### CORSO DI DEEP LEARNING EX WEB MINING E RETRIEVAL

Corso di Laurea in Informatica, Ing. Gestionale, Ing. Internet, ICT Internet Engineering, Ing. Informatica (a.a. 2024-2025)

Roberto Basili

# **Overview**

- Motivations for the Course
- The origin of Machine Learning
- ML and Web Data
- Generative AI:
  - Origin and Historical Development
  - Training and Using GenAI methods
  - Limitations and Risks

### **Overview**

- Deep Learning and Web Applications: Motivations & Perspectives
  - Web, User-generated contents, Social Media
  - The role of *learning*. What is Machine Learning?
  - Neural Network training: the role of Depth
- Advanced Deep Learning methods
  - Convolutional Networks for Image Processing
  - Transformers and Data Encoding
  - Generative AI: Large Language Models and Fondational Models
  - Multimodal Deep Learning architectures
- Applications
  - Sentence and Text Classification, Textual Inference
  - Generative AI, Dialogue and Human Robotic Interfaces
  - Prompt Engineering for Problem Solving, Reasoning and Domain Adaptation

### **Internet statistics**

50

(source: https://datareportal.com/global-digital-overview)



SOURCES: KEPIOS ANALYSIS; ITU; GSMA INTELLIGENCE; EUROSTAT, GOOGLE'S ADVERTISING RESOURCES; CNNIC; KANTAR & IAMAI; GOVERNMENT RESOURCES; UNITED NATIONS. COMPARABILITY: SOURCE AND BASE CHANGES. ALL FIGURES USE THE LATEST AVAILABLE DATA, BUT SOME SOURCES DO NOT PUBLISH REGULAR UPDATES, SO FIGURES FOR RECENT PERIODS MAY UNDER-REPRESENT ACTUAL USE. SEE NOTES ON DATA

are. **(O)**Meltwater



### DIGITAL GROWTH

CHANGE IN THE USE OF CONNECTED DEVICES AND SERVICES OVER TIME











#### **MEDIA USE**

THE PERCENTAGE OF INTERNET USERS AGED 16+ WHO CONSUME EACH MEDIA TYPE



97.8% **INTERNET: MOBILE PHONE** DATAREPORTAL 93.0% **SOCIAL MEDIA** 90.6% **INTERNET: LAPTOP, DESKTOP, OR TABLET** 89.7% **TV: LINEAR OR BROADCAST** 77.1% **TV: STREAMING OR ONLINE** 67.9% **PRESS: ONLINE** 66.4% **RADIO: BROADCAST** 65.3% **MUSIC STREAMING** 63.7% PODCASTS 62.2% **PRESS: PHYSICAL PRINT** 

28

SOURCE: GWI [Q3 2024]. COMPARABILITY: CHANGES IN AUDIENCE COMPOSITION AND SURVEY METHODOLOGY. SEE NOTES ON DATA. POTENTIAL MISMATCHES: THE VALUES SHOWN HERE ARE BASED ON THE TIME THAT PEOPLE SAY THEY SPEND CONSUMING EACH MEDIA TYPE, AND MAY DIFFER FROM SIMILAR DATA POINTS SHOWN ELSEWHERE IN THIS REPORT THAT REFLECT OTHER BEHAVIOURS SUCH AS USE OF DEVICES TO ACCESS THE INTERNET...



### MAIN REASONS FOR USING THE INTERNET

PRIMARY REASONS WHY INTERNET USERS AGED 16+ USE THE INTERNET









### TOP TYPES OF WEBSITES VISITED AND APPS USED

PERCENTAGE OF INTERNET USERS AGED 16+ WHO HAVE VISITED OR USED EACH KIND OF DIGITAL PROPERTY IN THE PAST MONTH





105 SOURCE: GWI (Q3 2024), COMPARABILITY: CHANGES IN AUDIENCE COMPOSITION AND SURVEY METHODOLOGY. SEE NOTES ON DATA. THE TREATMENT OF THE QUESTION USED TO INFORM THIS CHART OF THE COMPARABLE WITH VALUES PUBLISHED IN PREVIOUS REPORTS.



### **MOST COMMON LANGUAGES FOR WEB CONTENT**





#	LANGUAGE		SHARE OF WEBSITES	SHARE OF POPULATION	#	LANGUAGE		SHARE OF WEBSITES	SHARE OF POPULATION
01	ENGLISH		49.4%	18.5%	12	PERSIAN		1.2%	1.0%
02	SPANISH		6.0%	6.8%	13	CHINESE		1.2%	18.8%
03	GERMAN	D	5.6%	1.6%	14	VIETNAMESE	D	1.1%	1.1%
04	JAPANESE		5.0%	1.5%	15	INDONESIAN		1.1%	2.4%
05	FRENCH		4.4%	3.8%	16	CZECH		1.0%	0.2%
06	RUSSIAN		3.9%	3.1%	17	KOREAN		0.8%	1.0%
07	PORTUGUESE		3.8%	3.2%	18	UKRAINIAN		0.6%	0.5%
08	ITALIAN		2.7%	0.8%	19	HUNGARIAN		0.6%	0.2%
09	DUTCH, FLEMISH		2.1%	0.3%	20	romanian		0.5%	0.3%
10	POLISH		1.8%	0.5%	21	ARABIC		0.5%	9.3%
11	TURKISH		1.8%	1.1%	22	SWEDISH		0.5%	0.2%

160

SOURCES: W3TECHS; ETHNOLOGUE; U.N; KEPIOS ANALYSIS. NOTES: WEBSITE LANGUAGES AS DEFINED BY W3TECHS. LANGUAGES INCLUDE SUB-LANGUAGES (E.G. "CHINESE" INCLUDES MANDARIN, YUE, ETC.). FIGURES IN THE "SHARE OF WEBSITES" COLUMN ARE BASED ON W3TECH'S ANALYSIS OF TRAFFIC FOR THE TOP 10 MILLION WEBSITES, AS RANKED BY ALEXA INTERNET. FIGURES IN THE "SHARE OF POPULATION" COLUMN COMPARE THE TOTAL NUMBER OF FIRST-AND SECOND-LANGUAGE SPEAKERS OF EACH LANGUAGE (INCLUDING SUB-LANGUAGES) TO THE TOTAL GLOBAL POPULATION.



### **SOCIAL MEDIA TIMELINE: USER IDENTITIES**



NUMBER OF SOCIAL MEDIA USER IDENTITIES OVER TIME

FEB 2025

345



SOURCES: KEPIOS ANALYSIS; COMPANY ADVERTISING RESOURCES AND ANNOUNCEMENTS; CNNIC; BETA RESEARCH CENTER; MEDIASCOPE; OCDH; ARAB SOCIAL MEDIA REPORT; TECHRASA; CAFEBAZAAR NOTES: BASED ON ACTIVE USER FIGURES FOR THE LARGEST SOCIAL NETWORK OR INSTANT MESSAGING SERVICE BY GEOGRAPHY AT THE START OF EACH YEAR. ADVISORY: SOCIAL MEDIA USER IDENTITIES MAY NOT REPRESENT UNIQUE INDIVIDUALS, COMPARABILITY: SOURCE AND METHODOLOGY CHANGES; BASE REVISIONS, SEE NOTES ON DATA. are. social



SOURCES: KEPIOS ANALYSIS, COMPANY ADVERTISING RESOURCES AND ANNOUNCEMENTS; CNNIC; BETA RESEARCH CENTER; OCDH GUIDANCE; A VALUE OF 100 MEANS THAT THE SHARE OF TOTAL USERS EQUALS THE SHARE OF GLOBAL POPULATION. A VALUE BELOW 100 MEANS THAT THE SHARE OF TOTAL USERS IS BELOW THE SHARE OF GLOBAL POPULATION. A VALUE ABOVE 100 MEANS THE SHARE OF TOTAL USERS USERS IS ABOVE THE SHARE OF GLOBAL POPULATION. ADVISORY: SOCIAL MEDIA USER IDENTITIES MAY NOT REPRESENT UNIQUE INDIVIDUALS. NOTES: FIGURES MAY NOT SUM TO 100% DUE TO ROUNDING. REGIONS BASED ON THE UNITED NATIONS GEOSCHEME. COMPARABILITY: SOURCE AND METHODOLOGY CHANGES; BASE REVISIONS. SEE NOTES ON DATA.

353

are social **(O)** Meltwater

### MAIN REASONS FOR USING SOCIAL MEDIA

PRIMARY REASONS WHY SOCIAL MEDIA USERS AGED 16+ USE SOCIAL MEDIA PLATFORMS



KEEPING IN TOUCH WITH FRIENDS AND FAMILY				50.8%
FILLING SPARE TIME	DATAREPORTAL	Gvvi.	39.0%	
READING NEWS STORIES			34.5%	
FINDING CONTENT (E.G. ARTICLES, VIDEOS)		30.5%		
SEEING WHAT'S BEING TALKED ABOUT		29.2%		
FINDING PRODUCTS TO PURCHASE		27.3%		
FINDING INSPIRATION FOR THINGS TO DO AND BUY		27.2%		
WATCHING OR FOLLOWING SPORTS	23.6%			
WATCHING LIVE STREAMS	23.4%			
SEEING CONTENT FROM YOUR FAVOURITE BRANDS	22.5%			
MAKING NEW CONTACTS	22.1%			
WORK-RELATED NETWORKING OR RESEARCH	22.1%			
SHARING AND DISCUSSING OPINIONS WITH OTHERS	21.9%			
FOLLOWING CELEBRITIES OR INFLUENCERS	19.7%			
AVOIDING MISSING OUT ON THINGS (FOMO)	19.5%			

362

SOURCE: GWI (Q3 2024). NOTE: FIGURES REPRESENT THE SHARE OF INTERNET USERS AGED 16+ WHO REPORT USING AT LEAST ONE SOCIAL MEDIA PLATFORM OR MESSENGER SERVICE IN THE PAST MONTH. COMPARABILITY: CHANGES IN AUDIENCE COMPOSITION AND SURVEY METHODOLOGY. SEE NOTES ON DATA.



### **FAVOURITE SOCIAL MEDIA PLATFORMS**

PERCENTAGE OF ACTIVE SOCIAL MEDIA USERS WHO SAY THAT EACH OPTION IS THEIR "FAVOURITE" SOCIAL MEDIA PLATFORM

NOTE: YOUTUBE IS NOT AVAILABLE AS AN ANSWER OPTION IN THE SURVEY QUESTION THAT INFORMS THESE TABLES

#### FAVOURITE SOCIAL MEDIA PLATFORMS AMONGST FEMALE INTERNET USERS

#### FAVOURITE SOCIAL MEDIA PLATFORMS AMONGST MALE INTERNET USERS

SOCIAL AGE SOCIAL AGE AGE AGE AGE AGE AGE AGE AGE AGE PLATFORM 16 - 2425 - 3435 - 4445-54 16 - 2425 - 3435 - 4445 - 5455 - 64PLATFORM 55 - 64INSTAGRAM 24.8% 20.8% 16.0% 13.9% 10.2% INSTAGRAM 26.6% 18.4% 11.7% 9.4% 6.1% WHATSAPP 12.7% 14.3% 15.4% 17.3% 21.1% WHATSAPP 14.1% 14.7% 17.3% 19.5% 21.9% GWI. GWI. FACEBOOK 5.9% 10.6% 13.2% 14.9% 17.1% FACEBOOK 7.8% 13.5% 14.7% 16.4% 17.5% WECHAT 8.2% 11.7% 15.5% 13.6% 12.0% WECHAT 8.2% 11.8% 16.0% 14.4% 14.4% TIKTOK 15.5% 11.3% 7.8% 6.9% 5.0% TIKTOK 9.6% 7.1% 5.7% 5.3% 4.3% 8.1% 5.4% 8.5% DOUYIN 6.6% 9.8% 6.4% DOUYIN 6.5% 8.2% 6.8% 6.6% Х 3.1% 2.1% 1.8% 1.6% 1.7% Х 4.2% 4.6% 4.0% 3.8% 3.3% TELEGRAM 2.6% 2.1% 2.0% 2.1% 2.2% TELEGRAM 4.0% 3.2% 3.2% 2.6% 2.5% 1.7% 2.5% 2.5% 2.7% 3.0% 1.8% 2.4% 2.2% 2.4% MESSENGER MESSENGER 2.3% LINE 0.5% 0.8% 1.2% 2.6% 3.5% LINE 0.7% 1.0% 1.2% 2.4% 2.9%

SOURCE: GWI (Q3 2024). NOTES: YOUTUBE IS NOT AVAILABLE AS AN ANSWER FOR THIS QUESTION IN GWI'S SURVEY. TABLE VALUES IN BOLD ORANGE TEXT IDENTIFY THE TOP VALUE FOR THE RESPECTIVE AGE GROUP. DATA ONLY INCLUDES INTERNET USERS WHO HAVE USED AT LEAST ONE SOCIAL MEDIA PLATFORM IN THE PAST MONTH. SURVEY RESPONDENTS COULD CHOOSE FROM OTHER OPTIONS NOT SHOWN ON THIS CHART, SO VALUES MAY NOT SUM TO 100%. WE REPORT GWI'S VALUES FOR TIKTOK IN CHINA SEPARATELY AS DOUYIN, AS PER BYTEDANCE'S CORPORATE REPORTING. COMPARABILITY: CHANGES IN AUDIENCE COMPOSITION AND SURVEY METHODOLOGY. SEE NOTES ON DATA.





### **USE OF SOCIAL MEDIA FOR BRAND RESEARCH**



PERCENTAGE OF INTERNET USERS AGED 16+ WHO USE SOCIAL MEDIA TO RESEARCH BRANDS AND PRODUCTS THEY'RE CONSIDERING BUYING



SOURCE: GWI (Q3 2024). NOTE: VALUES REPRESENT PEOPLE WHO SAY THAT THEY RESEARCH BRANDS AND PRODUCTS THAT THEY'RE CONSIDERING BUYING ON AT LEAST ONE OF THE FOLLOWING: SOCIAL NETWORKS, QUESTION AND ANSWER SITES (E.G. QUORA), FORUMS AND MESSAGE BOARDS, MESSAGING AND LIVE CHAT SERVICES, MICRO-BLOGS (E.G. X / TWITTER), BLOGS ON PRODUCTS / BRANDS, VLOGS (I.E. BLOGS RECORDED IN A VIDEO FORMAT), AND ONLINE PINBOARDS (E.G. PINTEREST). COMPARABILITY: CHANGES IN AUDIENCE COMPOSITION AND SURVEY METHODOLOGY. SEE NOTES ON



### SOCIAL MEDIA ACCOUNT TYPES FOLLOWED

PERCENTAGE OF SOCIAL MEDIA USERS IN EACH AGE GROUP WHO FOLLOW EACH TYPE OF ACCOUNT ON SOCIAL MEDIA

\_\_\_\_\_\_\_

GLOBAL OVERVIEW

AGE 16 TO 24		AGE 25 TO 3	4	AGE 35 TO 4	4	AGE 45 TO 5	4	AGE 55 TO 6	4	AGE 65+*		
Friends, family, or other People you know	48.8%	FRIENDS, FAMILY, OR OTHER PEOPLE YOU KNOW	46.8%	FRIENDS, FAMILY, OR OTHER PEOPLE YOU KNOW	48.6%	FRIENDS, FAMILY, OR OTHER PEOPLE YOU KNOW	49.8%	FRIENDS, FAMILY, OR OTHER PEOPLE YOU KNOW	51.0%	FRIENDS, FAMILY, OR OTHER PEOPLE YOU KNOW	51.3%	
ENTERTAINMENT, MEMES, OR PARODY ACCOUNTS	37.4%	ACTORS, COMEDIANS, OR OTHER PERFORMERS	32.8%	TV SHOWS OR CHANNELS	29.3%	TV SHOWS OR CHANNELS	27.1%	TV SHOWS OR CHANNELS	25.5%	TV SHOWS OR CHANNELS	16.5%	
ACTORS, COMEDIANS, OR OTHER PERFORMERS	35.5%	ENTERTAINMENT, MEMES, OR PARODY ACCOUNTS	31.8%	ACTORS, COMEDIANS, OR OTHER PERFORMERS	28.3%	RESTAURANTS, CHEFS, OR FOOD PERSONALITIES	24.3%	RESTAURANTS, CHEFS, OR FOOD PERSONALITIES	21.0%	RESTAURANTS, CHEFS, OR FOOD PERSONALITIES	12.5%	
BANDS, SINGERS, OR OTHER MUSICIANS	32.2%	TV SHOWS OR CHANNELS	29.1%	RESTAURANTS, CHEFS, OR FOOD PERSONALITIES	27.5%	ACTORS, COMEDIANS, OR OTHER PERFORMERS	23.9%	COMPANIES AND BRANDS YOU PURCHASE FROM	19.3%	SPORTS PEOPLE AND TEAMS	11.9%	
INFLUENCERS OR OTHER EXPERTS	28.3%	Bands, singers, or other musicians	28.4%	ENTERTAINMENT, MEMES, OR PARODY ACCOUNTS	26.1%	Bands, Singers, Or Other musicians	23.2%	BANDS, SINGERS, OR OTHER MUSICIANS	18.7%	COMPANIES AND BRANDS YOU PURCHASE FROM	11.8%	
SPORTS PEOPLE AND TEAMS	26.7%	RESTAURANTS, CHEFS, OR FOOD PERSONALITIES	28.1%	BANDS, SINGERS, OR OTHER MUSICIANS	25.7%	CONTACTS RELEVANT TO YOUR WORK	22.3%	CONTACTS RELEVANT TO YOUR WORK	18.0%	BANDS, SINGERS, OR OTHER MUSICIANS	10.6%	
TV SHOWS OR CHANNELS	25.8%	INFLUENCERS OR OTHER EXPERTS	25.0%	COMPANIES AND BRANDS YOU PURCHASE FROM	23.6%	COMPANIES AND BRANDS YOU PURCHASE FROM	22.0%	ACTORS, COMEDIANS, OR OTHER PERFORMERS	17.4%	JOURNALISTS OR NEWS COMPANIES	9.7%	
GAMING EXPERTS OR GAMING STUDIOS	24.7%	SPORTS PEOPLE AND TEAMS	24.5%	CONTACTS RELEVANT TO YOUR WORK	23.4%	ENTERTAINMENT, MEMES, OR PARODY ACCOUNTS	21.6%	COMPANIES AND BRANDS YOU'RE CONSIDERING PURCHASING FROM	16.2%	ACTORS, COMEDIANS, OR OTHER PERFORMERS	8.9%	
RESTAURANTS, CHEFS, OR FOOD PERSONALITIES	23.5%	COMPANIES AND BRANDS YOU PURCHASE FROM	21.8%	SPORTS PEOPLE AND TEAMS	22.9%	COMPANIES RELEVANT TO YOUR WORK	19.7%	JOURNALISTS OR NEWS COMPANIES	15.3%	EVENTS YOU'RE ATTENDING	8.9%	
BEAUTY EXPERTS	20.0%	CONTACTS RELEVANT TO YOUR WORK	21.1%	COMPANIES RELEVANT TO YOUR WORK	22.5%	SPORTS PEOPLE AND TEAMS	19.5%	SPORTS PEOPLE AND TEAMS	15.0%	COMPANIES AND BRANDS YOU'RE CONSIDERING PURCHASING FROM	8.6%	

403

SOURCE: GWI (Q3 2024). NOTES: ONLY INCLUDES INTERNET USERS IN EACH AGE GROUP WHO HAVE USED AT LEAST ONE SOCIAL MEDIA PLATFORM IN THE PAST MONTH. (\*) DATA FOR AUDIENCES AGED 65+ ARE NOT YET AVAILABLE. IN ALL COUNTRIES, SO FINDINGS FOR AUDIENCES AGED 65+ MAY NOT BE DIRECTLY COMPARABLE WITH THOSE FOR OTHER AGE GROUPS. COMPARABILITY: CHANGES IN AUDIENCE COMPOSITION AND SURVEY METHODOLOGY. SEE NOTES ON DATA.



### **MOST USED INSTAGRAM HASHTAGS**

HASHTAGS THAT HAVE BEEN USED IN THE GREATEST NUMBER OF INSTAGRAM POSTS (ALL TIME)



# HASHTAG	POSTS	# HASHTAG	POSTS	#	HASHTAG	POSTS	#	HASHTAG	POSTS
01 #LOVE	2.15 B	11 #NATURE	836.2 M	21	#TBT	589.8 M	31	#MUSIC	489.0 M
02 #INSTAGOOD	1.91 B	12 #TRAVEL	764.6 M	22	#INSTALIKE	574.7 M	32	2 #LIKE4LIKE	476.2 M
03 #INSTAGRAM	1.47 B	13 #trending	755.6 M	23	#REPOST	570.2 M	33	3 #LIFE	470.2 M
04 #FASHION	1.20 B	14 #EXPLORE	750.2 M	24	#SUMMER	570.2 M	34	#FAMILY	461.4 M
05 #PHOTOGRAPHY	1.12 B	15 #FOLLOW	748.1 M	25	#BEAUTY	565.0 M	35	5 <b>#</b> ME	460.0 M
06 #ART	1.10 B	16 #PICOFTHEDAY	737.3 M	26	#FITNESS	558.0 M	36	#FRIENDS	458.8 M
07 #PHOTOOFTHEDAY	1.09 B	17 #INSTADAILY	731.7 M	27	#FOLLOWME	557.0 M	37	′ #FUN	454.0 M
O8 #REELS	954.9 M	18 #HAPPY	726.8 M	28	#EXPLOREPAGE	547.7 M	38	8 #SELFIE	450.0 M
09 #VIRAL	868.6 M	19 #CUTE	687.0 M	29	#FOOD	536.9 M	39	9 #SMILE	446.6 M
10 #BEAUTIFUL	853.7 M	20 #STYLE	682.2 M	30	#рното	521.9 M	40	) #GIRL	432.7 M

SOURCE: KEPIOS ANALYSIS, BASED ON DATA PUBLISHED ON INSTAGRAM. NOTES: WHERE LETTERS ARE SHOWN NEXT TO FIGURES IN THE "POSTS" COLUMN, "K" DENOTES THOUSANDS (E.G. "123 K" = 123,000), "M" DENOTES MILLIONS (E.G. "1.23 M" = 1,230,000), AND "B" DENOTES BILLIONS (E.G. "1.23 B" = 1,230,000,000). WHERE NO LETTER IS PRESENT, VALUES ARE SHOWN AS IS. INSTAGRAM NO LONGER PUBLISHES POST COUNTS FOR SOME POPULAR HASHTAGS, SO THESE TAGS ARE NOT INCLUDED IN THIS RANKING. COMPARABILITY: POST COUNTS ARE LIABLE TO CONSTANT CHANGE. FIGURES CORRECT AT THAG OF EROPT PRODUCTION



475

### **TOP TIKTOK HASHTAGS: POSTS**

HASHTAGS THAT HAVE BEEN USED IN THE GREATEST NUMBER OF POSTS ON TIKTOK (ALL TIME)



GLOBAL OVERVIEW

# HASHTAG	тіктокѕ	# HASHTAG	тіктокѕ	#	HASHTAG	т	стока		# HASHTAG	тіктокѕ
01 #CAPCUT	7.20 B	11 #COMEDY	914.4 M	21	#TREND		362.9 M	3	اکسبلور# ۱	181.1 M
O2 #FYP	6.60 B	12 #PARATI	907.5 M	22	#XYZBCA		344.1 M		2 #GOVIRAL	180.3 M
03 #FORYOU	4.80 B	13 #VIRALVIDEO	868.8 M	23	#FYPAGE		327.0 M	3	3 #FOLLOW	Цлан 177.6 М
04 #FORYOUPAGE	3.50 B	14 #PEK	644.1 M	24	#РЕКОМЕНДАЦИИ		304.9 M		4 #HUMOR	160.7 M
05 #DUET	3.20 В	15 #FY	590.6 M	25	#XUHUONG		296.6 M	з	5 #EDIT	143.4 M
06 #FYPシ	2.70 В	16 #GREENSCREEN	468.9 M	26	#LIKE	2	280.8 M	3	6 #EXPLORE	141.9 M
07 #VIRAL	2.60 B	17 #FYPPPPPPPPPPPPP	PPPPP 456.0 M	27	#FUNNY	ъ.	279.3 M	3	7 #ANIME	140.6 M
08 #TIKTOK	1.60 B	18 #TIKTOKINDIA	409.5 M	28	#VIDEO		236.4 M	3	8 #MEME	129.4 M
09 #TRENDING	1.10 B	19 #LOVE	390.7 M	29	#POURTOI		197.2 M	з	9 #TIK_TOK	124.3 M
10 #FYPシ <sup>v</sup> IRAL	999.9 M	20 #VIRALTIKTOK	373.6 M	30	#TIKTOKLOVER		192.6 M	4	0 #DANCE	123.3 M

SOURCE: KEPIOS ANALYSIS, BASED ON DATA PUBLISHED ON TIKTOK. NOTES: THE SAME TIKTOK MAY INCLUDE MULTIPLE HASHTAGS, SO ONE TIKTOK MAY REGISTER IN THE TIKTOKS COUNT FOR MULTIPLE HASHTAGS, WHERE LETTERS ARE SHOWN NEXT TO FIGURES, "M" DENOTES MILLIONS, "B" DENOTES BILLIONS, AND "T" DENOTES TRILLIONS, WHERE NO LETTER IS PRESENT, VALUES ARE SHOWN AS IS. POST COUNTS ROUNDED AT SOURCE. ADVISORY: THERE IS NO SIMPLE WAY TO IDENTIFY TOP HASHTAGS ON TIKTOK, SO THIS LIST MAY INADVERTENTLY MISS HASHTAGS THAT OUTPERFORM SOME OF THOSE FEATURED IN THIS LIST. COMPARABILITY: POST COUNTS ARE LIABLE TO CONSTANT CHANGE. FIGURES CORRECT AT TIME OF REPORT PRODUCTION.



### **MOST FOLLOWED HASHTAGS ON LINKEDIN**

HASHTAGS WITH THE GREATEST NUMBER OF FOLLOWERS ON LINKEDIN

FEB 2025

489

#	HASHTAG	FOLLO	WERS	# HASHTAG	FOLLO	WERS	#	HASHTAG	FOLLO	WERS		# HASHTAG	FOLLO	WERS
01	#INDIA		66.8 M	11 #CAREERS		22.2 M	21	#BRANDING		17.8 M	3	1 #MONEY		13.9 M
02	#INNOVATION		38.3 M	12 #MARKETS	~	21.9 M	22	2 #PROFESSIONALV	WOMEN	17.6 M	3	2 #bankingindustry	~	13.4 M
03	#MANAGEMENT		35.4 M	13 #STARTUPS	D.,	20.9 M	23	#ADVERTISINGAN	NDMARKETING	16.9 M	3	3 #SUSTAINABILITY	D.,	13.1 M
04	#HUMANRESOURCES		32.8 M	14 #MARKETING		20.1 M	24	#gender		16.4 M	з	4 #ALTERNATIVEENERGY		12.6 M
05	#DIGITALMARKETING		27.0 M	15 #SOCIALMEDIA		19.4 M	25	5 #WOMENINSCIE	NCE	16.3 M	3	5 #PERSONALBRANDING	G	10.3 M
06	#TECHNOLOGY	~	26.1 M	16 #VENTURECAPIT	AL	19.0 M	20	6 #FEMINISM	~	16.0 M	з	6 #HIRINGANDPROMOT	IION	9.9 M
07	#CREATIVITY	D,,	24.8 M	17 #SOCIALNETWO	DRKING	18.7 M	27	= #MOTIVATION	D.,	14.5 M	3	7 #HEALTHCARE		9.6 M
08	#FUTURE		24.2 M	18 #LEANSTARTUPS		18.7 M	27	= #PERSONALDEVE	LOPMENT	14.5 M	3	8 #EDUCATION		9.5 M
09	#FUTURISM		23.1 M	19 #ECONOMY		18.4 M	29	#INVESTING		14.2 M	3	9 #CUSTOMERRELATION	IS	8.5 M
10	#ENTREPRENEURSHIP		22.4 M	20 #ECONOMICS		17.8 M	30	#JOBINTERVIEWS		14.1 M	4			8.3 M







**GLOBAL OVERVIEW** 

>

# Locating Information ...



### SOURCES OF BRAND DISCOVERY

PERCENTAGE OF INTERNET USERS AGED 16+ WHO DISCOVER NEW BRANDS, PRODUCTS, AND SERVICES VIA EACH CHANNEL OR MEDIUM









### ... its consequences

FEB 2025

595

SOCIAL MEDIA ADVERTISING OVERVIEW

SPEND ON SOCIAL MEDIA ADVERTISING (IN U.S. DOLLARS) AND ITS SHARE OF THE DIGITAL ADVERTISING MARKET





SOURCE: STATISTA ADVERTISING & MEDIA OUTLOOK. SEE STATISTA.COM. NOTES: FIGURES REPRESENT ESTIMATES FOR FULLYEAR 2024, AND COMPARISONS WITH EQUIVALENT VALUES FOR THE PREVIOUS CALENDAR YEAR. FINANCIAL VALUES ARE IN U.S. DOLLARS. PERCENTAGE CHANGE VALUES ARE RELATIVE [I.E. AN INCREASE OF 20% FROM A STARTING VALUE OF 50% WOULD EQUAL 60%, NOT 70%]. "BPS" VALUES REPRESENT BASIS POINTS, AND INDICATE ABSOLUTE CHANGE. COMPARABILITY: BASE CHANGES. FIGURES ARE NOT COMPARABLE WITH PREVIOUS REPORTS.



### Did you know 2025? (on You Tube)

# More than 4,000 new books are published every day



# Do you know

Contains more information than a person was likely to come across in a lifetime in the 18th century...





# Dealing with real Social media data



### What is Data and Web Mining?

- Web Mining refers to a body of technologies currently needed for the exploitation of publicly available information from the Web and the IoT
  - Contents: data but also ... PEOPLE, LOCATIONS, EVENTS, CONCEPTS, TEMPORAL INFORMATION ...
    - **Relations:** 
      - Links within structured networks (retweets, follows, ...)
      - Thematic, interpersonal and semantic associations
      - Similarities and Analogies among people, behaviours, preferences
  - On-Line Structured and semi-structured resources (e.g. Wikipedia)
  - Textual, Multimedia and Multilingual Contents
  - Trends e time-related information (community on-line behaviours)
  - Opinions, Preferences, Expectations

# Why IR?

- The volumes involved in Web Mining pose the crucial problem of locating information beforehand
- Automatic information access is possible only if we solve the two major challenges
  - What is relevant
  - Where the relevant information is located
- Searching information corresponds to computing an <u>uncertain</u> <u>function</u> that models the <u>mapping</u> between information needs and the targeted data

# Machine Learning vs IR?

- Web mining involve heterogeneous information that is characterized search as strongly uncertainy process
- The available information is characterized by:
  - Incompleteness:
    - Short queries as an incomplete description of the information need
  - <u>Variability</u>: Wealth of data vs. heterogeneity of formats and access modes
    - Contents are dispersed in various forms across data sources
  - <u>Vague Requirements</u>
    - Information is often implicit (i.e. partially and qualitatively expressed) in the operational contexts
  - Subjectivity
    - Relevance depends on the user and not just on the contents
  - <u>Timeliness</u>
  - <u>Authority</u>

# Machine Learning vs. IR

- Uncertainty os so pervasive that exhaustive solutions (i.e. global optima) are not available or even not existing
- "Finding diamonds in the rough" (Fan Chung, UCSD)



### Machine Learning vs. IR

- ML technologies offer a wide variety of algorithms, strategies and techniques for the induction of sub-optimal, but surprisingly effective, solutions from available data
- Through *learning* data can be effectively used to suggest retrieval hypothesis, that are models of the *mapping* function (Learning to search)
- What is the target of the learned function? To improve computational aspects of the currently applied processes, such as
  - Decision Accuracy (i.e. best answers first): it depends on data semantics
  - System Responsiveness (i.e. reducing speed of the retrieval process)
  - Resource usage (i.e. more effective with less memory or input data)

# **Machine Learning**

- Machine learning is the study of computer algorithms that allow computer programs to automatically improve through experience. (Tom Mitchell, *Machine Learning*, McGraw-Hill, 1997)
- The evidence of the success of a learning process corresponds to the possibility of observing a measurable increment  $\Delta P$  of performances in solving a task C on the basis of experiences *E* that the agent is able to gather during its lifecycle.
- The nature and complexity of the learning ability is fully confined to the ability of characterizing the primitive notions here involved:
  - TASK C
  - PERFORMANCE P
  - EXPERIENCE E

### **Experience and Learning**

- Forms of experiences
- In chess games:
  - Data on previous matches, such as won challenges (or defeats) able to gather the utility (o inadequacy) of the strategies or moves there carried out.
  - Evaluation about individual moves offered by an external teacher (oracle, guide).
  - Adequacy of individual behaviour derived from self-observation, such as the capablility of analyising matches against itself based on a existing explicit model of the rules and strategies of the game.

### ML: a visual introduction

 See URL: <u>http://www.r2d3.us/visual-intro-to-machine-learning-part-</u> <u>1/?imm\_mid=0d76b4&cmp=em-data-na-newsltr\_20150826</u>



# The mathemathics of Learning

- Learning corresponds to the induction of mathematical function (i.e. the decision rules) that may have a discrete as well as a continuous behaviour:
  - Logical functions, (ad es., decision trees)
    - Learning the rules that better explain the data
    - Induction: Recursive search for necessary and sufficient conditions.
  - Probabilistic Approaches:
    - Learning what is *most likely* to be the better decision, according to an hypothesis about the input distribution (e.g., Bayesian classification)
    - Induction: Estimate the Posterior Probability (as parameters of known laws)
  - Metric Approaches
    - Decision as discrimination in metric spaces (e.g. linear and non linear functions)
      - K-NN
      - Linear Classifiers, perceptrons, Neural Networks, Support Vector Machines,...
      - Modeling as vectorial embedding, spectral analysis (space transformations)
    - Induction: determine the optimal parametrization from specific function classes (e.g. multilayer networks, polynomials of degree *n*)









# **Es. Decision Tree Learning**



# **Unsupervised Learning**

- When no oracle or knowledge of the task is available, learing may still be applied in several approaches:
  - Improve the current world model (knowledge acquisition/discovery)
  - Improve the efficiency of the currently available algorithms, through computational optimization
    - Better data structures representing the problem and the domain
    - Reduction of the processing steps required by the current models

# **Unsupervised Learning**

The induced hierachical model expresses a system of classes and relations able to imptove future interaction with the song collection

It has been discovered from data

No top-down design has been applied, as in knowledge engineering, but only bottom-up inferences (i.e. generalization from data)





# **Deep Learning**

#### See next lesson slides



### References

- Mitchell, Tom. M. 1997. *Machine Learning*. New York: McGraw-Hill.
- <u>Kernel machines, neural networks and graphical models</u>,
  P. Frasconi, A. Sperduti, A. Starita, Rivista AI\*IA Numero speciale per i "50 anni di IA", 2007.
- Very good video lecture by Andrew Ng (Stanford)
  http://academicearth.org/courses/machine-learning