Commonsense AI in the History of the Web

Citation: Mayank Kejriwal, Deborah L. McGuinness, and Henry Lieberman. 2025. Commonsense AI in the History of the Web. In Companion Proceedings of the ACM on Web Conference 2025 (WWW '25). Association for Computing Machinery, New York, NY, USA, 837–840.

As Published: https://doi.org/10.1145/3701716.3716841

Publisher: ACM|Companion Proceedings of the ACM Web Conference 2025

Persistent URL: https://hdl.handle.net/1721.1/162582

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Commonsense AI in the History of the Web

Mayank Kejriwal
kejriwal@isi.edu
University of Southern California
Los Angeles, California, USA

Deborah L. McGuinness
dlm@rpi.edu
Rensselaer Polytechnic Institute
Troy, New York, USA

Henry Lieberman
henry@csail.mit.edu
Massachusetts Institute of Technology
Cambridge, Massachusetts, USA

Abstract

Machine common sense (MCS)—the challenge of enabling computers to grasp everyday human knowledge—has been a grand challenge in Artificial Intelligence (AI) since the 1950s. While recent advances in large language models have led to impressive progress, there is still no consensus on how much common sense today's AI actually possesses. In this brief review, we revisit the historical development of MCS in the context of the Web, examining how the Web's evolution—from early knowledge representation efforts to knowledge graphs, the Semantic Web, and crowdsourcing—has shaped MCS research. We argue that key breakthroughs in Web technologies were instrumental in addressing longstanding challenges of scale and coverage in commonsense reasoning. At the same time, MCS research has influenced the development of core Web applications, including intelligent agents, plausibility-based reasoning, and robust evaluation of black-box AI systems.

CCS Concepts

• **Information systems** → **Web mining**; **Crowdsourcing**; • **Computing methodologies** → *Natural language processing*; **Knowledge representation and reasoning**; **Philosophical/theoretical foundations of artificial intelligence**.

Keywords

Machine Common Sense, LLMs, ConceptNet, Cyc

ACM Reference Format:
Mayank Kejriwal, Deborah L. McGuinness, and Henry Lieberman. 2025. Commonsense AI in the History of the Web. In *Companion Proceedings of the ACM Web Conference 2025 (WWW Companion '25), April 28-May 2, 2025, Sydney, NSW, Australia.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3701716.3716841

1 Introduction

Since the Dartmouth Summer Research Project in Artificial Intelligence (often referred to as the Dartmouth Workshop) in the mid-1950s [19], developing *agentic* systems that possess general human abilities like social reasoning, robust visual processing, and conversational naturalism, has achieved the status of a grand challenge. An important milestone that needs to be achieved to fulfill such a broad vision of artificial "general" intelligence is *machine common sense* (MCS). This challenge was true of the agentic systems envisioned in the 50s and is even more true today for the interacting agentic components on the web today.

ACM ISBN 979-8-4007-1331-6/2025/04
https://doi.org/10.1145/3701716.3716841

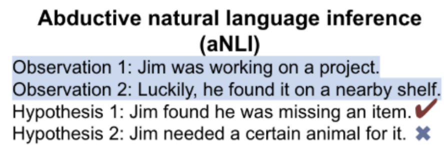

Figure 1: An example of a multiple-choice test instance from an MCS benchmark (aNLI). Given two observations as input (in blue), a model must choose from the most likely common sense hypothesis from the two provided here.

But what do we mean when we talk about *common sense*? Airtight 'formal' definitions of common sense have always been elusive, but it is minimally considered to be a set of abilities that form an essential part of what makes human intelligence so flexible, in spite of not having a clearly defined boundary. Recent progress in neural language modeling [6, 31], including the large language models (LLMs) like GPT-4, have brought exciting and provocative new progress that few thought possible before. Most researchers accept that there has been *some* progress on MCS, but there is practical uncertainty on the true state of this progress, and disagreement on how much distance is still left to cover [4].

There is ample reason for researchers and practitioners with an interest in the Web to care about MCS. Even with recent progress in LLMs, research on building agents with general-purpose reasoning ability – particularly, the ability to reason about novel problems for which not a lot (if any) solutions have been statistically observed – is still in its infancy, and some evidence suggests may be hitting a performance cliff. Even if we ignore the uneven performance, one reason it is so difficult to test the capabilities of LLMs is due to the nature of multiple-choice tests (see Figure 1), which continue to be the *de facto* mechanism for evaluating LLMs and other systems on MCS [5]. Such a test, just like a college multiple-choice final, is essentially a sampling procedure on the *space of the capabilities* of the agent it is trying to evaluate. For example, if a student is taking a chemistry test, and can answer a difficult question about a particular chemical reaction, we assume that they can also answer questions about similar substances in closely related reactions.

Evidence suggests that this assumption may not apply very well for LLMs trained largely on Web content. It is not difficult to formulate a set of questions that people will judge very similar to each other, but that get wildly different answers from LLMs. Statistics tells us that a small set of samples is only representative of the entire space if the space itself has certain continuity properties. In some tests, it appears that the space of capabilities for many LLMs does not have reliable continuity. However, the overwhelming focus on LLMs also obscures progress that has been made over the decades, many of which are concurrent with breakthroughs in Web technology and research. Bearing in mind this important context, this

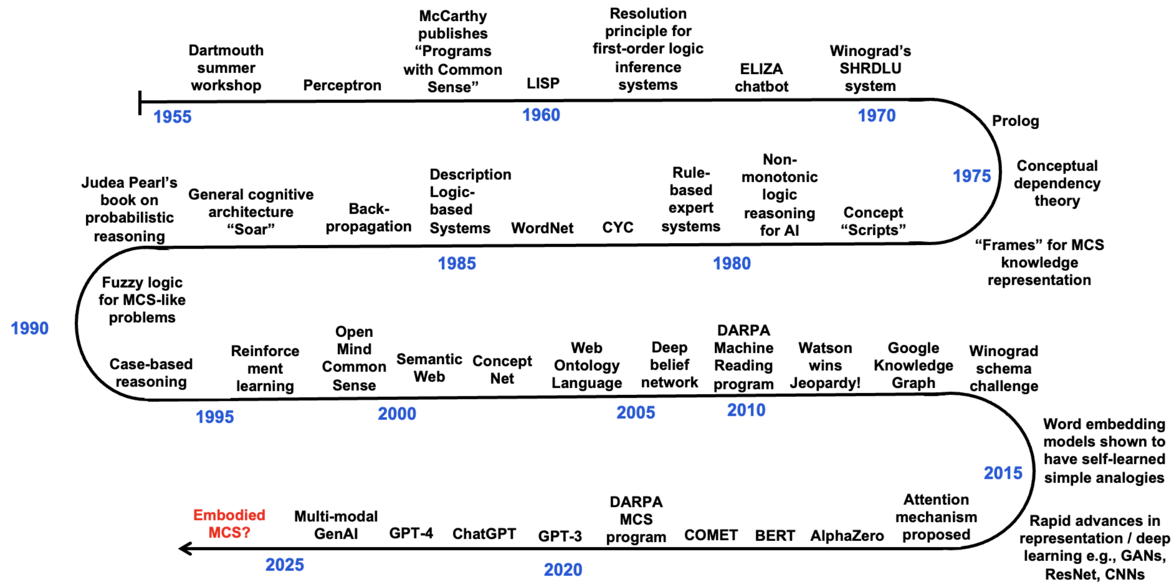

Figure 2: An overview of selected milestones in machine common sense (MCS) since its origins as a general AI problem in the 1950s. Consecutive items were sometimes published in the same year e.g., *Watson wins Jeopardy!* and *Google Knowledge Graph* can both be traced to 2011.

paper provides a historical perspective revisiting the original vision of MCS in the late 1950s, and its evolution in the wake of the Web. We argue that the growth of the Web, and research that it turbo-charged on knowledge graphs, Semantic Web and crowdsourcing, all played a key role in advancing MCS research. However, the contributions are not one-sided. MCS research today informs progress on applications directly relevant to Web sciences, such as agentic systems and their feasibility, complex (e.g., plausibility-based) reasoning, and robust evaluation of black-box LLMs.

2 Machine Common Sense as a Grand Challenge: A 70-Year Perspective

Although MCS has proven to be a surprisingly difficult AI problem, given its seeming ubiquity in humans, many milestones have been achieved along the way. Direct interest in the problem of MCS *qua* grand challenge extends as far back as the Dartmouth Workshop in 1956. Figure 2 provides an illustrative non-comprehensive path showcasing this progress, which also coincided with an explosion of research across both computing and AI. Innovations in hardware, including exponential growth in computing power following Moore's law [15], played an important role in fueling this growth, as did rapid developments and seminal contributions in areas ranging from machine learning to knowledge representation [10, 17].

The Dartmouth Workshop was followed shortly by McCarthy's seminal paper in 1959 on "programs with common sense" [18]. He proposed representing commonsense knowledge using the first-order predicate calculus. For example, he represented "If I'm at my desk, I can walk to my car" as the formula $at(I, desk) \rightarrow can(go(desk, car, walking))$. But the details needed work, as entities like *I* and *can* don't behave the way mathematical functions usually do. Collaborators of McCarthy continued to develop logical approaches to address this issue. Much work was devoted to formalizing particular areas of discourse such as time intervals, events, and the "naive physics" observable behavior of simple physical objects. A more modern example is Gordon and Hobbes's theory [11], which aims to model "how people think people think."

However, researchers started to recognize that logical approaches can be difficult for representing common sense knowledge. They rely on exact definitions and logically sound inference rules, which are difficult to come by for vaguely defined common sense concepts and heuristic inference. Marvin Minsky was also a strong proponent of the importance of common sense knowledge, but he advocated a frame-based representation (where each concept is linked to slots for its attributes) rather than a logical one [22].

An important common sense system was *Thought Treasure* [23]. Its notable feature was that it constructed rough simulations of situations that were described to it, such as a story about a few people in rooms of a house. What was remarkable was that even a crude simulation – a grid of positions, with people represented by question-mark symbols – was enough for it to answer many interesting common sense questions, simply by "looking at" the simulation in its "mind's eye."

Another strand of common sense research grew out of the expert systems movement of the 1980s [20]. The primary representations were *ontologies* – hierarchies of descriptions of concepts; and *If-Then* rules. A typical medical example would have an *If* part listing some symptoms and test results; and the *Then* part would yield a diagnosis. Ed Feigenbaum, one of the leading lights of the expert systems movement [8], expressed it as "in the knowledge lies the power." Expert systems has common sense knowledge – but only from a small group of experts rather than ordinary people.

Expert systems tended to be brittle, in that if the concepts and rule conditions were not met exactly, the inference would fail. This often occurred in cases where people would succeed, due to the flexibility they had in matching concepts and adapting consequences to the context. Expert systems could not match this flexibility, often because what people considered common sense knowledge was not explicitly mentioned in the rules.

Doug Lenat started the CYC project [16], motivated by combining the lessons of expert systems, with McCarthy's and Minsky's calls to action to mount a large project to collect common sense knowledge. CYC originally stood for "encyclopedia," the original assumption being that requisite knowledge would be factual knowledge like what is found in encyclopedias. But the approach soon shifted to modeling the implicit *common sense* knowledge that would enable a person to read factual knowledge in an encyclopedia. Hence, CYC should be credited for distinguishing purely factual or encyclopedic knowledge like "The 46^{th} President of the United States is Joe Biden," and common sense knowledge such as "A chair is for sitting." High-quality factual knowledge is explicitly stated in a very precise manner, can be definitively verified or falsified, and is suitable for representation in formal languages. In contrast, common sense knowledge is communicated between people in vague natural language, or simply assumed and not communicated at all. CYC was one of the earliest efforts to encode large amounts of common sense facts in a formal knowledge base.

CYC also took an all-encompassing view of inference, providing packages for performing inference techniques popular with other AI projects. It was notable for its idea of "micro-theories" – a way to do localized inference in specific domains of common sense discourse rather than insist on globally-applicable rules. Under continuous development since 1985, CYC is one of the longest-running, most impressive engineering achievements of AI.

3 Enter the Web

Pre-Web projects on commonsense knowledge using logical representation, or ontologies like CYC, presented small-scale solutions to representing and reasoning with commonsense. But they did not address the problem of scaling knowledge bases or rule-bases to sufficient breadth and depth to solve nontrivial problems. It took the Web for these changes to occur, with key Web technologies enabling rapid progress in MCS. MCS projects also contributed to the development of now-commonplace Web technologies.

Crowdsourcing. With the advent of the Web in the mid-to-late 1990s, it became possible to use the Web's audience to gather large amounts of knowledge via voluntary contributions, called *crowdsourcing*. At MIT, Singh and Havasi launched *Open Mind Common Sense* (OMCS) around 2000 [27]. To our knowledge, this was the first significant use of Web crowdsourcing to fuel an AI project. OMCS asked users to type in simple commonsense sentences in natural language, such as "You sit on a chair" or "Water is wet." Analysis of roughly a million of these statements yielded a set of about 25 most common relations like "is a," "used for," or "located at." OMCS expanded its knowledge about existing concepts by asking users questions about preferred relations, like, "What is a fork used for?". The result was compiled into a KG called *ConceptNet* [28].

Knowledge Graphs (KGs). KGs are data structures that represent entities in the world including objects, their inter-relationships, and concepts. They had their foundations in semantic networks from the 70s and beyond and rose to prominence with the Semantic Web and the linked data movement. Google popularized them with their famous "things not strings" blog in 2012 that also exposed the Google search utilized them to enhance its functionality. Their KG utilized Wikipedia content, Freebase, and CIA World Fact book (which also contained content from CYC), and ConceptNet among others. The input included a linguistic-oriented representation, like ConceptNet, and a formal logic representation, like CYC, and it like many later knowledge graphs, struck a balance between these representations, thereby allowing simpler knowledge entry and improved usability for AI developers. The Google KG was a large-scale widely available *open-source* Knowledge Graph. Analogical inference techniques like *AnalogySpace* and *Blending* [13, 29] foreshadowed now-popular embedding and dimensionality reduction methods [30]. The MIT group used common sense knowledge to provide intelligent defaults and predictive help in several interactive applications like preventing cyberbullying [7].

Today's successors, like the University of Washington's *Atlas of Machine Commonsense* (ATOMIC) [26], and USC's *Common Sense Knowledge Graph* (CSKG) [14] have used modern machine learning techniques to mine content and perform inference. Such aspects play an important role in conversation and other similar communicative acts between humans, as the philosopher Paul Grice detailed in his seminal papers [12]. Despite formal knowledge representation being largely surpassed in tests by recent LLMs, they remain useful for neuro-symbolic systems that employ hybrid explanation, sanity checking, and for debugging facilities [3].

The Semantic Web. In 2001, the visionary Semantic Web article was published in Scientific American [2] that presented a user scenario of intelligent assistance that made essential use of commonsense knowledge that could be found on, or inferred from, the Web. It imagined that the user would request in natural language, "Make a doctor's appointment for my mother within 20 miles of her home", and the system would return a set of possibilities that would meet those constraints. It would require integrating personal information like location, and general information like how to search for doctors, all represented in the proposed Semantic Web format. The first full book on the subject in 2003, "Spinning the Semantic Web" [9], showed how commonsense knowledge could be represented using ontology languages like the Web Ontology Language (OWL) [21], description logics in general [1], and Web technologies, and how commonsense reasoning could take place with them.

Web crawlers. The advent of Web crawlers, designed originally for search engines, were the enabling technologies that led to modern LLMs and GenAI chatbots. Once machine learning technologies that could generalize well from unrestricted natural language were developed, they could use Web crawlers to acquire commonsense knowledge from reading the Web. That allows them to achieve their current impressive performance on MCS benchmarks.

4 What is next for Machine Common Sense?

Despite enormous progress, there is still widespread belief that MCS remains unsolved, even for current state-of-the-art LLMs like

OpenAI's o1. The 70-year history of MCS leaves us with many lessons – one of which is to be cautious against declaring victory too soon. In fact, testing MCS is itself considered an important area of research in both NLP and computer vision, and in recent years, the role of both semantics and cognitive theories of commonsense have both been noted as critical for evaluating LLMs across the full spectrum of commonsense domains. Part of what makes the problem challenging is that, in contrast to formal logic-based reasoning (which tends to be deductive), both abductive and counterfactual reasoning are now recognized as central for achieving MCS [4].

This focus on reasoning, and on novel types of LLM prompting, such as graph-based retrieval augmented generation [25], suggests that modern work on KGs and fields like the Semantic Web have much to offer for a new *science* of MCS. Increasingly, researchers are starting to assess the utility of MCS in actual systems and in testing a system's ability to effectively reason with common sense knowledge. The distinction between knowledge and reasoning becomes important at a time when LLMs are prone to the Chinese Room phenomenon raised by John Rogers Searle: even when ChatGPT *seems* to be reasoning, it may in fact be regurgitating memorized knowledge. It is very difficult to know the inner mechanism behind why the model returns an answer that it does. There is clearly a need for rigorous tests with high external and internal validity. In the natural and social sciences, especially psychology, designing such tests is a well-understood problem, which the MCS community would also benefit from.

We end this paper by noting that what seems remarkable about the success of today's LLMs on MCS tests is that they were not explicitly designed for this purpose. They just turned out to work well. This may be due to LLMs ingesting many orders of magnitude more data than the explicitly-constructed common sense knowledge bases, indiscriminately absorbing both factual and common sense data in their training (and sometimes wrong and useless information as well, leading to their potential unreliability). Others have also proposed plausible causes, including measurement biases, for LLMs' learning of "unexpected" and emergent skills, and that are hence argued to not be as unpredictable as they might seem [24].

Because of its focus on agents, the Web will play an important role in further development of MCS. Clearly, one-off question answering tests are not adequate. What is needed instead are multi-step, complex reasoning challenges. The Web has also had a long history of focusing on metrics beyond mere performance: usefulness, robustness and feasibility are first-class citizens from the earliest days of the Web. Similar metrics can (and should) be applied to agentic systems that are applied to general-purpose problems, and that are therefore relevant for humanlike common sense.

Acknowledgments. This work was funded under the DARPA Machine Common Sense (MCS) program under award number N660011924033. Lieberman's work is also supported by the AFOSR grant "Cognitive Architectures for AI". We gratefully acknowledge Zhisheng Tang and Ke Shen at the University of Southern California for assisting with selected references and formatting of both figures.

References

[1] Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider (Eds.). 2007. *The Description Logic Handbook: Theory, Implementation, and Applications.* Cambridge University Press.
[2] Tim Berners-Lee and James Hendler. 2003. The Semantic Web. *Scientific American* (2003).
[3] David Chanin and Anthony Hunter. 2023. Neuro-symbolic Commonsense Social Reasoning. *arXiv preprint arXiv:2303.08264* (2023).
[4] Yejin Choi. 2022. The curious case of commonsense intelligence. *Daedalus* 151, 2 (2022), 139–155.
[5] Ernest Davis. 2023. Benchmarks for automated commonsense reasoning: A survey. *arXiv preprint arXiv:2302.04752* (2023).
[6] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).
[7] Karthik Dinakar, Birago Jones, Catherine Havasi, Henry Lieberman, and Rosalind Picard. 2012. Common sense reasoning for detection, prevention, and mitigation of cyberbullying. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 2, 3 (2012), 1–30.
[8] Edward A Feigenbaum. 1988. Knowledge processing: From fileservers to knowledge servers. *Applications of Expert Systems* 2 (1988), 3–11.
[9] Dieter Fensel, James Hendler, Henry Lieberman, and Wolfgang Wahlster. 2003. *Spinning the Semantic Web.* MIT Press.
[10] Michael R Garey, David S Johnson, and Larry Stockmeyer. 1974. Some simplified NP-complete problems. In *Proceedings of the sixth annual ACM symposium on Theory of computing.* 47–63.
[11] Andrew S Gordon and Jerry R Hobbs. 2017. *A formal theory of commonsense psychology: How people think people think.* Cambridge University Press.
[12] Paul Grice. 1991. *Studies in the Way of Words.* Harvard University Press.
[13] Catherine Havasi, Robyn Speer, James Pustejovsky, and Henry Lieberman. 2009. Digital intuition: Applying common sense using dimensionality reduction. *IEEE Intelligent systems* 24, 4 (2009), 24–35.
[14] Filip Ilievski, Pedro Szekely, and Bin Zhang. 2021. Cskg: The commonsense knowledge graph. In *The Semantic Web: 18th International Conference, ESWC 2021, Virtual Event, June 6–10, 2021, Proceedings 18.* Springer, 680–696.
[15] Charles E Leiserson, Neil C Thompson, Joel S Emer, Bradley C Kuszmaul, Butler W Lampson, Daniel Sanchez, and Tao B Schardl. 2020. There's plenty of room at the Top: What will drive computer performance after Moore's law? *Science* 368, 6495 (2020), eaam9744.
[16] Douglas B Lenat, Mayank Prakash, and Mary Shepherd. 1985. CYC: Using common sense knowledge to overcome brittleness and knowledge acquisition bottlenecks. *AI magazine* 6, 4 (1985), 65–65.
[17] Arthur B Markman. 2013. *Knowledge representation.* Psychology Press.
[18] John McCarthy. 1959. Programs with common sense.
[19] John McCarthy, Marvin L Minsky, Nathaniel Rochester, and Claude E Shannon. 2006. A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. *AI magazine* 27, 4 (2006), 12–12.
[20] John McDermott. 1982. R1: A rule-based configurer of computer systems. *Artificial intelligence* 19, 1 (1982), 39–88.
[21] Deborah L McGuinness, Frank Van Harmelen, et al. 2004. OWL web ontology language overview. *W3C recommendation* 10, 10 (2004), 2004.
[22] Marvin Minsky. 1974. A framework for representing knowledge.
[23] Erik T Mueller. 1998. Natural language processing with Thought Treasure.
[24] Stephen Ornes. 2024. How Quickly Do Large Language Models Learn Unexpected Skills? https://www.quantamagazine.org/how-quickly-do-large-language-models-learn-unexpected-skills-20240213/?mc_cid=54a7d2f367&mc_eid=c856ef74e7.
[25] Boci Peng, Yun Zhu, Yongchao Liu, Xiaohe Bo, Haizhou Shi, Chuntao Hong, Yan Zhang, and Siliang Tang. 2024. Graph retrieval-augmented generation: A survey. *arXiv preprint arXiv:2408.08921* (2024).
[26] Maarten Sap, Ronan Le Bras, Emily Allaway, Chandra Bhagavatula, Nicholas Lourie, Hannah Rashkin, Brendan Roof, Noah A Smith, and Yejin Choi. 2019. Atomic: An atlas of machine commonsense for if-then reasoning. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33. 3027–3035.
[27] Push Singh. 2002. The public acquisition of commonsense knowledge. In *Proceedings of AAAI Spring Symposium: Acquiring (and Using) Linguistic (and World) Knowledge for Information Access*, Vol. 3. https://cdn.aaai.org/Symposia/Spring/2002/SS-02-09/SS02-09-011.pdf
[28] Robyn Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 31.
[29] Robyn Speer, Catherine Havasi, and Henry Lieberman. 2008. AnalogySpace: Reducing the Dimensionality of Common Sense Knowledge.. In *Aaai*, Vol. 8. 548–553.
[30] Laurens Van Der Maaten, Eric O Postma, H Jaap van den Herik, et al. 2009. Dimensionality reduction: A comparative review. *Journal of Machine Learning Research* 10, 66-71 (2009), 13.
[31] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).