Artificial intelligence history, and libraries: History and Legacy of Library Contributions to machine learning



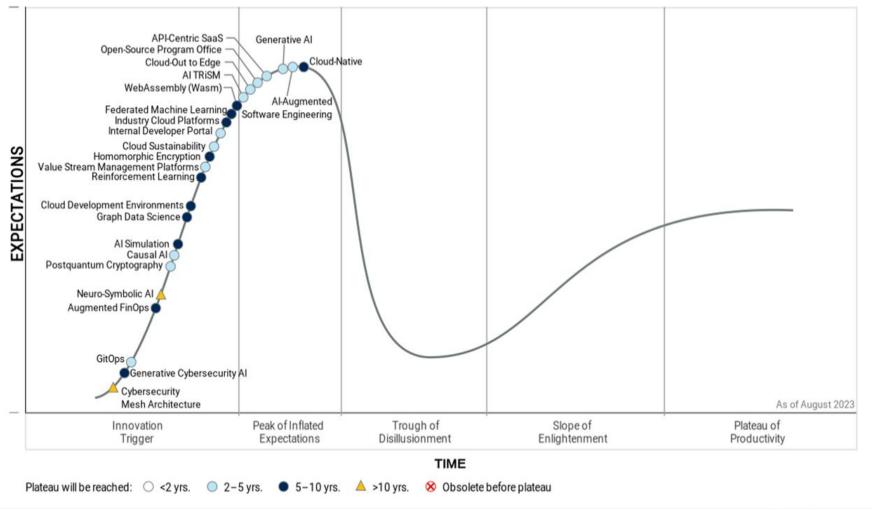
Wilhelmina Randtke, Head of Libraries Technologies and Systems, Georgia Southern University Libraries NASIG Autumn, October 17, 2023, Virtual



The "AI" buzz What is really new?

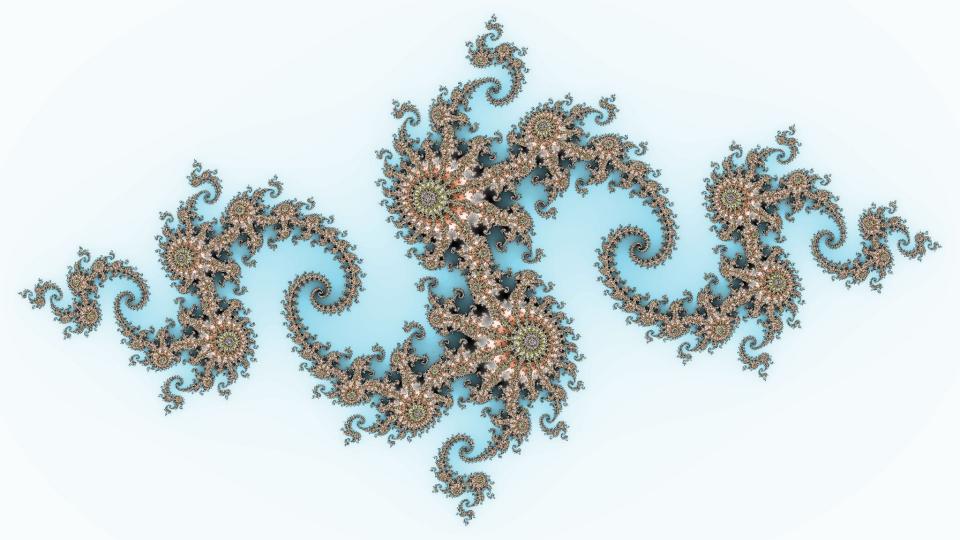
What is really new?

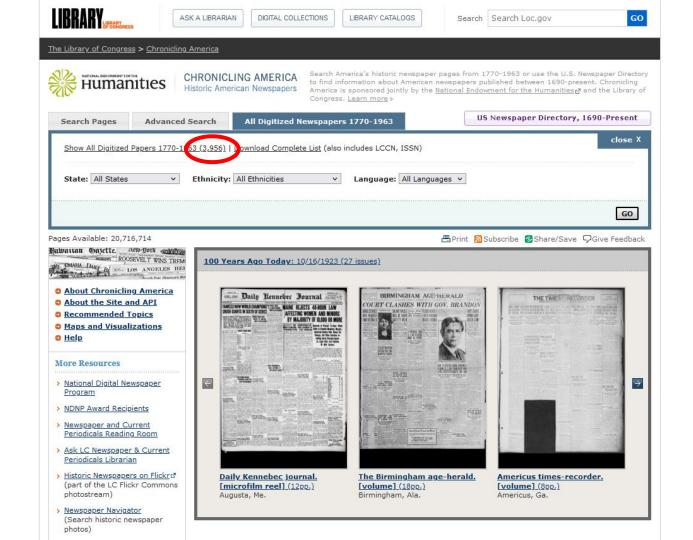




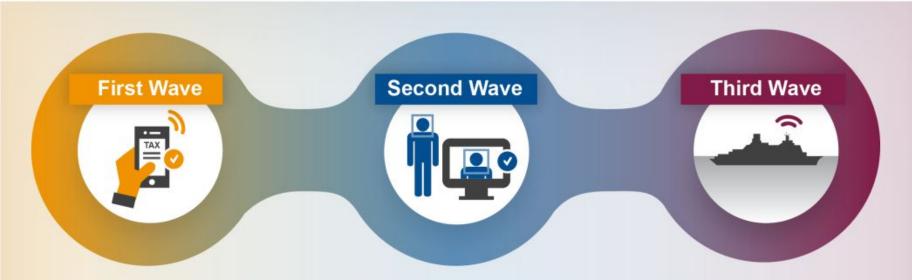
Gartner.

What is really new? Speed, scale





What is really new? Culture and attitudes?



Expert knowledge or criteria and logical reasoning

The first wave of AI is represented by expert knowledge or criteria developed in law or other authoritative sources and encoded into a computer program in the form of an expert system.

Machine/Statistical learning

Second-wave AI technology is based on machine learning, or statistical learning, and includes voice recognition, naturallanguage processing, and computer-vision technologies, among others.

Contextual adaptation

Third-wave AI technology combines the strengths of first- and second-wave AI, and is also capable of contextual sophistication, abstraction, and explanation.

Example: Online tax preparation

Example: Face-recognition technology

Example: Autonomous ships

Diagram from: Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities, Government Accountability Office (GAO) (June 2021).

Who Qualifies as Your Dependent

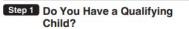
Dependents, Qualifying Child for Child Tax Credit, and Credit for Other Dependents

Follow the steps below to find out if a person qualifies as your dependent and to find out if your dependent qualifies you to take the child tax credit or the credit for other dependents. If you have more than four dependents, check the box under *Dependents* on page 1 of Form 1040 or 1040-SR and include a statement showing the information required in columns (1) through (4).



The dependents you claim are those you list by name and SSN in the Dependents section on Form 1040 or 1040-SR.

Before you begin. See the definition of Social security number, later. If you want to claim the child tax credit or the credit for other dependents, you (and your spouse if filing jointly) must have an SSN or ITIN issued on or before the due date of your 2022 returm (including extensions). If an ITIN is applied for on or before the due date of a 2022 return (including extensions) and the IRS issues an ITIN as result of the application, the IRS will consider the ITIN as issued on or before the due date of the return.



A qualifying child is a child who is your...

Son, daughter, stepchild, foster child, brother, sister, stepbrother, stepsister, half brother, half sister, or a descendant of any of them (for example, your grandchild, niece, or nephew)



was ... Under age 19 at the end of 2022 and younger than you (or your spouse if filing jointly)

Under age 24 at the end of 2022, a student (defined later), and younger than you (or your spouse if filing jointly)

Any age and permanently and totally disabled (defined later)



Who didn't provide over half of their own support for 2022 (see Pub. 501)



Who isn't filing a joint return for 2022 or is filing a joint return for 2022 only to claim a refund of withheld income tax or estimated tax paid (see Pub. 501 for details and examples)

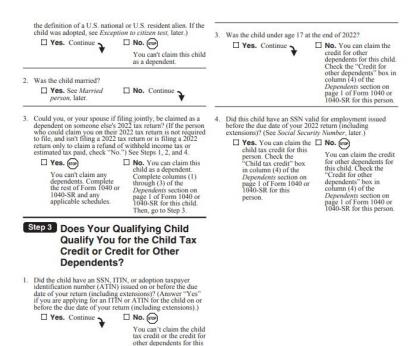


INTUIT turbotax Is Tom's other parent claiming Tom per your legal agreement?

Yes, Tom's other parent will claim Tom for 2018 per our legal agreement.

No, Tom's other parent will not claim Tom for 2018 per our legal agreement.

What if I don't know?



Need more information or forms? Visit IRS.gov.

child

□ No. (100)

child.

You can't claim the child

tax credit or the credit for other dependents for this

2. Was the child a U.S. citizen, U.S. national, or U.S. resident

to citizen test. later.)

Ves. Continue

alien? (See Pub. 519 for the definition of a U.S. national or U.S. resident alien. If the child was adopted, see *Exception*

-18-



Machine Learning... ...not so new

Machine Learning toolkits, in use today

- WEKA
- MALLET
- OpenNLP

These toolkits are listed in The National Artificial Intelligence Research and Development Strategic Plan 2019 Update , Select Committee on Artificial Intelligence of the National Science and Technology Council (June 2019), available at https://trumpwhitehouse.archives.gov/wp-content/uploads/2019/06/National-AI-Research-and-Development-Strategic-Plan-2019-Update-June-2019.pdf .



Machine Learning toolkits: Dates of public release

- WEKA:
 - (Waikato Environment for Knowledge Analysis)
 - established circa 1997
- MALLET
 - (Machine Learning for Language Toolkit)
 - established circa 2002
- OpenNLP
 - established circa 2004



Additional Machine Learning toolkits: Dates of public release

- GATE:
 - (General Architecture for Text Engineering)
 - established circa 1995
- RapidMiner
 - (formerly called Yet Another Learning Environment)
 - released for sale 2001
- SPSS Modeler
 - (formerly called Clementine)
 - released for sale 1994



Additional Machine Learning toolkits: Dates of public release

• SAS

- \circ released for sale 1972
- InfiniteInsight
 - (formerly KXEN)
 - released circa 1998
- KNIME
 - released for sale 2006
- ... and many more



History of Machine Learning in Libraries

"It's worth noting that the term "artificial intelligence" can create discomfort in the computer science community. The phrase has moved in and out of fashion over the decades and is used more in marketing than by researchers. "Machine learning" is more commonly used in the technical literature. Yet the nomenclature of AI is often embraced during funding application season, when venture capitalists come bearing checkbooks, or when researchers are seeking press attention for a new scientific result. As a result, the term is both used and rejected in ways that keep its meaning in flux." -Crawford, K. Atlas of AI (2021).

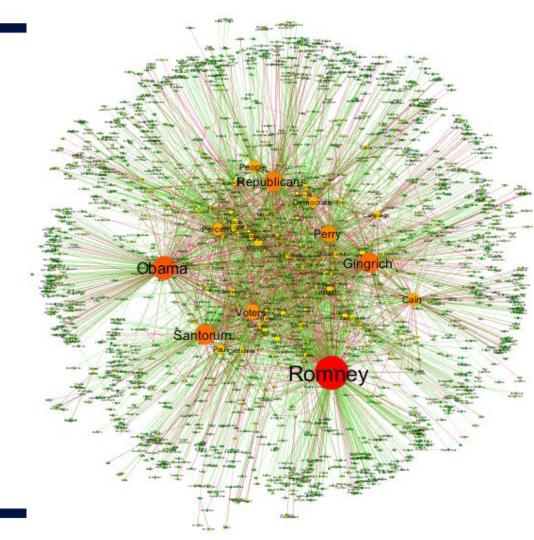


Buzzwords of past hype

Big Data (2013 hype)

Data Mining / Text Mining (2003 hype)

Predictive Coding



Toolkits (they are all old!)

- WEKA (Waikato Environment for Knowledge Analysis)
- MALLET (Machine Learning for Language Toolkit)
- OpenNLP
- GATE (General Architecture for Text Engineering)
- SAS

- RapidMiner (formerly called Yet Another Learning Environment)
- SPSS Modeler (formerly called Clementine)
- InfiniteInsight (formerly KXEN)
- KNIME



≡	G <mark>oogle</mark> Scholar	"general architecture for text engineering" "dublin core"			
٠	Articles	About 162 results (0.09 sec)			
	Any time Since 2023 Since 2022 Since 2019 Custom range	Grid-enabling natural language engineering by stealth <u>B Hughes</u> , <u>S Bird</u> - arXiv preprint cs/0304028, 2003 - arxiv.org Our model leverages existing lin0 guistic resource description and discovery mechanisms based on extended Dublin Core metadata. In addition, application design is flexible, allowing ☆ Save 𝔊 Cite Cited by 19 Related articles All 5 versions ≫	[PDF] arxiv.org		
	Sort by relevance Sort by date Any type Review articles	[PDF] Tools and trends in corpora use for teaching and learning <u>R Godwin-Jones</u> - 2001 - scholarspace.manoa.hawaii.edu The Dublin Core project began in 1995 to develop OLAC uses the core 15 elements of the Dublin Core and extends are GATE ("General Architecture for Text Engineering") which sets ☆ Save 99 Cite Cited by 37 Related articles All 3 versions SN	[PDF] hawaii.edu		
	☐ include patents✓ include citations	KIM—a semantic platform for information extraction and retrieval B Popov, <u>A Kiryakov</u> , D Ognyanoff Natural language, 2004 - cambridge.org The KIM platform provides a novel Knowledge and Information Management framework and	[PDF] proquest.com		
	Create alert	services for automatic semantic annotation, indexing, and retrieval of documents. It provides ☆ Save ワワ Cite Cited by 409 Related articles All 5 versions			
		Ontology-driven semantic digital library SA Noah, NAR Alias, NA Osman, Z Abdullah Technology: 6th Asia, 2010 - Springer An adaptation of the Dublin core metadata can be used to represent the semantic resources by employing the GATE (General Architecture for Text Engineering) engine toolkit [8]. The ☆ Save 奶 Cite Cited by 14 Related articles All 9 versions	[PDF] academia.edu		
		Content enrichment with expressive document modelling to leverage the understanding of unstructured data	[PDF] matec-conferences.org		

≡	Google Scholar	opennlp MARC "subject headings"	
٠	Articles	About 108 results (0.07 sec)	
	Any time	[PDF] OUC's participation in the 2012 INEX Book and Linked-Data Tracks	[PDF] uam.mx
	Since 2023	M Preminger, R Nordlie, D Massey cG2012 remains with, 2012 - ilitia.cua.uam.mx	
	Since 2022	The choice this time fell on the opennlp package of the Apache project. The package was	
	Since 2019	paper we try to build upon this, and look at subject headings extracted from the MARC data	
	Custom range	☆ Save ワワ Cite Cited by 4 Related articles All 8 versions ≫	
	Sort by relevance	[HTML] CDAPubMed: a browser extension to retrieve EHR-based biomedical	[HTML] biomedcentral.com
	Sort by date	literature	
		D Perez-Rey BMC medical, 2012 - bmcmedinformdecismak	
	Any type	selected for CDAPubMed was the Medical Subject Headings (MeSH) thesaurus [28], , OpenNLP [32], extracting sentences and words (or tokens) from the free text content. OpenNLP is	
	Review articles	☆ Save 50 Cite Cited by 21 Related articles All 19 versions SS	
	include patents		
	rinclude citations	[HTML] A framework for exploring associations between biomedical terms in PubMed	[HTML] nih.gov
	Create alert <u>H Yang</u> , L Zhao, Y Z	H Yang, L Zhao, Y Zhang, H Ju, D Wang, Y Hu Oncotarget, 2017 - ncbi.nlm.nih.gov	
		The text of PubMed were segmented into sentences by Apache OpenNLP first, and then	
		terms of sentences were recognized by MGREP. After that two terms occurring in a common	
		☆ Save 99 Cite Related articles All 8 versions	
		Understanding Natural Language Metadata	[PDF] unitn.it
		A Autayeu, F Giunchiglia, P Andrews - 2010 - eprints.biblio.unitn.it	400 - 25 B
		It is a thesaurus of subject headings maintained by the US as the basic building blocks of	
		the LC formulas. Any controlled We have chosen the NER algorithm from OpenNLP and we	
		A Cause RR Other Other the A Related anticles All Queenings AS	

☆ Save ワワ Cite Cited by 1 Related articles All 8 versions ≫

Example uses

Optical Character Recognition

Husene er for det meste af træ. I bydelen omkring Sannesund er gaderne trange og krogede; bebyggelsen bestaar af smaa træhuse omkring den store S. melkefabrik (kondenseret melk) m. fl. fabriker. Her ligger ogsaa toldboden, dampskibsbryggen. I den nye bydel langs Glommens (Sarpfossens) vestbred de store bygninger for Borregaards fabriker. S. er en af Norges ældste byer. Den hed oprindelig Borg og anlagdes af Olav den helmed elektricitet fra Hafslunds elektricitetsverk ved Sarpfossen. Foruden folkeskole og kommunal høiere almenskole er der en teknisk aftenskole. To blade («Glommen» og «Sarpen»). Byen vælger én stortingsrepræsentant. Ved Smaalensbanen staar S. i forbindelse med Kra. (Se pl. Sarpsborg.) [Litt.: B. Christophersen, «S.s historie fra 1016» (1901); «Norges næringsliv. S. og omegn. Med indledning af N. S. Olsen» (1908).]

and the second s			
lassen - @ leave dut, omit -	sion; expulsion - (f) exclusion f,	(u. at) ogs. à moins que ne, à	
(f) omettre, oublier.	retranchement m.	moins de.	out of - (f) hors (au, en dehors) de;
udeladelse - (t) Aus., West-	udelukkende - (t) ausschlies-	udenad - (t) auswendig - (e) by	devant (la porte osv.); (au) dehors.
lassung f - e omission - f	lich — (e) exclusive; entirely,	heart - (f) par cœur; (efter hukom-	udenforstaaende – @ un-
omission f, oubli m.	wholly - (f) exclusif; entierement,	melsen) de mémoire.	beteiligt - (e) outside(r), impartial
udelukke - (t) ausschliessen	tout.	udenbys - (t) ausser der Stadt;	spectator - (f) étranger à, hors de
- @ exclude, debar - (f) exclure,	uden - (t) phne, ausser : (ingen	auswårts; auswärtig - (e) out of	l'affaire; (de u.) ceux du dehors.
retrancher; être incompatible avec.	u.) niemand als - (e) without;	town; not resident in town - (f)	udenfra - (t) von aussen -
udelukkelse - (t) Ausschlles-	(ingen u.) none (nobody) but -	(qui demeure) hors de la ville.	e from without, from outside -
sen n. schliessung f - (e) exclu-	(f) sans; (ingen u) personne sinon;	udenfor — (t) ausser(halb);	(f) de dehors.

Shoffner, R. M. (1971). Some Implications of Automatic Recognition of Bibliographic Elements. Journal of the American Society for Information Science, 22(4), 275–282. https://doi.org/10.1002/asi.4630220406



Extracting metadata fields from digitized records

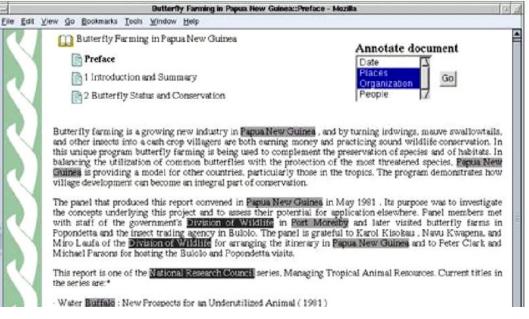
C:/Users/18503/Downloads/Altamaha%20Raft%20Festivals%20collection.pdf

e, osers, rosos, bownioads, ritamana / zonanc/ zon estivals//ze			
Draw 🗸 🖉 🗊 Read aloud Ask Bing Al	- + 🕶 2 of 4 🖓 (B	Q ⊖ B ₽ 2 ∅	
	GEORGIA SOUTHERN UNIVERSITY LIBRARIES Henderson Library and Lane Library		
	OVERVIEW OF COLLECTION		
	Title: Altamaha Raft Festivals Collection		
	Date: 1982		
	Extent: 1 box		
	Creator/Collector: Delma Eugene Presley and Norris Kemp Mabry, collectors		
	Language: English		
	Repository : Zach S. Henderson Library Special Collections, Georgia Southern University, Statesboro, GA. <u>specolle@georgiasouthern.edu</u> . 912-358-4331. <u>library.georgiasouthern.edu</u> .		
	Processing Note: Finding Aid revised in 2019.		

Shoffner, R. M. (1971). Some Implications of Automatic Recognition of Bibliographic Elements. Journal of the American Society for Information Science, 22(4), 275–282. https://doi.org/10.1002/asi.4630220406



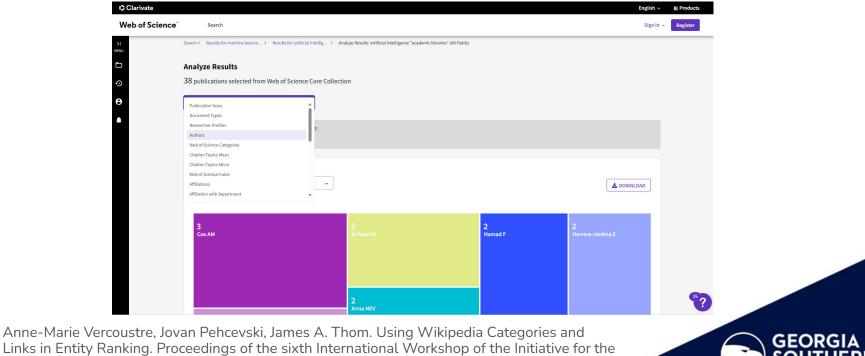
Tag names of people, places, and organizations



Witten, I. H., Don, K. J., Dewsnip, M., & Tablan, V. (2004). Text mining in a digital library. International Journal on Digital Libraries, 4(1), 56–59. https://doi.org/10.1007/s00799-003-0066-4



Entity ranking



Evaluation of XML Retrieval (INEX 2007), Dec 2007, Schloss Dagstuhl, Germany. pp. 321-335, ff10.1007/978-3-540-85902-4_28ff. ffinria-00192489.

UNIVERSITY LIBRARIES

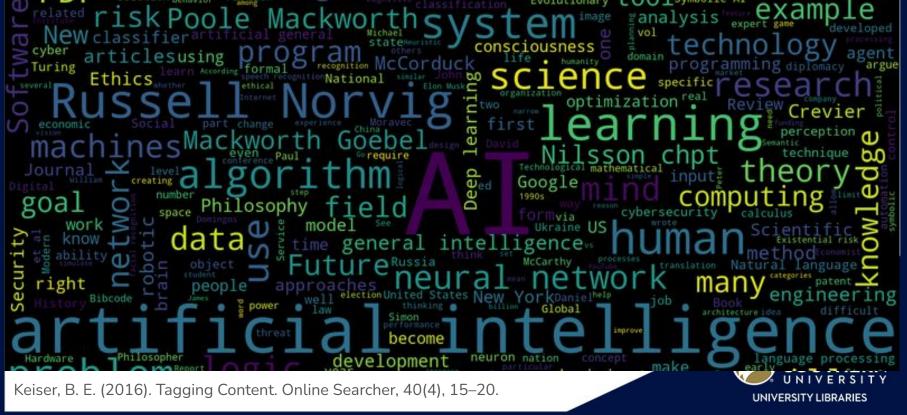
Classify by readability (hard vs easy to read)

File Home Ir	nsert Draw Design Layout References Mailings Review	View Help Acrobat			🖓 Comments	⁹ Editing ~ 🛛 🖻 외	ihare 👻
Paste S A	$\begin{array}{c c} \operatorname{ori} (\operatorname{Body}) & & & & & \\ \hline I & \sqcup & \to \to & x_2 & x^2 & A_{\mathcal{D}} \\ \hline & & \swarrow & \to & x_2 & x^2 & A_{\mathcal{D}} \\ \hline & & & \blacksquare & & & \\ \hline & & & \blacksquare & & & \\ \hline & & & \blacksquare & & & \\ \hline & & & & \blacksquare & & & \\ \hline & & & & \blacksquare & & & \\ \hline & & & & & \blacksquare & & & \\ \hline & & & & & \blacksquare & \\ \hline & & & & & \blacksquare & & \\ \hline & & & & & \blacksquare & \\ \hline & & & & \blacksquare & \\ \hline & & & & & \blacksquare & \\ \hline & & & & & \blacksquare & \\ \hline & & & & & \blacksquare & \\ \hline & & & & & \blacksquare & \\ \hline & & & & & \blacksquare & \\ \hline & & & & & \blacksquare & \\ \hline & & & & & & \\ \hline & & & & & & \\ \hline & & & &$	Normal No Spacing Styles	Heading 1	Create PDF Create PDF and share link share via Outlo		Editor Add-in	
Clipboard 😼	Font 🛛 Paragraph 🕞	Styles	ال <u>د</u> ا	Adobe Acrobat	voice	Editor Add-In	
		Readability Statistics	? ×	Ed	litor		~ ×
		Counts			Formality	2	
	Machine learning seems to be <u>newly</u> everywhere. It's no	Words Characters	101 577		Inclusiveness	Ø	
	newly useful. Imagine an automated cataloging program	Sentences	9		Design of the Comment		
	takes a week to run. Increased processing speed is a sub				Punctuation Convent	ions i	
	history of libraries and artificial intelligence. First, teasin libraries. High quality results and concrete applications c explored and published for decades. Over time, faster pr	Words per Sentence	9.0 11.2 5.5		Resume	ø	
	library and metadata work <u>contributes</u> to machine learn	Readability Flesch Reading Ease	39.6		Sensitive Geopolitical	References 🧳	
		Flesch-Kincaid Grade Level Passive Sentences	10.5 11.1%		Vocabulary	Ø	
			ОК				

Leroy, G., Miller, T., Rosemblat, G., & Browne, A. (2008). A balanced approach to health information evaluation: A vocabulary-based naïve Bayes classifier and readability formulas. Journal of the American Society for Information Science & Technology, 59(9), 1409–1419. https://doi.org/10.1002/asi.20837



Tagging (subject headings)



Locate and tag all citations in a piece of text

Thaler v. Vidal United States Court of Appeals Federal Circuit Aug

United States Court of Appeals, Federal Circuit. August 5, 2022 43 F.4th 1207 (Approx. 9 pages)

Document Filings (14) Negative Treatment (0) History (3) Citing References (96) - Table of Authorities Powered by KeyCife

numan enuues. That non-numans may immige patents does not ten us anything about whether non-numans may also be inventors of patents. The question before us inevitably leads back to the Patent Act's definition of "inventor," which uses the word "individual" – and does not use "whoever."⁵ Furthermore, as we noted already, the Dictionary Act establishes that Congress uses "whoever" as a much broader term than "individual." See 1 U.S.C. § 1.

Second, Thaler contends that AI software programs must qualify as inventors because otherwise patentability would depend on "the manner in which the invention was made," in contravention of 35 U.S.C. § 103. Section 103 is not about inventorship. Instead, it provides, in relevant part, that inventions may still be nonobvious even if they are discovered during "routine" testing or experimentation. *See Honeywell Int'l Inc. v. Mexichem Amanco Holding S.A. de C.V.*, 865 F.3d 1348, 1356 (Fed. Cir. 2017); *see also Graham v. John Deere Co.*, 383 U.S. 1, 15, 86 S.Ct. 684, 15 L.Ed.2d 545 (1966) (explaining that second sentence of § 103 was intended to clarify that "flash of creative genius" is unnecessary for patentability). This statutory provision relates to *how* an invention is made and does not trump a provision that specifically addresses *who* may be an inventor.

Third, Thaler emphasizes that the term "inventor" must be interpreted with attention to the "context in which that language is used[] and the broader context of the statute as a whole." Yates v. United States, 574 U.S. 528, 537, 135 S.Ct. 1074, 191 L.Ed.2d 64 (2015) (internal quotation marks omitted). We have undertaken precisely this task. For the reasons explained above, the Patent Act, when considered in its entirety, confirms that "inventors" must be human beings.

В

Our holding today that an "inventor" must be a human being is supported by our own precedent. See Univ. of Utah v. Max-Planck-

Arguello, J., Callan, J., & Shulman, S. (2008). Recognizing Citations in Public Comments. Journal of Information Technology & Politics, 5(1), 49–71. https://doi.org/10.1080/19331680802153683



Q- AA 🖂 - "="

Go to +

Recognizing author's emotions at time of writing



Facebook said it conducted the study to gauge users' response to content

Facebook is facing criticism after it emerged it had conducted a psychology experiment on nearly 700,000 users without their knowledge.

The test saw Facebook "manipulate" news feeds to control which emotional expressions the users were exposed to.

The research was done in collaboration with two US universities to gauge if "exposure to emotions led people to change their own posting behaviours".

Facebook said there was "no unnecessary collection of people's data".

"None of the data used was associated with a specific person's Facebook account," the social networking giant added.

Cornell University and the University of California at San Francisco were involved in the study.

Investopedia

INVESTING SIMULATOR

SIMULATOR BANKING

PERSONAL FINANCE NE

irump s 2016 presidential campaign.

 Although Cambridge Analytica closed its doors in 2018, several of the company's key personnel have since founded or moved to similar companies.

Understanding Cambridge Analytica

Founded in 2013, Cambridge Analytica was the offshoot of a British company called SCL Group (formerly Strategic Communication Laboratories) and was set up as a U.S. <u>subsidiary company</u> of SCL. During its early days, it targeted Facebook and other <u>social media platforms</u> to run advertisements based on user profiles, including ads for political campaigns.^[3]

In 2018, the company became the subject of widespread criticism following reports that it purchased data from a British academic concerning the personal psychological profile of roughly 270,000 Facebook users.^{[4] [5]} Due to the terms and conditions of the <u>Facebook platform</u> at that time, the company was able to indirectly access data from the Facebook friends of those users.^[6] In this

Boiy, E., & Moens, M.-F. (2009). A machine learning approach to sentiment analysis in multilingual Web texts. Information Retrieval Journal, 12(5), 526–558. https://doi.org/10.1007/s10791-008-9070-z



Recognizing emotions based on cell phone use



Burns, M. N., Begale, M., Duffecy, J., Gergle, D., Karr, C. J., Giangrande, E., & Mohr, D. C. (2011). Harnessing Context Sensing to Develop a Mobile Intervention for Depression. Journal of Medical Internet Research, 13(3), 11. https://doi.org/10.2196/jmir.1838



And many more...





Contributions of Libraries to Machine Learning

Lots of Words



Ontologies

CLASSIFICATION ADOPTÉE.

Dicotylédonées

PLANTES PHANÉROGAMES & pourvues de Cotylédons

Monocotylédonées

CLASSES. Pag. Dialypétales hypogynes..... périgynes..... 81 Gamopétaleshypogynes..... 175 G périgynes..... 241 ASSIFICATION Apétales non amentacées.... 299 amentacées..... 331 Gymnospermes ou Conifères.. 353 Pétaloïdes supérovariées..... 367 ADOPTÉE inférovariées..... 382 Glumacées exalbuminées..... 397 albuminées..... 404

Provenance



WordNet Search - 3.1 - WordNet home page - Glossary - Hel

Word to search for: mannish

Search WordNet

Display Options: (Select option to change) V Change

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations Display options for sense: (gloss) "an example sentence"

Adjective

- <u>S:</u> (adj) mannish (resembling or imitative of or suggestive of a man rather than a woman) "a mannish stride"
- <u>S</u>: (adj) mannish (characteristic of a man as distinguished from a woman) "true mannish arrogance"

Artificial intelligence history, and libraries: History and Legacy of Library Contributions to machine learning



Wilhelmina Randtke, Head of Libraries Technologies and Systems, Georgia Southern University Libraries NASIG Autumn, October 17, 2023, Virtual

