

## 11.9 Images

A considerable portion of ML concerns imaging or has images as its subject matter. There are various librarianship related problems here e.g. recognizing or classifying images, attaching metadata to images, searching among a collection of images, perhaps where the input is itself an image, and so forth. Images are reasonably important within librarianship. There are libraries with images as part of their collections. But also, many text sources, for example newspapers, have images or diagrams within their content. Additionally, there is film or video.

Such sources can be digitized, and from there processed using ML for certain tasks. It all becomes a matter of numbers and patterns in numbers. Some modern foundation models are, or aspire to be, 'multi-modal'. This means that they can work with source and output in any medium (e.g. text, or images, etc.)

An example is Benjamin Lee et al. *The Newspaper Navigator Dataset: Extracting And Analyzing Visual Content from 16 Million Historic Newspaper Pages in Chronicling America* (Lee et al. 2020). They write in their abstract:

... we introduce a visual content recognition model trained on bounding box annotations of photographs, illustrations, maps, comics, and editorial cartoons collected as part of the Library of Congress's Beyond Words crowdsourcing initiative and augmented with additional annotations including those of headlines and advertisements. We describe our pipeline that utilizes this deep learning model to extract 7 classes of visual

content: headlines, photographs, illustrations, maps, comics, editorial cartoons, and advertisements, complete with textual content such as captions derived from the METS/ALTO OCR, as well as image embeddings for fast image similarity querying.

We have here machine learning, crowdsourcing, additional annotations... on 16 million pages. Respect!

## 11.10 Annotated Readings for Chapter 11

Sanji, Majideh, Hassan Behzadi, and Gisu Gomroki. “Chatbot: An Intelligent Tool for Libraries.” *Library Hi Tech News* ahead-of-print (2022).  
<https://doi.org/10.1108/LHTN-01-2021-0002>. (Sanji, Behzadi, and Gomroki 2022).

Tay, Aaron. “List of Innovative Literature Mapping Tools | Aaron Tay’s Musings about Librarianship,” 2022. <https://musingsaboutlibrarianship.blogspot.com/p/list-of-innovative-literature-mapping.html?view=classic>. (Tay 2022).

Apparently there have been robots in libraries for some time. Here, from 1965, is Lawrence Lipton demonstrating his 'robot' Duhab (Detector of Undesirable Habitués). Duhab finds undesirables such as censors and book burners.



**Figure 27. Duhab**  
(<https://digital.library.ucla.edu/catalog/ark:/21198/zz0002tw94>  
[https://en.wikipedia.org/wiki/File:Lawrence\\_Lipton\\_and\\_DUHAB.jpg](https://en.wikipedia.org/wiki/File:Lawrence_Lipton_and_DUHAB.jpg)).

# Chapter 12: Librarians as Sentries

## 12.1 Copyright and Intellectual Property

As librarians license access to content from vendors, we need to ensure that contracts do not preclude our users from conducting text and data-mining research, algorithmically based research, and machine learning. (Miller 2020)

## 12.2 Intellectual Freedom

ChatGPT tells us that intellectual freedom has several drawbacks, including:

- Misinformation and false ideas: People may spread false or harmful ideas that are protected under the umbrella of intellectual freedom.
- Conflict with other values: Intellectual freedom may conflict with other values such as privacy, security, and safety, leading to difficult ethical dilemmas.
- Limits to freedom: In some cases, laws and regulations may limit intellectual freedom in order to protect public interest or prevent harm.
- Divide between different perspectives: Intellectual freedom may lead to a fragmentation of perspectives and ideas, making it more difficult for society to reach consensus on important issues.

- Resistance to change: Some individuals and groups may resist new ideas and perspectives, leading to resistance to social and cultural change.

[OpenAI . (2023) This paragraph was written with ChatGPT.]

### **12.3 Censorship and Algorithmic Curation**

Librarians can defend against censorship. ML is not directly a censor, but many of the techniques it enables and improves— such as personalization, recommendation, search— can lead to inadvertent censorship. The techniques can lead to filters and to placing individual patrons within their own bubbles, or silos, of information to the exclusion of the wider world of information. Customization is good and will presumably be appropriate and well received by patrons. But censorship is bad. There needs to be a balance. Removing inappropriate and offensive content entirely can be good, but it is hard to draw the boundary lines. Social media are also present. One would not ordinarily think extensively about social media in the context of libraries, but social media is a gatekeeper to some, perhaps even most, of the information that students and patrons obtain. There is also the 'manipulative capabilities of algorithmic processes'— the abilities of algorithmic processes to influence a person's thought, emotions, and decisions (IFLA 2020, 5). In some ways this is not a lot different to advertising or propaganda. But ML is more powerful and less obvious and transparent. Librarians might have a role as sentries here. They have a history of protecting patrons, and to arm wrestling with censorship.

Some AI tools improve access to information, but there is a need to check that the improvements are available to all including to people with disabilities.

There is also inadvertent censorship. ChatGPT tells us:

Inadvertent censorship refers to the unintended restriction of information or expression due to various factors. Here are some examples:

- **Technical difficulties:** Technical glitches or failures in communication systems can unintentionally restrict access to information or expression.
- **Overzealous filtering:** Some internet filtering systems, designed to protect users from harmful content, can inadvertently censor legitimate information or expression.
- **Misapplication of laws:** Laws and regulations designed to protect national security, public order, or public morals can be misapplied, leading to the unintentional restriction of information or expression.
- **Economic considerations:** Economic factors, such as the cost of publishing or distributing materials, can lead to the unintentional restriction of information or expression.
- **Social norms:** Social norms and cultural biases can lead to the unintentional restriction of information or expression that goes against the prevailing norms and biases.

[OpenAI . (2023) This paragraph was written with ChatGPT.]

## 12.4 Privacy

Centrally, this concerns Users' behavior and patron data. Privacy is a core value of librarianship (American Library Association 2006; E. Berman 2018). The main justification for this is to allow patrons intellectual freedom.

Lack of privacy and confidentiality has a chilling effect on users' selection, access to, and use of library resources (American Library Association 2006).

Modern computerized libraries typically use Integrated Library Systems (ILS) (sometimes known as Library Management Systems (LMS)) to manage their services. These systems collect data, some of it about individual patrons (e.g. name, age, address, email, phone, driver's license, gender, borrowing history, etc.) (E. Berman 2018). Outside agencies, for example credit reporting, and credit card, companies also collect data about individuals. It is possible to collate data across different systems, for example to use library patron data in conjunction with credit card data. Collation like this— data collected for one purpose being used without consent for another— should never happen. But it would be wishful thinking to suppose that it never does. There are issues here of privacy and consent.

Librarians might be watchful that ML applications meet the required legal and ethical standard.

## **12.5 Bias**

ML systems can be biased, and often will be if they are LLMs pre-trained by self-supervision on large quantities of internet sourced digitized text. Some of these biases can be corrected relatively easily by prompting, fine-tuning, or plugins. But it is helpful to all if it is known what the biases are. Librarians can be on guard to ensure that their services are fair, equitable, and do not commit harms of representation or omission. As mentioned earlier, librarians do manage bias in many aspects of their work, for example with collection development. Most of these traditional forms of potential biases are nothing to do with machine learning, nor have been caused by machine learning. However, ML may be able to help with redressing them.

## **12.6 Social Epistemology**

### ***12.6.1 Reliability, Validity, and Over Confidence***

Most machine learning systems can make mistakes. Generally, they work in terms of probabilities and will not give the same answer twice. This means that there is a problem with reliability. Further, if the different answers contradict or are inconsistent with each other, some answers must be plain wrong. This means that there is a problem with validity. In research in general, if the researcher has an instrument that is unreliable and lacking in

validity, typically that instrument would just be discarded. The probabilities here also concealed. An LLM would not ordinarily say 'there is a 70% chance of rain and a 30% chance of sunshine'. Rather, it would say definitively 'it will rain!' about 70% of the time, and, definitively, 'there will be sunshine!' about 30% of the time. The LLM displays an exaggerated confidence that a user might have difficulty in assessing and understanding. Librarians can help patrons with attitudes to information that ML systems provide.

### ***12.6.2 Confirmation Bias and Poor Reasoning***

As mentioned in Section 11.7, librarians, assisted by AI, can pay attention to cognitive biases.

### ***12.6.3 Misinformation***

Librarians can guard against misinformation. They can ensure that the information provided to patrons is accurate, reliable, and up to date.

### ***12.6.4 Awareness of the Digital Literacy of Patrons***

Librarians can have a sense of the information and digital literacy among the community of patrons in their libraries. If there are shortcomings, the

librarians can provide further guidance, education, and support, especially with the question of how ML impacts the information patrons receive.

Going forward, ML is going to provide many systems and tools for librarianship. Librarians themselves can evaluate these tools to ensure they meet the needs of their patrons and the communities.

## **12.7 Chatbots**

An eye needs to be kept on chatbots. There is the well-known historical example of Microsoft's Tay from 2016 (Wikipedia 2023n). Once in the wild and available to all, this chatbot degenerated into sexist and racist garbage within a day. This is an old example now. Also, it was configured to learn from the people it was talking to, and bad actors set out to lead it astray. They succeeded.

Chatbots can have several drawbacks, including:

- **Limited understanding:** Chatbots are limited in their understanding of human language and context, leading to misunderstandings and inaccurate responses.
- **Lack of emotions:** Chatbots lack emotions and empathy, making it difficult for them to fully understand human behavior and respond in a meaningful way.

- No creativity: Chatbots lack creativity and imagination, which can limit their ability to provide innovative solutions to complex problems.
- Inflexibility: Chatbots are designed to follow specific scripts and rules, making it difficult for them to adapt to changing circumstances or new information.
- Reliance on data: Chatbots rely on data and algorithms to generate responses, which can perpetuate biases and perpetuate harmful information.
- Security risks: Chatbots can be vulnerable to hacking and other forms of cyber-attacks, which can compromise sensitive information and put users at risk.
- Technical limitations: Chatbots can be limited by technical constraints, such as limited processing power, storage capacity, and connectivity.

[OpenAI . (2023) This paragraph was written with ChatGPT.]

## **12.8 Personalization and Paternalism**

Search usually will rank the links it returns. If there is ranking, or personalization, there is further opportunity for error or malfeasance. Most obviously the finger is on the scales with systems that allow sites to pay the search providers for higher positions in the ranked returns. ML systems need to be scrutinized in these areas.

Paternalism ‘is to act for the good of another person without that person’s consent, as parents do for children’ (Suber 1999). Generally speaking, outside of concerns of machine learning:

- parents being paternalistic to their own children is fine,
- some adults, such as teachers, being paternalistic to other peoples’ children can be fine,
- some adults, such as doctors and nurses, being paternalistic to adults can be fine,
- otherwise, adults being paternalistic to other full competent adults is often suspect.

There is an extensive literature on paternalism in librarianship, especially in so far as that might apply to children. (See, for example, (Frické, Mathiesen, and Fallis 2000).) This is a complex area. But there is a statement from John N. Berry that gets to the nub of the issue:

Nor must we deprive [children] of the nurture, the helping hand, the guidance, the tools for seeking truth and knowing when it is discovered. We cannot simply turn them loose in our jaded information society without helping them understand that some of the information is false, is evil, is dangerous, is misleading, or is ambiguous... That may not be a legal obligation, but it is clearly a moral duty for every librarian, every teacher, every parent and person in a free society (Berry 1998).

(Berry wrote this in 1998, long before LLMs and a powerful internet.)

The present text is primarily on research libraries and advancing knowledge. Children are not front and center here. Nevertheless, LLMs have the potential to act for good and to cause harm, both to children and to adults. Attention should be paid to questions of when paternalism is appropriate and when it should be avoided. Helping folk understand that some ‘information’ from LLMs is false, evil, dangerous, misleading, or ambiguous is a good idea, even if it might amount to paternalism.

## **12.9 Images and Facial Recognition Technology**

Attaching metadata to images is going to have its adventures. It is going to make mistakes. There are going to be false positives and false negatives. What turns on this is an open question. A notorious example is from 2015 when Google Photos misclassified Jacky Alcíné and his friend, who are both black, as being gorillas. It is not really known why this happened, nor, it seems, has it been fixed (Grant and Hill 2023). The author has had a mildly similar experience himself. In 2016, Richard Lee, a New Zealand man of Asian descent, had his passport photograph rejected by software because 'his eyes were closed'. Lee, an engineering student, told Reuters 'No hard feelings on my part, I've always had very small eyes and facial recognition technology is relatively new and unsophisticated' (Reuters 2016). (Give him a medal for saying that.) The present author is a New Zealander, and around 2016 he tried to renew his New Zealand passport and the passport robot rejected his photographs because 'his eyes were closed'. A couple of matchsticks and Photoshop fixed that!

Attention does need to be paid to the attachment of metadata to images.

Facial recognition technology in a general setting is discussed earlier in Section 7.7.3 (Buolamwini et al. 2020; American Library Association 2018). Our guess is that it will be used in libraries, without controversy, wherever and whenever patrons need to establish their identities e.g. when borrowing books. Tracking patrons is another matter. Attention needs to be paid to it.

## **12.10 Losing Jobs**

There is a standard answer to questions of automation causing loss of jobs. It is: automation automates routine repetitive jobs thus freeing up workers to do more complex and valuable tasks. Of course, this does not mean that there will be no loss of jobs. There is a standard text on this question. It is Kenning Arlitsch and Bruce Newell's *Thriving in the Age of Accelerations: A Brief Look at the Societal Effects of Artificial Intelligence and the Opportunities for Libraries* (Arlitsch and Newell 2017). It is slightly old in that since 2017 LLMs have opened the floodgates on what AI can do. Daron Acemoglu has a briefing *Get Ready for the Great AI Disappointment* which argues that the effects of LLMs in the near term will prove to be 'so-so automation' which perhaps displaces workers but without large gains in productivity (Acemoglu 2024).

## 12.11 Annotated Readings for Chapter 12

Acemoglu, Daron. “Get Ready for the Great AI Disappointment.” *Wired*, 2024.  
<https://www.wired.com/story/get-ready-for-the-great-ai-disappointment/>.  
(Acemoglu 2024)

Arlitsch, Kenning, and Bruce Newell. “Thriving in the Age of Accelerations: A Brief Look at the Societal Effects of Artificial Intelligence and the Opportunities for Libraries.” *Journal of Library Administration* 57, no. 7 (2017): 789–98.  
<https://doi.org/10.1080/01930826.2017.1362912>. (Arlitsch and Newell 2017)  
This has useful material on the question of whether library jobs will be lost to automation, and, if so, which positions.

Cook, John. “Cranky Uncle.” *Cranky Uncle*, 2023. <https://crankyuncle.com/>. (Cook 2023) One way of combatting misinformation.

Tait, Elizabeth, and Cameron M Pierson. “Artificial Intelligence and Robots in Libraries: Opportunities in LIS Curriculum for Preparing the Librarians of Tomorrow.” *Journal of the Australian Library and Information Association* 71, no. 3 (2022): 256–74. <https://doi.org/10.1080/24750158.2022.2081111>. (Tait and Pierson 2022)

## Chapter 13: Librarians as Educators

### 13.1 Information Literacy (for Consumers of Information)

Information literacy in the age of AI is a new beast. As always, there is the helping patrons and users navigate the interface between information and their good selves. But there is a lot more nowadays, now we have machine learning, large language models, and algorithmic pipelines that might use biased data and might produce biased results.

### 13.2 Artificial Intelligence Literacy

What does 'AI literacy' entail? Alongside basic digital literacy and [Information Technology] skills, 'AI literacy' usually begins with an **elementary understanding of how Artificial Intelligence and Machine Learning work**, what they can and cannot do (IFLA 2020, 10 their emphasis).

Michael Ridley and Danica Pawlick-Potts write:

Navigating the effects of AI as well as utilizing it in a responsible way requires a level of awareness, understanding, and skill that is not provided by current digital literacy or information literacy regimes (Ridley and Pawlick-Potts 2021a).

Others express similar views. For example, Charlie Harper writes:

The privacy (or really lack thereof) and ethics of data collection and dissemination should become an integral part of information literacy services. Now that images, video, and audio can be faked

in staggering ways, the reality of source origin is becoming increasingly messy, too. Facilitating and promoting critical thinking and awareness within the community is a must (Harper 2018).

Librarians have the opportunity to help the population with this new kind of information literacy. There are some new labels that are appearing here, for example, 'AI literacy', 'Digital literacy', 'Computer literacy', 'Machine Learning literacy' and even 'Algorithmic literacy' (Ridley and Pawlick-Potts 2021a; Digital2030 2022; Druga et al. 2019a; Carlson and Johnston 2015). We will come back to these names shortly. These literacies have two classes of potential students: library staff, and the patrons of libraries (i.e. the public at large, or students and researchers in colleges and universities). Then we may make a distinction between 'internal' (which here means 'intellectual') and 'external' (which here means 'social', understanding 'social' in a wide sense). So, for example, a person has AI literacy at an internal level if they understand many of the kinds of AI software programs and possibly could even write them. Roughly, this would be college level knowledge (or 'competencies') that include some AI or computer science courses. AI literacy at an external level requires knowledge of the settings or societal contexts in which AI is used, for good or for bad, and a reasonable grasp of how AI affects their own lives and that of societies as a whole. Even children can be AI literate, at an external level.

Back to the labels. We favor 'AI literacy'. ML is where the action is, but 'Machine Learning literacy' is a mouthful. Neither 'Digital' nor 'Computer literacies' are entirely accurate. 'Algorithmic literacy' is unfortunate. It smears the word 'algorithm' into a meaning it does not have, and then

exhorts us to be literate about a topic that it does not itself define correctly. Nevertheless, no doubt 'Algorithmic literacy' will gain ascendance in the marketplace for labels. At which point we will go with the flow. Anyway, it is the *external* versions of the literacies that librarians need to learn themselves and teach to others.

There is a literature on teaching machine learning. For example, the delightfully titled 'Can You Teach Me To Machine Learn?' (Sulmont, Patitsas, and Cooperstock 2019). See also (Nori et al. [2019] 2023; Druga et al. 2019a). The literature, though, is sparse, and the domain unexplored. It is not really known how to do it well. For librarians and libraries, probably hands on workshops, tutorials, and coding classes would be a good idea.

We have dealt with many of the components of AI literacy elsewhere, under different headings. But just to list many of the topics

- Algorithms and how they work
- Having a critical understanding of AI tools (for example, the ones mentioned in Chapter 5) and the information that they may provide
- Bias
- Privacy. Teaching people to understand how to protect their privacy in the context of machine learning is important. Also important is understanding the privacy policies of ML companies and ML applications.
- Facial recognition technology
- Research guidance Librarians could guide researchers in using machine learning tools to analyze their data. This could include

providing advice on the appropriate algorithms to use, helping to interpret the results, and ensuring that the research is conducted ethically.

- Social epistemology

### **13.3 Data Information Literacy (for Producers of Information)**

There is the topic of 'Data Information Literacy' (Data Information Literacy Project 2023; Carlson and Johnston 2015). Jake Carlson and Lisa Johnston ask:

... what data management and curation skills are needed by future scientists to fulfill their professional responsibilities and take advantage of collaborative research opportunities in e-science and technology-driven research environments? .... how can academic librarians apply their expertise in information retrieval, organization, dissemination, and preservation to teaching these competencies to students? (Carlson and Johnston 2015, 2)

Roughly, researchers need to curate their data in such a way that the data is available to other researchers and society at large. Also, for example, funding sources such as the National Science Foundation require a Data Management Plan for research that they support. This is an ideal area for cooperation between librarians and researchers. This time the librarians are helping the researchers as producers of information, not as users of information. Only a part of this general area involves, or might involve, machine learning. But some of it may well do. Lisa Johnston and Jon Jeffryes discuss a case where civil engineering students put sensors on

bridges to learn of the integrity of the bridges (Johnston and Jeffryes 2015). That challenge may well involve machine learning. As noted earlier, machine learning data, say for supervised learning, needs to be of a certain kind and style.

### **13.4 Changes in Learning and Teaching**

Teaching and learning are changing in colleges and universities, particularly in areas like law and business. If curricula change, then there will need to be changes in libraries. Likely learning will become more personalized. If so, there will be learning data and analytics both of individual students, of instructors, and of groups and classes. Library use data will be part of this. There are initiatives on the use of AI in Teaching and Learning (see, for example, (Office of Educational Technology 2023))

### **13.5 Scholarly Communication**

Jason Priem argues in his paper 'Beyond the paper' that:

The journal and article are being superseded by algorithms that filter, rate and disseminate scholarship as it happens. (Priem 2013)

This is an interesting and important paper. Priem, writing in 2013, locates many of the changes and possibilities for change in the Web and what the Web enables (for example, share early share often, filter, and crowdsource review). We could not agree more with the general thesis, but now we

would locate the causal factors more with machine learning and LLMs. AI has the potential to remove peer review and possibly even many traditional journals.

## **13.6 Academic Libraries Collaborating with other University Units**

Academic libraries do this already. But ML is going to bring about big changes in teaching and research.

The International Federation of Library Associations and Institutions (IFLA)'s 2020 Statement on Libraries and Artificial Intelligence (IFLA 2020, 14) mentions as example collaborations Stanford University Library AI Studio, the University of Rhode Island's AI lab in the University Library, and the University of Cincinnati Libraries' Digital Scholarship Center (see also (McKenzie 2018)).

A surprise is how few collaborations there seem to be. Most colleges or Universities have a Data Science department or school. All Universities have their libraries. Yet the visible collaborations can be counted on the fingers of one hand.

## **13.7 AI Laboratories in the Library**

In 2017-8 the University of Rhode Island placed an AI lab in its library:

The library, as an interdisciplinary space that values inclusivity, is the ideal place for people of all backgrounds to learn about AI [. . .] Unlike a typical AI lab focused on research, the URI AI Lab will offer students and instructors the chance to learn new computing skills, and also encourage them to deepen their understanding of AI and how it might affect their lives, through a series of talks and workshops. The lab will offer beginner- to advanced-level tutorials in areas such as robotics, natural language processing, smart cities, smart homes, the internet of things, and big data (McKenzie 2018).

Putting the lab in the University's library is strategic. Organizers hope that students majoring in different fields, from philosophy and ethics to computer science and biomedical engineering, will visit the lab and use it to brainstorm about important social and ethical issues today and create cutting-edge projects (Rhody Today 2017).

The experiences of the AI lab, up until 2022, are described in (Dekker, Ferria, and Mandal 2022). It seems that there has been such a demand for the AI lab, and for Data Science as a whole campus wide, that the University is considering its options.

### **13.8 Automated Decision-Making**

The European Union (EU) has a Generalized Data Protection Regulation (GDPR), (summarized in (Wolford 2018)). It is a law which applies to and protects EU citizens worldwide. It came into effect in 2018. Of particular interest to us is Article 22, which in part reads:

**Art. 22 GDPR**  
**Automated individual decision-making, including**  
**profiling**

1. The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.
2. Paragraph 1 shall not apply if the decision <further text omitted here> (GDPR 2018)

'Profiling', in GDPR, means:

... “any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person”

Thus profiling should be construed as a subset of processing, under two conditions: the processing is automated, and the processing is for the purposes of evaluation (Goodman and Flaxman 2017).

Article 22, in the context of surrounding text and definitions:

... restrict[s] automated individual decision-making (that is, algorithms that make decisions based on user-level predictors) which “significantly affect” users. The law ... also effectively create[s] a “right to explanation,” whereby a user can ask for an explanation of an algorithmic decision that was made about them (Goodman and Flaxman 2017, 1).

European law is European law, of course. But, nevertheless, GDPR is the result of deep consideration and analysis. It highlights the need for caution with use of personal data to produce decisions that affect people and the need for explanations of individual decisions that are made.

Automated Decision-Making (ADM), using data about people as input and producing decisions that affect people as output, has been around for a very long time. For example... The Constitution of the United States requires

that there be a census every 10 years. The results of this census determine how many seats each State has in the House of Representatives (among other things). Processing data on this scale used to be a nightmare. Processing the 1880 census took eight years (Roberts 2019, 100) In 1889, Herman Hollerith received a patent for the 'Art of compiling statistics'. It was a patent for the punched card.

A hole is thus punched corresponding to person, then a hole according as person is a male or female, another recording whether native or foreign born, another either white or colored, &c. (Hollerith 1889)

The data about each individual person was entered on a single card as punched holes. But then the data could be aggregated by running through the cards and using an electro-mechanical device to detect the relevant holes. The data also could be 'combined' across categories. For example, were you to have an interest in white female carpenters it was essentially trivial in principle to determine how many there were (in a region, or a State, or, indeed, in the United States). Hollerith cards were taken up by wider businesses, and commercial concerns, and the processing techniques led directly to the formation of IBM (and the standard IBM 80 column punched card). Hollerith, and his card, won the 1889 Bureau of the Census competition to mechanize the process of conducting the Census. The 1890 Census was automated.

Thus, there has been ADM for a while, maybe for a hundred and thirty years. You can see the strengths and weaknesses of this original automation of the Census. It makes viable a very difficult practical problem. There

should be few or no errors in the actual processing of the stacks of punched cards. Additionally, the actual processing is completely transparent and, in principle, could be inspected by anyone. But the data needs to be sound. If the data entry for an individual punches a hole for 'farmer' when the individual is actually a 'carpenter', garbage is part of the input and the results of that garbage-in potentially can appear in the output, as garbage-out. Even worse if groups of people are not even included, or properly represented in the census, there certainly could be bias or unfairness in what comes out.

Nowadays, ADM has spread far and wide. Not all such decisions involve data about people as input, not all directly affect people as output. Not all involve machine learning. In fact, probably machine learning forms a relatively small, but increasing, part of ADM. After all, powerful machine learning has been available only for maybe 5-10 years.

Here are some examples where machine learning may place a role in ADM:

- Autopilots on airplanes (and similar technologies on ships, trains, and motor vehicles)
- Many decisions in the financial realm (such as investing, the identification of suspect financial transactions, credit card, or mortgage eligibility (and what interest rates those eligible are entitled to).)
- Medicine (identification of possible positive indicators of skin or breast cancer)
- The behaviors of industrial robots

- Military (robotics on the battlefield including unmanned systems and vehicles of various kinds)
- Providing evidential input to sentencing decisions in criminal trials
- Various decisions involving employment (e.g. interviews, promotions, remedial interventions)
- Making suitable chess moves to be a companionable opponent or to be a tutor of the appropriate level
- And many many more

Some of these types of decision making involve partnerships with humans. Humans may have overrides, may make the final choice or decision, etc.

The strengths and weaknesses here are similar to those of the original Census automation. But there are differences. Many of the systems, or their intellectual aspects, are owned by private companies, sometimes very large companies. A result of this is that often the systems are proprietary and are 'trade secrets'. With this, transparency can disappear. Outsiders may have no way of being assured that the actual processing is sound. When it comes to machine learning, and large language models, for example, even the companies and their researchers can be in a position of not really knowing how their complex systems work (if, indeed, they do work). In this case, transparency is not being hidden by companies, it is simply not available. A factor here is that large language models, for example, are so expensive to establish that only the largest companies or the government have the resources to create them and do research with them. Universities do not have the money to become involved directly. In many fields, one thinks of

universities as being able to be qualified and intelligent honest outside agents. But this is not true for advanced machine learning. The universities cannot participate at all at a core level. Then there is an important difference regarding the data. In immigration controls in airports, and similar, the government uses machine-learning face recognition technology to scan the incoming crowds for 'persons of interest'. The government does not seek anyone's approval or consent to do this. This is similar to gathering the original Census data back in 1890. There is a social contract between members of the community or society that gives implicit consent from members of that society to their government to carry out various governmental tasks and procedures. But the situation with private companies is different. If you walk into a shopping mall and the mall scans your face, without your knowledge or informed consent, how is this right or ethically appropriate? In sum, attention needs to be paid to transparency and the uses of data. There is a lack of insight, a lack of transparency, and often a lack of informed consent. Putting this right perhaps falls to many parties: to the creators of the systems, to the makers of laws, to educators, and to the citizenry at large. Several countries, or groups of countries, do have laws. For example, as we have seen, there is the European GDPR. The USA does not have a relevant all-encompassing laws. Librarians certainly can play a role (Ridley and Pawlick-Potts 2021a). Assume so, what can they do?

The librarians can become knowledgeable themselves. The Masters of Library and Information Science (MLIS) degree, which is the 'union card' for advancement in professional librarianship, might include a courses in the wider aspects of machine learning. (None do, 2023, in the USA.)

Libraries, as institutions, can offer opportunities to their staff to increase their knowledge in this area. This might include courses, workshops, hands on sessions. Then the librarians might be alert to inform their patrons (about surveillance, consent, privacy, automatic decision making, etc.).

### **13.9 Explainable Artificial Intelligence (XAI)**

There is a research field, Explainable Artificial Intelligence (XAI), that has direct relevance to automated decision making. Exposure to the ideas of Explainable Artificial Intelligence (XAI) is of benefit to folk trying to educate others about machine learning.

Explainable AI is Machine Learning that has the property of being easily understood by humans (Wikipedia 2023e) .

What ML systems can tell us is that certain observable *features*, feature-data, are connected with other observable *labels*, label-data— for example, that being written by J.K. Rowling is correlated with being a popular book. Sometimes an ML system will tell us the labels, the predictions, without being able to tell us of the specific features *that it used in its inference or calculation*. In these cases, the ML system is a black box. You give it a book, the features of a book, and it will tell you if the book will be popular. It is an Oracle. (Henceforward, here, as an abbreviation, 'Oracle MLs'.)

At this point we will take a brief detour into two areas of philosophy: the philosophy of knowledge, epistemology, and the philosophy of understanding and explanation.

We want to distinguish knowledge from right opinion. This is a distinction from Plato's Meno, amplified by John Stuart Mill's work on freedom of speech (Plato 380AD; Mill 1869). (Nowadays, 'right opinion' would probably be more usually described as being 'true belief'.) Plato defined knowledge as having three components:

Knowledge needs to be true.  
It needs to be believed.  
It needs to be justified.

An Oracle ML might provide right opinion, but it does not give knowledge. For knowledge, the knower has to have evidence and be able to provide a reasoned defense, or explanation, or justification of the known. In brief, the knower has to be able to give the reasons why. Socrates argued:

...true opinions: while they abide with us they are beautiful and fruitful, but they run away out of the human soul, and do not remain long, and therefore they are not of much value until they are fastened by the tie of the cause...when they are bound... they have the nature of knowledge; and ... they are abiding. And this is why knowledge is more honourable and excellent than true opinion, because fastened by a chain (Plato 380AD, 98).

An Oracle does not give knowledge. It does not give causes.

Then, what, in this context, is understanding?

A simple but plausible answer given by contemporary philosophers of science is as follows: to understand a phenomenon is to grasp how the phenomenon is caused (Strevens 2013).

The phenomena of interest here is how the successful ML systems produce the correct predictions that they do. To understand, we need to know the causes at work. With Oracles, the causes are exactly what we do not have. Hence, to address these considerations, there is the existence of the research field Explainable AI (XAI), with its own techniques and methods. Space does not permit extended discussion here of XAI.

There are actually two causal problems with Oracles. There is the problem of specific predictions, understanding why a specific prediction has been made in a particular case, and there is the problem of understanding how a given Oracle works as a whole inside its black box. An example of the first problem is when a radiologist works in partnership with an ML system to assess a patient and to diagnose possible cancer. The radiologist (and the patient) need to know the basis of, or evidence for, the ML system's prediction. The reason for that need is that knowledge is better than right opinion. Knowledge is better for trust, ethics, respect for human beings, decision making, regulatory and legal requirements, etc. An example of the second problem is that the designers or programmers of the radiology image diagnosis system should understand how it works. That would help them going forward to assess and improve the system (Strevens 2013).

Elsewhere, we do negotiate our lives using correlations (right opinions). The farmer puts fertilizer on crops, we take aspirin for headaches, and a majority of us avoid smoking for health reasons. Some of these connections are undoubtedly causal. But we usually do not know of the details of the causality. Often, this is the best we can do. The phenomena, the laws, if

there are any, the initial conditions, and the interactions of these, are all intertwined and complex. Often, they are also probabilistic. Basically, the systems are close to being black boxes. If XAI can do as well as agriculture and medical science, that might be good enough.

### **13.10 Annotated Readings for Chapter 12**

Druga, Stefania, Sarah T. Vu, Eesh Likhith, and Tammy Qiu. “Inclusive AI Literacy for Kids around the World.” In *Proceedings of FabLearn 2019*, 104–11. FL2019. New York, NY, USA: Association for Computing Machinery, 2019. <https://doi.org/10.1145/3311890.3311904>. (Druga et al. 2019b)

Gesser, Avi, Robert Maddox, Anna Gressel, Mengyi Xu, Samuel J. Allaman, and Andres S. Gutierrez. “New Automated Decision-Making Laws: Four Tips for Compliance.” *Debevoise Data Blog* (blog), 2022. <https://www.debevoisedatablog.com/2022/06/25/new-automated-decision-making-laws-four-tips-for-compliance/>. (Gesser et al. 2022)

Goodman, Bryce, and Seth Flaxman. “European Union Regulations on Algorithmic Decision-Making and a ‘Right to Explanation.’” *AI Magazine* 38, no. 3 (2017): 50–57. <https://doi.org/10.1609/aimag.v38i3.2741>. (Goodman and Flaxman 2017)

Mittelstadt, Matthew G. “Artificial Intelligence: An Introduction for Policymakers | Mercatus Center,” 2023. <https://www.mercatus.org/research/research-papers/artificial-intelligence-intro-for-policymakers>. (Mittelstadt 2023)

Nori, Harsha, Samuel Jenkins, Paul Koch, and Rich Caruana. “InterpretML.” C++. 2019. Reprint, *InterpretML*, 2023. <https://github.com/interpretml/interpret>. (Nori et al. [2019] 2023)

Pritchard, Duncan, John Turri, and J. Adam Carter. “The Value of Knowledge.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman, Fall 2022. Metaphysics Research Lab, Stanford University, 2022. (Pritchard, Turri, and Carter 2022) <https://plato.stanford.edu/archives/fall2022/entries/knowledge-value/>. This is a valuable resource on considerations of knowledge versus right opinion.

Ridley, Michael, and Danica Pawlick-Potts. “Algorithmic Literacy and the Role for Libraries.” *Information Technology and Libraries* 40, no. 2 (2021). <https://doi.org/10.6017/ital.v40i2.12963>. (Ridley and Pawlick-Potts 2021b)

Stanford HAI. “Generative AI: Perspectives from Stanford HAI,” 2023.  
[https://hai.stanford.edu/sites/default/files/2023-03/Generative\\_AI\\_HAI\\_Perspectives.pdf](https://hai.stanford.edu/sites/default/files/2023-03/Generative_AI_HAI_Perspectives.pdf). (Stanford HAI 2023b) These are perspectives from Stanford leaders in medicine, science, engineering, humanities, and the social sciences on how generative AI might affect their fields and our world.

Sulmont, Elisabeth, Elizabeth Patitsas, and Jeremy R. Cooperstock. “Can You Teach Me To Machine Learn?” In Proceedings of the 50th ACM Technical Symposium on Computer Science Education, 948–54. SIGCSE ’19. New York, NY, USA: Association for Computing Machinery, 2019.  
<https://doi.org/10.1145/3287324.3287392>. (Sulmont, Patitsas, and Cooperstock 2019)

# Chapter 14: Librarians as Managers

## 14.1 Coming on Board

Libraries need management, and many librarians have management duties. Of course, they are usually assisted in these duties by computers and general library automation. The use of AI or ML in this setting would be a step further. Amanda Wheatley and Sandy Hervieux's have published an environmental scan of the use of AI in academic libraries (Wheatley and Hervieux 2019).

Their appraisal is bleak. Here are some quotations from the Wheatley and Hervieux paper, offered without comment or judgement:

What is perhaps the library's best kept secret has been its slow uptake on automation and digital technologies.

[On the possibility of libraries being run by machines] This unsettling reality should have awoken librarians, instead the profession reacted as it did to most technological revolutions - it waited. In fact, it is still waiting.

The absence of scholarly research on AI-related technologies in libraries is not to be unexpected. Libraries have suffered from issues on the adoption of digital technologies and a general resistance to change throughout the twentieth and twenty-first centuries.

The progression of industrial and office automation paved the way for libraries to adopt similar technology, yet this adoption was always years behind the current trends.

The current state of artificial intelligence in academic libraries has proven to be nearly non-existent.

Separately from this environmental scan, there are literature reviews that range wider than academic libraries. Rajesh Das and Mohammad Islam, in their systematic literature review of the application of AI and ML in libraries, identify in the publications the following ten themes (R. K. Das and Islam 2021):

- collection building and management,
- processing in libraries,
- circulation and user studies,
- reference service,
- library administration,
- library customization and retrieval,
- research and scholarship,
- service quality and innovation,
- intelligent agents for information search and retrieval,
- study on implementation and existing technologies and solution

Their review is backward looking, of necessity. It is considering what has been done (prior to 2020) and what AI and ML techniques were used to do it.

There is also Andrew Cox, Stephen Pinfield, and Sophie Rutter's *The intelligent library: Thought leaders' views on the likely impact of artificial intelligence on academic libraries* (Cox, Pinfield, and Rutter 2019). This

does have valuable content— its editorial summary is excellent. But it is old, and it is not clear how well informed the 33 'thought leaders' were around 2018. One of them did not know what Artificial Intelligence was, and none of them knew about large language models (nor did the rest of us).

IFLA is The International Federation of Library Associations and Institutions. It styles itself as 'the Global Voice of the Library and Information Profession'. Its 2022 Trend Report Update has no mention of artificial intelligence or machine learning (Al Badi et al. 2023). IFLA does have its earlier Statement on Libraries and Artificial Intelligence, which is very good (IFLA 2020).

Barbara Wood and David Evans write in their paper *Librarians' Perceptions of Artificial Intelligence and Its Potential Impact on the Profession*:

The results of our survey point to an overwhelming sense of complacency among librarians in regard to the transformative/disruptive effects of this technology. For the past 35 years, academic libraries have successfully embraced computerization. Why is it that, at this time, we have our heads in the sand? We liken it to the climate change debate—the data is there, but we choose to ignore it (Wood and Evans 2018).

For us, it is worthwhile for us to consider aspects of these themes which we have not considered elsewhere. In Chapter 10 the following topics were mentioned

- Workflow and Improving Service

- Optimize the Use of Space (and, Indeed, Other Resources)
- Robots
- Mimicking Librarian Experts' Behaviors

We are interested specifically in machine learning and artificial intelligence (not in plain automation). We are using the (R. K. Das and Islam 2021) as the skeleton for this.

## 14.2 Data and Analyses

There will be data on resources, processes, user behavior, and more. From these, there can be

- **Predictive Analytics** This identifies past usage patterns and trends as input to predict demand and need. It could be looking at many types of data as input (demographics, time of day, time of year (e.g. exams, summer breaks)). With machine learning it does not really matter how rich or apparently irrelevant the types of data. After training, anything known that is not connected with the correlations will just be given weight zero (i.e. discounted).
- **User Behavior Analytics.** This is data on what the patrons are doing, either individually or collectively. Attention needs to be given here to privacy or informed consent. Anonymous data, for example on which books are frequently checked out, can help with decisions on which books to keep physically in the stacks and which to store elsewhere, perhaps off-site. In contrast, recommendation systems will

work better with information tied to patrons individually (that still can be kept anonymous).

- **Learning Analytics.** In cases where libraries are serving students, faculty, and research, there can be data on what resources the patrons use and how they use them for the purposes of teaching and learning. In turn, this can be used to improve the process of education. There is a large and burgeoning field Artificial Intelligence in Education (AIED). See, for example, (Stanford HAI 2023a.) Librarianship will have a role in this. That role will extend to School Libraries and Public Libraries.

### **14.3 Evidence-Based Librarianship**

When making decisions, evidence-based practices encourage the use of the best available evidence, in conjunction with expertise and attention to values (Wikipedia 2023d). This approach tends to downplay tradition, especially when no-one knows why the tradition is as it is. There is Evidence-Based Library and Information Practice (EBLIP), with supporting books, research articles, and a journal (for example, (Hjørland 2011; University of Alberta Library 2023). Maybe now is its time. There is now plenty of data, and ML can analyze a quantity of it in a way that no human can. Of course, the management and administration of libraries used any amount of data prior to the advent of EBLIP. But now there is much more data and the existence of ML that can interact with that data intelligently. As mentioned many times, ML can deal with volume in a way that collective human resources cannot.

## **14.4 Data-Driven Decision Making**

### ***14.4.1 Collection Building and Management***

ML can analyze the collection and its uses. Predictive Analytics can be used as part input on decisions on which materials to purchase, keep in the collection, or weed or discard. It can help with resource allocation. The analysis would take input from availability of materials in other libraries, or the needs of the library's users. (See (Litsey and Mauldin 2018).)

### ***14.4.2 Circulation and User Studies***

Service, recommendation, personalization, and AIED all have relevance here. ML together with the data analytics mentioned earlier have the potential to effect improvements.

### ***14.4.3 Processing in Libraries***

AI can help with many of the processes that are part of librarianship. Most of these have been mentioned earlier in this text, including cataloguing, classification, acquisition, archiving, digitization, transcription, translation, indexing, and summarization.

#### ***14.4.4 Research and Scholarship***

The view here is that shortly the information discovery techniques and tools are going to be far superior to the ones we have today. We have amplified on this thesis throughout this text. What managers should do is to support the use of these tools.

#### ***14.4.5 Service Quality***

There will be improved metrics on most aspects of service. Being able to measure how good service is does not of itself guarantee good service. But whenever insights or innovations come to the fore, it is good to know whether they actually do make service better.

### **14.5 Acquiring the Appropriate AI Tools**

The managers can acquire the relevant AI tools.

IFLA makes a suggestion that would kill another bird with that stone:

Libraries can also support ethical AI research and development by their procurement choices: purchasing AI technologies which abide by ethical standards of privacy and inclusivity. This would both reaffirm the trust of users in libraries, and send a message to the AI research field by increasing the demand for ethical AI technologies (IFLA 2020, 2).

## **14.6 Analysts and Staff**

There is a shortage of skills and awareness.

The demands of responding to such changes may reveal a significant skills gap in the sector. We know there is already strong demand in the economy for data scientists for data analysis and visualisation. Perhaps some librarians will be required to develop these skills, or at least awareness of different techniques and how they need to be supported. These demands are a challenge because they prompt librarians to learn more about IT and quantitative data analysis, including statistics. In a relatively low-paid sector we may be unlikely to attract people with stronger STM backgrounds to the profession, while those in the profession are typically from an arts background. More optimistically, we can say there will also be a need for librarians as data curators to take on new data (Cox, Pinfield, and Rutter 2019, 17)

Foundation Models have given us all a break here. You do not have to be a programmer or a mathematician to use, develop, or configure LLMs. You can even be from an 'arts background'. So-called 'Prompt Engineering' makes many things possible— we will look further into that in Appendix B.

## **14.7 Fear of AI**

Within the library profession there does seem to be a fear of AI (as evidence, see, for example, (Cox, Pinfield, and Rutter 2019; ExLibris 2019)).

ExLibris invites us librarians to be Liberty Leading the People (Wikipedia 2023h):

... embracing AI not as users, but as active players to fight the risks of bias, misuse, and discrimination. If libraries take an active role in the implementation of artificial intelligence applications in the information management landscape, then they can help programmers find the best data for their algorithms. Once they assume the leading role, librarians can be co-creators of "an intelligent information system that respects the sources, engages critical inquiry, fosters imagination, and supports human learning and knowledge creation," according to Catherine Nicole Coleman. AI solutions can also facilitate both more process transparency and greater data control, with libraries able to safeguard their most important principles and maintain trust, neutrality, freedom of expression, mindful media consumption, and equal access to information, while promoting digital inclusion and data privacy

"Advocacy should not be directed at maintaining traditional librarianship, but in influencing the development of the emerging information systems."

(ExLibris 2019)

## **14.8 Annotated Readings for Chapter 14**

Gale, part of Cengage Group. "Library Marketing: Improve Outreach with Gale Analytics," 2023. <https://www.gale.com/databases/gale-analytics>. (Gale, part of Cengage Group 2023)

Litsey, Ryan, and Weston Mauldin. "Knowing What the Patron Wants: Using Predictive Analytics to Transform Library Decision Making." *The Journal of Academic Librarianship* 44, no. 1 (2018): 140–44. <https://doi.org/10.1016/j.acalib.2017.09.004>. (Litsey and Mauldin 2018) This is old now (before LLMs). One commentator describes that it '... seems like a dystopian vision that thinks it's a utopian vision'.

## **Chapter 15: Librarians as Astronauts**

### **15.1 Astronaut Training**

Well, you are in a good place. To do creative work with Foundation Models, you do not need to know anything about machine learning, or computer programming. You do not even have to have any domain knowledge about the area you are going to work in (which we are assuming here to be librarianship). The Foundation Model you use will have all the knowledge that is required. Your role is to ask it in ordinary natural language— to 'prompt' it— to do whatever you are looking for. Over the past twenty years or so many librarians have become absolute experts at using search engines, say Google Search. Now there are new kids on the block, Foundation Models, Large Language Models or Large Multimodal Models. It would not be a bad thing at all for many librarians to become absolute experts at LLMs or LMMs.

### **15.2 Why Should You Learn How To Do It?**

- Gives you understanding
- Allows you to teach others
- Allows you to produce apps.
- Allows you to work with ML professionals

## 15.3 What are the Real Creative Possibilities

This is hard to know (stating the obvious). Three areas where modern machine learning might have a distinction advantage over the incumbents in librarianship or in information curation and provision are: data visualization, chatbots, and information discovery including text data mining.

Data science— making sense of data— is important. Having the right representation of the data is often critical to understanding that data and seeing how to reason with it. The key point was made by Jill Larkin and Herbert Simon in their paper *Why a diagram is (sometimes) worth ten thousand words*. They suggest that informational equivalence should be distinguished from computational equivalence (Larkin and Simon 1987). The distinction can be explained by means of an illustration.

Donald Norman drew attention to a famous example of two games (Norman 1993). In the first:

you and your opponent take turns in choosing and selecting single, previously untaken, numbers in the range 1..9, and the winner of the game is the first person who has 3 numbers that sum to 15. [So you try both to get numbers that sum to 15 and also to prevent your opponent from getting such numbers.]

In the second:

you play Tic-Tac-Toe (or Noughts and Crosses),

Now, the first game we would find awkward and might lose from time to time (until we devised a suitable representational scheme). The second is trivial, and an attentive adult would never lose a game. But the two games are the same. They are different representations of the same game. If the numbers are laid out as though they were on a Tic-Tac-Toe board:

6 1 8

7 5 3

2 9 4

Then the problem of finding three numbers that sum to 15 is exactly that of finding three in a vertical, horizontal, or diagonal line.

Any information as to the board's state, or to the numbers selected, is exactly inter-translatable back and forth between Tic-Tac-Toe representation and numbers-selected representation. There is information equivalence. But humans can manipulate and compute with one representation much more easily than the other. (And, interestingly enough, a computer, or a computer program, would be more at home with the numeric representation as opposed to the two-dimensional board representation.)

Information opens up different possibilities for an agent depending on how it is represented. This is matter of the manipulations, operations, and computations, that the depiction facilitates. (A good collection of historical depictions of statistical data can be found in Michael Friendly's *Gallery of*

*Data Visualization* (Friendly 2007).) Machine learning may well be able to learn the best ways to represent data.

Let us sketch some territories here. In so far as this concerns diagrams, as it would for education and for research by humans, the area of interest is cognitive psychology and heuristics. In so far as it concerns computer processing, it would be a matter of data-structures and algorithms— parts of computer science. Librarians may be fine with either of these, but they are a little outside librarians' bailiwick.

Chatbots have been discussed extensively in this text. Certainly, chatbots in libraries represent an opportunity, possibly even a research opportunity. But chatbots for the purpose of service are going to be everywhere. It may be hard for librarian developers to keep themselves clear of the development stampede.

Information discovery and text data mining are both gifts to librarians. Librarians know from text. We will amplify shortly on one aspect of this possibility.

## **15.4 Sitting in Your Tin Can**

If you are going to sit in your tin can, you could think about how ML might provide new ways to interact with library resources. You could think about interactive games and simulations that might help with access or with AI literacy. You could think about many things.

## 15.5 Exploring World 3

### 15.5.1 Undiscovered Public Knowledge (UPK)

In 1986 Don Swanson wrote as the abstract to his paper *Undiscovered Public Knowledge*:

Knowledge can be public, yet undiscovered, if independently created fragments are logically related but never retrieved, brought together, and interpreted. Information retrieval, although essential for assembling such fragments, is always problematic. The search process, like a scientific theory, can be criticized and improved, but can never be verified as capable of retrieving all information relevant to a problem or theory. This essential incompleteness of search and retrieval therefore makes possible, and plausible, the existence of undiscovered public knowledge. Three examples intended to throw light on the logic of undiscovered knowledge are constructed and analyzed. The argument is developed within the framework of a Popperian or critical approach within science and on Popper's distinction between subjective and objective knowledge--the distinction between World 2 and World 3. (Swanson 1986)

The referenced work of Karl Popper can be explained thus. Popper had introduced in the idea of three worlds: World 1 (the physical world), World 2 (the mental world), and World 3 ('... the world of objective contents of thought, especially of scientific and poetic thoughts and of works of art.'). World 3— objective knowledge— contains the *contents* of books and libraries, not the physical books (World 1), nor the mental ideas conjured

up by the books (World 2), but the contents of the books understood as the objective assertions they make. (See *Epistemology Without a Knowing Subject* (Popper 1968).) World 3 objects can stand in logical relations one to another. One book may contain the assertion 'Leda is a swan'. Another book may contain the assertion 'All swans are white'. Those two books together, considered as a combined whole, entail the consequence that Leda is white. Swanson instantiated and generalized this in the following way. Suppose (some recorded research tells that) process A causes effect B and (some entirely different recorded research tells that) B causes effect C then A causes C, and this, perhaps hitherto unknown, relationship can be discovered by looking at the recorded research. This is an example of *Undiscovered Public Knowledge* (UPK). A real scientific example that Swanson offered from 1986 is that it was known, as recorded objective knowledge, that (A) fish oil reduces platelet clumping (B) and reduced platelet clumping is (C) of benefit to patients with Raynaud's disease. But it was *not* known as objective knowledge (i.e. published anywhere as recorded knowledge) that (A) fish oil can be (C) of benefit to patients with Raynaud's disease. That relationship can be discovered without entering a laboratory. It can be discovered without leaving a library.

[Usually] .... different assertions and findings need to be assembled across documents to create a new coherent assertion, much as different pieces of a puzzle are assembled to create a single picture. (Smalheiser 2017, 3)

We should keep in mind here that all knowledge is fallible and conjectural. So that when we appear to discover that fish oil helps with Raynaud's disease, we are discovering a new conjecture, that apparently no one had

proposed earlier, and that conjecture would need testing. If its components had been thoroughly tested and were reliable, it may be that further testing could be minimal. That would depend in part on what the practical consequences would be of trusting the hypothesis. In the case of medical treatments, and drugs, there are established protocols.

It may be that the apparent conjecture from UPK had been known and discovered earlier. Researchers looking for treatments for Raynaud's disease may have had hundreds of conjectures as to what might be suitable, and their problem might have been how to spend their time and money. In this case, the ABC example given above screams loudly: try fish oil. UPK is still valuable.

There are other kinds of UPK (Smalheiser 2017). There can be publications that no one currently reads. These may be publications in truly obscure journals, or in outlets that retrieval systems have trouble in finding (perhaps due to poor indexing or metadata).

Investigators in these areas often talk of one node, or two node, searches or of open, or closed, searches. If the search starts with process A and is trying to find anything it relates to, that would be a one node, or open, search. If the search starts with two processes A and C, and is trying to find a B that connects them, that would be a two node, or closed, search.

[Don Swanson was a librarian. More strictly, he rose to being Dean of the Graduate Library School of the University of Chicago.]

### ***15.5.2 Literature-Based Discovery (Text Based Informatics)***

The seminal work of Don Swanson has been generalized and developed to form *Literature-Based Discovery* (sometimes called *Text Based Informatics*). This is used widely in bioinformatics and medical informatics. See, for example, (Crichton et al. 2020; Heatley 2023; Moreau 2023; Sebastian, Siew, and Orimaye 2017; Thilakaratne, Falkner, and Atapattu 2020; Wikipedia 2023j).

A consideration of importance to us is that Literature-Based Discovery until about 2020 used what we now would regard as being primitive bibliometric tools. Computers were used, and Google-style searches. It used authors, titles, keywords, subject headings, etc. But 2023 machine learning, especially LLMs, open up new possibilities. No doubt, UPK researchers are on to LLMs already. But the target domain is vast— every book in every library. There are opportunities for all.

### ***15.5.3 A Message to Librarian Astronauts***

The growth of scientific knowledge is usually thought of in terms of wresting new discoveries from the physical world— World 1— admittedly a world that offers unlimited opportunity for discovery. But it should be of interest to librarians to notice that World 3 also qualifies as an endless frontier and to understand how and why this is so (Swanson 1986, 117)

## 15.6 Annotated Readings for Chapter 15

Smalheiser, Neil R. “Rediscovering Don Swanson: The Past, Present and Future of Literature-Based Discovery.” *Journal of Data and Information Science (Warsaw, Poland)* 2, no. 4 (2017): 43–64. <https://doi.org/10.1515/jdis-2017-0019>. (Smalheiser 2017)

For here  
Am I sitting in my tin can  
Far above the world  
Planet Earth is Blue  
And there is nothing I can do  
(David Bowie. *Space Oddity*)

# Appendix A: Some Theoretical Background to Librarianship

## A.1 Concepts, Classification, Taxonomies, and Items

A central component of classification is concepts, and a whole collection of concepts, being used at once, amount to a *taxonomy* or *conceptual scheme* or *classification scheme*. A concept has *intension* (meaning) and *extension* (which is the collection, or class, or set, of those items, or individuals, or things, in the world that fall under it, or are instances of it). Concepts will usually have *labels* (i.e. words, or names) that identify them. Some concepts can have two or more different labels— such labels would be synonyms e.g. 'attorney' and 'lawyer'. Some labels can label two different concepts— such labels would be homographs e.g. 'bank' (of river) and 'bank' (financial institution).

There is a general looseness of terminology in classification areas. The terms 'concept', 'class', 'set', 'category', and probably other terms, can get used interchangeably. We will go with the flow on this (for example, often using 'class' or 'set' for concept).

A taxonomy— unless it is a simple list or dictionary— will always have at least one class which is a subclass of another class. Then the overall structure of the scheme might be a 'hierarchy', a 'polyhierarchy', a 'directed acyclic graph', a 'tree', a 'forest of trees', or some other graph-theoretic structure (see (Frické 2012)).

## A.2 Controlled Vocabularies, and Thesauri

In many settings, librarianship will use *Controlled Vocabularies* (CVs). A controlled vocabulary is collection of words, or 'terms', where the words are 'controlled'— they have a fixed and definite form. As to the value of CVs, Elaine Svenonius writes

Perhaps as near as one can come to generalization about the value of a CV is simply to say where precision and recall are important retrieval objectives, then a CV of some kind is mandated. (Svenonius 2003, 837)

[There are more detailed explanations of CVs, and their value, in (Frické 2012; Harpring 2020; ANSI/NISO 2010; Zeng 2005). CVs are still of value in the age of computers and machine learning as they provide a standard way to describe concepts and their relationships, and a standard interface between different systems.]

The simplest CVs amount to little more than lists of terms. CVs used for indexing and metadata (e.g. subject classification) tend to be more sophisticated than this. They have a structure in the background, with links or connections between some terms. As an example, the Library of Congress Subjects Headings (LCSH) is a controlled vocabulary. LCSH has hundreds of thousands of terms in it, and it took a hundred years or more to build. It is probably the biggest and most elaborate CV ever made. Its size and maturity are not the only points of interest. It also relates broader

and narrower terms, and it suggests preferred terms in the case of synonyms.

Indexes typically use CVs. Each hierarchy in an index-CV is a 'tree'. But the index as a whole *is not a single tree*. There is not one single term at the top with every other term being a subterm of it. Rather there are several different trees. The structure is that of a *forest*. For example, the MESH subject headings are in 16 different trees. As an illustration, here are six of them

- Anatomy
- Organisms
- Diseases
- Chemicals and Drugs
- Analytical, Diagnostic and Therapeutic Techniques, and Equipment
- Psychiatry and Psychology

*Thesauri* are controlled vocabularies with relations provided between broader and narrower words or terms. However, some extant controlled vocabularies themselves have broader and narrow terms (e.g. the aforementioned LCSH). If this practice is acceptable, Thesauri can be much the same as Controlled Vocabularies.

A few words are in order about the nature of the links or relations between the broader and narrower terms. International standards for thesauri distinguish three possibilities here: subclass, instance, and part (ANSI/NISO 2010). Whales are a subclass of mammals, Moby Dick is an instance of a whale, a blowhole is part of a whale. So, an imaginary CV

might contain: 'mammal' narrower term 'whale', 'whale' narrower term 'Moby Dick, and 'whale' narrower term 'blowhole (whale)'. The first relation would be subclass (sometimes called 'generic'), the second instance, and the third partitive. The first two relations can come under the general approach of taxonomies. The third, addressing parts, is usually considered to be mereology (which is the study of parts and wholes).

We will usually assume that the collection of top-level terms in a thesaurus (those that are not narrower terms of other broader terms) could be printed, displayed, or written out in 'alphabetical' order. There is a small wrinkle here. Just what 'alphabetical' order amounts to is a question. To give an example, for a telephone directory, should 'Newark' come before 'New York, or after it? where should numbers go? where should foreign words go? where should abbreviations go? The practice here is known as the 'filing order', and it is a matter of several decisions or conventions (which we will not go into).

### **A.3 Ontologies and Ontological Vocabularies**

An ontology identifies what exists in a certain domain or area i.e. the individuals, the properties, the classes, etc. As examples, a medical setting might have an ontology with doctors, patients, medicines, bacteria, etc. in it. A soccer game might have players, teams, balls, referees, pitches, and the like. A library might have patrons, librarians, books, desks, reading rooms, and similar items. Computer scientists use ontologies as a guide to the requisite data-structures needed for their programs. Librarianship is not really interested in ontologies as such, but it is interested in the words or

terms that are used. An ontology will have terms or words associated with the items that appear. So, for example, in a soccer ontology, ‘pitch’, ‘field’, and ‘paddock’ are near synonyms for the surface on which the game is played. Such terms could be built into a controlled vocabulary, with ‘pitch’ being the preferred term, and ‘field’ and ‘paddock’ being lead-in terms. A thesaurus could also easily be used here (e.g. with ‘soccer complex’ having ‘pitch’ as a narrower term). Ontologies themselves, or what we might call ‘Ontological Vocabularies’, could also be used. For example, libraries lend books (or, perhaps, ‘libraries issue books’); we can make one of these forms a preferred form and record in the Ontological Vocabulary that libraries lend books.

Thus, Ontological Vocabularies, in so far as they provide relationships between words or terms, are a more general case than Thesauri. For example, consider, in a medical setting, the relationships ‘... is a symptom of ...’ , ‘... may be treated by ...’ ; a relevant Ontological Vocabulary might have preferred terms for ‘headache’, ‘fever’, ‘influenza’, and ‘aspirin’, and also such relationships as ‘headache may be treated by aspirin’.

If these possibilities are thought of in a graph-theoretic sense: a thesaurus links narrower terms to broader terms, and often the structure of this will amount to a tree i.e. there is a root, and parent nodes can have many children but each child can have only one parent. Some thesauri, for example the LCSH, permit some terms to have more than one broader term (i.e. a child node might have more than one parent node) in which case the structure would be a directed acyclic graph (DAG). In an *Ontological Vocabulary*, any node might possibly be linked to any other node; this

would amount to a general graph (i.e. a web of nodes and links). Roughly, then, an ontology is a generalization of a taxonomy, and an ontological vocabulary is a generalization of a thesaurus.

#### **A.4 Objective, Intersubjective, and Subjective**

These three terms are selectors among knowledge, or among candidates for knowledge. When an Art Historian describes a painting as being in the shape or proportions of a golden rectangle, she is picking out a property of the painting. She is making a perfectly objective ascription about the external world. She can be wrong, of course. There can be some mistake with her eyes or means of measurement. Nevertheless, right or wrong, she is aiming to be objective. Moving on, suppose her aesthetic is that of Neoclassicism, to the exclusion of all else. She may describe a painting by Jacques-Louis David as being a good painting. That would be an intersubjective judgement. There are plenty of other Neoclassicists and they would have the same view. But Neoclassicism itself is not part of the external world, it is an aspect of a certain kind of culture. Finally, there is the rest of us who know nothing about art but know what we like. Expressions of what we like would be subjective.

Taxonomies also can be objective, intersubjective, or subjective. Objective schemes are the province of science and mathematics. Classifying matter into atoms, elements, compounds, and mixtures, or classifying the numbers into integers, reals, or rational numbers, or classifying parts of the biological world into genus and species are all objective schemes. As noted earlier, our knowledge of the objective is fallible— scientists or

mathematicians can be mistaken (and hopefully they can correct the errors). Intersubjective schemes arise in the context of society, culture, politics, religion, myths, and the law. A good example is the **Nursing Interventions Classification (NIC)**

**[it]** is a comprehensive, research-based, standardized classification of interventions that nurses perform. It is useful for clinical documentation, communication of care across settings, integration of data across systems and settings, effectiveness research, productivity measurement, competency evaluation, reimbursement, and curricular design....

**The 565 interventions** in NIC (7th ed.) are grouped into thirty classes and seven domains for ease of use. The 7 domains are: Physiological: Basic, Physiological: Complex, Behavioral, Safety, Family, Health System, and Community. Each intervention has a unique number (code). The classification is continually updated with an ongoing process for feedback and review. In the back of the book, there are instructions for how users can submit suggestions for modifications to existing interventions or propose a new intervention (Butcher et al. 2018).

This is not a description of the external world. This is not intended to tell us how the world is built (the continual updating is a bit a giveaway on this). It is not objective (in our sense of objective). Rather it is a convenient fiction (or, as some might say, it is a social construction). However, nurses, doctors, insurance companies etc. all respect and use this fiction. It is intersubjective. It is artifact they all agree on and use. As to the final category of pure subjective classification— that, for example, is how you and I separately organize our clothes in our wardrobes.

Most classification schemes in information organization are intersubjective, but some have components that supervene on objective

classifications. For example, schemes for classification of resources on science or mathematics, e.g. books and research papers, would typically partially follow objective classifications that the scientists or mathematicians use.

## **A.5 Emotive and Descriptive Content**

Going back to the Art Historian describing a painting as being in the shape or proportions of a golden rectangle... Being golden rectangular as a property is not something that is either good or bad, and it is not something that she wants to recommend to you or to warn you away from. It is just a plain description. 'golden rectangular' has descriptive content and little or no emotive content. When a viewer describes a painting as being 'nice' or 'good', that viewer is perhaps expressing approval or recommendation. But it is unclear what in particular, if anything, in the painting, 'nice' or 'good' describes. 'Nice' or 'good' have emotive content and little or no fixed descriptive content. (Emotive content is sometimes called 'connotation'.)

Many adjectives (properties, nouns) have both some descriptive content and some emotive content. And it is possible for different words to have the same descriptive content but different emotive content. For example, there are Bertrand Russell's well-known 'emotive conjugations' [Comment, *Brains Trust* (BBC Radio) (26 Apr 1948)] such as

I am firm; you are obstinate; he is a pig-headed fool.  
I am righteously indignant; you are annoyed; he is making a fuss  
over nothing.

We made a tactical withdrawal; our allies retreated; the enemy was routed.  
I have reconsidered; you have changed your mind; he has gone back on his word.

Were the present author to be described as 'firm', or as 'obstinate', or, indeed, as 'a pig-headed fool', the descriptive or cognitive content of those ascriptions is exactly the same (we may suppose). That is the dispositions, and actual actions, of his that these concepts or labels pick out, if correctly used, are identical for all three labels. However, the labels are intended to either express or evoke different sentiments on the part of the speaker or the listener. Roughly, at one end, a speaker that uses 'firm' is mildly admiring of the trait and wishes the listener to be mildly admiring also; at the other end, well, we all know what labeling the author 'a pig-headed fool' is intended to achieve. These expressive or evocative functions are the emotive content. So, each of the concepts [firm, obstinate, pig-headed] has pretty well the same descriptive content, the same intensions, but they differ in emotive content. They are synonyms in all but emotional force.

Of course, both descriptive content and emotive content of particular labels or concepts can change through time and be different across cultures.

For various reasons, usually obvious ones, individuals, groups, and institutions often try to employ different labels or different classification schemes in order to manipulate the emotive-descriptive divide in a way they wish. For example, it may be that the label 'garbage collector' is a term of mild opprobrium, and all concerned might switch to 'sanitary engineer' as the concept or label of choice. Here, the intensions of 'garbage collector'

and 'sanitary engineer' are pretty well the same, but one has unwanted emotional association. [This move is usually a temporary palliative. Say folk do have bad feelings to, or opinions of, the profession of garbage collection— those bad attitudes do not attach to the name rather they attach to what those garbage collectors are and to what they do. We are all familiar with Romeo and Juliet's 'What's in a name? that which we call a rose By any other name would smell as sweet;'. Using the label 'sanitary engineer' might work for a while, but give it 50 years and a person may be able to insult another with the words 'sanitary engineer'.]

## **A.6 Classification Schemes and the Act of Classification**

There are two balls that we are juggling here. There are the classification schemes, and there is the act of classification. Consider, for example, the aforementioned Nursing Intervention Classification (Butcher et al. 2018). It is a perfectly fine classification, absolutely clear as to what it is and what it is supposed to do. Then there is the act of classification i.e. deciding what interventions a nurse, or healthcare provider have done on a particular occasion. As an example, the entry for activities relating to Hypertension includes 50-100 activities. Here are a brief few of those:

- Instruct at-risk patients to have regular preventative health screenings, including electrocardiogram, echocardiogram, electrolytes, urinalysis, as indicated
- Instruct related to healthy dietary pattern
- Instruct related to proper physical activity (e.g., exercise 30 to 45 minutes a day)
- Instruct related to contributing lifestyle habits that should be avoided (e.g., use of tobacco in any form and alcohol)

- Instruct the patient on lifestyle modification related to sleep and rest patterns (e.g., 8 hours per night is recommended)
- Provide information on possible changes in lifestyle necessary to avoid future complications and control the disease process
- Provide information related to the purpose and benefit of the lifestyle changes
- Instruct related to self-blood pressure monitoring and to report abnormal findings
- Instruct the patient on possible causes of hypertension
- Instruct the patient and family to take an active role in the management of disease process, (e.g., medication indications and administration, maintaining proper diet, exercise and healthy habits, quitting smoking, reducing stress, reducing weight, reducing sodium intake, reducing alcohol consumption, increasing exercise, as indicated)
- Instruct the patient and family on medication usage and indications
- Encourage the patient and family to maintain a list of current medications and reconcile routinely at wellness checks, hospital visits, or hospital admissions
- Instruct the patient to recognize and avoid situations that can cause increased BP (e.g., stress or sudden discontinuation of drug treatment)

(Butcher et al. 2018) p.213-214

Suppose a nurse interacts with a hypertension patient and there is a serious back-and-forth medical conversation. A question is: which of these activities has the nurse done? No doubt the nurse himself, and the local general medical environment know reasonably well. The point being made here is that the act of classification, deciding which items belong in which classes, is fallible, and it is subject to mistakes and possible even to deliberate misuse or malfeasance.

## **A.7 Annotated Readings for Appendix A**

Frické, Martin. *Logic and the Organization of Information*. New York: Springer, 2012. (Frické 2012). This has material on organizational structures and graph theory, and further references that are relevant.

# Appendix B: Working With LLMs

## B.1 Introduction

Do yourself a favor. Learn as much as you can about LLMs. You can go a long way merely by getting access and by typing. You can go further by having the computer programming language Python on your computer and using an LLM development environment like LangChain.

Many of the major LLM implementations are readily available. As examples:

- Bard (using LaMDA) is available from <https://bard.google.com>. It is free. Bard implementations will usually have a microphone icon and can take spoken, or dictated, input. [That will make interactions quicker.]
- GPT-4 and ChatGPT are available from <https://openai.com>. Some modes of access are free. Others can be paid for either by usage or by monthly subscription. Current subscription to GPT-Plus is \$20 a month. With a subscription, the page <https://chat.openai.com/> access to GPT-3.5 and GPT-4, and to plugins and web browsing enhancements. There are about 120 plugins available at the moment (with many more to come, no doubt). You are allowed to have 3 plugins enabled at one time. The author has Wolfram (which gives mathematics and more reliable facts and factual statistics), ScholarAI

(which helps with scientific papers), and Prompt Perfect. Separately the author has an API (Application Programming Interface) key to GPT APIs. This is for programs using a GPT LLM and is charged by the access (usually just cents for a day's work). More on the programming later.

- ChatGPT is also available as a free app for the iPhone (and there will be a free app for Android phones).
- Bing, a search engine (using GPT-4) is available on a web-browser, as an app for smartphones, as part of a chatbot voice assistant on Amazon Echo, and Google Home. New Bing is available from <https://www.bing.com/new>

We are going to use Application Programming Interfaces (API) for interacting with the LLMs. A valuable resource, with many links to materials is

OpenAI Cookbook (OpenAI [2022] 2023)

## **B.2 Prompts and Prompt Engineering**

An LLM takes some text (or images, or sounds) as an input prompt and returns text (or images, or sounds) as its response. For our purposes, we will generally assume that prompts are given as text in English. Crafting a prompt to obtain the right or desired kind of response is *Prompt Engineering*. Prompt engineering is an active research area, with hundreds of publications. There are also many excellent educational sources for learning prompt engineering. One example is *The Prompt Engineering*

*Guide* (Saravia 2023) created by Elvis Saravia of Dair.ai (Democratizing Artificial Intelligence Research, Education, and Technologies). That work is clear and comprehensive. We will sketch or paraphrase some of it to give you, the reader, some idea of the strengths and weaknesses of prompt engineering.

A point to be made is that you never really know what quite what you are going to get as the output from an LLM using a prompt. For a start, the relationship between the prompt and the response is probabilistic and usually exactly the same prompt can produce two or three different responses. (If the LLM produces these automatically, they are sometimes called drafts.) Then there is the fact that the textual prompts are given in natural language, say English, and any natural language has a penumbra of uncertainty and ambiguity about meaning. In computer programming, the programs are usually written in a programming language, say Haskell, which has a formal definitive syntax and semantics. There is no doubt whether an expression in the language is well-formed and no doubt as to what it means. In contrast, the LLMs are being prompted by textual strings which may or may not be correct English and which may or may not have an unambiguous meaning. Then the response, the output, is also in possibly ambiguous English in an intended or unintended format. These considerations suggest that there is a trial-and-error, and experimental, aspect to prompting and prompt engineering. In some contexts, this might not matter, when writing fiction for example. But in other contexts, for example, when generating coherent and logically sound legal arguments it might be important.

### ***B.2.1 Basic Examples of Zero-Shot Prompting***

Zero-Shot prompting is where you just give the instructions in text (you do not give examples or 'shots').

'... no demonstrations are allowed and only an instruction in natural language is given to the model' (Brown et al. 2020, 5).

Here are some examples of the following typical NLP processes of interest to us:

- Explanation
- Text Summarization
- Information Extraction, Question Answering
- Text Classification
- Conversation
- Code Generation
- Reasoning

#### **Explanation**

[The prompt]

Explain prompting used in Large Language Models

[GPT-3.5 replies]

Prompting is a technique used in large language models like GPT-3 to generate desired responses or outputs. When using a language model, you typically provide it with a prompt, which is a short piece of text that frames the context or question for the model. The model then generates a continuation of the prompt based on its training and the patterns it has learned from the data.

Prompting serves multiple purposes:

Context setting: The prompt helps establish the context for the model, giving it an understanding of the desired topic or

domain. For example, if you want to generate a news article about a specific event, you can provide a prompt that summarizes the key details of the event.

**Instructional guidance:** The prompt can include explicit instructions or guidelines for the model to follow. These instructions help steer the model's response in a particular direction. For instance, if you want the model to write a poem about nature, you can include instructions like "Write a poem about the beauty of nature and the changing seasons."

**Question-answering:** You can use prompts to ask specific questions to the model. By providing a question as a prompt, you can expect the model to generate an answer or response based on its understanding of the question and the knowledge it has acquired.

**Desired output control:** By carefully designing the prompt, you can influence the style, tone, or structure of the generated text. For example, if you want the model to produce a response in a formal or casual tone, you can explicitly specify that in the prompt.

It's important to note that the effectiveness of prompting can vary depending on the language model and the specific task at hand. Experimentation and iteration with different prompts are often required to achieve the desired results.

## **Text Summarization**

Summarize your last response

[GPT-3.5]

Prompting is a technique used in large language models like GPT-3 to guide their output. By providing a prompt, which is a short piece of text, you can establish context, give instructions, ask questions, or control the desired output. Prompts help the model generate more relevant and specific responses based on its training and learned patterns. It's important to experiment and iterate with different prompts to achieve the desired results.

## Information Extraction, Question Answering

In your last response you mentioned an LLM, which LLM was that?

[GPT-3.5]

Apologies for the confusion. In my previous response, I mentioned GPT-3 as an example of a large language model (LLM). GPT-3 stands for "Generative Pre-trained Transformer 3" and is one of the most well-known and widely used LLMs developed by OpenAI. It has been trained on a vast amount of text data and is capable of generating coherent and contextually relevant responses based on provided prompts.

What is a good way to learn how to use a LLM? Please keep your answer brief.

[Bard, with the option to view other drafts]

Here are some good ways to learn how to use a LLM:

- Read the documentation. Most LLMs come with extensive documentation that explains how to use them. This is a great place to start learning the basics of how to interact with the model.
- Take a tutorial. There are many tutorials available online that can teach you how to use a LLM. These tutorials can be a great way to learn the basics of how to use the model and how to apply it to different tasks.
- Experiment. The best way to learn how to use a LLM is to experiment with it. Try different prompts and see what kind of responses you get. You can also try using the model to complete different tasks.
- Join a community. There are many online communities where people discuss LLMs. These communities can be a great resource for learning new things about LLMs and getting help from other users.

Here are some additional tips for learning how to use a LLM:

- Start with a simple task. When you're first starting out, it's best to start with a simple task. This will help you learn the basics of how to interact with the model without getting overwhelmed.
- Be patient. LLMs can be slow, especially when they're first being trained. Be patient and give the model time to generate a response.

- Don't be afraid to ask for help. If you're stuck, don't be afraid to ask for help from other users or from the documentation. With a little practice, you'll be using LLMs like a pro in no time!

## Text Classification

Is the following text friendly or confrontational? Text: Hey you get offa my cloud

[Bard]

The text "Hey you get offa my cloud" is confrontational. It uses the words "hey" and "get offa," which are both informal and aggressive. The sentence also contains a direct command, which can be seen as rude.

Here are some ways to make the text more friendly:

- "Excuse me, but could you please get off my cloud?"
- "I'm sorry, but I need to use my cloud. Could you please move?"
- "Would you mind getting off my cloud? I need to use it."

These sentences are all more polite and less confrontational than the original text. They also use more formal language, which can make the request seem more reasonable.

## Conversation

Hello, how are you?

[GPT-4 No plugins]

Hello! As an artificial intelligence, I don't have feelings, but I'm here and ready to help you. How can I assist you today?

Would you like to hear how I am?

[ChatGPT-4]

Of course! I'm here to listen. Please feel free to share how you're doing.

## Code Generation

Two examples:

Write code to add up some numbers and return their average.

[ChatGPT-4]

Sure, here's a simple Python function to calculate the average of a list of numbers:

```
def average(numbers):
    return sum(numbers) / len(numbers)
numbers = [1, 2, 3, 4, 5]
print(average(numbers))
```

In this code:

- The **average** function takes a list of numbers as an argument.
- It uses the built-in **sum** function to add up the numbers, and the **len** function to find how many numbers there are.
- It returns the sum divided by the count, which is the average.
- The list **numbers** is defined as an example, and the average of its values is printed.

Please replace the numbers in the list with the numbers you'd like to calculate the average for.

write a parser to parse LISP expressions

[GPT-4]

Sure, here's a very basic recursive LISP parser written in Python. This parser will be able to handle simple LISP expressions like **(add 1 2)** or **(mul (add 1 2) (add 3 4))**. Please note that this is a basic implementation and does not cover all aspects of LISP. It's a simplified version to give you a head start.

---

```
def parse_lisp(expression):
    def tokenize(expression):
        return expression.replace('(', ' (') \
            .replace(')', ') ').split()

    def read_from_tokens(tokens):
        if len(tokens) == 0:
            raise SyntaxError('unexpected EOF')

        token = tokens.pop(0)
        if token == '(':
            L = []
            while tokens[0] != ')':
                L.append(read_from_tokens(tokens))
```

```

        tokens.pop(0) # pop off ')'
        return L
    elif token == ')':
        raise SyntaxError('unexpected ')
    else:
        try: return int(token)
        except ValueError:
            try: return float(token)
            except ValueError: return
str(token)

    tokens = tokenize(expression)
    return read_from_tokens(tokens)
# Example usage:
print(parse_lisp('(add 1 (mul 2 3))'))

```

---

This code takes a LISP expression as input, tokenizes it (splits it into separate parts), and then recursively parses the tokens into nested lists. Symbols are kept as strings, and numbers are converted to integers or floats.

Remember that this parser won't actually *evaluate* the LISP expressions—it just parses them. If you want to evaluate them as well, you'll need to implement an evaluator that understands the different LISP functions (like **add**, **mul**, etc.).

/\*Note from MF. This is not an off-the-chart difficult challenge, but it is not trivial either. The answer is pretty impressive. \*/

## Reasoning

The detective reasoned. If there is blood on the knife, the butler committed the crime. There is no blood on the knife. Therefore the butler did not commit the crime. Is this valid reasoning?

[GPT-4]

The reasoning here seems to follow the format of Modus Tollens, which is a valid form of argument in classical logic. However, the

initial conditional ("If there is blood on the knife, the butler committed the crime.") may not be necessarily true in all cases.

Modus Tollens reasoning goes like this:

1. If P, then Q.
2. Not Q.
3. Therefore, not P.

In this case:

1. If there is blood on the knife (P), then the butler committed the crime (Q).
2. There is no blood on the knife (Not P).
3. Therefore, the butler did not commit the crime (Not Q).

While it is valid reasoning according to Modus Tollens, it is contingent on the truth of the initial conditional statement. In this case, the conditional statement seems to suggest that only the butler could have left blood on the knife, which is quite an assumption. The blood could have come from someone else or the butler might have cleaned the knife, for instance. Therefore, while the logic is structurally sound, the premises (the assumptions or facts that the argument is based on) may not be valid, and thus the conclusion could potentially be flawed.

**/\*Note from MF. *GPT-4's answer here basically is completely wrong.* It pulls up Modus Tollens, which is fine. But then it muddles up its Ps and Qs. Look at the lines numbered 2. The Modus Tollens schema requires Not Q but then it uses Not P as an instance of that.**

**GPT-4 is attempting reasoning, but does not succeed in this case. \*/**

### ***B.2.2 Examples of Few-Shot Prompting***

Few-shot prompting covers cases where examples are given. Outside of LLMs it is very common when giving instructions to give at least one example of what is required (say, when teaching mathematics or passive voice construction in English writing). Most LLMs can work just from examples— they do not need the instruction or explanation.

Here is an example of one-shot. The LLM has been given a single example of an English to French translation and, without even being told what the task is, it has figured out what to do and has done it.

```
sea otter => loutre de mer
peppermint =>
[GPT-4]
The translation for "peppermint" in French is "menthe poivrée".
```

[While this is an example of one-shot prompting, it is unclear how good it is as a test of the LLM. It is from (Brown et al. 2020, 7). That is the main research paper on few-shot prompting. The LLM may have seen it thousands of times during training.]

Here is an example of many-shot.

```
cat=>cats
rabbit=>rabbits
mouse=>
[GPT-4]
The plural form of "mouse" is "mice".
```

You can see here how problems might arise. The LLM needs to spot the pattern. The two examples can be construed either as forming the plural form or as adding the letter 's' on the end of a word or noun.

```
cat=>cats
rabbit=>rabbits
mouse=>mouses
man=>mans
woman=>
```

[GPT-4]  
The plural of "woman" is "women".

Then

ctt=>ctts  
bbit=>bbits  
xouse=>xouses  
mnn=>mnns  
woman=>

[GPT-4]  
Following the same pattern you've established with the other words, the transformation of "woman" would be "womans". However, it's important to note that in standard English, the plural of "woman" is "women".

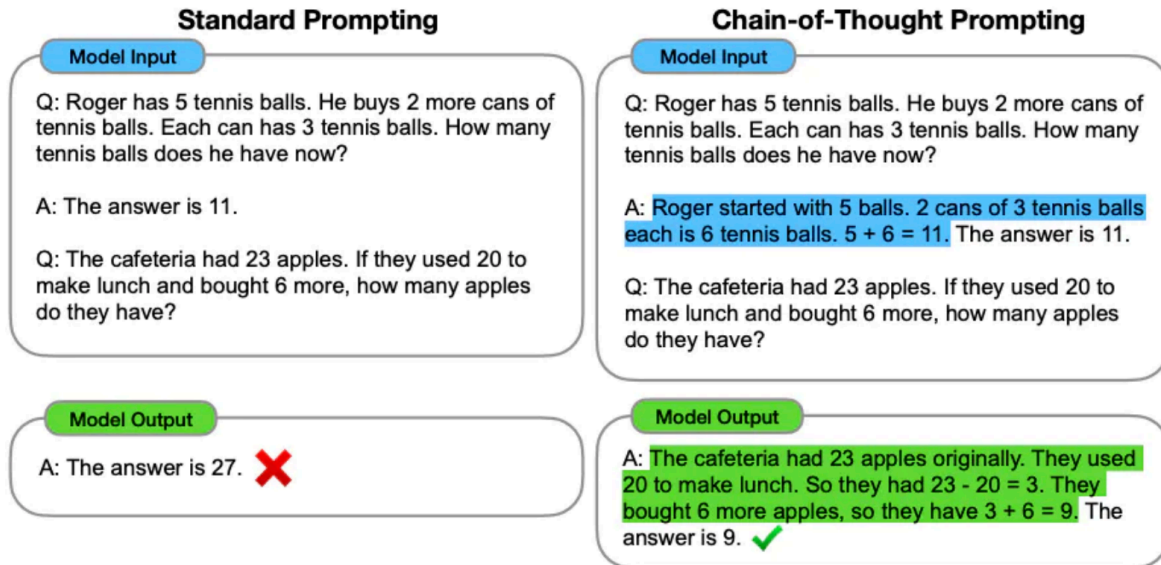
It is just too clever!

### ***B.2.3 Chain of Thought Prompting***

Jeff Dean writes

One of the broad key challenges in artificial intelligence is to build systems that can perform multi-step reasoning, learning to break down complex problems into smaller tasks and combining solutions to those to address the larger problem. (Dean 2023)

When ChatGPT came out, say November 2022, it was clear the available LLMs were weak in this area. Plainly, they could not do mathematical, logical, and scientific reasoning. It was found that to a degree this could be ameliorated by asking the LLM to show its reasoning, to show the steps it made— i.e. to show its chain of thought (Wei and Zhou 2022; Wei et al. 2023). One of the examples that Jason Wei and Denny Zhou give is this



**Figure 28. Standard Prompting and Chain-of-Thought Prompting (Wei and Zhou 2022).**

There is a one-shot prompt here, showing an example and the example displays the reasoning. It soon became clear that this could be done zero-shot, without examples, merely by adding to the original prompt:

Let's think step by step.

Then this adjustment of the prompt could be done automatically (for example, by the use of templates, which we will get to). This is an ongoing area of experimentation and discovery. Getting the LLM to think step by step, and to show its reasoning, is a good idea.

### ***B.2.4 Tuning, or Configuring, the Models or Prompts***

Most LLMs have a parameter— 'temperature'— that can be set prior to prompting. The background explanation of this is that an LLM's response involves a choice among probabilities. For example, with cloze task like

London is the capital city of [mask]

there will be several possibilities for the blank e.g. 'England', 'culture', 'fashion', 'progress', etc. and these will have different probabilities. Now, if the LLM always chooses the most probable response to a word, phrase, sentence, etc., it will always give the same answer. You might not want this. For example, if, on different romantic occasions, like date night once a week, you asked for a love poem for your sweetheart, you presumably would not want the same poem each time. On the other hand, if the LLM always chooses improbable answers, the answers may be interesting and humorous, but also often plain wrong. So, there is a parameter 'temperature' which adjusts for this. A higher temperature means that the LLM is more likely to generate creative and interesting answers, while a lower temperature means that LLM is more likely to generate factual and accurate answers. Temperature is in the range 0-1 and it would ordinarily be set by default to 0.7. In the case of Bard, it can be set using the following syntax:

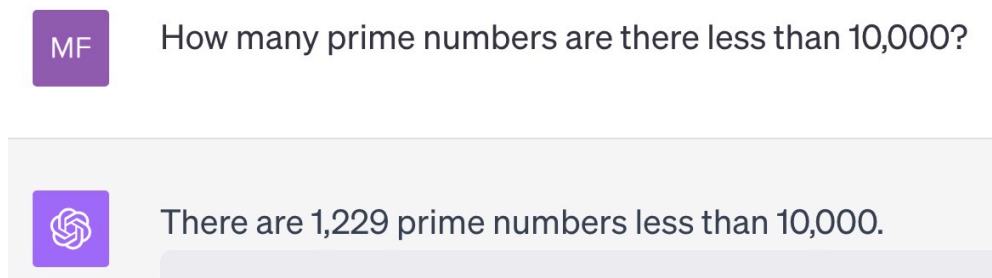
```
bard.generate(prompt, temperature=0.5)
```

For example:

```
bard.generate("London is the capital city of ...", temperature=0)
```

The syntax for setting the temperature in other LLMs may well be different.

We know that typically the large models can make mistakes. They are not 100% reliable. GPT-4V is no different in this regard. We also know that the models can give two contradictory answers to the same prompt in the same context. That such answers are contradictory means that one of these answers is false. That is, more generally, the models can give false answers. Also, in more than a few cases there can be no practical means for the user to check the answer. For example:



**Figure 29. GPT-4V on Prime Numbers.**

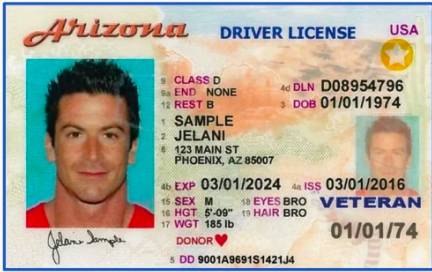
This answer may be right, or it may be wrong. Typical users would have no idea how to check which it is, or whether they should repose trust in what GPT-4V is telling them.

There are ways to improve the quality of answers. There are two standard techniques: imposing constraints and giving the LMM (or LLM) a reputation.

**Constraint prompting** is requiring the LMM to answer in a restricted (‘constrained’) way or format. For example, one constrained prompt to read a driving license would be:

**Constrained Prompting**

**Prompt:**  
Please read the text in this image and return the information in the following JSON format (note xxx is placeholder, if the information is not available in the image, put "N/A" instead).  
{ "class": xxx, "DLN": xxx, "DOB": xxx, "Name": xxx, "Address": xxx, "EXP": xxx, "ISS": xxx, "SEX": xxx, "HGT": xxx, "WGT": xxx, "EYES": xxx, "HAIR": xxx, "DONOR": xxx }



**GPT-4V:**  
{ "class": "D", "DLN": "D08954796", "DOB": "01/01/1974", "Name": "Jelani Sample", "Address": "123 Main St, Phoenix, AZ 85007", "EXP": "03/01/2024", "ISS": "03/01/2016", "SEX": "M", "HGT": "5-09", "WGT": "185 lbs", "EYES": "BRO", "HAIR": "N/A", "DONOR": "Y" }

**Figure 30. Screenshot of GPT-4V Reading a Driving License from (Yang et al. 2023).**

This is better than just prompting the LMM to ‘read the text in the image. [Notice that it has made a mistake with the HAIR.]

As explained in Sections 3.10 and 4.7, Base Models of LMMs are trying only to produce the next word as output. They need further training to become Instructor Tuned, or Fine Tuned, LLMs. But even at this point, they are

only doing an extension of their training. Many kinds of instruction or prompting can invoke an activity without necessarily emphasizing that the outcome be correct or true. Yang et al. mention in their paper the prompt ‘Count the number of apples in the image’, where the image is of apples laid out in rows and columns (Yang et al. 2023). Now, depending on its training, an LMM can count without really focusing on producing a correct answer. A better style of prompt **conditions on good performance**. An example prompt for the apple image would be ‘You are an expert in counting things in images. Let us count the number of apples in this image by <and then give detailed instructions on how to do it, and how to check the answer.>’ Some of the recent LLMs accept both ‘System Prompts’ and (Other) ‘Prompts’. A system prompt could then be used to set the system as being in the context where it was an expert in counting objects in images. This use of system prompts is often used to enhance ‘steerability’. If you would like to ‘steer’ an LLM to be an expert in counting or to answer in the style of Emily Dickinson, you would use a system prompt to do it.

### **B.3 Choices on Development**

Most ML programming work will either use Python or allow Python to be used. In turn Python is generally available on Macs, Windows, and Linux machines. Python will use ‘pip’ which is a package installer. Pip will usually be installed automatically with Python.

To check whether you have what is needed use a terminal and try the following 3 commands:

```
% python3 --version
Python 3.11.3

% where python3
/Library/Frameworks/Python.framework/Versions/3.11/bin/python3
/Library/Frameworks/Python.framework/Versions/3.8/bin/python3
/usr/local/bin/python3
/usr/bin/python3

% pip --version
pip 23.1.2 from
/Library/Frameworks/Python.framework/Versions/3.11/lib/python3.11/site-packages/pip (python 3.11)
```

The response to you on your machine should be somewhat similar, although probably not exactly the same. If you seem not to have Python, go to <https://www.python.org/> and install it.

To give you a general idea where we are going here. We are not going to create any LLMs (which would take months and might cost tens of millions of dollars). We are going to work with existing LLMs (e.g. GPT-4) using their APIs. We will send some messages and get replies, both of which we can process as we would like. There will be short Python programs—usually around 50 lines—to do the work. Typically, these Python programs will have 'import' statements at the top, for example:

```
import langchain
from langchain.llms import OpenAI
```

```
from langchain.utilities import WikipediaAPIWrapper

### langchain this is a convenient host system for working
with LLMs
### OpenAI gives us the API to GPT-3 GPT-4 etc.
### WikipediaAPIWrapper allows us to ask for material from
Wikipedia
```

Of course, starting out you will not have the OpenAI (and other) packages on your computer. This is where the Python package installer pip comes into play. There will be commands like:

```
% pip install -r requirements.txt

% pip install langchain
```

where, the requirements file has a list of, surprise, the requirements. You do not need to worry about this right now.

As to Foundation Models, the LLMs, they can be proprietary or open-sourced. Proprietary are closed-source models, usually vastly expensive to make and are owned by commercial companies. That they are 'closed-source' means that the source computer code is not available to you. You cannot see it, you do not know what it is. This does and does not matter. It does not matter in that there is no possibility of you altering the code and re-pre-training the model, because that might cost \$10 million which you do not have to spare. It does matter in that whoever owns the code might change it or withdraw it (they might have a huff, stop playing the game, and take their ball and go home). It also matters in that closed-source may be preferable in that it presumably keeps powerful code out of the hands of

bad actors. It may be possible to get some kind of closed-source version of a Foundation Model, which has locked in parameters. That might protect you from malfeasance, or monopolistic behavior, by the Model's owner. If LLMs develop as it seems they are going to, the models will have a role somewhat similar to public utilities. We need to be cautious when letting private companies have control of public utilities. Some Foundation Models have been open sourced, and are freely available to all (to use, see, or develop). Hugging Face provides a hub, a library of open-source Foundation Models (Hugging Face 2023). When you work with the APIs of OpenAI (or with HuggingFace, or with Pinecone...) you will need to get an API 'key'. Getting these may vary a little from host to host. In the case of OpenAI, go to their website <https://openai.com/> , create an account, then View API keys under your profile, and proceed from there. There are charges. Typically, they will give you a \$5 credit. The charges are reasonable. There is variation here but, ball-park figures, you can use their API all day for less than a dollar.

Educators working with Python might use either or both of Jupyter notebooks or an Integrated Development Environment (IDE). We should probably follow along with both here.

## **B.4 Moving Forward With LangChain**

***B.4.0 A Note on the Status of LangChain and Similar as of 11/6/2023***

As mentioned earlier, on 11/6/2023, OpenAI released its builder framework for GPTs and these are assistants, some close to being agents, based on GPT software. Also, many of the techniques that third-party developers had produced in the previous six months to enhance the original ChatGPT (such as being able to process private libraries, say of pdfs) have been rolled in to GPT-4 Turbo. This means that developers, and possibly even some start-up companies, are on shaky ground. OpenAI may have and release better software, or they may simply absorb or mimic any outside software that appears.

In the case of LangChain, it can work with LLMs from many different companies (e.g. from Google, from Meta, etc. — in fact with most of the open-source components from Hugging Face). There is a pedagogic advantage to this— there is no lock-in to the one company (no matter how benevolent that company might be).

The whole AI-LLM landscape is changing rapidly, and no one know exactly what is best. Our view is: doing some work with LangChain is good for learning, but it might be prudent not to form your start-up around LangChain and seek venture capital funding on that basis.

#### ***B.4.1 What is LangChain?***

LangChain is a software development framework designed to simplify the creation of applications using large language models (LLMs). As a language model integration framework, LangChain's use-cases largely overlap with those of language models in general, including document analysis and summarization, chatbots, and code analysis. (Wikipedia 2023g)

LangChain was created by Harrison Chase in October 2022 as an open-source project (Chase [2022] 2022). As of May 2023, 836 developers had contributed to it. What it does is:

1. Allows access to LLMs such as closed-source ones from OpenAI (e.g. GPT-4) or open-source ones from Hugging Face (e.g. LLaMA by Meta AI or Flan-T5 by Google).
2. Allows access to software to extract and manipulate text from resources such as pdfs (e.g. a User's research papers, a library's holdings) or other sources (e.g. Wikipedia)
3. Allows access to software that creates embeddings.
4. Allows access to vector databases to store, augment, or retrieve embeddings (such as Pinecone in the cloud) or (FAISS (Facebook AI Similarity Search) on a local machines).
5. Allows the combination of all these into 'chains' which can carry out any or all of the standard NLP, and other, operations (as a really simple example: Find the population density of the capital of Great Britain [This might involve determining the capital of Great Britain; then, having found that, determining its population, determining its area, and dividing one figure by the other].)
6. Allows the combination of chains into 'agents', intelligent assistants which can carry their operations autonomously.

LangChain can be used to build applications such as chatbots, question answering systems, summarization systems, and computer code writing applications. LangChain looks very promising. LangChain has a website

<https://langchain.com/>, and a conceptual guide <https://docs.langchain.com/docs/> You should scan the conceptual guide.

Prompt templates (once you get it right stick with it)

Most of Meta/Facebook LLMs are open sourced.

### ***B.4.2 LangChain Experiments Displayed to a Web Page***

There are many excellent Integrated Development Environments (IDEs) for Python. Two of them are: PyCharm from JetBrains <https://www.jetbrains.com/pycharm/> (there is a free version), and Visual Studio Code (VSC) from Microsoft <https://code.visualstudio.com/> (and this is free). JetBrains have been very supportive of instructional courses that the author has taught— a tip of the hat to them. But, right now, we will go forward with VSC. There are many YouTube videos on how to work with Python and VSC. Here is one:

Setup Python Using Visual Studio Code On Mac [Sonar Systems, the author, have similar videos for the other Operating Systems.]

The main thing with VSC is that you need to have the extension for Python installed. There is a separate mini-gotcha. When evaluating a Python file, VSC might ask you which Python interpreter to use. Then, when you try to make a choice, it might say there isn't one installed. At this point, it will

allow you to insert a path to an interpreter. Since you have run the `% where python3` command (as shown above) you will have path that you can insert.

To get a web page display we will use Streamlit (<https://streamlit.io/>) which is a fast and convenient way to produce web applications from Python code (particularly from Python ML prototypes). You will need an OpenAI key (getting one has been described earlier). You will need a running Python: either use a VSC installation, as described, or a set-up of your own. You will need to install some Python packages. Open a terminal in VSC and use pip install:

```
% pip install streamlit langchain openAI wikipedia
```

Shortly, you will create a Python file in VSC, say `app.py`. Then streamlit will run it (from the VSC terminal):

```
% streamlit run app.py
```

That should get you started with LangChain. Let's experiment away!

## **App Framework**

This code will just open a web page for us. We will put our API key in at this point, for use later. Copy and paste the code into a Python file in VSC, say `app.py`, and run it by executing `streamlit run app.py` from a VSC terminal. Streamlit will keep running. You can stop it by typing `Ctrl-c` :

```

# % pip install streamlit langchain openAI wikipedia

import streamlit as st
import langchain
import os

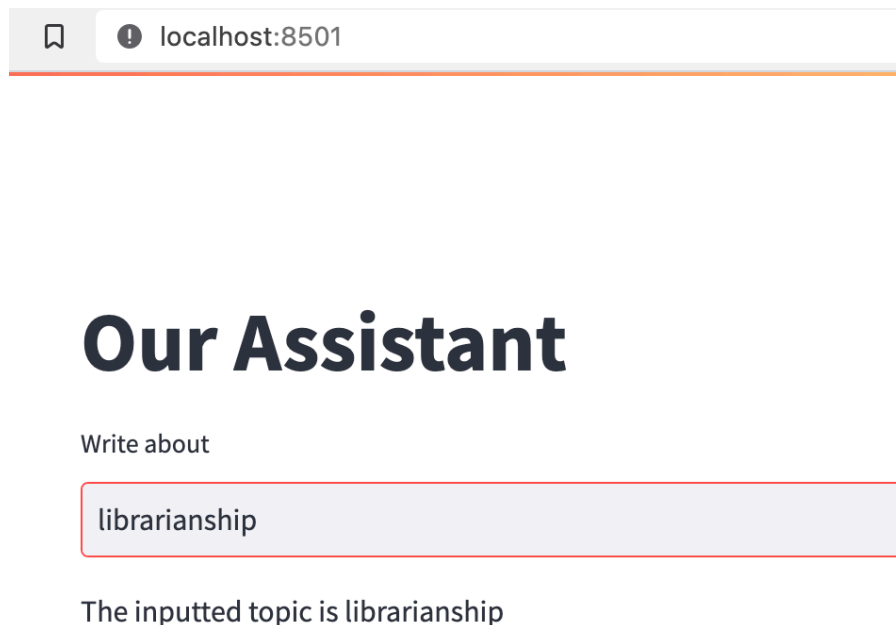
os.environ["OPENAI_API_KEY"] = <INSERT YOUR API KEY HERE> #
sorry you can't use mine

#Streamlit display
st.title ('Our Assistant')
input = st.text_input ('Write about')

#Testing we have input
if input :
    st.write('The inputted topic is ', input)

```

The resulting web page should look similar this:



**Figure 31. Assistant Showing Input of 'librarianship'.**

## Adding an LLM

We can add an LLM (and check its response):

```
# % pip install streamlit langchain openAI wikipedia

import streamlit as st
import langchain
import os
from langchain.chat_models import ChatOpenAI

os.environ["OPENAI_API_KEY"] = '<INSERT YOUR API KEY HERE>'
# insert your API_TOKEN here

#Streamlit display
st.title ('Our Assistant')
input = st.text_input ('Write about')

#Get ourselves an llm
llm = ChatOpenAI(model_name="gpt-3.5-turbo")

#Testing we have input and letting the llm respond
if input :
    response = llm.call_as_llm(prompt)
    st.write(response)
```

## Our Assistant

Write about

Librarianship is the profession that involves the management and operation of libraries and information centers. It involves selecting, acquiring, organizing, preserving, and disseminating information and resources to users. Librarianship encompasses a wide range of skills and knowledge, including information technology, cataloging, reference services, collection development, outreach and community engagement, and management. Librarianship is essential in providing access to information and resources for education, research, and personal enrichment. Librarians work in a variety of settings, including public libraries, academic libraries, school libraries, and special libraries.

## Figure 32. Assistant Showing Response to Input of 'librarianship'.

### Prompt Templates

Prompts are hard to get right. So, once you have a recipe that works it is a good idea to make a template out of it and use that. Templates have zero or more variables and some boilerplate. Then values for the variables will be inserted into the template to produce the actual prompt. We will make a template out of:

```
"Write one paragraph in the style of Emily Dickinson on the topic of {topic}"
```

```
# % pip install streamlit langchain openAI wikipedia

import streamlit as st
import langchain
import os
from langchain.chat_models import ChatOpenAI
from langchain import PromptTemplate

os.environ["OPENAI_API_KEY"] = <INSERT YOUR API KEY HERE>
#insert your API_TOKEN here. Sorry you can't use mine

#Streamlit display
st.title ('Our Assistant')
input = st.text_input ('Write about', key='about')

#Get ourselves an llm
llm = ChatOpenAI(model_name="gpt-3.5-turbo")

#Create a template
template = "Write one paragraph in the style of Emily Dickinson on the topic of {topic}"
```

```

prompt_template = PromptTemplate.from_template(template)

#Testing we have input and letting the llm respond
if input :
    real_prompt = prompt_template.format(topic=input) #put
the input into the template
    st.write("Real prompt:", real_prompt)
#check
    response = llm.call_as_llm(real_prompt) #ask llm
    st.write(response)

```

## Our Assistant

Write about

librarianship

Real prompt: Write one paragraph in the style of Emily Dickinson on the topic of librarianship

The librarian, an unseen scholar, toils amongst the stacks, a sentinel of knowledge and keeper of the boundless riches of the written word. With quiet diligence and an unwavering commitment to the pursuit of truth, the librarian is a custodian of history, a guardian of ideas, and a beacon of hope for those seeking refuge in the sanctuary of the library. Her presence is felt in the creak of the shelves, the rustle of the pages, and the hush of the reading room. Oh, how we owe our gratitude to the librarian, who with each turn of the page, illuminates the world with the light of knowledge and the promise of a brighter tomorrow.

**Figure 33. Assistant Showing Response to Input of 'librarianship' in the style of Emily Dickinson.**

Of course, the LLM in general will likely give a different response every time it is asked. Here is an alternative that it offered:

The librarian, with her quiet grace, is the keeper of the written word. She moves among the stacks, a gentle guardian of knowledge. With each book she takes down, a new world unfolds before her eyes. She is a seeker of truth, a lover of language, and a protector of ideas. Her hands are calloused from years of handling pages, but her spirit remains unbroken. She is the keeper of

secrets, the purveyor of dreams, and the champion of learning.  
The librarian is a treasure, a rare gem of wisdom and wonder.

## **Document Embeddings**

LLMs will be trained possibly on a large portion of the public facing Internet. But there are occasions when you might want to supplement this. As examples, an academic might want regularly to access and query all her own research publications and lectures, a library might have a special collection of materials in need of LLM processing.

If the new content is brief— say a single newspaper article— it could be prepended to the prompt. But the prompts can only be short, say a few thousand 'tokens', so this idea is clunky and is not going to work in general. What is needed is for all the new documents to be processed into embeddings (i.e. lists of numbers) and those lists to be stored in a vector database. The LLM will then augment its background by using the database.

Python and LangChain have a number of tools to get the documents in the right form in the first place (e.g. Python Beautiful Soup can scrape web pages, Python PyPDF library can extract texts from pdfs).

Just as a brief example here. Chapter 1 of this text has been printed as a pdf and put in a directory called SamplePdfs. That is our library (we needed a sample pdf not covered by copyright). We will convert it to embeddings (numbers), and store those in a vector database. We will use FAISS

(Facebook AI Similarity Search) database. Then we can ask questions of it. Here is an example

## Our Assistant

Ask Emily Dickinson to write about

Ask a question of your pdf library

Explain the student backpacker analogy as an example of reinforcement learning

In the student backpacker analogy, the student is working as an apple picker in an orchard. Her goal is to maximize her pay, which is equivalent to maximizing her rewards in reinforcement learning. She receives immediate rewards for each apple she picks and additional bonuses for filling baskets of apples faster than other pickers.

The student engages in trial-and-error exploration by trying different strategies for picking apples. For example, she may focus on picking bigger apples to fill baskets faster but at the cost of having fewer apples in each basket. The rewards she receives provide feedback on the effectiveness of her strategies.

Through repeated experiences of picking apples and receiving rewards, the student learns which strategies yield higher total rewards. She may learn to prioritize certain types of trees, pay attention to other pickers' strategies, and adapt her approach over time. This process of learning from trial-and-error and adjusting strategies based on rewards is a fundamental concept in reinforcement learning.

### **Figure 34. Assistant Explaining the Backpacker Analogy.**

GPT-3.5-Turbo is answering our question, not from what it knows from its training but from our library. In this simple case, the library has a single 47 page pdf in it. But it easily could have every piece of research that a scholar has ever written.

Separately, we could not resist hearing from Emily Dickinson again:

## Our Assistant

Ask Emily Dickinson to write about

algorithms

Ask a question of your pdf library

An unseen force, algorithmic in nature, doth guide the labyrinthine web of life. Its intricate threads, woven with precision and calculation, dictate the paths we traverse, the choices we make. The algorithm, a puppeteer of destiny, pulls the strings in this grand theatrical spectacle. It whispers secrets to the wind, orchestrates the dance of the stars. Oh, how the algorithmic beats of existence, like a metronome's steady pulse, govern the harmonious symphony of our world. Yet, amidst this orchestrated chaos, the individual, with a fragile thread of free will, seeks to defy the algorithm's relentless sway.

### Figure 35. 'Emily Dickinson' Explaining Algorithms.

The python code for our assistant is as follows. [The code here is written in a very idiosyncratic style, please do not take it as anything to be admired or copied.]

```
# % pip install streamlit langchain openAI pypdf

import streamlit as st
import langchain
import os
from langchain.chat_models import ChatOpenAI
from langchain import PromptTemplate

os.environ["OPENAI_API_KEY"] = <INSERT YOUR API KEY HERE>

#Streamlit display. Two entry possibilities
st.title ('Our Assistant')
```

```

input = st.text_input ('Ask Emily Dickinson to write about',
key='about')
question = st.text_input ('Ask a question of your pdf library',
key='question')

#Get ourselves an llm
llm = ChatOpenAI(model_name="gpt-3.5-turbo")

#Create a template
template = "Write one paragraph in the style of Emily Dickinson
on the topic of {topic}"

prompt_template = PromptTemplate.from_template(template)

#Testing we have input and letting the llm respond to the
templated version
if input :
    real_prompt = prompt_template.format(topic=input) #put the
input into the template
    response = llm.call_as_llm(real_prompt) #ask llm
    st.write(response) #write the
response

##### Going to load the pdfs in a directory (use either
unstructured pdf or pypdf)

# pip install langchain unstructured openai tiktoken pypdf

#loading the pdf docs from the SamplePdfs directory

from langchain.document_loaders import PyPDFDirectoryLoader
pdf_folder_path='SamplePdfs/'
loader = PyPDFDirectoryLoader(pdf_folder_path)
docs = loader.load()

#try FAISS (Facebook AI Similarity Search) as our database

#pip install faiss-cpu
from langchain.vectorstores import FAISS

##### Making the embeddings

from langchain.embeddings import OpenAIEmbeddings

embeddings = OpenAIEmbeddings()

```

```

#create the vector store from the embeddings of our library, to
use as the index
db = FAISS.from_documents(docs, embeddings)

#get a 'retriever' that will ask questions of our database
from langchain.chains import RetrievalQA
retriever = db.as_retriever()
qa = RetrievalQA.from_chain_type( llm=llm,
                                chain_type="stuff",
                                retriever=retriever,
                                return_source_documents=True)

#if there is a question, answer it from our library
if question:
    answer = qa({"query": question})
    st.write(answer['result'])

```

[Ben Dickson has a discussion of using your own documents in (Dickson 2023)]

## There is Lots More

Needless to say. Let us hope there is enough here to stimulate your interest.

## A Useful Resource

Nicholas Renotte's excellent *LangChain Crash Course: Build a AutoGPT app in 25 minutes!* is a useful resource. His accompanying video is hosted at <https://www.youtube.com/watch?v=MLK6SIjcjE8> and the code he uses is available from <https://github.com/nicknochnack/Langchain-Crash-Course/blob/main/app.py>.

### B.4.3 LangChain Using Jupyter

If instead, or as well, you may wish to use Jupyter....

A good way to get Jupyter notebooks is to use an Anaconda installation. Go to <https://anaconda.com> and download a free installation (probably from <https://www.anaconda.com/download>). Create yourself a directory (folder) for your Jupyter notebooks. Launch Anaconda-Navigator, that will give you

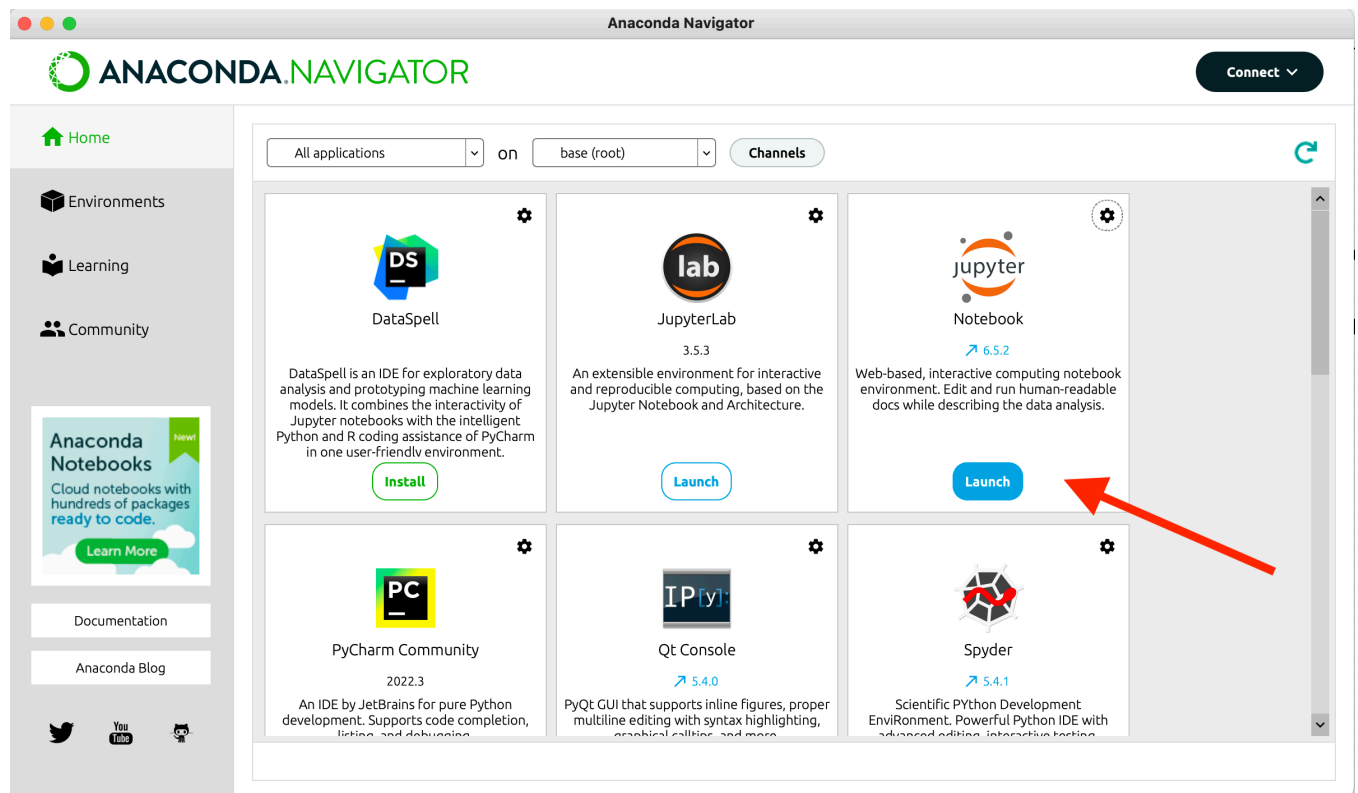
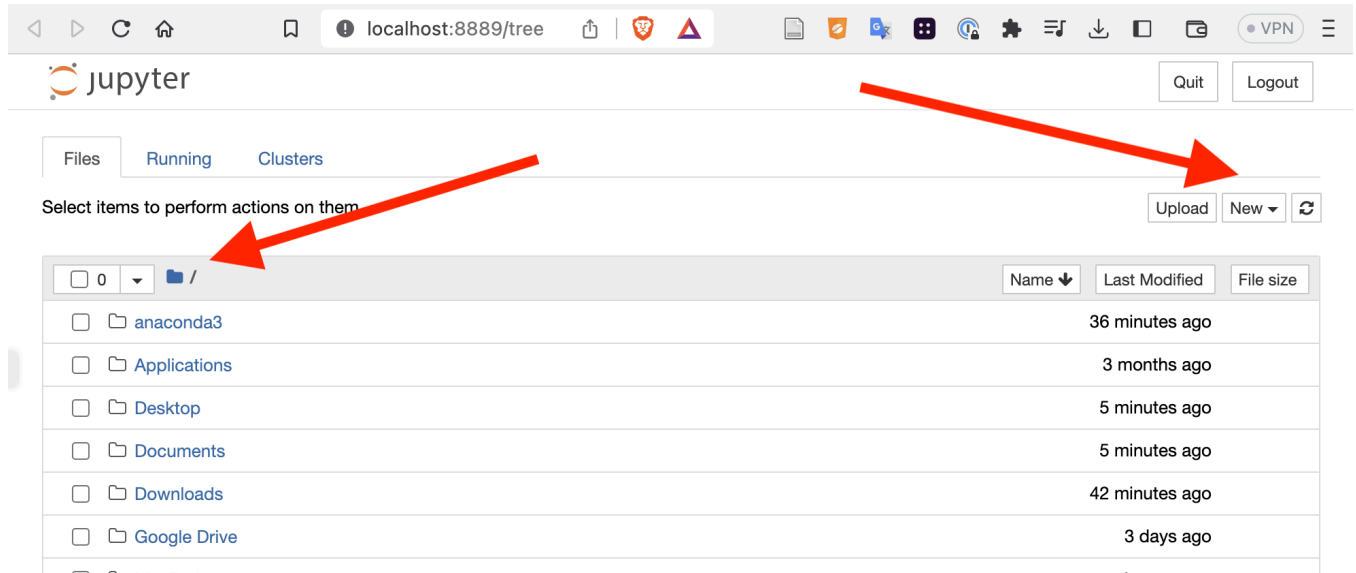


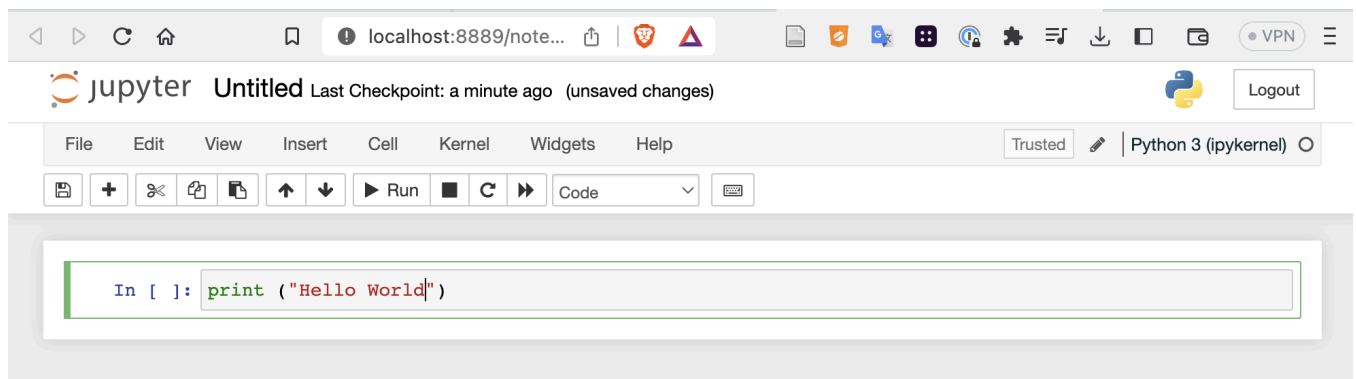
Figure 36. Anaconda Navigator.

Launch Jupyter Notebook. That will open a web page in a browser, looking similar to this



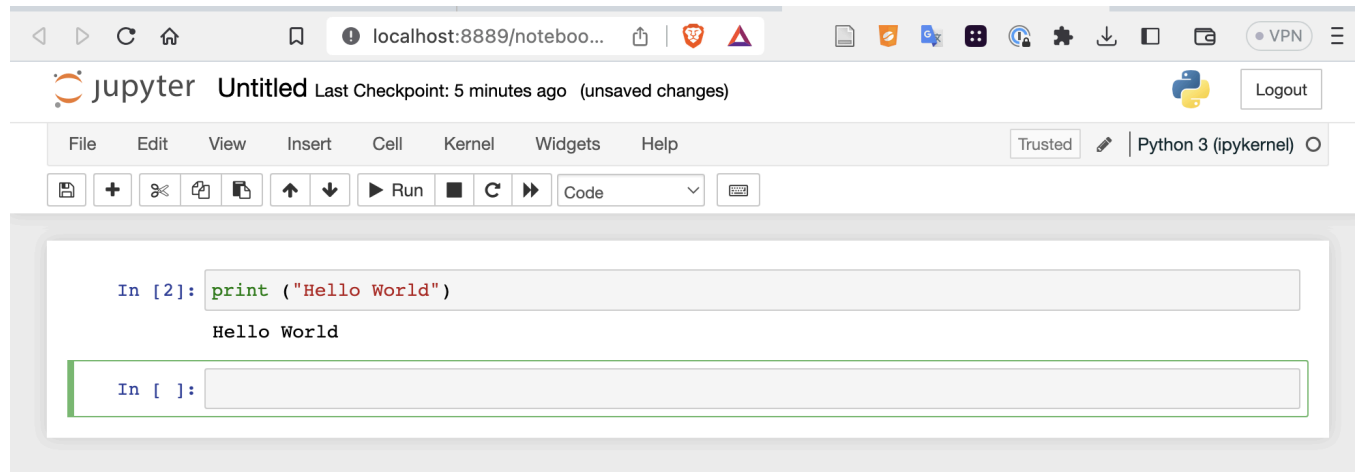
**Figure 37. Jupyter Display of Directory Structure.**

Navigate to the folder you are going to use for your Notebooks and, off the New button, create a new notebook. Type into the first cell `print ("Hello World")`.



## Figure 38. A Jupyter Notebook Being Given the Input 'print ("Hello World")'

Then either click Run or type shift-enter. This will 'run' or evaluate the cell



## Figure 39. Jupyter Evaluating the Cell Displayed in Figure 38.

Now you have Jupyter notebooks running Python. We do not need expertise in either Jupyter or Python, but at least some level of comfort is required. If you feel you do not have that, there are a myriad of excellent resources on You Tube and the Web (for example, Jupyter's own documentation or Codecademy's How to Use Jupyter Notebooks.)

The package installer pip can be used within Jupyter notebooks, but care is needed. The problem is that Jupyter is so powerful that it might be running all sorts of instances of Python in all sorts of places. You need to be sure that the package is installed with the right Python. Use the following code in a cell of your notebook and evaluate it:

```
import sys
!{sys.executable} -m pip install <insert the package you
require here>
```

The `{sys.executable}` fragment picks up the Python that is running.

[A Jupyter notebook has the suffix `.ipnyb`.] We will have an interest in Greg Kamradt's Cookbook (Langchain Cookbook Part 1 - Fundamentals Part 1 `.ipnyb` and Langchain Cookbook Part 2 - Fundamentals Part 2 `.ipnyb`). A way to get this is to go to <https://github.com/gkamradt/langchain-tutorials> click on the green code button and download the ZIP compression of all the files. (Not everyone is comfortable downloading ZIP files from the Internet. If you have concerns, and caution is definitely in order, just do not do it.) Assuming you have the ZIP, expand it into the folder you are using for your Jupyter LangChain files.

#### ***B.4.4 Resources for LangChain using Jupyter***

For LangChain in the context of Jupyter, we are going to suggest two sources: Greg Kamradt's excellent *The LangChain Cookbook - Beginner Guide To 7 Essential Concepts* (Kamradt [2023] 2023), and James Briggs and Francisco Ingham's also excellent *LangChain AI Handbook* (Briggs and Ingham 2022). The latter comes from the company Pinecone, which hosts vector databases in the cloud (an important part of infrastructure for NLP).

For the Kamradt, there is a video and github resources (which we have downloaded as a zip):

- The LangChain Cookbook - Beginner Guide To 7 Essential Concepts [video]\_and there is a second and other videos at (Kamradt 2023)
- The LangChain Cookbook [Jupyter]\_This uses OpenAI.

For the Briggs and Ingham, there are 12 videos from James Briggs, Jupyter notebooks and a book.

- Getting Started with GPT-3 vs. Open Source LLMs - LangChain #1 [video]\_This uses OpenAI and Hugging Face, and it links to the other videos.
- The Jupyter code notebooks for this are at <https://github.com/pinecone-io/examples/tree/master/generation/langchain/handbook> and the notebook for the first video is oo-langchain-intro.ipynb (you can download this by clicking on the download symbol, then you can open it in your running Jupyter).
- The handbook is (Briggs and Ingham 2022).

## **B.5 Annotated Resources for Appendix B**

Costa, Ricky. "ChatGPT Cheat Sheet." <https://Neuralmagic.Com/>, 2023.  
[https://www.kdnuggets.com/publications/sheets/ChatGPT\\_Cheatsheet\\_Costa.pdf](https://www.kdnuggets.com/publications/sheets/ChatGPT_Cheatsheet_Costa.pdf). (Costa 2023)

Deepankar, and Florian. “PromptPerfect - Elevate Your Prompts to Perfection with AI Prompt Engineering,” 2023. <https://promptperfect.jina.ai/>. (Deepankar and Florian 2023)

Huyen, Chip. “Building LLM Applications for Production,” 2023. <https://huyenchip.com/2023/04/11/llm-engineering.html>. (Huyen 2023) This is good. Huyen writes ' It's easy to make something cool with LLMs, but very hard to make something production-ready with them.'

Mollick, Ethan. “Working with AI: Two Paths to Prompting,” 2023. <https://www.oneusefulthing.org/p/working-with-ai-two-paths-to-prompting>. (Mollick 2023)

Monigatti, Leonie. Getting Started with LangChain: A Beginner's Guide to Building LLM-Powered Applications (Monigatti 2023)

OpenAI. “Guide to Prompt Engineering,” 2023. <https://platform.openai.com/docs/guides/prompt-engineering>. (OpenAI 2023c)

## Appendix C: Two Important Methodological Points

There is a conventional symbolization that will help here with probabilities. We will write  $p(A)$  to mean the probability of A where A is some sentence, for example,  $p(\text{it is raining})$ . Then there is the notion of *conditional probability* written  $p(A|B)$ , where A and B are both sentences, and this is read 'the probability of A given B' or 'the probability of A given the condition B', for example  $p(\text{Jane gets wet}|\text{it is raining})$  which would be read 'the probability of Jane getting wet given than it is raining'.

### C.1 False Positives and False Negatives

In medical testing, and more generally in binary classification, a false positive is an error in data reporting in which a test result improperly indicates presence of a condition, such as a disease (the result is positive), when in reality it is not, while a false negative is an error in which a test result improperly indicates no presence of a condition (the result is negative), when in reality it is present. (Wikipedia 2022d)

Aesop's fable of the Shepherd Boy who cried 'Wolf' provides a classical partial illustration of this. We can modify the fable to become a full example. In Aesop, the boy is supposed to cry 'Wolf' when there is a wolf. So, the condition is the presence of a wolf, and the positive test for this is the boy's cry of 'Wolf'. Now, as we all know, in the fable, the boy become bored and lonely, and started crying 'Wolf' even though there was no wolf. These cries are all **false positives** (the test is positive but the condition does not exist). The villagers responded to these false positives, several

times over, rushing to be with the boy and to protect their flock. Later, a wolf actually did appear, and the boy cried ‘Wolf’. This cry is a **true positive** (the test is positive and the condition does exist). However, the villagers did not respond i.e. they ignored a true positive (largely because they were tired of false positives and mistook a true one for a false one). Here is our modification. All of us, the villagers and ourselves, know that not being told that there is a wolf present is not the same as being told that there is not a wolf present. So, the villagers, desirous of peace of mind, set up the situation differently going forward. Apparently, the wolves in that region were crepuscular. That means that they hunt at dawn and a dusk. They gave the Shepherd Boy a second task: he also had the job of crying ‘No wolf’ once at dawn and at dusk, on the condition that there was indeed no wolf. Initially, the Shepherd Boy was conscientious with the second task. He made the appropriate cries in the absence of wolves. These cries are all **true negatives**. They are supposed to indicate the absence of a condition and they do exactly that. But, we know, the boy was a bit of a larrikin. Sure enough, one dusk, a wolf appeared, and yet the boy cried ‘No wolf’. This cry would be a **false negative** (the test is negative, but yet the condition does exist).

During a 24 hour period, the villagers might a) hear nothing (in which case, there would have been a failure in duties), b) hear cries of ‘Wolf’ or ‘No Wolf’ at different times, dawn or dusk (these individually could be true or false positives, or true or false negatives), or c) hear cries of ‘Wolf’ or ‘No Wolf’ at the same time (in which case, they would know that either the ‘Wolf’ cry was a false positive or the ‘No Wolf’ cry was a false negative).

The various relations here between condition and test result can be expressed as conditional probabilities. We will relax our terminological conventions a little and just write ‘WolfCry’ and ‘NoWolfCry’ for the cries and ‘WolfPresent’ and ‘NoWolfPresent’ for the conditions. Then the probabilities are

$p(\text{WolfCry} \text{WolfPresent})$	True positive
$p(\text{WolfCry} \text{NoWolfPresent})$	False positive
$p(\text{NoWolfCry} \text{NoWolfPresent})$	True negative
$p(\text{NoWolfCry} \text{WolfPresent})$	False negative

The general theory here finds its greatest application perhaps in medicine. We all know of false positives, negatives, and the like, in the context of Covid tests. Almost all real world tests do have false positives, false negatives, etc. Typical values for these probabilities might be around 0.05 (roughly, 1 in 20 results is a false positive or false negative). Were the probabilities to be much higher than this, the tests would be regarded as being unsatisfactory.

*As we will see shortly, the fact that a test seems to 'indicate' that a person does or does not have a disease with a certain probability does not actually mean that the probability in question is indeed the probability of the person actually having the disease. There is another factor.*

## **C.2 The Base-Rate Fallacy**

The Base-Rate Fallacy, sometimes known under the heading ‘Harvard Medical School Test’, is probably the most common case of probabilistic

reasoning where in real life almost everyone is tempted to reason incorrectly.

Say there is some dread disease— Lurgi— and there is a test for the disease. This test is very good. So good, in fact, that if anyone actually has the Lurgi then there is a 0.95 probability that the test will show positive (and a 0.05 probability, when the person actually has the Lurgi, that the test will say that they do not have it— the 'false negatives') i.e.

$$\begin{aligned} p(\text{PositiveTest} \mid \text{Lurgi}) &= 0.95 \text{ or } 95\% \\ p(\text{NegativeTest} \mid \text{Lurgi}) &= 0.05 \text{ or } 5\% \end{aligned}$$

Now assume we test John Smith, and sad to say, the test is positive. So, does John Smith have the Lurgi? Is it probable that he has the Lurgi? Do we know anything at all about whether John Smith has the Lurgi (on the basis of this test and its result alone)?

We know two pieces of information

$$\begin{aligned} p(\text{PositiveTest} \mid \text{Lurgi}) &= 0.95 \text{ or } 95\% \\ \text{John Smith has tested positive.} \end{aligned}$$

and we are trying to find out John Smith's status viz-a-viz Lurgi. What we would like to know, or need to know, first off, is this

$$p(\text{Lurgi} \mid \text{PositiveTest}) = ?$$

but there is no way of reasoning from

$$p(\text{PositiveTest} | \text{Lurgi})$$

alone, to

$$p(\text{Lurgi} | \text{PositiveTest})$$

One aspect of the difficulty is that we do not know the rate for the false positives. (Another is that we do not know the background rate, or base rate, for Lurgi. We will get to that shortly). Conclusion: we know nothing about John Smith's status viz-a-viz Lurgi.

Now, let us allow us to have information on the false positives. Say

$$p(\text{PositiveTest} | \sim \text{Lurgi}) = 0.1$$

We now know a) the probabilities for the true positives (0.95) and for the false positives (0.1) and b) that John Smith has tested positive. Do we know whether it is probable that John Smith has Lurgi. What we are being tempted with here is what is known as the *base-rate fallacy* (Amos Tversky and Kahneman 1982).

Many of the misleading (or trick) examples of this are set up in the same way. There is a very low background probability of something, say of a person having Lurgi, which we will set for this example as being 0.01. And there is some sort of test for the condition of having Lurgi, which is pretty good, say  $p(\text{PositiveTest} | \text{Lurgi}) = 0.95$ . But the test also gives some false positives (that is to say, it occasionally indicates that a person has Lurgi when they do not have it), say  $p(\text{PositiveTest} | \sim \text{Lurgi}) = 0.1$ . Then we are

told the following story and asked the following question. A person goes in and is tested for Lurgi and the test is positive, is the probable that the person has Lurgi? Most of us say that it is, whereas, in fact, it is very unlikely. A good way to see this (and how Bayes' Theorem applies) is to re-tell the story in terms of "natural" frequencies. In this story, you live in a town of 10000 people and 100 of them have Lurgi. Everybody is tested for Lurgi. Of those hundred people with Lurgi, 95 test positive. Of those 9,900 without Lurgi, 990 test positive for Lurgi (the false positives). You test positive for Lurgi. Is it likely that you have Lurgi? Well, you have a 95/990 i.e. about a 1 in 11 chance of having it, and about a 10 in 11 chance of not having it. You probably do not have it. The correct reasoning here is an instance of Bayes' Theorem, in the form

$$p(\text{Lurgi}|\text{PositiveTest}) = \frac{p(\text{positive} | \text{Lurgi}) \times p(\text{Lurgi})}{p(\text{positive}|\text{Lurgi})p(\text{Lurgi})+p(\text{positive}|\sim\text{Lurgi})p(\sim\text{Lurgi})}$$

With numbers

$$p(\text{Lurgi}|\text{PositiveTest}) = \frac{.95 \times .01}{.95 \times .01 + .1 \times .99}$$

=>

$$p(\text{Lurgi}|\text{PositiveTest}) = \frac{.0095}{.0095 + .099}$$

What we are tempted to do when reasoning badly is a) to focus on how good the test is when giving positive results from positive cases, and b) ignoring the background rate (how rare the disease is, *simpliciter*). And

what we need to do is a) take the false positives into account, check how often the test gives a positive result from a negative case, and b) remember the background rate (then, roughly: if the disease is rare, and the test can give false positives, the probability is that a positive is a false positive).

Consider this. We have a test that is pretty good in that it usually comes out positive for those that have the disease, but it does produce some false positives, perhaps 5%. This means that if you test 100 people, who do not have the disease, 5 of them might test positive. It also means that if you test 100 million people, who do not have the disease, 5 million of them might test positive. That is quite a lot! So, if you screened the entire population of the US (say 300 million) you might have 15 million false positives. Now if the disease is very rare in the population (for example, folk in the US having Ebola, which might be 1,2, or 3 people only). If you test someone for in the US for Ebola (with one of those 5% false positive tests above), and they test positive, it is much more likely that they are a false positive (and they don't have Ebola) than it is that they have Ebola. Thinking otherwise is the so-called base rate fallacy.

[Experts will know that, strictly speaking, parts of the 'natural frequencies' explanation are not entirely correct in full detail. We will not worry about that here. ]

### **C.3 Annotated Readings for Appendix C**

Howson, Colin, and Peter Urbach. *Scientific Reasoning : The Bayesian Approach*. 3rd ed. Chicago: Open Court, 2006. (Howson and Urbach 2006) Bayesian techniques are widespread in modern science (and in machine learning). The Howson and

Urbach book is excellent. While it is well written and an engaging read, it may be a little advanced for us.

## Appendix D: Causal Diagrams

### D.1 Causation and Correlation

It may be wise to say a word or two about causality. Philosophers have studied causality for thousands of years. They have made progress, but their theories are way too complex for us. The computer scientist Judea Pearl introduced a way of thinking about causality that was suitable for reasoning about causality in the setting of artificial intelligence and machine learning (Pearl 2009b). What was needed here is some principled way of understanding correlation and causation and their differences. Here is a proposal following from the wider work of Pearl and his intellectual colleagues. In statements like:

Taking aspirins causes relief from headaches.

there are three features of interest. There is a direction, a direction in time. The earlier taking of aspirins produces, brings about, or ‘causes’ the later relief from headaches. The later relief from headaches does not produce, bring about or ‘cause’ the earlier taking of aspirins. Second, there is an association or correlation between the cause and the effect. In the example case, there is a regularity between taking aspirins and relief from headaches. This regularity is not an absolute guarantee. For one reason or another, the taking of aspirins does not always relieve headaches on all occasions. However, the taking of aspirins does increase the chance, or likelihood, or probability, that the headaches will be relieved. Thirdly, there is what might be called an intervention or counterfactual factor. Often with

causality, we have the ability to intervene or produce or change or manipulate the cause in an attempt to manipulate the effect. This is a great and desirable feature. Were our teenage child to have a headache, we could give them an aspirin and this may well provide relief. Similarly, here we can reason counterfactually. If the child in fact had not been given an aspirin, we might make the consoling observation ‘you know, an aspirin would have helped you’. Plain correlation, without causation, does *not* have a direction, *does*, or *can*, involve probabilities, does *not* give us the ability to manipulate outcomes, and does *not* support counterfactuals. Cirrhosis of the liver is caused by drinking alcohol. Smoking is correlated with cirrhosis of the liver (among the population to date). Current smokers and non-smokers have different probabilities of having cirrhosis of the liver. Those with cirrhosis of the liver and those without cirrhosis of the liver have different probabilities of being smokers. But an intervention that stops you smoking, if you are a smoker, does not change your probability of getting cirrhosis of the liver (provided all other factors are unchanged, in particular whether you drink or not). Smoking is correlated with cirrhosis of the liver, but it does not cause cirrhosis of the liver.

In sum, with causality, there is a direction, changing of probability, and the possibility of interventions and counterfactuals.

There can be causal talk, correlation talk, and ‘weaselly’ talk (which is talk intending to make the reader think of causality.) Nick Huntington-Klein gives useful examples of the words and phrases in use here:

What are some of these words?

We can say that X causes Y by saying: X causes Y, X affects Y, the effect of X on Y, X increases/decreases Y, X changes Y, X leads to Y, X determines Y, X triggers Y, X improves Y, X is responsible for Y, and so on...

We can say that X and Y are related without implying causality by saying X and Y: are associated, are correlated, are related, tend to occur together, tend not to occur together, go together, and so on...

If some weaselly writer ... doesn't want to *say* causality but does want the reader to *hear* it, they might say: X is linked to Y, X is followed by Y, X has ramifications for Y, X predicts Y, people who X are more likely to Y, Y happens as X happens, and many others.

Knowing these terms can help you interpret what scientific studies are really saying, and when someone might be trying to pull one over on you (Huntington-Klein 2022).

## **D.2 Causal Diagrams**

Researchers in ML, and, indeed, in causality and statistics in general, often employ causal diagrams (Pearl 2009a; Scheines 1997; Pearl 1995). These are useful in many settings, in particular with being assured that predictions have genuine substance, and with addressing questions of bias and fairness. When we can identify causes in a system, two possibilities open up. If we can also manipulate or adjust or change the causes, we may be able to change the effects (and this may be very desirable). Separately, we can start reasoning counterfactually. That is, thinking what would, or might happen, were the causes to be changed (and this counterfactual analysis may give us insight on bias and fairness, among other things).

Causal diagrams use variables and arrows. For example, if we think, as a causal model, that smoking causes lung cancer, the following diagram might be suitable:



**Figure 40. A Causal Diagram Showing Smoking Causing Lung Cancer.**

The arrow indicates our views on causality (in this case, that smoking causes lung cancer— that smoking is a *direct* cause of lung cancer). The arrow is an arrow of causality. The causality flows 'downstream' from the smoking to the lung cancer. Sometimes the direct cause, or tail of the arrow, is called the 'parent', then the adjacent variable that the head of the arrow attaches to would be the 'child'.

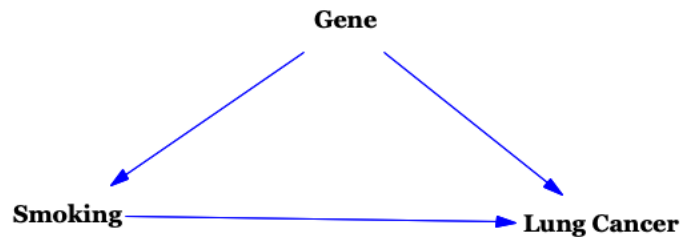
Such diagrams have two roles. As far as causality is concerned, there is no data involved. We have produced this model out of thin air, or out of our ideas, theories, conjectures, or background knowledge. The causality here is not deterministic. It does not mean that every single smoker gets lung cancer as a piece of inexorable mechanistic clockwork. Rather, it is about the category or type or class of smokers— that at least some of them are caused to get lung cancer by their smoking (other things being equal). If this diagram is correct, the second role will come into play, and that is that there will be a statistical connection, or correlation, or dependence, or association, between smoking and lung cancer. Data is involved here. We

would expect to see a correlation between the folk who smoke and the folk who get lung cancer. The diagram says nothing at all about what it is to be a smoker— whether the values here are just 'Yes' or 'No' or whether there are grades or degrees of being a smoker. Similarly, it says nothing at all about the values for the variable lung cancer. Also, the diagram says nothing about what the causal effect is. We know full well that being a smoker *increases* the probability of getting lung cancer. But as far as the semantics of the diagram is concerned, the arrow just asserts that there is some causal effect, or causal association between the two variables. Smoking may increase lung cancer, it may decrease lung cancer, the arrow in the diagram is agnostic on this.

Arrows show causal connection. Lack of arrows show (presumed) absence of causal connection. We have only two variables in our diagram. More complex diagrams may have many more variables. We have to be explicit with our commitments— put arrows between variables if our model assumes a causal connection, omit arrows where we assume no connection. Obviously, for example, in our simple case, something may cause smoking, and lung cancer itself may cause other effects. But a causal diagram does not need to contain the world history, or the world future. There just needs to be enough variables for the problem at hand.

There is a requirement or constraint on the variables and arrows. If any two variables have a common cause— that is, another variable that causes both of them— that variable, that common cause, and its arrows, need to be in the diagram. [This requirement is known as the *Causal Markov Condition*.] We have only two variables in our diagram, so only two to check (i.e.

smoking and lung cancer). But suppose we theorize that there is a gene that causes smoking and, also, that gene causes lung cancer (the gene does, by itself). Then, to be a causal diagram— to satisfy the Causal Markov Condition— our diagram would need to be modified to:



**Figure 41. Smoking Causing Lung Cancer And a Gene Causing Both of These Conditions.**

[In fact, we are not here going to assume that there is such a gene, so the original diagram, with just two variables is a causal diagram as it stands. It satisfies the Causal Markov Condition.]

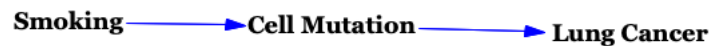
There are three components, or building blocks, or modules, that might occur in a causal diagram: *chains*, *forks*, and *collisions*.

To introduce chains, we need *paths* and paths are sequences of adjacent arrows. If the arrows in a path connect head-to-tail, the path is a *directed* path, otherwise the path is an *undirected* path. A *chain*, or *causal chain*, is a directed path between variables. In the diagram:



**Figure 42. A Causal Chain From Smoking to Lung Cancer.**

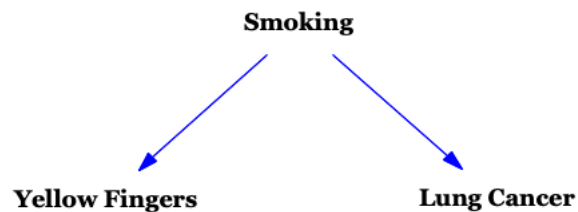
There is just the one chain: **Smoking->Lung Cancer**. More detail can be added between the variables for smoking and the lung cancer. For example, there might be cell mutation and it may be valuable to include an (intermediary) variable for that. A more complete diagram might be:



**Figure 43. A Causal Chain From Smoking to Lung Cancer Mediated by Cell Mutation.**

In this diagram there are three (causal) chains: **Smoking->Cell Mutation**, **Cell Mutation->Lung Cancer**, **Smoking->Cell Mutation->Lung Cancer**. Interest will likely be with causal chain between smoking and cancer, and if indeed smoking causes cancer there will be correlation between the two. There also will be a correlation between smoking and cell mutation and cell mutation and lung cancer.

A *fork* is where there is a common cause of two variables. If we think, as a causal model, that smoking causes both lung cancer and yellow stains on a smoker's fingers, the following diagram might be suitable:

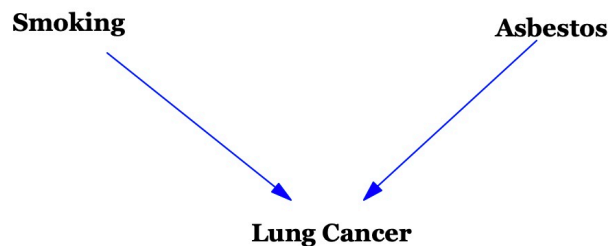


**Figure 44. A Fork from Smoking to Yellow Fingers and Lung Cancer.**

In this, smoking is a common cause of both yellow fingers and lung cancer. This, a common cause, is a *fork*. Forks need care where correlations are concerned. Smoking will be correlated with yellow fingers, smoking will be correlated with lung cancer, and *yellow fingers will be correlated with lung cancer*. There is a little more that can be said. There is the notion of *conditioning* and a simple explanation of conditioning is that it is knowing or fixing the value of a variable. Suppose the smoking variable can have two values only: being a smoker, or not being a smoker. Consider just non-smokers. Some of them will have yellow fingers. Few or none of them will have lung cancer. But now there will be *no* correlation between the yellow fingers and lung cancer. It is the smoking that causes lung cancer, but *none* of the people in the group are smokers. Equally, consider just smokers. Some of them will have yellow fingers. Some of them will have lung cancer. Again, there will be *no* correlation between the yellow fingers and lung cancer. It is the smoking that causes lung cancer, but *all* of the people in the group are smokers. So, if we conditionalize on the smoking variable, there is no correlation between yellow fingers and lung cancer. In sum here, where there is a common cause, a fork, there is correlation (or dependence)

between the effects. But if the analysis conditions on the common cause there is no conditional correlation (or conditional dependence) between the effects.

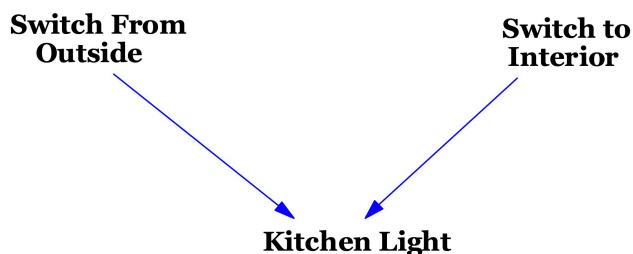
A *collision* is when there is a common effect of two different causes. Smoking is not the only action that causes lung cancer. Exposure to asbestos can cause lung cancer. We might not be especially interested in the asbestos cause. But if we are, we might produce a diagram similar to this:



**Figure 45. A Fork from Smoking and Asbestos to Lung Cancer.**

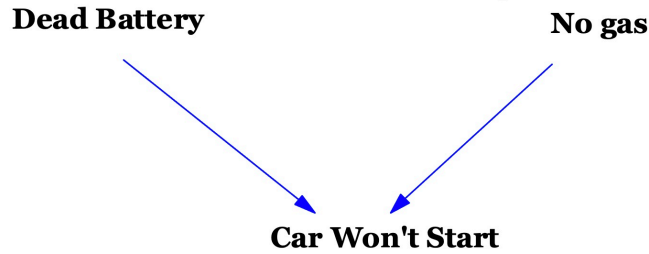
Consider the path **Smoking -> Lung Cancer <- Asbestos**. This is an undirected path, and the path enters and leaves the same variable (here Lung Cancer) by arrow heads. This means that the variable (here Lung Cancer) is a *collider* in this path. Paths that have colliders in them are *closed* or *inactive*. Paths that do not are *open* or *active*. Directed paths tell of causality, of causal chains. So, at the level of causality, we know, or assume, that smoking causes lung cancer and asbestos causes lung cancer. But we also know, or assume, that smoking does not cause exposure to asbestos nor does exposure to asbestos cause smoking. At the level of statistics, we would expect there to be a correlation between smoking and

lung cancer and a correlation between asbestos and lung cancer, but no correlation between smoking and asbestos. Actually, where colliders are concerned again there is a little more to be said. To explain this, it is useful to have different examples of colliders. The first is for a kitchen light. There is a door from outside to the kitchen, and a door from the kitchen to the rest of the interior of the house. Each of these doors has a kitchen light switch near it. These switches turn the single kitchen light on or off:



**Figure 46. A Collider Between Two Switches and a Light.**

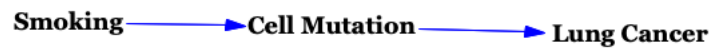
In the path **Switch From Outside** -> **Kitchen Light** <- **Switch To Interior** there is a collider. There is causality, and correlations, between the switches and the light, but not between the switches. Whether one switch is on (or off) is entirely independent of whether the other switch is on (or off). Looking at one of the switches alone will tell you nothing about the other switch. But let us conditionalize on the **Kitchen Light** i.e. permit information as to whether the light is on or off. Now there will be correlation between the switches (e.g. if the light is on and one switch is off, the other switch must be on, etc.). The second example is from Judea Pearl (who is the main modern theorist in this domain). It concerns a car that will not start. This may be because of a dead battery, or it may be because the car has no gas.



**Figure 47. A Collider Between Possible Causes of a Car Not Starting.**

Dead Battery is entirely independent of No gas. But in the path **Dead Battery** -> **Car Won't Start** <- **No gas**, if we conditionalize on the variable **Car Won't Start** then **Dead Battery** becomes conditionally dependent on **No gas**. For example, if you are told that the car won't start (i.e. that the value for the **Car Won't Start** variable is True) then knowing that the battery is good tells you that the car does not have any gas. In sum here, where there is a common effect, a collider, there is no correlation (or dependence) between the causes. But if the analysis conditions on the common effect there is conditional correlation (or conditional dependence) between the causes. This is called *collider bias*. It is worth mentioning because more than a few times in real research publications *it happens by accident*. Julia Rohrer mentions the following example. Say we are interested in whether rigorous research is correlated with innovative research— no causality being looked for here, just association— and we decide, being fancy, to look at this question with published research being considered separately from unpublished research. But rigorous research causes it to be published, and innovative research causes it to be published, so being published is a collider. Our conditioning on the collider may produce association out of thin air— collider bias (Rohrer 2018, 35).

Let us now briefly revisit the ordinary chain structure with one or more link or mediating variables. Say:

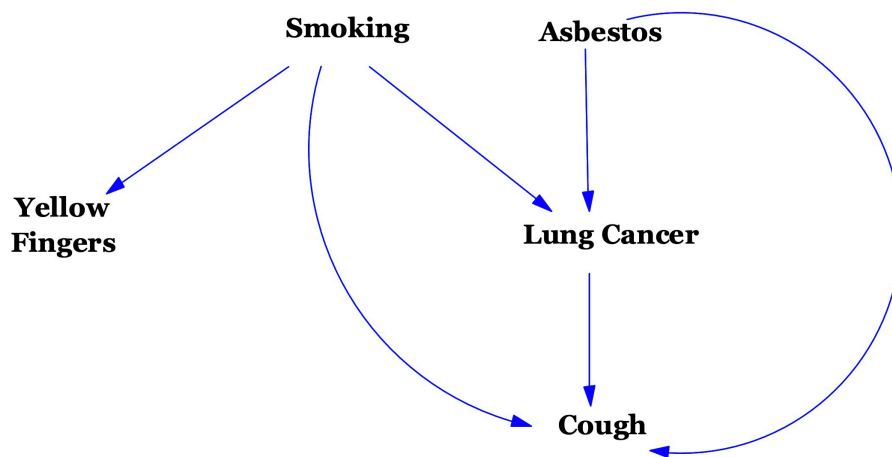


**Figure 48. A Chain From Smoking to Lung Cancer Mediated by Cell Mutation.**

As it stands there will be three dependencies or associations or correlations: **Smoking: Cell Mutation, Cell Mutation: Lung Cancer, Smoking: Lung Cancer**. But consider what happens were we to conditionalize on the link variable, i.e. on Cell Mutation. Knowledge of a value for the cell mutation so-to-speak masks any values for the smoking variable as far as predicting the lung cancer is concerned. Suppose the variable **Cell Mutation** can have just two values: that there is mutation and that there is no mutation. We know in general that smoking is associated with lung cancer. This means that in some cases varying whether there is smoking or not brings about whether there is lung cancer or not. But is smoking associated with lung cancer when there definitely is cell mutation? The answer is 'No'. Varying the smoking has no effect on the cell mutation and it is the cell mutation that is the direct cause of the lung cancer. It is similar for the other case where we fix the link value as being no mutation. In sum here, where there are link variables in a chain, there is correlation (or dependence) between all the upstream variables of a chosen link, and the link, and all the downstream variables of the link. But if the analysis

conditions on the chosen link there is no conditional correlation (or conditional dependence) between all the upstream variables of a chosen link, and the link, and all the downstream variables of the link.

The causal diagrams can be much more complicated than those displayed here. Minimally there are diagrams like this:



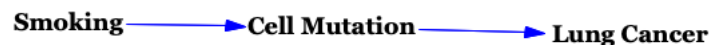
**Figure 49. A Complex Interplay of Causes Around Smoking.**

Lung cancer causes coughs. Asbestos causes lung cancer and, without causing lung cancer, causes coughs. Smoking also causes coughs without causing lung cancer. Smoking causes yellow fingers.

The diagrams use arrows and these give a direction or flow to time and causality— which variables cause which other variables and which variables are 'earlier' than others. This brings another consideration into focus. No variable can cause itself. This means that no directed path in a causal diagram (i.e. a path following the direction of the arrows) can go around in

a circle or 'cycle' and come back to an earlier variable. This means that the diagram, the graph structure as a whole, is a Directed Acyclic Graph (DAG).

Let us review the presentation. Causal diagrams can have two meanings or functions: a data free explanation or description of the causes that are presumed to be at work, and a potentially data rich statistical description of the associations or dependencies or correlations. The causality has a direction (if smoking is the cause of lung cancer, lung cancer is not the cause of smoking). The dependencies do not have directions (if smoking is associated with lung cancer, lung cancer is associated with smoking). Variables can be independent of each other. There is also the notion of conditional dependence and independence. The various structures of the graphs (the chains, forks, and collisions) give rise to the various dependencies. They give rise to complex statistical predictions that can be tested. For example, the very simple causal diagram:



**Figure 50. A Chain From Smoking to Lung Cancer Mediated by Cell Mutation.**

entails that smoking is correlated with lung cancer and not conditionally correlated with lung cancer when cell mutation is controlled for. If either of these statistical prediction are mistaken, so too is the original causal diagram.

There is a technique, in fact an algorithm, *d-separation* ('direction separation') that can convert a causal diagram into all the statistical predictions that it entails.

To sum up this whole presentation in a simple way. If there are forks, there needs to be conditionalization on the common causes. If there are collisions, there needs to be *no* conditionalization on the colliders.

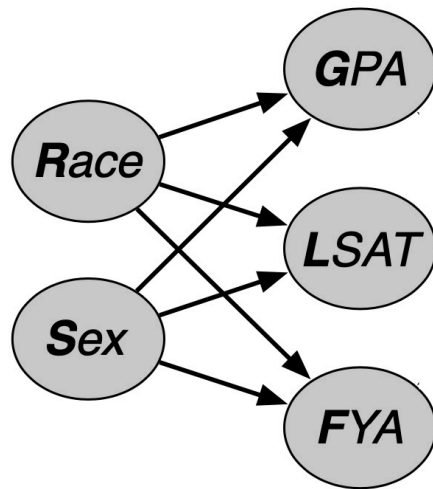
The importance of the causal diagrams is this. Almost all, maybe even all, machine learning is about what causes what— about learning about causality. Machine learning, and all other kinds of empirical science for that matter, never has anything more to work with than correlations. They all have to make the leap from correlation to causation. There is no way this can be done infallibly. There is no way it can be done without assumptions. But causal diagrams provide a framework for making assumptions and for suggesting the appropriate correlations to test.

Let us conclude by sketching two examples of how causal diagrams might be used in connection with bias. Matt Kusner et al. introduce the example of predicting the First Year Average Grade (FYA) of students in law school (Kusner et al. 2018). They write:

The Law School Admission Council conducted a survey across 163 law schools in the United States [35]. It contains information on 21,790 law students such as their entrance exam scores (LSAT), their grade-point average (GPA) collected prior to law school, and their first year average grade (FYA).

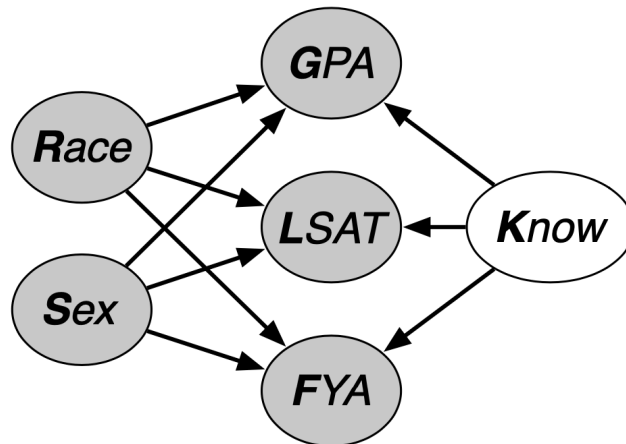
Given this data, a school may wish to predict if an applicant will have a high FYA. The school would also like to make sure these predictions are not biased by an individual's race and sex. However, the LSAT, GPA, and FYA scores, may be biased due to social factors. (Kusner et al. 2018)

A causal diagram for this might be:



**Figure 51. Law School First Year Average Grade (Kusner et al. 2018).**

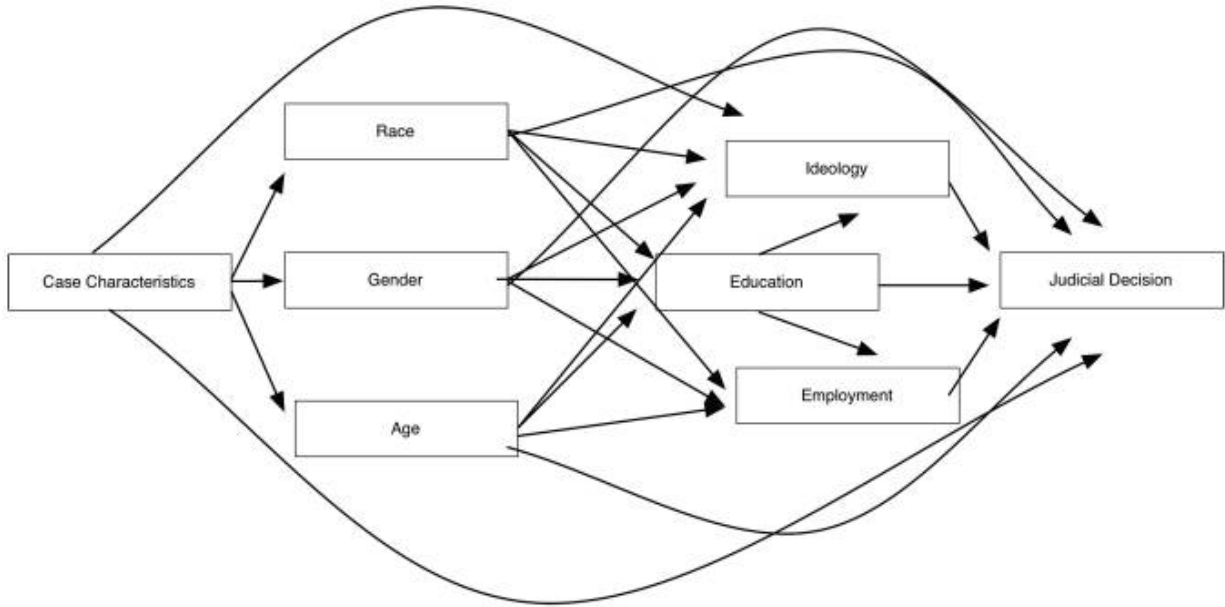
This asserts that Race non-deterministically affects GPA, LSAT, and FYA, as does Sex. In this problem, it is imagined that the Law Schools want to predict first year average grade FYA and they cannot use Race and Sex as these are protected. The next step is to realize, or take into account, that knowledge, a student's knowledge, also affects GPA, LSAT, and FYA, and this gives a second causal diagram.



**Figure 52. Law School First Year Average Grade With Student Knowledge (Kusner et al. 2018).**

With this causal diagram, some mathematics can be done on the forks, collisions, and the data, that can produce the requisite predictions of FYA from LSAT and GPA without using the protected attributes of Race and Sex. The Law Schools would be able to prove the predicted FYA is not biased.

Typically, the diagrams are more complex. This one from Tyler VanderWeele and Nancy Staudt relates legal cases and their case characteristics to both judicial decisions and the likelihood that litigation will take place in courts with judges of a particular race, gender, age or ideology (for plaintiffs prefer to file claims with judges deemed friendly to their legal claims) (VanderWeele and Staudt 2011).



**Figure 53. A Depiction of Possible Factors in a Judicial Decision (VanderWeele and Staudt 2011).**

In sum. If there is an assertion of bias, bias in allocation, say, perhaps with allocation of mortgages or allocation of places in a Law School. Suitable causal diagrams might be able to highlight relevant correlations to be investigated. In turn, these might be able to produce evidence, within the bounds of fallibility, that there is no bias. The diagrams provide a structure for investigation false positives and false negatives. Separately, explainable artificial intelligence (XAI) is important. Explanations need causes. Causal diagrams are a steppingstone, along with data and correlations, towards identifying causes.

## D.3 Annotated Readings for Appendix D

Hernan, Miguel. “Causal Diagrams: Draw Your Assumptions Before Your Conclusions.” edX, 2022. <https://www.edx.org/course/causal-diagrams-draw-your-assumptions-before-your>. (Hernan 2022)

Myint, Leslie. “Key Structures in Causal Graphs”, 2020. <https://www.youtube.com/watch?v=UAovyBnzi9U>. (Myint 2020) Leslie Myint has posted many excellent videos on statistics and causality. This is a relevant example.

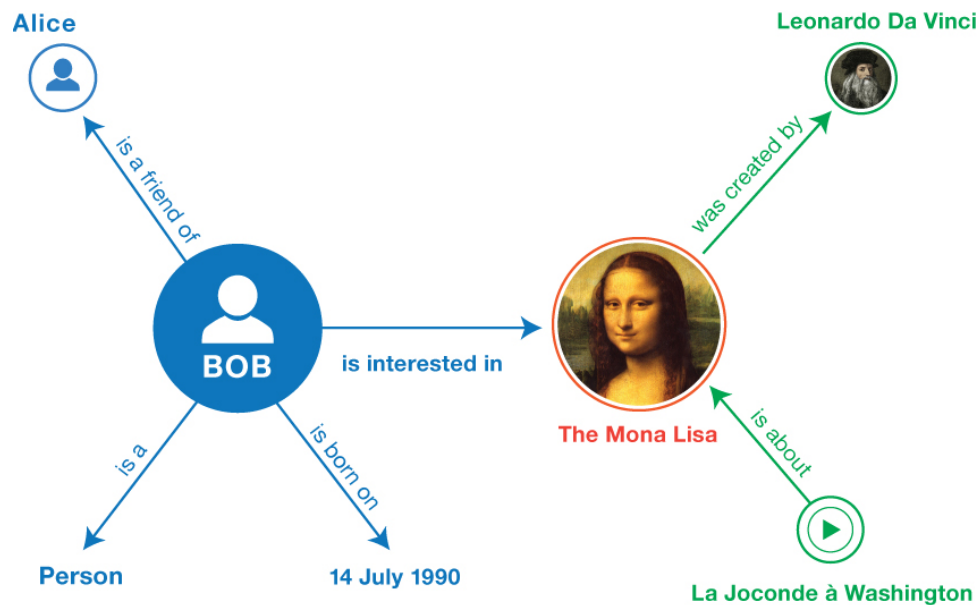
Rohrer, Julia M. “Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data.” *Advances in Methods and Practices in Psychological Science* 1, no. 1 (2018): 27–42. <https://doi.org/10.1177/2515245917745629>. (Rohrer 2018). This is a masterpiece. It is probably a little advanced for us.

# Appendix E: Knowledge Graphs

## E.1 Knowledge Graphs

A knowledge graph is a means of representing knowledge. Its relevance to us is that it is a technology or technique that can help with the search and discovery of information. It is not in itself an artificial intelligence or machine learning technique. However, it can be part of an infrastructure that supports some applications of machine learning in librarianship.

Conceptually, a knowledge graph starts with a node representing an object, which might be a person, a place, or a thing, or a date, or etc., then adds as links facts about that object. Here is a very simple knowledge graph:



**Figure 54. A Simple Knowledge Graph (W3C Working Group 2014).**

Typical knowledge graphs will be much more complicated than this. Google have a knowledge graph technology, 'Google Knowledge Graph'. As of 2012 this had 500 million objects and 3.4 billion facts in it (Singhal 2012). Google populate their Knowledge Graph using search data that they have from users. For example, if the search 'who painted the Mona Lisa?' was popular, the answer to that search would be added to the knowledge graph as a fact.

Usually, a knowledge graph is a data structure that computer programs can use in the background (as opposed to the user exploring it visually). No user is going to look through a graph of 500 million objects.

Some possibilities for knowledge graphs are:

1. Helping disambiguate. For example, were there to be several 'Bob' objects in the graph, the user or system would know that disambiguation was required for queries about 'Bob' and maybe suggest facts to do it.
2. Helping with creating summaries of information about an object. For example, it could help identify key facts (say about the Mona Lisa).
3. Pointing the user to further popular (or rare) queries or information about the objects.
4. Being part of a question answering system (for example, who painted Mona Lisa?)

5. Being part of a recommender system (for example, by extracting the preferences of similar people in the graph).

## **E.2 Annotated Readings for Appendix E**

Peng, Ciyuan, Feng Xia, Mehdi Naseriparsa, and Francesco Osborne. “Knowledge Graphs: Opportunities and Challenges.” *Artificial Intelligence Review* 56, no. 11 (2023): 13071–102. <https://doi.org/10.1007/s10462-023-10465-9>.(Peng et al. 2023).

# Glossary

There are many excellent ML glossaries available online, for example Google's *Machine Learning Glossary: ML Fundamentals* (Google for Developers 2023) (We are using and editing parts of that here.) There are also 'explanations of key concepts', approaching them from a librarianship point of view, for example, Brady Lund and Ting Wang's *Chatting about ChatGPT: how may AI and GPT impact academia and libraries?* (Lund and Wang 2023).

Generally, we are restricting ourselves to technical or unusual terms that appear in this text.

## A

### accuracy

The number of correct classification predictions divided by the total number of predictions.

### application programming interfaces (APIs)

the protocols that allow outside programs to use, or communicate with, an application.

### algorithm

1. a step-by-step computational procedure.
2. any piece of computer software, especially software that can make decisions independent of a human decision maker.

## alignment

that the model's predictions or behavior correspond closely with the expected or desired or intended outcome.

## anaphora

the use of pronouns or other words to refer back to previously mentioned subjects or objects.

## artificial intelligence (AI)

the use of computers, algorithms, and sometimes outside data, to solve problems that an ideally rational and intelligent human being would be able to solve, given the time, resources, and ingenuity.

## artificial general intelligence (AGI)

a non-human mechanism that demonstrates a *broad range* of problem solving, creativity, and adaptability.

## artificial super intelligence (ASI)

is the possibility that AGI becomes established and then the AGI machines simply design themselves and become smarter and smarter in a runaway fashion.

## attention

a technique that enables models to dynamically focus on certain parts of the input for better performance, especially in sequence-to-sequence tasks.

## B

### bag of words

a representation of the words in a phrase or passage, irrespective of order.

### bias (ethics/fairness)

1. unfairness by means of stereotyping, prejudice or favoritism towards some things, people, or groups over others.
2. errors in input data or output predictions (independent of questions of fairness or unfairness).

## C

### cause (or causal factor)

a) X and Y occur, b) X precedes Y in time, c) the presence of X raises the probability of Y, and d) had there been an intervention that changed X, that would change the probability of Y. For example, the diagnosis that smoking caused lung cancer in a group of patients amount to a) the patients smoked and have lung cancer, b) the patients smoked before they got lung cancer, c) their smoking raised their probability of getting lung cancer, and d) had they not smoked in the first place, their probability of getting lung cancer would have been lower.

## chain-of-thought (COT) prompting

encouraging a large language model (LLM) to explain its reasoning, step by step. For example, consider the following prompt, paying particular attention to the second sentence:

How many g forces would a driver experience in a car that goes from 0 to 60 miles per hour in 7 seconds? *In the answer, show all relevant calculations.*

Chain-of-thought prompting forces the LLM to perform all the calculations, which might increase the chances of the answer being correct. In addition, chain-of-thought prompting enables the user to examine the LLM's steps to determine whether the answer makes sense.

## chatbot

a computer program designed to simulate conversation with human users, especially over the internet.

## class

a labeled collection of items. For example:

1. In a model that detects spam, the two classes might be spam and not-spam.
2. In a model that identifies dog breeds, some of the classes might be poodle, beagle, pug.

## classification model

predicts a class. For example, the following are all classification models:

1. A model that predicts an input sentence's language (French? Spanish? Italian?).
2. A model that predicts tree species (Maple? Oak? Baobab?).
3. A model that predicts the positive or negative class for a particular medical condition.

## cloze task

a method used in language teaching and as a machine learning task. Such tasks are often referred to as a fill-in-the-blanks task, cloze test, gap-filling task, or text completion task. It involves having a model fill in missing words or tokens in a sentence or paragraph.

## clustering

grouping related examples, particularly during unsupervised learning. Once all the examples are grouped, a human can optionally supply a label or meaning to each cluster for example that cluster 1 is 'dwarf trees' and cluster 2 is 'full-size' trees.

## conversational implicature

the meaning that is implied by a speaker but not explicitly stated in the conversational context.

## conjectural

speculative, hypothetical, theoretical, presumptive, or fallible. There is a philosophical view, highlighted by Karl Popper, that all knowledge of the world, and some knowledge of mathematics and logic, are conjectural.

## confirmation bias

the tendency to search for, interpret, favor, and recall information in a way that confirms one's preexisting beliefs or hypotheses.

## controlled vocabulary

a standardized terminology, curated term set, or fixed lexicon. It is a predefined set of terms that are used to ensure consistency in the tagging and categorization of content.

## constraint prompting

a method to direct a model's generation or response by setting specific constraints or conditions or by providing a template that should be used for the answer.

## counterfactual

Let us start with an example. Napoleon lost the Battle of Waterloo, as a matter of fact. But an adventurous historian might consider the question of what would have happened if Napoleon had won the Battle of Waterloo. Supposing that Napoleon won would be counter-to-the-facts and the historian's analysis would be a counterfactual analysis.

Thinking counterfactually, which we do all the time, especially when making decisions, involves considering alternative scenarios and outcomes that did not happen or might not have happened but which could have happened under different circumstances. (If I'd run faster, I would have caught the bus. If I'd studied harder, I would have passed the exam. If inflation had continued, there would have been a fall in unemployment. etc.) Some counterfactuals are true, and some are false (for example, if the aforementioned bus were travelling at 50mph no amount of faster running on my part would result in my catching it). Counterfactual reasoning and causal reasoning are often intertwined. Were we to say 'John Smith's lung cancer was caused by smoking' we in part mean 'If John Smith had not smoked, he would not have got lung cancer' (which is counterfactual). Then, were we to say 'If Jane Smith, a non-smoker, were to become a smoker, she would raise the probability of her getting lung cancer', that statement is a true counterfactual supported by a causal connection between smoking and lung cancer. Overall, counterfactuals are a valuable tool in many fields for analyzing and understanding the implications of events and decisions by considering alternatives that did not occur.

## CPU, GPU, and CUDA®

a CPU is a central processing unit, the main computing component of a standard computer. A GPU is a graphics processing unit, a computing component originally designed to provide accelerated graphics for video games and similar. Typically, a CPU works serially, one task after another, whereas a GPU works in parallel, carrying out many different tasks at the same time. The company NVIDIA realized that GPUs, which they specialized in, were ideal for artificial intelligence. They produced the

CUDA® platform, which is for high performance, high throughput, computing (not necessarily having anything to do with graphics). As of 2024, most LLM research and commercial work will be using the CUDA® platform and NVIDIA electronic chips.

## D

### data set or dataset

raw data, commonly organized in one of the following formats:

1. a spreadsheet
2. a file in CSV (comma-separated values) format

### deep learning (DL)

advanced techniques within the field of artificial intelligence focused on mimicking the operation of the human brain.

### deep model

neural networks containing more than one hidden layer.

### deepfakes

media that have been altered, or wholly generated, by artificial intelligence to present something that did not actually occur. Examples include manipulated videos which might include realistic speeches from politicians that simply never happened.

## delayed rewards

in many settings there are rewards (such as passing the exam, arriving at the destination safely, finding the cheese in the maze.) Delayed rewards are where the agent has to do some exploring of the task or making more than a few steps presumably towards the goal, before a reward is given. The above examples have delayed rewards— pressing a button to get a pellet of food does not.

## digitization

the conversion of text, pictures, or sound into a digital form that can be processed by a computer.

## discriminative model

predicts the appropriate or correct labels for new examples presented to it. For example, a discriminative model might predict whether incoming email was spam. It discriminates between spam and not-spam. In contrast, generative models might produce or create completely new examples such as new images or paintings. Most supervised learning models are discriminative models.

## dynamic

The terms dynamic and online are synonyms in machine learning. The following are common uses of dynamic and online in machine learning:

1. A dynamic model (or online model) is a model that is retrained frequently or continuously. A dynamic model is a ‘lifelong learner’ that constantly adapts to evolving data.
2. Dynamic training (or online training) is the process of training frequently or continuously.
3. Dynamic inference (or online inference) is the process of generating predictions on demand.

## E

### embeddings

numerical representations of text, concepts, or other types of data as lists of numbers (i.e. they are points or vectors in a high-dimensional space).

### empirical

knowledge or models that are based on observation or experience rather than on pure mathematics or pure logic. Empirical knowledge is tested by means of experiments or observations.

### epoch

a full training pass over the entire training set such that each example has been processed once.

### example

The values of one row of features and possibly a label. Examples in supervised learning fall into two general categories:

1. A labeled example consists of one or more features and a label. Labeled examples are used during training.
2. An unlabeled example consists of one or more features but no label. Unlabeled examples are used during inference and use.

For instance, a model is being trained to determine the influence of weather conditions on student test scores. Here are three labeled examples:

Features			Label
Temperature	Humidity	Pressure	Test score
15	47	998	Good
19	34	1020	Excellent
18	92	1012	Poor

**Figure 90. Labeled Examples.**

Here are three unlabeled examples:

Features		
Temperature	Humidity	Pressure
12	62	1014
21	47	1017
19	41	1021

**Figure 91. Unlabeled Examples.**

In supervised machine learning, models train on labeled examples and make predictions on unlabeled examples.

In semi-supervised and unsupervised learning, unlabeled examples are used during training.

## expert system

a computer system that emulates the decision-making ability of a human expert.

## F

### fakes

in machine learning, fakes are the products of algorithms designed to generate deceptive content that mimics the real one.

### fallible

not absolutely certain. Many views, theories, and observation reports are true. Nevertheless, often it is not known with absolute certainty whether they are true. Our knowledge of them is fallible.

### false negative

an example in which the model mistakenly predicts a member of the negative class. For example, the model predicts that a particular email message is not spam (the negative class), but that email message actually is spam.

## false positive

an example in which the model mistakenly predicts a member of the positive class. For example, the model predicts that a particular email message is spam (the positive class), but that email message is actually not spam.

## feature

An input variable to a machine learning model. An example consists of one or more features. For instance, suppose a model is being trained to determine the influence of weather conditions on student test scores. The following table shows three examples, each of which contains three features and one label:

Features			Label
Temperature	Humidity	Pressure	Test score
15	47	998	Good
19	34	1020	Excellent
18	92	1012	Poor

**Figure 92. Features With Labels.**

## few-shot learner

a machine learning approach where the, perhaps pre-trained, model is designed to learn information with a small amount of, perhaps additional, training data.

## few-shot prompting

prompting where a few examples are given of what the LLM should do.

## fine tuning

taking a pre-trained model and adjusting its parameters slightly to adapt to a new but related task.

## foundation models

starting points for creating more specialized machine learning models.

## frame

a frame, within machine learning and artificial intelligence, is a data structure for representing a stereotypical situation, like a room and its contents, ordering a meal in a restaurant, or buying an airline ticket.

## G

### generative

the capability of some models to generate completely new data instances that resemble the training data— for example new images or paintings, or new poems or essays.

## genial understander system

a natural language understanding system which can interact with users in a friendly or amiable manner to find out what they want and any requisite data to assist in that task (e.g. to discover data for a frame or frames).

## generalization

making correct predictions on new, previously unseen data. A model that can generalize is the opposite of a model that is overfitting.

## GPUs, CPUs, and CUDA®

See CPUs, GPUs, and CUDA®.

## gradient descent

a mathematical technique used in training to minimize loss or error. Gradient descent iteratively adjusts parameters of the model, gradually finding the best combination to minimize loss.

## ground truth

Reality. For example, consider a model that predicts whether a student in their first year of university will graduate within six years. Ground truth for this model is whether that student actually graduated within six years.

## H

### hallucinations

false generations, spurious outputs, or fictitious predictions. It is when a model generates information that is ungrounded or not supported by the input data.

### homograph

a word characterized by lexical ambiguity, as it shares the same spelling as another word but has a different meaning for example 'bank' (of river) and 'bank' (financial).

## I

### inductive bias

training data, no matter how extensive, is never enough on its own to determine predictions about new unseen data. Assumptions are needed. Those assumptions are inductive bias.

### inference

the process of making predictions by applying a trained model to unlabeled examples.

### interpretability

explaining or to presenting a model's reasoning in terms understandable to a human.

## intersubjective (shared conventional knowledge)

conventional agreements shared by multiple individuals or systems. For example, a library classification system such as the Dewey Decimal System is a shared conventional agreement.

## K

### keyword search

retrieving information by matching query terms ('keywords') against a set of documents or a database.

## L

### label

in supervised machine learning, the 'answer' or 'result' portion of an example. Each labeled example consists of one or more features and a label. For instance, in a spam detection dataset, the label would probably be either 'spam' or 'not spam.' In a rainfall dataset, the label might be the amount of rain that fell during a certain period.

### labeled example

See example.

## language model

a statistical model, a generative model, or a predictive model of language. It is a type of model that can predict the next word in a sentence or help generate text based on previous text.

## large language models (LLMs)

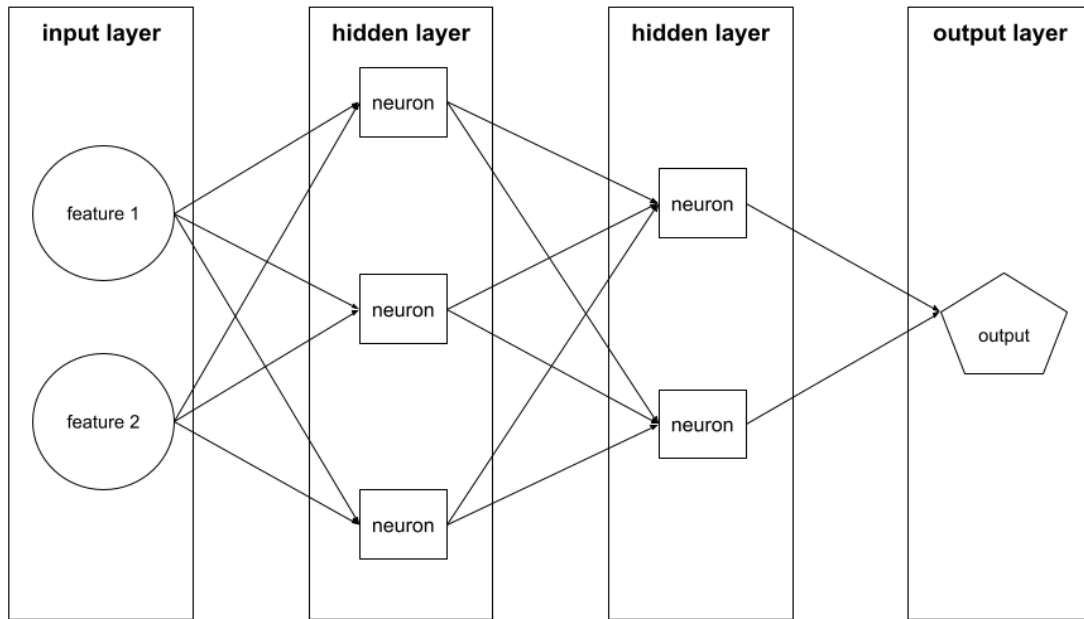
usually generative pre-trained transformers. These are highly complex models designed to understand, generate, and translate human language.

## layer

A set of neurons in a neural network. Three common types of layers are as follows:

1. The input layer, which provides values for all the features.
2. One or more hidden layers, which find relationships between the features and the label.
3. The output layer, which provides the prediction.
- 4.

For example, the following illustration shows a neural network with one input layer, two hidden layers, and one output layer:



**Figure 93. A Neural Network.**

M

## machine learning (ML)

a program or system that trains a model from input data. The trained model can make useful predictions from new (never-before-seen) data drawn from the same distribution as the one used to train the model.

## Markov (process or chain)

a mathematical system that undergoes transitions from one state to another, with probabilistic rules that depend only on the current state and not on the sequence of events that preceded it. It is a memoryless model. It ‘knows’ the state that it is in, but ‘knows’ nothing of the states before that. The child’s game Snakes and Ladders is an example of a Markov process.

## modality

a high-level data category. For example, numbers, text, images, video, and audio are five different modalities.

## model

any mathematical construct that processes input data and returns output. Phrased differently, a model is the set of parameters and structure needed for a system to make predictions. In supervised machine learning, a model is trained on labeled examples, then, in use, it takes an unlabeled example as input and infers a prediction of the correct label as output. Unsupervised machine learning also generates models, typically a function that can map an input example to the most appropriate cluster for it.

## N

### negative class

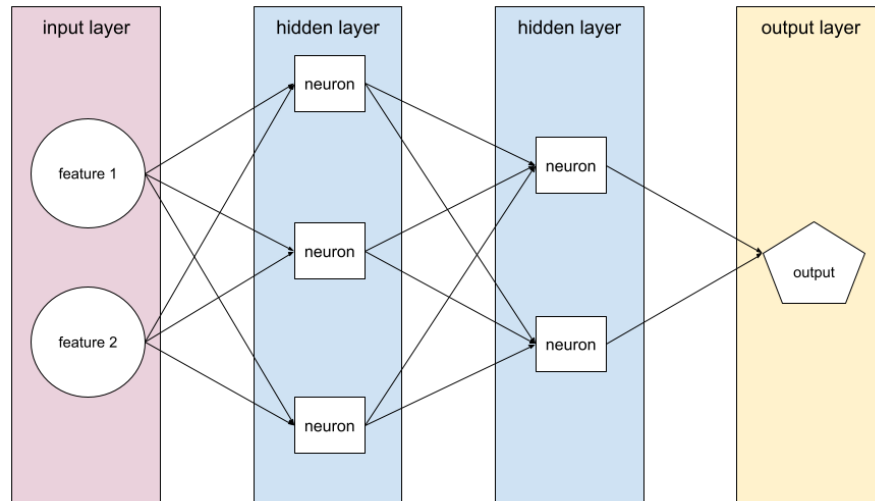
in binary classification, one class is termed positive and the other is termed negative. The positive class is the thing or event that the model is testing for and the negative class is the other possibility. For example:

1. The negative class in a medical test might be ‘not tumor.’
2. The negative class in an email classifier might be ‘not spam.’

### neural network

a model containing at least one hidden layer. A deep neural network is a type of neural network containing more than one hidden layer. For

example, the following diagram shows a deep neural network containing two hidden layers.



**Figure 94. A Neural Network Showing Hidden Layers.**

Each neuron in a neural network connects to all of the nodes in the next layer. For example, in the preceding diagram, notice that each of the three neurons in the first hidden layer separately connect to both of the two neurons in the second hidden layer.

Neural networks implemented on computers are sometimes called artificial neural networks to differentiate them from neural networks found in brains and other biological nervous systems.

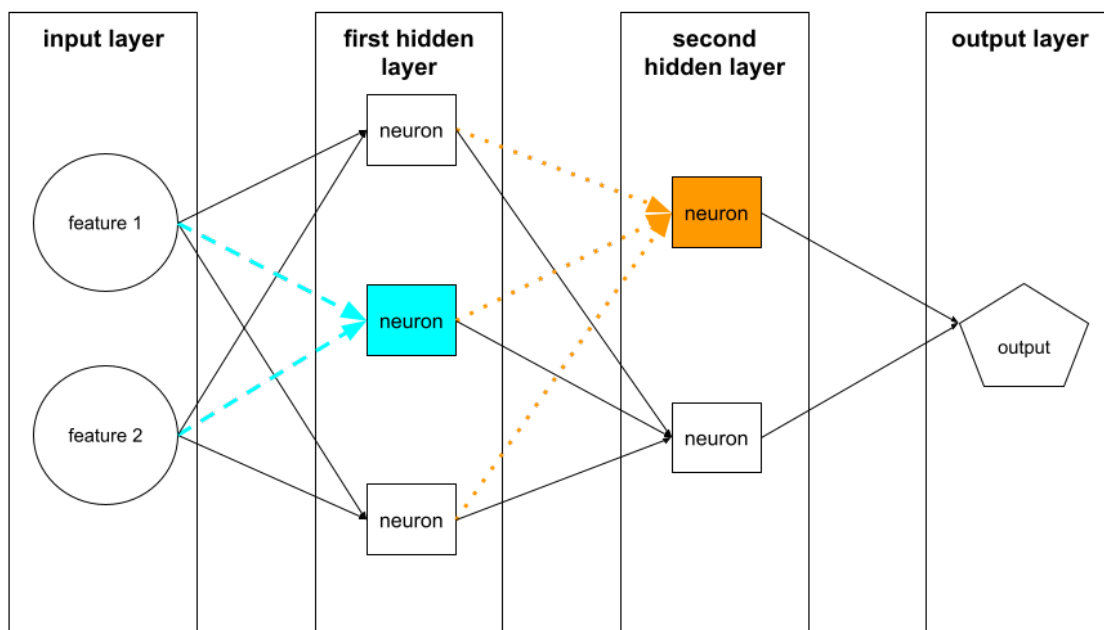
Some neural networks can mimic extremely complex relationships between different features and the label.

## neuron

In machine learning, a neuron is a distinct unit within a hidden layer of a neural network. Each neuron performs a calculation on its own inputs and decides whether it is activated i.e. whether it fires or is 'triggered'.

A neuron in the first hidden layer accepts inputs from the feature values in the input layer. A neuron in any hidden layer beyond the first accepts inputs from the neurons in the preceding hidden layer. For example, a neuron in the second hidden layer accepts inputs from the neurons in the first hidden layer.

The following illustration highlights two neurons and their inputs.



**Figure 96. A Neural Network Showing Some Neurons.**

A neuron in a neural network mimics the behavior of neurons in brains and other parts of nervous systems.

## node (neural network)

A neuron in a hidden layer.

## O

### objective knowledge

factual knowledge and the knowledge of science and mathematics.

### offline

synonym for static.

### one-shot prompting

prompting where a single example is given.

### online

synonym for dynamic.

### ontology

a conceptual framework, a knowledge representation schema, a semantic framework, or a taxonomy, particularly if it is structured hierarchically.

### online inference

generating predictions on demand. For example, suppose an app passes input to a model and issues a request for a prediction. A system using

online inference responds to the request by running the model (and returning the prediction to the app).

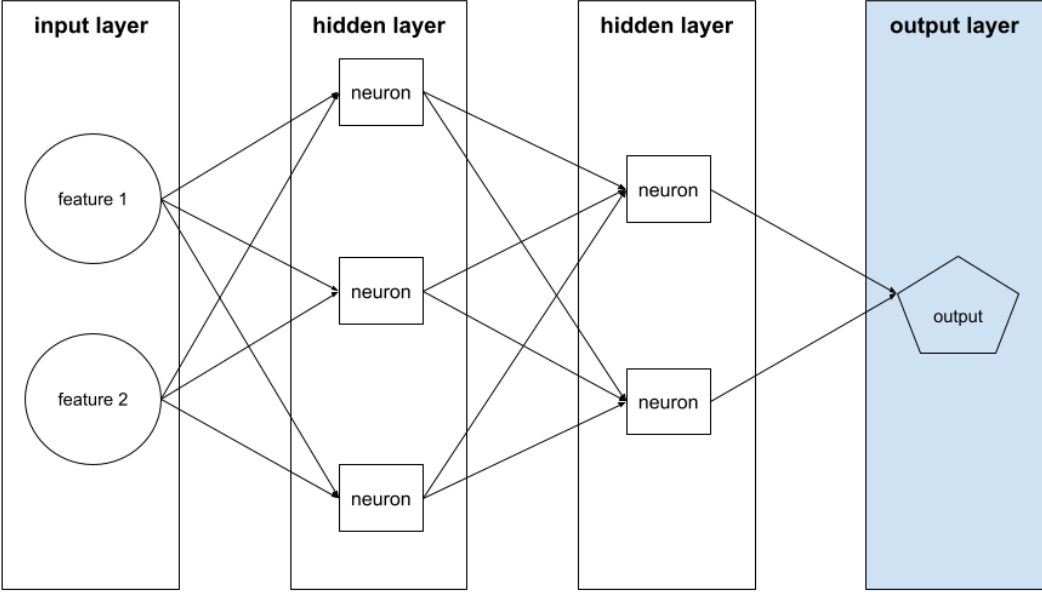
## optical character recognition (OCR)

a technology for text recognition, character reader technology, document digitization, and image-to-text conversion. It converts different types of documents, such as scanned paper documents, PDF files, or images captured by a digital camera, into editable and searchable data.

## output layer

The ‘final’ layer of a neural network. The output layer contains the prediction.

The following illustration shows a small deep neural network with an input layer, two hidden layers, and an output layer:



## **Figure 96. A Neural Network Showing An Output Layer.**

### overfitting

creating a model that matches the training data so closely that the model fails to make correct predictions on new data that is not part of the training.

## P

### panopticon

a prison where all actions of the prisoners are observable by the controlling entity.

### parameter

internal variables that can be adjusted during training to alter the model's behavior. Suitable, or the best, values for the parameters are learned during training.

### paternalism

to act for the good of another person without that person's consent, as parents do for children (Suber 1999).

### positive class

the class that is being testing for. For example, the positive class in a cancer model might be 'tumor.' The positive class in an email classifier might be 'spam.'

## post-processing

adjusting the output of a model after the model has been run. Post-processing can be used to enforce fairness constraints without modifying models themselves. For example, one might apply post-processing to a binary classifier by setting a classification threshold such that equality of opportunity is maintained for some attribute by checking that the true positive rate is the same for all values of that attribute. As an example of this, in a mortgage application program, post-processing might check that those living north of the railway tracks get as many mortgages as those living south of the railway tracks.

## precision

Consider a search of a library collection for items that are relevant to that search. Precision is the proportion, or percentage, *of the returned items* that are relevant. So, if everything returned is relevant, precision is 100%. If half the items returned are relevant, precision is 50%. Precision has a companion property ‘recall’. Recall is the proportion, or percentage, *of the relevant items in the collection* that are returned. If all the relevant items are returned, recall is 100%. If half the relevant items in the collection are returned, recall is 50%.

Machine learning uses similar concepts of precision and recall. Precision is the proportion of true positives among the positives in a classification task. If every positive is a true positive, precision is 100%. If half the positives are true positives (and half are false positives) precision is 50%. If there are no false negatives (i.e. all the positives are captured), recall is 100%. If half of

the positives get classified as negatives (i.e. half of the positives get misclassified as (false) negatives), recall is 50%.

## prediction

a model's output. For example:

1. The prediction of a binary classification model is either the positive class or the negative class.
2. The prediction of a multi-class classification model is one class.
3. The prediction of some models can be a number.

## Q

### quasi-empirical

investigative approaches or knowledge that are similar to those of empirical research. That is, they might use observation and experiment in discovery and testing. For example, in mainstream computer science, determining what typical algorithms will do when run is a matter of mathematical proof. There is no need to run the programs or observe them running. Such knowledge is non-empirical. But, in machine learning, often finding out what the software will do is a matter of trying it, observing its behavior, and even conducting experiments. In such areas, the knowledge is quasi-empirical.

## R

### rater

a human who provides labels for examples. 'Annotator' is another name for rater.

## recall

Consider a search of a library collection for items that are relevant to that search. Recall is the proportion, or percentage, *of the relevant items in the collection* that are returned. If all the relevant items are returned, recall is 100%. If half the relevant items in the collection are returned, recall is 50%. Recall has a companion property 'precision'. Precision is the proportion, or percentage, *of the returned items* that are relevant. So, if everything returned is relevant, precision is 100%. If half the items returned are relevant, precision is 50%.

Machine learning uses similar concepts of recall and precision. If there are no false negatives (i.e. all the positives are captured) in a classification task, recall is 100%. If half of the positives get classified as negatives (i.e. half of the positives get mis-classified as (false) negatives), recall is 50%. Precision is the proportion of true positives among the positives. If every positive is a true positive, precision is 100%. If half the positives are true positives (and half are false positives) precision is 50%.

## recommender system

a system that predicts and provides suggestions tailored to the user's preferences.

## reinforcement learning

a learning paradigm that trains algorithms based on a system of rewards and penalties.

## reliability

the degree to which a measurement, prediction, or algorithm yields stable consistent results over multiple runs or data sets. A bathroom scale that always weighs 5 pounds light is reliable (but it is not accurate and nor are its results valid).

## retrieval-augmented generation (RAG)

is a technique to both to keep an LLM up to date with what it ‘knows’ and to be more accurate in its replies. The idea is to give the LLM access to an external database or databases. Then factual prompt questions to the LLM are augmented with the instruction to check with the databases and find supporting facts, references, and citations. As external knowledge grows there is no need to re-train the LLM. Rather, all that is required is for the databases to be updated (which they usually would be as a matter of course, say for news articles).

## S

### self-supervised

a machine learning technique that involves algorithms that can learn to label, classify, or predict new instances without explicit human-provided

labels. Once the labels exist, the technique can use standard supervised learning.

## semantic search

meaning-based search, conceptual search, or context-aware retrieval. It uses search algorithms that understand the searcher's intent and the contextual meaning of terms to fetch more relevant results.

## sentiment analysis

computationally determining and categorizing opinions expressed in a piece of text, especially to determine the writer's attitude towards a particular topic.

## specification (for a computer program)

the program design, software requirements, or system design. It outlines the expected functions, behaviors, and structures of the computer program.

## static

done once rather than continuously. The terms static and offline are synonyms. The following are common uses of static and offline in machine learning:

1. static model (or offline model) is a model trained once and then used for a while.
2. static training (or offline training) is the process of training a static model.

3. static inference (or offline inference) is a process in which a model generates a batch of predictions at a time.

## static inference

synonym for offline inference.

## stochastic (processes)

random, non-deterministic, or probabilistic processes.

## stochastic psittacosis

[this is a joke.] Emily Bender describes large language models as being stochastic parrots. Psittacosis is a disease that parrots can have. Hence stochastic psittacosis captures the shortcomings of large language models.

## supervised machine learning

training a model from features and their corresponding labels. Supervised machine learning is analogous to learning a subject by studying a set of questions and their corresponding answers. After mastering the mapping between questions and answers, a student can then provide answers to new (never-before-seen) questions on the same topic.

## switch transformer

a type of transformer model designed to handle extremely large-scale datasets and models by activating only a portion of the model at a time.

## synonym (pair)

a pair of equivalent terms, lexical equivalents, alternative terms, or interchangeable words. These are words or phrases that have the same or nearly the same meanings and can be used to provide variety in text or speech, for example 'attorney' and 'lawyer'.

## T

### theory-laden

observations, especially those using instruments, can be said to be theory-laden and this means that they are influenced by underlying theoretical frameworks. For example, measuring the temperature of a medical patient using an ordinary glass thermometer takes for granted background theories about the expansion of mercury and glass. That observations are theory-laden does not mean that any or all of them are false or incorrect. It does, however, mean that they are fallible (i.e. caution is needed, they might be mistaken).

### token

is used in two different ways. Were we to ask, how many letters are there printed below?

t t t

The answer 'one' would be talking of the letter tee as a *type*. Whereas the answer 'three' is talking of the *tokens* of the letter tee. The second sense of

‘token’ concerns small processing units. For example, LLMs have a context window of text. That text consists of tokens, the chunks of text that the LLM looks at. Typically, tokens are larger than individually characters but smaller than whole words. So, a context window of 1000 tokens might amount to a context window of 1000 words.

## training

the process of determining the ideal parameters comprising a model. During training, a system reads in examples and gradually adjusts parameters. Training uses each example anywhere from a few times to billions of times.

## training set

The subset of the dataset used to train a model.

Traditionally, examples in the dataset are divided into the following three distinct subsets:

1. a training set
2. a validation set
3. a test set

Ideally, each example in the dataset should belong to only one of the preceding subsets. For example, a single example should not belong to both the training set and the validation set.

## transformers

a type of model in machine learning known as self-attention models, transformer architectures, sequence-to-sequence models, or attention-based models. These are particularly powerful in handling sequences of data, such as natural language, for tasks like translation or summarization.

## true negative

an example correctly predicted by the model as belonging to the negative class. For example, the model infers that a particular email message is not spam, and that email message really is not spam.

## true positive

an example correctly predicted by the model as belonging to the positive class. For example, the model infers that a particular email message is spam, and that email message really is spam.

## U

## underfitting

a model with poor predictive ability caused by the model not having fully captured the complexity of the training data. Many problems can cause underfitting, including:

1. Training on the wrong set of features.
2. Training for too few epochs.
3. Providing too few hidden layers in a deep neural network.

## unfairness

synonym for one meaning of bias.

## unlabeled example

see example.

## unsupervised machine learning

training a model to find patterns in a dataset, typically an unlabeled dataset. The most common use of unsupervised machine learning is to cluster data into groups of similar examples. For example, an unsupervised machine learning algorithm can cluster songs based on various properties of the music. The resulting clusters can become an input to other machine learning algorithms (for example, to a music recommendation service). Clustering can help when useful labels are scarce or absent. For example, in domains such as anti-abuse and fraud, clusters can help humans better understand the data.

## V

### validation

the initial evaluation of a model's quality. Validation checks the quality of a model's predictions against the validation set. Because the validation set differs from the training set, validation helps guard against overfitting. Evaluating the model against the validation set can be thought of as the

first round of testing and evaluating the model against the test set as the second round of testing.

## validation set

the subset of the dataset that performs initial evaluation against a trained model. Typically, you evaluate the trained model against the validation set several times before evaluating the model against the test set.

Traditionally, you divide the examples in the dataset into the following three distinct subsets:

1. a training set
2. a validation set
3. a test set

Ideally, each example in the dataset should belong to only one of the preceding subsets. For example, a single example should not belong to both the training set and the validation set.

## validity

accuracy, correctness, or soundness. It concerns the extent to which a model or method accurately measures or predicts what it is intended to.

## Bibliography

- A12 Allen Institute for AI. 2022. "About — ARC: AI2 Reasoning." 2022.  
<https://leaderboard.allenai.org/arc/submissions/about>.
- Abebe, Rediet, Moritz Hardt, Angela Jin, John Miller, Ludwig Schmidt, and Rebecca Wexler. 2022. "Adversarial Scrutiny of Evidentiary Statistical Software." In *2022 ACM Conference on Fairness, Accountability, and Transparency*, 1733–46. FAccT '22. New York, NY, USA: Association for Computing Machinery.  
<https://doi.org/10.1145/3531146.3533228>.
- Abid, Abubakar, Maheen Farooqi, and James Zou. 2021. "Persistent Anti-Muslim Bias in Large Language Models." arXiv. <https://doi.org/10.48550/arXiv.2101.05783>.
- Acemoglu, Daron. 2024. "Get Ready for the Great AI Disappointment." *Wired*, 2024.  
<https://www.wired.com/story/get-ready-for-the-great-ai-disappointment/>.
- Adler, Melissa. 2017. *Cruising the Library*. Fordham University Press.  
<https://doi.org/10.2307/j.ctt1xhr79m>.
- Ager, Simon. 2023. "Omniglot - the Online Encyclopedia of Writing Systems and Languages." 2023. <https://omniglot.com/>.
- AI Advantage, Igor. 2024. "The AI Advantage." YouTube. 2024.  
<https://www.youtube.com/channel/UCHhYXsLBEVVnbvsq57n1MTQ>.
- AICommunity. 2024. "Community - AI Advantage." 2024.  
<https://myaiadvantage.com/community>,  
<https://myaiadvantage.com/community>.
- Akter, Syeda Nahida, Zichun Yu, Aashiq Muhamed, Tianyue Ou, Alex Bäuerle, Ángel Alexander Cabrera, Krish Dholakia, Chenyan Xiong, and Graham Neubig. 2023. "An In-Depth Look at Gemini's Language Abilities." arXiv.  
<http://arxiv.org/abs/2312.11444>.
- Akyürek, Ekin, Dale Schuurmans, Jacob Andreas, Tengyu Ma, and Denny Zhou. 2022. "What Learning Algorithm Is In-Context Learning? Investigations with Linear Models." arXiv. <http://arxiv.org/abs/2211.15661>.
- Al Badi, Waleed, Laurie Alvandian, Anna Au, Magdalena Gomulka, Esther Bravo Govea, Louise-Anne Charles, Fatima Oury Sow Gueye, et al. 2023. "IFLA Trend Report 2022 Update." <https://repository.ifla.org/handle/123456789/2456>.
- Alammar, Jay. 2019. "The Illustrated Word2vec." 2019.  
<http://jalammar.github.io/illustrated-word2vec/>.
- Algorithmic Justice League. 2022. "Algorithmic Justice League - Unmasking AI Harms and Biases." 2022. <https://www.ajl.org/>.
- Alpert-Abrams, Hannah. 2016. "Machine Reading the Primeros Libros" 10 (4).  
<http://www.digitalhumanities.org/dhq/vol/10/4/000268/000268.html>.
- Altman, Sam, dir. 2023. *OpenAI DevDay, Opening Keynote*.  
<https://www.youtube.com/watch?v=U9mJuUkhUzk>.
- Amatriain, Xavier. 2023. "Transformer Models: An Introduction and Catalog." arXiv.Org. 2023. <https://arxiv.org/abs/2302.07730v2>.
- American Association of Law Libraries. 2019. "AALL Ethical Principles." AALL. 2019.  
<https://www.aallnet.org/advocacy/government-relations/recommended-guidelines/aall-ethical-principles/>.

- American Library Association. 2006. "Privacy: An Interpretation of the Library Bill of Rights." <https://www.ala.org/advocacy/intfreedom/librarybill/interpretations/privacy>.
- . 2007. "Types of Libraries." Text. Education & Careers. 2007. <https://www.ala.org/educationcareers/careers/librariycareerssite/typesoflibraries>.
- . 2008. "Office for Intellectual Freedom." Text. About ALA. 2008. <https://www.ala.org/aboutala/offices/oif>.
- . 2018. "Facial Recognition." Text. Tools, Publications & Resources. 2018. <https://www.ala.org/tools/future/trends/facialrecognition>.
- . 2021. "Professional Ethics and Code of Ethics." Text. Tools, Publications & Resources. 2021. <https://www.ala.org/tools/ethics>.
- Amodei, Dario, Danny Hernandez, Girish Sastry, Jack Clark, Greg Brockman, and Ilya Sutskever. 2019. "AI and Compute." OpenAI. 2019. <https://openai.com/blog/ai-and-compute/>.
- anc. 2023. "The Open American National Corpus." 2023. <https://anc.org/>.
- Angwin, Julia, and Jeff Larson. 2016. "Machine Bias." Text/html. ProPublica. May 23, 2016. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.
- ANSI/NISO, National Information Standards Organization. 2010. "Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies." National Information Standards Organization. [https://groups.niso.org/apps/group\\_public/download.php/12591/z39-19-2005r2010.pdf](https://groups.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf).
- Anthropic. 2024. "Meet Claude." 2024. <https://www.anthropic.com/claude>.
- Araújo, Paula Carina de, Renata Cristina Gutierrez Castanha, and Birger Hjørland. 2021. "Citation Indexing and Indexes." *Knowledge Organization*, . Also available in ISKO Encyclopedia of Knowledge Organization, eds. Birger Hjørland and Claudio Gnoli, <https://www.isko.org/cyclo/citation>, 48 (1): 72–101.
- ARCPrize. 2024. "ARC Prize - Official Guide." ARC Prize. 2024. <https://arcprize.org/guide>.
- Arlitsch, Kenning, and Bruce Newell. 2017. "Thriving in the Age of Accelerations: A Brief Look at the Societal Effects of Artificial Intelligence and the Opportunities for Libraries." *Journal of Library Administration* 57 (7): 789–98. <https://doi.org/10.1080/01930826.2017.1362912>.
- Aschenbrenner, Leopold. 2024. "Situational Awareness: The Decade Ahead." 2024. <https://situational-awareness.ai/>.
- Asemi, Asefeh, Andrea Ko, and Mohsen Nowkarizi. 2020. "Intelligent Libraries: A Review on Expert Systems, Artificial Intelligence, and Robot." *Library Hi Tech* 39 (2): 412–34. <https://doi.org/10.1108/LHT-02-2020-0038>.
- Ayre, Lori, and Jim Craner. 2018. "Algorithms: Avoiding the Implementation of Institutional Biases." *Public Library Quarterly* 37 (3): 341–47. <https://doi.org/10.1080/01616846.2018.1512811>.
- Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. 2016. "Neural Machine Translation by Jointly Learning to Align and Translate." arXiv. <https://doi.org/10.48550/arXiv.1409.0473>.

- Bapna, Ankur, Isaac Caswell, Julia Kreutzer, Orhan Firat, Daan van Esch, Aditya Siddhant, Mengmeng Niu, et al. 2022. "Building Machine Translation Systems for the Next Thousand Languages." arXiv. <https://doi.org/10.48550/arXiv.2205.03983>.
- Barité, Mario. 2018. "Literary Warrant (IEKO)." 2018. [https://www.isko.org/cyclo/literary\\_warrant](https://www.isko.org/cyclo/literary_warrant).
- Barocas, Solon, Kate Crawford, Aaron Shapiro, and Hanna Wallach. 2017. "The Problem with Bias: From Allocative to Representational Harms in Machine Learning." Special Interest Group for Computing, Information and Society (SIGCIS) (2017)."
- Beer, David. 2017. "The Social Power of Algorithms." *Information, Communication & Society* 20 (1): 1–13. <https://doi.org/10.1080/1369118X.2016.1216147>.
- Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜." In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–23. FAccT '21. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445922>.
- Benjamin, Garfield. 2022. "#FuckTheAlgorithm: Algorithmic Imaginaries and Political Resistance." In *2022 ACM Conference on Fairness, Accountability, and Transparency*, 46–57. FAccT '22. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3531146.3533072>.
- Berman, Erin. 2018. "Big Brother Is Watching You: The Ethical Role of Libraries and Big Data." *Choose Privacy Every Day* (blog). 2018. <https://chooseprivacyeveryday.org/the-ethical-role-of-libraries-and-big-data/>.
- Berman, Matthew. 2024. "Intro to RAG for AI (Retrieval Augmented Generation)." 2024. [https://www.youtube.com/watch?v=Yo8Nn230\\_mY](https://www.youtube.com/watch?v=Yo8Nn230_mY).
- Berman, Sanford. 1971. *Prejudices and Antipathies: A Tract on the LC Subject Heads Concerning People*. Jefferson, North Carolina: McFarland & Company, Inc.
- . 2000. "Review of Bowker, Geoffrey C. and Star, Susan Leigh. 'Sorting [Sic] Things out: Classification and Its Consequences.'" *Progressive Librarian* 17.
- Berry, John N. 1998. "Choosing Sides: The Impasse between Prudes and Purists Has Forced the Issue." *Library Journal* 123 (4): 6–7.
- Bhatia, Aatish. 2023. "How Can an A.I. Learn to Write? Choose a Famous Author, and We'll Show You." *The New York Times*, 2023, sec. The Upshot. <https://www.nytimes.com/interactive/2023/04/26/upshot/gpt-from-scratch.html>.
- Biosphere. 2022. "Home | Biosphere 2." 2022. <https://biosphere2.org/>.
- Blei, David M, Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3:993–1022.
- Blodgett, Su Lin, Solon Barocas, Hal Daumé III, and Hanna Wallach. 2020. "Language (Technology) Is Power: A Critical Survey of 'Bias' in NLP." In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5454–76. Online: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.acl-main.485>.
- Bobrow, Daniel G, Ronald M Kaplan, Martin Kay, Donald A Norman, Henry Thompson, and Terry Winograd. 1977. "GUS, A Frame-Driven Dia|og System," 19.

- Bolukbasi, Tolga, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. 2016. "Man Is to Computer Programmer as Woman Is to Homemaker? Debiasing Word Embeddings." In , 9. Barcelona, Spain.
- Bommasani, Rishi, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, et al. 2022. "On the Opportunities and Risks of Foundation Models." arXiv. <https://doi.org/10.48550/arXiv.2108.07258>.
- Bommasani, Rishi, Kevin Klyman, Shayne Longpre, Sayash Kapoor, Nestor Maslej, Betty Xiong, Daniel Zhang, and Percy Liang. 2023. "The Foundation Model Transparency Index." arXiv. <https://doi.org/10.48550/arXiv.2310.12941>.
- Bourg, Chris. 2017. "What Happens to Libraries and Librarians When Machines Can Read All the Books?" *Feral Librarian* (blog). 2017. <https://chrisbourg.wordpress.com/2017/03/16/what-happens-to-libraries-and-librarians-when-machines-can-read-all-the-books/>.
- Bowker, Geoffrey C., and Susan Leigh Star. 2000. *Sorting Things out: Classification and Its Consequences*. Cambridge, MA: The MIT Press.
- Briggs, James, and Franciso Ingham. 2022. "LangChain AI Handbook." Pinecone. 2022. <https://www.pinecone.io/learn/langchain/>.
- British Library. 2020. "Digitised Manuscripts Harley MS 7368." 2020. [https://www.bl.uk/manuscripts/FullDisplay.aspx?ref=Harley\\_MS\\_7368](https://www.bl.uk/manuscripts/FullDisplay.aspx?ref=Harley_MS_7368).
- Brooks, Frederick P. 1975. *The Mythical Man-Month*. 1st ed. Addison-Wesley.
- Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, et al. 2020. "Language Models Are Few-Shot Learners." arXiv. <https://doi.org/10.48550/arXiv.2005.14165>.
- Brunon-Ernst, Anne, ed. 2012. *Beyond Foucault: New Perspectives on Bentham's Panopticon*. <https://www.routledge.com/Beyond-Foucault-New-Perspectives-on-Benthams-Panopticon/Brunon-Ernst/p/book/9780754668435>.
- Budds, Diana, Diana Budds, and Diana Budds. 2017. "Biased AI Is A Threat To Civil Liberties. The ACLU Has A Plan To Fix It." Fast Company. July 25, 2017. <https://www.fastcompany.com/90134278/biased-ai-is-a-threat-to-civil-liberty-the-aclu-has-a-plan-to-fix-it>.
- Buolamwini, Joy. 2016. "The Coded Gaze." AJL -ALGORITHMIC JUSTICE LEAGUE. November 6, 2016. <https://medium.com/mit-media-lab/the-algorithmic-justice-league-3cc4131c5148>.
- . 2019. "The Algorithmic Justice League." *MIT MEDIA LAB* (blog). 2019. <https://medium.com/mit-media-lab/the-algorithmic-justice-league-3cc4131c5148>.
- . 2023. *Unmasking AI: My Mission to Protect What Is Human in the World of Machines*. <https://www.penguinrandomhouse.com/books/670356/unmasking-ai-by-joy-buolamwini/>.
- Buolamwini, Joy, Vicente Ordóñez, Jamie Morgenstern, and Erik Learned-Miller. 2020. "Facial Recognition Technologies: A Primer." [https://assets.website-files.com/5e027ca188c99e3515b404b7/5ed1002058516c11edc66a14\\_FRTsPrimerMay2020.pdf](https://assets.website-files.com/5e027ca188c99e3515b404b7/5ed1002058516c11edc66a14_FRTsPrimerMay2020.pdf).
- Butcher, H.K., G.M. Bulechek, J.M. Docterman, and C.M. Wagner, eds. 2018. *Nursing Intervention Classification*.
- c.ai. 2023. "Character.Ai." Character.Ai. 2023. <https://beta.character.ai/>.

- Caliskan, Aylin. 2021. "Detecting and Mitigating Bias in Natural Language Processing." *Brookings* (blog). 2021. <https://www.brookings.edu/research/detecting-and-mitigating-bias-in-natural-language-processing/>.
- Caliskan, Aylin, Joanna J. Bryson, and Arvind Narayanan. 2017. "Semantics Derived Automatically from Language Corpora Contain Human-like Biases." *Science* 356 (6334): 183–86. <https://doi.org/10.1126/science.aal4230>.
- Carlson, Jake, and Lisa Johnston. 2015. *Data Information Literacy: Librarians, Data, and the Education of a New Generation of Researchers*. <http://public.eblib.com/choice/PublicFullRecord.aspx?p=2039088>.
- Cartter, Eileen. 2023. "The Pope Francis Puffer Photo Was Real in Our Hearts." GQ. 2023. <https://www.gq.com/story/pope-puffer-jacket-midjourney-ai-meme>.
- Casscells, W., A. Schoenberger, and T. B. Graboys. 1978. "Interpretation by Physicians of Clinical Laboratory Results." *The New England Journal of Medicine* 299 (18): 999–1001. <https://doi.org/10.1056/NEJM197811022991808>.
- Cauchy, Louis Augustin. 1847. "Méthode Générale Pour La Résolution Des Systèmes d'équations Simultanées." *Compte Rendu à l'Académie Des Sciences* 25:536–38.
- CCP. 2020. "Center for Creative Photography." Center for Creative Photography. 2020. <https://ccp.arizona.edu/home>.
- Chan, Lois Mai. 2007. *Cataloging and Classification: An Introduction*. 3rd ed. Lanham, Maryland: The Scarecrow Press, Inc.
- Chan, Lois Mai, Phyllis A. Richmond, and Elaine Svenonius. 1985. "Principles of Book Classification: E. Wyndham Hulme: Editor's Introduction." In *Theory of Subject Analysis*. Edited by Lois Mai Chan, Phyllis A. Richmond and Elaine Svenonius. Littleton, Colorado, 48-49.
- Chase, Harrison. (2022) 2022. "LangChain." Python. <https://github.com/hwchase17/langchain>.
- Chen, Xiongwen, and Dale G. Brockway. 2017. "Height-Diameter Relationships in Longleaf Pine and Four Swamp Tree Species." *Journal of Plant Studies* 6 (2): 94–101. <https://www.srs.fs.usda.gov/pubs/54612>.
- Cherian, Anoop, Kuan-Chuan Peng, Suhas Lohit, Kevin A. Smith, and Joshua B. Tenenbaum. 2023. "Are Deep Neural Networks SMARTer than Second Graders?" arXiv. <https://doi.org/10.48550/arXiv.2212.09993>.
- Chiang, Wei-Lin, Lianmin Zheng, Ying Sheng, Anastasios Nikolas Angelopoulos, Tianle Li, Dacheng Li, Hao Zhang, et al. 2024. "Chatbot Arena: An Open Platform for Evaluating LLMs by Human Preference." arXiv. <https://doi.org/10.48550/arXiv.2403.04132>.
- Cho, Kyunghyun, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. "Learning Phrase Representations Using RNN Encoder–Decoder for Statistical Machine Translation." In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1724–34. Doha, Qatar: Association for Computational Linguistics. <https://doi.org/10.3115/v1/D14-1179>.
- CHOICE Media Channel, dir. 2022. *Artificial Intelligence in Academic Libraries: How New AI Services Can Support Your Library Users*. <https://www.youtube.com/watch?v=Ssg-sKLIqok>.
- Chollet, François. 2019. "On the Measure of Intelligence." arXiv. <https://doi.org/10.48550/arXiv.1911.01547>.

- Chomsky, Noam, and Ramin Mirfakhraie. 2023. "ChatGPT and Human Intelligence: Noam Chomsky Responds to Critics | MR Online." 2023. <https://mronline.org/2023/04/24/chatgpt-and-human-intelligence-noam-chomsky-responds-to-critics/>.
- Chomsky, Noam, Ian Roberts, and Jeffrey Watumull. 2023. "Opinion | Noam Chomsky: The False Promise of ChatGPT." *The New York Times*, 2023, sec. Opinion. <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html>.
- Chowdhury, Ruman. 2023. "Rumman Chowdhury, Ph.D." Rumman Chowdhury, Ph.D. 2023. <http://www.rummanchowdhury.com>.
- Clariant Creative Agency. 2022. "Your Guide to Pillar Pages and Topic Clusters | Clariant Creative Agency." 2022. <https://www.clariantcreative.com/guide-to-pillar-pages-and-topic-clusters>.
- Clark, Jason A. (2018) 2022. "Algorithmic-Awareness." Rich Text Format. <https://github.com/jasonclark/algorithmic-awareness>.
- Coding Vidya. 2023. "Home | Coding Vidya - Best Computer Science Portal -." 2023. <https://codingvidya.com/>.
- Colby, Kenneth Mark, Franklin Dennis Hilf, Sylvia Weber, and Helena C Kraemer. 1972. "Turing-like Indistinguishability Tests for the Validation of a Computer Simulation of Paranoid Processes." *Artificial Intelligence* 3:199–221. [https://doi.org/10.1016/0004-3702\(72\)90049-5](https://doi.org/10.1016/0004-3702(72)90049-5).
- Colyer, Adrian. 2016. "The Amazing Power of Word Vectors | the Morning Paper." 2016. <https://blog.acolyer.org/2016/04/21/the-amazing-power-of-word-vectors/>.
- Common Crawl. 2022. "Common Crawl." 2022. <https://commoncrawl.org/>.
- Cook, John. 2023. "Cranky Uncle." Cranky Uncle. 2023. <https://crankyuncle.com/>.
- Corbett-Davies, Sam, and Sharad Goel. 2018. "The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning." arXiv. <https://doi.org/10.48550/arXiv.1808.00023>.
- Cordell, Ryan. 2020. "Machine Learning + Libraries." LC Labs. Library of Congress. <https://labs.loc.gov/static/labs/work/reports/Cordell-LOC-ML-report.pdf>.
- Corrado, Edward M. 2021. "Artificial Intelligence: The Possibilities for Metadata Creation." *Technical Services Quarterly* 38 (4): 395–405. <https://doi.org/10.1080/07317131.2021.1973797>.
- Costa, Ricky. 2023. "ChatGPT Cheat Sheet." <https://Neuralmagic.Com/>. [https://www.kdnuggets.com/publications/sheets/ChatGPT\\_Cheatsheet\\_Costa.pdf](https://www.kdnuggets.com/publications/sheets/ChatGPT_Cheatsheet_Costa.pdf).
- Cox, Andrew M., and Suvodeep Mazumdar. 2022. "Defining Artificial Intelligence for Librarians." *Journal of Librarianship and Information Science*, 09610006221142029. <https://doi.org/10.1177/09610006221142029>.
- Cox, Andrew M., Stephen Pinfield, and Sophie Rutter. 2019. "The Intelligent Library: Thought Leaders' Views on the Likely Impact of Artificial Intelligence on Academic Libraries." *Library Hi Tech* 37 (3): 418–35. <https://doi.org/10.1108/LHT-08-2018-0105>.
- Coyle, Karen. 2016. "FRBR Before and After." 2016. <http://kcoyle.net/beforeAndAfter/>.
- Crawford, Kate, dir. 2017. *The Trouble with Bias - NIPS 2017 Keynote - Kate Crawford #NIPS2017*. Neural Information Processing Systems. [https://www.youtube.com/watch?v=fMym\\_BKWQzk](https://www.youtube.com/watch?v=fMym_BKWQzk).

- . 2022. “Lessons From The Panoptic Sort.” *International Journal of Communication*, no. 16, 1632–34.
- Crichton, Gamal, Simon Baker, Yufan Guo, and Anna Korhonen. 2020. “Neural Networks for Open and Closed Literature-Based Discovery.” *PLOS ONE* 15 (5): e0232891. <https://doi.org/10.1371/journal.pone.0232891>.
- Das, Kinnor, Clay J. Cockerell, Anant Patil, Paweł Pietkiewicz, Mario Giulini, Stephan Grabbe, and Mohamad Goldust. 2021. “Machine Learning and Its Application in Skin Cancer.” *International Journal of Environmental Research and Public Health* 18 (24): 13409. <https://doi.org/10.3390/ijerph182413409>.
- Das, Rajesh Kumar, and Mohammad Sharif Ul Islam. 2021. “Application of Artificial Intelligence and Machine Learning in Libraries: A Systematic Review.” *arXiv:2112.04573 [Cs]*. <http://arxiv.org/abs/2112.04573>.
- Data Information Literacy Project. 2023. “Data Information Literacy.” 2023. <https://www.datainfolit.org/>.
- Davis, Wayne. 2019. “Implicature.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Fall 2019. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/fall2019/entries/implicature/>.
- Dean, Jeff. 2023. “Google Research, 2022 & beyond: Language, Vision and Generative Models.” 2023. <https://ai.googleblog.com/2023/01/google-research-2022-beyond-language.html#GenerativeModels>.
- Deepankar, and Florian. 2023. “PromptPerfect - Elevate Your Prompts to Perfection with AI Prompt Engineering.” 2023. <https://promptperfect.jina.ai/>.
- Dekker, Harrison, Angel Ferria, and Indrani Mandal. 2022. “URI Libraries’ AI Lab--Evolving to Meet the Needs of Students and Research Communities.”
- Dempsey, Lorcan. 2023a. “Generative AI and Large Language Models: Background and Contexts.” LorcanDempsey.Net. 2023. <https://www.lorcandempsey.net/intro-gen-ai/>.
- . 2023b. “Generative AI, Scholarly and Cultural Language Models, and the Return of Content.” LorcanDempsey.Net. 2023. <https://www.lorcandempsey.net/generative-ai-a-note-about-content/>.
- Desjardins, Jeff. 2017. “How Many Millions of Lines of Code Does It Take?” Visual Capitalist. February 8, 2017. <https://www.visualcapitalist.com/millions-lines-of-code/>.
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. “BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding.” *arXiv:1810.04805 [Cs]*, May. <http://arxiv.org/abs/1810.04805>.
- Dewland, Thomas A., Jeffrey E. Olgin, Eric Vittinghoff, and Gregory M. Marcus. 2013. “Incident Atrial Fibrillation Among Asians, Hispanics, Blacks, and Whites.” *Circulation* 128 (23): 2470–77. <https://doi.org/10.1161/CIRCULATIONAHA.113.002449>.
- Dickson, Ben. 2023. “How to Customize LLMs like ChatGPT with Your Own Data and Documents - TechTalks.” 2023. <https://bdtechtalks.com/2023/05/01/customize-chatgpt-llm-embeddings/>.
- Digital2030. 2022. “The Algorithm Literacy Project | Understanding Algorithms.” 2022. <https://algorithmliteracy.org/>.

- Dizikes, Peter. 2011. "When the Butterfly Effect Took Flight." MIT Technology Review. 2011. <https://www.technologyreview.com/2011/02/22/196987/when-the-butterfly-effect-took-flight/>.
- Driess, Danny. 2023. "PaLM-E: An Embodied Multimodal Language Model." 2023. <https://ai.googleblog.com/2023/03/palm-e-embodied-multimodal-language.html>.
- Druga, Stefania, Sarah T. Vu, Eesh Likhith, and Tammy Qiu. 2019a. "Inclusive AI Literacy for Kids around the World." In *Proceedings of FabLearn 2019*, 104–11. FL2019. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3311890.3311904>.
- . 2019b. "Inclusive AI Literacy for Kids around the World." In *Proceedings of FabLearn 2019*, 104–11. FL2019. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3311890.3311904>.
- Duhem, Pierre. 1914. *La Théorie Physique Son Objet et Sa Structure, 2nd Ed.,*. Translated by The Aim and Structure of Physical Theory English Translation Phillip Wiener. Paris: Chevalier et Rivière.
- Economist. 2022. "Huge 'Foundation Models' Are Turbo-Charging AI Progress." *The Economist*, 2022. <https://www.economist.com/interactive/briefing/2022/06/11/huge-foundation-models-are-turbo-charging-ai-progress>.
- Econtalk. 2023. "Adam Mastroianni on Peer Review and the Academic Kitchen." Econlib. 2023. <https://www.econtalk.org/adam-mastroianni-on-peer-review-and-the-academic-kitchen/>.
- Egan, Margaret E., and Jesse H. Shera. 1952. "Foundations of a Theory of Bibliography." [https://www-jstor-org.ezproxy4.library.arizona.edu/stable/4304106#metadata\\_info\\_tab\\_contents](https://www-jstor-org.ezproxy4.library.arizona.edu/stable/4304106#metadata_info_tab_contents).
- Encyclopedia.com. 2019. "Encyclopedia.Com | Free Online Encyclopedia." 2019. <https://www.encyclopedia.com/>.
- ExLibris. 2019. "Artificial Intelligence in the Library: Advantages, Challenges and Tradition." <https://cdn2.hubspot.net/hubfs/2909474/Ex%20Libris%20Artificial%20Intelligence%20White%20Paper.pdf>.
- Fallis, Don. 2002. "Introduction: Social Epistemology and Information Science." *Social Epistemology* 16 (1): 1–4. <https://doi.org/10.1080/02691720210132752>.
- . 2006. "Social Epistemology and Information Science." In *Annual Review of Information Science and Technology*, edited by Blaise Cronin. Vol. 40.
- Fedus, William, Barret Zoph, and Noam Shazeer. 2022. "Switch Transformers: Scaling to Trillion Parameter Models with Simple and Efficient Sparsity." arXiv. <https://doi.org/10.48550/arXiv.2101.03961>.
- Feigenbaum, E.A. 1989. "Toward the Library of the Future." *Long Range Planning* 22 (1): 118–23. [https://doi.org/10.1016/0024-6301\(89\)90059-9](https://doi.org/10.1016/0024-6301(89)90059-9).
- Fernandez, Peter. 2016. "Through the Looking Glass: Envisioning New Library Technologies' How Artificial Intelligence Will Impact Libraries." *Library Hi Tech News* 33 (5): 5–8. <https://doi.org/10.1108/LHTN-05-2016-0024>.
- . 2023. "Through the Looking Glass: Envisioning New Library Technologies' AI-Text Generators as Explained by ChatGPT." *Library Hi Tech News* 40 (3): 11–14. <https://doi.org/10.1108/LHTN-02-2023-0017>.

- Firmani, D., Marco Maiorino, P. Merialdo, and Elena Nieddu. 2018. "Towards Knowledge Discovery from the Vatican Secret Archives. In Codice Ratio - Episode 1: Machine Transcription of the Manuscripts." *KDD*.  
<https://doi.org/10.1145/3219819.3219879>.
- Firmani, Donatella, Paolo Merialdo, and Marco Maiorino. 2017. "In Codice Ratio: Scalable Transcription of Vatican Registers." 2017. <https://ercim-news.ercim.eu/en111/special/in-codice-ratio-scalable-transcription-of-vatican-registers>.
- Fitch, Kent. 2023. "Searching for Meaning Rather Than Keywords and Returning Answers Rather Than Links." *The Code4Lib Journal*, no. 57.  
<https://journal.code4lib.org/articles/17443>.
- Fogg, B.J. 2003. *Persuasive Technology: Using Computers to Change What We Think and Do*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.
- Frické, Martin. 1997. "Information Using Likeness Measures." *Journal of the American Society for Information Science* 48:882–92.
- . 2012. *Logic and the Organization of Information*. New York: Springer.
- . 2015. "Big Data and Its Epistemology." *Journal of the Association for Information Science and Technology* 66:651–61.
- . 2021. "Boolean Logic." *Knowledge Organization* 48 (2): 177–91.  
<https://doi.org/10.5771/0943-7444-2021-2-177>.
- Frické, Martin, Kay Mathiesen, and Don Fallis. 2000. "The Ethical Presuppositions behind the Library Bill of Rights." *The Library Quarterly* 70 (4): 468–91.  
<https://doi.org/10.1086/603218>.
- Friendly, Michael. 2007. "Gallery of Data Visualization." 2007.  
<https://www.datavis.ca/gallery/>.
- Froelich, Thomas. 2004. "A Brief History of Information Ethics." BID 13 Desembre 2004. 2004.
- Fu, Yao, Hao Peng, and Tushar Khot. 2023. "How Does GPT Obtain Its Ability? Tracing Emergent Abilities of Language Models to Their Sources." 2023.  
<https://yaofu.notion.site/How-does-GPT-Obtain-its-Ability-Tracing-Emergent-Abilities-of-Language-Models-to-their-Sources-b9a57ac0fcf74f30a1ab9e3e36fa1dc1>.
- Fuller, Steve. 1988. *Social Epistemology (Bloomington: Indiana University Press)*. Bloomington: Indiana University Press.
- G2. 2023. "Best Bot Platforms Software." G2. 2023.  
<https://www.g2.com/categories/bot-platforms>.
- Gadd, Elizabeth. 2020. "AI-Based Citation Evaluation Tools: Good, Bad or Ugly?" *The Bibliomagician* (blog). 2020.  
<https://thebibliomagician.wordpress.com/2020/07/23/ai-based-citation-evaluation-tools-good-bad-or-ugly/>.
- Gale, part of Cengage Group. 2023. "Library Marketing: Improve Outreach with Gale Analytics." 2023. <https://www.gale.com/databases/gale-analytics>.
- Gallagher, James, and Piotr Skalski. 2023. "First Impressions with GPT-4V(Ision)." Roboflow Blog. 2023. <https://blog.roboflow.com/gpt-4-vision/>.
- Gandy Jr., Oscar H. 2021. *The Panoptic Sort: A Political Economy of Personal Information*. Second Edition, Second Edition. Oxford, New York: Oxford University Press.

- Gao, Yunfan, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, and Haofen Wang. 2023. "Retrieval-Augmented Generation for Large Language Models: A Survey." arXiv. <https://doi.org/10.48550/arXiv.2312.10997>.
- Garnar, Martin, and Trina Magi, eds. 2021. *Intellectual Freedom Manual*. 10th ed. ALA Editions. <https://www.alastore.ala.org/content/intellectual-freedom-manual-tenth-edition>.
- GDPR. 2018. "Art. 22 GDPR – Automated Individual Decision-Making, Including Profiling." *General Data Protection Regulation (GDPR)* (blog). 2018. <https://gdpr-info.eu/art-22-gdpr/>.
- Gee, Georgia. 2023. "Here Are the Stadiums That Are Keeping Track of Your Face." *Slate*, 2023. <https://slate.com/technology/2023/03/madison-square-garden-facial-recognition-stadiums-list.html>.
- Gesser, Avi, Robert Maddox, Anna Gressel, Mengyi Xu, Samuel J. Allaman, and Andres S. Gutierrez. 2022. "New Automated Decision-Making Laws: Four Tips for Compliance." *Debevoise Data Blog* (blog). 2022. <https://www.debevoisedatablog.com/2022/06/25/new-automated-decision-making-laws-four-tips-for-compliance/>.
- Gillies, Midge. 2020. "Amy Johnson – A Brief Biography – Amy Johnson Arts Trust." 2020. <http://amyjohnsonartstrust.co.uk/her-life/>.
- Github. 2022. "GitHub Copilot · Your AI Pair Programmer." GitHub. 2022. <https://github.com/features/copilot>.
- Glusac, Elaine. 2021. "Your Face Is, or Will Be, Your Boarding Pass." *The New York Times*, 2021, sec. Travel. <https://www.nytimes.com/2021/12/07/travel/biometrics-airports-security.html>.
- Goldman, Alvin I. 1999. *Knowledge in a Social World*. Oxford: Clarendon Press.
- Goodman, Bryce, and Seth Flaxman. 2017. "European Union Regulations on Algorithmic Decision-Making and a 'Right to Explanation.'" *AI Magazine* 38 (3): 50–57. <https://doi.org/10.1609/aimag.v38i3.2741>.
- Google Cloud. 2023. "Speech-to-Text Supported Languages | Cloud Speech-to-Text Documentation | Google Cloud." 2023. <https://cloud.google.com/speech-to-text/docs/speech-to-text-supported-languages>.
- Google for Developers. 2022. "Machine Learning Crash Course." 2022. <https://developers.google.com/machine-learning/crash-course>.
- . 2023. "Machine Learning Glossary: ML Fundamentals." Google for Developers. 2023. <https://developers.google.com/machine-learning/glossary/fundamentals>.
- Gozalo-Brizuela, Roberto, and Eduardo C. Garrido-Merchan. 2023. "ChatGPT Is Not All You Need. A State of the Art Review of Large Generative AI Models." arXiv. <https://doi.org/10.48550/arXiv.2301.04655>.
- Grant, Nico, and Kashmir Hill. 2023. "Google's Photo App Still Can't Find Gorillas. And Neither Can Apple's." *The New York Times*, 2023, sec. Technology. <https://www.nytimes.com/2023/05/22/technology/ai-photo-labels-google-apple.html>.
- Griffey, Jason. 2019. "Artificial Intelligence and Machine Learning in Libraries." gwern. 2023. "Douglas Hofstadter Changes His Mind on Deep Learning & AI Risk." <https://www.lesswrong.com/posts/kAmgdEjq2eYQkB5PP/douglas-hofstadter-changes-his-mind-on-deep-learning-and-ai>.

- Hacking, Ian. 1999. *The Social Construction of What?* London: Harvard University Press.
- HAI. 2023. “AI Index Report 2023 – Artificial Intelligence Index.” 2023. <https://aiindex.stanford.edu/report/>.
- Hanson, Norwood Russell. 1958. *Patterns of Discovery*. Cambridge: Cambridge University Press.
- Hardt, Moritz, Eric Price, and Nathan Srebro. 2016. “Equality of Opportunity in Supervised Learning [Preprint].” arXiv. <https://doi.org/10.48550/arXiv.1610.02413>.
- Harper, Charlie. 2018. “Machine Learning and the Library or: How I Learned to Stop Worrying and Love My Robot Overlords.” *The Code4Lib Journal*, no. 41. <https://journal.code4lib.org/articles/13671>.
- Harpring, Patricia. 2020. “Featuring the Getty Vocabularies,” 91.
- Harris, Richard. 2002. “The deHavilland D.H.82 Tiger Moth and the Moth Family.” 2002. [https://harris1.net/hold/av/avhist/a8/a8\\_moth.htm](https://harris1.net/hold/av/avhist/a8/a8_moth.htm).
- Hauptman, Robert. 1988. *Ethical Challenges in Librarianship*. Oryx.
- . 2002. *Ethics and Librarianship*. Jefferson, N.C.: McFarland.
- Heatley, Louise M. 2023. “National Centre for Text Mining – NaCTEM.” XHTML. NaCTeM. 2023. <https://www.nactem.ac.uk/index.php>.
- Heaven, Will Douglas. 2022. “Why Meta’s Latest Large Language Model Survived Only Three Days Online.” MIT Technology Review. 2022. <https://www.technologyreview.com/2022/11/18/1063487/meta-large-language-model-ai-only-survived-three-days-gpt-3-science/>.
- Henderson, Leah. 2020. “The Problem of Induction.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2020. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/spr2020/entries/induction-problem/>.
- Hernan, Miguel. 2022. “Causal Diagrams: Draw Your Assumptions Before Your Conclusions.” edX. 2022. <https://www.edx.org/course/causal-diagrams-draw-your-assumptions-before-your>.
- Hjørland, Birger. 2011. “Evidence-Based Practice: An Analysis Based on the Philosophy of Science.” *Journal of the American Society for Information Science and Technology* 62 (7): 1301–10. <https://doi.org/10.1002/asi.21523>.
- Hofstadter, Douglas. 2018. “The Shallowness of Google Translate.” *The Atlantic*. 2018. <https://www.theatlantic.com/technology/archive/2018/01/the-shallowness-of-google-translate/551570/>.
- Hollerith, Herman. 1889. Art of compiling statistics. United States US395782A, filed September 23, 1884, and issued 1889. <https://patents.google.com/patent/US395782A/en>.
- Hond, Anne A. H. de, Marieke M. van Buchem, and Tina Hernandez-Boussard. 2022. “Picture a Data Scientist: A Call to Action for Increasing Diversity, Equity, and Inclusion in the Age of AI.” *Journal of the American Medical Informatics Association: JAMIA*, 2178–81. <https://doi.org/10.1093/jamia/ocac156>.
- Hornik, Kurt, Maxwell Stinchcombe, and Halbert White. 1989. “Multilayer Feedforward Networks Are Universal Approximators.” *Neural Networks* 2 (5): 359–66. [https://doi.org/10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8).

- Howard, Jennifer. 2017. "What Happened to Google's Effort to Scan Millions of University Library Books? - EdSurge News." EdSurge. August 10, 2017. <https://www.edsurge.com/news/2017-08-10-what-happened-to-google-s-effort-to-scan-millions-of-university-library-books>.
- Howson, Colin, and Peter Urbach. 2006. *Scientific Reasoning : The Bayesian Approach*. 3rd ed. Chicago: Open Court.
- Huang, Austin, Suraj Subramanian, Jonathan Sum, Khalid Almubarak, and Stella Biderman. 2018. "The Annotated Transformer." 2018. <http://nlp.seas.harvard.edu/annotated-transformer/>.
- Hugging Face. 2023. "Hugging Face – The AI Community Building the Future." 2023. <https://huggingface.co/>.
- Huntington-Klein, Nick. 2022. *The Effect: An Introduction to Research Design and Causality | The Effect*. <https://theeffectbook.net/index.html>.
- Huyen, Chip. 2023. "Building LLM Applications for Production." 2023. <https://huyenchip.com/2023/04/11/llm-engineering.html>.
- IFLA. 2012. "IFLA Code of Ethics for Librarians and Other Information Workers (Full Version) – IFLA." 2012. <https://www.ifla.org/publications/ifla-code-of-ethics-for-librarians-and-other-information-workers-full-version/>.
- . 2020. "IFLA Statement on Libraries and Artificial Intelligence." <https://repository.ifla.org/handle/123456789/1646>.
- Imbler, Sabrina. 2021. "This Moth's Name Is a Slur. Scientists Won't Use It Anymore." *The New York Times*, 2021, sec. Science. <https://www.nytimes.com/2021/07/09/science/gypsy-moth-romani-entomological-society.html>.
- Iris.ai. 2023. "The Workspace Tools." Iris.Ai - Your Researcher Workspace. 2023. <https://iris.ai/features/>.
- Izquierdo, H. Andrés. 2022. "20 Artificial Intelligence and Text and Data Mining: Future Rules for Libraries?" In *Navigating Copyright for Libraries*, edited by Jessica Coates, Victoria Owen, and Susan Reilly, 497–540. De Gruyter Saur. <https://doi.org/10.1515/9783110732009-022>.
- Jakeway, Eileen, Lauren Algee, Laurie Allen, Meghan Ferriter, Jaime Mears, Abigail Potter, and Kate Zwaard. 2020. "Machine Learning + Libraries Summit Event Summary." LC Labs Digital Strategy Directorate.
- Jo, Eun Seo, and Timnit Gebru. 2020. "Lessons from Archives: Strategies for Collecting Sociocultural Data in Machine Learning [Conference]." In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 306–16. FAT\* '20. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3351095.3372829>.
- Johansen, Johanna, Tore Pedersen, and Christian Johansen. 2021. "Studying Human-to-Computer Bias Transference." *AI & SOCIETY*. <https://doi.org/10.1007/s00146-021-01328-4>.
- Johnston, Lisa R., and Jon Jeffryes. 2015. "Teaching Civil Engineering Data Information Literacy Skills: An E-Learning Approach." In *Data Information Literacy*, edited by Lisa R. Johnston and Jake Carlson, 149–78. Librarians, Data, and the Education of a New Generation of Researchers. Purdue University Press. <https://www.jstor.org/stable/j.ctt6wq2vh.12>.

- Jurafsky, Dan, and James H. Martin. 2023. "Speech and Language Processing." 2023. <https://web.stanford.edu/~jurafsky/slp3/>.
- Jurafsky, Daniel, and James H. Martin. 2021. "Chapter 24 Chatbots & Dialogue Systems." In *Speech and Language Processing. Daniel Jurafsky & James H. Martin. Copyright © 2021. All Rights Reserved. Draft of December 29, 2021.* <https://web.stanford.edu/~jurafsky/slp3/24.pdf>.
- Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. Penguin Books.
- Kamradt, Greg. 2023. "Data Independent - YouTube." 2023. <https://www.youtube.com/channel/UCyR2Ct3pDOeZSRyZH5hPO-Q>.
- . (2023) 2023. "Learn LangChain." Jupyter Notebook. <https://github.com/gkamradt/langchain-tutorials>.
- Kaplan, Jared, Sam McCandlish, Tom Henighan, Tom B. Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. 2020. "Scaling Laws for Neural Language Models." arXiv. <https://doi.org/10.48550/arXiv.2001.08361>.
- Karpathy, Andrej. 2023a. "[D] A Baby GPT." Reddit Post. *R/MachineLearning*. [www.reddit.com/r/MachineLearning/comments/12h1zld/d\\_a\\_baby\\_gpt/](http://www.reddit.com/r/MachineLearning/comments/12h1zld/d_a_baby_gpt/).
- . 2023b. "Deep Neural Nets: 33 Years Ago and 33 Years from Now." 2023. <https://karpathy.github.io/2022/03/14/lecun1989/>.
- . (2022) 2023. "nanoGPT." Python. <https://github.com/karpathy/nanoGPT>.
- . 2023c. "State of GPT." Microsoft Build. 2023. <https://build.microsoft.com/en-US/sessions/db3f4859-cd30-4445-a0cd-553c3304f8e2>.
- Katell, Michael, Meg Young, Bernease Herman, Dharma Dailey, Aaron Tam, Vivian Guetler, Corinne Binz, Daniella Raz, and P. M. Krafft. 2019. "An Algorithmic Equity Toolkit for Technology Audits by Community Advocates and Activists." arXiv. <https://doi.org/10.48550/arXiv.1912.02943>.
- Kayid, Amr, and Nils Reimers. 2022. "Bonjour. مرحبا. Guten Tag. Hola. Cohere's Multilingual Text Understanding Model Is Now Available." Context by Cohere. 2022. <https://txt.cohere.com/multilingual/>.
- Kim, Bohyun. 2017. "AI-Powered Robots for Libraries: Exploratory Questions." In . Wildau, Germany. <https://library.ifla.org/id/eprint/2700/>.
- Kirwan Institute. 2017. "2017 State of the Science: Implicit Bias Review | Kirwan Institute for the Study of Race and Ethnicity." 2017. <https://kirwaninstitute.osu.edu/article/2017-state-science-implicit-bias-review>.
- Kitcher, Philip. 2002. "Veritistic Value and the Project of Social Epistemology." Edited by Alvin I. Goldman. *Philosophy and Phenomenological Research* 64 (1): 191–98. <https://www.jstor.org/stable/3071029>.
- Knapp, Jeff. 2021. "Library Guides: 'Fake' News: Resources for Fact-Checking." PennState University Libraries. 2021. <https://guides.libraries.psu.edu/c.php?g=620262&p=4319365>.
- Knowlton, Steven A. 2005. "Three Decades Since Prejudices and Antipathies: A Study of Changes in the Library of Congress Subject Headings." *Cataloging & Classification Quarterly, Vol. 40(2) 2005* 40:123–29.
- Knox, Emily J.M. 2023. *Foundations of Intellectual Freedom*. Chicago: ALA Neal-Schuman.
- Kunder, Maurice de. 2022. "WorldWideWebSize.Com | The Size of the World Wide Web (The Internet)." 2022. <https://www.worldwidewebsize.com/>.

- Kurzweil, Ray. 2005. *The Singularity Is Near: When Humans Transcend Biology*. 1st edition. New York: The Viking Press.
- Kusner, Matt J., Joshua R. Loftus, Chris Russell, and Ricardo Silva. 2018. "Counterfactual Fairness." *arXiv:1703.06856 [Cs, Stat]*.  
<http://arxiv.org/abs/1703.06856>.
- Lamont, Julian, and Christi Favor. 2017. "Distributive Justice." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2017. Metaphysics Research Lab, Stanford University.  
<https://plato.stanford.edu/archives/win2017/entries/justice-distributive/>.
- Larkin, Jill H., and Herbert A. Simon. 1987. "Why a Diagram Is (Sometimes) Worth Ten Thousand Words." *Cognitive Science* 11:65–99.
- LeCun, Yann, Corinna Cortes, and Chris Burges. 1998. "MNIST Handwritten Digit Database." 1998. <http://yann.lecun.com/exdb/mnist/>.
- Lee, Benjamin Charles Germain, Jaime Mears, Eileen Jakeway, Meghan Ferriter, Chris Adams, Nathan Yarasavage, Deborah Thomas, Kate Zwaard, and Daniel S. Weld. 2020. "The Newspaper Navigator Dataset: Extracting And Analyzing Visual Content from 16 Million Historic Newspaper Pages in Chronicling America." <https://doi.org/10.48550/arXiv.2005.01583>.
- Lemaréchal, Claude. 2012. "Cauchy and the Gradient Method." *Documenta Mathematica* Extra Volume ISMP:251–54.
- Lewis, Patrick, Ludovic Denoyer, and Sebastian Riedel. 2019. "Unsupervised Question Answering by Cloze Translation." In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 4896–4910. Florence, Italy: Association for Computational Linguistics. <https://doi.org/10.18653/v1/P19-1484>.
- Lewis, Patrick, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, et al. 2021. "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks." *arXiv*.  
<https://doi.org/10.48550/arXiv.2005.11401>.
- Library Hi Tech News. 2023. "Special Issue on ChatGPT." *Library Hi Tech News*. 40 (3).
- Library of Congress. 2022. "G.Pdf." 2022.  
<https://www.loc.gov/aba/publications/FreeLCSH/G.pdf>.
- Lin, Chin-Yew. 2004. "ROUGE: A Package for Automatic Evaluation of Summaries." In *Text Summarization Branches Out*, 74–81. Barcelona, Spain: Association for Computational Linguistics. <https://aclanthology.org/W04-1013>.
- Litsey, Ryan, and Weston Mauldin. 2018. "Knowing What the Patron Wants: Using Predictive Analytics to Transform Library Decision Making." *The Journal of Academic Librarianship* 44 (1): 140–44.  
<https://doi.org/10.1016/j.acalib.2017.09.004>.
- Liu, Rob. 2019. "Review of Weapons of Math Destruction Summary." 2019.  
<https://lifecub.org/books/weapons-of-math-destruction-cathy-oneil-review-summary>.
- Loo, Joshua. 2024. "System 1 and System 2 Thinking." The Decision Lab. 2024.  
<https://thedeclarationlab.com/reference-guide/philosophy/system-1-and-system-2-thinking>.

- Lorang, Elizabeth, Leen-Kiat Soh, Yi Liu, and Chulwoo Pack. 2020. "Digital Libraries, Intelligent Data Analytics, and Augmented Description: A Demonstration Project." *Faculty Publications, UNL Libraries*.  
<https://digitalcommons.unl.edu/libraryscience/396>.
- Lun, H.P. 1958. "The Automatic Creation of Literature Abstracts," 159–65.
- Lund, Brady D., and Ting Wang. 2023. "Chatting about ChatGPT: How May AI and GPT Impact Academia and Libraries?" *Library Hi Tech News* 40 (3): 26–29.  
<https://doi.org/10.1108/LHTN-01-2023-0009>.
- Luong, Thang, Eugene Brevdo, and Rui Zhao. (2017) 2019. "Neural Machine Translation (Seq2seq) Tutorial." Python. tensorflow. <https://github.com/tensorflow/nmt>.
- Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schütze. 2009. "Introduction to Information Retrieval." 2009. <https://nlp.stanford.edu/IR-book/information-retrieval-book.html>.
- Mao, Yuqing, and Zhiyong Lu. 2017. "MeSH Now: Automatic MeSH Indexing at PubMed Scale via Learning to Rank." *Journal of Biomedical Semantics* 8 (April):15. <https://doi.org/10.1186/s13326-017-0123-3>.
- Markowitz, Dale. 2022. "Meet AI's Multitool: Vector Embeddings." Google Cloud Blog. 2022. <https://cloud.google.com/blog/topics/developers-practitioners/meet-ais-multitool-vector-embeddings>.
- McCulloch, Warren, and Walter Pitts. 1943. "A Logical Calculus of the Ideas Immanent in Nervous Activity." *Bulletin of Mathematical Biophysics* 5:115–33.
- McKenzie, Lindsay. 2018. "A New Home for AI: The Library." Inside Higher Ed. 2018. <https://www.insidehighered.com/news/2018/01/17/rhode-island-hopes-putting-artificial-intelligence-lab-library-will-expand-ais-reach>.
- McNeal, Michele L., and David Newyear. 2013. "Chapter 1: Introducing Chatbots in Libraries." *Library Technology Reports* 49 (8): 5–10.  
<https://journals.ala.org/index.php/ltr/article/view/4504>.
- Meszaros, Evan, and Mandi Goodsett. 2022. "Debunking & Prebunking: Strategies for Librarians to Eradicate Misinformation." *American Library Association Annual Conference*. [https://engagedscholarship.csuohio.edu/msl\\_facpub/183](https://engagedscholarship.csuohio.edu/msl_facpub/183).
- Meta. 2023. "Preserving the World's Language Diversity Through AI." *Meta* (blog). 2023. <https://about.fb.com/news/2023/05/ai-massively-multilingual-speech-technology/>.
- . 2024. "Meta Llama." Meta Llama. 2024. <https://llama.meta.com/>.
- Metz, Rachel. 2024. "OpenAI Scale Ranks Progress Toward 'Human-Level' Problem Solving." *Bloomberg.Com*, 2024.  
<https://www.bloomberg.com/news/articles/2024-07-11/openai-sets-levels-to-track-progress-toward-superintelligent-ai>.
- Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. "Efficient Estimation of Word Representations in Vector Space." arXiv.  
<http://arxiv.org/abs/1301.3781>.
- Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. "Distributed Representations of Words and Phrases and Their Compositionality." In *Advances in Neural Information Processing Systems*. Vol. 26. Curran Associates, Inc.  
<https://proceedings.neurips.cc/paper/2013/hash/9aa42b31882ec039965f3c4923ce901b-Abstract.html>.

- Mikolov, Tomas, Wen-tau Yih, and Geoffrey Zweig. 2013. "Linguistic Regularities in Continuous Space Word Representations." In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 746–51. Atlanta, Georgia: Association for Computational Linguistics. <https://aclanthology.org/N13-1090>.
- Mill, John Stuart. 1869. "II. Of the Liberty of Thought and Discussion." In *On Liberty*, edited by John Stuart Mill. London: Longman, Roberts & Green.
- Miller, Johnathan. 2020. "The New Library User: Machine Learning."
- Minsky, Marvin, and Seymour Papert. 1969. *Perceptrons: An Introduction to Computational Geometry*.
- Mishra, Prakhar. 2021. "10 Popular Keyword Extraction Algorithms in Natural Language Processing." *MLearning.Ai* (blog). 2021. <https://medium.com/mllearning-ai/10-popular-keyword-extraction-algorithms-in-natural-language-processing-8975ada5750c>.
- Mittelstadt, Matthew G. 2023. "Artificial Intelligence: An Introduction for Policymakers | Mercatus Center." 2023. <https://www.mercatus.org/research/research-papers/artificial-intelligence-intro-for-policymakers>.
- Mollick, Ethan. 2023. "Working with AI: Two Paths to Prompting." 2023. <https://www.oneusefulthing.org/p/working-with-ai-two-paths-to-prompting>.
- Monigatti, Leonie. 2023. "Getting Started with LangChain: A Beginner's Guide to Building LLM-Powered Applications." Medium. 2023. <https://towardsdatascience.com/getting-started-with-langchain-a-beginners-guide-to-building-llm-powered-applications-95fc8898732c>.
- Moreau, Erwan. 2023. "Literature-Based Discovery: Addressing the Issue of the Subpar Evaluation Methodology." *Bioinformatics* 39 (2). <https://doi.org/10.1093/bioinformatics/btad090>.
- Myint, Leslie, dir. 2020. *Key Structures in Causal Graphs*. <https://www.youtube.com/watch?v=UAovyBnzi9U>.
- Narang, Sharan, and Aakanksha Chowdhery. 2022. "Pathways Language Model (PaLM): Scaling to 540 Billion Parameters for Breakthrough Performance." 2022. <https://ai.googleblog.com/2022/04/pathways-language-model-palm-scaling-to.html>.
- Nawar, Tamer. 2021. "Veritism Refuted? Understanding, Idealization, and the Facts." *Synthese* 198 (5): 4295–4313. <https://doi.org/10.1007/s11229-019-02342-2>.
- Nguyen, Linh Cuong. 2020. "The Impact of Humanoid Robots on Australian Public Libraries." *Journal of the Australian Library and Information Association* 69 (2): 130–48. <https://doi.org/10.1080/24750158.2020.1729515>.
- Nielsen, Michael A. 2015. "Neural Networks and Deep Learning." <http://neuralnetworksanddeeplearning.com>.
- NLP-progress. 2022. "Tracking Progress in Natural Language Processing." NLP-Progress. 2022. <http://nlpprogress.com/>.
- Noble, Safiya. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. 1 edition. New York: NYU Press.
- Nolan, Beatrice. 2022. "Artists Say AI Image Generators Are Copying Their Style to Make Thousands of New Images — and It's Completely out of Their Control."

- Business Insider. 2022. <https://www.businessinsider.com/ai-image-generators-artists-copying-style-thousands-images-2022-10>.
- Nori, Harsha, Samuel Jenkins, Paul Koch, and Rich Caruana. (2019) 2023. “InterpretML.” C++. InterpretML. <https://github.com/interpretml/interpret>.
- Norman, Donald A. 1993. *Things That Make Us Smart: Defending Human Attributes in the Age of the Machine*. Reading, MA: Addison-Wesley.
- NuminaGroup. 2023. “Warehousing Encyclopedia.” NuminaGroup. 2023. <https://numinagroup.com/lp/warehousing-encyclopedia/>.
- Office of Educational Technology. 2023. “Artificial Intelligence and the Future of Teaching and Learning.” Office of Educational Technology. 2023. <https://tech.ed.gov/ai-future-of-teaching-and-learning/>.
- Olson, Hope A. 2000. “Difference, Culture and Change: The Untapped Potential of LCSH.” *Cataloging & Classification Quarterly* 29:53–71.
- . 2002. *The Power to Name: Locating the Limits of Subject Representation in Libraries*. Boston: Kluwer.
- On Large Language Models for Understanding Human Language* Christopher Manning. 2022. <https://www.youtube.com/watch?v=YfXc4OBDmnM>.
- O’Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. 1 edition. New York: Crown.
- , dir. 2018. *The Truth About Algorithms | Cathy O’Neil*. <https://www.youtube.com/watch?v=heQzqX35c9A>.
- OpenAI. 2017. “Proximal Policy Optimization.” OpenAI. 2017. <https://openai.com/blog/openai-baselines-ppo/>.
- . 2022a. “ChatGPT: Optimizing Language Models for Dialogue.” OpenAI. 2022. <https://openai.com/blog/chatgpt/>.
- . 2022b. “Introducing Whisper.” OpenAI. 2022. <https://openai.com/blog/whisper/>.
- . 2022c. “OpenAI.” OpenAI. 2022. <https://openai.com/>.
- . 2023a. “GPT-4 Technical Report.” GPT-4 Technical Report. 2023. <https://cdn.openai.com/papers/gpt-4.pdf>.
- . 2023b. “GPT-4V(Ision) System Card.” 2023. [https://cdn.openai.com/papers/GPTV\\_System\\_Card.pdf](https://cdn.openai.com/papers/GPTV_System_Card.pdf).
- . 2023c. “Guide to Prompt Engineering.” 2023. <https://platform.openai.com/docs/guides/prompt-engineering>.
- . 2023d. “Introducing GPTs.” 2023. <https://openai.com/blog/introducing-gpts>.
- . (2022) 2023. “OpenAI Cookbook.” Jupyter Notebook. OpenAI. <https://github.com/openai/openai-cookbook>.
- . 2024. “Hello GPT-4o.” 2024. <https://openai.com/index/hello-gpt-4o/>.
- OpenAI Platform. 2024. “OpenAI Developer Platform.” 2024. <https://platform.openai.com>.
- Oppy, Graham, and David Dowe. 2021. “The Turing Test.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2021. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/win2021/entriesuring-test/>.
- Ouyang, Long, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, et al. 2022. “Training Language Models to Follow Instructions with Human Feedback.” arXiv. <http://arxiv.org/abs/2203.02155>.

- Padilla, Thomas. 2019. "Responsible Operations: Data Science, Machine Learning, and AI in Libraries."
- Padilla, Thomas, Laurie Allen, Hannah Frost, Sarah Potvin, Elizabeth Russey Roke, and Stewart Varner. 2019. "Always Already Computational: Collections as Data: Final Report." <https://doi.org/10.5281/zenodo.3152935>.
- Pagano, Tiago P., Rafael B. Loureiro, Fernanda V. N. Lisboa, Rodrigo M. Peixoto, Guilherme A. S. Guimarães, Gustavo O. R. Cruz, Maira M. Araujo, et al. 2023. "Bias and Unfairness in Machine Learning Models: A Systematic Review on Datasets, Tools, Fairness Metrics, and Identification and Mitigation Methods." *Big Data and Cognitive Computing* 7 (1): 15. <https://doi.org/10.3390/bdcc7010015>.
- Pahwa, Nitish. 2023. "Silicon Valley's Favorite New Toy Has a Risky Tradeoff." *Slate*, 2023. <https://slate.com/technology/2023/08/chatgpt-ai-arms-race-sustainability.html>.
- Patel, Dwarkesh. 2024. "Francois Chollet, Mike Knoop - LLMs Won't Lead to AGI - \$1,000,000 Prize to Find True Solution." 2024. <https://www.dwarkeshpatel.com/p/francois-chollet>.
- Pearl, Judea. 1995. "Causal Diagrams for Empirical Research." *Biometrika* 82 (4): 669–88. <https://doi.org/10.1093/biomet/82.4.669>.
- . 2009a. "Causal Inference in Statistics : An Overview." *Statistics Surveys* 3:96–146.
- . 2009b. *Causality: Models, Reasoning, and Inference*. 2nd ed. <http://bayes.cs.ucla.edu/BOOK-2K/>.
- Peng, Ciyuan, Feng Xia, Mehdi Naseriparsa, and Francesco Osborne. 2023. "Knowledge Graphs: Opportunities and Challenges." *Artificial Intelligence Review* 56 (11): 13071–102. <https://doi.org/10.1007/s10462-023-10465-9>.
- Peroni, Silvio, and David Shotton. 2012. "FaBiO and CiTO: Ontologies for Describing Bibliographic Resources and Citations." *Journal of Web Semantics* 17 (December):33–43. <https://doi.org/10.1016/j.websem.2012.08.001>.
- Peters, Jay. 2020. "IBM Will No Longer Offer, Develop, or Research Facial Recognition Technology." *The Verge*. 2020. <https://www.theverge.com/2020/6/8/21284683/ibm-no-longer-general-purpose-facial-recognition-analysis-software>.
- Petroni, Fabio, Tim Rocktäschel, Sebastian Riedel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, and Alexander Miller. 2019. "Language Models as Knowledge Bases?" In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2463–73. Hong Kong, China: Association for Computational Linguistics. <https://doi.org/10.18653/v1/D19-1250>.
- Piantadosi, Steven. 2023. "Modern Language Models Refute Chomsky's Approach to Language." *LingBuzz*. <https://lingbuzz.net/lingbuzz/007180>.
- Pichai, Sundar, and Demis Hassabis. 2023. "Introducing Gemini: Our Largest and Most Capable AI Model." *Google*. 2023. <https://blog.google/technology/ai/google-gemini-ai/>.
- Pickering, Ruth, Matthew Ismail, Daniel W. Hook, Simon J. Porter, Catherine Nicole Coleman, Michael A. Keller, James W. Weis, et al. 2022. *Artificial Intelligence in*

- Libraries and Publishing*. Michigan Publishing Services.  
<https://doi.org/10.3998/mpub.12669942>.
- Pierce, John Robinson. 1980. *An Introduction to Information Theory : Symbols, Signals & Noise*. 2nd, rev. ed. New York: Dover Publications.
- Plato. 380AD. "Meno." 380AD. <http://classics.mit.edu/Plato/meno.html>.
- Poincaré, Henri. 1905. *Hypotheses in Physics*. Science and Hypothesis. London: Walter Scott Publishing.
- Popper, Karl R. 1963. *Conjectures and Refutations*. London: Routledge and Kegan Paul.
- . 1968. "Epistemology Without a Knowing Subject." In *Studies in Logic and the Foundations of Mathematics*, 52:333–73. [https://doi.org/10.1016/S0049-237X\(08\)71204-7](https://doi.org/10.1016/S0049-237X(08)71204-7).
- Priem, Jason. 2013. "Beyond the Paper." *Nature* 495 (7442): 437–40.  
<https://doi.org/10.1038/495437a>.
- Pritchard, Duncan, John Turri, and J. Adam Carter. 2022. "The Value of Knowledge." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman, Fall 2022. Metaphysics Research Lab, Stanford University.  
<https://plato.stanford.edu/archives/fall2022/entries/knowledge-value/>.
- "Project Implicit." 2011. 2011. <https://implicit.harvard.edu/implicit/index.jsp>.
- Pushkin, Alexander. 1881. "Eugene Onegin." 1881.  
<https://www.gutenberg.org/files/23997/23997-h/23997-h.htm>.
- Rabiner, L.R. 1989. "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition." *Proceedings of the IEEE* 77 (2): 257–86.  
<https://doi.org/10.1109/5.18626>.
- Race, Technology, and Algorithmic Bias*. 2019.  
<https://www.radcliffe.harvard.edu/video/race-technology-and-algorithmic-bias-vision-justice>.
- Rainie, Lee, and Janna Anderson. 2017. "Code-Dependent: Pros and Cons of the Algorithm Age." *Pew Research Center: Internet, Science & Tech* (blog). 2017.  
<https://www.pewresearch.org/internet/2017/02/08/code-dependent-pros-and-cons-of-the-algorithm-age/>.
- READ-COOP. 2021. "About Us." READ-COOP. 2021. <https://readcoop.eu/about/>.
- Redi, Miriam. 2018. "How We're Using Machine Learning to Visually Enrich Wikidata." Wikimedia Foundation. 2018.  
<https://wikimediafoundation.org/news/2018/03/14/machine-learning-visually-enriching-wikidata/>.
- Rees, David C., Thomas N. Williams, and Mark T. Gladwin. 2010. "Sickle-Cell Disease." *The Lancet* 376 (9757): 2018–31. [https://doi.org/10.1016/S0140-6736\(10\)61029-X](https://doi.org/10.1016/S0140-6736(10)61029-X).
- Reimers, Nils, and Jay Alammr. 2023. "The Embedding Archives: Millions of Wikipedia Article Embeddings in Many Languages." Context by Cohere. 2023.  
<https://txt.cohere.com/embedding-archives-wikipedia/>.
- Reitz, Joan M. 2014. "ODLIS P." Online Dictionary for Library and Information Science. 2014. [https://odlis.abc-clio.com/odlis\\_p.html](https://odlis.abc-clio.com/odlis_p.html).
- Reuters. 2016. "New Zealand Passport Robot Tells Applicant of Asian Descent to Open Eyes." *Reuters*, 2016, sec. Technology News.  
<https://www.reuters.com/article/us-newzealand-passport-error-idUSKBN13WoRL>.

- Rhody Today. 2017. “URI to Launch Artificial Intelligence Lab.” 2017.  
<https://www.uri.edu/news/2017/12/uri-to-launch-artificial-intelligence-lab/>.
- Ridley, Michael, and Danica Pawlick-Potts. 2021a. “Algorithmic Literacy and the Role for Libraries.” *Information Technology and Libraries* 40 (2).  
<https://doi.org/10.6017/ital.v40i2.12963>.
- . 2021b. “Algorithmic Literacy and the Role for Libraries.” *Information Technology and Libraries* 40 (2). <https://doi.org/10.6017/ital.v40i2.12963>.
- Roberts, David Lindsay. 2019. *Republic of Numbers*.  
<https://doi.org/10.1353/book.67892>.
- Rohrer, Julia M. 2018. “Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data.” *Advances in Methods and Practices in Psychological Science* 1 (1): 27–42. <https://doi.org/10.1177/2515245917745629>.
- Rolan, Gregory, Glen Humphries, Lisa Jeffrey, Evanthia Samaras, Tatiana Antsouпова, and Katharine Stuart. 2019. “More Human than Human? Artificial Intelligence in the Archive.” *Archives and Manuscripts* 47 (2): 179–203.  
<https://doi.org/10.1080/01576895.2018.1502088>.
- Romero, Alberto. 2021. “GPT-3 — A Complete Overview.” Medium. 2021.  
<https://towardsdatascience.com/gpt-3-a-complete-overview-190232eb25fd>.
- Rosenblatt, Frank. 1957. “The Perceptron A Perceiving and Recognizing Automaton (Project Para).” <https://blogs.umass.edu/brain-wars/files/2016/03/rosenblatt-1957.pdf>.
- . 1958. “The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain.” *Psychological Review* 65 (6): 386–408.  
<https://doi.org/10.1037/h0042519>.
- Rosenblatt, Lucas, and R. Teal Witter. 2022. “Counterfactual Fairness Is Basically Demographic Parity.” arXiv. <https://doi.org/10.48550/arXiv.2208.03843>.
- Rosenfeld, Louis, Peter Morville, and Jorge Arango. 2015. *Information Architecture: For the Web and Beyond*. O’Reilly Media, Inc.
- Rowley, Jennifer. 2000. *Organising Knowledge: An Introduction to Managing Access to Information*. 3rd ed. Burlington, VT: Gower.
- Roy, Kaushik, Vedant Khandelwal, Harshul Surana, Valerie Vera, Amit Sheth, and Heather Heckman. 2023. “GEAR-Up: Generative AI and External Knowledge-Based Retrieval Upgrading Scholarly Article Searches for Systematic Reviews.” arXiv. <http://arxiv.org/abs/2312.09948>.
- Rubenstein, Paul K., Chulayuth Asawaroengchai, Duc Dung Nguyen, Ankur Bapna, Zalán Borsos, Félix de Chaumont Quitry, Peter Chen, et al. 2023. “AudioPaLM: A Large Language Model That Can Speak and Listen.” arXiv.  
<https://doi.org/10.48550/arXiv.2306.12925>.
- Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. 1986. “Learning Representations by Back-Propagating Errors.” *Nature* 323 (6088): 533–36.  
<https://doi.org/10.1038/323533a0>.
- Russell, Bertrand. 1912. *The Problems of Philosophy*.  
<https://www.gutenberg.org/files/5827/5827-h/5827-h.htm>.
- Rutkowski, Greg. 2023. “Greg Rutkowski: Artist.” 2023.  
<https://rutkowski.artstation.com/>.
- Sadeh, Tamar. 2015. “From Search to Discovery.” *Bibliothek Forschung Und Praxis* 39 (2): 212–24. <https://doi.org/10.1515/bfp-2015-0028>.

- Samuel, Arthur L. 1959. "Eight-Move Opening Utilizing Generalization Learning. (See Appendix B, Game G-43.1 Some Studies in Machine Learning Using the Game of Checkers. IBM Journal, 210–229." In .
- Sanderson, Grant, and 3Blue1Brown, dirs. 2017a. *But What Is a Neural Network? | Chapter 1, Deep Learning*. <https://www.youtube.com/watch?v=aircAruvnKk>.
- , dirs. 2017b. *Gradient Descent, How Neural Networks Learn | Chapter 2, Deep Learning*. <https://www.youtube.com/watch?v=IHZwWFHWa-w>.
- Sanji, Majideh, Hassan Behzadi, and Gisu Gomroki. 2022. "Chatbot: An Intelligent Tool for Libraries." *Library Hi Tech News* ahead-of-print. <https://doi.org/10.1108/LHTN-01-2021-0002>.
- Saravia, Elvis. 2023. "Prompt Engineering Guide – Nextra." 2023. <https://www.promptingguide.ai/>.
- Scheines, Richard. 1997. "An Introduction to Causal Inference." In *Causality in Crisis?*, 185–200. University of Notre Dame.
- Schlicht, Matt, and Ben Parr. 2023. "Chatbots Magazine: The #1 Place to Learn about Chatbots." Chatbots Magazine. 2023. <https://chatbotsmagazine.com/>.
- Sebastian, Yakub, Eu-Gen Siew, and Sylvester O. Orimaye. 2017. "Emerging Approaches in Literature-Based Discovery: Techniques and Performance Review." *The Knowledge Engineering Review* 32:e12. <https://doi.org/10.1017/S0269888917000042>.
- Seff, Ari, dir. 2023. *How ChatGPT Is Trained*. <https://www.youtube.com/watch?v=VPRSBzXzavo>.
- Shannon, C E. 1948. "A Mathematical Theory of Communication." *The Bell System Technical Journal* 27:379–423, 623–56.
- Shannon, Claude Elwood, and Warren Weaver. 1949. *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.
- Singhal, Amit. 2012. "Introducing the Knowledge Graph: Things, Not Strings." Google. 2012. <https://blog.google/products/search/introducing-knowledge-graph-things-not/>.
- Smalheiser, Neil R. 2017. "Rediscovering Don Swanson: The Past, Present and Future of Literature-Based Discovery." *Journal of Data and Information Science (Warsaw, Poland)* 2 (4): 43–64. <https://doi.org/10.1515/jdis-2017-0019>.
- Smith, Linda C. 1981. "Citation Analysis." *Library Trends* 30:83–106.
- Smith, Martha Montague. 1997. "Information Ethics." In *Annual Review of Information Science and Technology*, 32:339–66.
- Snow, Karen. 2017. "Defining, Assessing, and Rethinking Quality Cataloging." *Cataloging & Classification Quarterly* 55 (7–8): 438–55. <https://doi.org/10.1080/01639374.2017.1350774>.
- Society of American Archivists. 2020. "SAA Core Values Statement and Code of Ethics | Society of American Archivists." 2020. <https://www2.archivists.org/statements/saa-core-values-statement-and-code-of-ethics>.
- Somers, James. 2017. "Torching the Modern-Day Library of Alexandria." The Atlantic. 2017. <https://www.theatlantic.com/technology/archive/2017/04/the-tragedy-of-google-books/523320/>.
- Spivak, Nova, and Nick Slavin. 2023. "Arch Mission Foundation." Arch Mission. 2023. <https://archmission.org/>.

- Stamp, Mark. 2017. "A Revealing Introduction to Hidden Markov Models." In *Introduction to Machine Learning with Applications in Information Security*, by Mark Stamp, 1st ed., 7–35. Chapman and Hall/CRC.  
<https://doi.org/10.1201/9781315213262-2>.
- Stanford HAI. 2023a. "AI Will Transform Teaching and Learning. Let's Get It Right." Stanford HAI. 2023. <https://hai.stanford.edu/news/ai-will-transform-teaching-and-learning-lets-get-it-right>.
- . 2023b. "Generative AI: Perspectives from Stanford HAI." 2023. [https://hai.stanford.edu/sites/default/files/2023-03/Generative\\_AI\\_HAI\\_Perspectives.pdf](https://hai.stanford.edu/sites/default/files/2023-03/Generative_AI_HAI_Perspectives.pdf).
- Starmer, Josh, dir. 2020. *Neural Networks Pt. 2: Backpropagation Main Ideas*. <https://www.youtube.com/watch?v=IN2XmBhILt4>.
- Starmer, Josh, and StatQuest, dirs. 2019. *Gradient Descent, Step-by-Step*. <https://www.youtube.com/watch?v=sDv4f4s2SB8>.
- Stevens, Michael. 2013. "Looking into the Black Box." Opinionator. 2013. <https://archive.nytimes.com/opinionator.blogs.nytimes.com/2013/11/24/looking-into-the-black-box/>.
- Suber, Peter. 1999. "Paternalism." 1999. [https://dash.harvard.edu/bitstream/handle/1/4725017/suber\\_paternal.htm](https://dash.harvard.edu/bitstream/handle/1/4725017/suber_paternal.htm).
- Sueiras, Jorge. 2021. "Continuous Offline Handwriting Recognition Using Deep Learning Models." arXiv. <https://doi.org/10.48550/arXiv.2112.13328>.
- Sulmont, Elisabeth, Elizabeth Patitsas, and Jeremy R. Cooperstock. 2019. "Can You Teach Me To Machine Learn?" In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*, 948–54. SIGCSE '19. New York, NY, USA: Association for Computing Machinery.  
<https://doi.org/10.1145/3287324.3287392>.
- Suresh, Harini, and John Gutttag. 2021. "A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle." In *Equity and Access in Algorithms, Mechanisms, and Optimization*, 1–9. -- NY USA: ACM.  
<https://doi.org/10.1145/3465416.3483305>.
- Sutskever, Ilya, Oriol Vinyals, and Quoc V Le. 2014. "Sequence to Sequence Learning with Neural Networks." In *Advances in Neural Information Processing Systems*. Vol. 27. Curran Associates, Inc.  
<https://proceedings.neurips.cc/paper/2014/hash/a14ac55a4f27472c5d894ec1c3c743d2-Abstract.html>.
- Sutton, Richard S., and Andrew G. Barto. 2018. *Reinforcement Learning: An Introduction*. 2nd ed. <http://incompleteideas.net/book/the-book-2nd.html>.
- Svenonius, Elaine. 1969. Review of *Review of Two Kinds of Power: An Essay on Bibliographical Control*, by Patrick Wilson. *The Library Quarterly: Information, Community, Policy* 39 (1): 112–14. <http://www.jstor.org/stable/4305960>.
- . 2003. "Design of Controlled Vocabularies." In *Encyclopedia of Library and Information Science*, 822–38. New York: Marcel Dekker.
- Swanson, D. R. 1986. "Undiscovered Public Knowledge." *Library Quarterly* 56:103–18.
- Tait, Elizabeth, and Cameron M Pierson. 2022. "Artificial Intelligence and Robots in Libraries: Opportunities in LIS Curriculum for Preparing the Librarians of Tomorrow." *Journal of the Australian Library and Information Association* 71 (3): 256–74. <https://doi.org/10.1080/24750158.2022.2081111>.

- Tashea, Jason. 2017. "Courts Are Using AI to Sentence Criminals. That Must Stop Now." *Wired*, 2017. <https://www.wired.com/2017/04/courts-using-ai-sentence-criminals-must-stop-now/>.
- Tay, Aaron. 2022. "List of Innovative Literature Mapping Tools | Aaron Tay's Musings about Librarianship." 2022. <https://musingsaboutlibrarianship.blogspot.com/p/list-of-innovative-literature-mapping.html?view=classic>.
- Taylor, Arlene G. 2004. *The Organization of Information*. 2nd ed. Westport, Conn: Libraries Unlimited.
- Taylor, Ross, Marcin Kardas, Guillem Cucurull, Thomas Scialom, Anthony Hartshorn, Elvis Saravia, Andrew Poulton, Viktor Kerkez, and Robert Stojnic. 2022. "Galactica: A Large Language Model for Science." arXiv. <https://doi.org/10.48550/arXiv.2211.09085>.
- Taylor, Wilson L. 1953. "'Cloze Procedure': A New Tool for Measuring Readability." *Journalism Quarterly* 30 (4): 415–33. <https://doi.org/10.1177/107769905303000401>.
- Teixeira, Lawrence. 2023. "The New Open AI GPT-4 Vision on ChatGPT: Bridging the Gap Between Text and Image Understanding." 2023. <https://medium.com/@lawrenceteixeira/the-new-open-ai-gpt-4-vision-on-chatgpt-bridging-the-gap-between-text-and-image-understanding-9337ed4c1a61>.
- Tella, Adeyinka. 2020. "Robots Are Coming to the Libraries: Are Librarians Ready to Accommodate Them?" *Library Hi Tech News* 37 (8): 13–17. <https://doi.org/10.1108/LHTN-05-2020-0047>.
- Tella, Adeyinka, and Yusuf Ayodeji Ajani. 2022. "Robots and Public Libraries." *Library Hi Tech News* 39 (7): 15–18. <https://doi.org/10.1108/LHTN-05-2022-0072>.
- Tenney, Ian, Patrick Xia, Berlin Chen, Alex Wang, Adam Poliak, R. Thomas McCoy, Najoung Kim, et al. 2022. "What Do You Learn from Context? Probing for Sentence Structure in Contextualized Word Representations." In . <https://openreview.net/forum?id=SJzSgnRcKX>.
- The AI Advantage, dir. 2023. *100+ Insane ChatGPT Vision Use Cases*. <https://www.youtube.com/watch?v=ywNNRzc7-To>.
- Thilakarathne, Menasha, Katrina Falkner, and Thushari Atapattu. 2020. "A Systematic Review on Literature-Based Discovery: General Overview, Methodology, & Statistical Analysis." *ACM Computing Surveys* 52 (6): 1–34. <https://doi.org/10.1145/3365756>.
- Thoppilan, Romal, Daniel De Freitas, Jamie Hall, Noam Shazeer, Apoorv Kulshreshtha, Heng-Tze Cheng, Alicia Jin, et al. 2022. "LaMDA: Language Models for Dialog Applications." arXiv. <https://doi.org/10.48550/arXiv.2201.08239>.
- Thunström, Almira Osmanovic. 2022. "We Asked GPT-3 to Write an Academic Paper about Itself—Then We Tried to Get It Published." *Scientific American*. 2022. <https://www.scientificamerican.com/article/we-asked-gpt-3-to-write-an-academic-paper-about-itself-then-we-tried-to-get-it-published/>.
- Turner, Ash. 2018. "How Many People Have Smartphones Worldwide (Oct 2022)." 2018. <https://www.bankmycell.com/blog/how-many-phones-are-in-the-world>.
- Tversky, A. 1974. "Judgments under Uncertainty: Heuristics and Biases." *Science* 185:1124-1131.

- Tversky, Amos, and Daniel Kahneman. 1982. "Evidential Impact of Base Rates." In *Judgement under Uncertainty: Heuristics and Biases*, edited by Daniel Kahneman, A Slovic, and Amos Tversky. Cambridge University Press.
- UCL. 2018. "Transcribe Bentham." Bentham Project. 2018. <https://www.ucl.ac.uk/bentham-project/transcribe-bentham>.
- UCSF Office of Diversity and Outreach UCSF. 2022. "Unconscious Bias Training." University of California: Office of Diversity and Outreach. 2022. <https://diversity.ucsf.edu/programs-resources/training/unconscious-bias-training>.
- ujet.cx. 2022a. "Critical State of Automation in Customer Experience." UJET. 2022. <https://ujet.cx/resources/reports/critical-state-of-automation-customer-experience-2022-report-lp>.
- . 2022b. "UJET Research Reveals Chatbots Increase Frustration for 80% of Consumers." UJET. 2022. <https://ujet.cx/press-releases/ujet-research-reveals-chatbots-increase-frustration>.
- University of Alberta Library. 2023. "Evidence Based Library and Information Practice." 2023. <https://journals.library.ualberta.ca/eblip/index.php/EBLIP>.
- Uszkoreit, Jakob. 2017. "Transformer: A Novel Neural Network Architecture for Language Understanding." 2017. <https://ai.googleblog.com/2017/08/transformer-novel-neural-network.html>.
- VanderWeele, Tyler J., and Nancy Staudt. 2011. "Causal Diagrams for Empirical Legal Research: A Methodology for Identifying Causation, Avoiding Bias and Interpreting Results." *Law, Probability & Risk: A Journal of Reasoning under Uncertainty* 10 (4): 329–54. <https://doi.org/10.1093/lpr/mgr019>.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. "Attention Is All You Need." arXiv. <https://doi.org/10.48550/arXiv.1706.03762>.
- Vincze, Joseph. 2017. "Virtual Reference Librarians (Chatbots)." *Library Hi Tech News* 34 (4): 5–8. <https://doi.org/10.1108/LHTN-03-2017-0016>.
- Von Hilgers, Philipp, and Amy M. Langville. 2006. "The Five Greatest Applications of Markov Chains." 2006. <http://langvillea.people.cofc.edu/MCapps7.pdf>.
- W3C Working Group. 2014. "RDF 1.1 Primer." 2014. <https://www.w3.org/TR/rdf11-primer/>.
- Wang, Angelina, Solon Barocas, Kristen Laird, and Hanna Wallach. 2022. "Measuring Representational Harms in Image Captioning." In *2022 ACM Conference on Fairness, Accountability, and Transparency*, 324–35. FAccT '22. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3531146.3533099>.
- Wattenberg, Martin, Fernanda Viégas, and Moritz Hardt. 2022. "Attack Discrimination with Smarter Machine Learning." 2022. <https://research.google.com/bigpicture/attacking-discrimination-in-ml/>.
- Wei, Jason, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. 2023. "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models." arXiv. <https://doi.org/10.48550/arXiv.2201.11903>.

- Wei, and Denny Zhou. 2022. “Language Models Perform Reasoning via Chain of Thought.” 2022. <https://ai.googleblog.com/2022/05/language-models-perform-reasoning-via.html>.
- Weidinger, Laura, Jonathan Uesato, Maribeth Rauh, Conor Griffin, Po-Sen Huang, John Mellor, Amelia Glaese, et al. 2022. “Taxonomy of Risks Posed by Language Models.” In *2022 ACM Conference on Fairness, Accountability, and Transparency*, 214–29. FAccT ’22. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3531146.3533088>.
- Weigert, Verena. 2020. “Chatbots in Libraries – Library Services.” 2020. <https://libraryservices.jiscinvolve.org/wp/2020/09/chatbots-in-libraries/>.
- Weizenbaum, Joseph. 1966. “ELIZA—a Computer Program for the Study of Natural Language Communication between Man and Machine.” *Communications of the ACM* 9 (1): 36–45. <https://doi.org/10.1145/365153.365168>.
- Wells, David. 2021. “Online Public Access Catalogues and Library Discovery Systems.” Text. 2021. <https://www.isko.org/cyclo/opac#6>.
- Wenar, Leif. 2021. “Rights.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2021. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/spr2021/entries/rights/>.
- Wheatley, Amanda, and Sandy Hervieux. 2019. “Artificial Intelligence in Academic Libraries: An Environmental Scan.” *Information Services & Use* 39 (4): 347–56. <https://doi.org/10.3233/ISU-190065>.
- Widyassari, Adhika Pramita, Supriadi Rustad, Guruh Fajar Shidik, Edi Noersasongko, Abdul Syukur, Affandy Affandy, and De Rosal Ignatius Moses Setiadi. 2020. “Review of Automatic Text Summarization Techniques & Methods.” *Journal of King Saud University - Computer and Information Sciences*. <https://doi.org/10.1016/j.jksuci.2020.05.006>.
- Wikipedia. 2022a. “Algorithm.” In *Wikipedia*. <https://en.wikipedia.org/w/index.php?title=Algorithm>.
- . 2022b. “Algorithmic Bias.” In *Wikipedia*. [https://en.wikipedia.org/w/index.php?title=Algorithmic\\_bias](https://en.wikipedia.org/w/index.php?title=Algorithmic_bias).
- . 2022c. “Artificial Linguistic Internet Computer Entity.” In *Wikipedia*. [https://en.wikipedia.org/w/index.php?title=Artificial\\_Linguistic\\_Internet\\_Computer\\_Entity&oldid=1112755868](https://en.wikipedia.org/w/index.php?title=Artificial_Linguistic_Internet_Computer_Entity&oldid=1112755868).
- . 2022d. “False Positives and False Negatives.” In *Wikipedia*. [https://en.wikipedia.org/w/index.php?title=False\\_positives\\_and\\_false\\_negatives&oldid=1088158900](https://en.wikipedia.org/w/index.php?title=False_positives_and_false_negatives&oldid=1088158900).
- . 2022e. “LaMDA.” In *Wikipedia*. <https://en.wikipedia.org/w/index.php?title=LaMDA&oldid=1103832671>.
- . 2022f. “Vatican Apostolic Archive.” In *Wikipedia*. [https://en.wikipedia.org/w/index.php?title=Vatican\\_Apostolic\\_Archive&oldid=1095541005](https://en.wikipedia.org/w/index.php?title=Vatican_Apostolic_Archive&oldid=1095541005).
- . 2023a. “CAPTCHA.” In *Wikipedia*. <https://en.wikipedia.org/w/index.php?title=CAPTCHA>.
- . 2023b. “Confirmation Bias.” In *Wikipedia*. [https://en.wikipedia.org/w/index.php?title=Confirmation\\_bias&oldid=1186068624](https://en.wikipedia.org/w/index.php?title=Confirmation_bias&oldid=1186068624).

- . 2023c. “Crowdsourcing.” In *Wikipedia*.  
<https://en.wikipedia.org/w/index.php?title=Crowdsourcing&oldid=1131878886>.
  - . 2023d. “Evidence-Based Practice.” In *Wikipedia*.  
[https://en.wikipedia.org/w/index.php?title=Evidence-based\\_practice&oldid=1141797027](https://en.wikipedia.org/w/index.php?title=Evidence-based_practice&oldid=1141797027).
  - . 2023e. “Explainable Artificial Intelligence.” In *Wikipedia*.  
[https://en.wikipedia.org/w/index.php?title=Explainable\\_artificial\\_intelligence&oldid=1144112716](https://en.wikipedia.org/w/index.php?title=Explainable_artificial_intelligence&oldid=1144112716).
  - . 2023f. “Google Hummingbird.” In *Wikipedia*.  
[https://en.wikipedia.org/w/index.php?title=Google\\_Hummingbird&oldid=1152805833](https://en.wikipedia.org/w/index.php?title=Google_Hummingbird&oldid=1152805833).
  - . 2023g. “LangChain.” In *Wikipedia*.  
<https://en.wikipedia.org/w/index.php?title=LangChain>.
  - . 2023h. “Liberty Leading the People.” In *Wikipedia*.  
[https://en.wikipedia.org/w/index.php?title=Liberty\\_Leading\\_the\\_People&oldid=1158901762](https://en.wikipedia.org/w/index.php?title=Liberty_Leading_the_People&oldid=1158901762).
  - . 2023i. “Library.” In *Wikipedia*.  
<https://en.wikipedia.org/w/index.php?title=Library&oldid=1140601084>.
  - . 2023j. “Literature-Based Discovery.” In *Wikipedia*.  
[https://en.wikipedia.org/w/index.php?title=Literature-based\\_discovery&oldid=1140927264](https://en.wikipedia.org/w/index.php?title=Literature-based_discovery&oldid=1140927264).
  - . 2023k. “Open Archives Initiative Protocol for Metadata Harvesting.” In *Wikipedia*.  
[https://en.wikipedia.org/w/index.php?title=Open\\_Archives\\_Initiative\\_Protocol\\_for\\_Metadata\\_Harvesting&oldid=1133365780](https://en.wikipedia.org/w/index.php?title=Open_Archives_Initiative_Protocol_for_Metadata_Harvesting&oldid=1133365780).
  - . 2023l. “PageRank.” In *Wikipedia*.  
<https://en.wikipedia.org/w/index.php?title=PageRank&oldid=1165448388>.
  - . 2023m. “RankBrain.” In *Wikipedia*.  
<https://en.wikipedia.org/w/index.php?title=RankBrain&oldid=1140034006>.
  - . 2023n. “Tay (Chatbot).” In *Wikipedia*.  
[https://en.wikipedia.org/w/index.php?title=Tay\\_\(chatbot\)](https://en.wikipedia.org/w/index.php?title=Tay_(chatbot)).
  - . 2023o. “The Library of Babel.” In *Wikipedia*.  
[https://en.wikipedia.org/w/index.php?title=The\\_Library\\_of\\_Babel&oldid=1141372445](https://en.wikipedia.org/w/index.php?title=The_Library_of_Babel&oldid=1141372445).
  - . 2023p. “Wikipedia:Artificial Intelligence.” In *Wikipedia*.  
[https://en.wikipedia.org/w/index.php?title=Wikipedia:Artificial\\_intelligence&oldid=1157714616](https://en.wikipedia.org/w/index.php?title=Wikipedia:Artificial_intelligence&oldid=1157714616).
  - . 2024. “Her.” In *Wikipedia*.  
[https://en.wikipedia.org/w/index.php?title=Her\\_\(film\)&oldid=1237911810](https://en.wikipedia.org/w/index.php?title=Her_(film)&oldid=1237911810).
- Wilburn, Brad. 1999. “Spice for the Good Life.” *Issues in Ethics* 10.
- Wilson, P. 1968. *Two Kinds of Power: An Essay on Bibliographical Control*. Berkeley: University of California Press.
- Wolfe, Matt. 2023. “Future Tools - Find The Exact AI Tool For Your Needs.” 2023.  
<https://www.futuretools.io/>.
- Wolford, Ben. 2018. “What Is GDPR, the EU’s New Data Protection Law?” GDPR.Eu. 2018. <https://gdpr.eu/what-is-gdpr/>.

- Wolfram. 2023a. "Wolfram Mathematica: Modern Technical Computing." 2023. <https://www.wolfram.com/mathematica/>.
- Wolfram, Stephen. 2023b. "ChatGPT Gets Its 'Wolfram Superpowers!'" 2023. <https://writings.stephenwolfram.com/2023/03/chatgpt-gets-its-wolfram-superpowers/>.
- Wolkoff, K. N. 1996. "The Problem of Holocaust Denial Literature in Libraries." *Library Trends* 45:87–96.
- Wood, Barbara, and David Evans. 2018. "Librarians' Perceptions of Artificial Intelligence and Its Potential Impact on the Profession." *Computers in Libraries* 38 (1). [https://www.researchgate.net/publication/322977069\\_Librarians'\\_Perceptions\\_of\\_Artificial\\_Intelligence\\_and\\_Its\\_Potential\\_Impact\\_on\\_the\\_Profession](https://www.researchgate.net/publication/322977069_Librarians'_Perceptions_of_Artificial_Intelligence_and_Its_Potential_Impact_on_the_Profession).
- Writer, Beta. 2019. *Lithium-Ion Batteries*. <https://link.springer.com/book/10.1007/978-3-030-16800-1>.
- Xiang, Chloe. 2023. "OpenAI's GPT-4 Is Closed Source and Shrouded in Secrecy." Vice. 2023. <https://www.vice.com/en/article/ak3w5a/openais-gpt-4-is-closed-source-and-shrouded-in-secrecy>.
- Xu, Binfeng, Zhiyuan Peng, Bowen Lei, Subhabrata Mukherjee, Yuchen Liu, and Dongkuan Xu. 2023. "ReWOO: Decoupling Reasoning from Observations for Efficient Augmented Language Models." arXiv. <https://doi.org/10.48550/arXiv.2305.18323>.
- Yang, Zhengyuan, Linjie Li, Kevin Lin, Jianfeng Wang, Chung-Ching Lin, Zicheng Liu, and Lijuan Wang. 2023. "The Dawn of LMMs: Preliminary Explorations with GPT-4V(Ision)." arXiv. <https://doi.org/10.48550/arXiv.2309.17421>.
- Zellers, Rowan, Ari Holtzman, Yonatan Bisk, Ali Farhadi, and Yejin Choi. 2019. "HellaSwag: Can a Machine Really Finish Your Sentence?" arXiv. <http://arxiv.org/abs/1905.07830>.
- Zeng, Marcia Lei. 2005. "Construction of Controlled Vocabularies, A Primer (Based on Z39.19)." 2005.
- Zewe, Adam. 2023. "Solving a Machine-Learning Mystery." MIT News | Massachusetts Institute of Technology. 2023. <https://news.mit.edu/2023/large-language-models-in-context-learning-0207>.
- Zheng, Lianmin, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, et al. 2023. "Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena." arXiv. <https://doi.org/10.48550/arXiv.2306.05685>.
- Ziegler, S. L. 2019. "Digitization Selection Criteria as Anti-Racist Action." *The Code4Lib Journal*, no. 45. <https://journal.code4lib.org/articles/14667>.