The Year in Search: Exploring User Search Behavior with Generative AI

Marlene van Ballegooie, University of Toronto Libraries

IGeLU Conference – Siena, Italy – September 17, 2025

Why Study User Search Queries?

Investigation into the use of AI to analyze user search queries from the Library's discovery system

Why study user search queries?

- Understand user needs and behaviors
- Highlight access and discovery barriers
- Refine discovery system relevance ranking
- Identify opportunities for library instruction
- Track emerging trends and research interests
- Inform collection development

Research Questions

Can AI be used to accurately categorize search queries and derive meaningful insights into user search behavior?

How do users interact with our discovery system?

- What subjects/topics do they search?
- What types of searches do they do? (i.e. title, author, subject/concept)
- What is the intent behind their search activity?

Project Scope

Proof of concept project

- Mixpanel – New data analytics platform for PrimoVE
- Google Gemini 2.5 Pro

Big data set

- From April 9, 2024 - April 8, 2025 = 11,388,948 user searches
- Analyzing ALL user queries was not possible due to sheer volume and current limitations with AI tools
- Analyzing only most frequent searches may hide nuances in user behavior

The Year in Search

Create a representative sample from the past year to analyze trends more comprehensively

- All queries with more than 5 searches (203,849 unique queries)
- Random sample of 1% of all queries searches 5 or less times (58,802 unique queries)
- Total data set = 262,651 unique queries, representing 2.5 million individual user queries (22% of all searches in past year)

Gemini Gems:
The AI Workhorses

Subject Categorizer

Assigned three subject terms to each search query (not based on LC Subject Headings)

Mapped each query to a top-level subject category using the Thema classification scheme

Each Thema classification mapped to a broader academic discipline:

Arts & Humanities	Clinical, Pre-Clinical & Health
Engineering & Technology	Life Sciences & Physical Sciences
Social Sciences	

Search Type Categorizer

Classified each search query into one of the following categories:

- Title (with a more granular ‘type’ attribute, i.e. journal article, book, database, etc.)
- Person/Entity
- Subject/Concept
- Unknown

User Intent Categorizer

Identified the underlying motivation behind the search

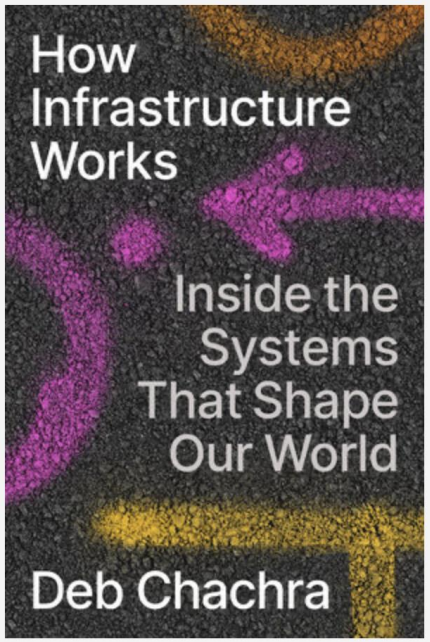
Categorized search queries into one of six user intent types:

- Specific Item Intent
- Known Entity Intent
- Specific Topic Intent
- Keyword-Driven Intent
- Exploratory Topic Intent
- Unknown Intent

AI in Action:
Categorization Examples

Examples of Title Searches

78401 | 6 | How Infrastructure Works | Civil Engineering | Urban Planning | Public Works | T - Technology, Engineering, Agriculture, Industrial processes | Engineering & Technology | title [Book] | Specific Item Intent



Add to Goodreads Look Inside

NAMED A BEST BOOK OF 2023 BY *PUBLISHERS WEEKLY*

"Revelatory, superbly written, and pulsing with wisdom and humanity, *How Infrastructure Works* is a masterpiece." —Ed Yong, author of *An Immense World*

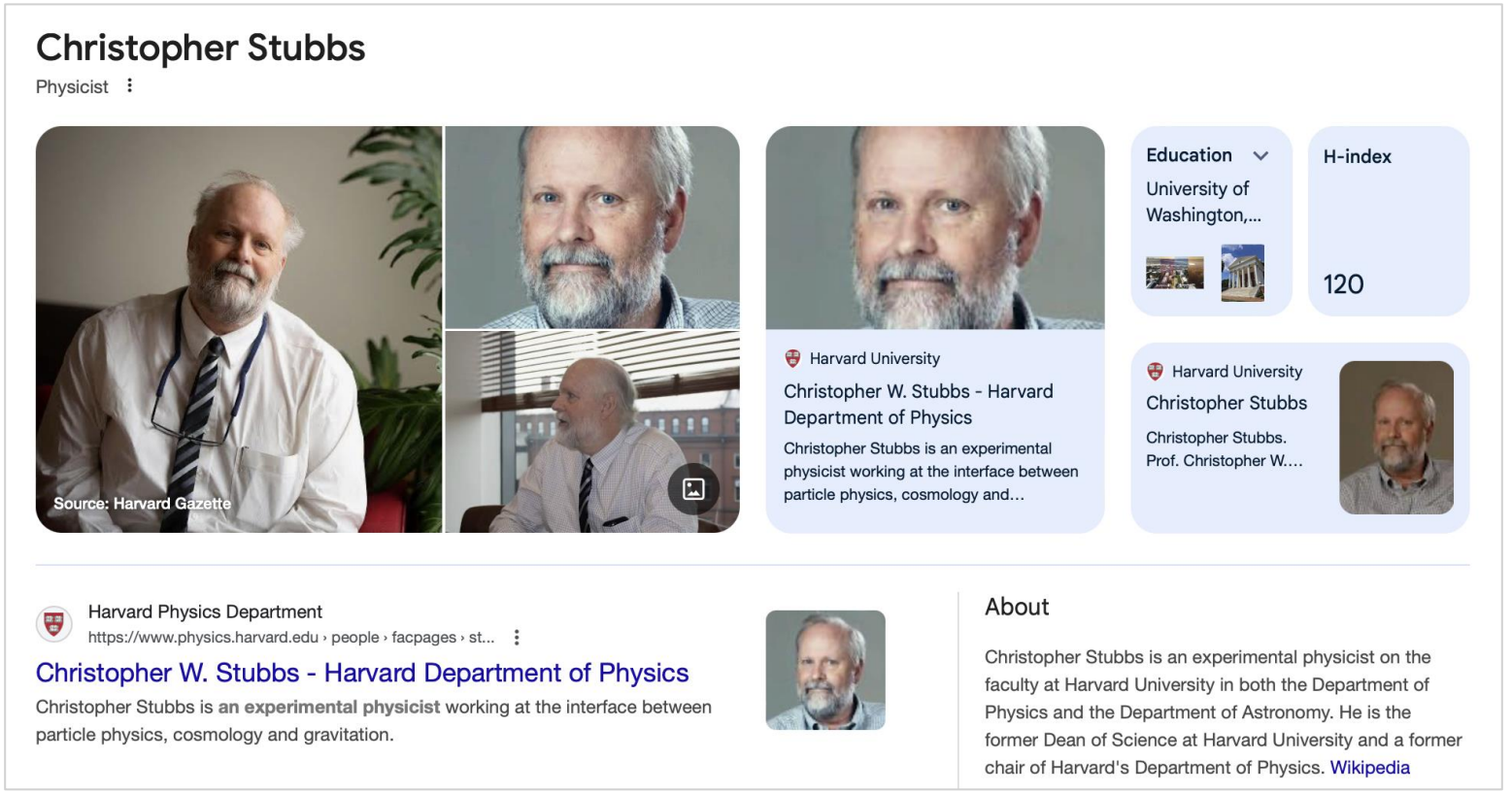
A new way of seeing the essential systems hidden inside our walls, under our streets, and all around us

Infrastructure is a marvel, meeting our basic needs and enabling lives of astounding ease and productivity that would have been unimaginable just a century ago. It is the physical manifestation of our social contract—of our ability to work collectively for the public good—and it consists of the most complex and vast technological systems ever created by humans.

A soaring bridge is an obvious infrastructural feat, but so are the mostly hidden reservoirs, transformers, sewers,... **Keep Reading**

Example of Person/Entity Searches

139052 | 11 | Stubbs, C. W. | Astrophysics | Cosmology | Experimental Physics | P – Mathematics and Science | Life Sciences & Physical Sciences | person/entity | Known Entity Intent



Example of Subject Searches

52627 | 15 | Dead Sea scrolls | Archaeology | Biblical Studies | Ancient Manuscripts | N – History and Archaeology | Arts & Humanities | subject/concept | Specific Topic Intent

What Did We Learn About Our Users?

Top 20 Searches for the Year

Number of Searches	Search String
3907	scientific method
3084	pubmed
2743	uptodate
2517	oxford english dictionary
2252	web of science
2251	alma991107272958406196
1782	oed
1668	globe and mail
1561	islam
1415	new york times
1408	toronto star
1288	chicago manual of style
1277	scopus
1149	Law of Work
1143	alma991107324123806196
1139	hinduism
1087	factiva
1028	the cambridge ancient history
916	nature
914	science

Search Frequency and User Behavior

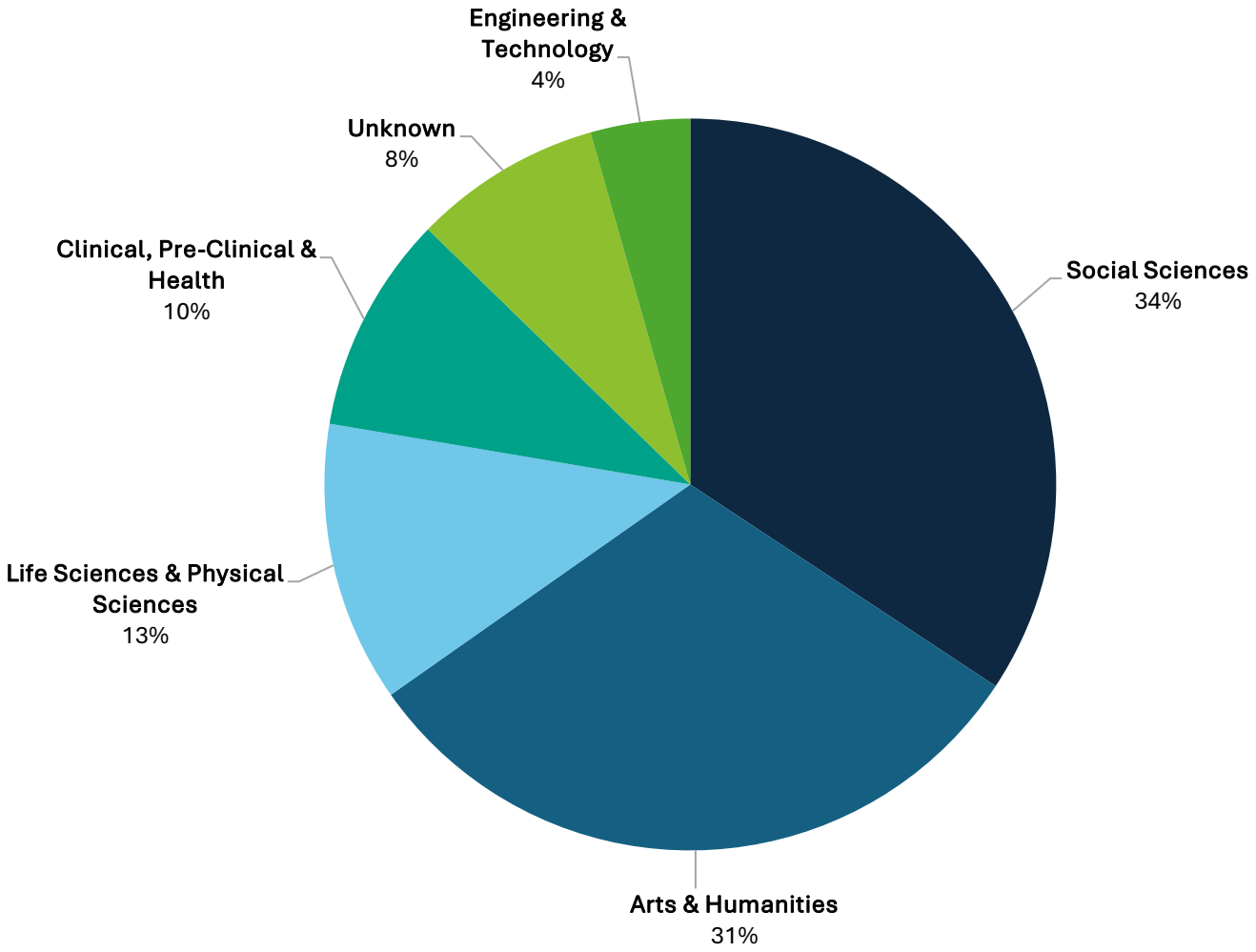
Highly skewed distribution:

- A striking 68% (4,145,357 out of 6,084,125) of search terms were only searched once, confirming a long-tail distribution.
- Only 32% of terms were searched more than once, with only 0.1% searched 10 or more times.

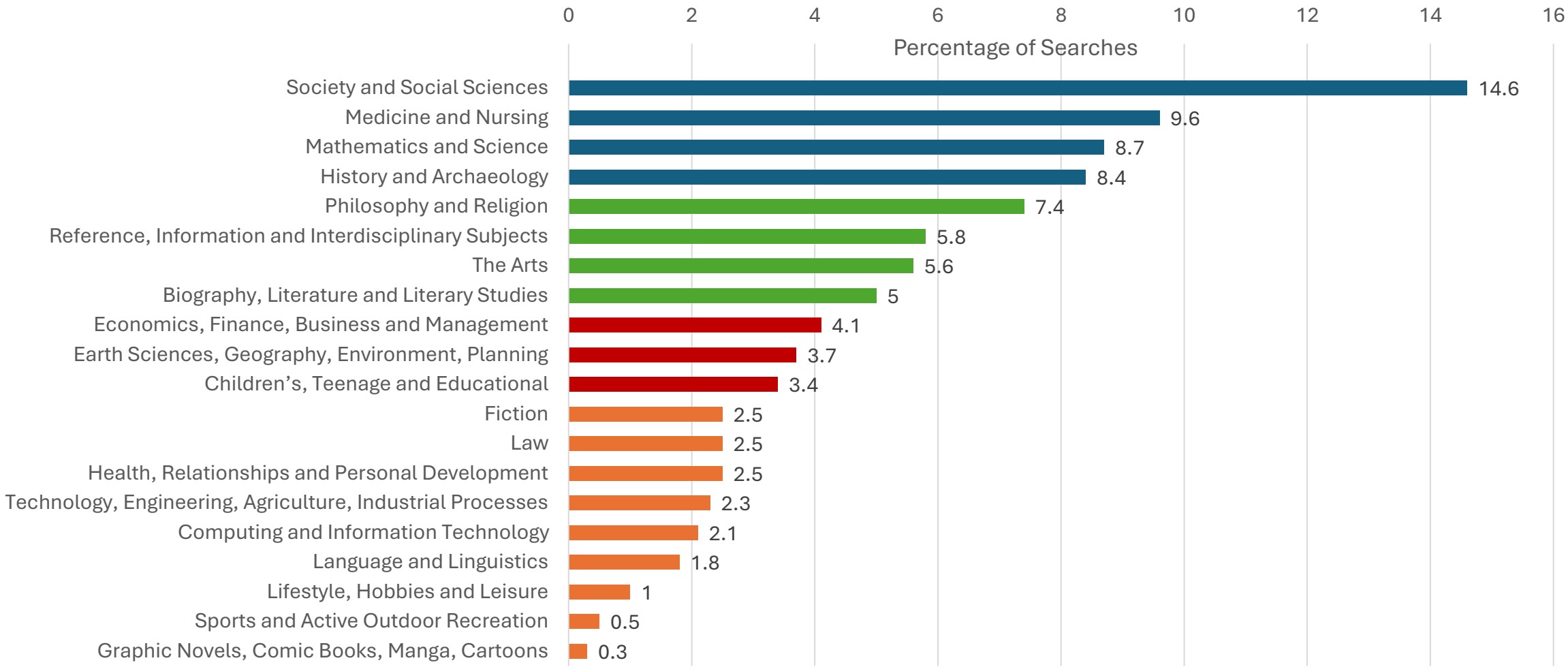
What does this distribution indicate?

- Users have highly individualized needs
- Lack of uniformity in search string construction, with many variations of the same query
- Clustering similar search queries would yield more definitive results

Categorization By Academic Discipline



Categorization By Subject Classification



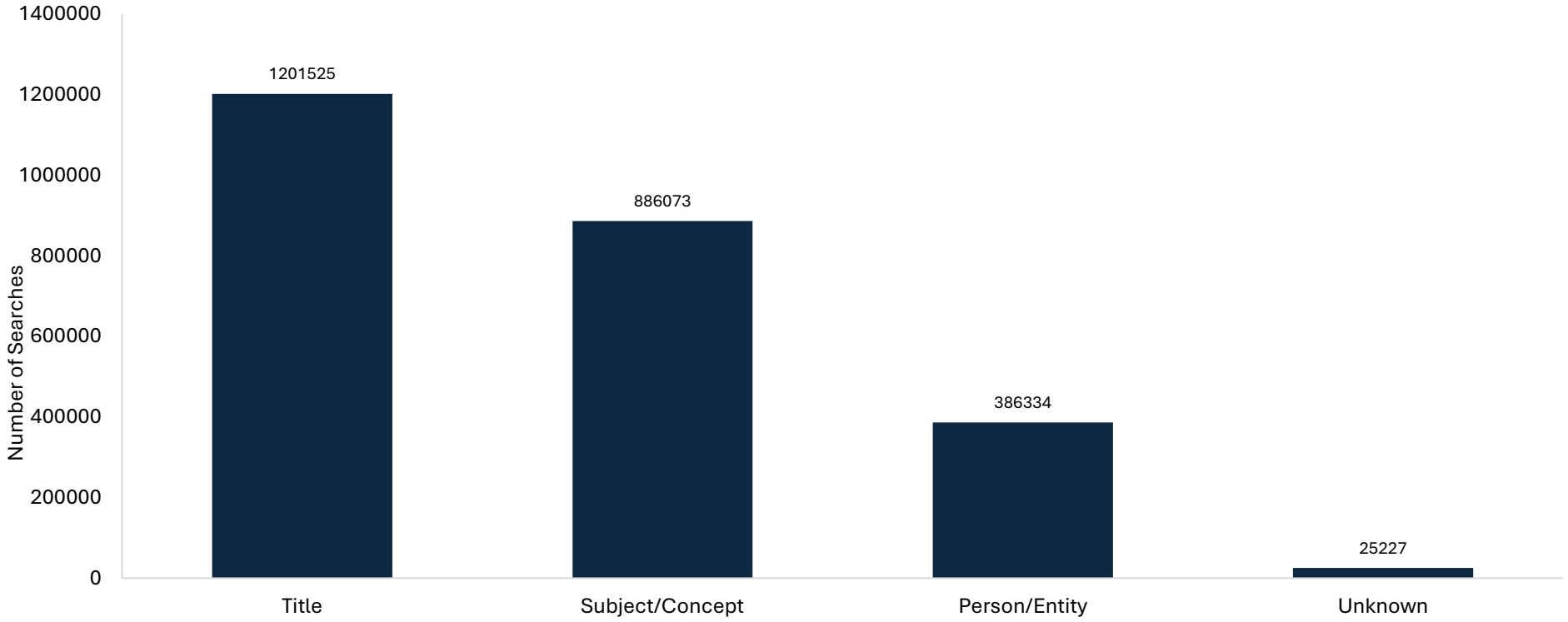
Note: About 8% of queries couldn't be assigned a category, usually because they were ambiguous personal names or lacked context.

Emerging Research Areas

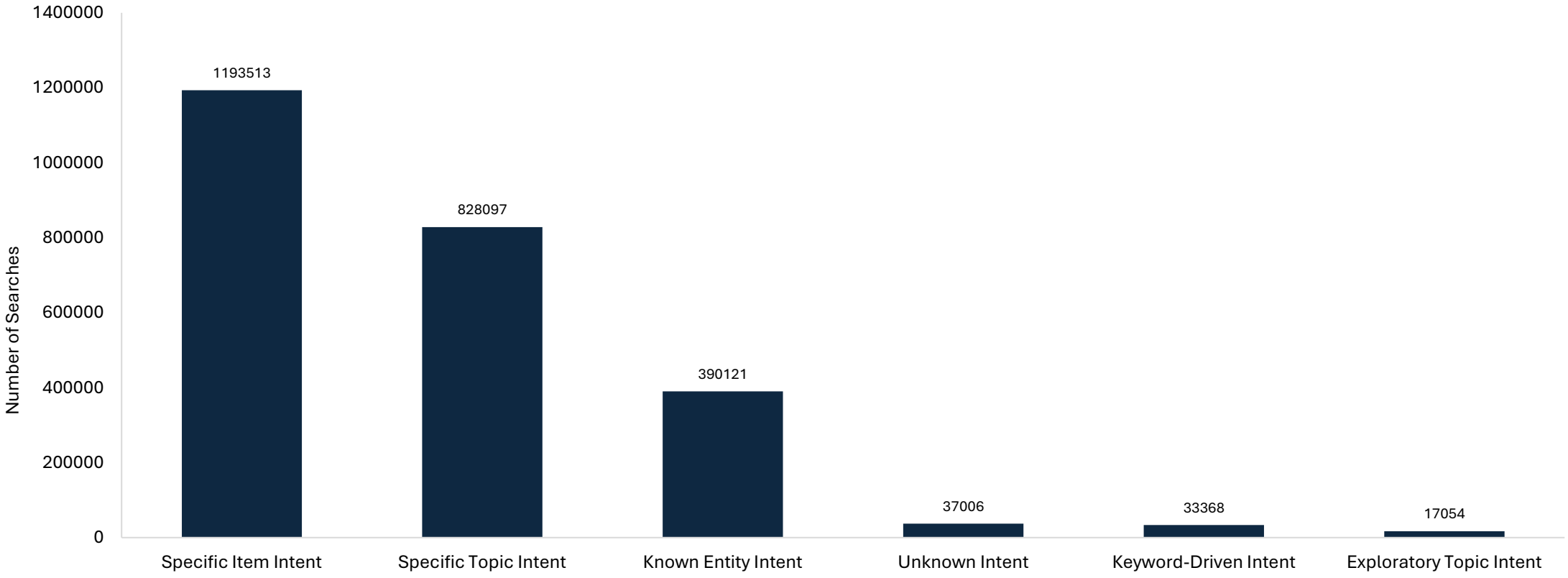
AI can assist in detecting popular and emerging areas of research:

- Artificial Intelligence & Society
- Indigenous Studies & Decolonization
- Climate Change & Sustainability
- Mental Health & Well-being
- Intersectionality & Social Justice
- Digital Media & Surveillance

Categorization By Search Type



Categorization By User Intent



Top Insights from the Research

Key Takeaways About User Behavior

- Search begins ELSEWHERE (Google, Google Scholar, PubMed, social media, etc.)
- Frequent full citations or author-title queries suggest known-item retrieval dominates over exploratory searching

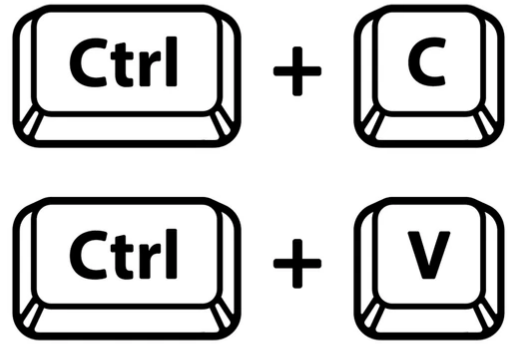
Key Takeaways About User Behavior

Users prioritize efficiency and frequently rely on copy-paste to enter search terms

Long, encoded query strings (e.g., %26 for &) suggest inputs are pasted from syllabi, citation managers, or bibliographies

Unexpected copy-paste patterns observed in queries:

- Full abstract text
- Publication extent statements (e.g., 256 p. ; ill.)
- The letter ‘v’ – unsuccessful paste action



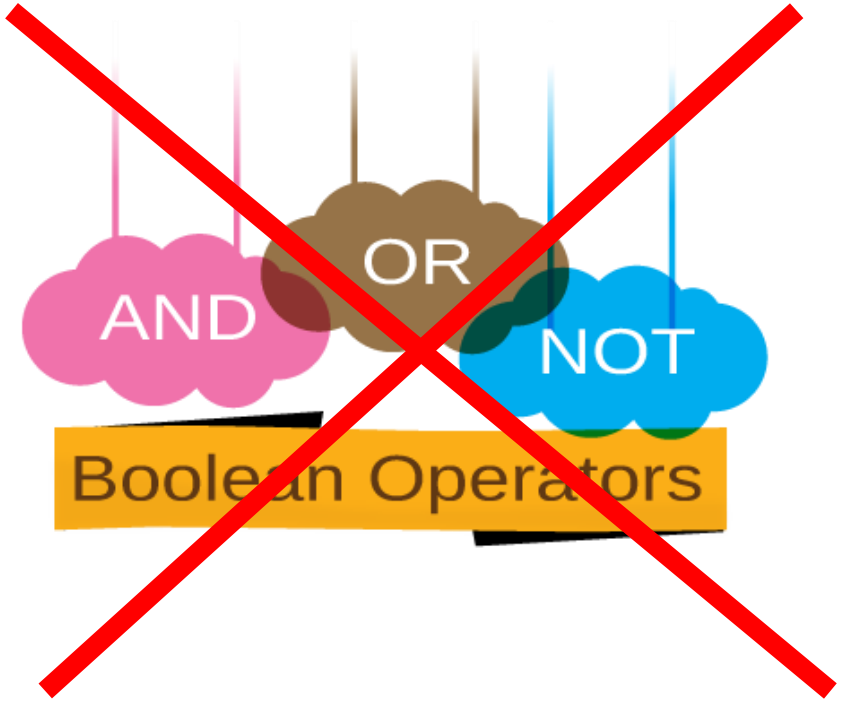
Key Takeaways About User Behavior

Single search box is heavily preferred; users rarely engage with advanced search interface

Fielded searching and Boolean logic (AND, OR, NOT) are infrequently used

Quotation marks and parentheses to control search logic are seldom applied

Precision search techniques remain uncommon despite their potential to improve results

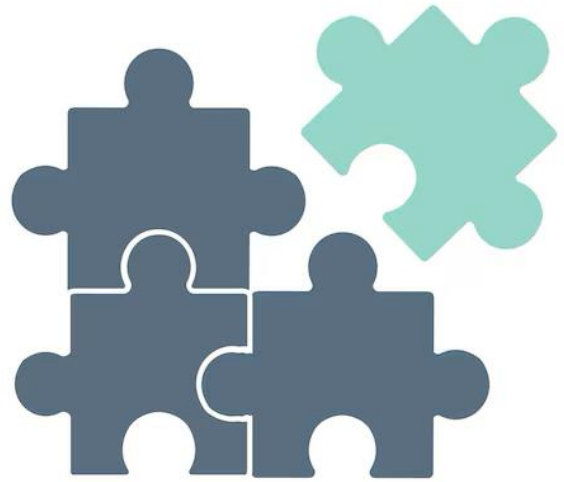


Current State of AI Research Tools

Most AI integrations today are siloed chatbots bolted onto existing research platforms.

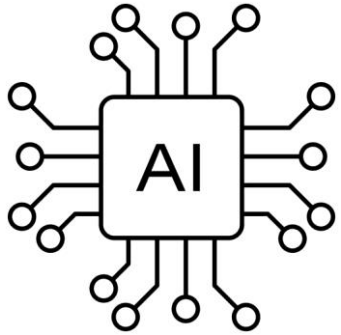
These implementations offer a quick win for vendors eager to showcase AI capabilities.

For users, the result is often a fragmented search experience, leading to uncertainty about where or how to begin their research.



A Better Path Forward

From siloed “research assistants” to integrated “research assistance”



AI Mediated Search

Focused result set based on identified search type and user intent

thank you!

Marlene van Ballegooie
m.vanballegooie@utoronto.ca