

Automatic Media Analysis: From Monitoring To Insights

Dr. Miguel Martinez

Co-Founder / Chief Data Scientist at Signal AI

research.signal-ai.com

miguel.martinez@signal-ai.com

@miguelmalvarez



What is Signal?

- 6 years old scale-up from London with > \$30M raised
- Present in Europe, US and Asia with 130+ people
- Supporting Business leaders making better decisions by applying real-time large-scale Text Analytics. Processing >3M docs daily
- Research at the core of the company with 8 researchers in-house
- Collaborations with many universities (MSc, Visiting Researchers, Publications, ...)
- Co-Organisers of NewsIR (in SIGIR this year)
- ECIR long relationship: Chairs, Reviews, Publications, Sponsoring 2019, ...



Goal of the talk

- Introduce some of the challenges in media analysis
- Explain how Signal has approached these problems
- Discuss future challenges in the space and impactful research lines for the community
- I would like it to be as interactive as possible, so please stop me and **ask questions as I go**



The Problem



Problem

Business understand their internal metrics and processes well,
but they are ill-suited for responding to changing **external** factors



Problem

Business understand their internal metrics and processes well, but they are ill-suited for responding to changing **external** factors

Information production and accessibility is exploding.

Data, not instinct, is driving decisions at the highest levels.

Organisations are being exposed to critical risks every moment.



Problem: Use Cases

Opportunity

Leading market shift
New markets
Interests in new products
Lead Generation

Risk

Regulation
Competitors Initiatives
People leaving
Market shift

Reputation

Customer feedback
Damaged Products
PR Crisis



Problem: Complex

- Underlying Strategic Questions
- Represented as Complex Information Needs
- Over multiple data types
- Changing over time



Problem: Complex

- Underlying Strategic Questions
 - Represented as Complex Information Needs
 - Over multiple data types
 - Changing over time
- *Should I invest more in Asian AI startups?*
 - Articles on AI, related to funding rounds, product launches and M&A in Asia
 - Amount raised in Asia vs the US
 - Report every month and alerted on changes
 - Suggestion for upcoming startups

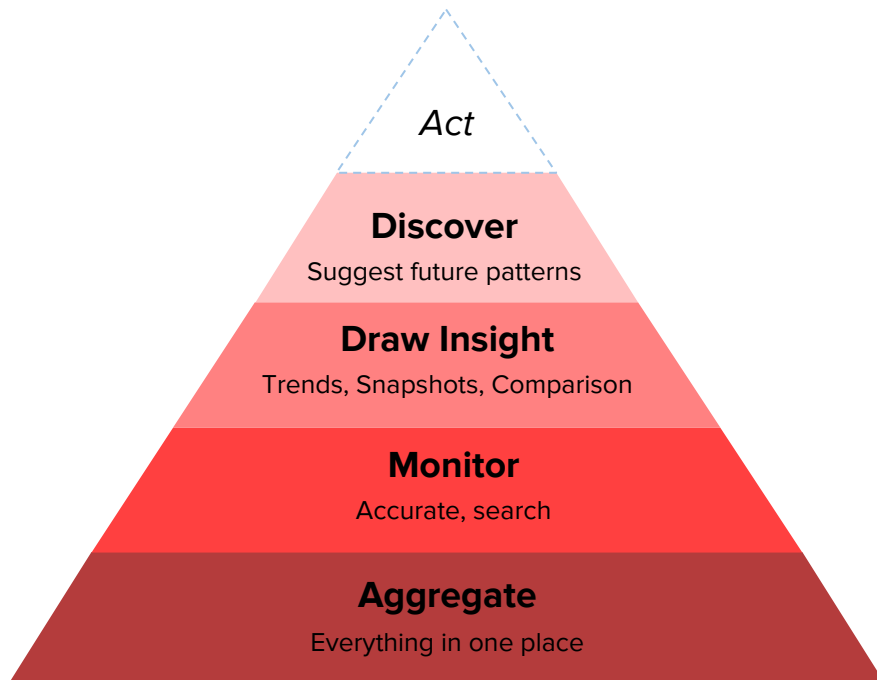


Signal's Perspective and Approach





Our Perspective



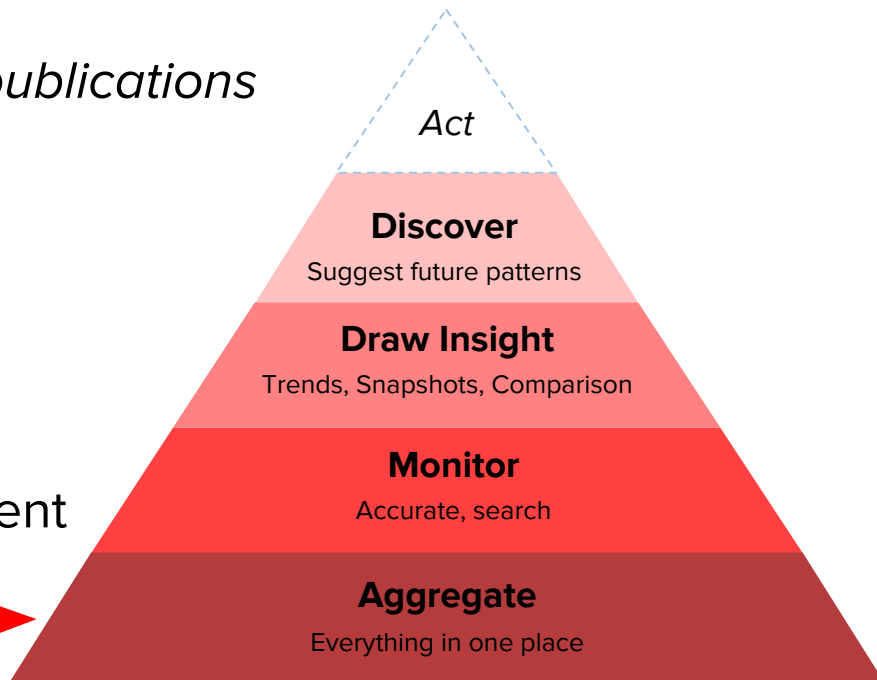


Our Perspective

Having complete coverage...

TechCrunch, VC Blogs, Investment publications

- Multiple Data Types
 - News, Blogs, Regulation, ...
- All the potentially relevant content
- In Real-time

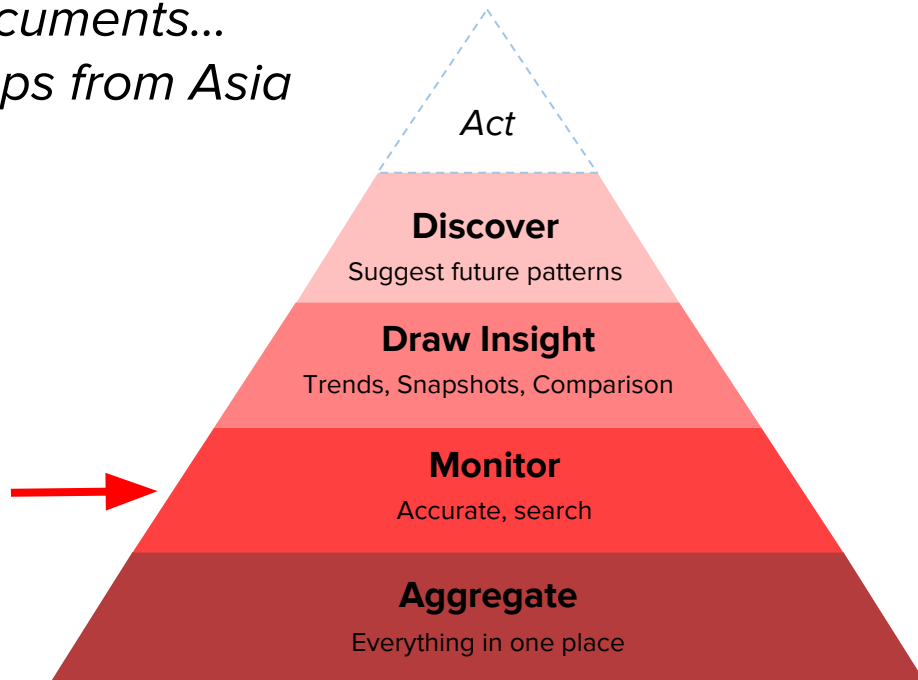




Our Perspective

Retrieving only and all relevant documents...
Docs about funding, VC and startups from Asia

- Information Filtering
- Based on complex queries
- Document-focused

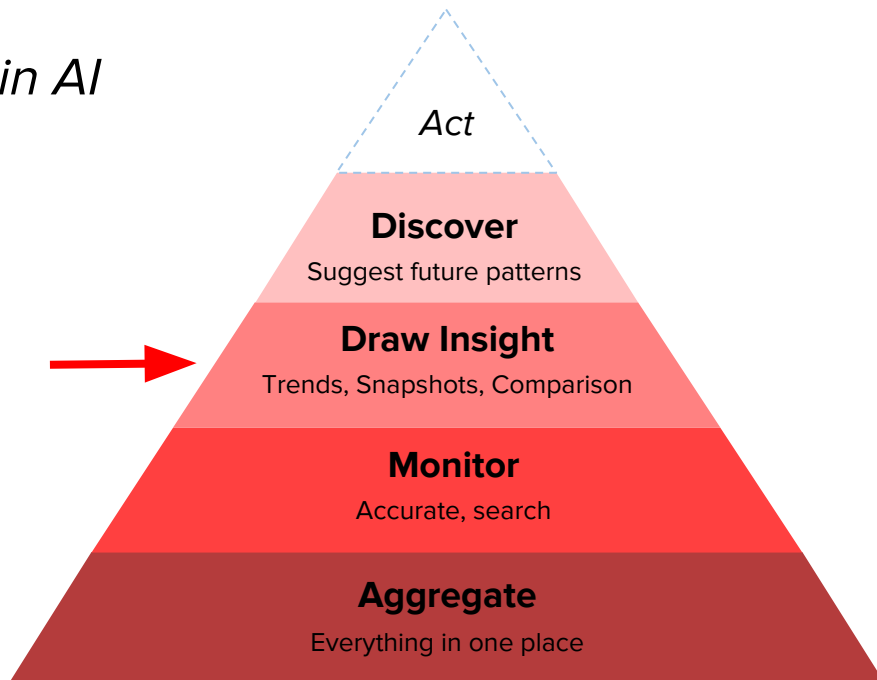




Our Perspective

When Text becomes a Story...
New trends, people and companies in AI

- Explore high-volumes quickly
- Focused on sets of documents
- Visualisations are key
- Trends and changes
- Comparisons

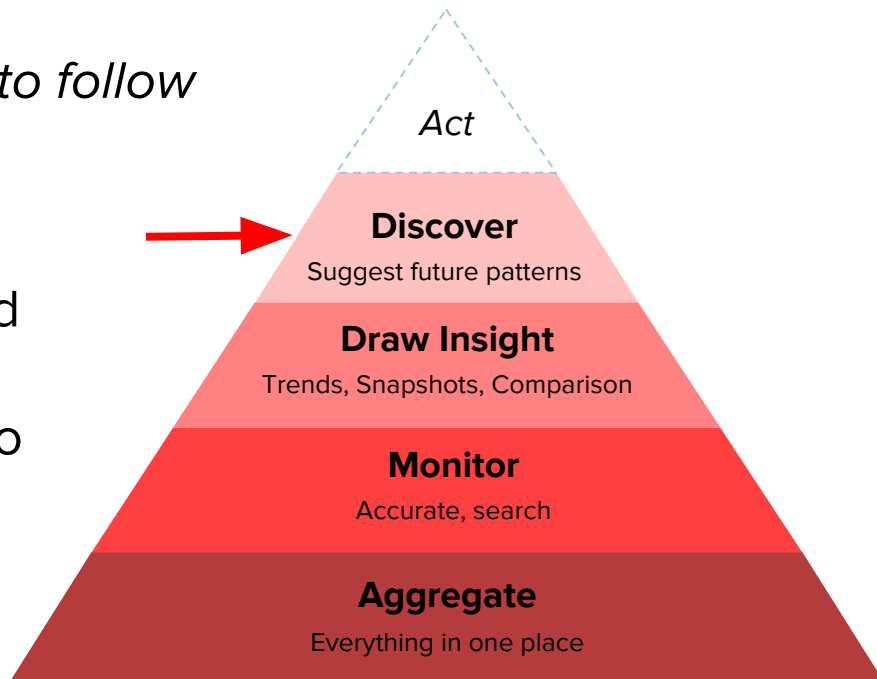




Our Perspective

Discover your Unknown unknowns...
Other fields, organisations or topics to follow

- Factors you should care about, but you do not know you should
- Recommendations outside echo chambers
- Horizon scanning

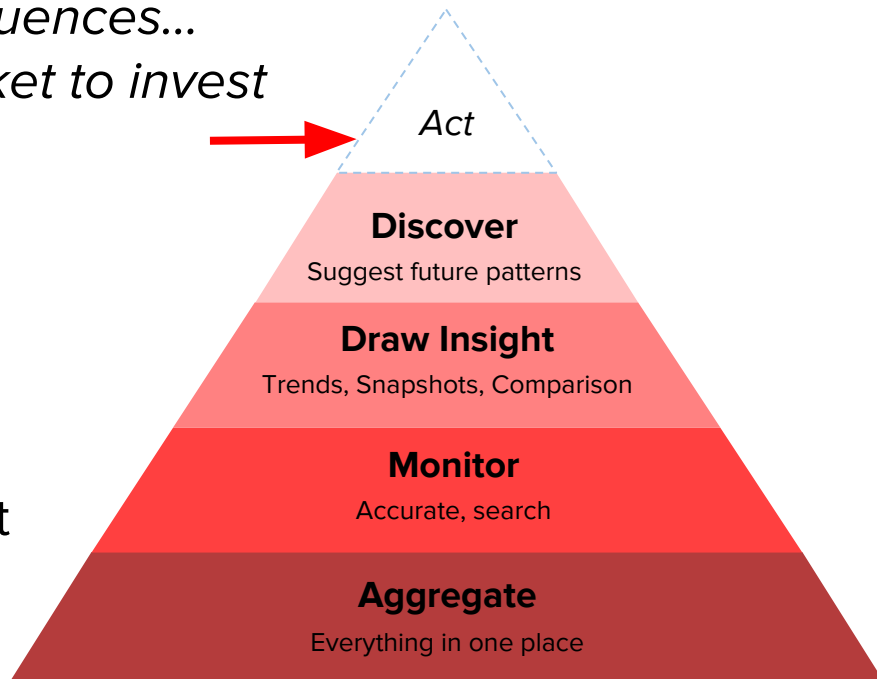




Our Perspective

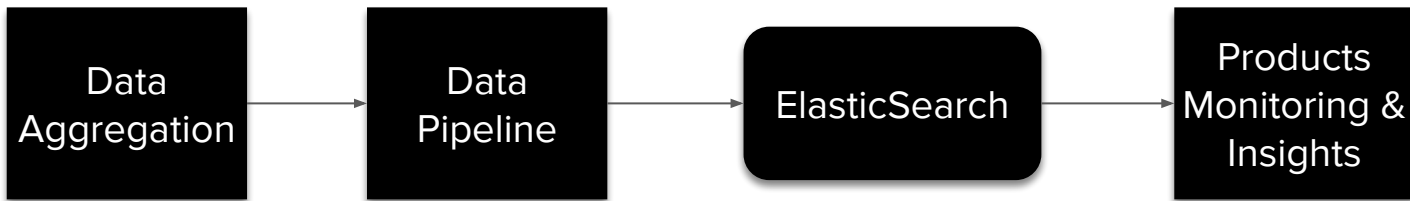
Suggest actions and predict consequences...
Best companies in the Asian AI market to invest

- Long term ambition
- Predictive Modelling & Forecast
- Digital “Consiglieri”





Our Solution





Data Aggregation

Print
& Online News

Social
& Blogs

Regulation

Broadcast

Alternative



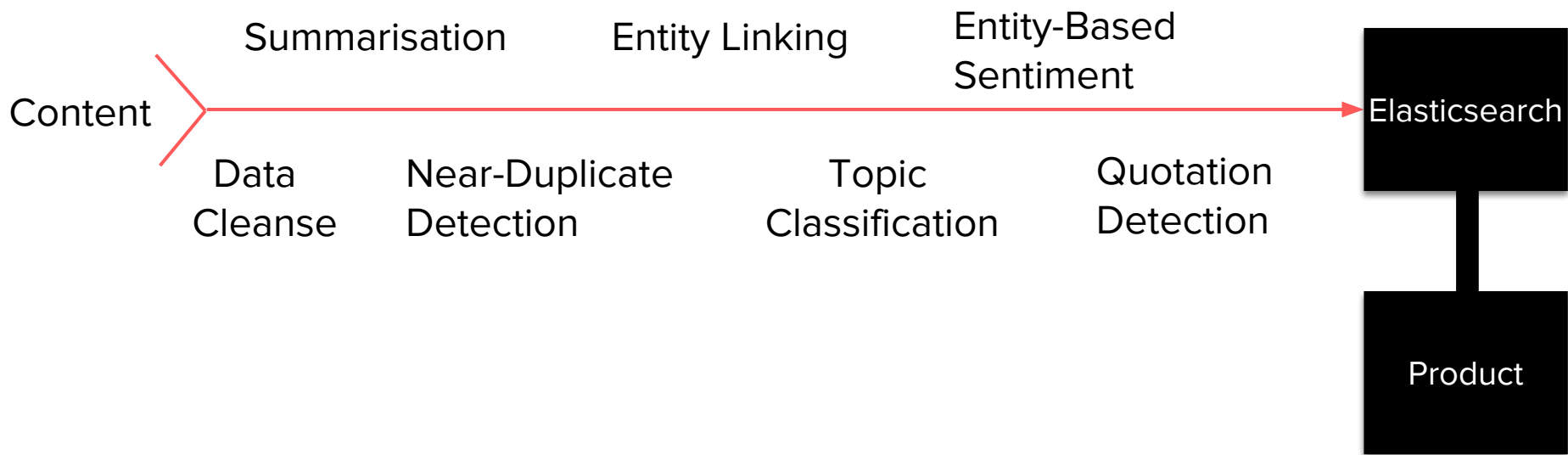
Data Aggregation



- Multiple data types is critical for some use-cases
 - Detecting future legislation changes based on social media unrest
 - Media Monitoring requires Tier 1 global publications
- The line between different types of data is blurring
 - Some blogs as reputable as the major newspapers
 - Some twitter account more influential than major newspapers

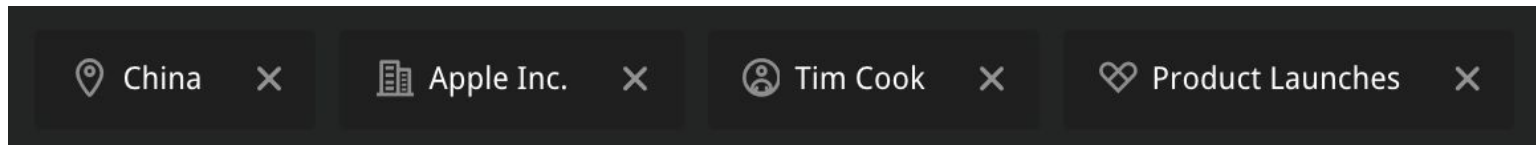


Data Pipeline





Product: Monitoring



Entity: Location

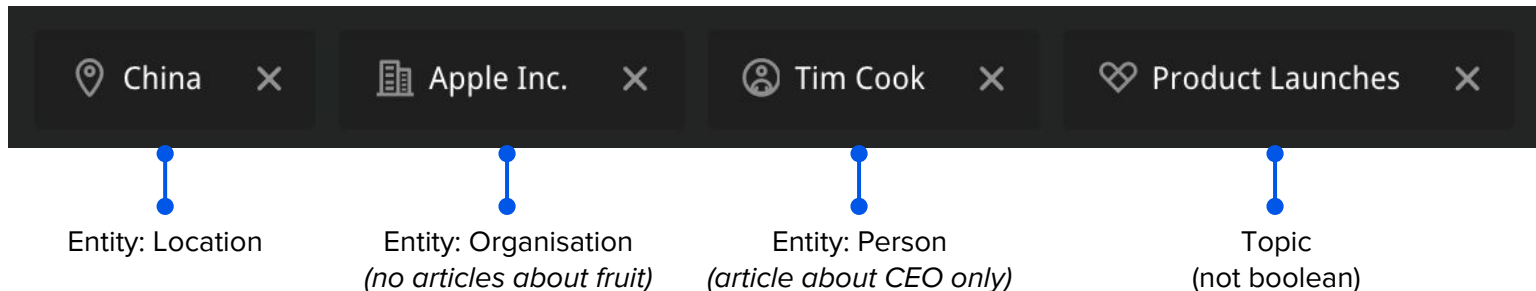
Entity: Organisation
(no articles about fruit)

Entity: Person
(article about CEO only)

Topic
(not boolean)



Product: Monitoring



- We index all the metadata from the pipeline
- Expected quality levels very high for some use-cases
- **The power is in the combination of fields (e.g., entities + topics + sentiment)**



Product: Insights

- Volumetric: Share of voice, brand comparison
- Competitors Sentiment in the industry and for given topics
- Popularity: Main related entities, topics, sources, ...
- **Useful, but the context, reason and narrative is missing**



Product: Insights Research Ahead

- Media Perception
 - How is the media representing an event/brand
 - Who/What is affecting the current trends
 - Defenders/Detractors: Main positive/negative stories (and why)
 - ...
- Overview given a context (event, brand, industry, ...)
 - Automatic event description and summarisation in real-time
 - Main stories given a user context and an information need
 - What is driving the media



Learnings

- User problems requiring multiple “*IR technologies*” to be used
- User-centric evaluation. Traditional methods might be biased
- Data changes all the time: Distribution, Volume, Velocity, Quality
- Comparisons are way more powerful than absolute numbers (e.g., Am I liked? vs Am I more liked than my competitors?)
- Explainability and/or ability to drill into individual articles is key

Automatic Media Analysis: From Monitoring To Insights

Dr. Miguel Martinez

Co-Founder / Chief Data Scientist at Signal AI

research.signal-ai.com

miguel.martinez@signal-ai.com

@miguelmalvarez