



Yale SCHOOL OF MEDICINE
Biomedical Informatics and Data Science



MedViz: A New Tool for Exploring Biomedical Literature using Large Language Models and Visual Analytics

Huan He, PhD

Research Scientist, Yale University



2025 ANNUAL SYMPOSIUM

#AMIA2025

DISCLOSURE

I and my spouse/partner have no relevant relationships with commercial interests to disclose.



LEARNING OBJECTIVES

At the conclusion of this presentation the learner will be able to:

- 1. Recognize how hybrid LLM and visual analytics agents enhance biomedical knowledge discovery for healthcare research**, by enabling context-aware, citation-backed responses, interactive hypothesis generation, and seamless integration of natural language queries with visual exploration.
- 2. Understand the architecture and workflow of MedViz to support such applications**, including its three core components—data pipeline, LLM-based agents, and visual analytics interface—and how they integrate to support exploration of large-scale biomedical literature.



Existing Biomedical Literature Search Engines

PubMed[®] (Glucagon-like peptide-1 receptor agonist) OR (GLP-1 receptor agonist) OI

[Advanced](#) [Create alert](#) [Create RSS](#) [User Guide](#)

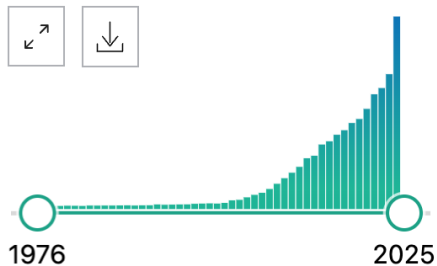
Sort by:

MY CUSTOM FILTERS

21,391 results

< < Page of 214 > >>

RESULTS BY YEAR



PUBLICATION DATE

- 1 year
- 5 years
- 10 years

- Semaglutide and Cardiovascular Outcomes in Obesity without Diabetes.**
1 Lincoff AM, Brown-Frandsen K, Colhoun HM, Deanfield J, Emerson SS, Esbjerg S, Hardt-Lindberg S, Hovingh GK, Kahn SE, Kushner RF, Lingvay I, Oral TK, Michelsen MM, Plutzky J, Tornøe CW, Ryan DH; SELECT Trial Investigators.
Share N Engl J Med. 2023 Dec 14;389(24):2221-2232. doi: 10.1056/NEJMoa2307563. Epub 2023 Nov 11. PMID: 37952131 Clinical Trial.
BACKGROUND: **Semaglutide**, a **glucagon-like peptide-1 receptor agonist**, has been shown to reduce the risk of adverse cardiovascular events in patients with diabetes. ...RESULTS: A total of 17,604 patients were enrolled; 8803 were assi ...

- Comparative effectiveness of GLP-1 receptor agonists on glycaemic control, body weight, and lipid profile for type 2 diabetes: systematic review and network meta-analysis.**
2
Cite Yao H, Zhang A, Li D, Wu Y, Wang CZ, Wan JY, Yuan CS



2025 ANNUAL SYMPOSIUM

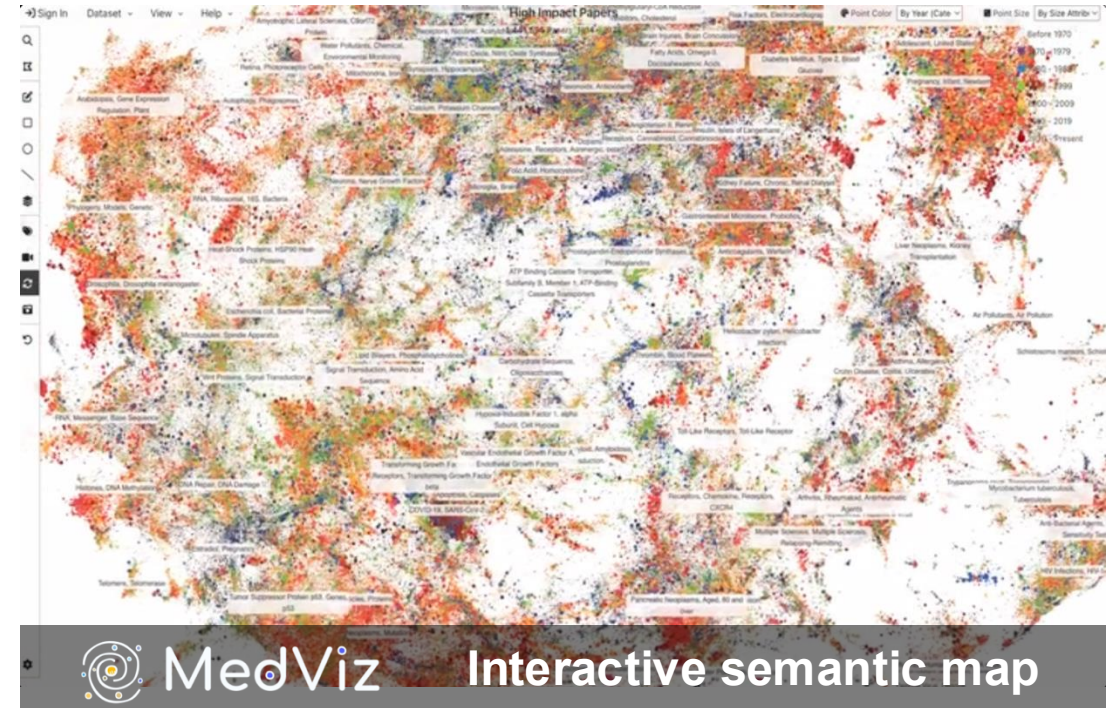
#AMIA2025

Our Vision - Redefine Biomedical Literature Search

The collage shows three overlapping search interfaces. The top interface is a search results page for "prostate cancer" with 4,370,000 results, featuring filters for fields of study, date range, and PDF availability. The middle interface is a PubMed search for "diabetes glp1" with 11,060 results, showing a bar chart of publication dates from 1988 to 2024. The bottom interface is a ranked list of documents for "diabetes glp1", listing titles like "Pharmacological treatment of hyperglycemia in type 2 diabetes" and "Cardiovascular, mortality, and kidney outcomes with GLP-1 receptor agonists in patients with type 2 diabetes: a systematic review and meta-analysis of cardiovascular outcome trials".

Ranked list of documents

Transform 
Agentic AI
Visual Analytics



- **Global landscape** of the knowledge space of interest
- Interactively explore the whole semantic space
- **Extract information** from focused papers
- Synthesize **insights** across papers
- Explore new directions or trends of **future**



2025 ANNUAL SYMPOSIUM

#AMIA2025

System Design

1. Data pipeline

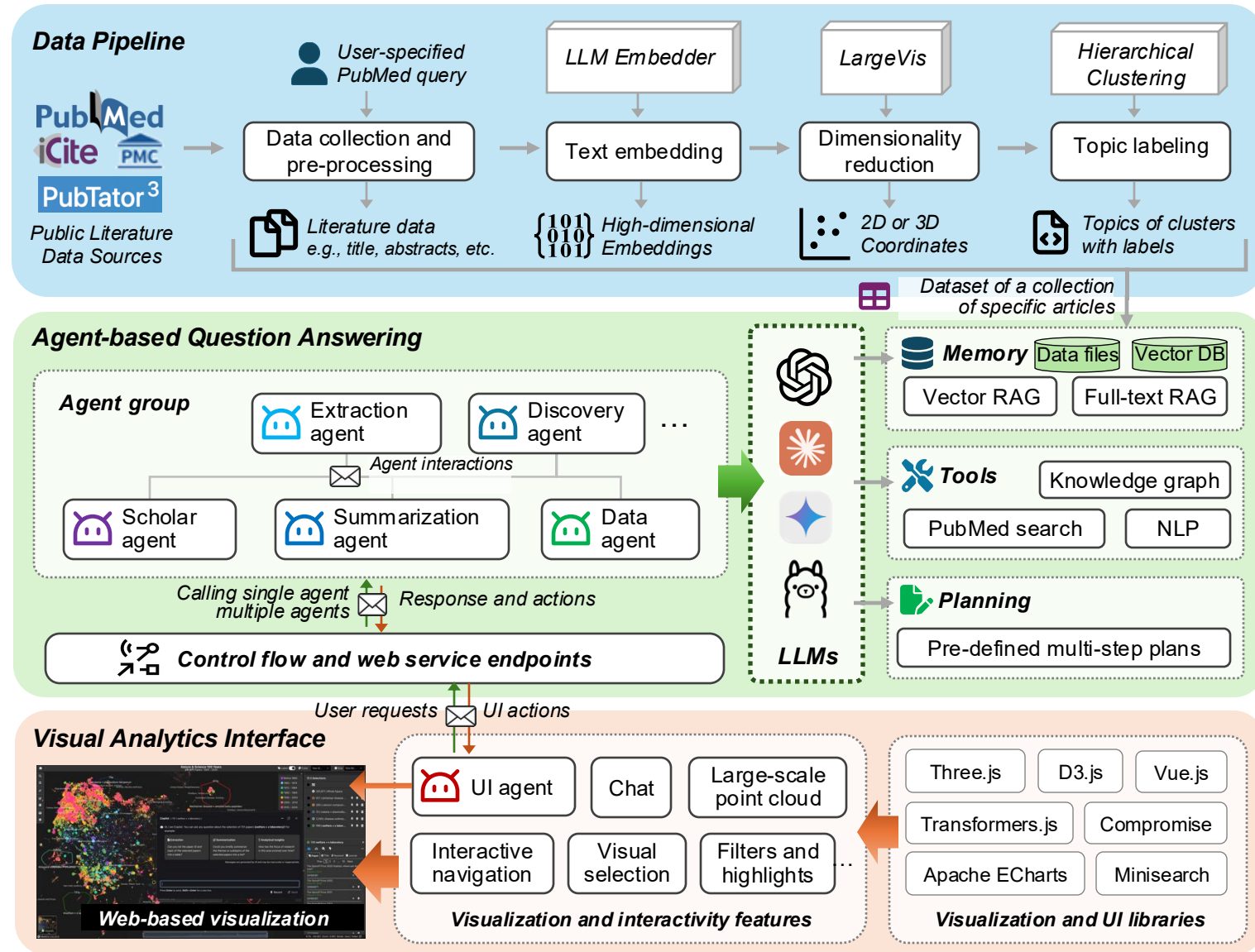
- LLM embedding
- Hierarchical clustering

2. Multiple agent-based question answering

- LLM-based agents
- Multi-agent collaboration

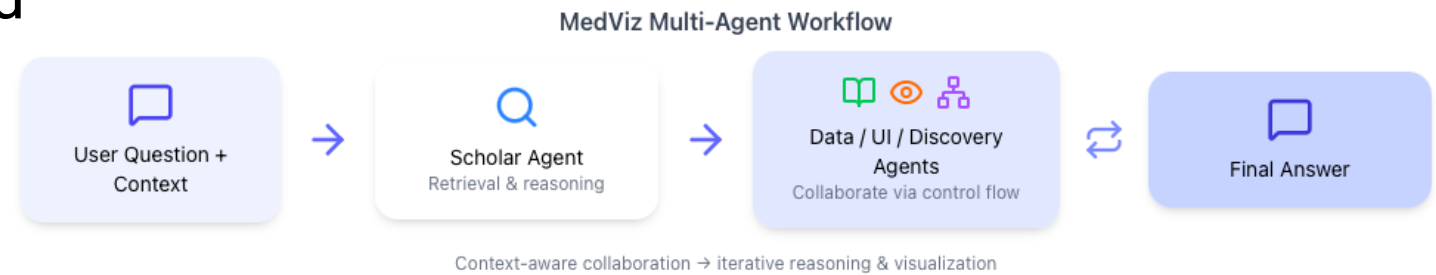
3. Interactive visual analytics interface

- Web GL rendering
- Large-scale points with interactivities



LLM + Tools

- Traditional LLM → one-shot Q&A; MedViz → **context-aware multi-agent collaboration**
- Each agent handles a specialized task (e.g., **Scholar, Discovery, Data, UI**)
- Agents combine **reasoning, retrieval, and visualization** for complex literature workflows



MedViz Visualization

1 Each paper is visualized as a point using its 2D embeddings



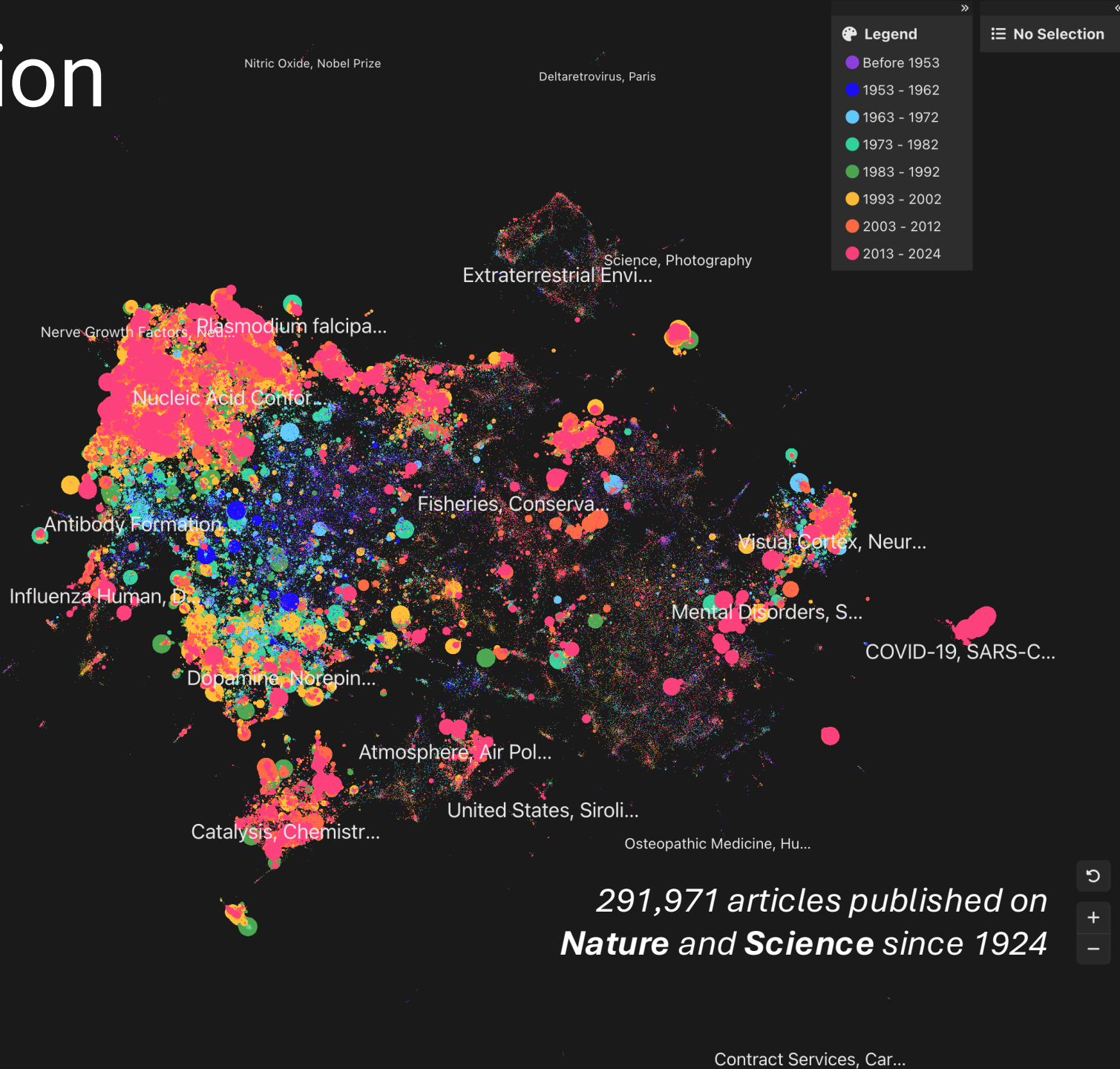
x: 0.734724,
y: 0.012341

2 Each point is colored by year or other attributes

Before 1965	Regulatory T Cell Induction and Immune Modulation
1965 - 1974	Antibody Binding and Protein Imaging Studies
1975 - 1984	Immune Biomarkers and Disease Correlations
1985 - 1994	Rare Case Reports of Malignant Lesions, Such as Ly
1995 - 2004	Immunotherapies and Combination Treatments for I
2005 - 2014	Cancer Immunotherapy Response (ICIs) and Surviv
2015 - 2024	Cellular Immune Responses and Immune Exhaustio
	Genomic Alterations and Tumor Mutational Burden
	Tumor Expression Profiles and Survival Prediction
	Gene Expression Signatures and Immune Pathways

3 Each point is resized by the number of citations or other

- Smaller point indicates fewer citations
- Bigger point indicates more citations



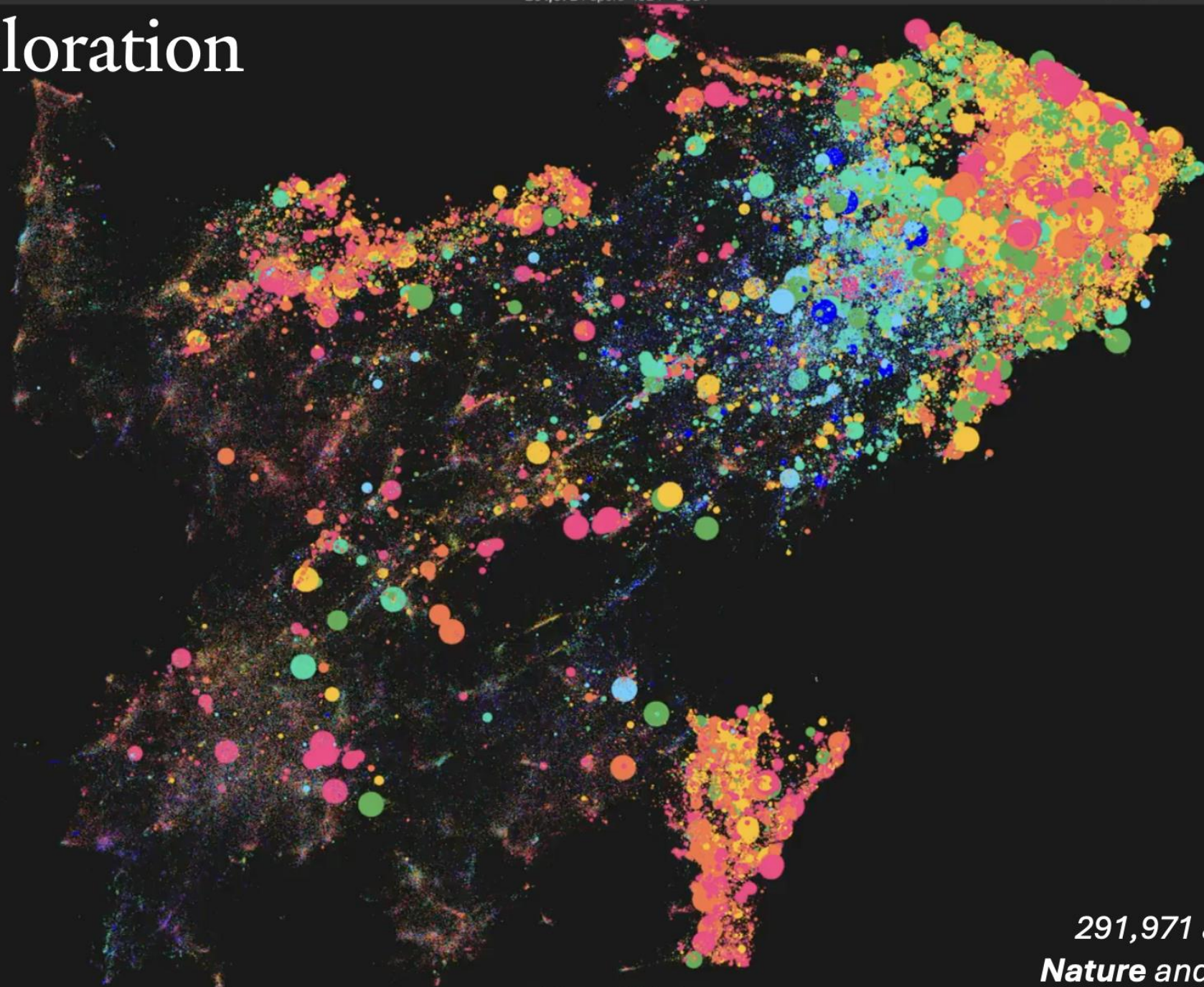
291,971 articles published on
Nature and Science since 1924

3D Exploration

Nature & Science 100 Years
291,972 Papers 1924 - 2024

Label Color Year (C... Size Size Att...

- Before 1953
 - 1953 - 1962
 - 1963 - 1972
 - 1973 - 1982
 - 1983 - 1992
 - 1993 - 2002
 - 2003 - 2012
 - 2013 - 2024
- No Selection



Layouts

2D

- Semantic
- Iris (Simple)

3D

- LargeVis 3D
- Iris 3D

291,971 articles published on
Nature and Science since 1924
-6130.70, -16440.86 | Zoom: 2.05 | Mode: view | Index: ☺

2010-Present

Human genomics and proteomics

Variant databases, mutational constraint, large-scale sequencing, phenotype-genotype links, reference genomes, etc.

Protein folding

AlphaFold, three-track network, proteome-scale prediction, protein interactions, AI-driven modeling, etc.

Plastic pollution

Plastic input quantification, pollution trajectory, microplastic health risks, mitigation gap, ocean impact, etc.

AI and deep learning

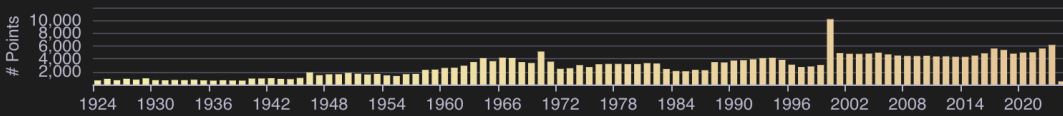
AlphaGo, deep learning foundations, ML trends, generative AI impact, machine learning, generative models, neural networks, etc.

COVID-19

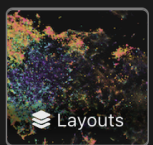
Risk factors, viral shedding, host-virus protein mapping, immune dysregulation, drug targets, clinical epidemiology, virology, host-pathogen interactions, etc.

- Before 1953
- 1953 - 1962
- 1963 - 1972
- 1973 - 1982
- 1983 - 1992
- 1993 - 2002
- 2003 - 2012
- 2013 - 2024

No Selection



291,971 articles published on
Nature and Science since 1924



Question answering

Visual Perception, ...

Mental Disorders, S...

Authors' names have 'astonishing' influence on peer reviewers.
36220906 Nature | 2022
Citations: 1.0

Calcium Co

Could you briefly summarize the themes or subtopics of the selected papers into a list?

Here are the summarized themes or subtopics of the provided papers on SARS:

1. Policy and Governance:

- China fires CDC head and launches investigations related to SARS.
- International collaboration (e.g., China and Germany) in response to the SARS outbreak.

2. Global Health Initiatives:

- WHO's call to action to maintain focus on combating SARS.

3. Frontline Response and Healthcare:

- Challenges faced by healthcare providers in managing SARS outbreaks.

4. Epidemiological Modeling:

- Efforts to understand and model the potential scope of the SARS epidemic.

5. Missed Opportunities and Lessons Learned:

- Analysis of China's missed opportunities in handling the SARS crisis.

6. Laboratory-Related Incidents:

- Investigations into lab infections associated with SARS cases.

7. Global Cooperation and Communication:

- International efforts to share resources and information in the fight against SARS.

8. Research and Scientific Discovery:

- Updates on research findings and discoveries related to the SARS virus one year after the outbreak.
- Animal studies indicating the resilience of the SARS virus.

9. Public Engagement and Awareness:

- The need for a clear agenda and public awareness regarding SARS.

These themes reflect the multifaceted nature of the response to the SARS outbreak, covering aspects from governance and policy to scientific research and public health.

Press **Enter** to send. **Shift + Enter** for a new line.

Record Send

Legend

- Before 1953
- 1953 - 1962
- 1963 - 1972
- 1973 - 1982
- 1983 - 1992
- 1993 - 2002
- 2003 - 2012
- 2013 - 2024

1 Selection

Whole Space

12 | severe acute respi...

12 severe acute respiratory syndrome + ...

Papers (12)

Prev 1 Next

SARS. China dumps CDC head, probes lab. Science (New York, N.Y.) | 2004 15247446

WHO issues rallying cry to keep fight against SARS on track. Nature | 2003 12827153 PMC7094894

Infectious diseases. Battling SARS on the frontlines. Science (New York, N.Y.) | 2003 12730562

China and Germany join forces over SARS. Nature | 2003 12815391 PMC7095333

SARS outbreak. Modelers struggle to grasp epidemic's potential scope. Science (New York, N.Y.) | 2003 12714711

SARS in China. China's missed chance. Science (New York, N.Y.) | 2003 12869735

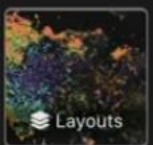
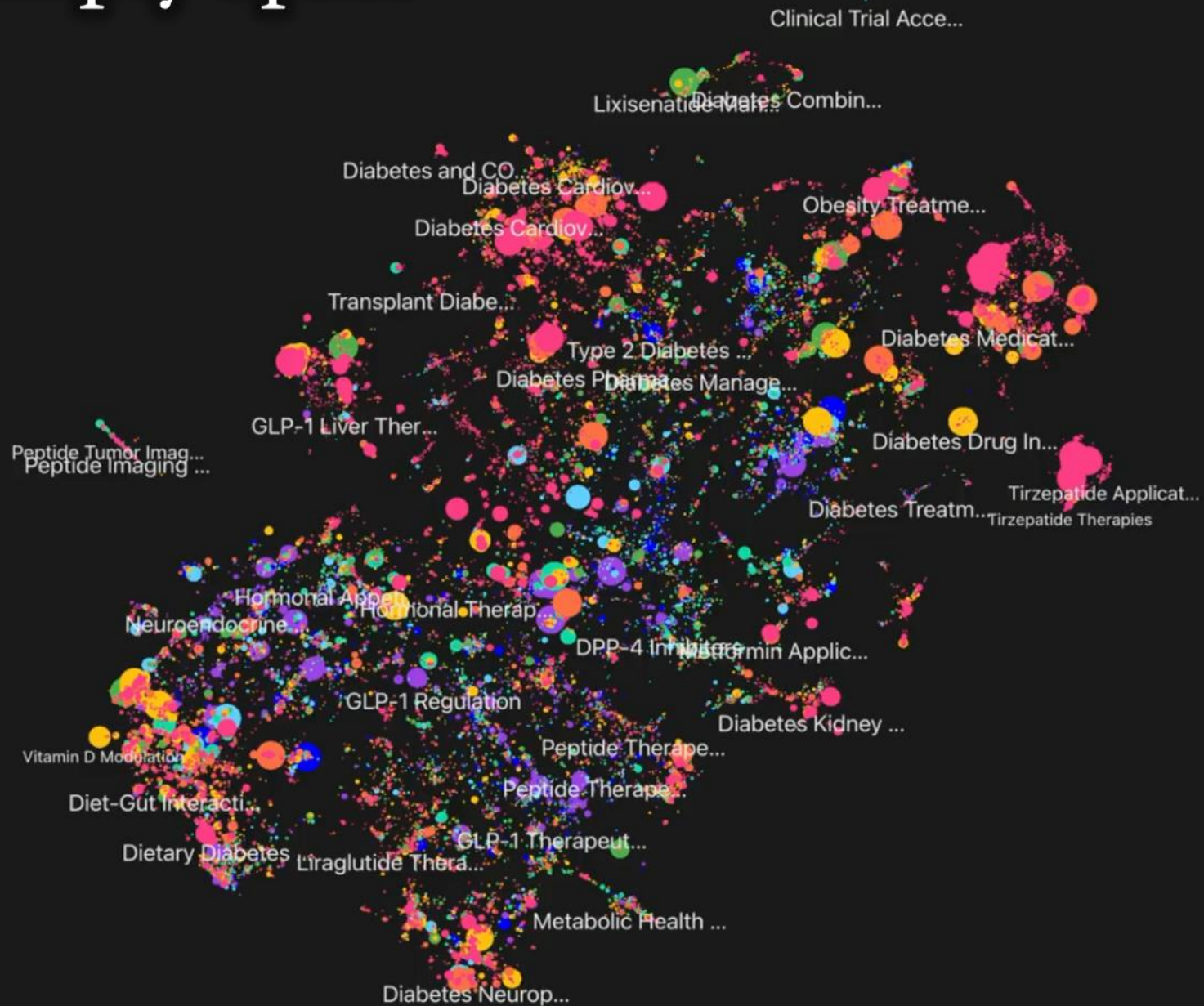
Discover Empty Space



Legend for Year (Color):

- Before 2009
- 2009 - 2010
- 2011 - 2012
- 2013 - 2014
- 2015 - 2016
- 2017 - 2018
- 2019 - 2020
- 2021 - 2024

No Selection

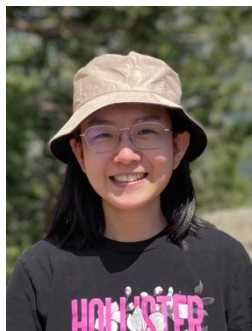


Acknowledgement

Postdocs, students, developers, designers, librarians, and collaborators from BIDS Clinical NLP Lab, YSM Medical library, and University of Michigan.



Xueqing Peng, PhD



Yutong Xie



Chia-Hsuan Chang,
PhD



Qijia Liu



Xingjian Zhang



Brian Ondov, PhD



Lei Wang



Alyssa Ann Grimshaw,
MLIS, MBA



Yujia Zhou



Daniela Weng



Chi Wing Ng



Eric Gossett



João Cardoso



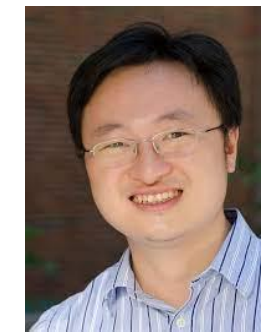
Bin Choi



Ira Einbinder



Kijana Richmond



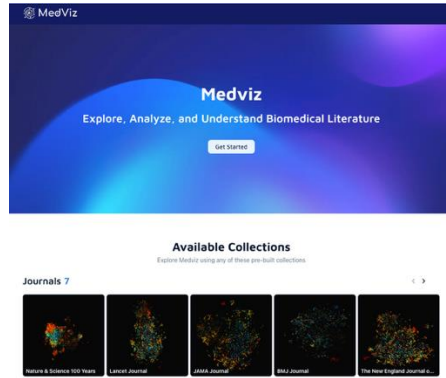
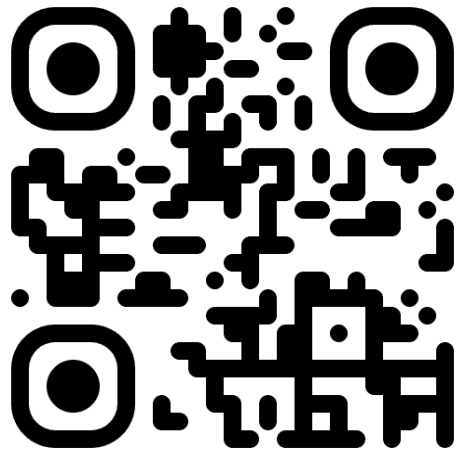
Qiaozhu Mei, PhD



Hua Xu, PhD

Questions?

Website with sample datasets

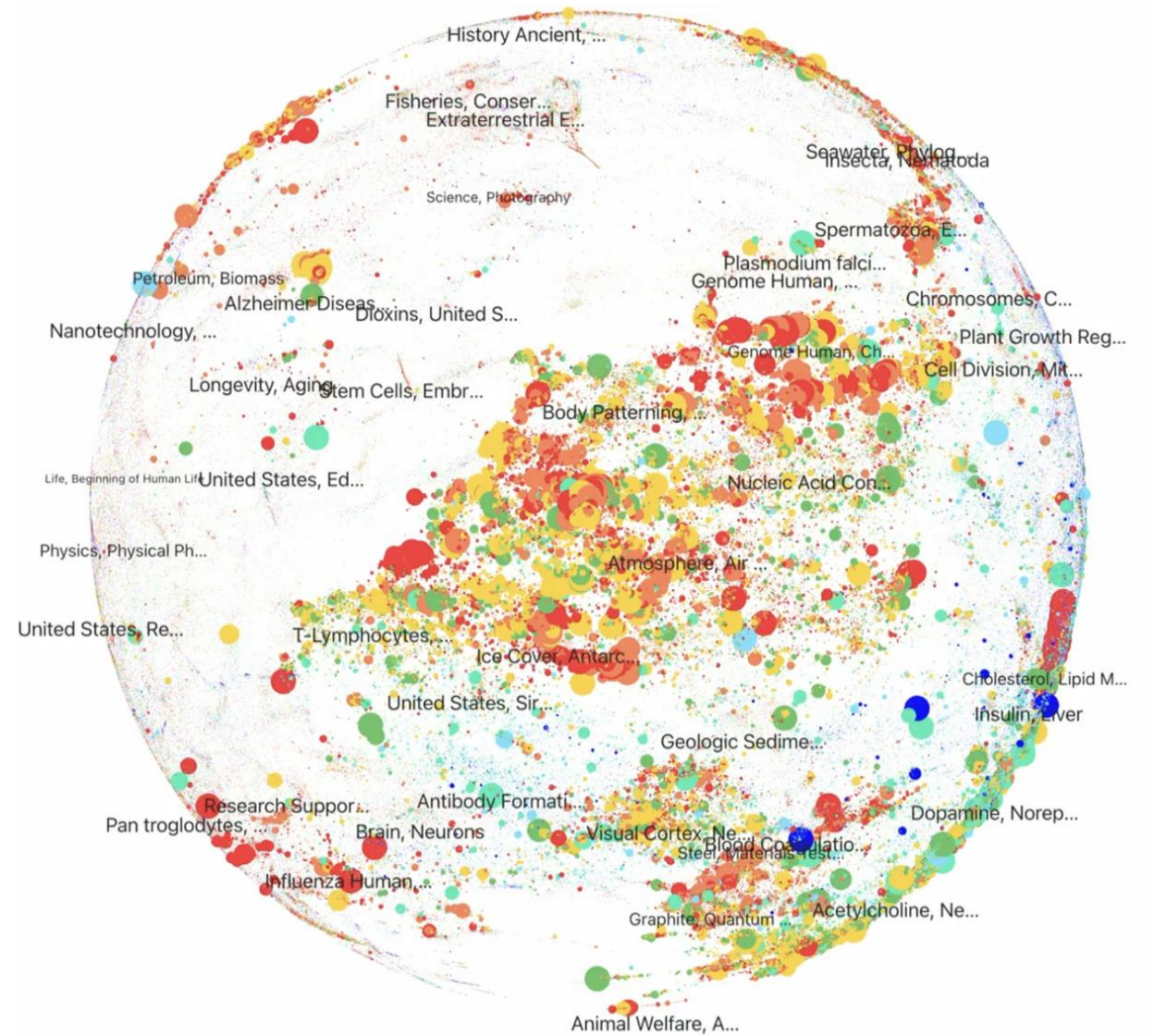


<https://medviz.org/>

Contact us

Huan He, PhD, huan.he@yale.edu

Hua Xu, PhD, hua.xu@yale.edu





INFORMATICS PROFESSIONALS. LEADING THE WAY.

