

# AI Insights Dataset Analysis

Colleen McKenzie and J. Bryce Hidysmith

## 1 AI Insights Dataset Analysis

### 1.1 Overview

This dataset collects a list of discoveries essential to the development of the current generation of artificial intelligence. We trace this development back as early as we have records of relevant intellectual progress—in this case, to Ancient Greek philosophy—looking for the foundations of complex new approaches.

Our goal in this project is to present an alternative to an achievement-based timeline of AI development. The development of new capabilities in machine intelligence frequently produces exciting tangible results, which dominate public conceptions of progress and are well-remembered in histories of the field, but such results are not always the product of novel technical insight. Some new abilities rely more on recombination of existing tools or on additional computational capacity—sufficient to solve complex problems efficiently, but not necessarily the most direct precursors of new tools for further recombination, or of new directions for future work without additional increases in computational efficiency. Our aim is to augment the current discourse of historically grounded forecasting by adding a specific discursive thread that tracks theoretical discoveries alone, complementing the more capability-focused accounts.

We've collected metadata about each discovery (and its discoverers) and looked for trends in their production. Our own findings are described below, but we encourage extension and additional evaluation of the data, and will publish our own ongoing work here.

### 1.2 Methods & Assumptions

Our metric for what constitutes an insight is an estimate of how novel or surprising the results of that insight are, using a comparison of LSTM and RNNs as a baseline—that is, given that RNNs exist, how surprising is the development of LSTM? All discoveries that seem equally or more surprising were included in our set. This is one of many possible thresholds for significance we could have chosen, and the rationale behind ours is subjective:

Our focus on theoretical progress means that the data do not include progress captured in implementing an existing theory, or applying it more directly. For example, punched-card programming of early computers had clear impacts on the progress of computing, but we attribute the insight behind that technology to the [18th-century textile worker](#) who developed a punched-card method for programming looms. From this perspective, the application of punched cards to electrical computers is a combination of existing theories, not a new development in itself.

We consulted two existing chronicles of the history of AI in an effort to flesh out parts of our data set that may have fallen in our blind spots: - Nilsson, N. J. (2009). *The quest for artificial intelligence*. Cambridge University Press. - Crevier, D. (1993). *AI: the tumultuous history of the search for artificial intelligence*. Basic Books.

A few additional notes on the data:

- **First to discover, not first to publish:** We credit an insight to the first person known to have described it, as best we can determine through reputable sources, not the first person to publish that description. This practice is motivated in part by our understanding that not all discoverers published work in the academic tradition (e.g. William Newcomb, credited with the conception of Newcomb's problem), and in part by some discoveries' having been kept confidential, pre-publication, in corporate and state departments.
- **Specific vs. general discoveries:** Sometimes an early discovery constitutes a special case of a later, more general discovery; in this case, we credit the more general discovery. If the special case itself seems to constitute an influential insight, we'll credit it as a separate line item; but if it was largely ignored, or was unavailable to discoverers of the more general version, we omit it from this list. For example, ideas from evolutionary biology have been rediscovered independently by AI researchers through simulationist and purely rationalistic investigation that were previously discovered empirically through biological field research before being formalized in such a manner that they could be used by AI researchers. We count the usable formalization, rather than the case studies that could be used as raw material for specifying the usable formalization.
- **Hardware agnosticism:** Our focus here is on information theoretical insights, not on insights that aid in the development of new hardware, the rationale being that artificial intelligence is a substrate independent pursuit, provided that the Church-Turing thesis holds. Further, charting the history of today's electronic computers alone would constitute a project of roughly equal magnitude, and one that aligns less closely with our area of expertise.

### 1.3 Data & Parameters

The full dataset (as json) and its schema are available at [mediangroup.org/research](https://mediangroup.org/research); this section provides a brief overview of the data for readers of this report.

The set includes 193 total insights.

For each insight, we record the following information (plus sources for each fact):

- Name
- Year of discovery
- First work in which the insight was referenced or published
- Discoverer(s)
- Institutional affiliation of discoverer(s) at time of discovery (if applicable)
- Institution(s) sector and type (see below)
- Nation of origin of discoverer(s)
- Additional notes

Institutions are assigned rough sectors and types from the following lists:

Sectors:

- Academic
- Military
- Industrial
- State
- Other R&D
- Religious
- N/A

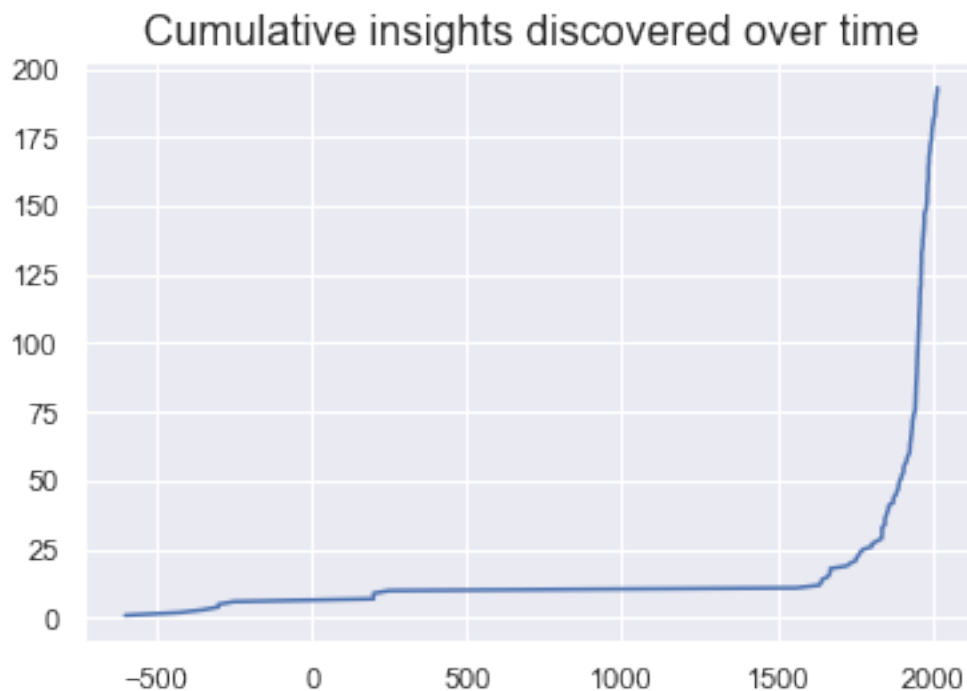
Types:

- Public university
- Private university
- For-profit entity
- Non-profit entity
- N/A

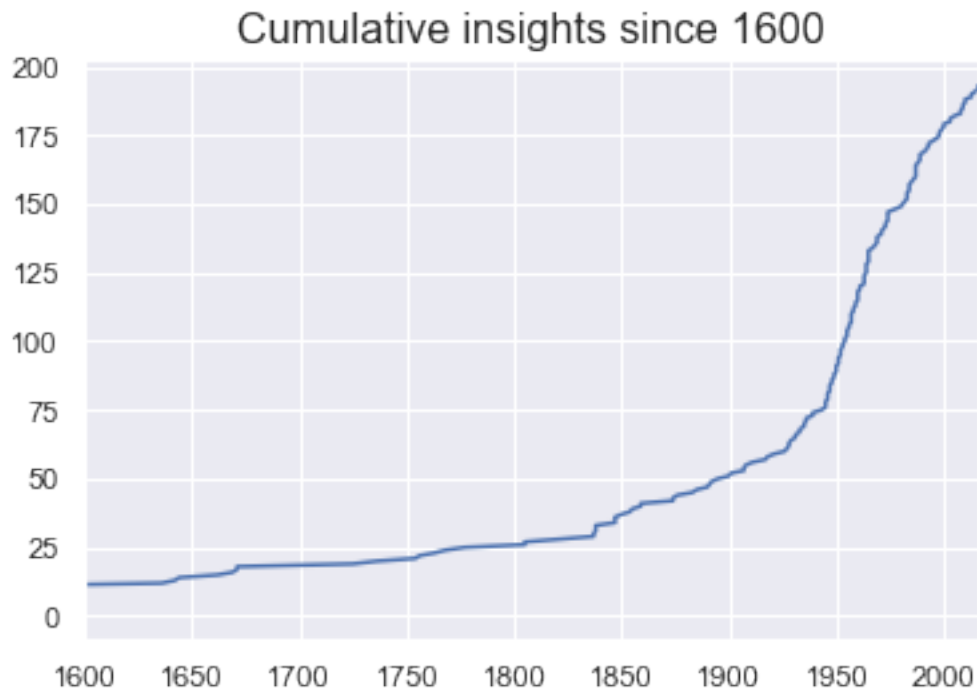
## 1.4 Analysis

For the purposes of this first analysis section, we will discuss the data as we have been able to make it available to us. Reflections on the limitations of the data and the extent to which they are relevant to readership follow in the *Methodological Reflections* section.

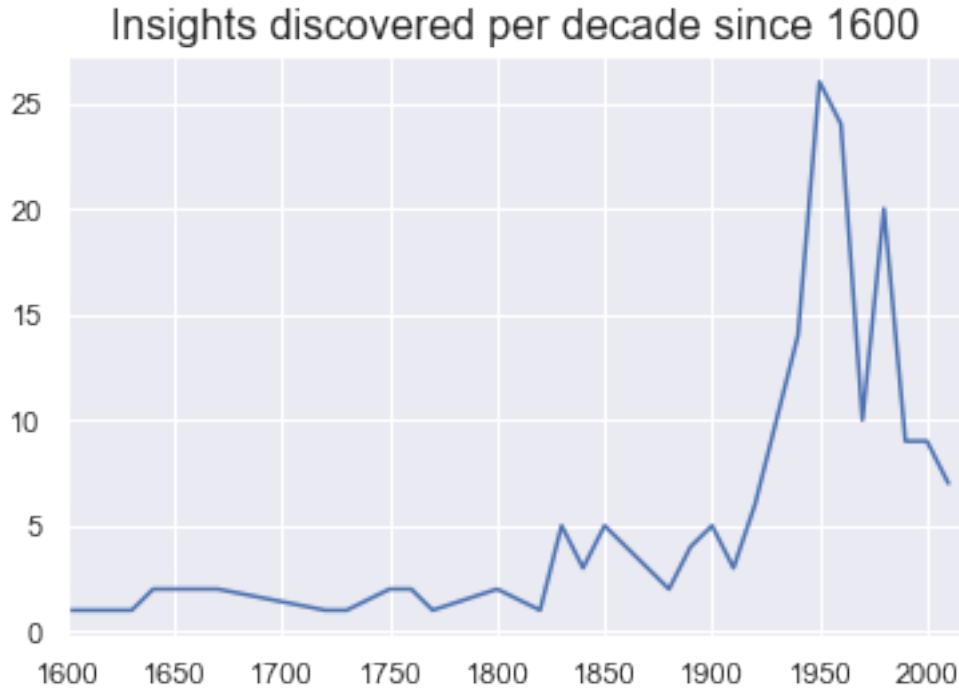
### 1.4.1 Temporal distribution



At a glance, the distribution of insights over time shows an expected uptick in progress contemporaneous with the European Enlightenment, followed by a sharp increase in the pace of discovery at the beginning of the 20th century.



A view of only the most recent centuries provides more detail on the arc of recent progress, showing a significant increase in production with introduction of digital computers available to major institutions in the late 1940s.



A breakdown of insight discoveries by decade clarifies the changing rates of discovery.

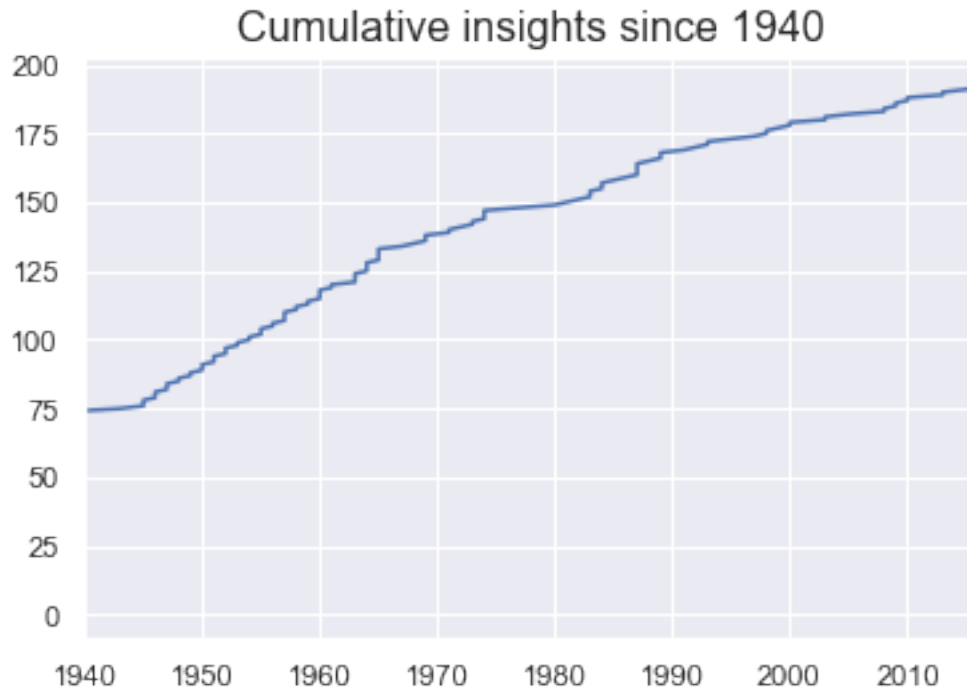
**Historical context** The uneven distribution of insights correlates clearly with the trajectory of Western intellectual and economic history. The theological and philosophical conservatism of the middle ages matches the flat section of the first curve above from c. 500 - 1500 AD. During the middle ages, there was no sense of how economic or political advantage might be gained from the use of power not derived from human, animal, or natural sources (e.g. waterwheels applied to rivers). The media of the middle ages, too, was meant for interpretation by individual human minds, as opposed to instantiation on larger groups or in artificial substrates. For example, medieval alchemical texts were written with the expectation that the text is exclusively human-readable, and so the informational payload of the text is found not within the structure of the text as an object itself, but in the interpretation and implementation of the text by its human reader; this can be contrasted with the periodic table, which describes a structure already present in nonhuman substrate. This norm continued throughout most of the Enlightenment, perhaps reaching its height during the efforts of the Encyclopedists led by Denis Diderot.

The philosophical ideas throughout the 18th and the first half of the 19th century provided the foundation for a scientific community able to take action during the industrial revolution, but the industrial revolution itself provided the necessary material tools for developing automation systems capable of directly realizing political, economic, or scientific goals. The technologies of the industrial revolution, including the social technologies of the factory and the joint stock company developed immediately prior during the late Enlightenment, differed from previous technologies in that they contained *implicit* instructions for their operation that were not simply transformations of a source of physical power such as a horse's muscles. A joint-stock company's organization can be thought of as a primitive attempt at a superhuman intelligence implemented on a collection of individual human beings.

The transition to automated systems that lacked a human in their control loop was the advent of the Jacquard loom (captured in our data as an earlier insight by Basile Bouchon). Earlier automata in the West, the Muslim World, and China, were limited to toys, and their effects limited to the thought that they inspired in their viewers. The Digesting Duck, an automaton from eighteenth century France that crudely pretended the metabolism of a waterfowl, is as inert in its physical effect on the world as a child's toy truck or Michaelangelo's David except by inspiring the behavior of human agents. The automata of the industrial period became actors in their own right. The Jacquard Loom or a modern 3D printer takes a machine-readable instruction and produces an effect that is as physically relevant as if it had been accomplished by human labor.

Though automation systems capable of acting as intelligent agents in their own right, rather than replacements for physical labor, could have been developed in the 19th century through the work of Babbage and Lovelace, the uses of cryptography and automatic aiming systems for artillery during the Second World War provided the industrial basis for their maturation. Blechley Park's work in decoding the German Enigma code was perhaps the first instance when computation at a scale not accomplishable in an individual human mind was required to counter an existential threat to a major power. Previous efforts in state computation by actors such as the US Government during the censuses of the 1870s and 1880s provided valuable technological precursors, but these calculating and tabulating machines were not true computers in the modern sense. Cryptographic and cryptoanalytic efforts during the wars provided the centralization of capital required to produce an abstract automation system in the form of Colossus, the first digital programmable computing device. Such an automation system that processes and transforms information rather than simply taking an instruction to produce a physical effect is obviously required for the development of modern artificial intelligence. We must further conjecture that the invention of programmable digital computing devices was required for the psychological conditions necessary for individual scientists and engineers to experiment in the fields that would eventually birth artificial intelligence. There was only one Turing, publishing his basic principles of computability in *On Computable Numbers* (1936), but there are an uncountable number of individuals able to experiment with the machines derived from these basic insights.

The massive upward trend in insights following the end of the Second World War can thus be attributed to access to modern computing machines following their invention, as well as the incentivization of computer science and artificial intelligence research instrumentally by the states locked in conflict during the Cold War for purposes of military advantage. Artificial intelligence, like any intellectual pursuit undertaken as more than a recreation by those with leisure time and intrinsic interest, advances at pace with its support by institutions tasked with the necessities security and prosperity.



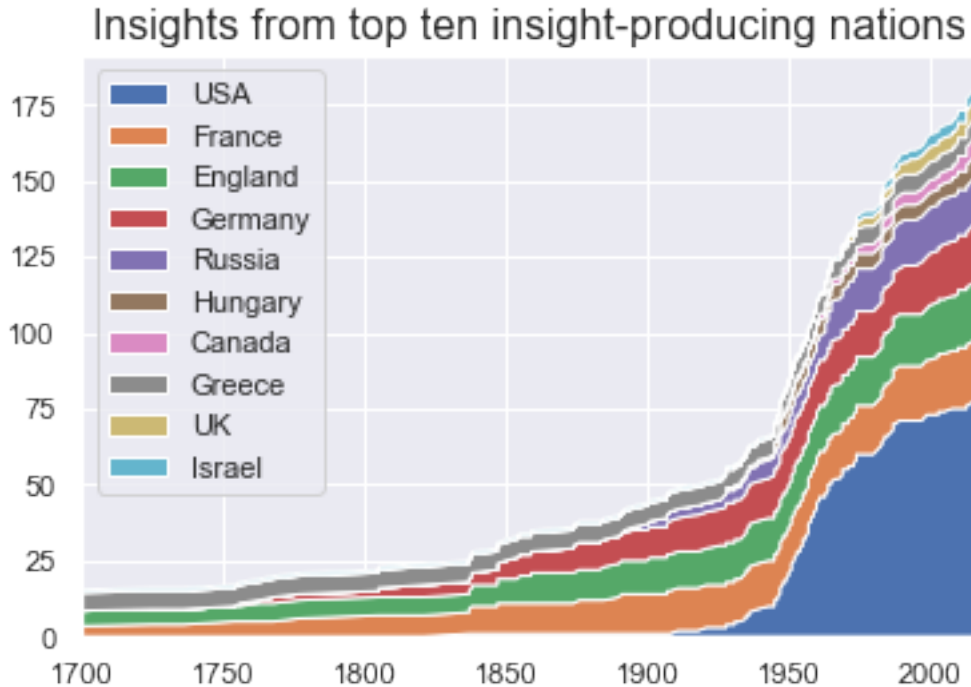
The distribution of insights from the beginning of the Cold War to the present shows only a marginal deceleration after the fall of the Soviet Union, compared to the roughly linear rate of increase at the beginning of the Cold War. The shallower slopes of curves in the mid-1970s and the late 80s through early 90s correspond to the two periods now described as AI winters, when progress slowed.

The former of these two periods appears to be a more substantial decline in insights we measured. We expect this effect is a result of the 1970s being a period of application and combination of existing insights to new problem areas, without substantial generation of additional insights, in addition to the effect of decreased AI-related research during this period.

As Nilsson puts it:

“Until about the early 1970s, most AI research dealt with what Seymour Papert called ‘toy’ problems. . . . However, soon after, AI efforts began a definite shift towards applications work, confronting problems of real-world importance. . . . One reason for the increasing interest in applications was that the power of AI methods had increased to the point where realistic applications seemed within reach.” (Nilsson, 207)

### 1.4.2 Geocultural distribution



(Note: data for “Russia” includes insights produced by Soviet scientists born within Russian borders and working at historically Russian institutions.)

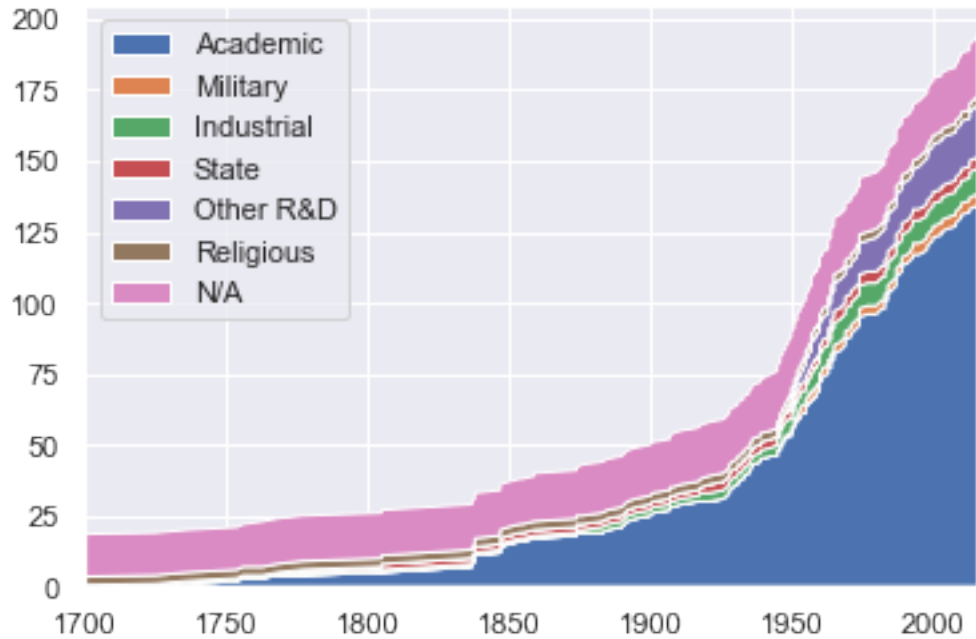
The distribution of insights by nationality of discover reflects a number of well-understood trends in intellectual history. Ancient Greece provided the intellectual foundations that England and France then built upon during the Enlightenment, before the rise of Germany as it industrialized and centralized politically.

During the twentieth century, the rise of both the Russian and American superpowers correlated to substantial advances, but the United States dramatically outcompeted the Soviet Union. Some of the gains that could be credited to the Soviet Union if one assumes a model based on spheres of influence are marked as Hungarian in origin, as well as a few other countries such as Moldova that did not make it into the top ten. The United States’ position as an aggregator of immigrants and refugees from Europe during and after the Second World War substantially increases the number of entries counted as American. The circumstances that made such devices available were not only economic but also politically strategic, as the political aims of military supremacy by both the United States and the Soviet Union during the Cold War enabled their local markets to distribute said computing machines. Additionally, given that the state of Israel was not established until 1949, entries there only date to that time, though many of the scientists who immigrated to Israel and established the scholarly communities that contributed to AI efforts originated in Soviet or American territory.

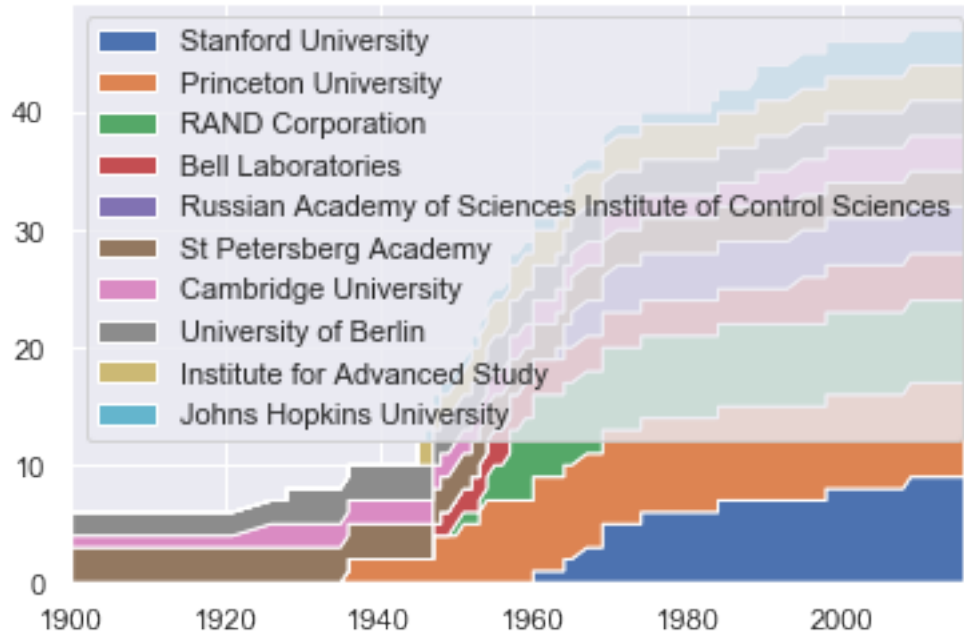


### 1.4.3 Institutional distribution

Insights by sector of origin



Insights from top ten insight-producing institutions

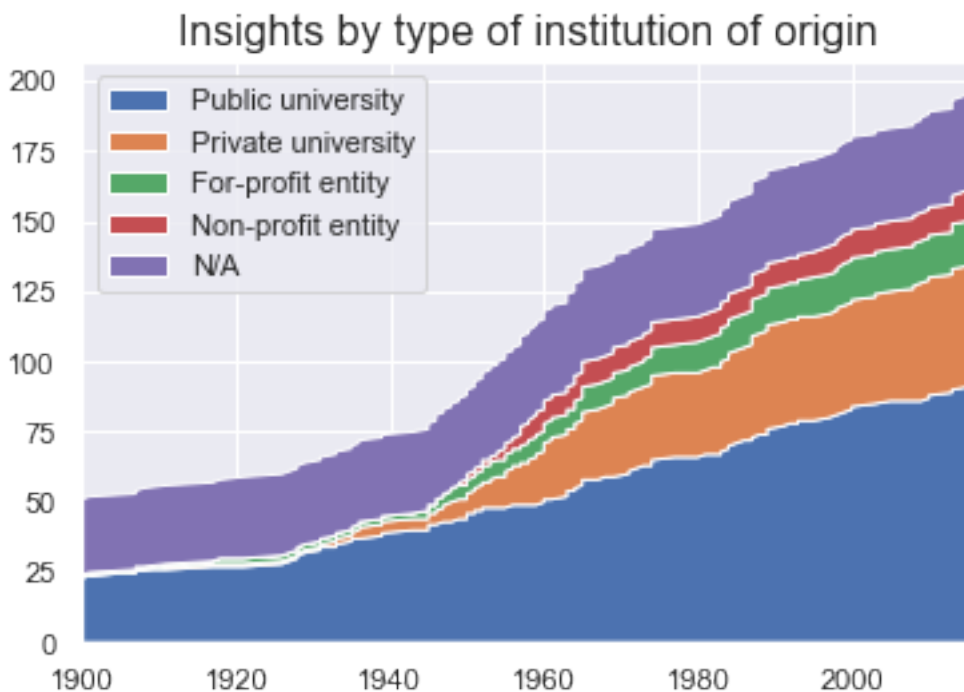


The vast majority of insights in our dataset, unsurprisingly, were discovered and developed by individuals at academic institutions. Early progress in mathematics and the sciences was due mostly to unaffiliated individuals—including most researchers from antiquity and the Renaissance—but as academic institutions gained prestige and resulting brilliance, they began to outcompete other institutional types.

Two trends stand out in the latter part of the 20th century: the increase in output from nonacademic R&D institutions, and the slight increase in insights derived within military institutions.

Nonacademic R&D includes institutions such as Bell Laboratories and the RAND Corporation, and appears in our data first in 1947 (an attribution to Bell Labs’s John Tukey). Bell Laboratories (or Bell Labs) was an unusual institution, in historical context, and in many ways the first of its kind: a result of a step-function increase in the engineering progress needed for telecommunications infrastructure, Bell Labs was incubated within the monopolistic Bell Telephone Company before its 1984 dissolution after a successful antitrust suit. Within it, researchers such as Shannon, Shockley, and Hamming developed some of the core advances of their fields, and it has continued to produce novel research for decades: the most recent insight to come out of Bell Labs, in our data, was from 1995. The RAND Corporation, in contrast, was developed in response to state officials suggesting potential benefits from private involvement in relevant research areas, and RAND itself developed out of an aircraft company, becoming a separate, nonprofit institution in 1948. While insights from RAND represent a shorter timespan in our dataset, limited to just the decade between 1950 and 1960, they outnumber those from any institution besides Stanford and Princeton universities, well-known powerhouses of AI development, by the present day. Dynamic programming (Bellman) and symbolic AI (Newell, Shaw, and Simon) are notable among its body of production.

The relative scarcity of military institutions in our data suggests a more indirect role due to military involvement. The development of artificial intelligence is deeply entangled with military concerns, but military institutions seem to have been sources of demand for progress, not sources of progress itself. DARPA, for example, often contracted independent entities such as SRI or teams of collaborating academics to execute on finding solutions in its areas of interest. We note when such entities were affiliated with military institutions, though few such instances occur in our data.



A final note on the breakdown of insights among institutional entities: while the credit owed to academic institutions is clear, nonacademic institutions have been contributing an increasing portion of total progress since the middle of the 20th century. We were surprised to see a near even distribution of insights between for-profit and non-profit entities—though RAND Corporation constitutes the bulk of the latter—in recent decades, and we suppose that this supports the importance of ongoing investment in independent organizations focused on AI whose mission is explicitly aligned with public good.

#### 1.4.4 Methodological reflections

Our methods generate a view of the past through a particular cultural lens—specifically, that of young American computer scientists. This project is not an attempt at the impossible task of aggregating all insights into artificial intelligence that exist, rather it is an attempt from our position to aggregate the ones interpretable to ourselves. As we are roughly in the same position socially as other artificial intelligence researchers in the Anglosphere, we expect that the amount of insights accessible to Anglosphere researchers will be analogous to those accessible to ourselves. There may be other paths to artificial intelligence development that are only accessible to individuals with differing social position, technical knowledge, or linguistic and cultural proficiency. Though it is certainly likely that such paths exist, we can make no comment on their rate of advancement.

Additionally, it seems prudent to make note of the likely presence of survivorship bias: we are limited to only the information that has survived to the point of being accessible to ourselves. It is possible that many of the insights of the past that we have analyzed—particularly the distant past—were developed with the aid of information sources that no longer exist, either dead with the minds that generated them or lost with the documents that once preserved them. As with any field, this lost information may compromise the ability of individuals in the present to correctly

interpret and use past insights. This case of survivorship bias is unavoidable, and mitigatable to the degree that present researchers remain aware of it. Related work on this topic is available from Samo Burja of Bismarck Analysis, in his essay [\*Intellectual Dark Matter\*](#).