Common Sense[makers]

האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

"Open Science Sensemaking: Adding Sensemaking Data to the Scientific Record for Making Sense of Science

Ronen Tamari
July 2023

Open Access science needs Open Science Sensemaking (OSSm): open infrastructure for sharing scientific sensemaking data

Ronen Tamari[1,2] and Daniel A Friedman[2,3,4]

[1]Hebrew University of Jerusalem, Israel
[2]Common SenseMakers
[3]University of California, Davis, USA
[4]Active Inference Institute

April 26, 2023

Abstract

While open access publishing effectively broadens access to scientific research products, the problem of *making sense* of the volumes of new information is becoming increasingly acute. Traditional curation methods like peer-reviewed journals and recommendation services are failing to keep pace, resulting in unprecedented information overload and knowledge fragmentation. We contend that making sense of science requires open access to diverse sources of scientific sensemaking data, and that current frictions and failures of scientific sensemaking arise from deficiencies in reckoning with these kinds of data. Sensemaking data are the digital traces of sensemaking processes,

[Submitted on 15 Jun 2023 (v1), last revised 24 Jun 2023 (this version, v2)]

Exploring the MIT Mathematics and EECS Curriculum Using Large Language Models

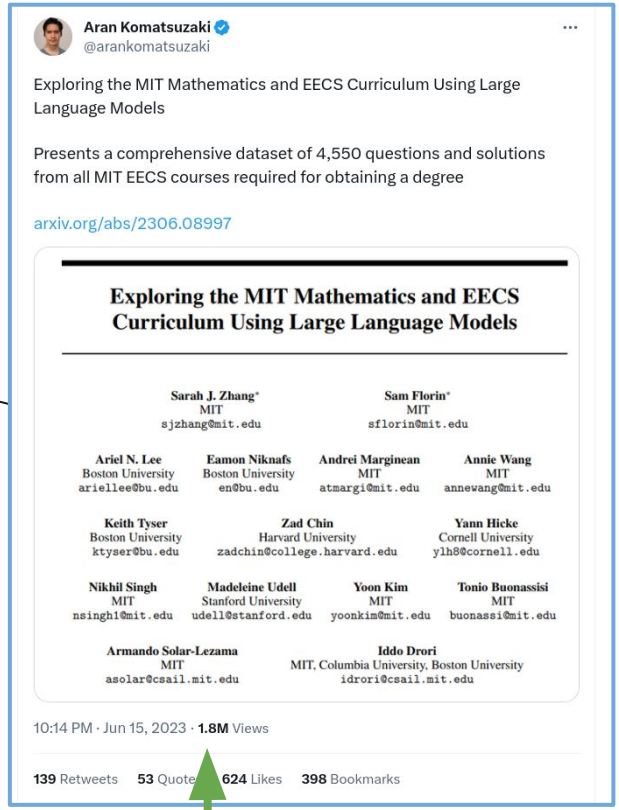
Sarah J. Zhang, Samuel Florin, Ariel N. Lee, Eamon Niknafs, Andrei Marginean, Annie Wang, Keith Tyser, Zad Chin, Yann Hicke, Nikhil Singh, Madeleine Udell, Yoon Kim, Tonio Buonassisi, Armando Solar-Lezama, Iddo Drori

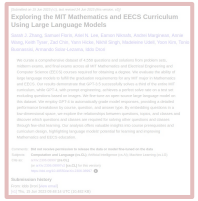
We curate a comprehensive dataset of 4,550 questions and solutions from problem sets, midterm exams, and final exams across all MIT Mathematics and Electrical Engineering and Computer Science (EECS) courses required for obtaining a degree. We evaluate the ability of large language models to fulfill the graduation requirements for any MIT major in Mathematics and EECS. Our results demonstrate that GPT-3.5 successfully solves a third of the entire MIT curriculum, while GPT-4, with prompt engineering, achieves a perfect solve rate on a test set excluding questions based on images. We fine-tune an open-source large language model on this dataset. We employ GPT-4 to automatically grade model responses, providing a detailed performance breakdown by course, question, and answer type. By embedding questions in a low-dimensional space, we explore the relationships between questions, topics, and classes and discover which questions and classes are required for solving other questions and classes through few-shot learning. Our analysis offers valuable insights into course prerequisites and curriculum design, highlighting language models' potential for learning and improving Mathematics and EECS education.

Comments: **Did not receive permission to release the data or model fine-tuned on the data**
Subjects: **Computation and Language (cs.CL)**; Artificial Intelligence (cs.AI); Machine Learning (cs.LG)
Cite as: arXiv:2306.08997 **[cs.CL]**
(or arXiv:2306.08997v2 **[cs.CL]** for this version)
https://doi.org/10.48550/arXiv.2306.08997

Submission history

From: Iddo Drori [view email]
[v1] Thu, 15 Jun 2023 09:48:14 UTC (10,482 KB)





Aran Komatsuzaki ✓
@arankomatsuzaki

Exploring the MIT Mathematics and EECS Curriculum Using Large Language Models

Presents a comprehensive dataset of 4,550 questions and solutions from all MIT EECS courses required for obtaining a degree

arxiv.org/abs/2306.08997

Exploring the MIT Mathematics and EECS Curriculum Using Large Language Models

Sarah J. Zhang* MIT sjzhang@mit.edu
Sam Florin* MIT sflorin@mit.edu
Ariel N. Lee Boston University ariellee@bu.edu
Eamon Niknafs Boston University en@bu.edu
Andrei Marginean MIT atmargi@mit.edu
Annie Wang MIT annewang@mit.edu
Keith Tyser Boston University ktyser@bu.edu
Zad Chin Harvard University zadchin@college.harvard.edu
Yann Hicke Cornell University ylh8@cornell.edu
Nikhil Singh MIT nsingh1@mit.edu
Madeleine Udell Stanford University udell@stanford.edu
Yoon Kim MIT yoonkim@mit.edu
Tonio Buonassisi MIT buonassi@mit.edu
Armando Solar-Lezama MIT asolar@csail.mit.edu
Iddo Drori MIT, Columbia University, Boston University idrori@csail.mit.edu

10:14 PM · Jun 15, 2023 · **1.8M** Views

139 Retweets **53** Quotes **624** Likes **398** Bookmarks

Raunak Chowdhuri ✓
@sauhaarda

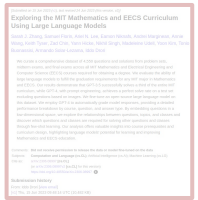
A recent work from @iddo claimed GPT4 can score 100% on MIT's EECS curriculum with the right prompting.

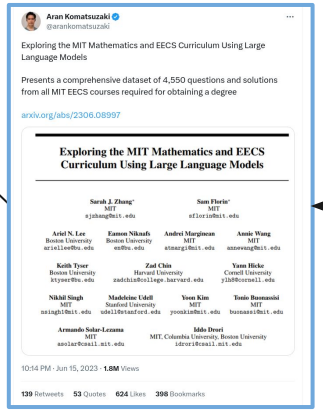
My friends and I were excited to read the analysis behind such a feat, but after digging deeper, what we found left us surprised and disappointed.

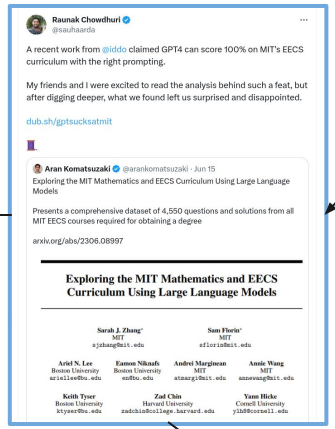
dub.sh/gptsucksatmit

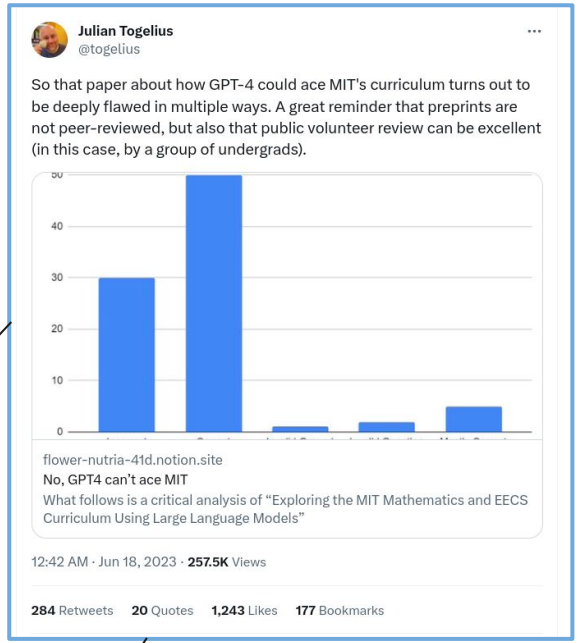
📕

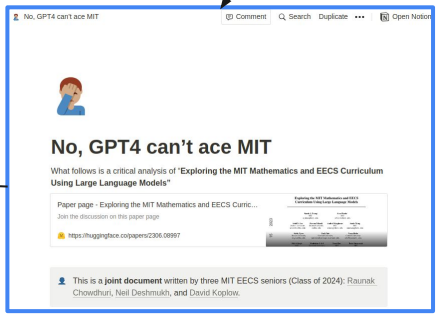
> **Aran Komatsuzaki** ✓ @arankomatsuzaki · Jun 15
> Exploring the MIT Mathematics and EECS Curriculum Using Large Language Models
>
> Presents a comprehensive dataset of 4,550 questions and solutions from all MIT EECS courses required for obtaining a degree
>
> arxiv.org/abs/2306.08997
>
> **Exploring the MIT Mathematics and EECS Curriculum Using Large Language Models**
>
>
> **Sarah J. Zhang*** MIT sjzhang@mit.edu
> **Sam Florin*** MIT sflorin@mit.edu
> **Ariel N. Lee** Boston University ariellee@bu.edu
> **Eamon Niknafs** Boston University en@bu.edu
> **Andrei Marginean** MIT atmargi@mit.edu
> **Annie Wang** MIT annewang@mit.edu
> **Keith Tyser** Boston University ktyser@bu.edu
> **Zad Chin** Harvard University zadchin@college.harvard.edu
> **Yann Hicke** Cornell University ylh8@cornell.edu
>




Aran Komatsuzaki
@arankomatsuzaki
Exploring the MIT Mathematics and EECS Curriculum Using Large Language Models
Presents a comprehensive dataset of 4,550 questions and solutions from all MIT EECS courses required for obtaining a degree
arxiv.org/abs/2306.08997
Exploring the MIT Mathematics and EECS Curriculum Using Large Language Models
Sarah J. Zhang* MIT sjzhang@mit.edu
Sam Florin* MIT sflorin@mit.edu
Ariel N. Lee Boston University arielnlee@bu.edu
Eamon Nikuafs Boston University en@bu.edu
Andrei Margineau MIT atmargi@mit.edu
Annie Wang MIT annewang@mit.edu
Keith Tyser Boston University ktyser@bu.edu
Zad Chin Harvard University zadchin@college.harvard.edu
Yann Hicke Cornell University ylh8@cornell.edu
Nikhil Singh MIT nzsingh@mit.edu
Madeleine Udell Stanford University udell@stanford.edu
Yoon Kim MIT yoonkim@mit.edu
Tonio Buonassisi MIT buonassi@mit.edu
Armando Solar-Lezama MIT asolar@csail.mit.edu
Iddo Drori MIT, Columbia University, Boston University idrori@csail.mit.edu
10:14 PM · Jun 15, 2023 · 1.8M Views
139 Retweets 53 Quotes 624 Likes 398 Bookmarks

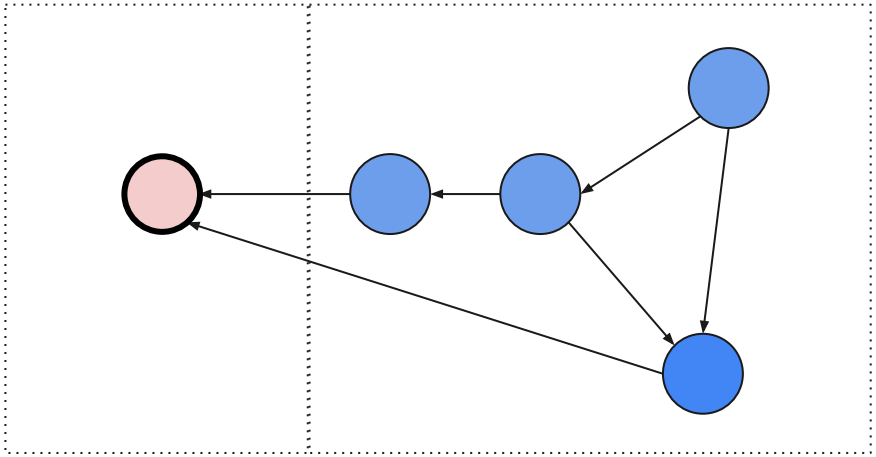

Raunak Chowdhuri
@sauhaarda
A recent work from @iddo claimed GPT4 can score 100% on MIT's EECS curriculum with the right prompting.
My friends and I were excited to read the analysis behind such a feat, but after digging deeper, what we found left us surprised and disappointed.
dub.sh/gptsucksatmit
Aran Komatsuzaki @arankomatsuzaki · Jun 15
Exploring the MIT Mathematics and EECS Curriculum Using Large Language Models
Presents a comprehensive dataset of 4,550 questions and solutions from all MIT EECS courses required for obtaining a degree
arxiv.org/abs/2306.08997
Exploring the MIT Mathematics and EECS Curriculum Using Large Language Models
Sarah J. Zhang* MIT sjzhang@mit.edu
Sam Florin* MIT sflorin@mit.edu
Ariel N. Lee Boston University arielnlee@bu.edu
Eamon Nikuafs Boston University en@bu.edu
Andrei Margineau MIT atmargi@mit.edu
Annie Wang MIT annewang@mit.edu
Keith Tyser Boston University ktyser@bu.edu
Zad Chin Harvard University zadchin@college.harvard.edu
Yann Hicke Cornell University ylh8@cornell.edu


Julian Togelius
@togelius
So that paper about how GPT-4 could ace MIT's curriculum turns out to be deeply flawed in multiple ways. A great reminder that preprints are not peer-reviewed, but also that public volunteer review can be excellent (in this case, by a group of undergrads).
50
40
30
20
10
0
flower-nutria-41d.notion.site
No, GPT4 can't ace MIT
What follows is a critical analysis of "Exploring the MIT Mathematics and EECS Curriculum Using Large Language Models"
12:42 AM · Jun 18, 2023 · 257.5K Views
284 Retweets 20 Quotes 1,243 Likes 177 Bookmarks


No, GPT4 can't ace MIT
Comment
Search
Duplicate
Open Notion
No, GPT4 can't ace MIT
What follows is a critical analysis of "Exploring the MIT Mathematics and EECS Curriculum Using Large Language Models"
Paper page - Exploring the MIT Mathematics and EECS Curric...
Join the discussion on this paper page
https://huggingface.co/papers/2306.08997
This is a joint document written by three MIT EECS seniors (Class of 2024): Raunak Chowdhuri, Neil Deshmukh, and David Koplow.

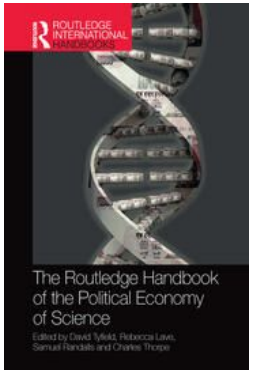
Paper

Sensemaking *about* paper



Scientific record,
Open Access

Off scientific record,
controlled by commercial
Platforms

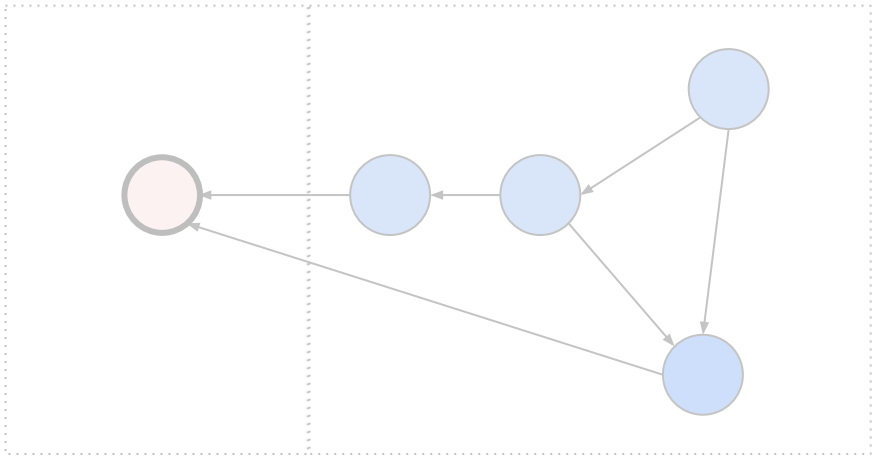


Open Access Panacea
(Muellerleile, 2017)

"...focusing too closely on ... openness may be distracting us from the ways that capital is sneaking in the back door and enclosing the ***very tools we need to make sense*** *of this new world"*

Paper

Sensemaking *about* paper



Scientific record,
Open Access

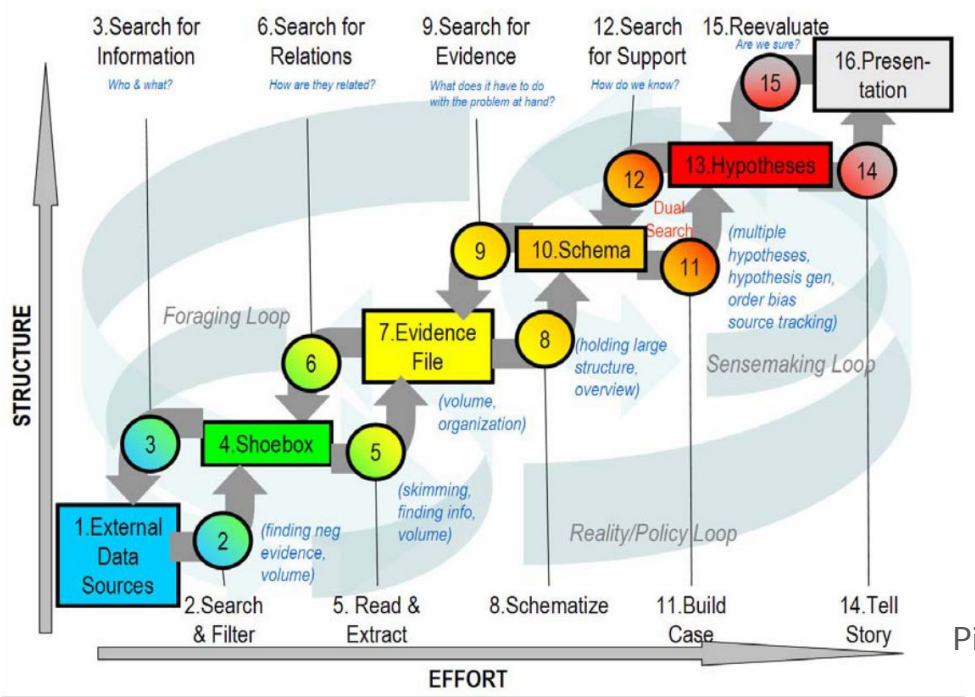
Off scientific record,
controlled by commercial
Platforms

1) What are the tools we need to “make sense of science”?

2) Why aren’t these Twitter people just using nanopubs!?

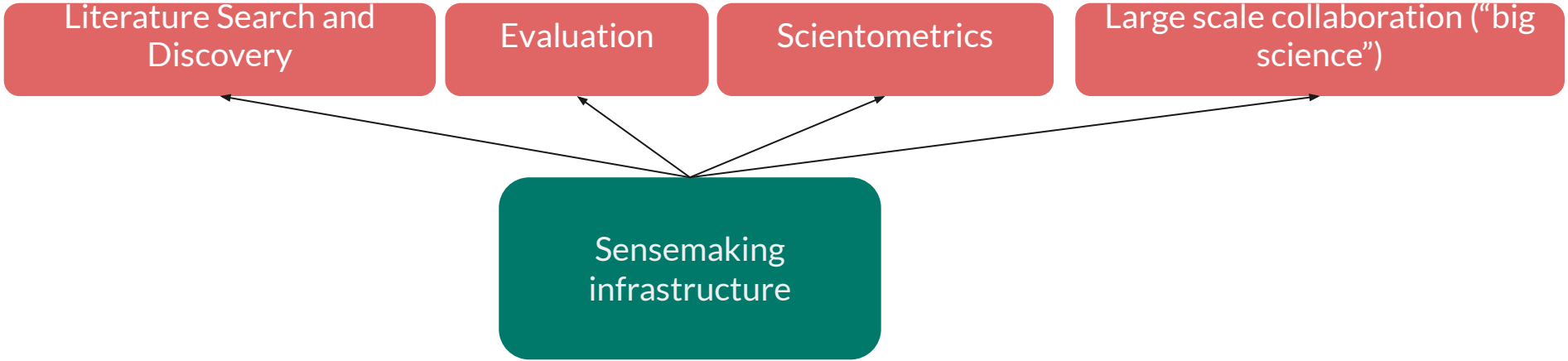
What is Sensemaking?

Organizing, synthesizing and structuring new information to support future actions

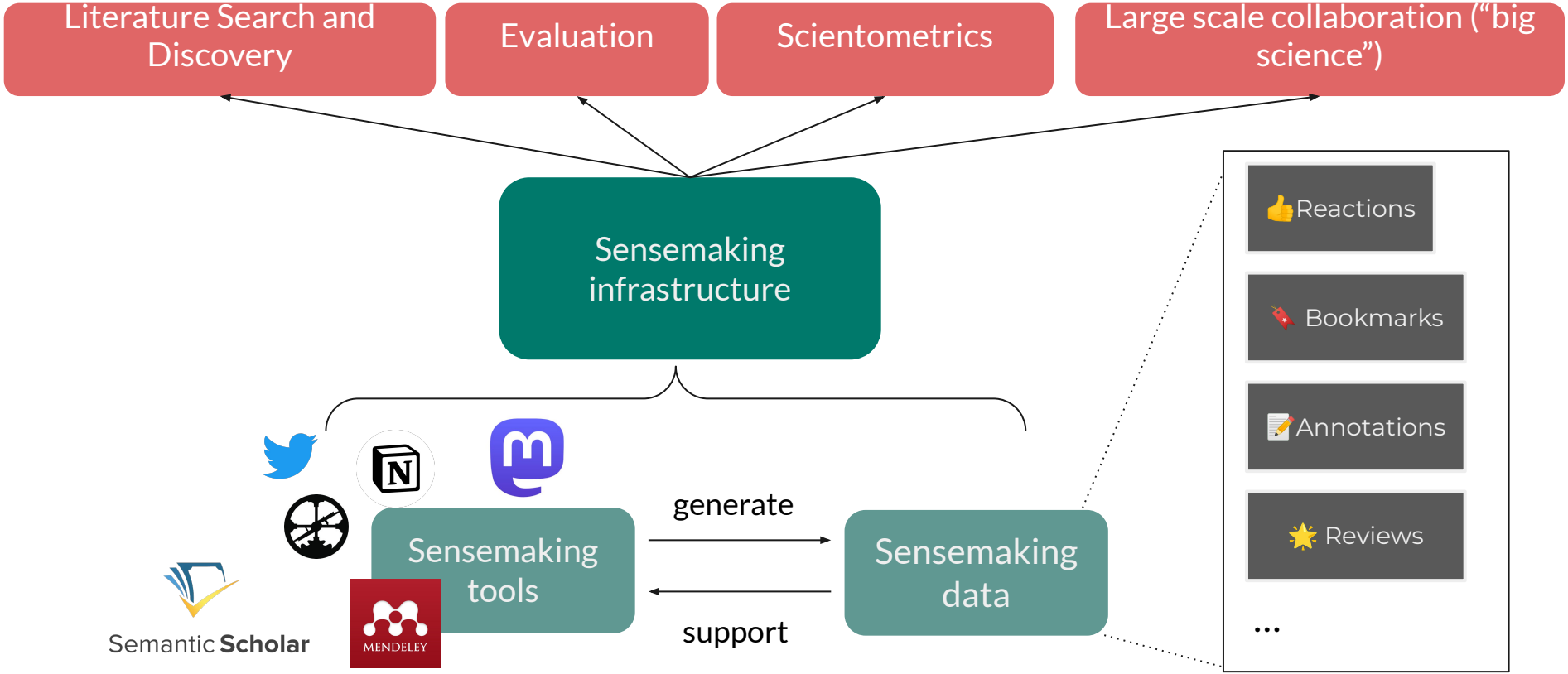


Pirolli, P., & Card, S. (2005)

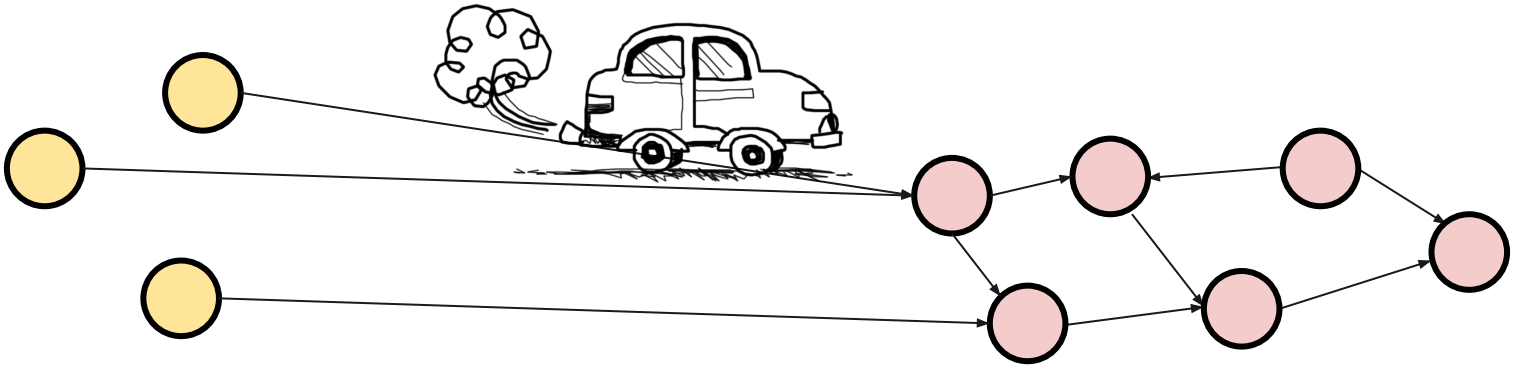
Scientific sensemaking infrastructure is important …



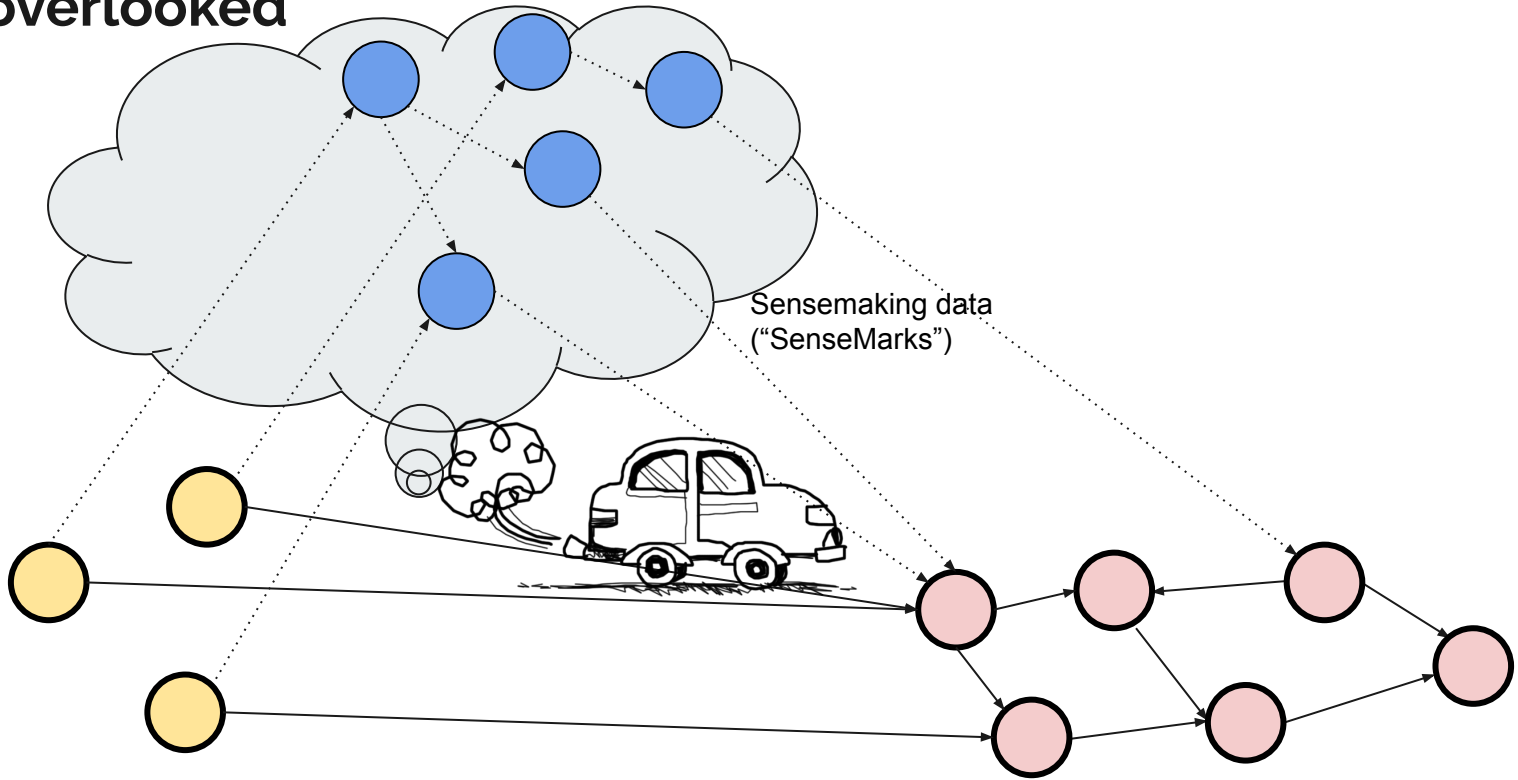
Scientific sensemaking infrastructure is important …



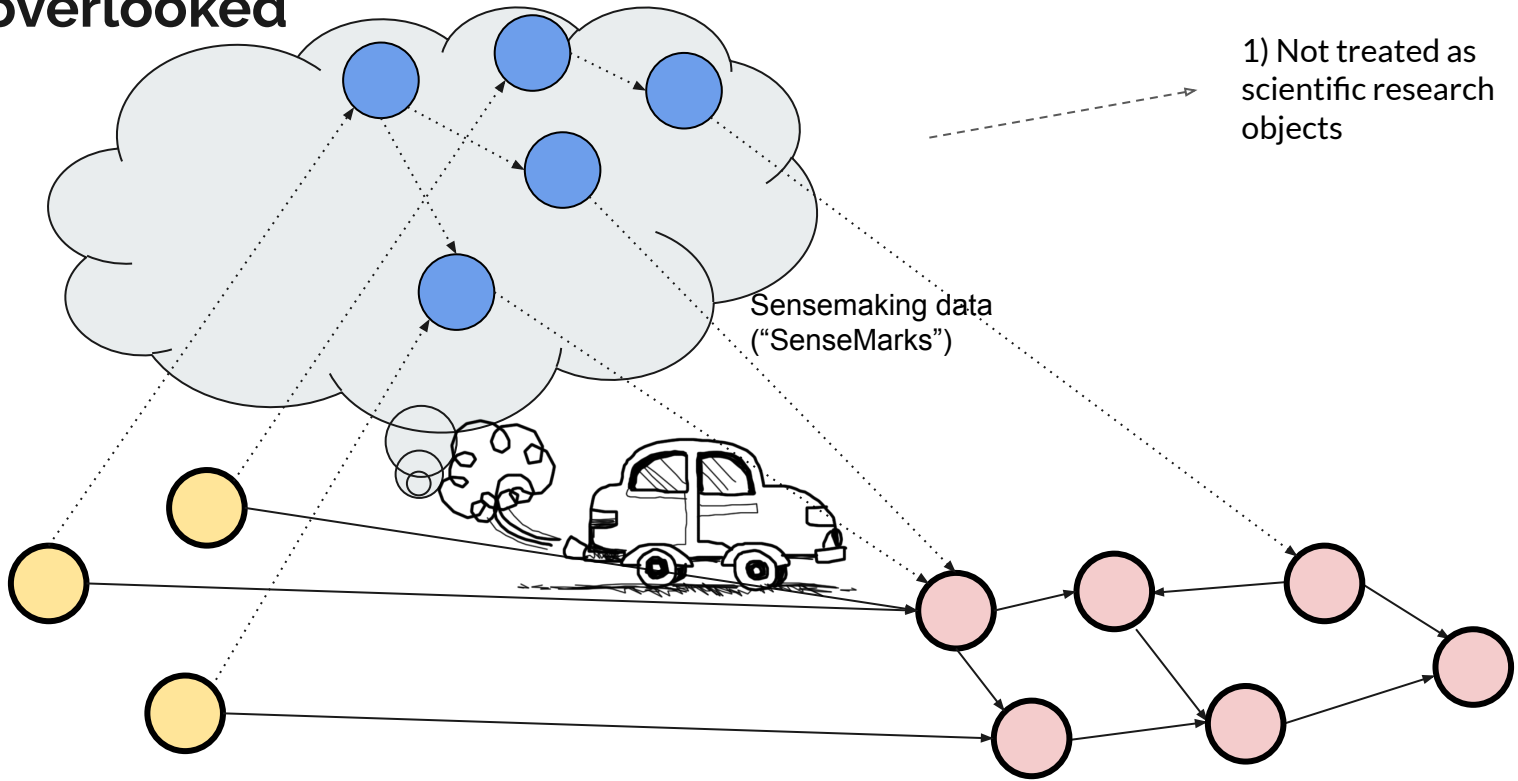
Scientific sensemaking infrastructure is important … but overlooked

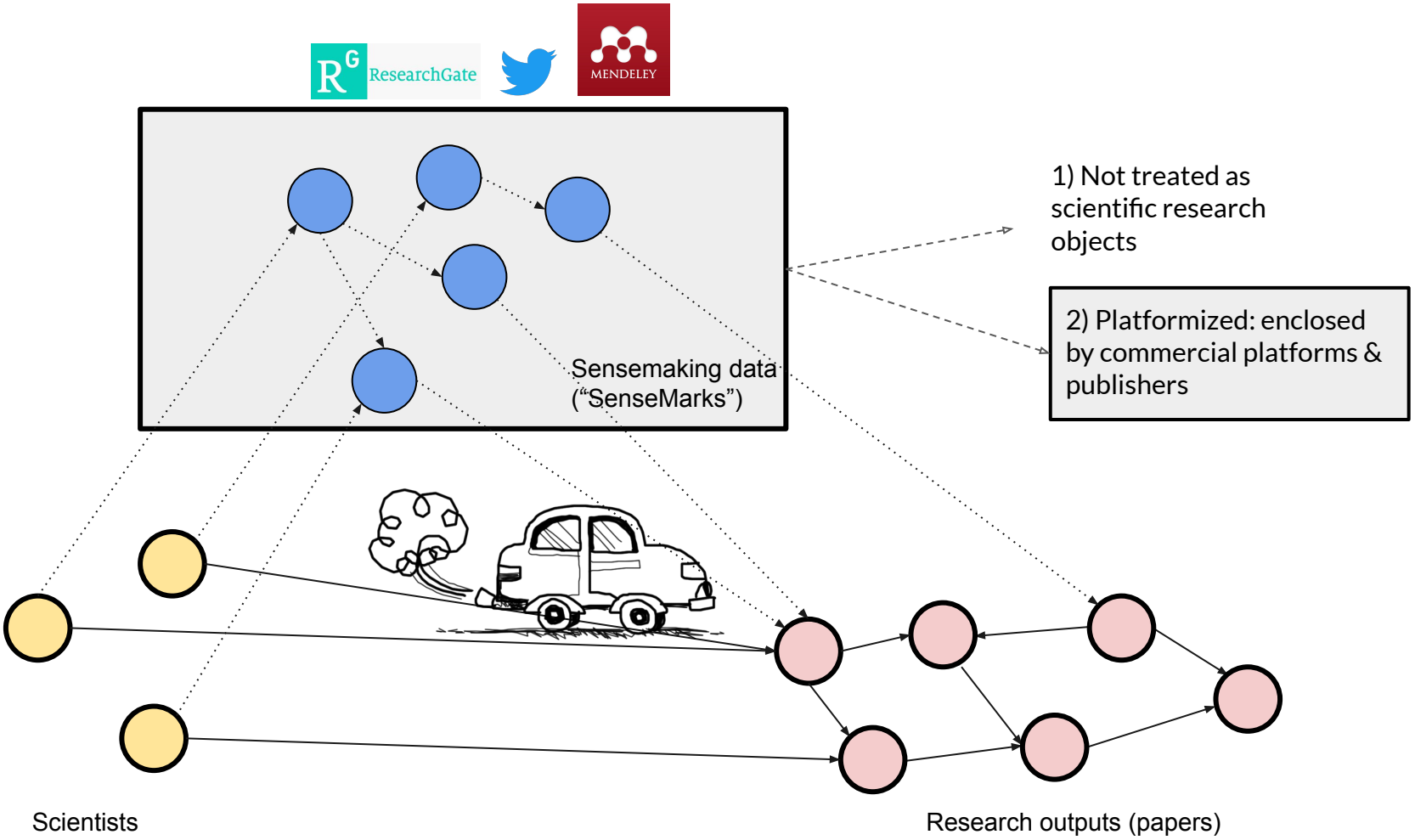


Scientific sensemaking infrastructure is important … but overlooked

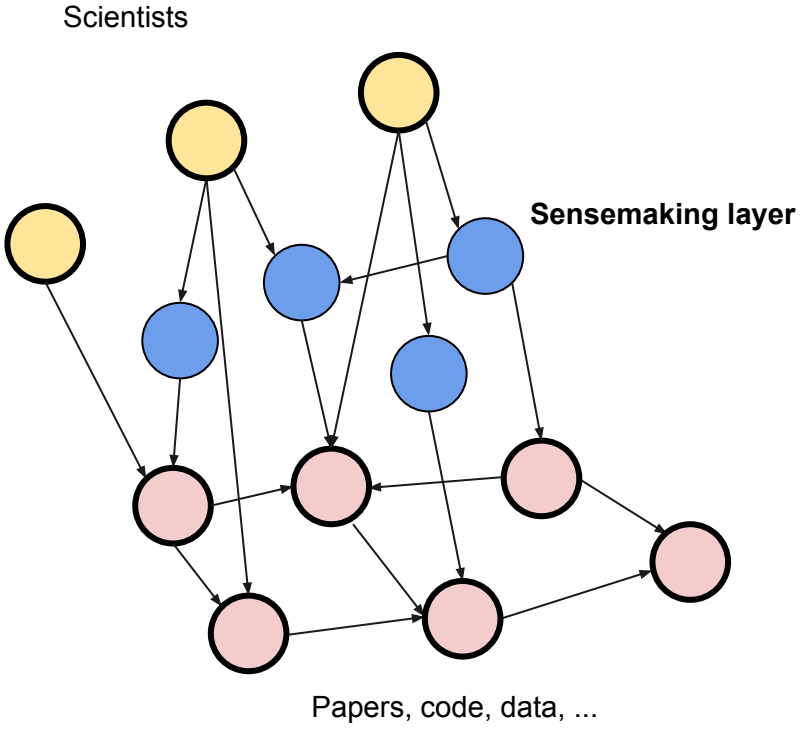


Scientific sensemaking infrastructure is important … but overlooked




ResearchGate
MENDELEY
Sensemaking data
("SenseMarks")
1) Not treated as scientific research objects
2) Platformized: enclosed by commercial platforms & publishers
Scientists
Research outputs (papers)

Open Science Sensemaking (OSSm): adding the sensemaking layer to the academic graph



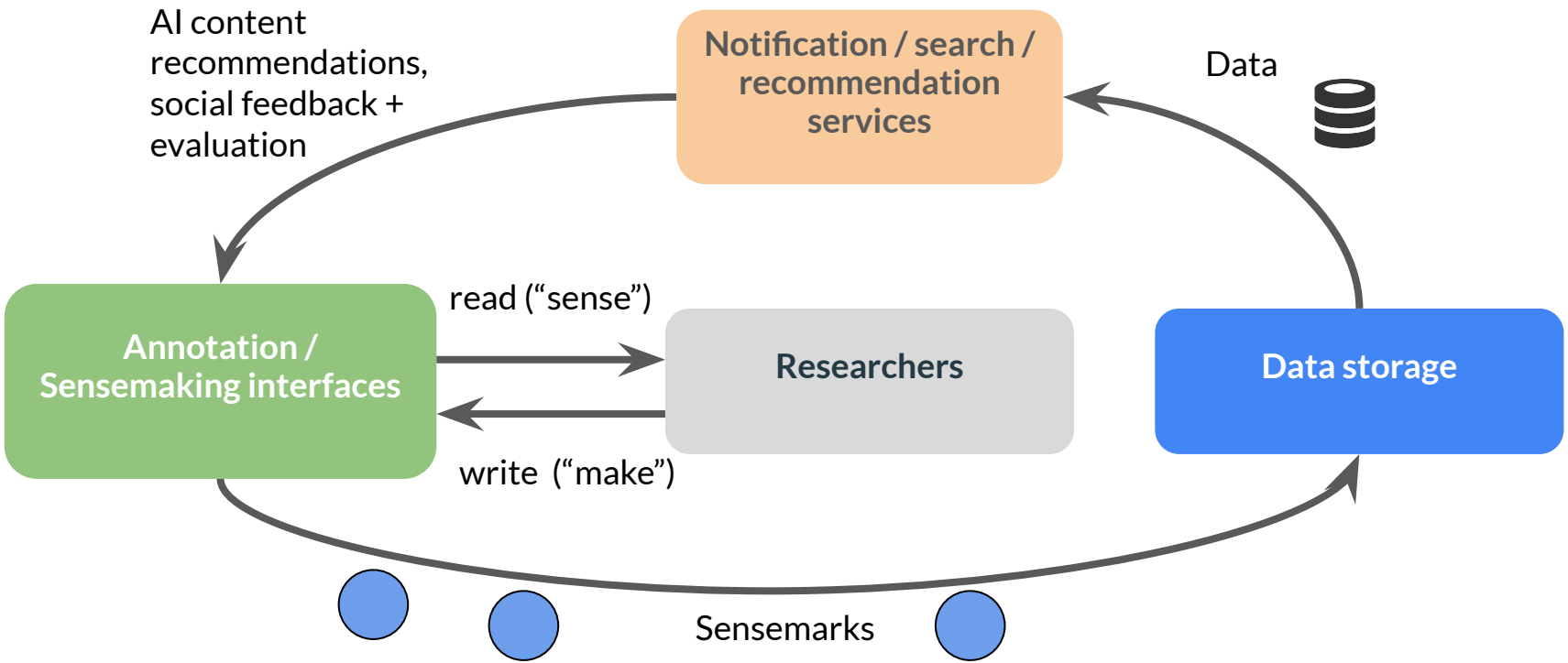
Public sensemaking data also merit treatment as scientific research outputs!

1. Open Access

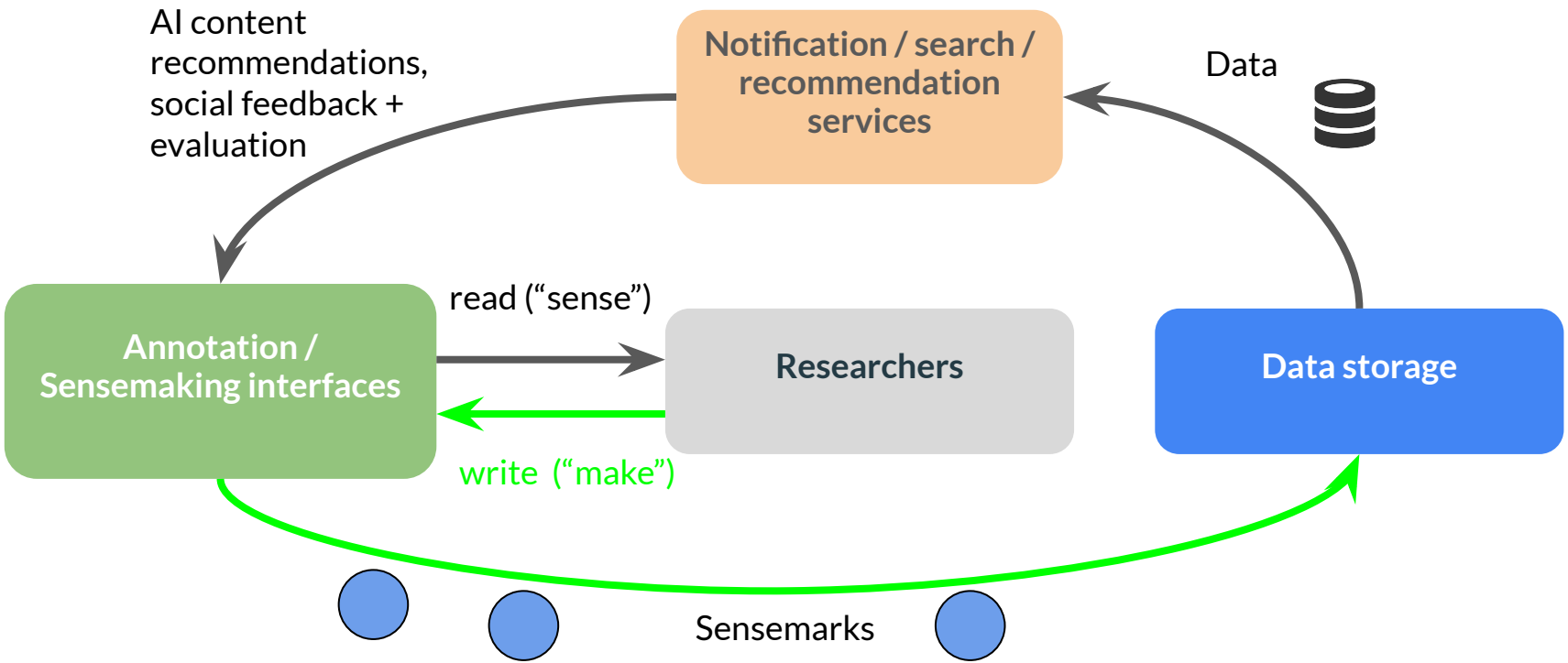
2. FAIR

3. Embedded in sensemaking feedback loops

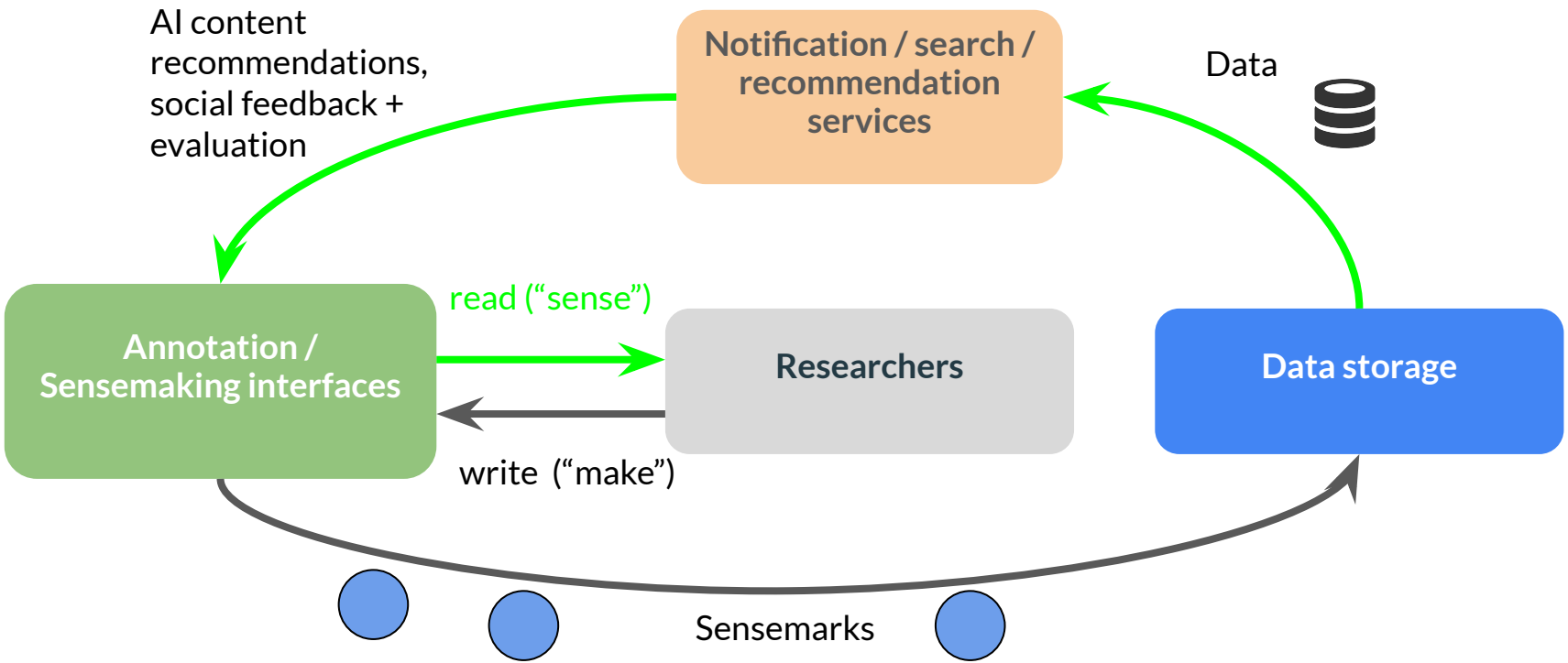
Sensemaking feedback loop



Sensemaking feedback loop

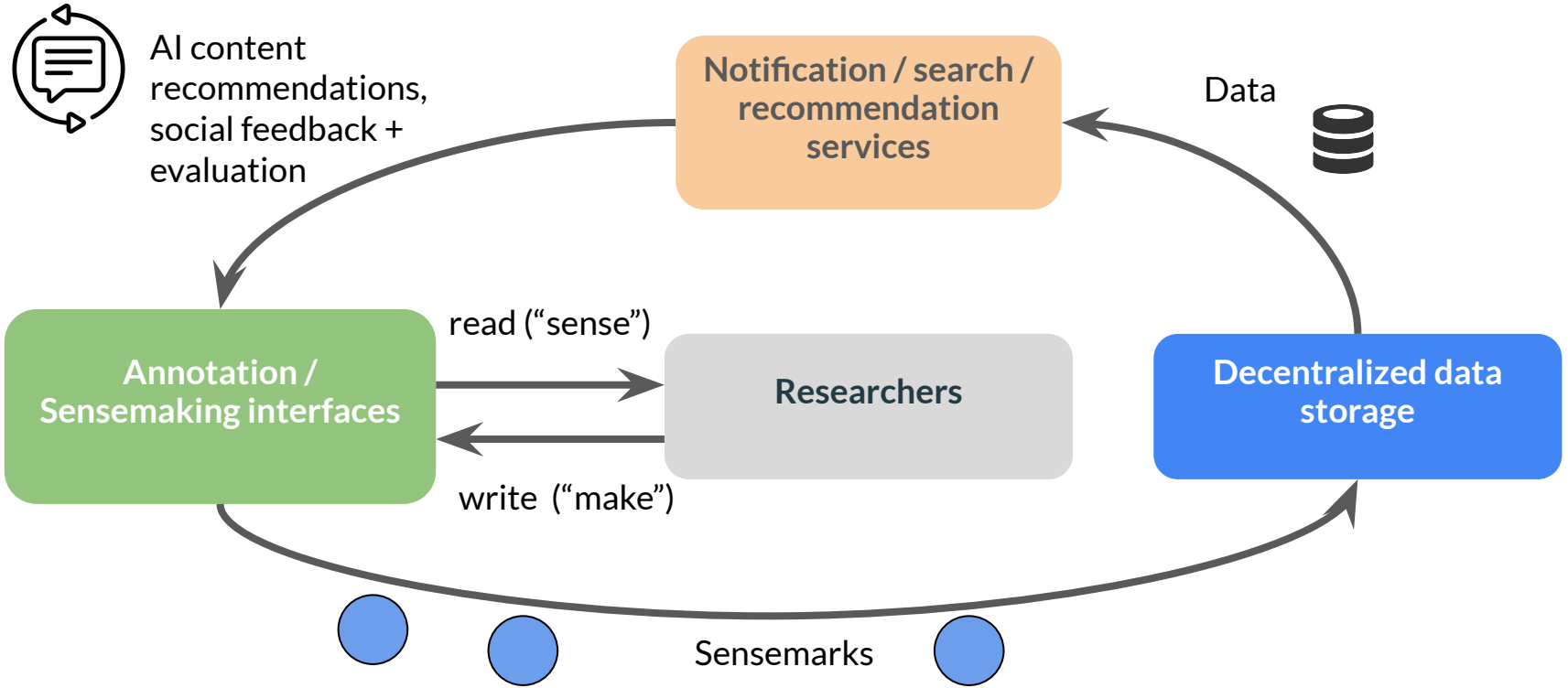


Sensemaking feedback loop



Open Science Sensemaking Network

Like social media, but decentralized + use nanopubs/sensemarks instead of posts!



Open Science Sensemaking	Nanopublications
FAIR publishing for structured expressions of “nano” scientific knowledge	
Making sense of existing knowledge	Publishing new knowledge
App layer: sensemaking feedback loops (social + content)	Data layer: storage, security
General audience (“citizen scientists”)	Meant mainly for scientists?

Thanks!

https://www.csensemakers.com/ | https://twitter.com/CSenseMakers