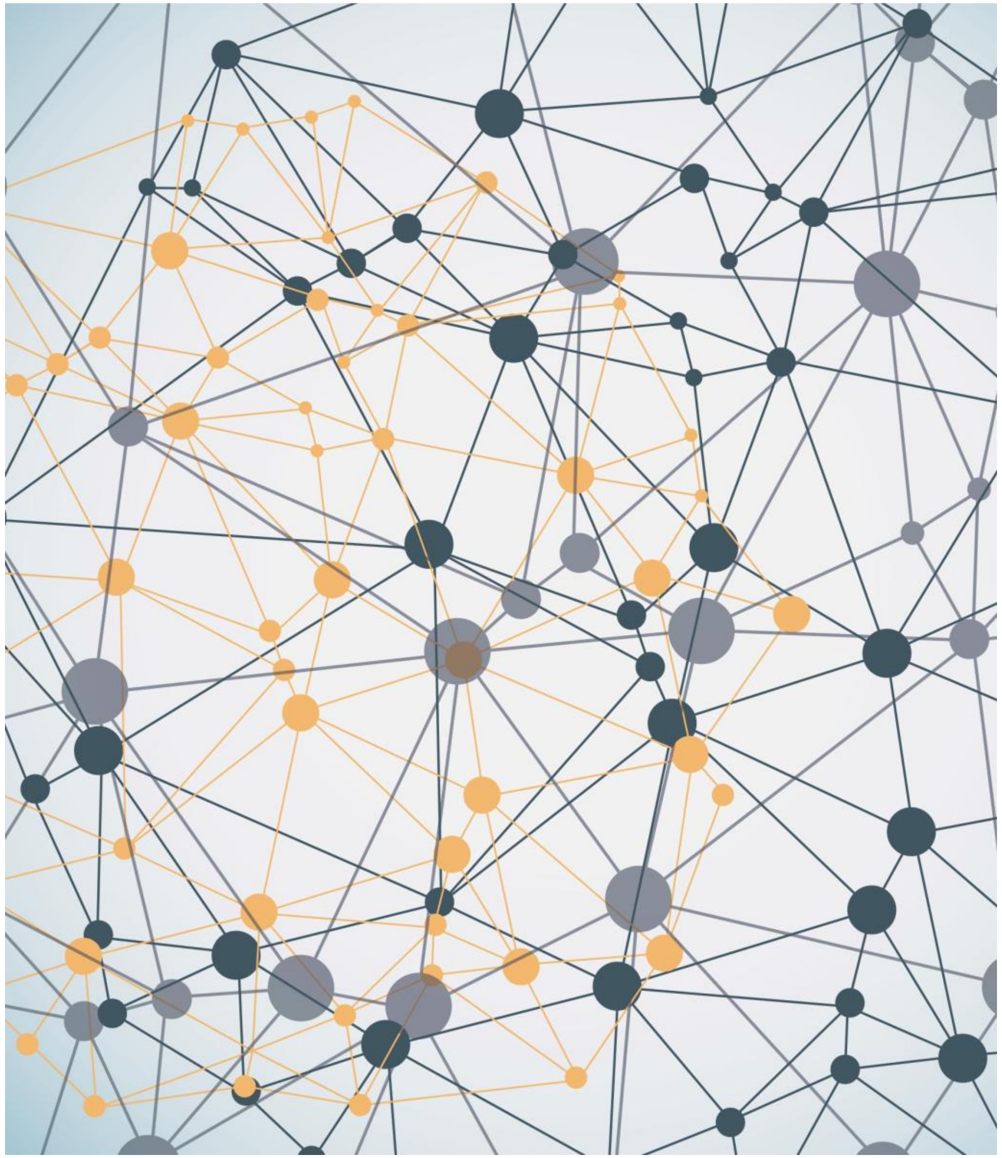
AI, NETWORKS & JOURNALISM



REPOSITORY

http://bit.ly/workshop-ai-journalist

[Berry Sanders](): Teacher Fontys, ICT (Media Design), Researcher and Project Lead

Coen Crombach: Teacher Fontys, ICT (Software), Researcher

Iman Mossavat : Teacher Fontys, ICT (Software), Researcher

WHICH IMAGE IS AI GENERATED?

A

B

CONTEXT

Large Language Models (e.g. ChatGPT)

Unprecedented abilities

Unprecedented issues

Large Interconnected Unstructured Data

Poor overview

Needle in the haystack effect

Image reference

CAN CHATBOTS UNDERSTAND?

Yes, they can understand

SKILL-MIX: A FLEXIBLE AND EXPANDABLE FAMILY OF EVALUATIONS FOR AI MODELS

Dingli Yu[1] Simran Kaur[1] Arushi Gupta[1]
Jonah Brown-Cohen[2] Anirudh Goyal[2] Sanjeev Arora[1]
[1]Princeton Language and Intelligence (PLI), Princeton University
[2]Google DeepMind

ABSTRACT

With LLMs shifting their role from statistical modeling of language to serving as general-purpose AI agents, how should LLM evaluations change? Arguably, a key ability of an AI agent is to flexibly combine, as needed, the basic skills it has learned. The capability to combine skills plays an important role in (human) pedagogy and also in a paper on emergence phenomena (Arora & Goyal, 2023).

This work introduces SKILL-MIX, a new evaluation to measure ability to combine skills. Using a list of N skills the evaluator repeatedly picks random subsets of k skills and asks the LLM to produce text combining that subset of skills. Since the number of subsets grows like N^k, for even modest k this evaluation will, with high probability, require the LLM to produce text significantly different from any text in the training set. The paper develops a methodology for (a) designing and administering such an evaluation, and (b) automatic grading (plus spot-checking by humans) of the results using GPT-4 as well as the open LLaMA-2 70B model.

Administering a version of SKILL-MIX to popular chatbots gave results that, while generally in line with prior expectations, contained surprises. Sizeable differences exist among model capabilities that are not captured by their ranking on popular LLM leaderboards ("cramming for the leaderboard"). Furthermore, simple probability calculations indicate that GPT-4's reasonable performance on $k = 5$ is suggestive of going beyond "stochastic parrot" behavior (Bender et al., 2021), i.e., it combines skills in ways that it had not seen during training.

We sketch how the methodology can lead to a SKILL-MIX based eco-system of open evaluations for AI capabilities of future models.

No, they cannot

On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?

Emily M. Bender*
ebender@uw.edu
University of Washington
Seattle, WA, USA

Timnit Gebru*
timnit@blackinai.org
Black in AI
Palo Alto, CA, USA

Angelina McMillan-Major
aymm@uw.edu
University of Washington
Seattle, WA, USA

Shmargaret Shmitchell
shmargaret.shmitchell@gmail.com
The Aether

ABSTRACT

The past 3 years of work in NLP have been characterized by the development and deployment of ever larger language models, especially for English. BERT, its variants, GPT-2/3, and others, most recently Switch-C, have pushed the boundaries of the possible both through architectural innovations and through sheer size. Using these pretrained models and the methodology of fine-tuning them for specific tasks, researchers have extended the state of the art on a wide array of tasks as measured by leaderboards on specific benchmarks for English. In this paper, we take a step back and ask: How big is too big? What are the possible risks associated with this technology and what paths are available for mitigating those risks? We provide recommendations including weighing the environmental and financial costs first, investing resources into curating and carefully documenting datasets rather than ingesting everything on the web, carrying out pre-development exercises evaluating how the planned approach fits into research and development goals and supports stakeholder values, and encouraging research directions beyond ever larger language models.

alone, we have seen the emergence of BERT and its variants [39, 70, 74, 113, 146], GPT-2 [106], T-NLG [112], GPT-3 [25], and most recently Switch-C [43], with institutions seemingly competing to produce ever larger LMs. While investigating properties of LMs and how they change with size holds scientific interest, and large LMs have shown improvements on various tasks (§2), we ask whether enough thought has been put into the potential risks associated with developing them and strategies to mitigate these risks.

We first consider environmental risks. Echoing a line of recent work outlining the environmental and financial costs of deep learning systems [129], we encourage the research community to prioritize these impacts. One way this can be done is by reporting costs and evaluating works based on the amount of resources they consume [57]. As we outline in §3, increasing the environmental and financial costs of these models doubly punishes marginalized communities that are least likely to benefit from the progress achieved by large LMs and most likely to be harmed by negative environmental consequences of its resource consumption. At the scale we are discussing (outlined in §2), the first consideration should be the

[On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? "1F99C (acm.org)]

Opinions vary significantly

LLM DRAWBACKS

Hallucinations: fabricate plausible-sounding but **incorrect answers**.

They quickly become outdated when trying to understand current events due to their **fixed training** data.

Pre-trained models lack access to **private** or proprietary organizational data.

HALLUCINATION

Factual Inaccuracies: The LLM produces a factually incorrect statement.

Unsupported Claims: The LLM generates a response that has **no basis in the input or context**.

Nonsensical Statements: The LLM produces a response that **doesn't make sense** or contradictory.

Improbable Scenarios: The LLM generates a response that describes an **implausible or highly unlikely event**.

TRUSTWORTHY AI

Requirements depend on the application, the list is extensive

In this talk we investigate

- Privacy
- Content adaptability
- Content reliability

Out of scope: accuracy, transparency, bias and fairness, ...

OLLAMA FOR PRIVATE LLMS

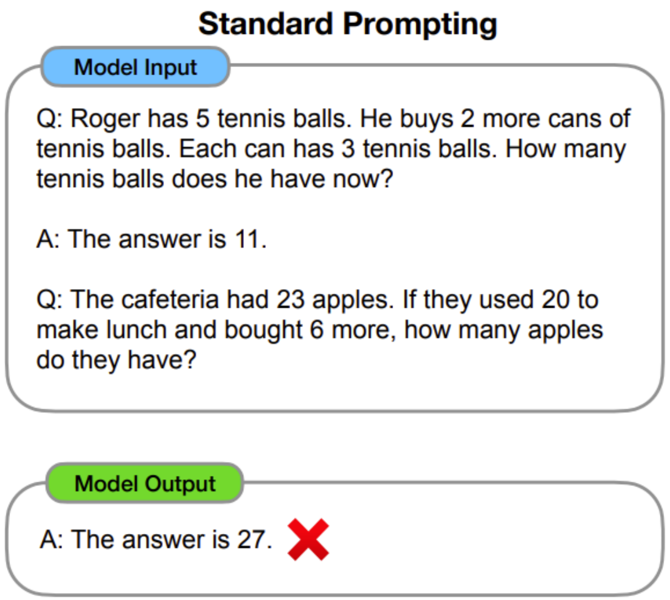
Get up and running with large language models, locally.

Run Llama 2, Code Llama, and other models.
Customize and create your own.

LLM ON A CONSUMER LAPTOP

Privacy Demo

CHAIN-OF-THOUGHT PROMPTS



2201.11903.pdf (arxiv.org)

RETRIEVAL ASSISTED GENERATION (RAG)

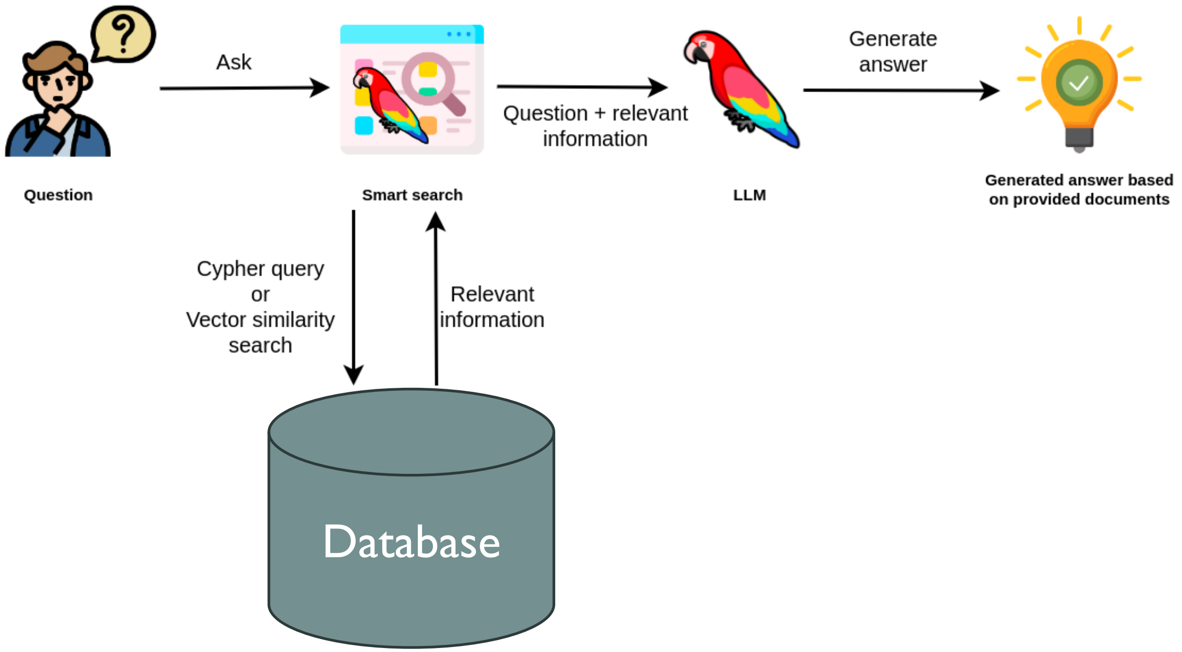


Image credit

RAG ON YOUR OBSIDIAN NOTES

Demo

main_llm_0.py | main_llm_kg.py | main_llm_kg_nebula.py

```python
    documents = ObsidianReader(obsidian_vault_path).load_data()  # Returns list of documents    index = VectorStoreIndex.from_documents(documents)

    print('indexing')

    index = VectorStoreIndex.from_documents(documents)
    # store it for later
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
(neo4j) PS C:\Users\imanm\OneDrive - Office 365 Fontys\Documenten\myCode\learn llms> & C:/ProgramData/anaconda3/envs/neo4j/python.exe "c:/Users/imanm/OneDrive - Office 365 Fontys/Documenten/myCode/learn llms/main_llm_0.py"
The full path to the PERSIST_DIR directory is: C:\Users\imanm\OneDrive - Office 365 Fontys\Documenten\myCode\learn llms\storage\testRAG
reading obsidian docs
indexing
indexing done
setting up the query engine
query response calculated
Enter your prompt (type 'exit' to quit):
```

Ln 56, Col 30 Spaces: 4 UTF-8 CRLF Python 3.10.11 ('neo4j': conda)

NETWORKS

identify influence

reveal hidden ties

map alliances and oppositions, …

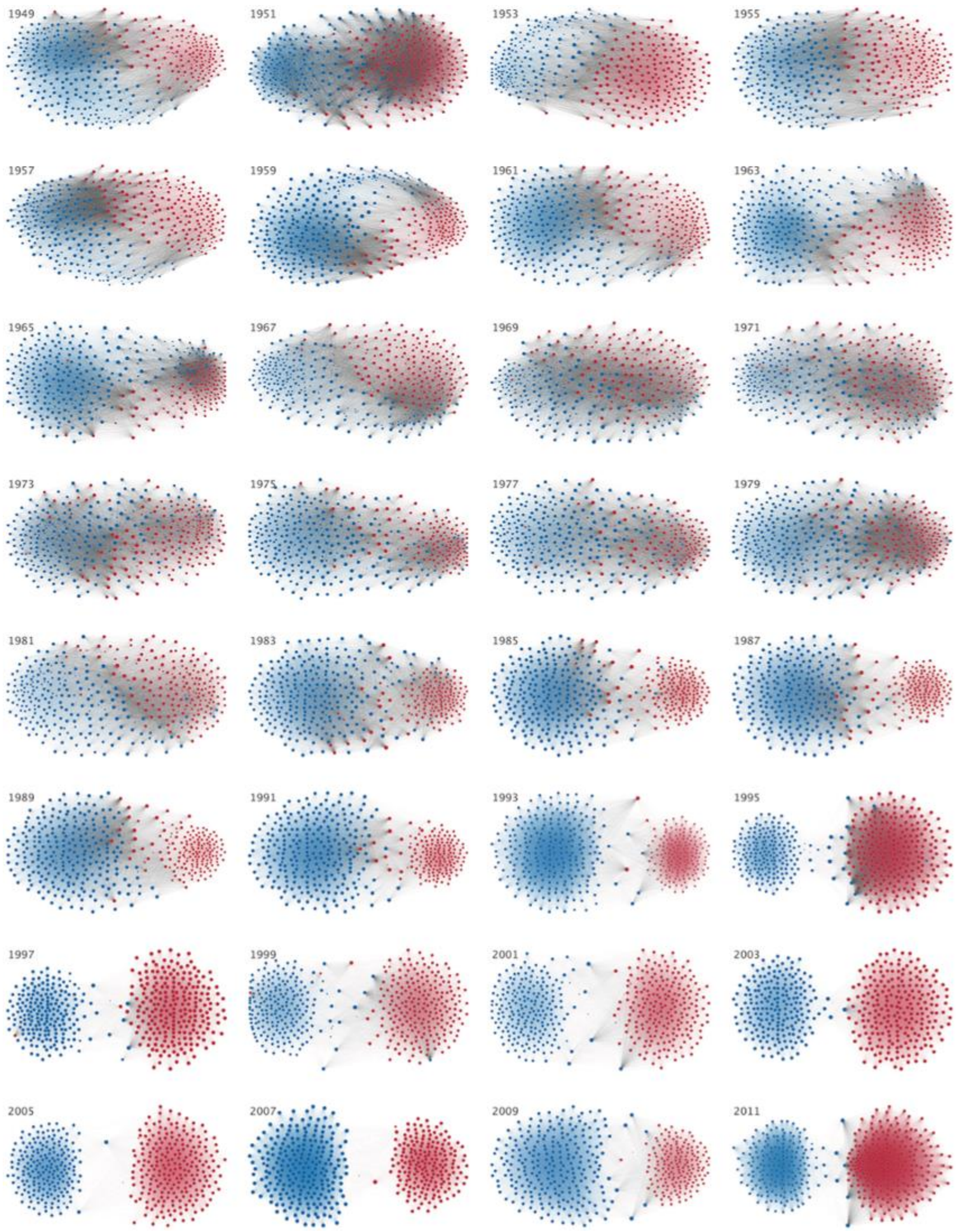
[Narrating Networks (tandfonline.com)]

EXAMPLE OF NARRATING STORIES WITH GRAPHS

Donald Trump, His Children, and 500+ Potential Conflicts of Interest - WSJ.com

IDENTIFYING INFLUENCE IN A CITATION NETWORK

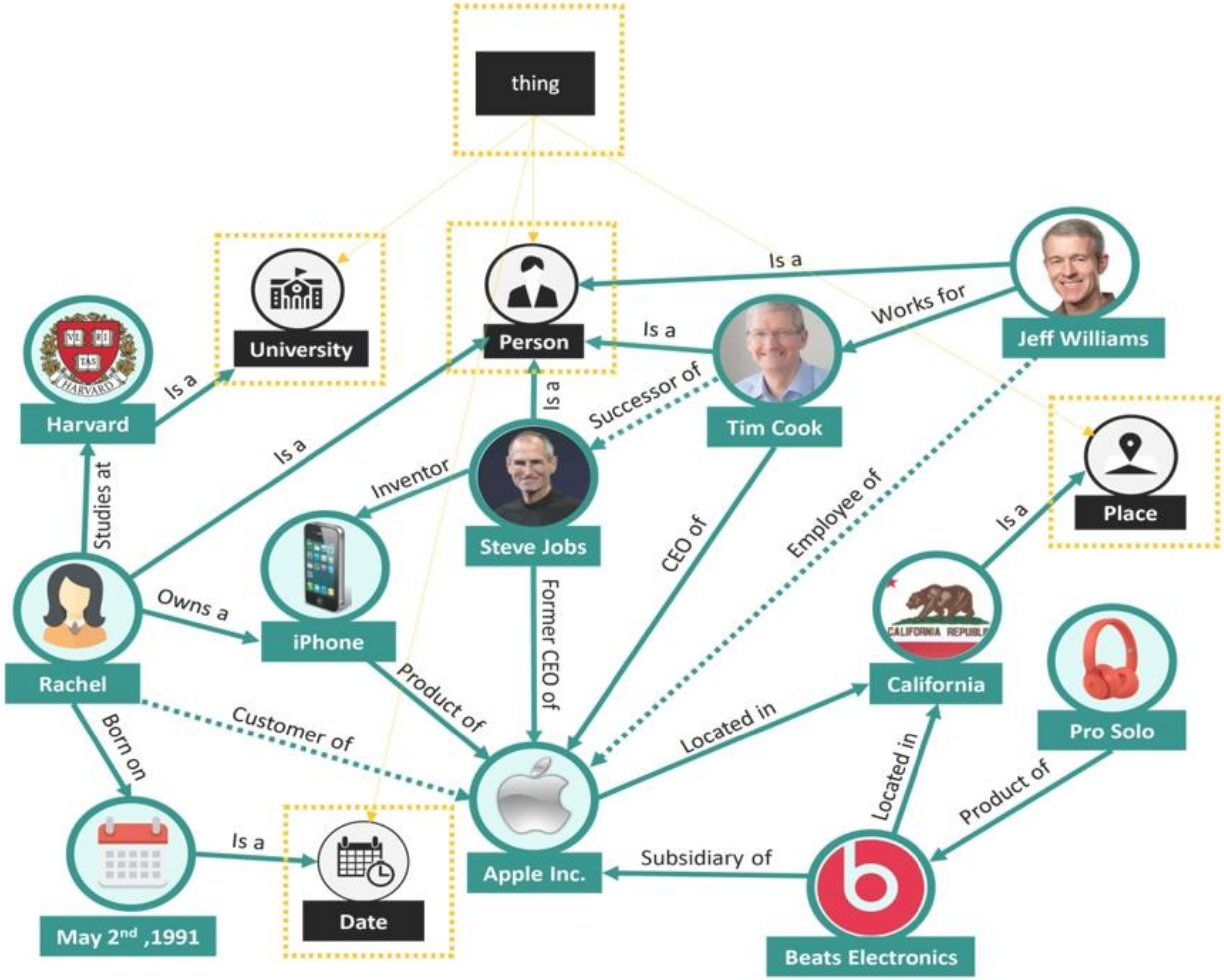
venue	year	title	centrality (in	centrality (out
and Machines	2018	AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principle	1,0689E-11	0,139721849
e	2018	The Moral Machine experiment	6,88847E-12	0,113700812
	2019	Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artifici	3,08793E-12	0,113664546
ics and Well-Being	2019	The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems	2,32783E-12	0,11366213
ial Intelligence: Foundations, Theory, and	2017	Towards a Code of Ethics for Artificial Intelligence	1,56772E-12	0,113659473
and Machines	2019	The Ethics of AI Ethics: An Evaluation of Guidelines	0,018865496	0,105415509
e Machine Intelligence	2019	The global landscape of AI ethics guidelines	0,01605566	0,105346211
:e	2018	How AI can be a force for good	2,32783E-12	0,094833462
	2019	From What to How: An Overview of AI Ethics Tools, Methods and Research to Translat	0,011409733	0,079432544
ophical Transactions of the Royal Society	2016	Faultless responsibility: on the nature and allocation of moral responsibility for distribu	1,56772E-12	0,079396093
SIGSOFT FSE	2018	Does ACM’s code of ethics change ethical decision making in software development?	3,08793E-12	0,079300572
ournal of applied psychology	2010	Bad apples, bad cases, and bad barrels: meta-analytic evidence about sources of unethi	3,08793E-12	0,07929429
	2017	Artificial Intelligence Policy: A Primer and Roadmap	8,07614E-13	0,079293633
e	2019	Don’t let industry write the rules for AI	1,56772E-12	0,079246509
Aach. Intell.	2019	Principles alone cannot guarantee ethical AI	0,015137611	0,068855827
Electronic Journal	2019	AI Ethics - Too Principled to Fail?	0,015137611	0,068777726
	2018	The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation	1,98103E-11	0,068772567
I@AAAI	2018	Linking Artificial Intelligence Principles	3,08793E-12	0,06877134
OCIETY	2017	Preparing for the future of Artificial Intelligence	1,56772E-12	0,068770637
ata Soc.	2017	Fairer machine learning in the real world: Mitigating discrimination without collecting se	2,32783E-12	0,068770587
	2016	The Social and Economic Implications of Artificial Intelligence Technologies in the Near	8,07614E-13	0,068770364
and Machines	2020	Publisher Correction to: The Ethics of AI Ethics: An Evaluation of Guidelines	1,56772E-12	0,068770131
	2016	Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Demo	8,07614E-13	0,068770075
	2017	AI Now 2017 Report	3,08793E-12	0,068769911
:e and Engineering Ethics	2016	Artificial Intelligence and the ‘Good Society’: the US, EU, and UK approach	4,60815E-12	0,06038587
and Information Technology	2009	The ethics of information transparency	2,32783E-12	0,060384697



Division of Democrat and Republican Party members over time 1949–2012. Edges are drawn between members who agree above the Congress' threshold value of votes.

(PDF) The Rise of Partisanship and Super-Cooperators in the U.S. House of Representatives (researchgate.net)

KNOWLEDGE GRAPHS



RAG WITH KNOWLEDGE GRAPHS

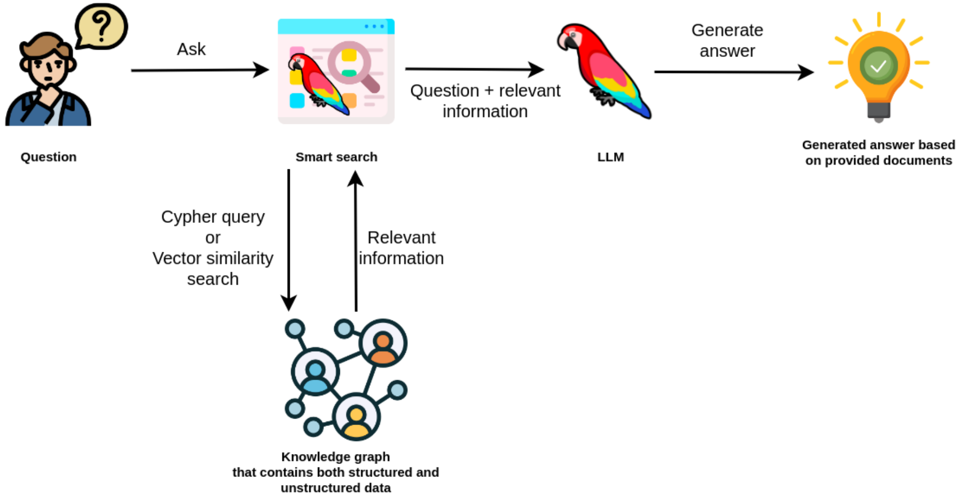
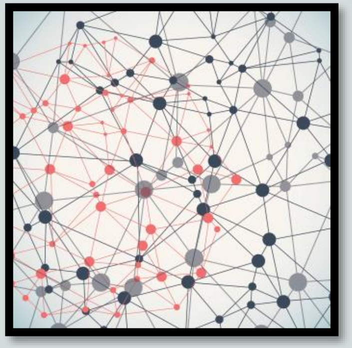


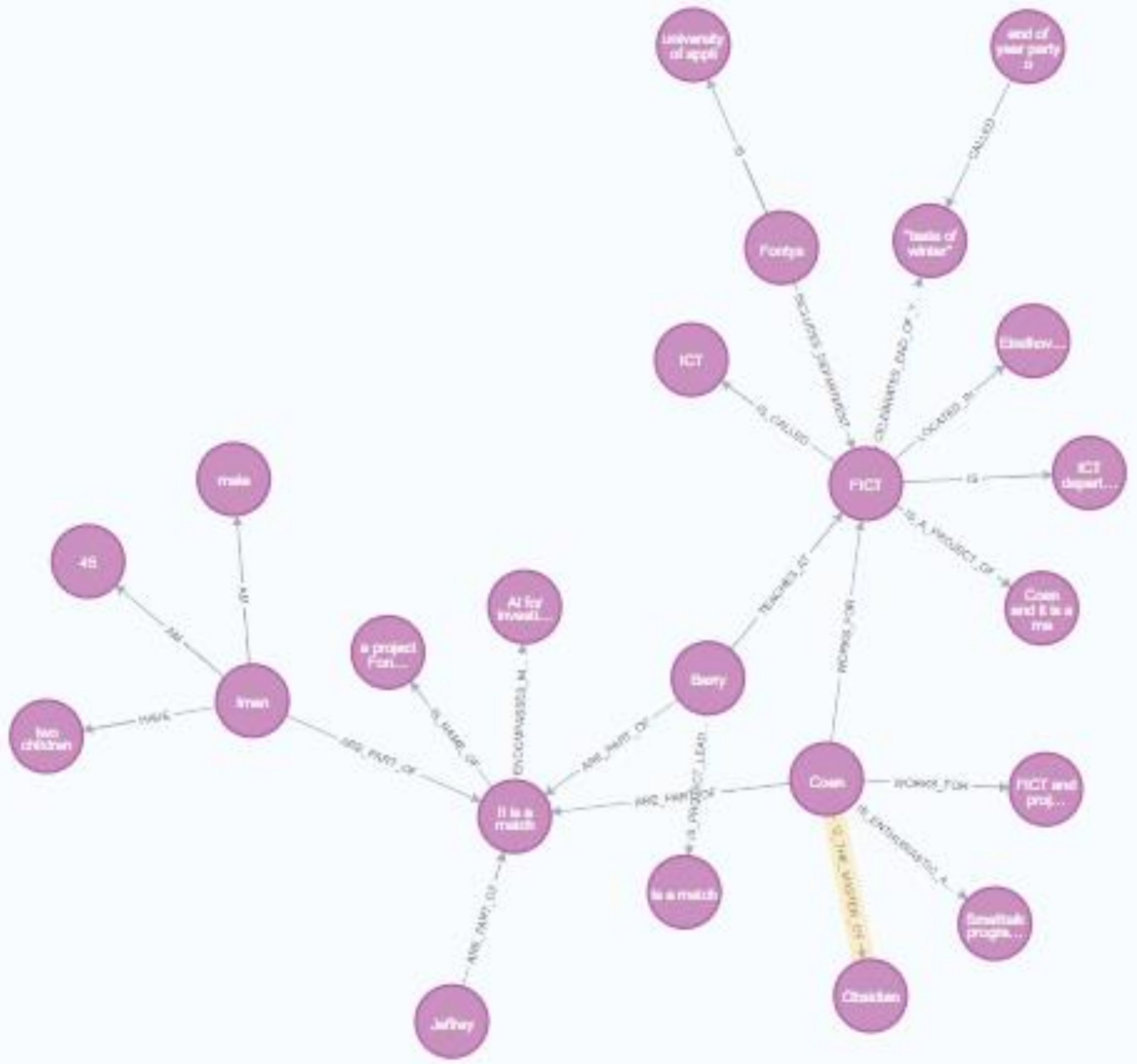
Image credit

NETWORKS AND LLMS

Networks are very relevant in the days of LLMs.

- Use LLM to build Knowledge-graphs
- Use Knowledge-Graphs for advanced RAG




university of appli
end of year party o
Fontys
"taste of winter"
ICT
Eindhov...
FICT
ICT depart...
Coen and it is a ma
male
45
Iman
two children
a project Fon...
AI for investi...
Barry
It is a match
Coen
FICT and proj...
is a match
Smalltalk progra...
Obsidian
Jeffrey
HAVE
AM
ARE_PART_OF
IS_CALLED
TEACHES_AT
WORKS_FOR
IS
IS_A_PROJECT_OF
CALLED
LOCATED_IN