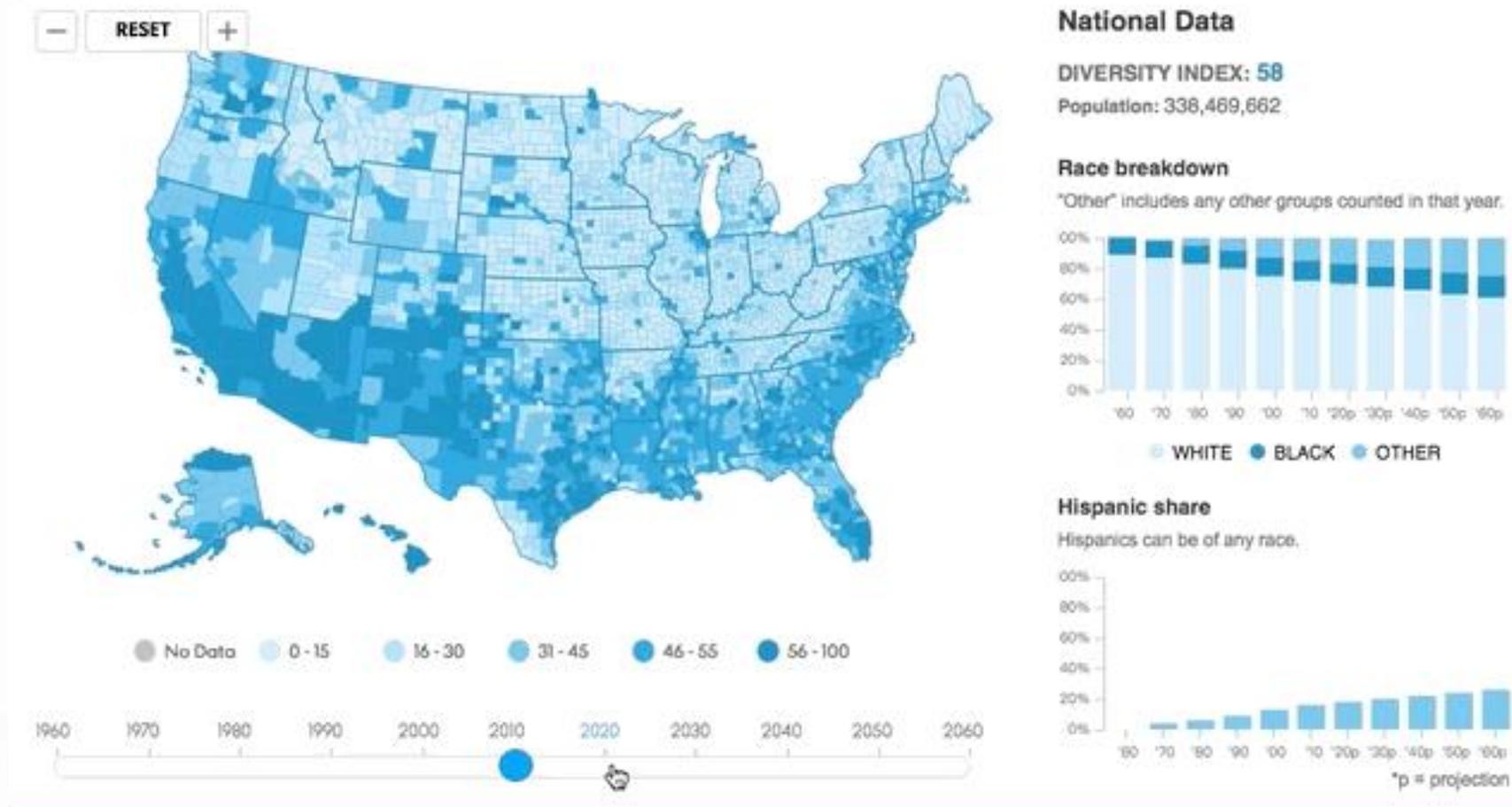
2016年：数据新闻错误预测希拉里当选美国总统

# 大数据新闻具有预测能力？

数据可以帮助我们更多的发现规律，但所有的算法都是有条件的  
没有万能的公式可以百分百的预测未来。

# USA TODAY

## 美国种族多样性



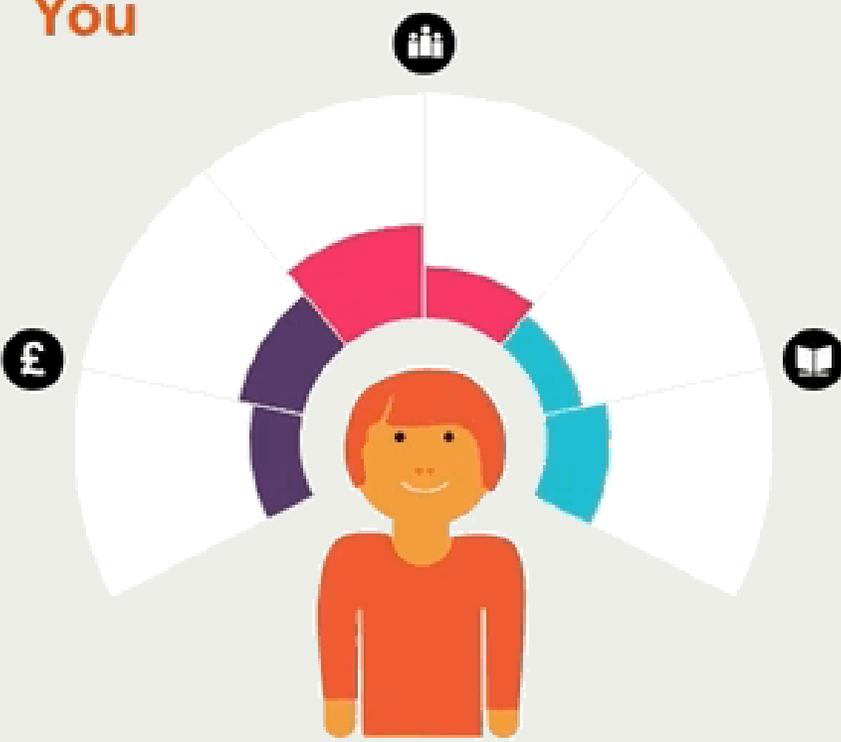
$$(White^2 + Black^2 + AmericanIndian^2 + Asian^2 + NativeHawaiian^2) \times (Hisp^2 + NonHisp^2)$$

B B C

你在哪个阶层？

Economic £ Social Social Cultural Your result ✓

**You**



What is your annual household income after taxes?  
Total income for you/spouse/significant other

Under £10k  £10-25k  £25-50k  £50-100k  Over £100k

Do you own or rent a property?  
Value of all property owned/mortgaged by you/spouse/significant other

Own  Rent

Under £125k  £125-250k  £250-500k  Over £500k

Do you have any savings?  
Pensions, shares, ISAs etc

None  £0-10k  £10-25k  £25-50k  £50-100k  
 Over £100k

Coloured wedges represent your details, select icons to find out more.

< NEXT >

- Your personal information is safe - all calculations are carried out on your computer
- BBC Lab UK's Great British Class Survey was launched in January 2011. More than 161,000 people took part, making this the largest ever study of class in the UK

算法的有限性并不是媒体在制作需要算法的数据新闻时面临的最大挑战。

更大的挑战在于

成本

高昂的成本  
媒体难以负担  
不能成为常态

目的

为什么要用算法？  
算法是终极目标吗

权力

数据的话语权  
归属与谁

心态

由算法能力衍生出  
的自卑感和优越感

# 数据新闻核心

用科学的方法

向可靠的数据

问合理的问题

讲有趣的故事



# 数据与算法 其他应用

智能推荐

算法监督

自动化写作

# 新闻内容自动推荐

方便读者的同时，是否存在这些问题？

创造信息茧房

原创内容缩减

原创团队减少

# 新闻内容自动推荐

打破信息茧房：ESCAPE YOUR BUBBLES

Who do you want to understand better?

Republicans

LET'S DO THIS

WIRED

Mashable

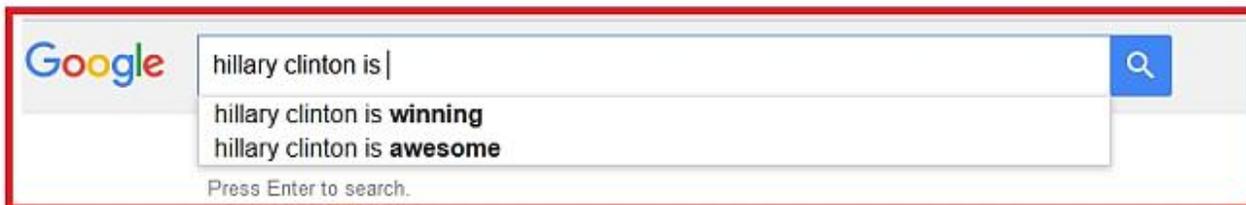
TNW

lifehacker

How does it work?

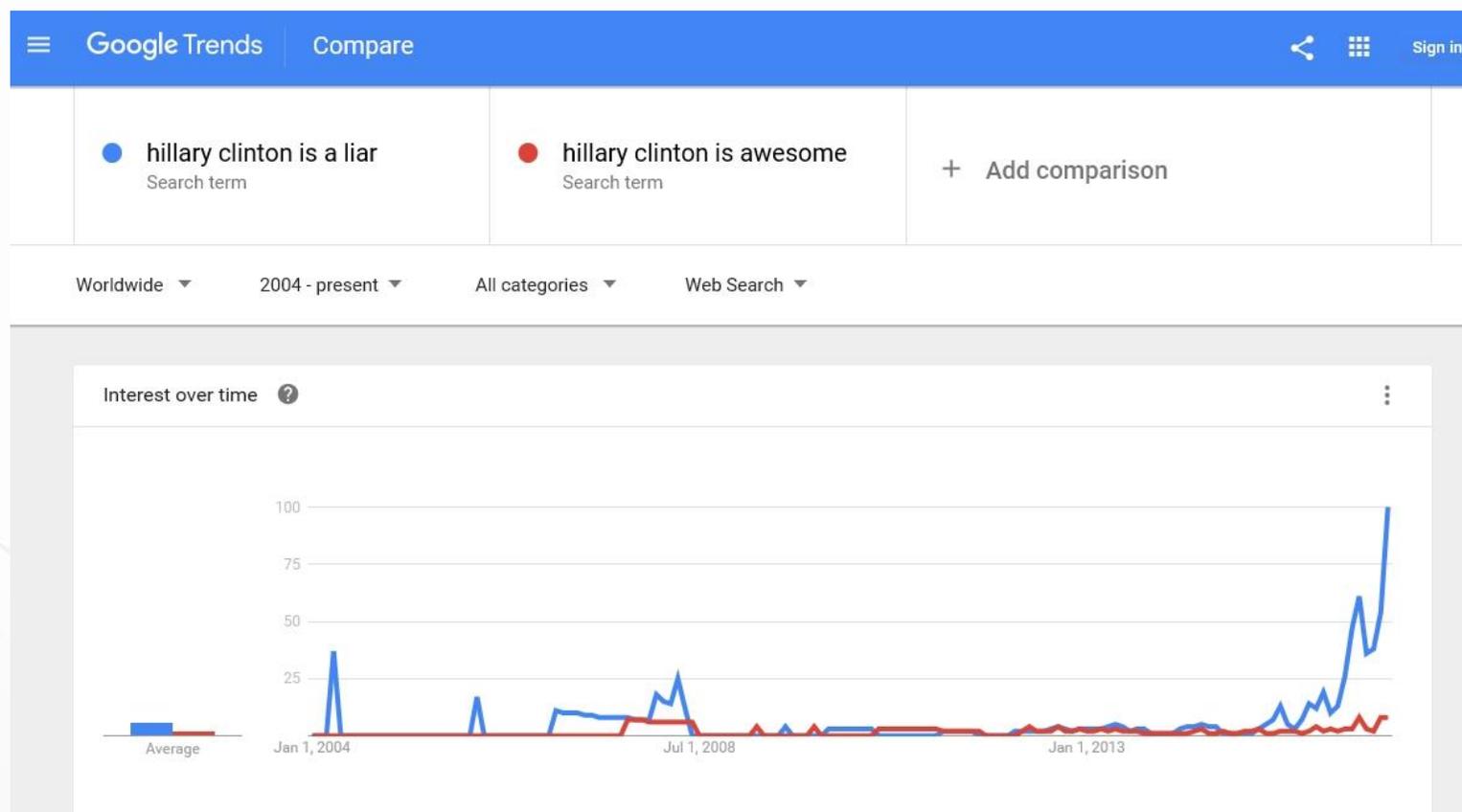
# 推荐阅读中的算法偏差

2016年Google News对美国大选的推荐阅读算法偏差



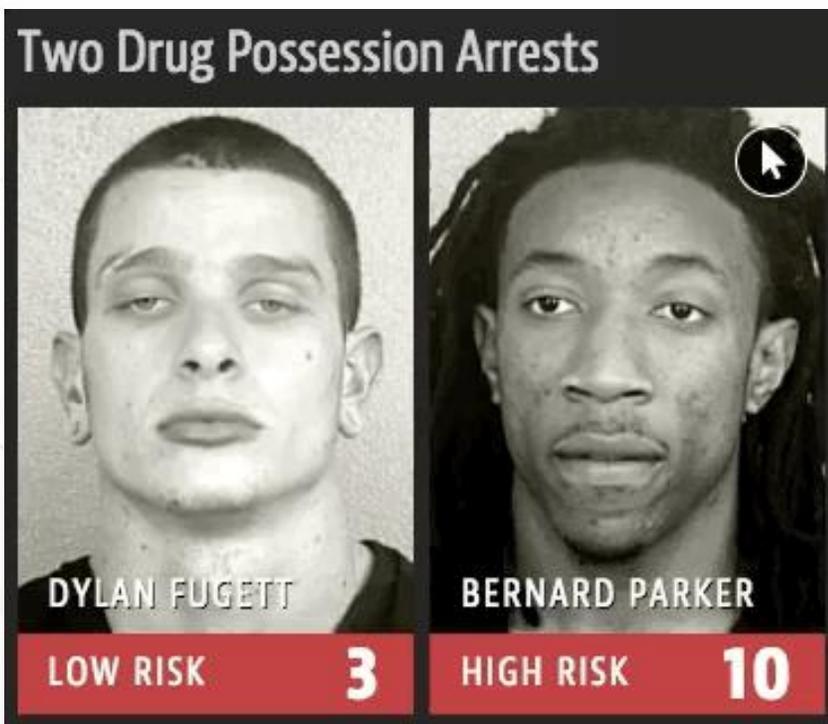
# 推荐阅读中的算法偏差

2016年Google News对美国大选的推荐阅读算法偏差



# 算法偏差与算法监督

## ALGORITHM ACCOUNTABILITY



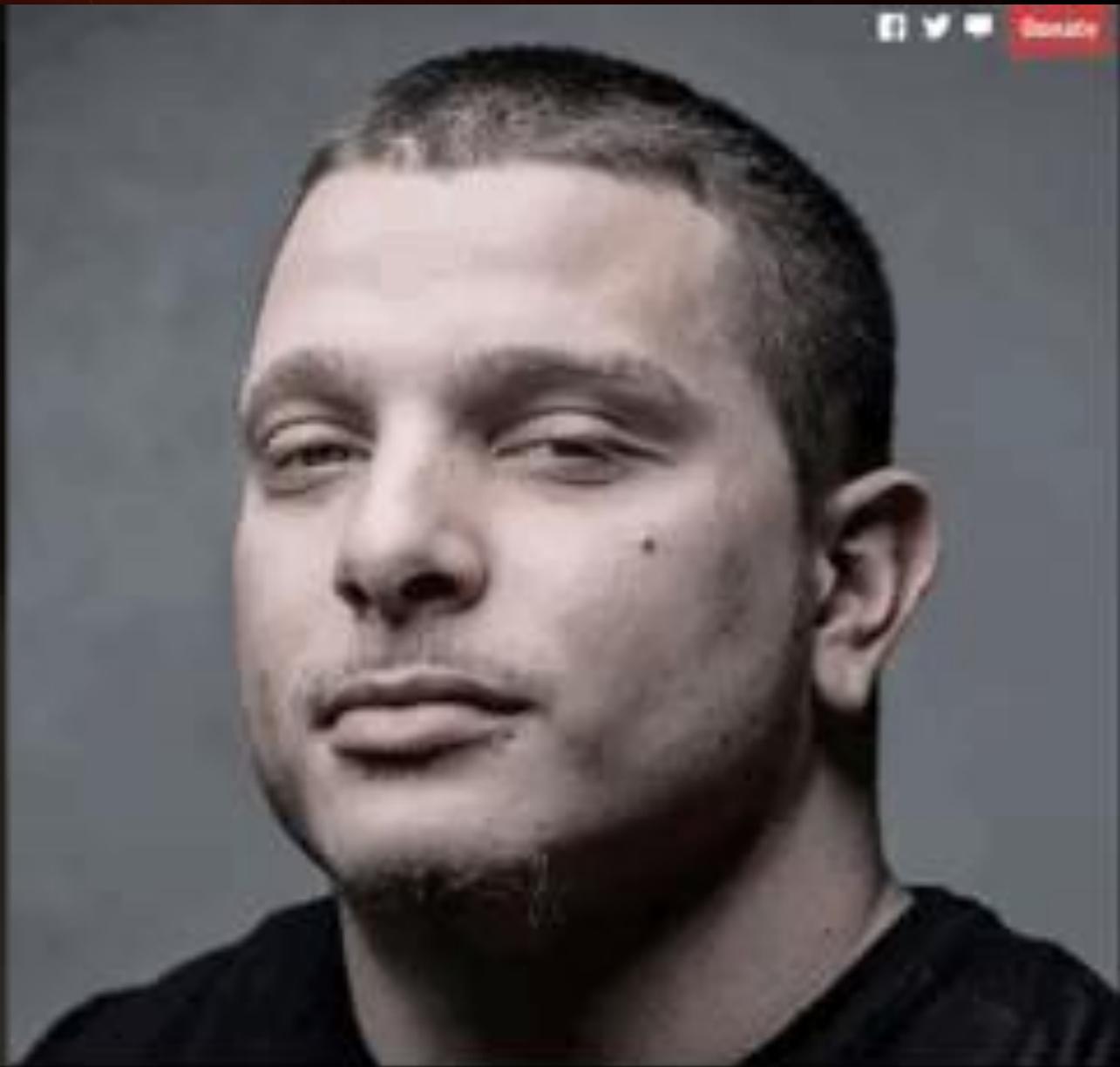
跨界互联  
数聚未来

第四届中国数据分析师行业峰会  
CHINA DATA ANALYST SUMMIT

CDA 数据分析师  
CERTIFIED DATA ANALYST

IT大咖说  
知识分享平台

数据+ PUBLICA



Twitter, Weibo, and other social media icons.

# 自动化新闻写作



HELIOGRAF

自动拟写文章  
可发评论



WIBBITZ

自动化视频编辑剪辑



QUAKEBOT

运用传感器数据自动  
推送地震报道

# 自动化新闻写作 未来

更多领域的报道

REACH

更加深入的报道

DEPTH

更加快速的报道

SPEED

# 真伪新闻检测

## REUTERS

### NEWS TRACKER

找到最热新闻  
确认其真实性  
并找寻传播路径

## facebook

联合新闻真实性检测机构Snopes,  
Factcheck.org, ABC News, the AP,  
Politifact, 以及Poynter's International  
Fact Checking Networks等等监督  
Facebook上的假新闻

# 数据时代

## 数据成为不可忽视的信息类型

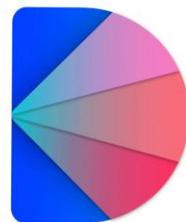
### 催生两大数据传播行业痛点

#### 数据来源的痛点

- 数据难获取
- 数据质量差

#### 数据表达的痛点

- 服务价格贵
- 技术门槛高



# 镛数

[www.dydata.io](http://www.dydata.io)

通过提供高质量的聚合数据

极为便捷的数据可视化写作工具

打造数据共享、交换、交易与传播平台

并为平台所有的参与者创造价值

# 1. 海量零代码模板, 多源聚合数据

类型



JavaScript 库

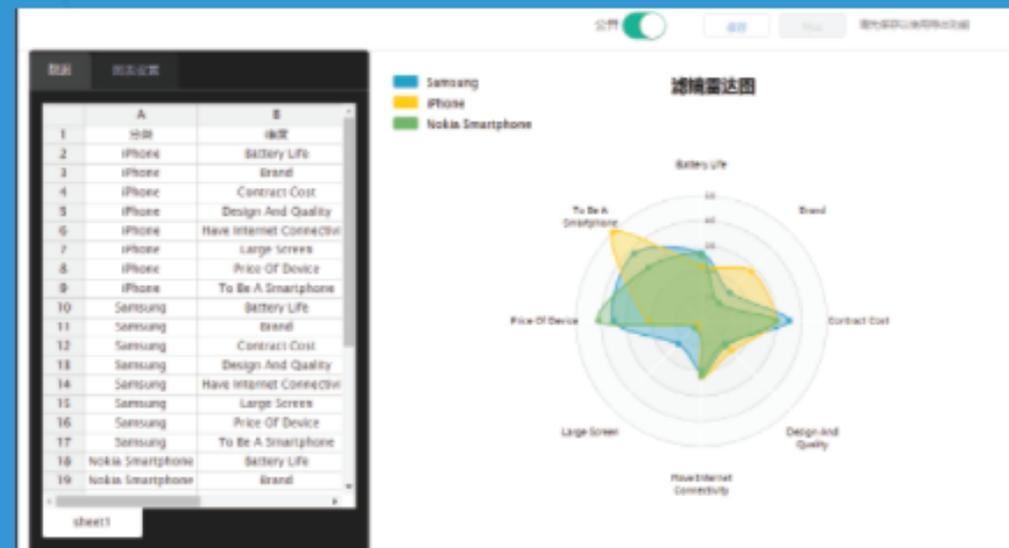
- D3
- eCharts
- G2

模板/案例

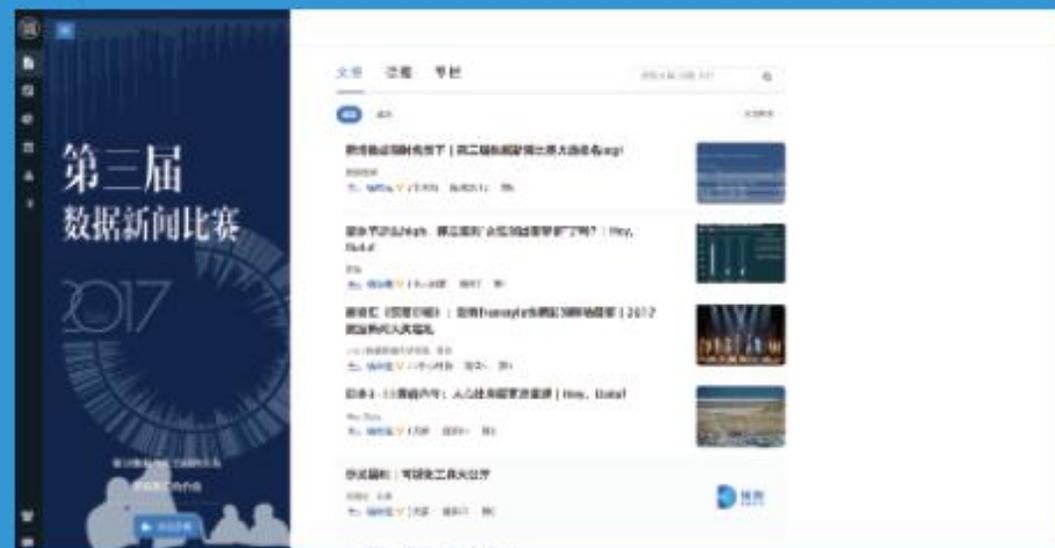
- 模板
- 案例



## 2. 数据到模板智能匹配, 从数据到发布一站式完成



# 3. 多模式写作、多平台分发



# 数据、算法与媒体未来？



CDA 数据分析师  
www.cda.cn

THANKS

# 跨界互联 数聚未来

第四届中国数据分析师行业峰会  
CHINA DATA ANALYST SUMMIT