

標普全球 | CONCERT

大數據與人工智慧(AI)對學術界的重要性

The Importance of Big Data and Artificial Intelligence (AI) to Academia

演講者: Adelaide Tam 譚琬而

落地數據業務經理 (亞太區), 標普全球市場財智

adelaide.tam@spglobal.com

S&P Global

Market Intelligence

演講主題 -

- **大數據與人工智慧(AI)對學術界的重要性**
- **學術界數據運用的痛點**
- **標普的資料庫和人工智慧的解決方案**

AI與大數據在學術界的應用趨勢

教授對於AI研究跟教學的正面應用:

technology executive council

TOP STARTUPS FOR THE ENTERPRISE | COUNCIL MEMBERS | FOUNDING MEMBERS | ADVISORY BOARD

TECHNOLOGY EXECUTIVE COUNCIL

Why some college professors are adopting ChatGPT AI as quickly as students

PUBLISHED SUN, APR 2 2023-9:13 AM EDT | UPDATED MON, APR 17 2023-1:24 AM EDT

Carolyn Chun

SHARE f t in

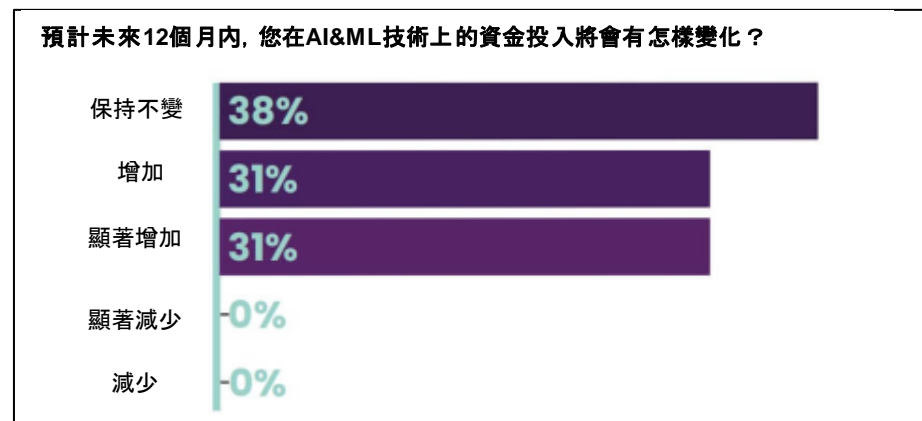
KEY POINTS

- A recent analysis by researchers at NYU, Princeton and the Wharton School finds that many of the jobs that will be most "exposed" to generative AI such as ChatGPT are in the college teaching profession.
- One of the first narratives to emerge from the sudden explosion in usage of ChatGPT is the risk of students cheating on writing assignments.
- But use by college teachers is growing quickly too, and adoption by educators may be critical to making the case that AI will augment the jobs humans are doing rather than replace them.

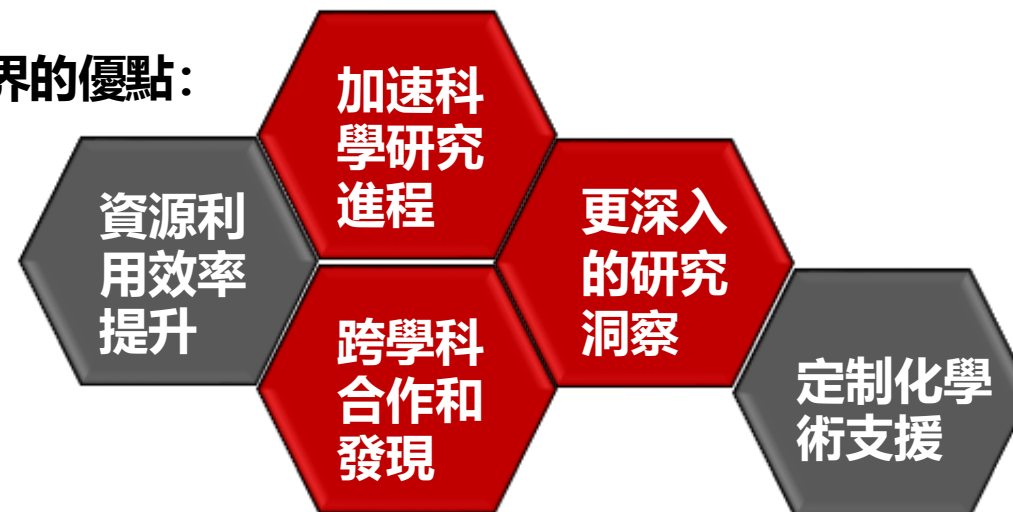
RELATED

Companies are learning that Gen Z isn't the easiest generation to work with

市場的關注與態度: 對AI與ML技術資金投入的調查:



大數據對於學術界的優點:



學術界數據運用的常見痛點 | 標普的解決方案

重複的手動數據清洗

自動化手動可重複任務



數據量太龐大

改進特定程式/介面 (例如Excel) 的限制



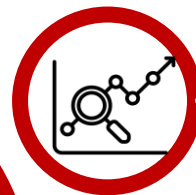
無效數據庫整理

利用 MI/CIQ Excel 外掛程式的資料庫數據



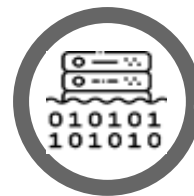
缺乏歷史數據

回溯測試因素/測試學術理論



沒有自建數據庫

構建雲端數據湖、應用程式、模型或演算法,



缺乏研究另類數據相關知識儲備

計算機科學學院的預測分析、定量或數據科學白皮書



人工智慧數據集 AI Datasets

- Textual Data**
- Machine Readable Transcripts
 - Machine Readable Broker Research
 - Machine Readable Filings
 - SCRIPTS Asia Transcripts
 - Textual Data Analytics
 - FiscalNote U.S Legislation & Regulation Data

- KENSHO人工智慧**
- Link 數據映射和實體關聯
 - Scribe (自動音訊轉錄)
 - NERD (實體識別)

Traditional Data

- Financials Point in Time data:**
- S&P Capital IQ Fundamentals
 - S&P Capital IQ Estimates
 - **Compustat® Financials**
 - Ownership
- Data Valuation Analytics**
- Economic Data
 - Country Risk
 - Short Interest
 - Dividend Forecasting
 - Research Signals – Short Squeeze
- Company Relationships:**
- **Professionals (ExecuComp)**
 - Company Intelligence
 - Business Relationships
 - Key Developments & Global Events
 - Private Equity
 - Transactions
 - Private Company Fundamentals
 - S&P Dow Jones Indices
 - ETF Style Classification
 - Alpha Factor Library

另類數據集 Alternative Datasets

- 永續發展數據 Sustainability Data**
- 環境數據
 - 物理風險
 - 碳收益 Carbon Earnings at Risk
 - 巴黎協定 Paris Alignment
 - 標準普爾全球 ESG 評分
 - Weather Source 氣候數據

Entity level identifiers
(e.g: DUNS, LEI, VAT, Local Registry)

Instrument level identifiers
(e.g. ISIN, Bloomberg Unique ID, SEDOL)

Industry Classification identifiers
(e.g. GICS, NACE, NAICS, ICB)

Suppliers / Customers / Strategic Alliances relationships mapping

Ultimate Parent / Parent / Child company identification and mapping

映射關聯服務 Cross-Reference

數據輸送技術 Delivery Technology

API

SFTP/SQL Database

WRDS

Snowflake

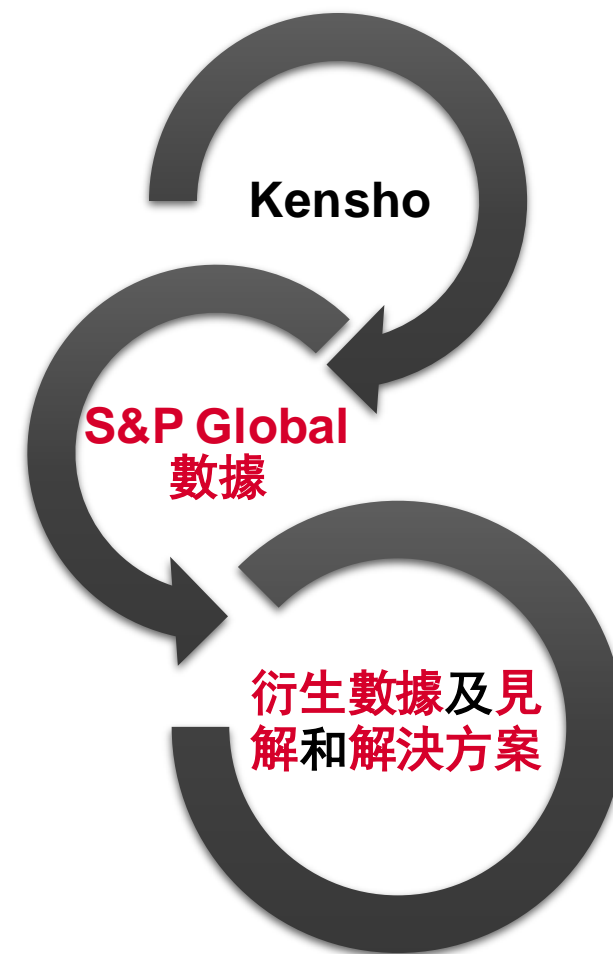
Marketplace Extract

Workbench

API Drive

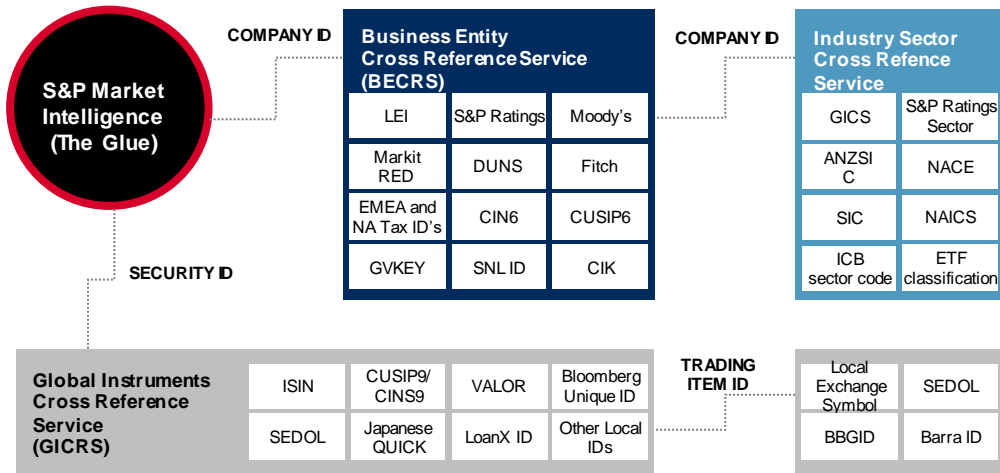
Kensho 是標普全球的人工智慧和創新中心。

- Kensho 開發改變機器學習數據應用上的尖端技術。
- Kensho 的工程人才利用機器學習的最新進展，以及標普全球提供的頂級數據廣度和深度，為學術研究者創造新的可行見解和解決方案。



強大的解決方案：新一代分析工具、機器學習模型和數據可視化

交叉映射：Cross Reference Services & Kensho Link



KENSHO Link

僅用於說明目的

Entity Information Input

Company Name* (required)

barclays egypt

Search

機器學習可以整理雜亂的數據並允許您快速準確地攝取新數據集。

Results & CapIQ IDs

Score	CapIQ ID	CapIQ Company Name
82.90%	13401047	Attijariwafa bank Egypt S.A.E.
2.70%	4481284	Centamin plc
1.60%	13393444	Telecom Egypt Company S.A.E.
1.40%	20530398	Egypt

文本數據和主題標記 Textual Data & Topic Tagging

KENSHO

Scribe NERD



Kensho Scribe：使用專為商業和金融音訊構建的機器學習進行自動語音辨識。

Kensho NERD：識別文本數據中的企業、人員、位置和其他實體並消除歧義。



For illustrative purposes only.

Kensho NERD 示例：命名實體識別和消除歧義應用在文本分析上

McCormick & Co. agreed to buy hot-sauce maker **Cholula**, seeking to capitalize on record demand for spicy condiments.

The seasonings company said Tuesday it reached a deal with **L Catterton**, the private-equity firm that has owned **Cholula** since 2019, to pay \$800 million for the brand. **The Wall Street Journal** first reported Monday that such a deal, culminating a auction of the business, could come together as soon as Tues

McCormick, famous for its red tins, owns brands including F and Thai Kitchen products.

It already owns another popular known for its buffalo-wing sa

McCormick, which is based in Hunt Valley, Md., and has a market value of around \$24 billion, sells both to consumers and restaurants

- 將 Cholula 視為 Cholula 食品公司，而不是墨西哥的城市。
- 將其鏈接到 CapitalIQ 中的實體

Entity Name: The Cholula Food Co. Inc.
Entity ID: 252373178
Confidence: 99%
Entity Type: Company

Tagging Source:

Wikimedia

NAIROBI, Jan 5 (**Reuters**) - **Kenyan** President **Uhuru Kenyatta** said on Friday he will retain **Henry Rotich** as his finance minister.

Trained as an economist, **Rotich** worked with the central bank and the ministry of finance, before his initial appointment to the role in 2013. He has presided over steady economic growth but he has been criticised for ramping up the national debt.

The president was sworn in for a second and final term last November, three months after the **Supreme Court** invalidated his initial re-election and ordered a re-run.

Tag entities from Wikimedia

在模稜兩可的情況下選擇一個實體：肯尼亞最高法院

Click [here](#) to learn more!

Kensho NERD 亞洲學術案例：媒體共同報導對公司股價的影響的研究結果

研究撮要：媒體共同報導(Media Co-coverage)產生“聚光燈”效應，市場對共同報導公司有相同的市場反應，最終影響股價走勢。這有可能成為有利可圖的交易策略。

Kensho NERD 在確定媒體共同報導和建立因果關係方面發揮著至關重要的作用。識別電話會議記錄中提到的公司名稱，然後結合股票市場和其他數據進行分析。

從研究個體到構建網絡，不同維度尋找“peer firm”來預測股價：

基於專利文本構建技術“peer firm”

基於提單文本構建航運“peer firm”

基於招聘廣告文本構建勞工“peer firm”

基於媒體文本構建共同報導“peer firm”

Link to Research Paper: ["Media Co-coverage and Overreaction in Cross-Industry Information Transfers"](#)

人工智能數據集 AI Datasets

Textual Data

- Machine Readable Transcripts
- Machine Readable Broker Research
- Machine Readable Filings
- SCRIPTS Asia Transcripts
- Textual Data Analytics
- FiscalNote U.S. Legislation & Regulation Data

KENSHO人工智慧

- Link 數據映射和實體關聯
- Scribe (自動音訊轉錄)
- NERD (實體識別)

Traditional Data

Financials Point in Time data:

- S&P Capital IQ Fundamentals
- S&P Capital IQ Estimates
- **Compustat® Financials**
- Ownership

Data Valuation Analytics

- Economic Data
- Country Risk
- Short Interest
- Dividend Forecasting
- Research Signals – Short Squeeze

Company Relationships:

- **Professionals (ExecuComp)**
- Company Intelligence
- Business Relationships
- Key Developments & Global Events
- Private Equity
- Transactions
- Private Company Fundamentals
- S&P Dow Jones Indices
- ETF Style Classification
- Alpha Factor Library

另類數據集 Alternative Datasets

永續發展數據 Sustainability Data

- 環境數據
- 物理風險
- 碳收益 Carbon Earnings at Risk
- 巴黎協定 Paris Alignment
- 標準普爾全球 ESG 評分
- Weather Source 氣候數據

Entity level identifiers

(e.g: DUNS, LEI, VAT, Local Registry)

Instrument level identifiers

(e.g. ISIN, Bloomberg Unique ID, SEDOL)

Industry Classification identifiers

(e.g. GICS, NACE, NAICS, ICB)

API

SFTP/SQL
Database

WRDS

Snowflake

Marketplace Extract

Workbench

API Drive

Suppliers / Customers / Strategic Alliances
relationships mapping

Ultimate Parent / Parent / Child
company identification and
mapping

數據輸送技術 Delivery Technology

映射關聯服務 Cross-Reference

利用文本資料推動學術研究出版和教學

當今超過90%的資料是非結構化的，主要是文字形式。人工智慧和機器學習的興起，越來越多企業和研究人員採用**自然語言處理 (NLP)** 來幫助從文字資料中發現更多見解。

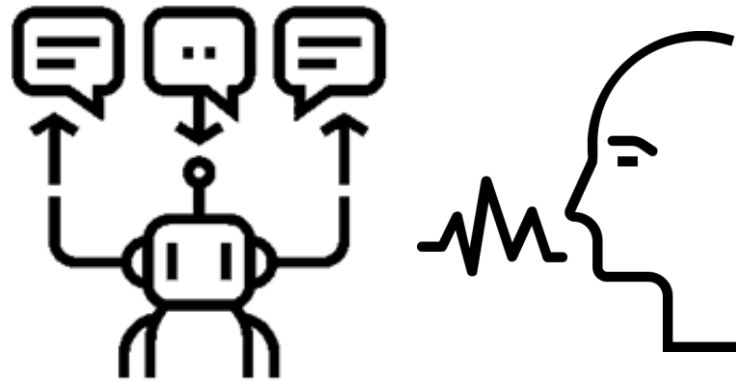
- 精準節省時間在傳統基本面研究、風險評估工作流程上，文字訊息一直由**人類** **手動**使用
- 通過結構化和機器可讀的格式，客戶可以：
 - **系統地**利用文本數據
 - 以**更快的**速度擴展傳統工作流程



文本數據產品

- **公司業績會議記錄** [Transcripts](#)
- **情緒分數和行為指標** [Textual Data Analytics \(TDA\)](#)
- **機器可讀的公司檔案** [Machine Readable Filings](#)
- **機器可讀券商研究報告** [Machine Readable Broker Research](#)
- **標普全球評等信用研究報告** [RatingsXpress®: Credit Research](#)
- **美國國會法規/立法*** [US Congressional Regulation/Legislation*](#)
- **Yewno 司法分析*** [Yewno Judicial Analytics*](#)
- **美國專利資料*** [US Patent Data*](#) (abstract data only)
- **Kensho 命名實體辨識和消除歧義** [Named Entity Recognition and Disambiguation \(NERD\)](#)

*除原始文本外還提供結構化分析



利用結構化的文字資料進行研究分析，釋放時間和精力，免除數據清洗之憂

Sample Data				
transcriptPersonName	speakerTypeName	transcriptComponentTypeName	componentText	KenshoNERDentityM...
Operator	Operator	Presentation Operator Message	Good day, everyone, and welcome to the Entegris' ...	
Bill Seymour	Executives	Presenter Speech	Good morning, everyone. Earlier today, we announ...	U.S. Securities and Ex...
Bertrand Loy	Executives	Presenter Speech	Thank you, Bill. I will make some comments on our...	Digital Specialty Che...
Bertrand Loy	Executives	Presenter Speech	Thank you, Bill. I will make some comments on our...	MPD Chemicals, LLC
Gregory Graves	Executives	Presenter Speech	Thank you, Bertrand. Our overall second quarter p...	MPD Chemicals, LLC
Gregory Graves	Executives	Presenter Speech	Thank you, Bertrand. Our overall second quarter p...	Versum Materials, Inc.
Operator	Operator	Question and Answer Operator Message	[Operator Instructions] We will now take our first ...	The Goldman Sachs G...
Toshiya Hari	Analysts	Question	I had 3, if I may. First, in terms of the operating mo...	
Bertrand Loy	Executives	Answer	Yes. So Toshiya, let me give you maybe a high-level...	Versum Materials, Inc.

Parsed, Structured, Well Linked

情緒分數和行為指標 (Textual Data Analytics) 示例：正面和負面演講示例

正面

Lauda, to be fair to them, they're doing a **good** job on the airline side. By any commercial measure, Lauda is doing very **well**. Load factor is running at **94%**, high on-time performance. We've restructured the aircraft leasing costs this year. We were very **fortunate**, I think, in the last year to be able to expand the fleet with leased A320s. And there is a sweet spot in the market at the moment for secondhand leased [CLOs], which are a bit -- are available at a **terrific** value like **\$170,000, \$180,000** a month on a lease basis.

- *Michael O'Leary, Group CEO & Executive Director, Ryanair Holdings*

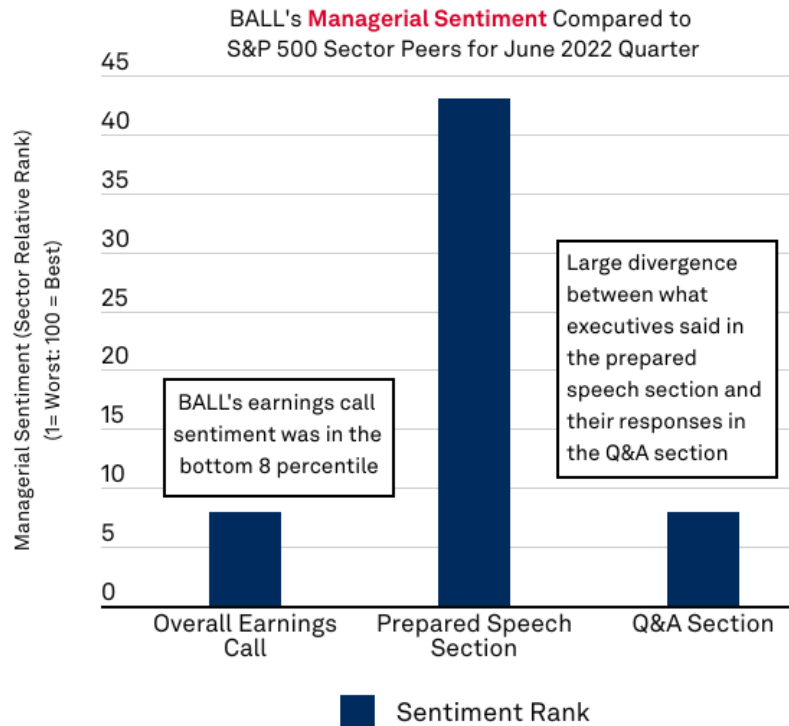
負面

So the MCAP design **failed** to deal with a boundary condition in an environment we should have known something about. The regulators made the same **mistake**. There are engineering disciplines that you can and should apply to that question that I think can get fixed relatively quickly. I don't think culture contributed to that **miss**. I think simply disciplines did. The question with respect to training recommendations and the **horrible** IMs that people would like to think all of Boeing rights really was **relegated** to a relatively small group of folks, but it wasn't detected.

- *David Calhoun, President, CEO & Director, The Boeing Company*

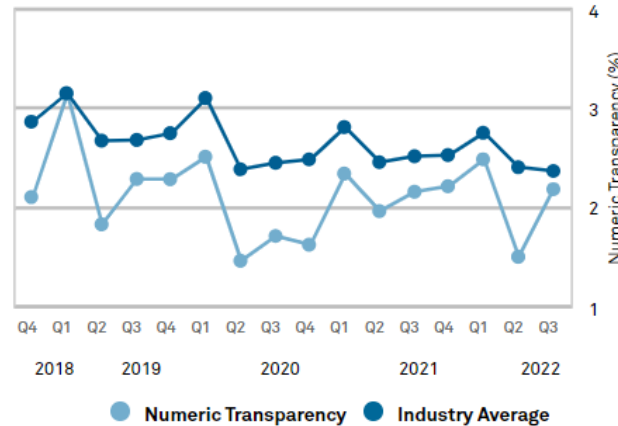
Net Positivity = (# of Positive Words - # of Negative Words) / Total Non-Stop Words Spoken

NLP分析800+指標，聚焦11,600+公司盈利電話，發現額外信號的基石



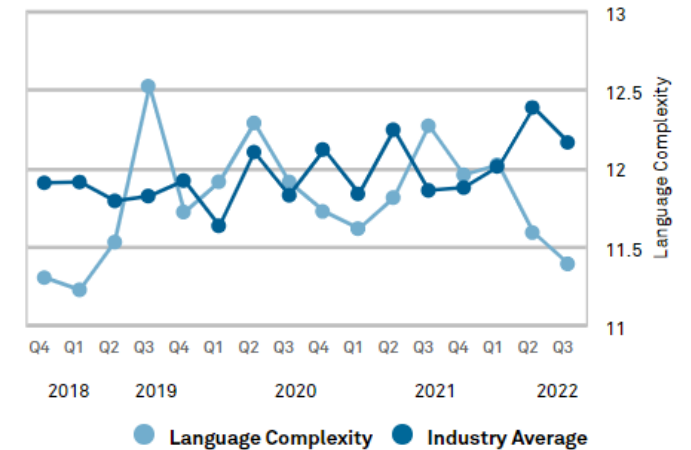
NUMERIC TRANSPARENCY ①

Q3 2022 Final Transcript Score: 2.19%
 Q3 2022 Materials Industry Avg. ①: 2.37%



LANGUAGE COMPLEXITY ①

Q3 2022 Final Transcript Score: 11.39
 Q3 2022 Materials Industry Avg. ①: 12.17



- Analyze the Net Positivity, Language Complexity, and Numeric Transparency scores based on **Executive** and **Analyst** speech
- Analyze the Net Positivity, Language Complexity, and Numeric Transparency at the **component** level:
 - Presentation operator message
 - Presenter speech
 - Question
 - Answer



Scores in the Document Viewer



- View the **NetPositivity**, **LanguageComplexity**, and **Numeric Transparency** metrics at the **total transcript level**
- Analyze the **trend** and compare company scores to the top level aggregated GICS or Market Intelligence industry score.
- View scores in a bar or line chart.
- Export scores and charts into excel.
- Access most scores **within 3 hours** of the Transcript being available on the platform

全球機器可讀文件(MRF)的ESG文檔

通過 ESG 檔進行的企業溝通將比現在更加普遍和頻繁

1.7萬

具有ESG文檔的公司

5.6萬+

ESG文檔

2006年

自 2014 年以來每年都有 2,000 多份文件

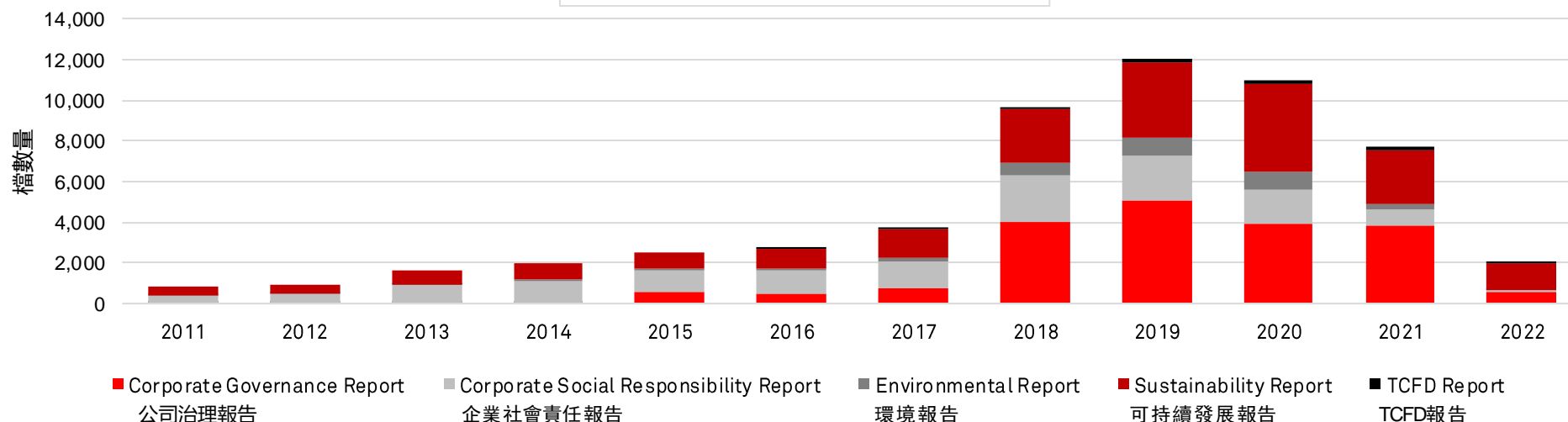
>92%

標普 1200 成分股已發佈 ESG 文件

主要亮點

- 對公司發佈的ESG內容運行NLP模型
- 分析和比較公司歷年的ESG計畫，並與同行進行比較
- 辨別與ESG相關引入的新舉措和戰略變化
- 利用公司檔案中的定性資訊優化定量ESG評分
- 根據公司過往的ESG目標和計畫的描述對其ESG成效作出評估並評分
- 創建專有ESG評估

按檔案類型劃分的 ESG 文件總覆蓋率



謝謝!

S&P Global
Market Intelligence

版權所有© 2023年 由S&P Global Inc 旗下公司S&P Global Market Intelligence編制。保留所有權利。

此等材料乃基於公眾一般可獲得且相信屬可靠來源的資料編制，僅供參考。未經S&P Global Market Intelligence或其附屬公司（統稱S&P Global）事先書面批准，不得以任何形式或方式修改、反向製作、複製或分發本材料所載內容（包括指數資料、評級、信用相關分析及資料、研究、模型、軟體或其他應用程式或其輸出結果）或其任何部分（“有關內容”），也不得將有關內容儲存於資料庫或檢索系統。有關內容不得用作任何非法或未經授權用途。S&P Global及任何協力廠商供應商（統稱“S&P Global各方”）並不保證有關內容的準確性、完整性、適時性或可用性。S&P Global各方概不對因使用有關內容而引致的任何錯誤或遺漏負責，不論原因為何。有關內容乃按“原樣”提供。S&P Global各方概不作出任何及所有明示或暗示的保證，包括但不限於保證用作任何特定目的或用途的適銷性或合適性；保證不存在程式缺陷、軟體錯誤或缺陷；保證有關內容的運行不會中斷；或保證有關內容可在任何軟體或硬體設定下運行。S&P Global各方在任何情況下概不就使用有關內容而引致的任何直接、間接、附帶、懲戒性、補償性、懲罰性、特殊或相應而生的損害、成本、開支、法律費用或損失（包括但不限於收入損失或利潤及機會成本損失或疏忽導致的損失）向任何人士承擔任何責任，即使在已獲知可能會發生該等損害的情況下亦然。

S&P Global Market Intelligence的任何意見、引述及信用相關及其他分析僅代表發佈當日的意見，而並非事實的陳述，亦不構成買入、持有或賣出任何證券或做出任何投資決定的建議，並不保證任何證券的適合性。S&P Global Market Intelligence可提供指數資料。指數不可直接投資，但可根據指數通過可投資工具投資於該指數所代表的資產類別。S&P Global Market Intelligence無義務在刊發後以任何方式或格式更新有關內容。當做出投資和其他商業決策時，不應依賴有關內容，且不應以有關內容代替使用者、其管理層、員工、顧問及/或客戶的技術、判斷和經驗。除非註冊為受託人或投資顧問，否則S&P Global Market Intelligence不以受託人或投資顧問身份行事。S&P Global將其各個分部的活動相互隔離，以保持其各自活動的獨立性和客觀性。因此S&P Global的某些分部也許擁有其他S&P Global分部沒有的資訊。S&P Global已制定政策和流程，確保對每個分析流程獲取的非公開信息保密。

S&P Global一般可能向證券發行人或承銷商或債務人收取評級和某些分析的報酬。S&P Global保留傳播其意見和分析的權利。S&P Global提供的公開評級和分析請參見其網站www.standardandpoors.com（免費）和www.ratingsdirect.com及www.globalcreditportal.com（訂閱），並可能通過其他管道（包括S&P Global出版物和協力廠商再分銷商）分銷。有關我們評級費用的其他資訊，請參閱www.standardandpoors.com/usratingsfees。

附錄

我們遍佈全球各個行業的解決方案屢獲殊榮。 Our award-winning solutions span industries across the globe.



- Global Data Provider of the Year 2022
- European Data Provider of the Year 2022
- APAC Data Provider of the Year 2022
- Regulatory Solution of the Year 2022



- Best Enterprise Data Management (EDM) Provider, 2022



- Best Overall Sell-Side Technology Provider, 2023
- Best Sell-Side Credit Risk Product, 2023



- RiskTech 100 – Top 10 Ranking
- Market Risk Category Winner: Buy Side Risk Solution (FRA)
- RiskTech Buyside 50 Buyside 50 - Top 5 Ranking RiskTech
- RiskTech Buyside 50 Investment/Market Risk - Credit



- Best Reference Data Provider, 2022
- Best Corporate Actions Data Provider, 2022
- Best Research Provider, 2022



- Most Innovative Diversity, Equity, and Inclusion Strategy
- Best Sanctions and PEPs Solutions



- InAsia Media - Corporate Actions
- InAsia Media - Data Provider of the Year
- InAsia Media - Initial Margin Compliance
- InAsia Media - Due Diligence and Third-party Risk



- Best Enterprise Data Management (EDM) System Provider, 2022
- Best Cloud Application Provider, 2022
- Best Performance Measurement and Attribution System Provider, 2022



- Best Environment Data Provider for ESG
- Best Analytics Provider for ESG

75+ Offerings Available across your **Data**, **Valuations** and **Analytics** needs



Standardized Data

Traditional Data

- [S&P Capital IQ Fundamentals](#)
- [S&P Capital IQ Estimates](#)
- [Compustat®](#)
- [Professionals](#)
- [Company Relationships](#)
- [Company Intelligence](#)
- [Business Relationships](#)
- [Key Developments & Global Events](#)
- [Private Equity](#)
- [Alpha Factor Library](#)
- [Research Signals – Short Squeeze*](#)
- [Ownership](#)
- [Transactions](#)
- [Private Company Fundamentals](#)
- [Economic Data*](#)
- [Country Risk*](#)
- [Securities Finance \(Short Interest\) dataset*](#)
- [Dividend Forecasting*](#)
- [S&P Dow Jones Indices](#)
- [ETF Style Classification](#)



Deep Industry & Cross Reference Services

Deep Industry Data

- [SNL Fundamentals](#)
- [SNL Financial Institutions Regulatory](#)
- [SNL Real Estate Property](#)
- [SNL U.S. Bank Branch](#)
- [SNL Depository Rates](#)
- [SNL Insurance](#)
- [SNL Energy](#)
- [SNL Metals & Mining](#)
- [SNL Media & Telecommunications](#)

Cross Reference Services

- [Business Entity Cross Reference Service](#)
- [Global Instruments Cross Reference Services](#)
- [Industry Sector Cross Reference Services](#)
- [Industry Classifications \(GICS®\)](#)
- [RED for CDS*](#)
- [Fixed Income Reference Data*](#)



Market Data & Textual Data

Market Data

- [ClariFI*](#)
- [CDS Pricing*](#)
- [OTC Derivates Data*](#)
- [Dynamic Tolerance Data*](#)
- [Initial Margin*](#)
- [Sector Curves*](#)
- [Liquidity Navigator*](#)
- [Pricing Data*](#)
- [Portfolio Valuation*](#)
- [Private Valuation*](#)

Textual Data

- [Machine Readable Transcripts](#)
- [SCRIPTS Asia Transcripts](#)
- [Textual Data Analytics](#)
- [Machine Readable Filings](#)
- [Machine Readable Broker Research](#)
- [Fiscal Note U.S. Legislation & Regulation Data](#)



Alternative Data

ESG Data

- [S&P Global ESG Scores](#)
- [Trucost Environmental Data](#)
- [Trucost Physical Risk](#)
- [Trucost Carbon Earnings at Risk](#)
- [Trucost Paris Alignment](#)
- [Weather Source Climate Data](#)
- [Crypto Data Services*](#)

Niche

- [Trading Analytics*](#)
- [IPquery Patent & Trademark Data](#)
- [Panjiva Supply Chain Intelligence](#)
- [Panjiva Macro: UN Comtrade](#)
- [MedMine Medical Device Transactions](#)
- [FDA Adverse Event Reporting System \(FAERS\)](#)
- [China Sentiment Analytics](#)
- [Preqin Alternative Asset Fund Performance](#)
- [BitSight Security Ratings](#)
- [451 Research Datacenter KnowledgeBase](#)

*Dataset originated from IHS Markit