

WILEY

ENABLING DISCOVERY | POWERING EDUCATION | SHAPING WORKFORCES

AI for Scholarly Research

Enabling a Deeper, Wider, Faster and More Personalized Research Experience

將AI助力於學研 - 為研究人員實現更深入、廣泛、快速且人性化的研究體驗

Dr. Hong Zhou
Director of Intelligent Services Group
& AI R&D

Presentation Overview

Context

- Digital Transformation with ABC
- AI is an inevitable trend
- From **Content** to **Knowledge**

AI for Researchers

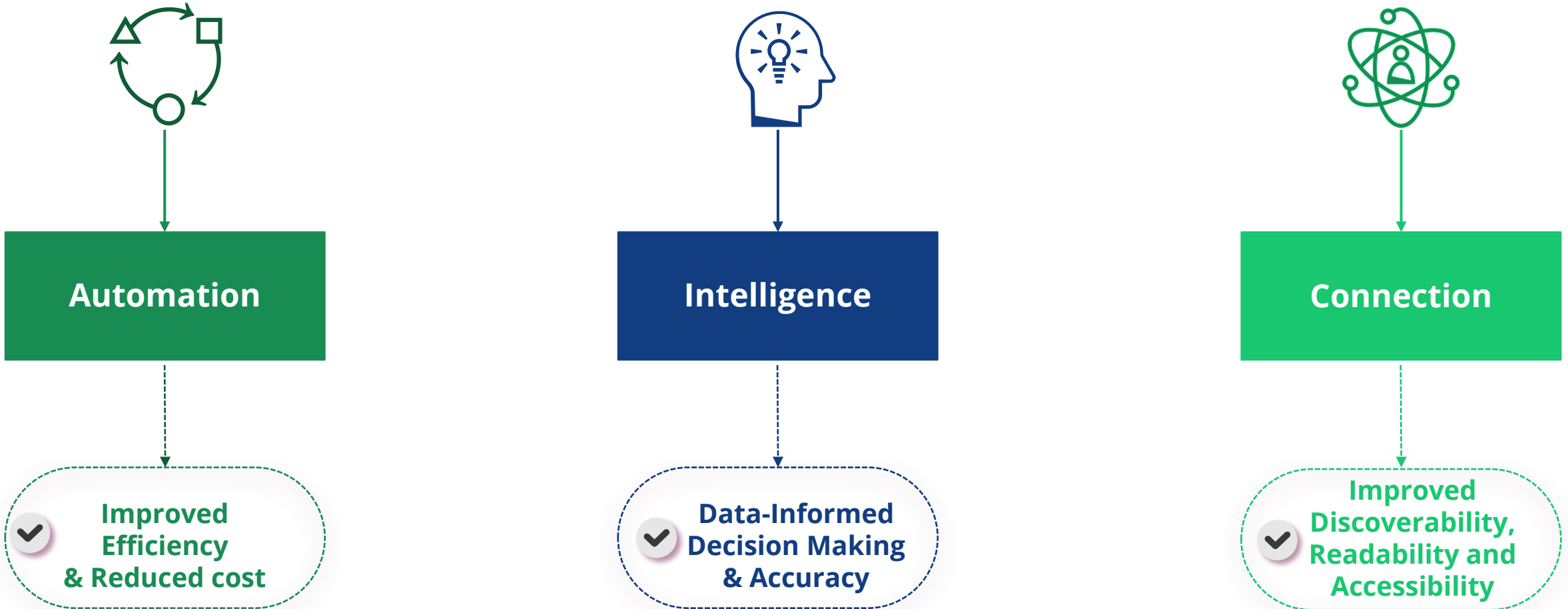
- Content **Generation**
- Information **Discovery**
- Content **Accessibility**

Future of Text & Data Mining

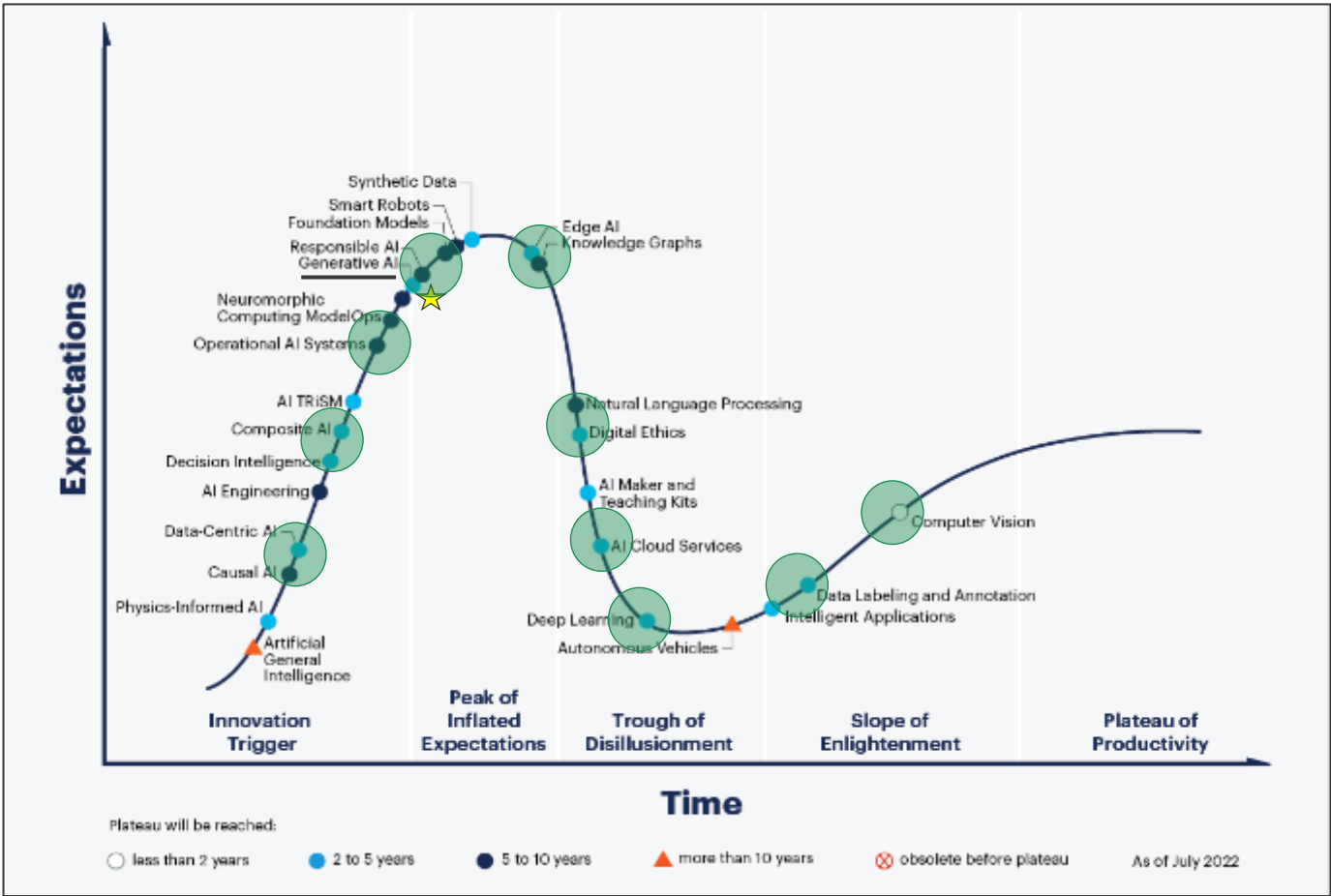
- The **Future** of AI in **Scholarly Publishing**
- How Organizations can Approach **Digital Transformation**

Digital Transformation: AI + Big Data + Cloud (ABC)

Goal: To Facilitate People (not replace them)



AI Development is an Inevitable Trend



Context:

- From Information Age to Intelligent Age
- Global Digital Transformation

3 key elements for AI:

- Big Data: **126 Million Fold Increased**
- Computing Power: **100 Million Fold Increased**
- AI Models: **1 Billion Fold Increased**

How ChatGPT is Related to Other AI Concepts

Now

Artificial Narrow Intelligence (ANI)

AI Generated Content (AIGC)

Output produced by generative AI in the form of text, images, video, music, code...



Generative AI such as ChatGPT, Bard, DALLE 2

Utilize LLMs as the underlying tech and train and finetune with more data to create content such as text, images, videos

Other Applications based on LLMs such as Language translation, Sentiment analysis, Information discovery...

Large Language Models(LLMs) such as GPT-X, BERT

Trained on vast amounts of text data and can understand human language and used for content generation, translation, summarization, classification, Q&A and more

Future Goal

Artificial General Intelligence(AGI)

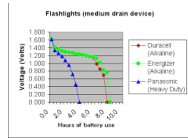
A type of AI that possesses general intelligence which is similar to human intelligence



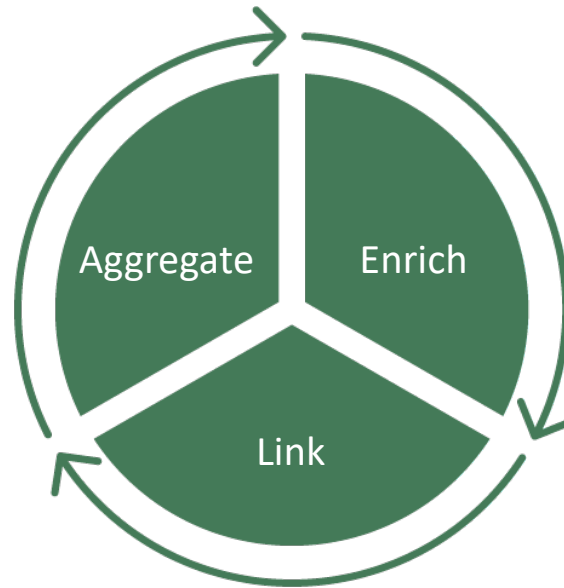
From a Content to a Knowledge

Content

Time min	Temp °C	Conductivity mS/cm	Concentration wt%	Phase Diagram wt%
0	-1.00	52.0	3.4	1.1
30	-2.00	52.1	3.4	2.2
50	-2.50	51.1	3.7	2.5
70	-2.50	62.3	4.1	2.8
90	-2.69	67.4	4.4	3.0
105	-2.53	72.7	4.8	3.2
120	-3.10	76.0	5.0	3.4

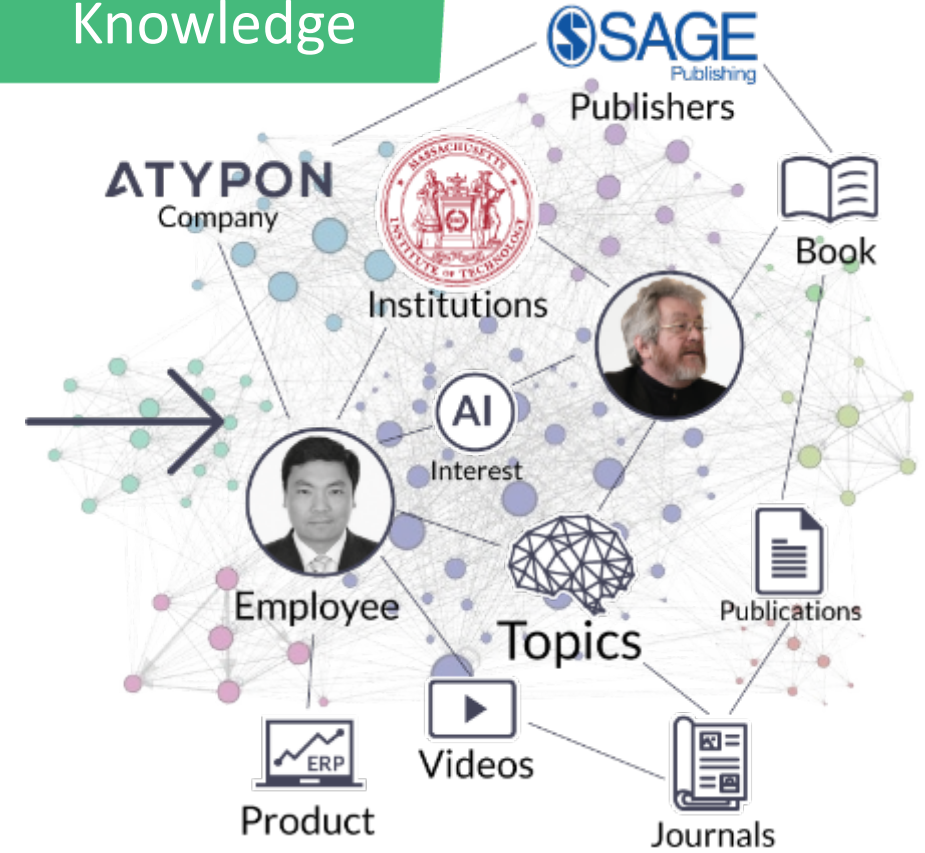


Microsoft Word document content snippet.

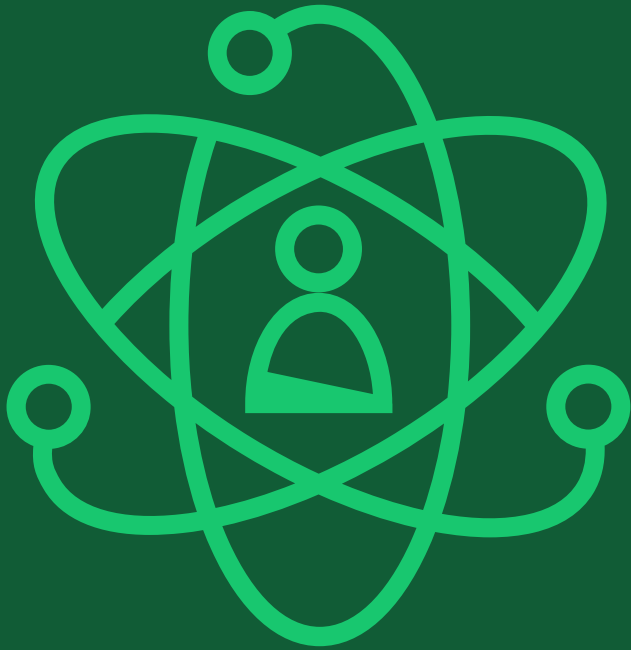


Name	UserId	PubId	PageId	ActionTime
1	251819348	41518104	1033	48:21.3
2	298382030	41219686	1033	48:22.4
3	298382032	41415282	1033	48:25.0

Knowledge



WILEY



Content Generation

Research Topic Suggestion

Ask ChatGPT/Bard to suggest research topics/directions for PhD study and for journal/SI development

Prompt: "I am a PhD student and focus on NLP area. I am really interested in large language models currently. Can you suggest me a good topic for PhD research and paper writing" and "I am a journal editor in computer science, I would like to create a new journal or special issue, please suggest some important and popular topics for the new journal and special issue to have enough submission to make the journal or special issue successful"

ChatGPT provides a list of research topics and directions for PhD study and for journal/SI development. The topics are:

1. **Multilingual and Cross-Lingual NLP (Journal):** Focus on models and algorithms that can work across multiple languages, helping bridge the gap between high-resource and low-resource languages.
2. **NLP for Low-Resource Languages (Special Issue):** There is a growing need for NLP technologies in languages that don't have as much data available. This issue could focus on novel techniques for data augmentation, transfer learning, and more for low-resource languages.
3. **Explainable NLP (Special Issue):** Understanding why NLP models make certain decisions is crucial, especially in sensitive applications. This issue could cover interpretability methods, evaluation metrics, and applications of explainable NLP.
4. **Bias and Fairness in NLP (Special Issue):** Focus on identifying and mitigating biases in language models. This includes understanding the sources of bias in NLP models and developing techniques to ensure fairness.
5. **NLP for Healthcare (Journal):** Application of NLP to extract insights from healthcare data such as clinical notes, research papers, and patient feedback.
6. **Neurosymbolic Approaches in NLP (Special Issue):** Combining deep learning with symbolic reasoning for more robust language understanding. This may include integrating knowledge graphs, logic, and reasoning into NLP models.
7. **NLP in Education (Special Issue):** Explore how NLP can be used to enhance education, such as automated grading, intelligent tutoring systems, or analyzing student writing responses for improvement.

ChatGPT give a good topic and detailed research objectives and also provide a general framework and methodology to help students to approach this.

Bard, here are some good topics for PhD research and paper writing in NLP with large language models:

- **The limitations of large language models:** Large language models have achieved state-of-the-art results in a variety of NLP tasks, but they still have limitations. For example, they can be brittle to adversarial examples, and they can be slow to train. A PhD research project could explore the limitations of large language models and develop methods to mitigate these limitations.
- **The use of large language models for new NLP tasks:** Large language models have been used for a variety of NLP tasks, such as question answering, summarization, and translation. However, there are still many other NLP tasks that could benefit from the use of large language models. A PhD research project could explore the use of large language models for new tasks, such as creative writing, code generation, or medical diagnosis.
- **The development of new methods for training large language models:** The training of large language models is a computationally expensive process. A PhD research project could develop new methods for training large language models that are more efficient and scalable. This could include techniques for data augmentation, transfer learning, or more efficient model architectures.
- **The understanding of large language models:** Large language models are complex systems, and we do not fully understand how they work. A PhD research project could focus on the understanding of large language models. This could include developing new methods for analyzing the internal workings of large language models, or could involve developing new ways to interpret the outputs of large language models.

There are just a few ideas for PhD research topics in NLP with large language models. There are many other possibilities, and the best topic for you will depend on your interests and expertise. I hope this helps!

Bard, here are some important and emerging topics in NLP that you could consider for your new journal or special issue:

- **Multilingual NLP:** As the world becomes increasingly interconnected, there is a growing need for NLP systems that can understand and process text in multiple languages. This is a challenging task, but there has been significant progress in recent years.
- **Low-code NLP:** Low-code NLP is a new approach to NLP that makes it easier for developers to build and deploy NLP applications. This is achieved by using pre-trained language models and other tools that can be combined to create powerful NLP applications without the need for extensive programming knowledge.
- **Explainable NLP:** Explainable NLP is a field of research that focuses on developing NLP systems that can explain their predictions. This is important for a number of reasons, such as ensuring that NLP systems are fair and unbiased, and for debugging and improving NLP systems.
- **Multimodal NLP:** Multimodal NLP is a field of research that focuses on developing NLP systems that can understand and process text, speech, and other modalities of information. This is a challenging task, but it has the potential to create powerful NLP systems that can understand and interact with the world in a more natural way.
- **Privacy-preserving NLP:** Privacy-preserving NLP is a field of research that focuses on developing NLP systems that can protect the privacy of users. This is important because NLP systems often process sensitive data, such as personal information.

There are just a few suggestions, and there are many other important and emerging topics in NLP that you could consider. The best topic to choose will depend on the interests and expertise of the journal or special issue editors, and what you think would be of interest to your target audience.

Bard gives more emerging and up to date ideas with less detail

Readability and Writing Quality Check and Improvement

ChatGPT can:

- Evaluate the writing quality and give reasonable feedback
- Understand specific terms and make correction (but not always)
- Re-write the manuscript with better sentence structure and words so that the readability is much improved:

Google Bard has similar capabilities, but it can also quantitatively measure readability and re-write the manuscript but not as good as ChatGPT

YI

How is the readability of following text: {In TMC, doctors write down what they did to help their patients get better. This is important for TMC to get better. The records are usually from books, cases, and the internet, but they don't have all the information. This paper talks about a way to sort the records using a computer program. First, the books are turned into computer files, and the words are put in groups. Then, a computer program is used to find out what the records are about. This helps to sort the records, and the new way is better than the old way.}



The readability of the text is fair, but it could be improved. Here's an example of how to make it more readable:

Traditional Chinese Medicine (TCM) doctors document their treatments to help their patients recover. However, records from books, cases, and the internet don't always contain all the necessary information.

To address this issue, this paper presents a novel way of sorting TCM records using a computer program. First, the books are digitized and the information is grouped into categories. Then, the computer program analyzes each record to determine its topic. This approach streamlines the sorting process and outperforms traditional methods.

In conclusion, this paper proposes an innovative approach to managing TCM records by utilizing computer technology. The new method improves efficiency and accuracy, addressing a long-standing challenge in TCM.

Title & Abstract Generation

ChatGPT can generate any length of abstract and title with logical explanation

Title: [Embracing Generative AI in Business, Research, and Society](#)

Explanation: "This title reflects the key points discussed throughout the entire text, highlighting the impact of generative AI and large language models in various aspects of our lives, from business strategies and applications to technology, ethics, and human-machine interaction. The title also emphasizes how the rapid evolution of AI is revolutionizing information discovery, education, research integrity, and AI research and development, while stressing the importance of trust, risk, and security management in AI systems."

Abstract: Generative AI is revolutionizing information discovery, domain-specific applications, and task-specific solutions, while fostering advancements in human-machine interaction. As AI becomes more accessible, it's crucial to balance its benefits with ethical and governance considerations, addressing privacy concerns and regulations. Personal thoughts on the Intelligence Revolution emphasize the potential of AI in education, research, and career development, as well as the need for maintaining research integrity and managing AI trust, risk, and security. The rapid development of AI technology presents both opportunities and challenges that individuals and organizations must adapt to and address for a successful transition into an AI-driven future.

Where to Submit – Journal Suggestion

ChatGPT and Journal Finder give some common suggestions. But dedicated Journal Finder gives the correct answer. ChatGPT and Google Bard recommend related top-tier journals only.

Based on the topic and findings of your paper, I would suggest considering the following journals for submission:

- Journal of Clinical Oncology (JCO):** JCO is a high-impact journal that focuses on clinical research in oncology. It publishes articles related to all aspects of breast cancer, including clinical trials and treatment strategies. Your study's phase 3 design and significant findings would make it a suitable fit for JCO.
- The Lancet Oncology:** The Lancet Oncology is a prestigious journal that publishes high-quality research in oncology. It covers a broad range of topics, including clinical trials, novel therapies, and treatment outcomes. Given the phase 3 trial design and the potential impact of your findings on hormone receptor-positive advanced breast cancer treatment, this journal could be a good fit.
- Breast Cancer Research and Treatment:** This journal specifically focuses on breast cancer research and treatment. It publishes clinical and translational studies, including clinical trials, therapeutic approaches, and novel targets. Since your study evaluates the efficacy and safety of capivasertib in hormone receptor-positive advanced breast cancer, it aligns well with the scope of this journal.
- Cancer:** Cancer is a highly regarded journal that publishes research in various aspects of cancer, including clinical trials and treatment outcomes. It covers a wide range of topics and welcomes studies that contribute to the understanding and management of different cancer types. Your paper's findings on the use of capivasertib in hormone receptor-positive advanced breast cancer may be of interest to the readership of Cancer.

Remember to carefully review the submission guidelines and scope of each journal to ensure your paper aligns with their focus. Additionally, consult with your co-authors or colleagues who are familiar with the field to determine the most appropriate journal for your work.

[This paper](#) is published in
The New England Journal of Medicine

ChatGPT

VS

Journal Finder

European Journal of Cancer

Publisher: Elsevier |
ISSN: 0959-8049
Relevance: 54%
Additional info: Open Access 605296

The New England Journal of Medicine

Publisher: New England Journal of Medicine |
ISSN: 0028-4793
Relevance: 53%
Additional info: Open Access 9191599

NPJ breast cancer

Publisher: Springer Nature |
ISSN: 2374-4677
Relevance: 49%
Additional info: Open Access 7506

Content Summarization

Extractive Summary: [Effects of Vitamin D on Endometriosis-Related Pain: A Double-Blind Clinical Trial](#)

- Before laparoscopy, the mean pelvic pain score in the vitamin D group was 4.05 ± 3.45 and 4.82 ± 4.1 ($p=0.513$) in the placebo group. Before laparoscopy, the mean dysmenorrhea pain score in the vitamin D group was 7.37 ± 2.61 and in placebo group it was 6.42 ± 3.04 ($p=0.325$).
- Table 2 shows a comparison between the 2 groups for severity of pelvic pain and/or dysmenorrhea at different time points (before laparoscopy, in second menses after laparoscopy, and at 24 weeks after laparoscopy). At the second menses after laparoscopy, there was no significant difference between the 2 groups for pelvic pain ($p=0.583$) and dysmenorrhea ($p=0.365$), and at 24 weeks after laparoscopy there was no significant difference between mean pain scores in the 2 groups. Mean pelvic pain at 24 weeks after laparoscopy in the vitamin D group was 0.84 ± 1.74 and in placebo group it was 0.68 ± 1.70 ($p=0.513$).
- We explored the relationship between vitamin D and endometriosis in a double-blind, randomized clinical trial looking at the effect of vitamin D supplementation on cessation of pain in proven endometriosis after laparoscopic diagnosis and treatment.
- There may be a relationship between vitamin D and pathogenesis of endometriosis, but in our study vitamin D was not effective in treatment of endometriosis-related pain.
- In this double-blind, randomized clinical trial, at 24 weeks after laparoscopic treatment of endometriosis there was no significant difference between effect of vitamin D3 (cholecalciferol) and placebo on severity of dysmenorrhea and/or pelvic pain.
- The remaining 39 cases were randomly assigned in vitamin D ($n=19$) or placebo treatment ($n=20$) groups.
- After authorization by the university Ethics Committee, eligible patients were assigned by simple randomization to receive either vitamin D or placebo. In the vitamin D group (D group), we prescribed oral vitamin D 50 000 iu/weekly for 12 weeks (capsule D-Vigel, vitamin D3 50 000 iu, Daana Pharma Co. Tabriz-Iran) and in the placebo group (P group) we prescribed 1 capsule of placebo (Daana Pharma Co. Tabriz-Iran) weekly for 12 weeks.
- Mean dysmenorrhea was 2.10 ± 2.33 in the vitamin D group and 2.73 ± 2.84 in the placebo group ($p=0.45$).

ChatGPT Summary

This paper presents the results of a double-blind, randomized clinical trial that investigated the effect of vitamin D supplementation on the cessation of pain in women with endometriosis. The study involved 39 women with proven endometriosis who were randomly assigned to receive either vitamin D or placebo treatment. The severity of dysmenorrhea (menstrual pain) and pelvic pain was assessed using a visual analogue scale (VAS) test before laparoscopy, at the second menstrual cycle after laparoscopy, and at 24 weeks after laparoscopy. The results showed that there was no significant difference in pain severity between the vitamin D and placebo groups at 24 weeks after laparoscopy. This study is the first randomized clinical trial to explore the potential relationship between vitamin D treatment and relief of endometriosis-related pain. However, it should be noted that the sample size was small, and the study did not measure the participants' vitamin D levels before intervention. Therefore, the authors suggest that the observed lack of significant difference in pain severity could be attributed to the possibility that the prescribed dose and duration of vitamin D supplementation were only beneficial for treating vitamin D deficiency rather than endometriosis itself.

Extraction + ChatGPT Summary

This paper presents the findings of a double-blind, randomized clinical trial that aimed to explore the relationship between vitamin D supplementation and endometriosis-related pain. The study focused on the effect of vitamin D on the cessation of pain in individuals with proven endometriosis after laparoscopic diagnosis and treatment. A total of 39 eligible patients were randomly assigned to either the vitamin D group ($n=19$) or the placebo group ($n=20$). The study measured the severity of pelvic pain and dysmenorrhea before laparoscopy in both groups, finding no significant difference between the two groups. Similar results were observed at the second menses after laparoscopy and at 24 weeks after laparoscopy, indicating that vitamin D supplementation did not have a significant impact on pain reduction compared to the placebo group. Mean pain scores for pelvic pain and dysmenorrhea at 24 weeks after laparoscopy were comparable between the vitamin D and placebo groups. Specifically, the mean pelvic pain score was 0.84 ± 1.74 in the vitamin D group and 0.68 ± 1.70 in the placebo group ($p=0.513$). Furthermore, the mean dysmenorrhea score was 2.10 ± 2.33 in the vitamin D group and 2.73 ± 2.84 in the placebo group ($p=0.45$). Although there is a potential relationship between vitamin D and the pathogenesis of endometriosis, the results of this study indicate that vitamin D supplementation was not effective in treating endometriosis-related pain. Therefore, the findings suggest that alternative treatment approaches should be considered for managing pain in individuals with endometriosis.

ChatGPT produces better and more fluent summary than Bard. But it is much more expensive and generates less detail than the extractive summary

Deeper Information Discovery

Challenge

Information Overload

Structured Knowledge
Embedded in Unstructured
Text

Solution

Key Entity Identification &
Extraction

Knowledge Mining & Search

Specific Entities Mining

Knowledge Mining takes valuable information from customers' existing content to create a more structured layout and generate new business opportunities.

Existing Content

Inogatran, which is a polar, cationic molecule with a molecular weight of 439 Da, is expected to be excreted in the bile [14]. Inogatran given after thrombolysis only partially prevents reocclusion because large amounts of thrombin generation occur during the early stages of thrombolysis. We investigated the micellar behavior of a series of asymmetric polystyrene-block-poly (4-vinylpyridine) (PS-b-P4VP) block copolymers in different P4VP-selective alcoholic solvents. The PS-b-P4VP block copolymer used had a molecular weight of 59 000 g mol⁻¹, where the volume fraction of PS block was roughly 0.3. The time-resolved (in situ) dynamic light scattering (DLS) measurements, in combination with (ex situ) electron microscopy imaging, revealed information about the aggregation state of PS-b-P4VP BCP in different alcohols and the effect of heat treatment. These are determined from the binding energies of CH₄, C₂H₂ and C₂H₆. Two sets of experimental values are considered according to the two possible (as yet undecided) values of the heat of evaporation of diamond of 141 and 170 kcal/mol. After the values of the integrals are fixed the binding energy of C₂H₄ and the dissociation energies CH₄ → CH₃ + H, C₂H₂ → C₂H + H and C₂H₆ → C₂H₅ + H can be calculated and compared with the experiments (Table 1).

Structured Data

#	Compound	Property	Value	Unit
1				
2				
3				

Knowledge Mining

Ask ChatGPT and Bard to extract the semantic triples from unstructured text

Can you extract triplets in the following sentences: {
Our previous work showed that high levels of lysophosphatidic acid (LPA) expression are positively correlated with LF hypertrophy. <https://biocreative.bioinformatics.udel.edu/tasks/biocreative-v/track-3-cdr/>
Men and women with a mutation in MLH1 have a 52-82% lifetime risk (up to age 70) to develop colon or rectal cancer. <https://geneticassociationdb.nih.gov/>
However, while cisplatin increased PID1 mRNA, it decreased PID1 protein. (Chemical-Gene) <https://biocreative.bioinformatics.udel.edu/>

ChatGPT

Certainly! Here are the triples with the entities and their corresponding relations found in the given sentences:

- Triple 1:
 - Entity 1: Lysophosphatidic acid (LPA)
 - Relation: Positively correlated with
 - Entity 2: LF hypertrophy
- Triple 2:
 - Entity 1: Mutation in MLH1
 - Relation: Lifetime risk of
 - Entity 2: Colon or rectal cancer
- Triple 3:
 - Entity 1: Cisplatin
 - Relation: Increased
 - Entity 2: PID1 mRNA
 - Relation: Decreased
 - Entity 3: PID1 protein

These triples represent the relationships between the entities mentioned in the given sentences, with each triple consisting of an entity, a relation, and another entity.

ChatGPT performs well in this experiment, it can even extract 2 triplets from the 3rd examples

Bard

Sure, I can do that. Here are the triplets extracted from the sentences:

- Sentence 1:
 - Our previous work - showed - high levels of lysophosphatidic acid(LPA) expression
 - high levels of lysophosphatidic acid(LPA) expression - are positively correlated with - LF hypertrophy
- Sentence 2:
 - Men and women - have - a 52-82% lifetime risk(up to age 70) to develop colon or rectal cancer
 - mutation in MLH1 - associated with - 52-82% lifetime risk(up to age 70) to develop colon or rectal cancer
- Sentence 3:
 - cisplatin - increased - PID1mRNA
 - PID1mRNA - decreased - PID1 protein

Bard perform even better to identify not only all domain specific triplets but also generic triplets

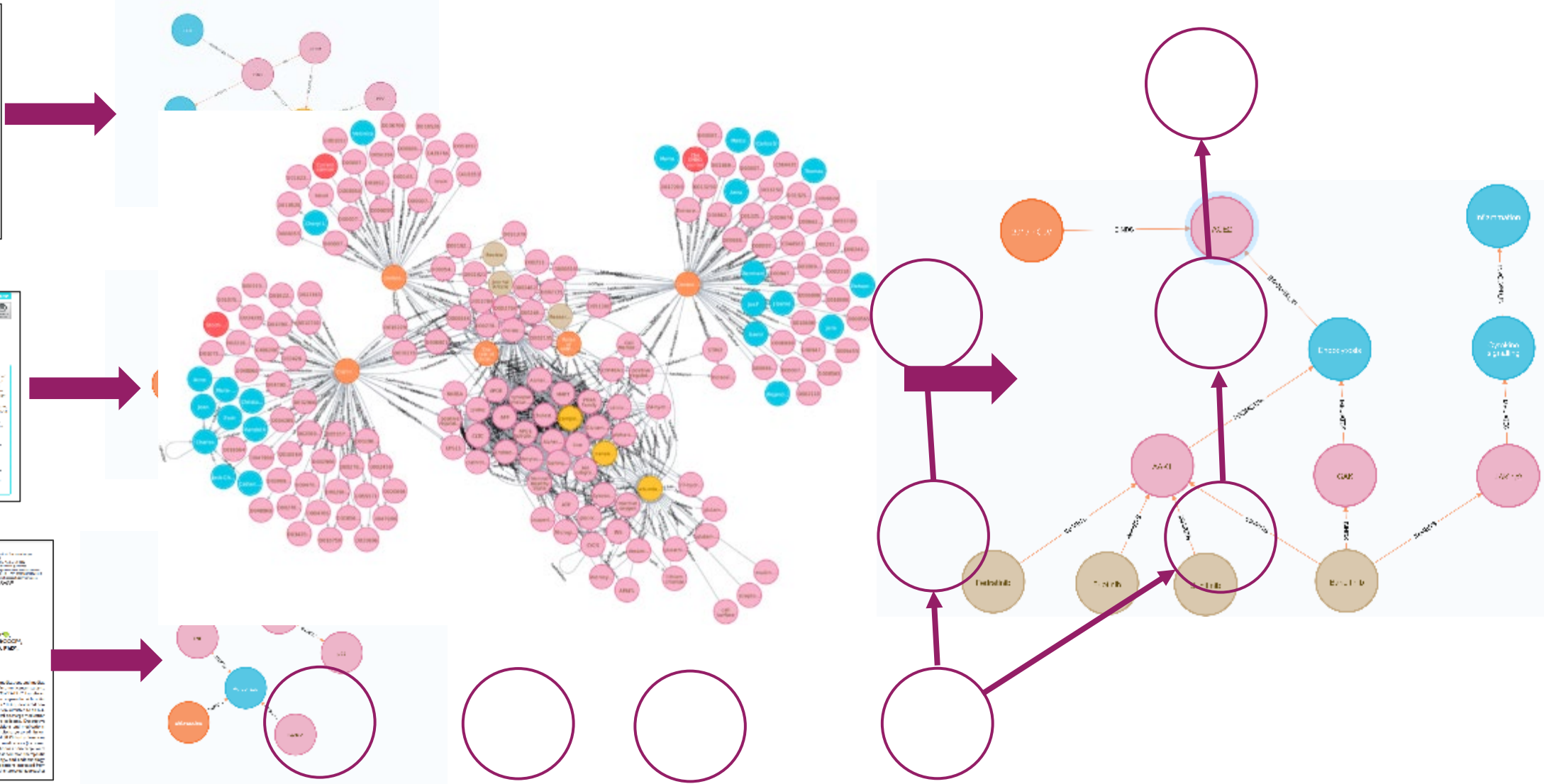
VS.

From Content to Knowledge

Interventions to control the disease
 "Interventions to control the disease" - A list of interventions and their effectiveness in controlling the disease. The text is dense and includes various medical terms and statistics.

Systemic identification of ACE2
 "Systemic identification of ACE2 expression modulators reveals cardiomyopathy as a risk factor for mortality in COVID-19 patients" - A research paper abstract discussing the identification of ACE2 expression modulators and their role in COVID-19 mortality.

Medication Changes and Diagnosis
 "Medication Changes and Diagnosis of New-Onset Hypoxia in COVID-19 Intensive Care Unit Survivors" - A study abstract focusing on medication changes and the diagnosis of new-onset hypoxia in COVID-19 ICU survivors.



Personalized and Richer Information Discovery

Challenge

Open Science Movement
with Rich Data

Researchers Must Find
Answers by Themselves from
a Long List of Search Results

Solution

Multimedia Search

Personalized
Recommendation

Natural Language Q&A

Multimedia Content Discovery

Multimedia content discovery includes topic, image, video and funder searches:

The screenshot shows a search interface with a search bar containing 'NIH' and a search button. Below the search bar, there are two tabs: 'Articles (93565)' and 'Databases (211)'. The 'Articles' tab is selected. Below the tabs, there are options for 'Download citations' and 'Download PDFs', and a dropdown menu for 'Order by Relevance' set to 'Relevance' and a dropdown for '10 per page'. The search results are displayed as a list of articles. The first article is titled 'Research misconduct, NSF v NIH: Its nature and prevalence and the impact of their respective methods of investigation and adjudication' by Donald S. Kornfeld, published in 'Accountability in Research, Volume 26, 2019 - Issue 6'. The second article is titled 'Is there a trade-off between academic research and faculty entrepreneurship? Evidence from US NIH supported biomedical researchers' by Dirk Czarnitzki & Andrew A. Toole, published in 'Economics of Innovation and New Technology, Volume 19, 2010 - Issue 5'.

Funder [Advanced search](#)

Articles (93565) Databases (211)

[Download citations](#) [Download PDFs](#)

[Research misconduct, NSF v NIH: Its nature and prevalence and the impact of their respective methods of investigation and adjudication](#) >

Donald S. Kornfeld

Accountability in Research, Volume 26, 2019 - Issue 6

Article | Published Online: 20 Aug 2019 | Views: 363 | Citations: 2

[Abstract](#) [Full Text](#) >

[Is there a trade-off between academic research and faculty entrepreneurship? Evidence from US NIH supported biomedical researchers](#) >

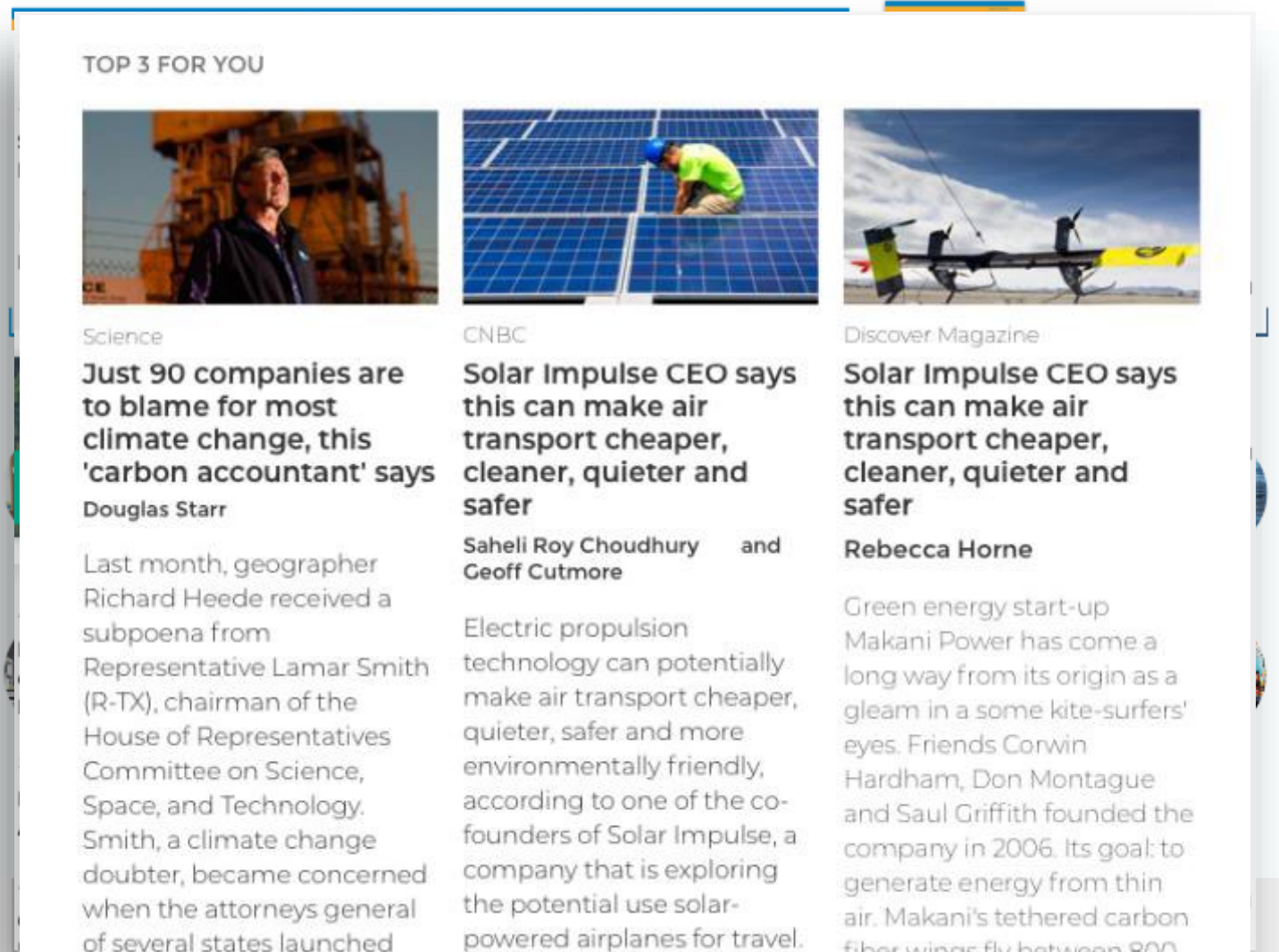
Dirk Czarnitzki & Andrew A. Toole

Economics of Innovation and New Technology, Volume 19, 2010 - Issue 5


Article | Published Online: 14 Jul 2010 | Views: 264 | Citations: 15

Recommendations & Personalized News Feeds

Recommendations is one of the most common applications of AI. It varies from suggesting relevant experts, to identifying interest groups and enable a personalized user experience based on individual interests.



TOP 3 FOR YOU




Science

Just 90 companies are to blame for most climate change, this 'carbon accountant' says

Douglas Starr

Last month, geographer Richard Heede received a subpoena from Representative Lamar Smith (R-TX), chairman of the House of Representatives Committee on Science, Space, and Technology. Smith, a climate change doubter, became concerned when the attorneys general of several states launched




CNBC

Solar Impulse CEO says this can make air transport cheaper, cleaner, quieter and safer

Saheli Roy Choudhury and Geoff Cutmore

Electric propulsion technology can potentially make air transport cheaper, quieter, safer and more environmentally friendly, according to one of the co-founders of Solar Impulse, a company that is exploring the potential use solar-powered airplanes for travel.



Discover Magazine

Solar Impulse CEO says this can make air transport cheaper, cleaner, quieter and safer

Rebecca Horne

Green energy start-up Makani Power has come a long way from its origin as a gleam in a some kite-surfers' eyes. Friends Corwin Hardham, Don Montague and Saul Griffith founded the company in 2006. Its goal: to generate energy from thin air. Makani's tethered carbon fiber wings fly between 800

Reviewer Suggestion

Journal editor ask AI to suggest some reviewers for giving title and abstract of [paper](#) as shown below

I am a journal editor in computer science area and I have following paper with title:{Bilinear joint learning of word and entity embeddings for Entity Linking} and abstract:{Entity Linking (EL) is the task of resolving mentions to referential entities in a knowledge base, which facilitates applications such as information retrieval, question answering, and knowledge base population. In this paper, we propose a novel embedding method specifically designed for EL. The proposed model jointly learns word and entity embeddings which are located in different distributed spaces, and a bilinear model is introduced to simulate the interaction between words and entities. We treat EL as a ranking problem, and utilize a pairwise learning-to-rank framework with features constructed with learned embeddings as well as conventional EL features. Experimental results show the proposed model produces effective embeddings which improve the performance of our EL algorithm. Our method yields the state-of-the-art performances on two benchmark datasets CoNLL and TAC-KBP 2010.}. Can you give me recommendations about reviewers for this paper?

ChatGPT

Certainly, finding the right reviewers for your paper is crucial for ensuring a fair and thorough review process. Given the topic of your paper, "Bilinear Joint Learning of Word and Entity Embeddings for Entity Linking" in the field of computer science, especially focusing on entity linking, you should look for experts who have a strong background in natural language processing (NLP), machine learning, and entity linking. Here are some recommendations for potential reviewers:

- Professor Jane Smith**
 - Affiliation: Department of Computer Science, University of XYZ
 - Expertise: NLP, Machine Learning, Entity Linking
 - Contact: janesmith@email.com
- Dr. John Doe**
 - Affiliation: Research Scientist, AI Research Lab, ABC Corporation
 - Expertise: NLP, Deep Learning, Information Retrieval
 - Contact: john.doe@email.com
- Professor Emily Brown**
 - Affiliation: Department of Information Sciences, University of LMN
 - Expertise: Information Retrieval, Knowledge Base Population, NLP
 - Contact: emily.brown@gmail.com
- Dr. David Johnson**
 - Affiliation: Senior Research Scientist, Institute for Machine Learning, DEF Institute
 - Expertise: Machine Learning, Embedding Models, NLP
 - Contact: davidjohnson@gmail.com

ChatGPT gives non exist reviewers

Bard

- Michael Berzheim: Berzheim is a professor of computer science at Stanford University. He is an expert in natural language processing and information retrieval, and has published over 100 papers in top academic journals and conferences.

- Yoo Myeong Kim: Kim is a professor of computer science at the National University of Singapore. He is an expert in machine learning and natural language processing, and has published over 100 papers in top academic journals and conferences.

- Kristina Toutanova: Toutanova is a research scientist at Google AI. She is an expert in natural language processing and machine learning, and has published over 100 papers in top academic journals and conferences.


Bard returns the authors with photos and source. But some affiliation info are incorrect

Our Own Reviewer Finder

Yoshiyuki Takafuji
Affiliation: [University of Tsukuba](#)
Expertise: NLP, Machine Learning, Entity Linking
Contact: yoshiyuki.takafuji@u.tsukuba.ac.jp

Joint Learning of the Embedding of Words and Entities for Named Entity Disambiguation
DOI: [10.1007/978-3-319-91313-5_15](https://doi.org/10.1007/978-3-319-91313-5_15)

Joint Learning of the Embedding of Words and Entities for Named Entity Disambiguation
DOI: [10.1007/978-3-319-91313-5_15](https://doi.org/10.1007/978-3-319-91313-5_15)

Learning Distributed Representations of Facts and Entities from Knowledge Base
DOI: [10.1007/978-3-319-91313-5_15](https://doi.org/10.1007/978-3-319-91313-5_15)

Hiroaki Takada
Affiliation: [University of Tsukuba](#)
Expertise: NLP, Machine Learning, Entity Linking
Contact: hiroaki.takada@u.tsukuba.ac.jp

Shuangshu Luo
Affiliation: [University of Tsukuba](#)
Expertise: NLP, Machine Learning, Entity Linking
Contact: shuangshu.luo@u.tsukuba.ac.jp

A Novel Approach for Analyzing Entity Linking Between Words and Entities for a Knowledge Base Weighted Summation Model
DOI: [10.1007/978-3-319-91313-5_15](https://doi.org/10.1007/978-3-319-91313-5_15)

Ours returns real relevant researchers with h-index, ORCID, relevancy score and relevant publication

Both **Bard** and **ChatGPT** don't perform well in this experiment and their results contain serious issues. **Dedicated reviewer suggestion** service give more reliable results with richer info!

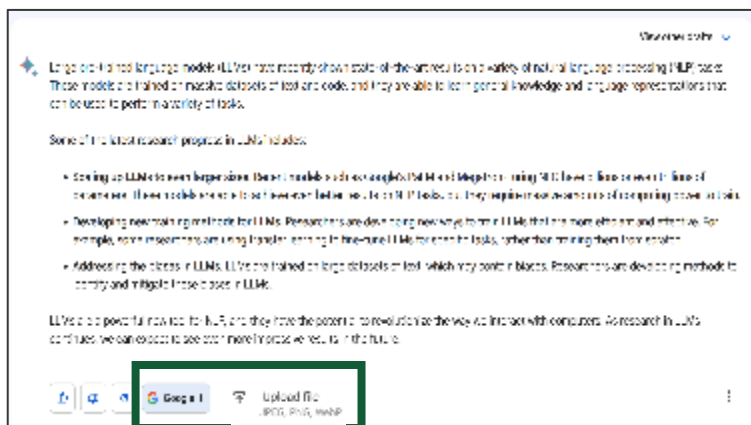
New Way of Search

Question: What is the latest study progress about large pre-trained language models?

For the usage of scholarly research-related cases, Bing gives a better result than the two others. Both Bard and ChatGPT limit their answers to their data while Bing generates the result based on the web search and gives the related articles with links as well (which are real latest articles)

Google Bard

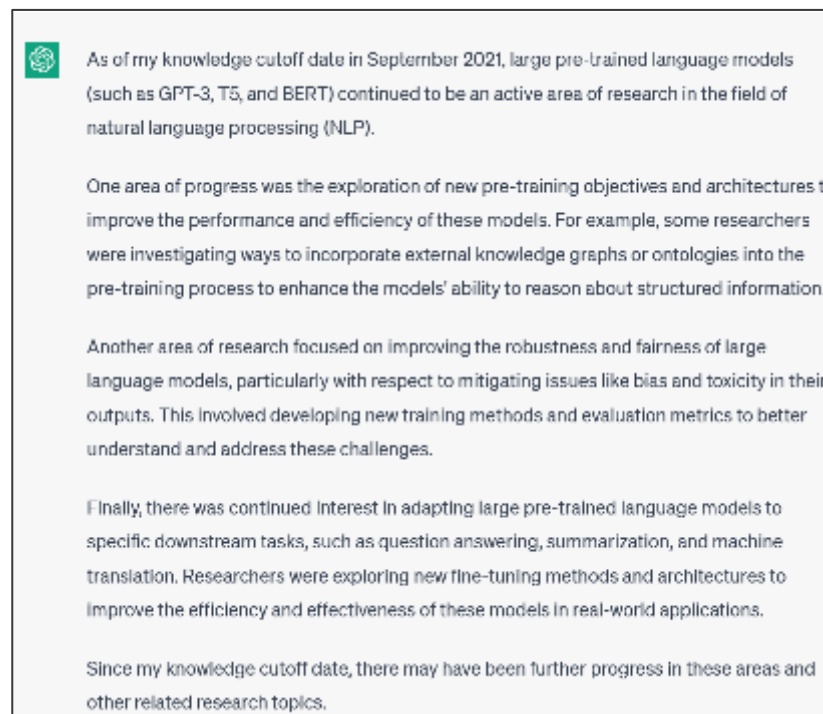
Add search in AI



Bard currently search in 135 languages, and it can **support image search**

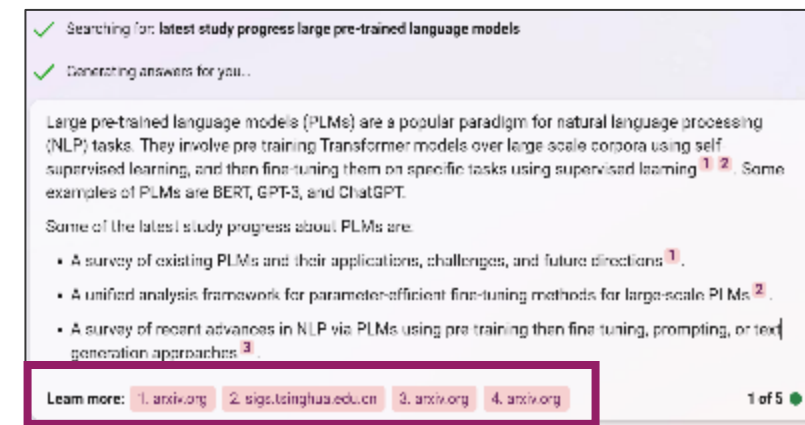
ChatGPT

Generate answer with knowledge before 2021



Microsoft Bing

Add AI in search



English search is better than Chinese search in Bing and **less hallucination**

Personalized Recommendations

VI I am a research scientist in NLP area. I am currently interested in large language models like GPT-3. Can you recommend some relative papers about this area

Certainly! Here are some papers related to large language models like GPT-3 that you may find useful:

1. "Language Models are Few-Shot Learners" by Tom B. Brown et al. (2020) - This paper introduces GPT-3, a large-scale language model that can perform a wide range of natural language processing tasks without task-specific training data.
2. "Scaling Laws for Neural Language Models" by Andrew M. Dai et al. (2021) - This paper investigates the scaling properties of neural language models and proposes a new metric for evaluating their performance.
3. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding" by Jacob Devlin et al. (2019) - This paper introduces BERT, a pre-trained language model that achieves state-of-the-art performance on a wide range of natural language processing tasks.
4. "XLNet: Generalized Autoregressive Pretraining for Language Understanding" by Zhilin Yang et al. (2019) - This paper introduces XLNet, a pre-trained language model that achieves state-of-the-art results on several natural language processing benchmarks.
5. "Reformer: The Efficient Transformer" by Nikita Kitaev et al. (2020) - This paper proposes the Reformer, an efficient variant of the Transformer architecture that can handle longer sequences than previous models.

I hope these papers help you in your research!

Both **ChatGPT** and **Bard** can provide personal recommendation based on conversations/prompts.

But they sometimes make up answers as shown below

VI Can you recommend some latest papers

Sure, here are some of the latest papers related to large language models and natural language processing:

1. "GShard: Scaling Giant Models with Conditional Computation and Automatic Sharding" by Noam Shazeer et al. (2021) - This paper introduces GShard, a new approach for scaling up giant language models that uses conditional computation and automatic sharding.
2. "Simplifying Pre-training of Large Language Models with Multiple Causal Language Modeling" by Sheng Shen et al. (2021) - This paper proposes a simplified pre-training approach that uses multiple causal language modeling objectives and achieves competitive performance on several natural language processing benchmarks.
3. "AdaM: Adapters for Modular Language Modeling" by Timo Schick et al. (2021) - This paper introduces AdaM, a new approach for modular language modeling that uses adapters to enable efficient fine-tuning of large pre-trained models.
4. "Hierarchical Transformers for Long Document Processing" by Zihang Dai et al. (2021) - This paper proposes a hierarchical transformer architecture for processing long documents, which achieves state-of-the-art performance on several document-level NLP tasks.
5. "Scaling up Sparse Feature Learning with Mega-Learning" by Jiawei Xu et al. (2022) - This paper introduces Mega-Learning, a new training technique for sparse feature learning that enables the efficient training of very large models with billions of parameters.

I hope these papers are helpful for your research!

Wider Information Discovery

Challenge

Data Silos

Need for Cross-
Disciplinary Research

Solution

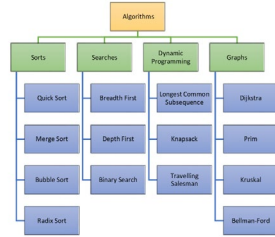
Classifications

Milestone Paper

Research Analytics

Content Classification

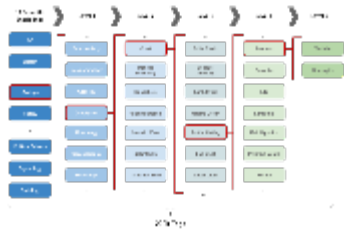
PUBLISHER-SPECIFIC TAXONOMIES



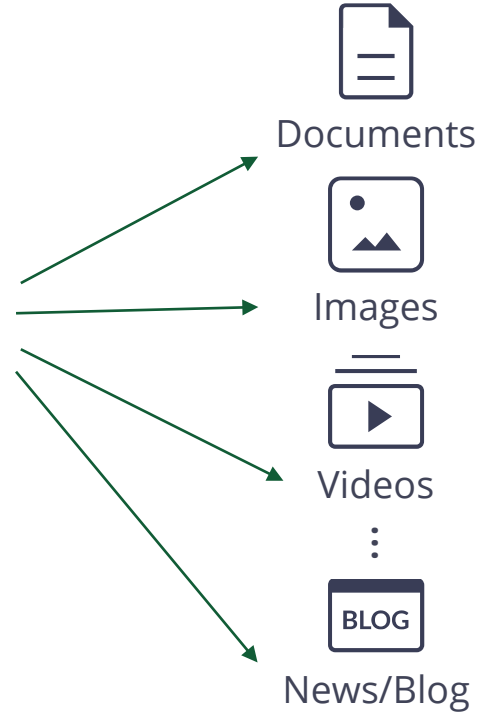
PUBLIC TAXONOMIES (MeSH, UN SDG)



GLOBAL TAXONOMY



AUTO TAGGER



Topic 1



Topic 2



Content Classification

ChatGPT can **NOT** tag content based on customer specific taxonomy and it often generates fake tags or IDs

“Evaluation of BNT162b2 Covid-19 Vaccine in Children Younger than 5 Years of Age

Although coronavirus disease 2019 (Covid-19) is generally mild in children younger than 5 years of age, severe disease, hospitalizations, and post-acute effects, including multisystem inflammatory syndrome in children (MIS-C), can occur.¹⁻³ In the United States, rates of Covid-19–associated hospitalization among children younger than 5 years of age peaked at...”

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9947923/>

Atypon Taxonomy tags

1. Medicine: 0.98
2. Pediatrics: 0.95
3. Pandemic: 0.93
4. Immunization: 0.93
5. Clinical trial: 0.92
6. Booster: 0.73
7. Coronavirus disease 2019: 0.65
8. Emergency use authorization: 0.6

Dedicated MeSH Auto Tagger

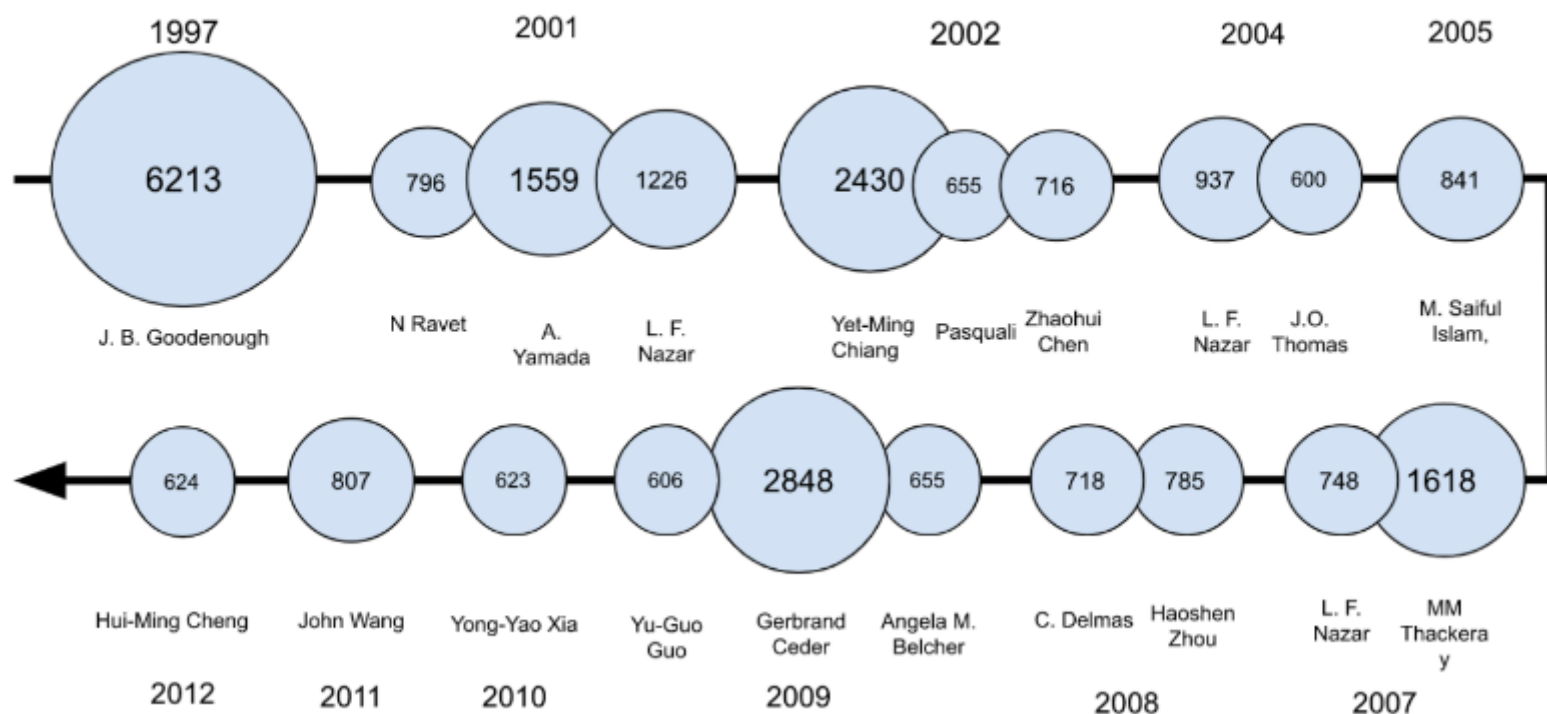
1. COVID-19: 0.99
2. Child, Preschool: 0.97
3. Pandemics: 0.93
4. Coronavirus Infections: 0.91
5. Infant: 0.89
6. COVID-19 Vaccines: 0.87
7. Viral Vaccines: 0.85
8. Pneumonia, Viral: 0.83
9. Hospitalization: 0.77
10. SARS-CoV-2: 0.69

ChatGPT MeSH Tagging

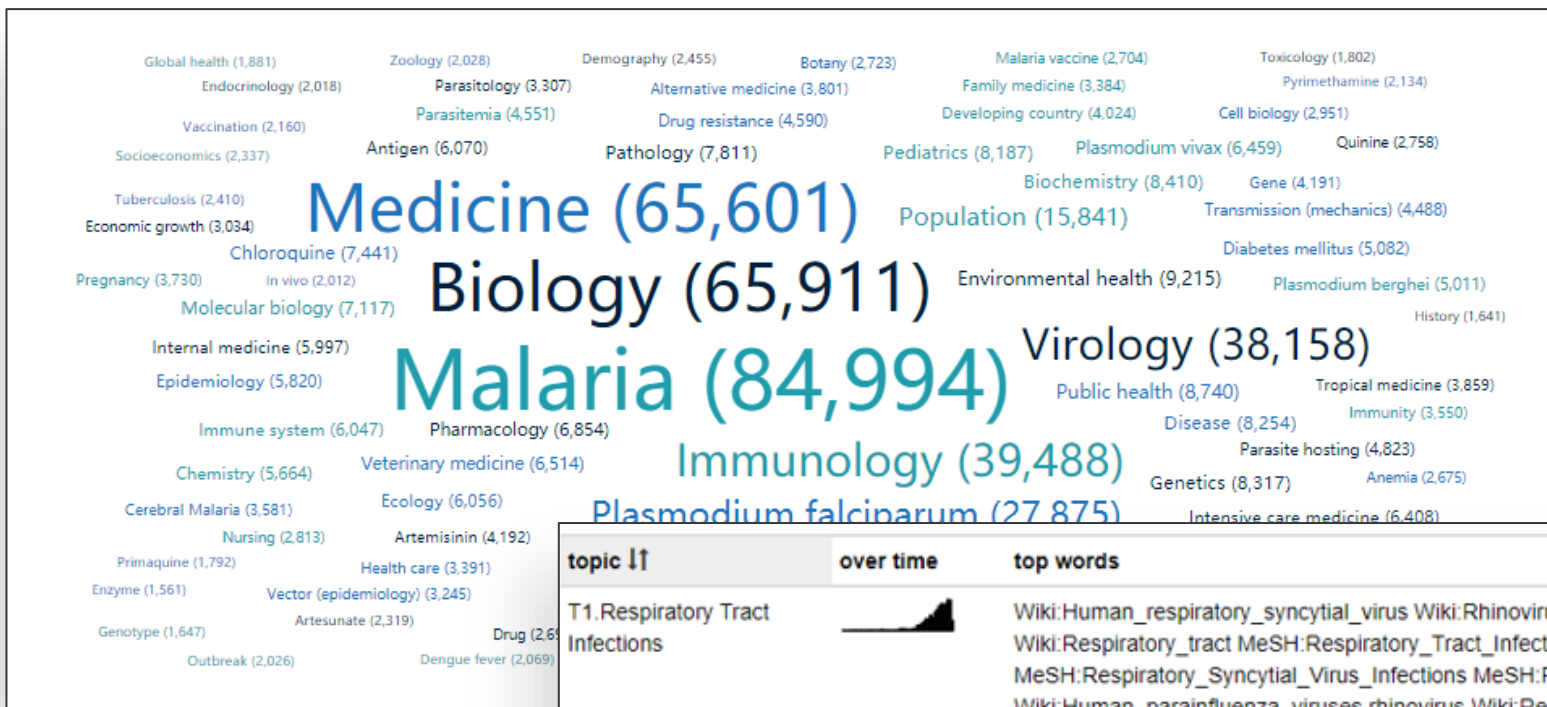
1. Vaccines - 0.95
2. Coronavirus disease 2019- 0.95
3. Child – 0.90
4. Clinical Trials, Phase I - 0.85
5. Clinical Trials, Phase II - 0.85
6. Clinical Trials, Phase III - 0.85
7. Immunogenicity, Vaccine – 0.80
8. Safety - 0.80
9. **Efficacy - 0.80**
10. Placebos - 0.75

Milestone Work Identification

Milestone papers relates to LiFePO4



Research Analytics



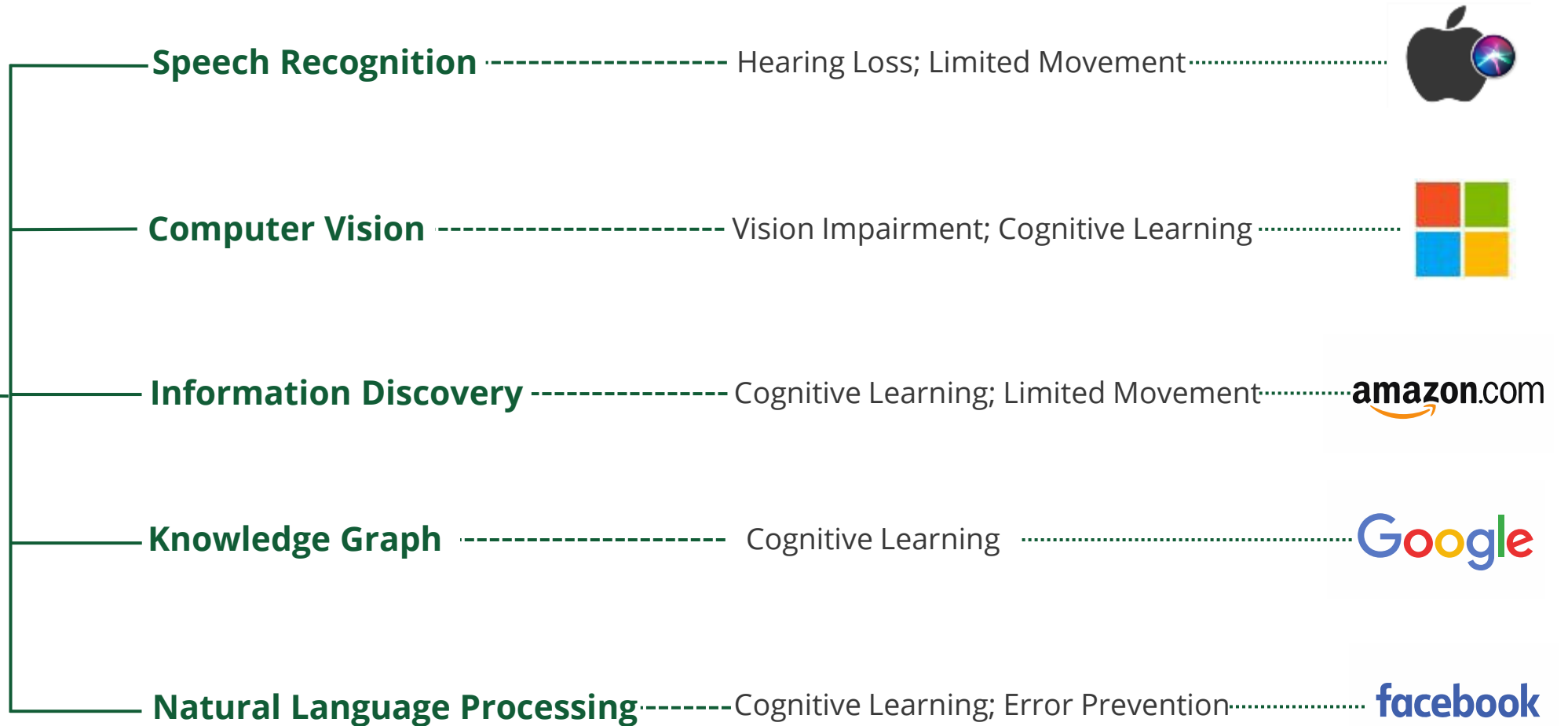
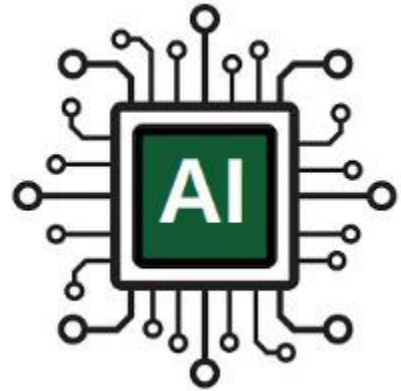
topic ↑↓	over time	top words	proportion of corpus
T1. Respiratory Tract Infections		Wiki:Human_respiratory_syncytial_virus Wiki:Rhinovirus MeSH:Child MeSH:Infant Wiki:Respiratory_tract MeSH:Respiratory_Tract_Infections children rsv MeSH:Preschool MeSH:Respiratory_Syncytial_Virus_Infections MeSH:Rhinovirus syncytial Wiki:Human_parainfluenza_viruses rhinovirus Wiki:Respiratory_tract_infection	3.7%
T10. Pulmonary Diseases & Hepatitis (mixed)		Wiki:Hepatitis_C_virus hcv Wiki:Hepatitis_B_virus Wiki:Chronic_obstructive_pulmonary_disease MeSH:Hepacivirus asthma MeSH:Asthma MeSH:Hepatitis_C hbv hepatitis exacerbations MeSH:Hepatitis_B_virus MeSH:Pulmonary_Disease MeSH:Chronic_Obstructive copd	1.0%
T11. Public health surveillance		surveillance MeSH:Global_Health Wiki:Disease_surveillance global MeSH:History international MeSH:Communicable_Diseases Wiki:Public_health_surveillance MeSH:International_Cooperation Wiki:Global_health countries MeSH:Emerging MeSH:Population_Surveillance Wiki:International MeSH:World_Health_Organization	3.8%
T12. Murine hepatitis virus		MeSH:Murine_hepatitis_virus Wiki:Hungaroton mhv mouse mice Wiki:Mouse_hepatitis_virus Wiki:Knockout_mouse murine Wiki:Murinae Wiki:Laboratory_mouse MeSH:Hepatitis MeSH:Inbred_C57BL hepatitis Wiki:Bundesautobahn_59 Wiki:Watch	1.6%
T13. Epidemiological studies & data analysis		Wiki:Confidence_interval Wiki:Odds_ratio ci MeSH:Child MeSH:Infant MeSH:Aged MeSH:Preschool MeSH:Case-Control_Studies MeSH:Risk_Factors incidence MeSH:Adult MeSH:Adolescent Wiki:Logistic_regression MeSH:Middle_Aged regression	2.4%

WILEY



Content Accessibility

Different Branches of AI



Alt Text for Image Example

Alt text (alternative text) describes the appearance or function of an image on a page for screen reader & SEO.

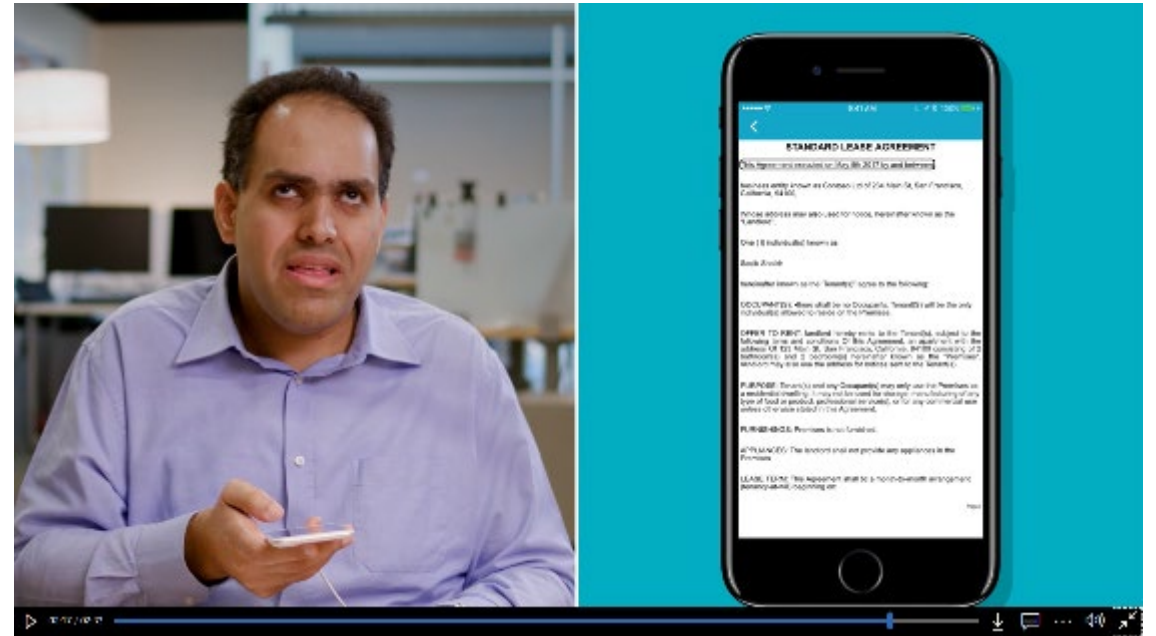


``

Content to Audio Example

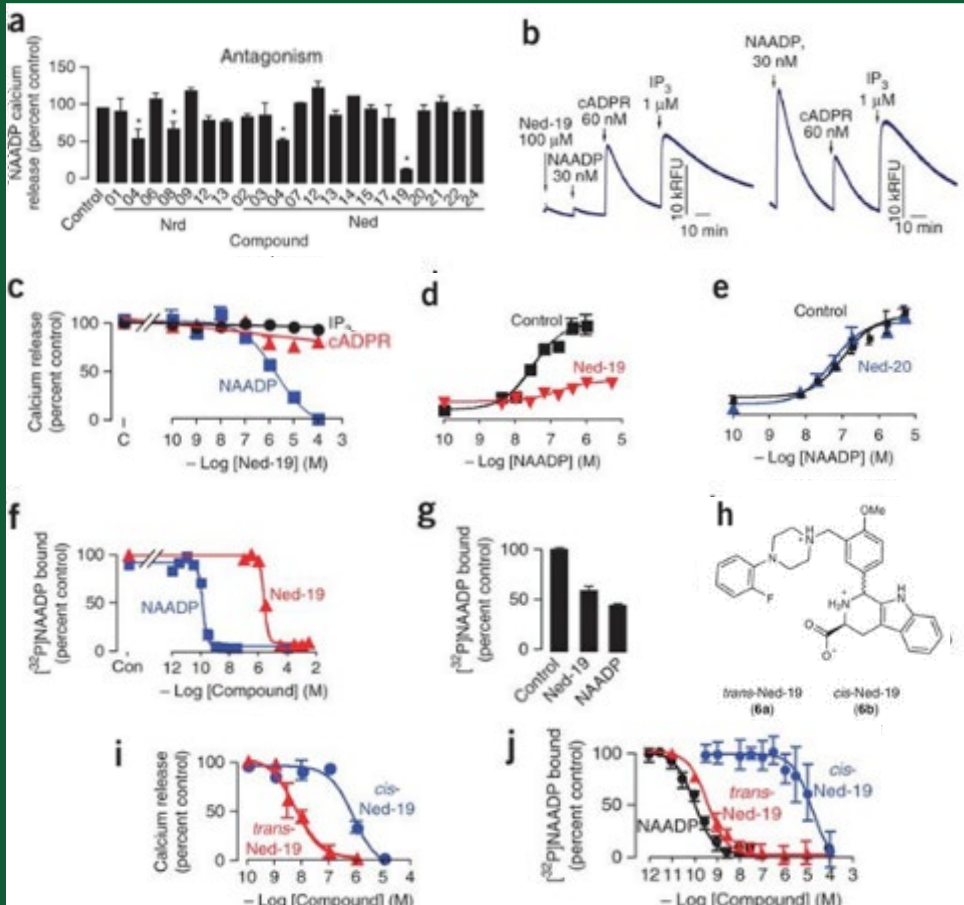
Microsoft, Google and Samsung all provides free apps that assists people with impaired vision by narrating surroundings.

- Microsoft Seeing-AI has already assisted people with > 20 million tasks
- Available in 6 languages
- <https://www.microsoft.com/en-us/ai/seeing-ai>



Compound Figure Segmentation Example

Antagonism



cADPR

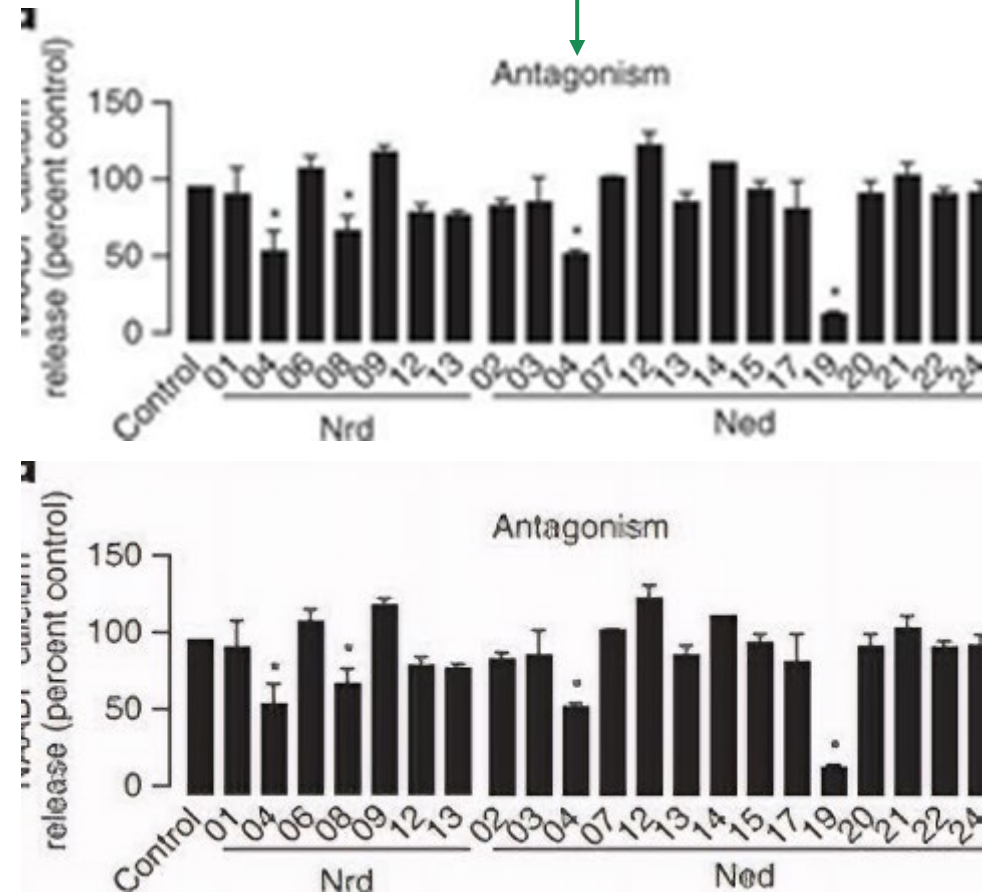
- a. NAADP compound
- b. NAADP IP₃
- c. Calcium release
- d. NAADP Log Ned-19
- e. NAADP Log Ned-20
- f. NAADP bound
- g. Control Ned-19
- h. cis-Ned-19
- i. Log compound M
- j. Percent control bound

Image Quality Enhancement Examples

AI can **automatically increase contrast** in images.



AI can **automatically enhance image resolution** and quality to improve readability.



Speech-to-Text Example

AI can listen to a video or sound file and then transcribe the spoken words into text.



The series that we're publishing this week on stillverse is the perfect example of great science, which can be implemented today and tomorrow to save the lives of literally millions of children around the world. My name is Richard Horton and I'm editor of the Lancet. There's been such an amazing success over the last decade in reducing maternal newborn and child health. But the last great Taboo in Women's and Children's Health is the epidemic of still. Us 2.6 million stillbirths around the world every year and we just don't talk about it. We want the Lancet to publish the very best science by the very best scientists in the world, but we don't want that science to sit on the shelf or the library. We want that science to lead to change social change action people believe wrongly that there's nothing still Bursar inevitable and that's just not the case ten countries in the world have over two-thirds of stillbirths. So if we could Target the political leaders the heads of state of those 10 countries and get those heads of state to make stillbirth our priority for their Nations. We could have a massive impact. We have the evidence right now to be able to end this Bases full fantastic. Advocacy advocacy by parents groups nongovernmental organizations Civil Society to be able to take the message to their political leaders. So we want to create a climate and environment

Accurate: ~90% transcription accuracy

Fast: Processes 30 seconds of audio in **12 seconds**

Audio Noise Reduction Example

AI can help people listen by enhancing speech of specific people while suppressing all other sounds.





Future Thoughts

- ✓ Authorship & Ownership Change of Content in Web 3.0
- ✓ Embracing AI Collaboration
- ✓ Defining the right problem first
- ✓ Conversational information discovery
- ✓ Personal Research Assistants
- ✓ Upholding Research Integrity
- ✓ AI Governance
- ✓ **Don't fully rely on AI**

Approachable Digital Transformation for Organizations

- ✓ **Business:** Use data to create clear strategies and monitor ROI
- ✓ **Product:** Make user-centric products and accelerate the time to value
- ✓ **People:** Find people with the right expertise and increase colleague understanding
- ✓ **Process:** Ensure agile, cross-functional collaboration and innovation
- ✓ **Data:** Central data station including data management & operations to break data silos
- ✓ **Technology:** You only need a working solution, not necessarily the most advanced one.

WILEY



Questions?

Thank you!

[Hong's Posts in the Scholarly Kitchen](#)

[The Survey need your input](#)