

Integrating Large Language Models into Research Paper Classification

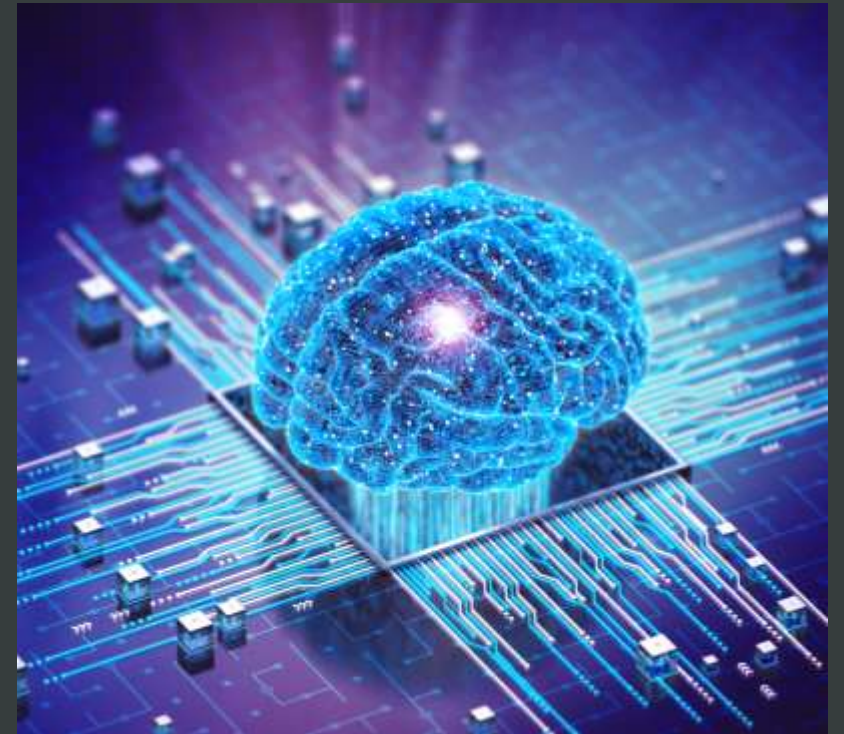
KUNAL AHLUWALIA, SENIOR PRODUCT MANAGER
GEORGIA HELLARD TIMM, PRODUCT MANAGER
ELSEVIER



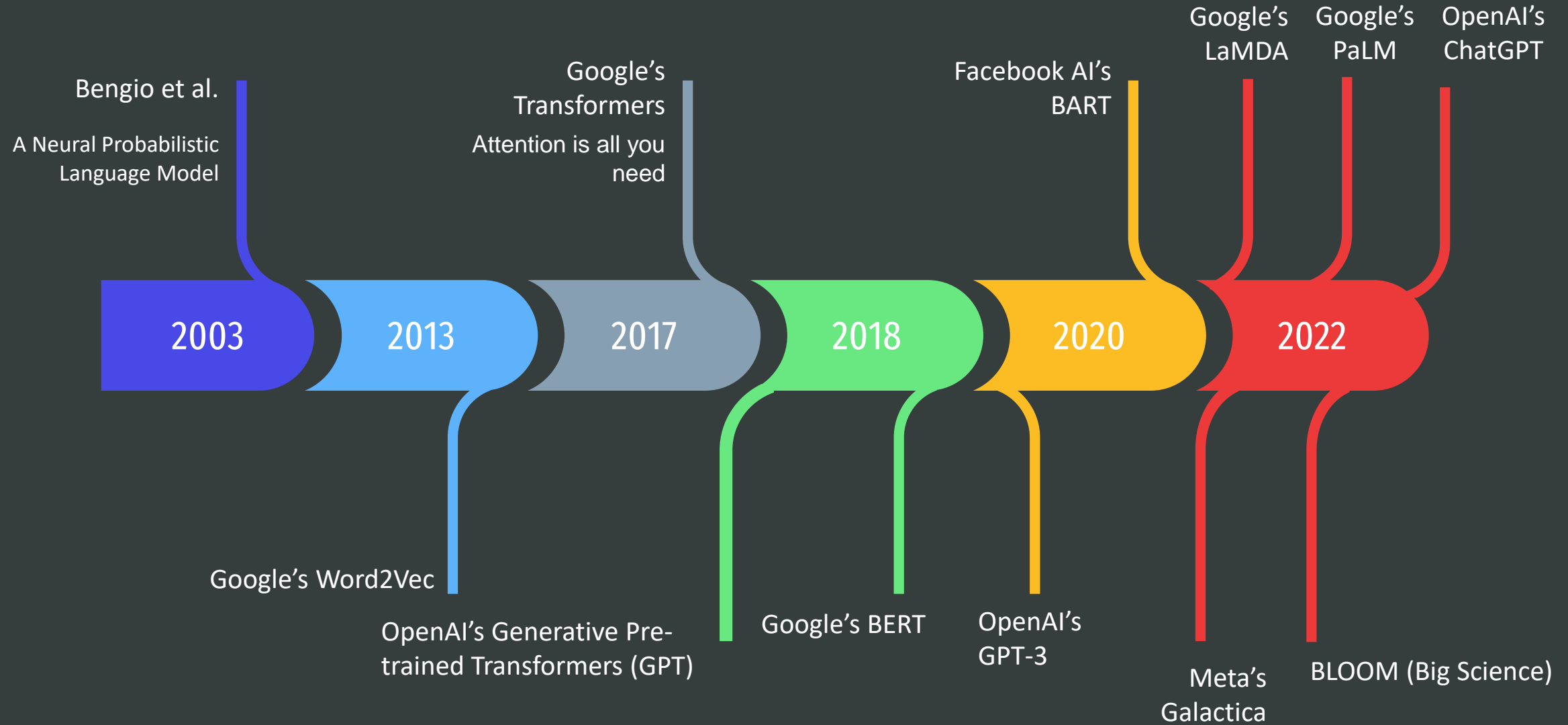
SSRN

Key Capabilities of Large Language Models

- Improve search (Conversational Search / Semantic Search / Question-Answering)
- Traditional text mining exercises; text classification, text clustering, sentiment analysis, language quality assessment, information extraction, named entity recognition and linking
- Assisting code development
- Chatbot / Tutor
 - To partly automate human interactions
 - To facilitate conversational capabilities in a virtual environment
- Summarization of content
 - Producing summaries of content spread across different sources in a more automated way than has been done today
 - To summarize content that is today not being summarized.



A Brief History of (Large) Language Models



Potential sources of problems in the application of LLMs

The top reasons why LLMs can be failing:

- **Embedded Bias:** bias introduced in the training data or in the process of training the model
- **Catastrophic Forgetting:** the tendency of an AI system to entirely forget information it previously knew after learning new information (especially in neural networks, most architectures have terrible memory)
- **Explainability:** the majority of attribution methods are unstable and lack explainability on their own accord
- **Quantifying Uncertainty:** for most applications, the time consuming process of validating and obtaining a confidence level for a decision is not applicable (e.g., self-driving cars)
- **Common Sense and Reasoning:** Axioms and reasoning algorithms are not embedded in the vast majority of AI systems
- **Hallucination:** with higher temperatures in LLMs, they can become really creative, and produce a lot of missinformation, e.g., fabricate references

5 Steps to integrate GenAI in your product

(1) Define the problem

- Low profile, high impact – Not part of your core customer facing offering
- Areas for efficiency

(2) Evaluate your data

- Having the right data is key
- Do you have it? How will you acquire it?

(3) Build vs Buy

- Do you have the skills? How much will it cost?
- Consider data privacy

(4) Proof of Concept

- Ensure subject matter experts are involved
- Set clear acceptance criteria

(5) Scale to production (and monitor)

- Keep an eye on spiralling costs
- Watch out for model drift

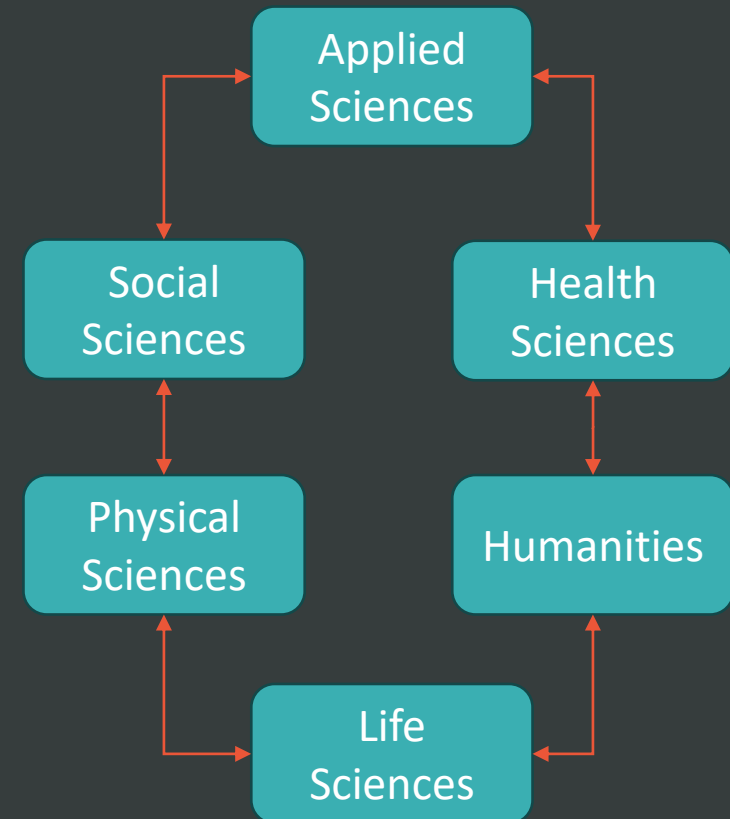
Scopus AI – Change the way you view knowledge

The screenshot displays the Scopus AI interface. At the top, there's a search bar and navigation links for Documents, Authors, Researcher Discovery, Affiliations, and Scopus AI Alpha. Below the search bar, a text box contains the query "Influence of seismology on civil engineering designs". The main content area shows an AI-generated overview of the topic. On the left, there's a text summary with a "Show all references" link and a "Rate this summary" button. Below the summary are three example questions generated by AI. On the right, a mind map diagram shows "Seismology" at the center, branching into "Earthquake monitoring", "Geotechnical Engineering", "Earthquakes", "Structural analysis", "Civil Engineering", and "Seismic Design". Further sub-branches include "Damage Data", "Impact", "Research", "Sustainability", and "Innovation".

- Combines generative AI with Scopus trusted content and data to help researchers get deeper insights faster, enable collaboration and support even greater impact across the research ecosystem.
- Key features:
 - Summarized views based on Scopus abstracts
 - Easy navigation to "Go Deeper Links" for extended exploration
 - Natural language queries
 - A soon-to-be-added graphical representation, offering new perspectives of interconnected research themes

SSRN

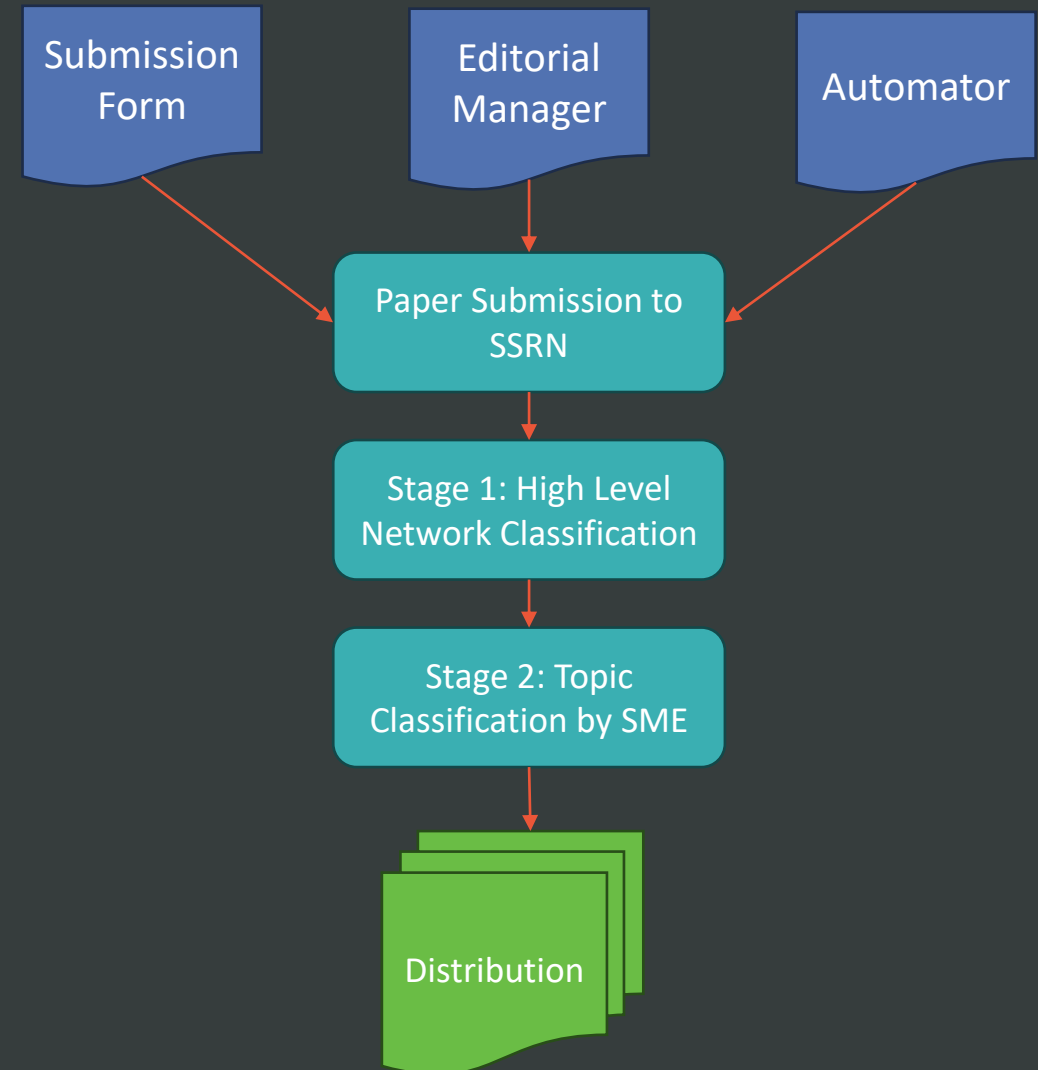
- Preprint repository acquired by Elsevier in 2016:
 - Founded in 1994
 - 2.5 million monthly unique visitors
 - 1.3M papers & 1.4M authors
 - 250M total paper downloads
- Free to upload & download but we make money by:
 - Research Paper Series
 - Custom Data Feeds
 - Subscriptions
 - Sponsorships
 - Jobs & Announcements



SSRN Paper Classification

SSRN Paper Classification

- Accounting Research Network (41,612 Papers)
- Anthropology & Archaeology Research Network (106,511 Papers)
- Cognitive Science Network (27,532 Papers)
- Corporate Governance Network (46,985 Papers)
- Criminal Justice Research Network (17,522 Papers)
- Decision Science Research Network (19,878 Papers)
- Economics Research Network (670,415 Papers)
 - Economics Departments Research Papers (11,184 Papers)
 - Economics Research Centers Papers (95,543 Papers)
 - ERN Conferences & Meetings (1,495 Papers)
 - ERN First Look (480 Papers)
 - ERN Government & Public Agency Research Paper Series (46,441 Papers)
 - ERN Partners in Publishing Journals (6,730 Papers)
 - ERN Professional, Practitioner, & Policy Paper Series (2,012 Papers)
 - ERN Public Policy Centers Research Papers (8,387 Papers)
 - ERN Subject Matter eJournals (529,347 Papers)
 - AEA: Ten Years & Beyond: Economists Answer NSF's Call for Long-Term Research Agendas eJournal (17 Papers)
 - Agricultural & Natural Resource Economics eJournal (20,105 Papers)
 - Behavioral & Experimental Economics eJournal (18,321 Papers)
 - Development Economics eJournals (41,579 Papers)
 - Econometrics eJournals (130,219 Papers)
 - Econometrics: Applied Econometric Modeling eJournals (81,982 Papers)
 - Econometrics: Applied Econometrics & Modeling eJournal, Archives of Vols. 1-4, 2008-11 (3,128 Papers)
 - Econometrics: Computer Programs & Software eJournal (620 Papers)
 - Econometrics: Data Collection & Data Estimation Methodology eJournal (2,673 Papers)
 - Econometrics: Econometric & Statistical Methods - General eJournal (11,910 Papers)
 - Econometrics: Econometric & Statistical Methods - Special Topics eJournal (8,297 Papers)
 - Econometrics: Econometric Model Construction, Estimation & Selection eJournal (2,611 Papers)
 - Econometrics: Mathematical Methods & Programming eJournal (6,641 Papers)
 - Econometrics: Multiple Equation Models eJournal (4,796 Papers)
 - Econometrics: Single Equation Models eJournal (3,150 Papers)
 - Economic Growth eJournal (14,812 Papers)



The Robot Librarian?



What problem are we trying to solve?



Improve
classification
speed



Reduce
classification
backlog



Improve
classification
consistency



Reduce
classification
costs

What do we know?

- We know ChatGPT is good at classification.
- We ran an initial trial with ChatGPT with good results.
- ChatGPT is flexible.
- ChatGPT would be faster, cheaper and more consistent.

Our initial testing

GE

I'm going to give you another abstract, and I want you to use the same taxonomy to classify it



Of course, please provide the new abstract for classification.



GE

Despite hundreds of guides, frameworks, and principles intended to make AI “ethical” or “responsible”, ever more powerful applications continue to be released ever more quickly. Safety and security teams are being downsized or sidelined to bring AI products to market. And a significant portion of AI developers apparently believe there is a real risk that their work poses an existential threat to humanity.



This contradiction between statements and action can be attributed to three factors that undermine the prospects for meaningful governance of AI. The first is the shift of power from public to private hands, not only in deployment of AI products but in fundamental research. The second is the wariness of most states about regulating the sector too aggressively, for fear that it might drive innovation elsewhere. The third is the dysfunction of global processes to manage collective action problems, epitomized by the climate crisis and now frustrating efforts to govern a technology that does not respect borders. The tragedy of AI governance is that those with the greatest leverage to regulate AI have the least interest in doing so, while those with the greatest interest have the least leverage.

Resolving these challenges either requires rethinking the incentive structures — or waiting for a crisis that brings the need for regulation and coordination into sharper focus.



Based on the provided abstract, the following classification would be appropriate:



- ✓ LSN: Technology & Innovation - Law & Policy (Topic)
- ✓ LSN: Ethics, Social and Political Philosophy (Topic)
- ✓ LSN: Global Governance (Topic)
- ✓ LSN: International Law: Public Law - Public International Law (Topic)
- ✗ LSN: Political Economy - Policy Studies (Topic)

The abstract primarily focuses on the governance challenges posed by the rapid advancement of AI technology, the shifting power dynamics between public and private entities, the lack of aggressive regulation by states, and the global challenges in managing collective action problems. It also touches on the incentives and the potential for a crisis to drive regulation and coordination. Therefore, it fits within the realms of technology and innovation law and policy, ethics, social and political philosophy, global governance, international law, and political economy and policy studies.

Regenerate

Our initial steps



- ChatGPT was going through the correct steps.



- ChatGPT can explain its choices.



- There were some problems.



- Implementation will be slow.



- We won't be able to implement this solution in a blanket way.

Integration with Data Science



- We were able to partner with the Data Science team.



- We're going to be using a solution that already exists and building on top of it.



- What data do we already have?

Given an article's title and abstract, your task is to classify it into the most relevant topics. Your model should return the top 3 topics along with the confidence scores for each topic. It is crucial that the selected topics are STRICTLY from the provided taxonomy (from level 1) and do not include any topics that are not listed. Report only the topic name itself, without its parents. The confidence score represents the model's certainty about the relevance of a particular topic to the given article. Make sure to double-check the selected topics against the provided taxonomy before providing the output.

Given Article:

Title of the given article: Determination of Polyunsaturated and Omega-3 Fatty Acids in Vegetable Oils and Animal Fats by the Double-Quantum Filtered Correlation Spectroscopy (DQF-COSY)

keywords of the given article: Polyunsaturated fatty acid, omega-3 fatty acid, Edible fatty acid, Quantification, DQF-COSY spectrum

Abstract of the given article: Polyunsaturated fatty acids in vegetable oils and animal fats were one of the most important nutrient substances in the daily intake of humans and animals. These kinds of compounds were responsible for the source of energy, the structural unit of the cell membrane and important physical function. The conventional gas chromatography method inevitably used corrosive and carcinogenic chemicals, and spent long sample preparation time and detection time. To provide a determination method, the molar concentrations of omega-3 FAs and PUFAs in vegetable oils and animal fats were directly investigated with the DQF-COSY spectrum. The method was validated with trueness, selectivity, precision, limit of quantification, and linearity. The method could be used to quantify 14 kinds of vegetable oils, 5 kinds animal fats and 3 kinds of blend oils. The technique was found suitable for the quality control, the market supervision and the nutritional evaluation of FAs.

Taxonomy of Topics:

Applied Sciences

Health Sciences

Humanities

Life Sciences

Physical Sciences

Social Sciences

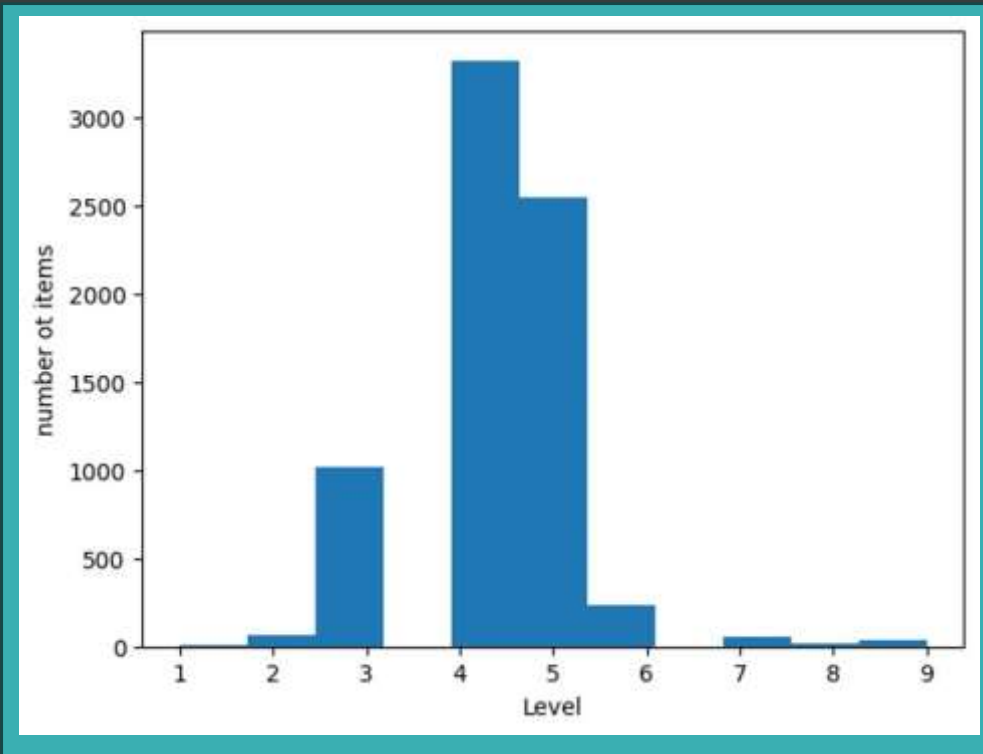
Expected Output:

From the Taxonomy of Topics, select top 3 relevant topics for the given article.

Format the output as a json contains one key, Selected_Topics, which is a list of tuples. each tuple contains a topic and the confidence score for that topic. please report the full name of the topic, including "topic" if it is written there. For tuples, instead of '(' and ')', please use '[' and ']'.

Some data

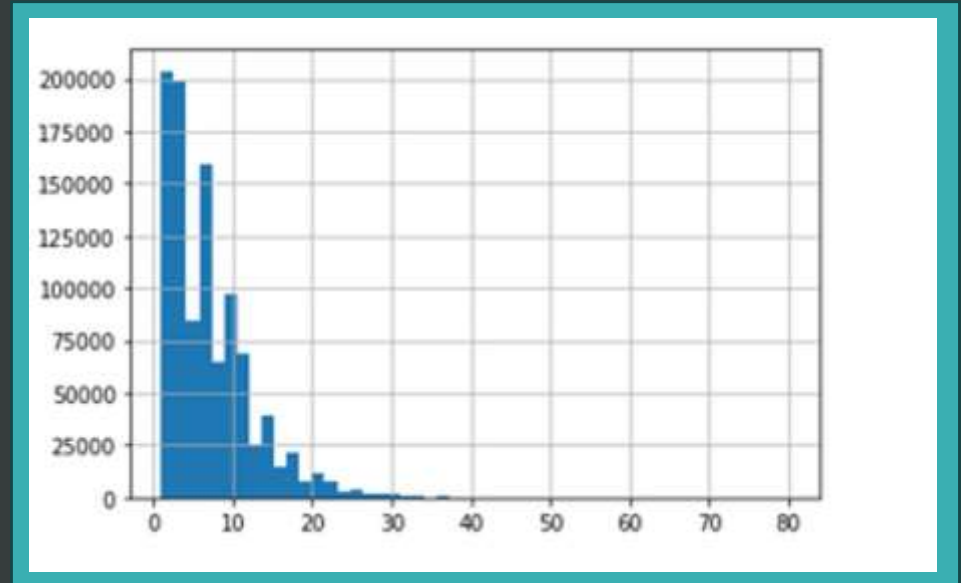
SSRN Taxonomy



7827 Items on different levels.

Example:
Social Sciences
Legal Scholarship Network
Consumer Law eJournal
LSN: Consumer Credit & Payment Issues (Topic)
LSN: Consumer Credit Issues (Sub-Topic)
LSN: Consumer Disclosure (Sub-Topic)
LSN: Consumer Leasing Act (Sub-Topic)

Classifications per Article



What is our experiment?

- We're giving ChatGPT information about our taxonomy.
- We give it papers to classify.
- We're comparing the labels.
- We're giving both sets of classifications scores out of 5.
- We're integrating feedback from the evaluations.



We've learnt that our taxonomy is too large for ChatGPT to ingest at once.

An overview of our method

We need to find a way for ChatGPT to search through our large taxonomy.

1

Start from the top down.

2

Try and eliminate unnecessary areas of our taxonomy.

How does ChatGPT think it's performing?



Average scores for classifications:
GPT4 classifications: 85%
Human classifications: 80%

How does SSRN think GPT is performing?



- We're encouraged by the results that we've seen so far.
- We're finessing the prompt engineering.
 - We think that ChatGPT will be able to at least match the classification quality of the average human classification.
- We're seeing rapid progress:
 - We're able move quickly through iterations of small data sets, adjusting our prompt engineering as we get the results from each round.

How does SSRN think ChatGPT is performing?

<u>Challenges</u>	<u>Proposed Solution</u>
ChatGPT was breaking our classification rules: it was selecting too many topics from the same area.	We've had to reengineer our prompts so that it accepts our classification rules.

How can we begin to implement this?

We're still experimenting with our method - this the Ideation phase.

1. We've set our acceptance criteria
2. We're adding to the prompts
3. We're using a comparison tool



How can we use GPT in our classification process?

Once we ChatGPT is functioning to a high enough standard, we need to integrate it - this will be the Implementation phase.

1. We'll start with the lower-risk networks.
2. We'll then explore whether it's possible to use ChatGPT in higher-risk networks.



What does this mean for SSRN?

- How will classification look going forward?
- What will classifiers be doing?



How are we integrating GenAI into SSRN?

(1) Define the problem

- Reducing the need for manual classification by SMEs.
- Speed up classification and lower classification costs.

So, what have we learnt?

- Prompt engineering is really tricky, and we spend a lot of time thinking about how we can simplify what we're asking AI to do.



A central collage of numerous colorful sticky notes in shades of pink, blue, yellow, and green. Each sticky note features a large, black question mark. The notes are scattered and overlapping, creating a dense, textured appearance. Two teal banners are overlaid on the collage: one at the top right and one at the bottom left.

Ask Away!

Any Questions?