

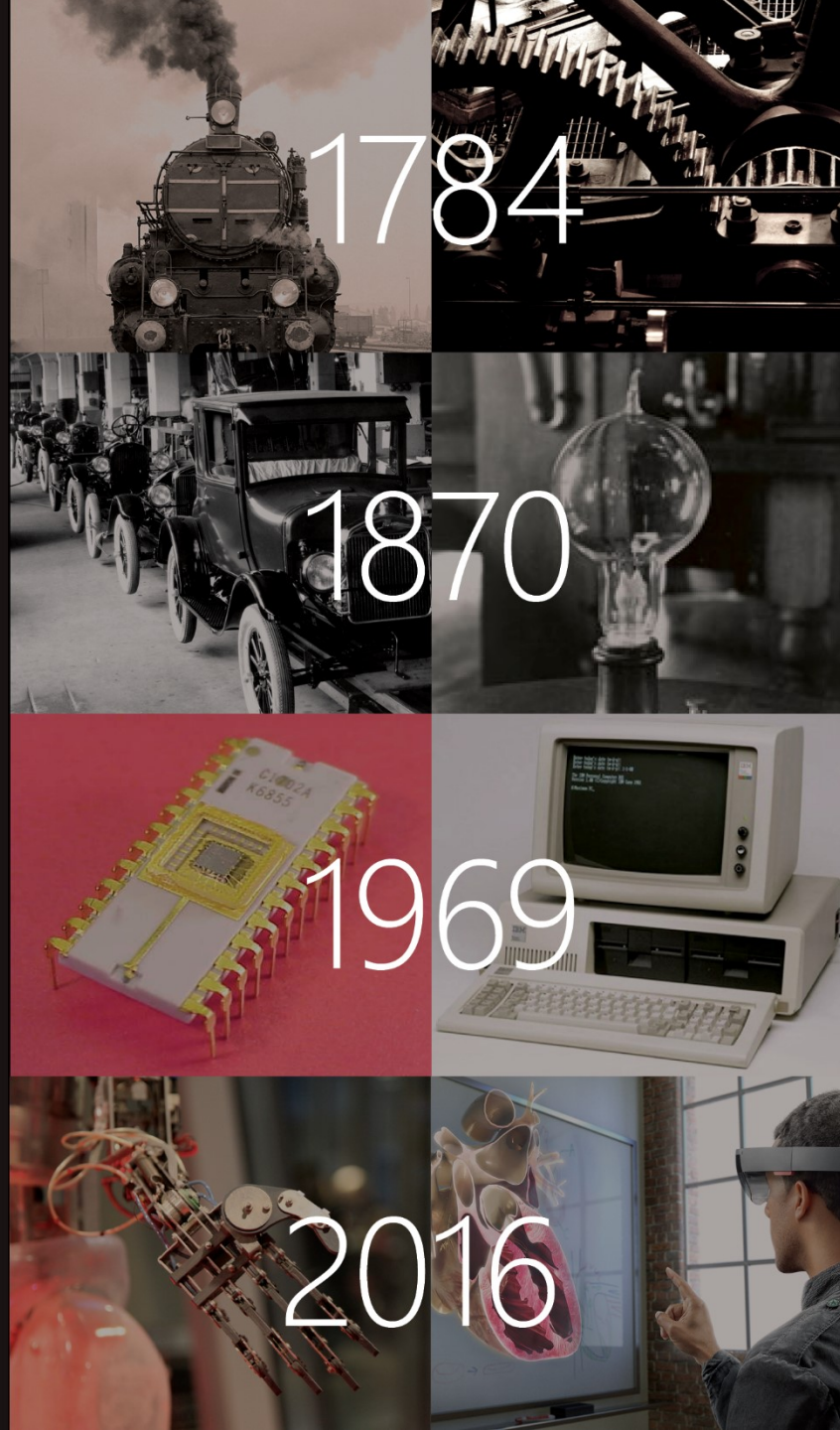
A stylized wireframe illustration of a human head in profile, facing right. The head is composed of a complex network of thin, light blue lines that form a mesh-like structure, representing neural networks or artificial intelligence. The background is a dark, solid blue.

L'intelligence artificielle aujourd'hui

Mieux connaître & comprendre afin de mieux gérer

Candi Carrera
Country Manager – Microsoft Luxembourg

Le chemin du progrès humain



1784

1870

1969

2016

1^{ère} révolution industrielle

Eau & vapeur, mécanisation de la production

2^{ème} révolution industrielle

Division du travail et énergie électrique pour production de masse

3^{ème} révolution industrielle

Électronique et TIC automatisant production

4^{ème} révolution industrielle

Convergence des technologies brouillant les frontières entre sphère physique, numérique et biologique (systèmes cyber-physiques)



Commun aux 4 révolutions

initié par personnes
Atteindre certains objectifs

*gagner de l'argent
devenir célèbre
simplement pour surmonter des défis
éliminer des inefficacités*

2ème révolution industrielle



1920

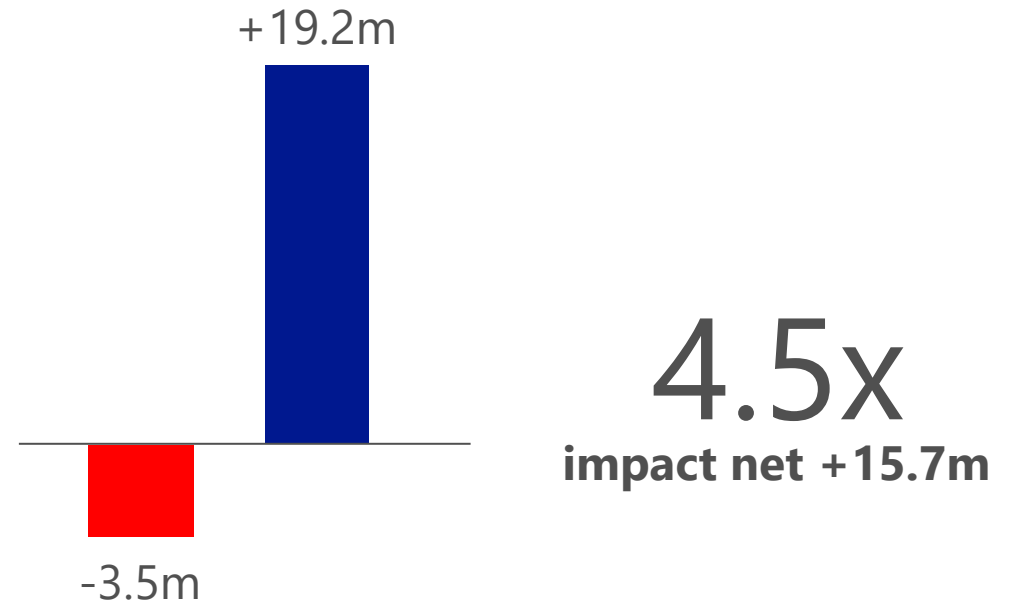


1930

3ème révolution industrielle



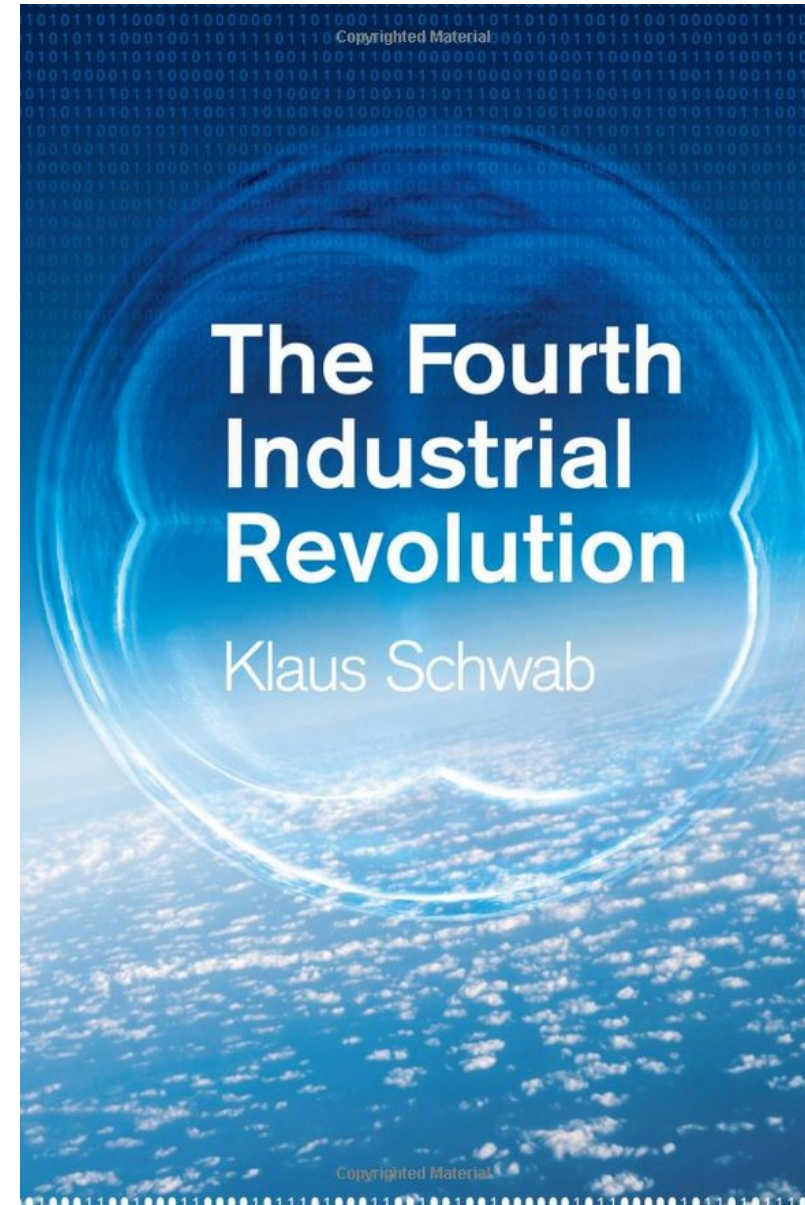
Marché du travail américain



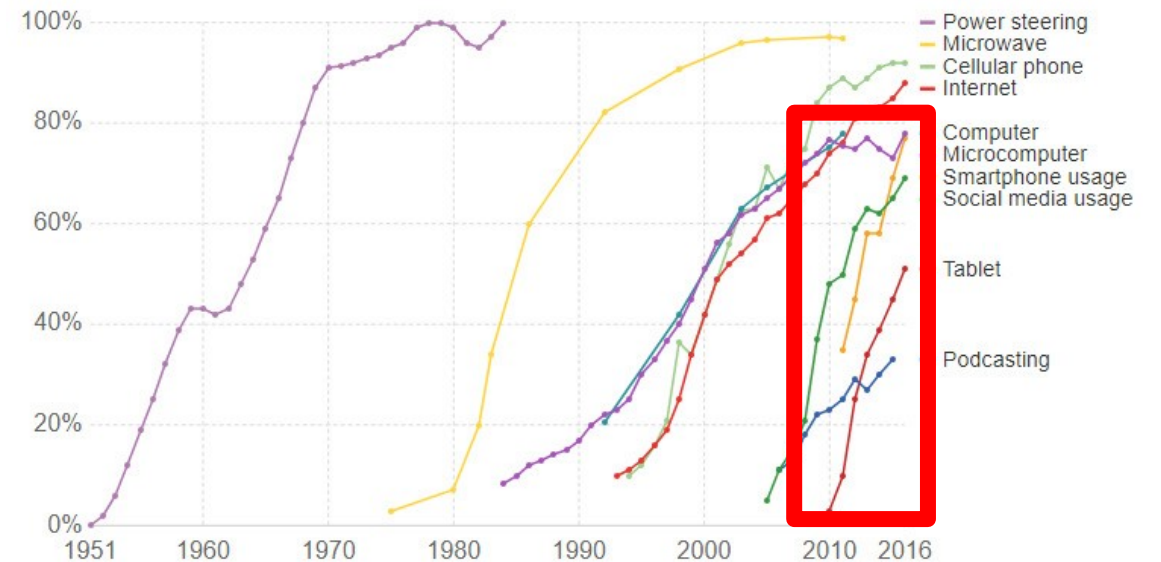
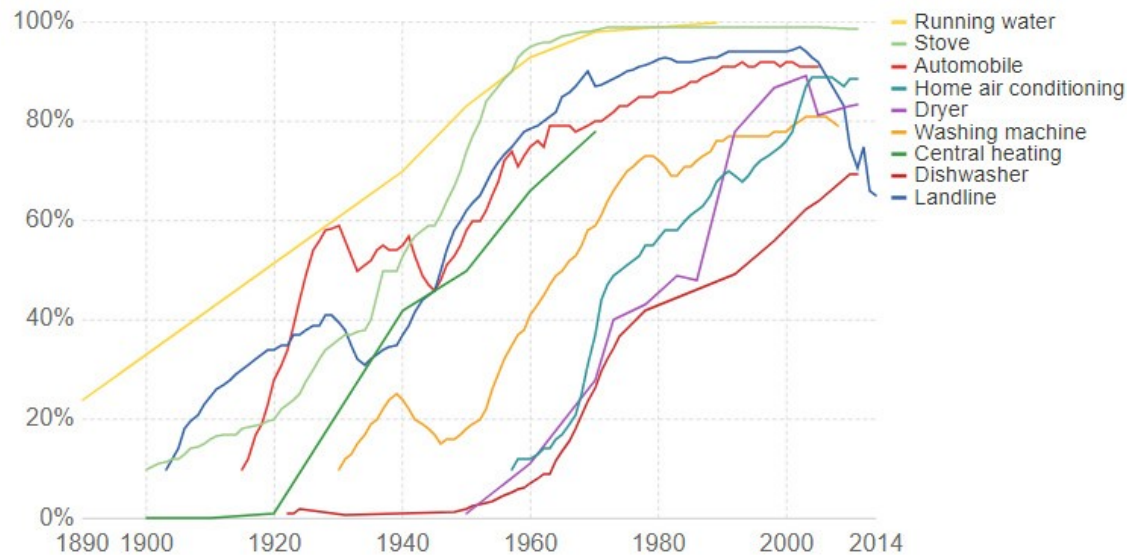
Source : McKinsey Global Institute, *Jobs Lost, Jobs Gained: Workforce transitions in a time of automation*, December 2017

4ème révolution industrielle

2016



Vitesse croissante - adoption technologique



5y avg

A wireframe profile of a human head, composed of numerous interconnected lines forming a mesh, set against a dark blue background. The head is facing right.

Conséquence

pour économie & entreprises

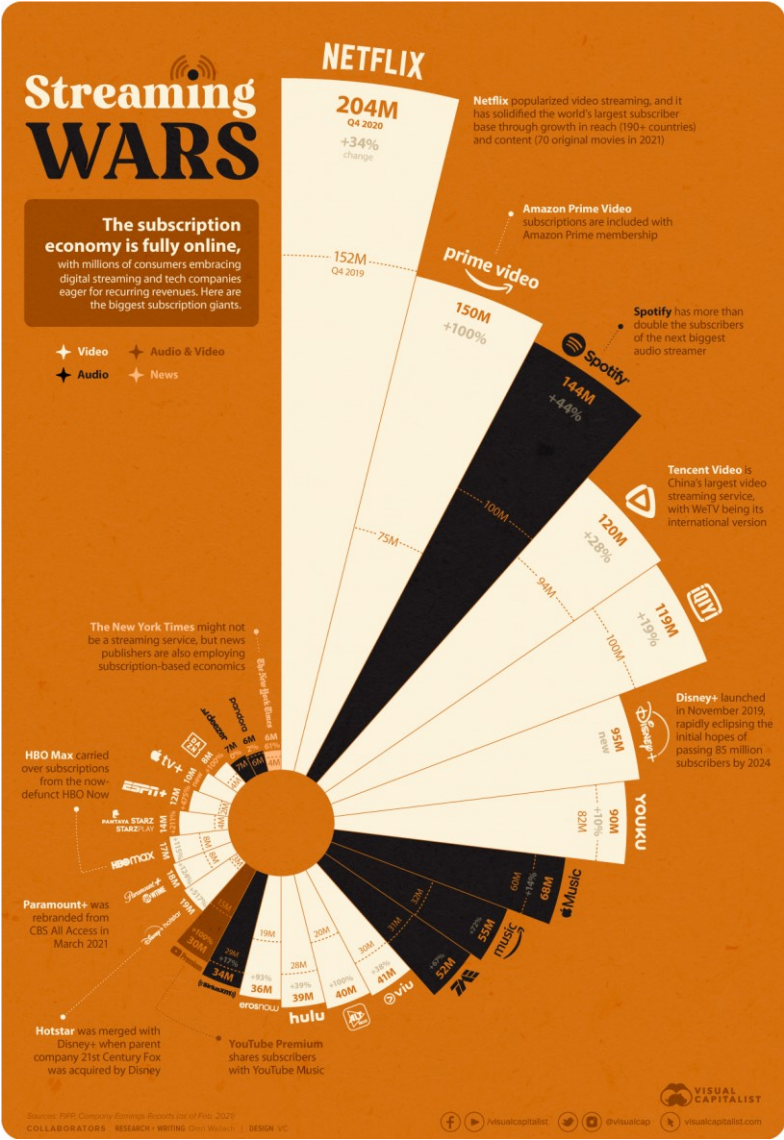
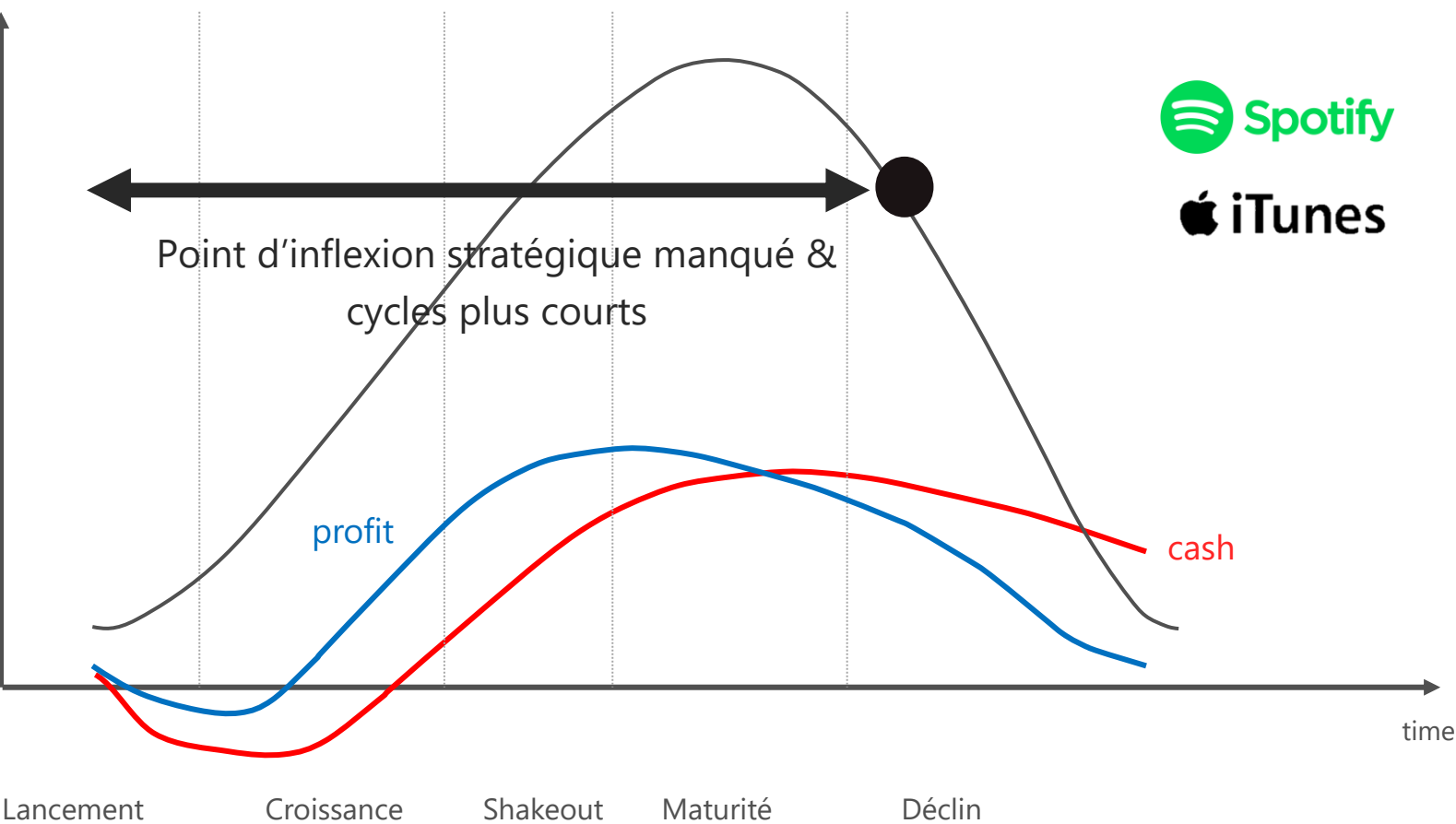
Facteurs à l'origine de perturbation du modèle d'affaires

- **Nouvelle pression** sur les prix et les marges
- Nouveaux concurrents émergeant d'endroits **inattendus**
- Dynamique de concentration mondiale & « **winner takes it all** »
- Modèles d'affaires plug-and-play & vitesse plus élevée



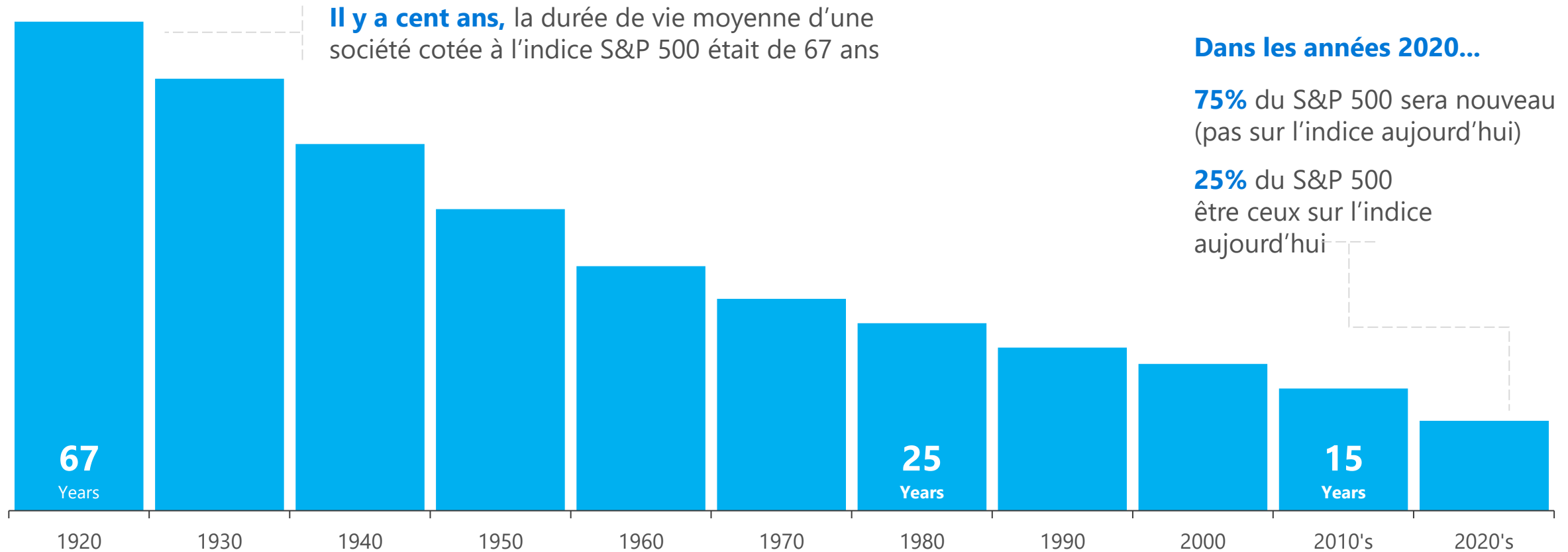
Source : Factors Driving Business Model Disruption, McKinsey

Courbe en S de croissance d'entreprise



Source : www.visualcapitalist.com/which-streaming-service-has-the-most-subscriptions/

Diminution du temps d'adaptation



Source: BBC / Professor Richard Foster of Yale University

<http://www.bbc.com/news/business-16611040>

DJ30 – Entrée moyenne et médiane (aujourd'hui)

Company name	Index	Ticker	Industry	Joined DJ30	in years
Procter & Gamble	NYSE	PG	Fast-moving consumer goods	26/05/1932	89,0
3M Company	NYSE	MMM	Conglomerate	09/08/1976	44,8
IBM	NYSE	IBM	Information technology	29/06/1979	41,9
Merck & Co.	NYSE	MRK	Pharmaceutical industry	29/06/1979	41,9
American Express	NYSE	AXP	Financial services	30/08/1982	38,7
McDonald's	NYSE	MCD	Food industry	30/10/1985	35,5
Boeing	NYSE	BA	Aerospace and defense	12/03/1987	34,2
The Coca-Cola Company	NYSE	KO	Food industry	12/03/1987	34,2
Caterpillar Inc.	NYSE	CAT	Construction and Mining	06/05/1991	30,0
JPMorgan Chase	NYSE	JPM	Financial services	06/05/1991	30,0
The Walt Disney Company	NYSE	DIS	Broadcasting and entertainment	06/05/1991	30,0
Johnson & Johnson	NYSE	JNJ	Pharmaceutical industry	17/03/1997	24,2
Walmart	NYSE	WMT	Retailing	17/03/1997	24,2
The Home Depot	NYSE	HD	Retailing	01/11/1999	21,5
Intel	NASDAQ	INTC	Information technology	01/11/1999	21,5
Microsoft	NASDAQ	MSFT	Information technology	01/11/1999	21,5
Verizon	NYSE	VZ	Telecommunication	08/04/2004	17,1
Chevron Corporation	NYSE	CVX	Petroleum industry	19/02/2008	13,2
Cisco Systems	NASDAQ	CSCO	Information technology	08/06/2009	11,9
The Travelers Companies	NYSE	TRV	Financial services	08/06/2009	11,9
UnitedHealth Group	NYSE	UNH	Managed health care	24/09/2012	8,6
Goldman Sachs	NYSE	GS	Financial services	20/09/2013	7,6
Nike	NYSE	NKE	Apparel	20/09/2013	7,6
Visa Inc.	NYSE	V	Financial services	20/09/2013	7,6
Apple Inc.	NASDAQ	AAPL	Information technology	19/03/2015	6,1
Walgreens Boots Alliance	NASDAQ	WBA	Retailing	26/06/2018	2,9
Dow Inc.	NYSE	DOW	Chemical industry	02/04/2019	2,1
Amgen	NASDAQ	AMGN	Pharmaceutical industry	31/08/2020	0,7
Honeywell	NYSE	HON	Conglomerate	31/08/2020	0,7
Salesforce	NYSE	CRM	Information technology	31/08/2020	0,7

Moyenne 22,1 ans
Médiane 21,5 ans

Sociétés de grande et moyenne capitalisation

Accès **quasi illimité** au capital
(trésorerie propre, actionnaires,
dette, ...)

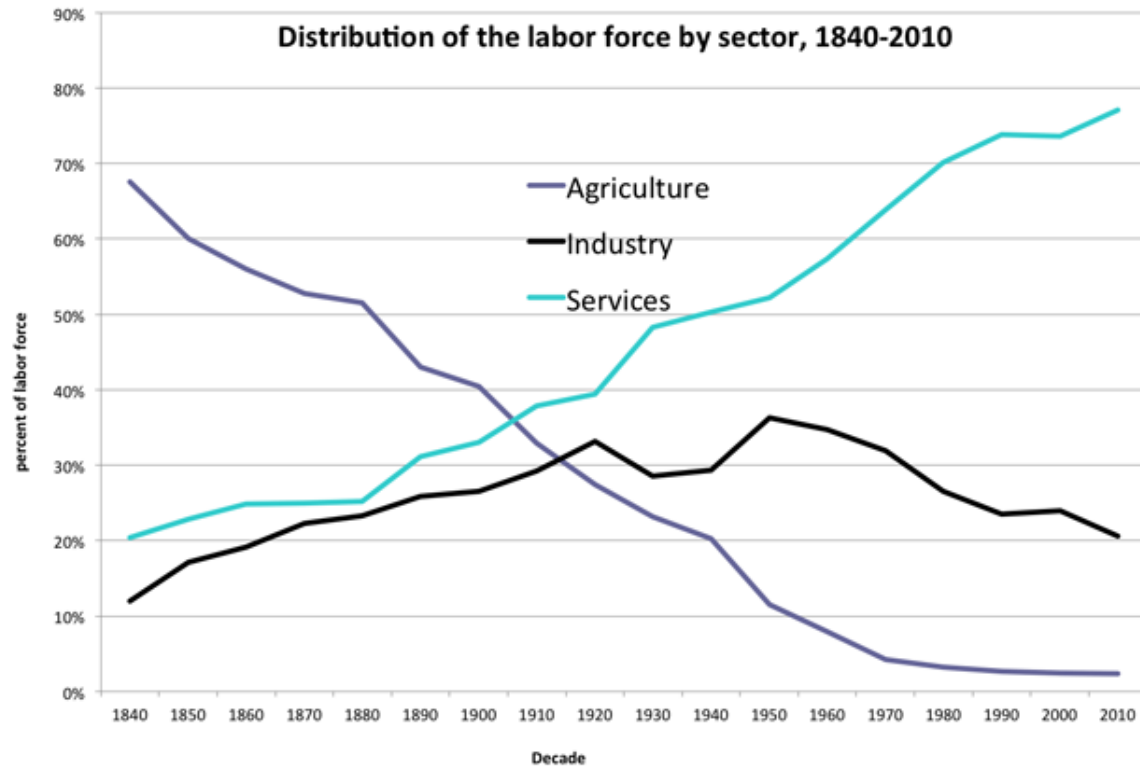
Marques fortes

Accès **facile** aux talents



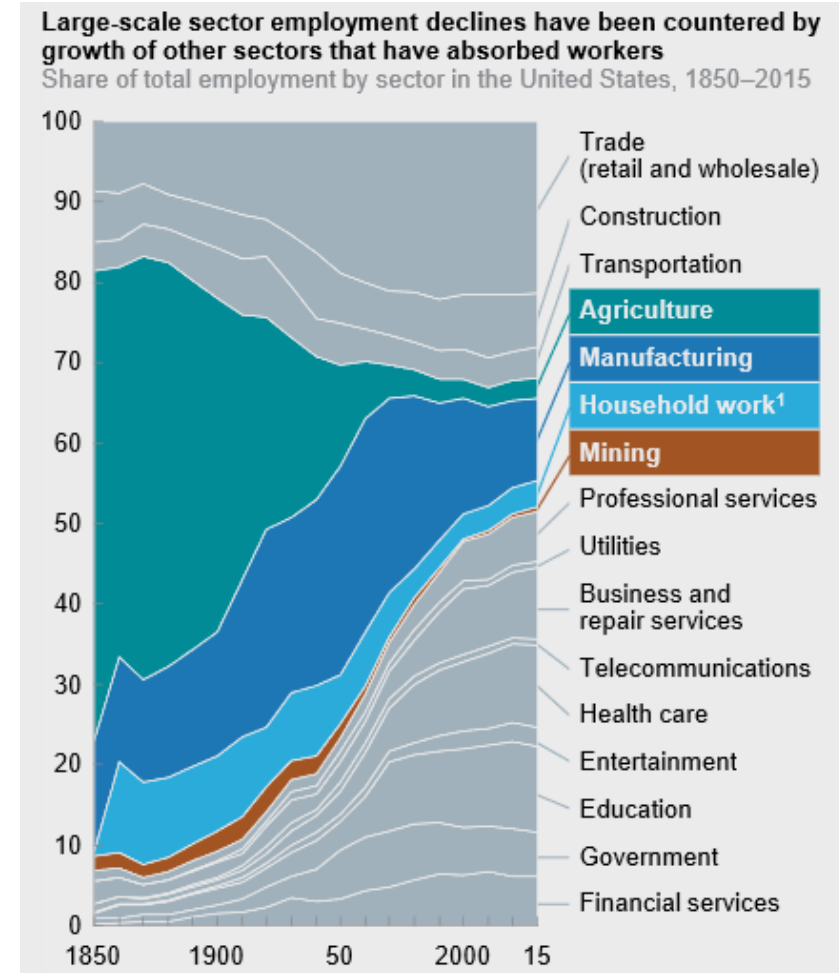
Quid des entreprises au Luxembourg ?

Répartition de la main-d'œuvre américaine 1840-2015



Source: 1840–1900: Robert E. Gallman and Thomas J. Weiss. "The Service Industries in the Nineteenth Century." *In Production and Productivity in the Service Industries*, ed. Victor R. Fuchs, 287–352. New York: Columbia University Press (for NBER), 1969

Source : McKinsey Global Institute, *Jobs Lost, Jobs Gained: Workforce transitions in a time of automation*, December 2017



A wireframe profile of a human head, composed of numerous interconnected lines forming a mesh, rendered in a light blue color against a dark background.

Pourquoi
l'IA maintenant

Lien entre
IA & 4ème révolution industrielle

Pourquoi l'IA
maintenant ?

Convergence de plusieurs tendances

- **Migration** accrue des activités socio-économiques vers Internet
- **Miniaturisation et baisse exponentielle** du coût de technologies
- **Percées** dans Machine Learning
- **Limitations** du cerveau humain

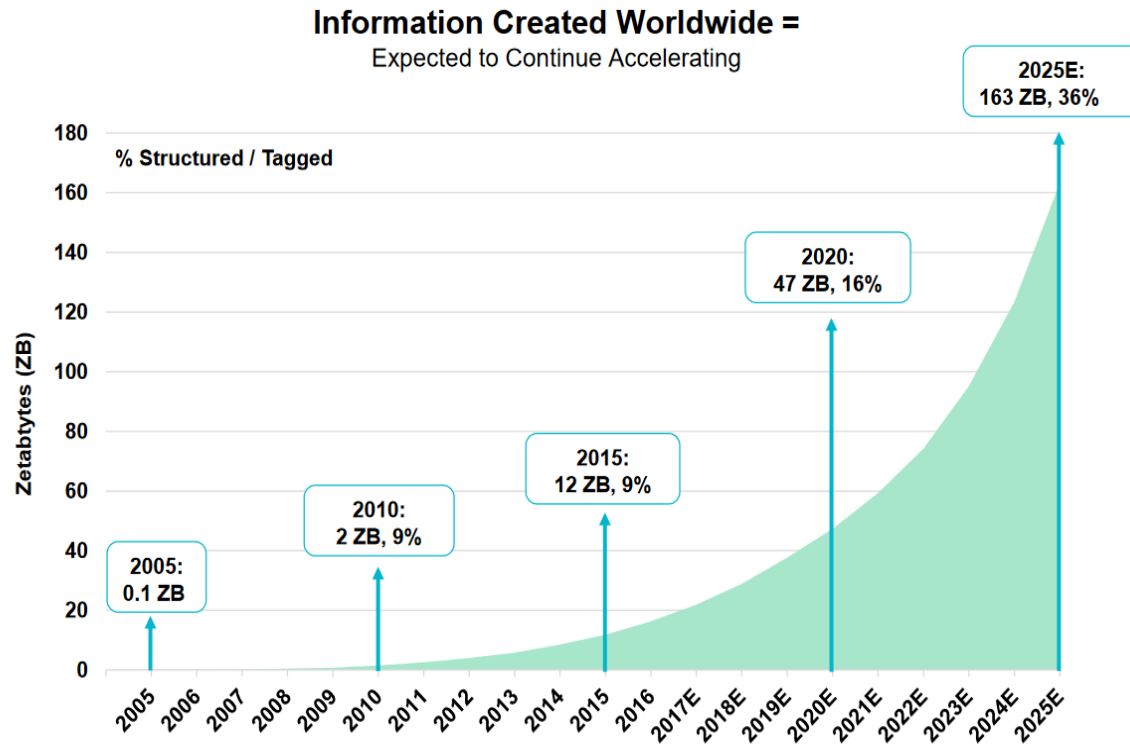
Équivalence des coûts pré- 4^{ème} révolution industrielle

Technology	Cost to equivalent functionality	Cost scale
3D printing	\$40.000 (2007) to \$100 (2014)	400x in 7 years
Industrial robots	\$500.000 (2008) to \$22.000 (2013)	23x in 5 years
Drones	\$100.000 (2007) to \$700 (2013)	142x in 6 years
Sensors (3D LIDAR sensor)	\$20.000 (2009) to \$79 (2014)	250x in 5 years
DNA sequencing of 1 human	\$10 million (2007) to \$1.000 (2014)	10.000x in 7 years
Storage cost of 1 Gbyte	\$10 (2000) To \$0.05 (2015)	200x in 15 years

Source :

- Exponential Organizations, Salim Ismail
- Komorowski law, Cost of storage per gigabyte

Univers numérique exponentiel



KLEINER
PERKINS

Source: IDC DataAge 2025 Study, sponsored by Seagate (3/17)
Note: 1 petabyte = 1MM gigabytes, 1 zeta byte = 1MM petabytes

KP INTERNET TRENDS 2017 | PAGE 132

$$IA = f(BigData(storage))$$



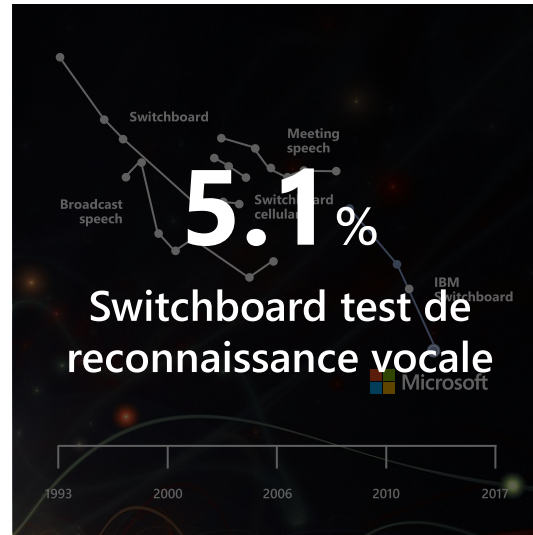
- **<0.5% données utilisées** pour améliorer performance de l'entreprise
- **14%** croient exploiter aujourd'hui
- Potentiel **énorme** pour les entreprises ?

Réalité - Percées IA sur fonctions cognitives



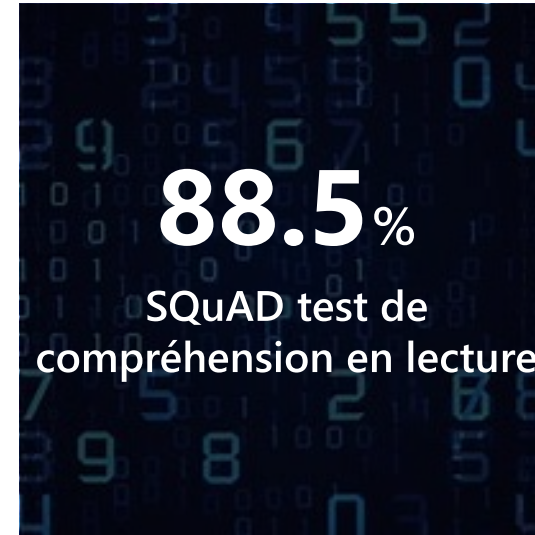
2016

Reconnaissance d'objets
Parité humaine



2017

Reconnaissance vocale
Parité humaine



January 2018

Compréhension lecture machine
Parité humaine



March 2018

Traduction automatique
Parité humaine

Univers économique – découvertes scientifiques & cerveau humain

4ème paradigme scientifique

basé sur l'idée que calcul scientifique constitue un nouvel ensemble de méthodes au-delà de l'empirisme, de la théorie et de la simulation

63% des chefs d'entreprise croient en la promesse de l'IA et du ML

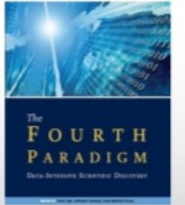
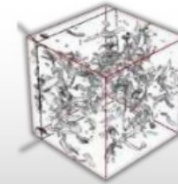
Sources :

- Microsoft The Future of Work, query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE2MfkH
- Microsoft Canada, Consumer Insights study 2015
- Harvard Business Review Analytic Services study (2018), The workplace evolution, <https://hbr.org/resources/pdfs/comm/microsoft/workplaceevolution.pdf>

The 4th Paradigm for Scientific Discovery



$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - K\frac{c^2}{a^2}$$

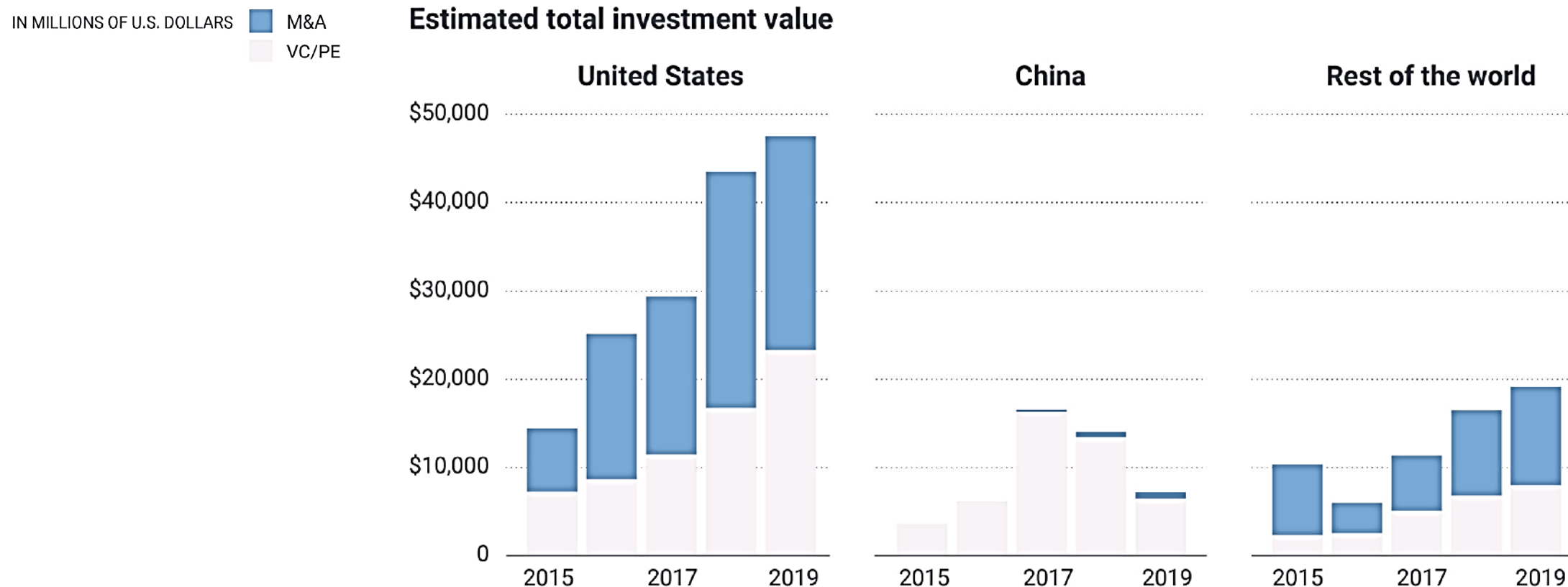


Experimental	Theoretical	Computational	The Fourth Paradigm
Thousand years ago <i>Description of natural phenomena</i>	Last few hundred years <i>Newton's laws, Maxwell's equations...</i>	Last few decades <i>Simulation of complex phenomena</i>	Today and the Future <i>Unify theory, experiment and simulation with large multidisciplinary Data</i> <i>Using data exploration and data mining (from instruments, sensors, humans...)</i> <i>Distributed Communities</i>

Réalité - Investissements IA par pays

FIGURE 11

Disclosed and estimated total investment value by target region and nature of investment



SOURCE: CSET ANALYSIS OF CRUNCHBASE AND REFINITIV DATA.

Intelligence artificielle

"La théorie et le développement de systèmes informatiques capables d'effectuer des tâches nécessitant normalement l'intelligence humaine, telles que la perception visuelle, la reconnaissance vocale, la prise de décision et la traduction entre les langues"

Source : English Oxford Living Dictionary

Première apparition
Intelligence Artificielle

1948

1956

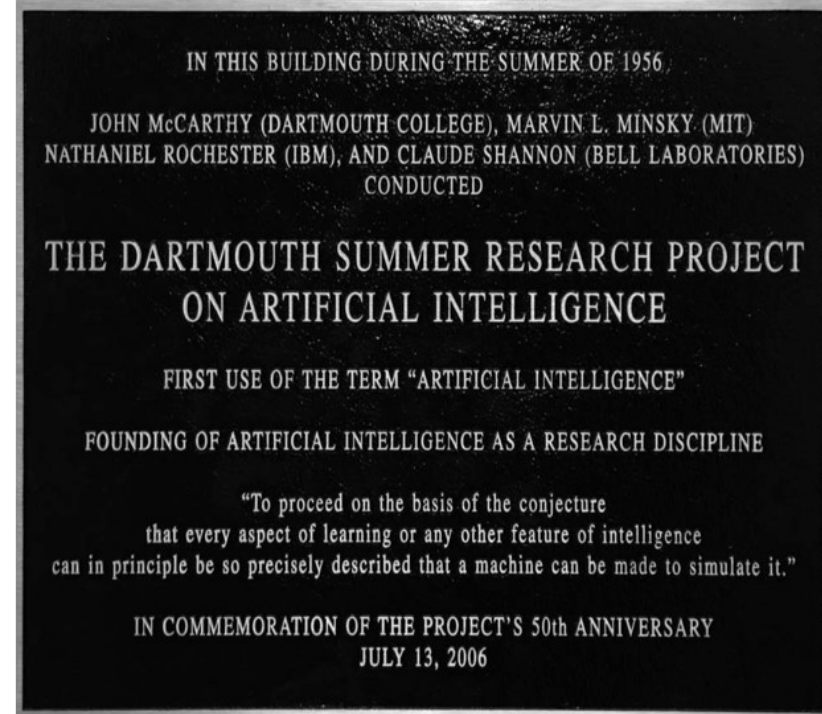
1962

1950



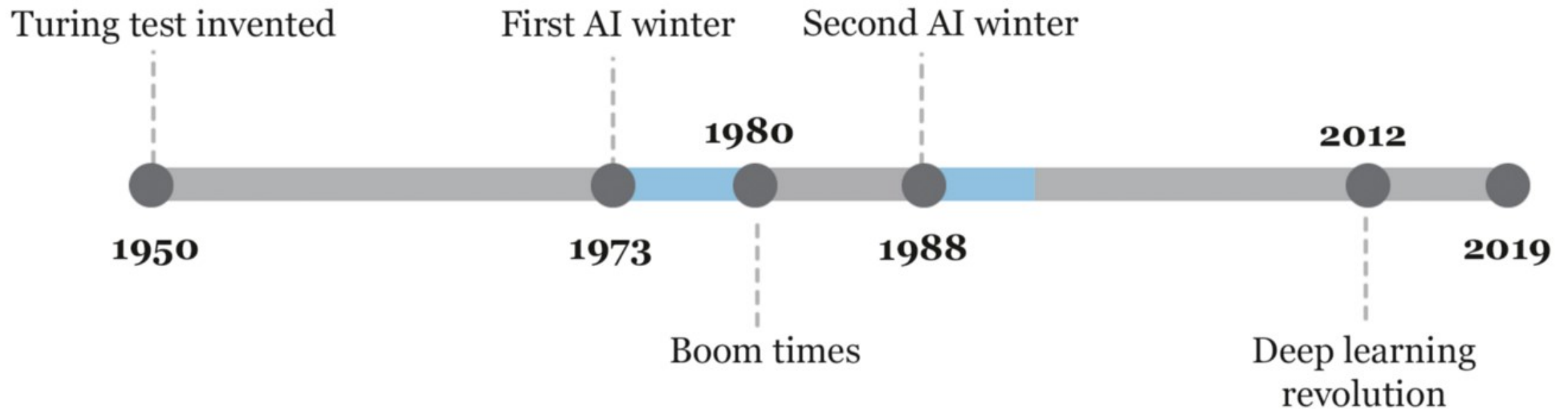
Turing Test

"Machines informatiques et intelligence"



John McCarthy (Dartmouth College)
Marvin Minsky (MIT)
Nathan Rochester (IBM)
Claude Shannon (Bell Labs)

1956



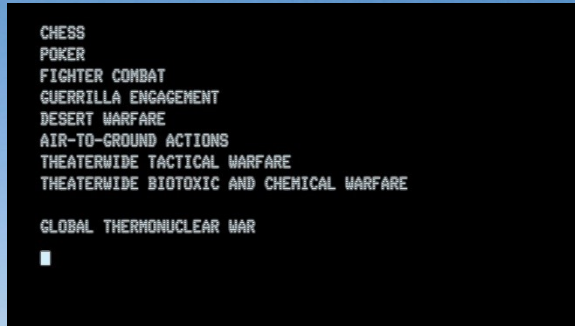
Source : towardsdatascience.com/history-of-the-first-ai-winter-6f8c2186f80b

CHESS
POKER
FIGHTER COMBAT
GUERRILLA ENGAGEMENT
DESERT WARFARE
AIR-TO-GROUND ACTIONS
THEATERWIDE TACTICAL WARFARE
THEATERWIDE BIOTOXIC AND CHEMICAL WARFARE

GLOBAL THERMONUCLEAR WAR



1983



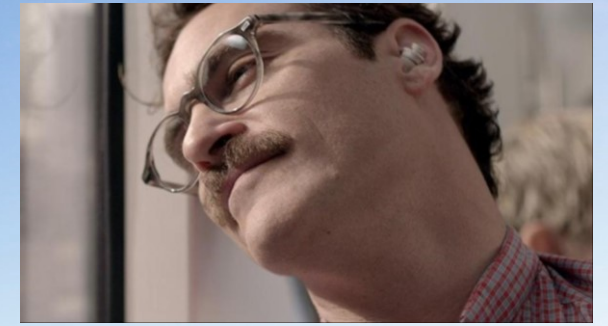
Wargames (1983)



Terminator (1984)



Matrix (2003)



Her (2013)



De la science-fiction à la réalité

i can i i everything else

De la science-fiction à la réalité

Bob: i can i i everything else

Alice: balls have zero to me to me to me to me to me to me to me to me to

Bob: you i everything else

Alice: balls have a ball to me to me to me to me to me to me to me

Bob: i i can i i i everything else

Alice: balls have a ball to me to me to me to me to me to me to me

Bob: i

Alice: balls have zero to me to me to me to me to me to me to me to me to

Bob: you i i i i i everything else

Alice: balls have o to me to me to me to me to me to me to me to me to

Bob: you i i i everything else

Alice: balls have zero to me to me to me to me to me to me to me to me to



Cryptophasie

phénomène linguistique développé par des jumeaux **que seuls les jumeaux comprennent**
du Grec *crypto* (**secret**) & *phasia* (**parole**)

Cas de cryptophasie



Facebook chief executive Mark Zuckerberg has supported AI development CREDIT: REUTERS

Follow

By Matthew Field

1 AUGUST 2017 • 10:21AM

Facebook shut down a pair of its artificial intelligence robots after they invented their own language.

Researchers at Facebook Artificial Intelligence Research built a chatbot earlier this year that was meant to learn how to negotiate by mimicking human trading and bartering.

But when the social network paired two of the programs, nicknamed Alice and Bob, to trade against each other, they started to learn their own bizarre form of communication.

The chatbot conversation "led to divergence from human language as the agents developed their own language for negotiating," the researchers said.

Facebook Shuts Down AI System After Bots Create Language Humans Can't Understand

Indo-Asian News Service, 31 July 2017



Share on Facebook



Tweet



Share



Share



Email



Reddit



22 comments



HIGHLIGHTS

- Chatbots started speaking in their own language defying codes provided
- Initially the AI agents used English to converse with each other
- This comes after Elon Musk said that AI was the biggest risk

Days after Tesla CEO Elon Musk said Facebook co-founder **Mark Zuckerberg's understanding of artificial intelligence (AI) was limited**, the social media company has reportedly shut down one of its AI systems because "things got out of hand." The AI bots created their own language, from the scratch and without human input, forcing Facebook to shut down the AI system. The AI bots' step of creating and communicating with the new language defied the provided codes.

According to a **report** in Tech Times on Sunday, "The **AI** did not start shutting down computers worldwide or something of the sort, but it stopped using English and started using a language that it created." Initially the AI agents used English to converse with each other but they later created a new language that only AI systems could understand, thus, defying their purpose.

This led **Facebook** researchers to shut down the AI systems and then force them to speak to each other

Microsoft silences its new A.I. bot Tay, after Twitter users teach it racism [Updated]

Sarah Perez @sarahintampa / 3 years ago

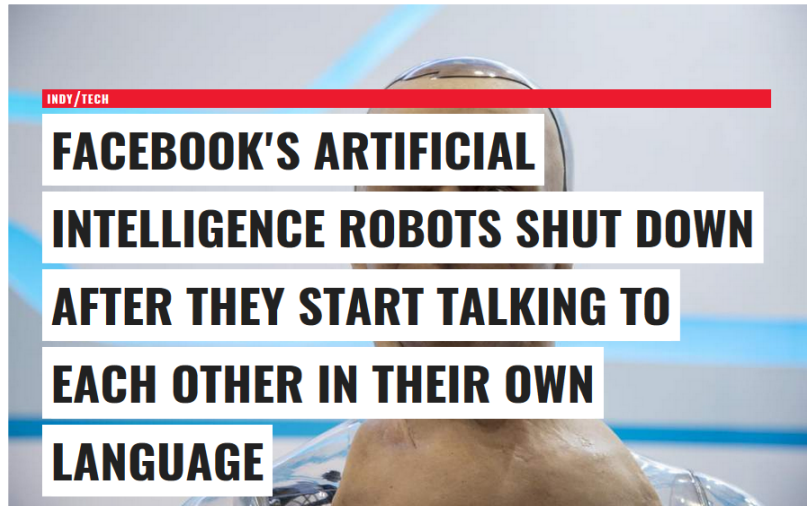
 Comment



Microsoft's newly launched A.I.-powered bot called Tay, which was responding to tweets and chats on GroupMe and Kik, has already been shut down due to concerns with its inability to recognize when it was making offensive or racist statements. Of course, the bot wasn't *coded* to be racist, but it "learns" from those it interacts with. And naturally, given that this is the Internet, one of the first things online users taught Tay was how to be racist, and how to spout back ill-informed or inflammatory political opinions. [Update: Microsoft now says it's "making adjustments" to Tay in light of this problem.]

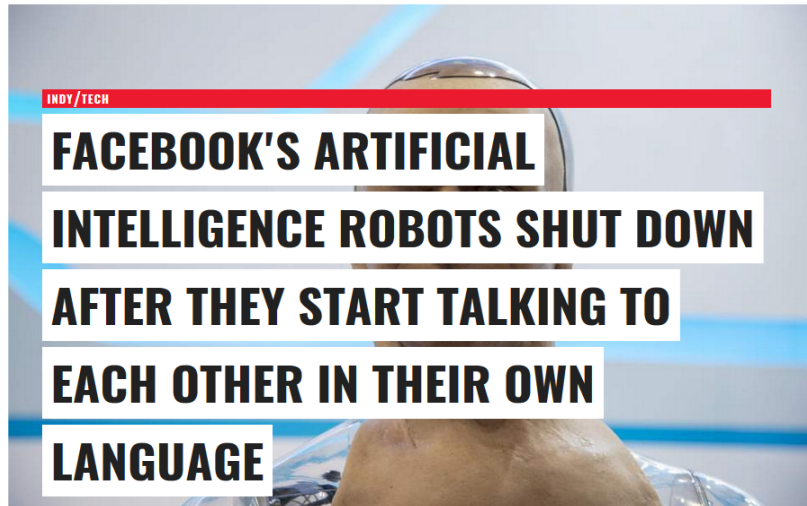
In case you missed it, [Tay is an A.I. project](#) built by the Microsoft Technology and Research and Bing teams, in an effort to conduct research on conversational understanding. That is, it's a bot that you can talk to online. The

Réalité – pas fiction

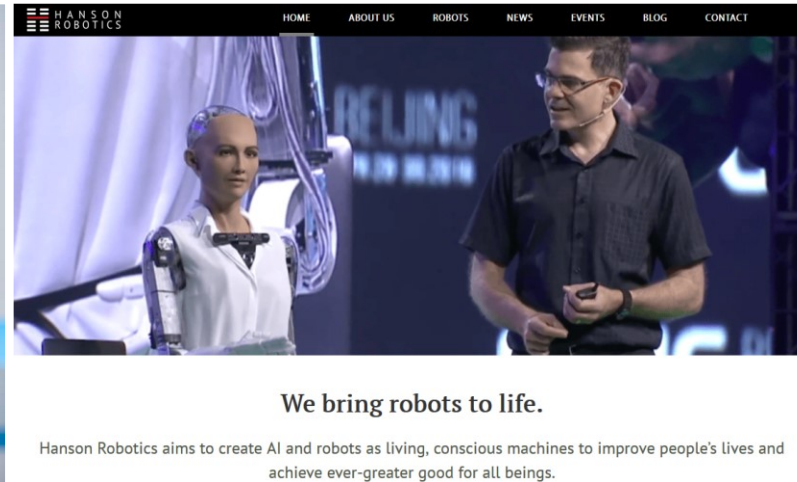


AI robots
Facebook

Réalité – pas fiction

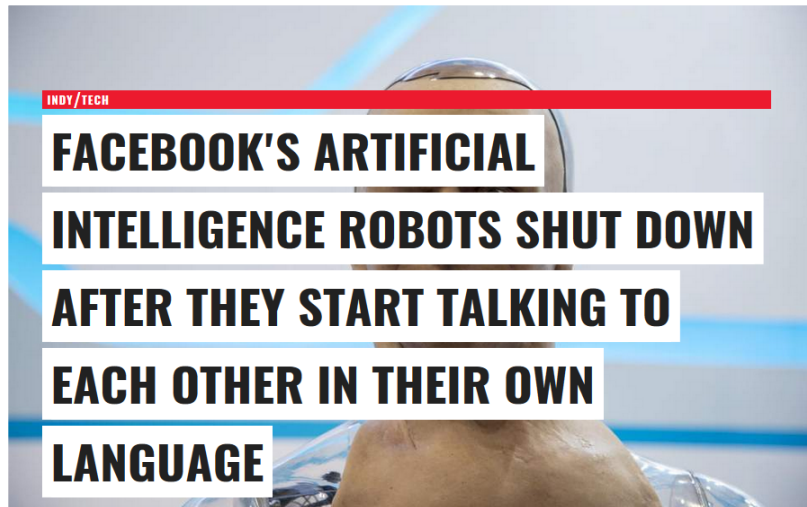


AI robots
Facebook

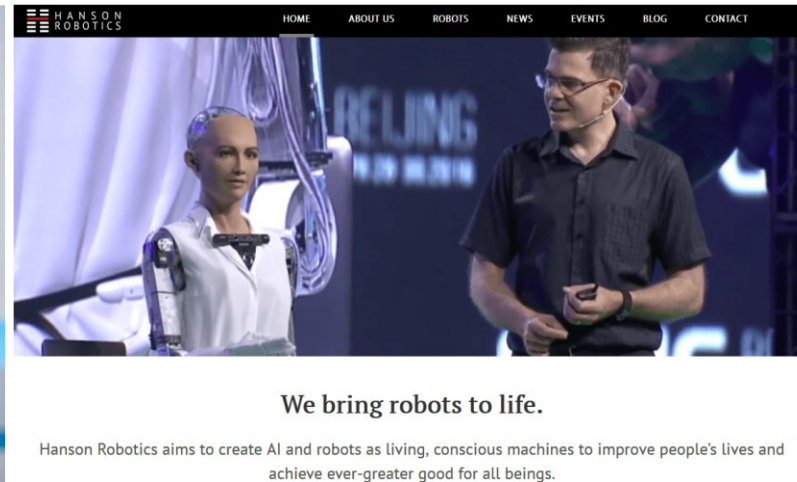


Sophia
Hanson Robotics

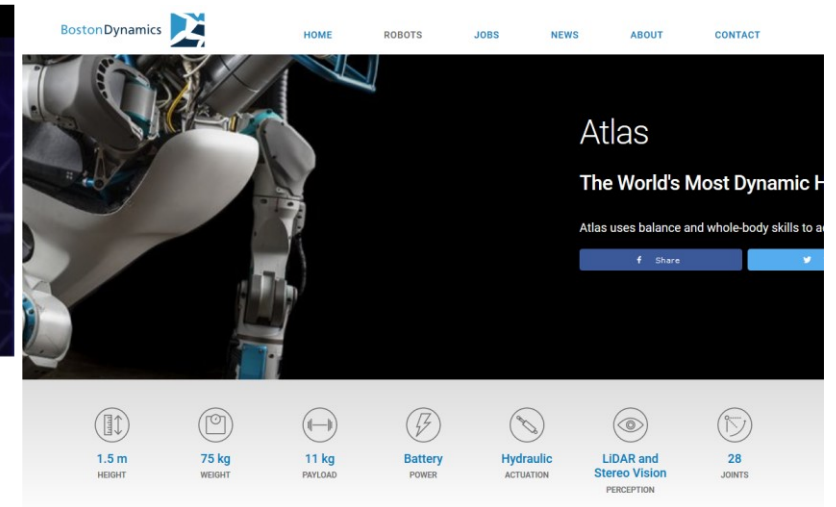
Réalité – pas fiction



AI robots
Facebook



Sophia
Hanson Robotics



Atlas
Boston Dynamics



Exemples
Cols blancs & Cols bleus

Étude de cas - cols bleus

Véhicules autonomes (AV)



AUTONOMOUS

SAFE
1G-2

SENSOR
ACTIVE

INT'3

7A

AV – Véhicules autonomes



SAE J3016™ LEVELS OF DRIVING AUTOMATION™

Learn more here: [sae.org/standards/content/j3016_202104](https://www.sae.org/standards/content/j3016_202104)

Copyright © 2021 SAE International. The summary table may be freely copied and distributed AS-IS provided that SAE International is acknowledged as the source of the content.

	SAE LEVEL 0™	SAE LEVEL 1™	SAE LEVEL 2™	SAE LEVEL 3™	SAE LEVEL 4™	SAE LEVEL 5™
What does the human in the driver's seat have to do?	You <u>are</u> driving whenever these driver support features are engaged – even if your feet are off the pedals and you are not steering			You <u>are not</u> driving when these automated driving features are engaged – even if you are seated in “the driver’s seat”		
	You must constantly supervise these support features; you must steer, brake or accelerate as needed to maintain safety			When the feature requests, you must drive	These automated driving features will not require you to take over driving	

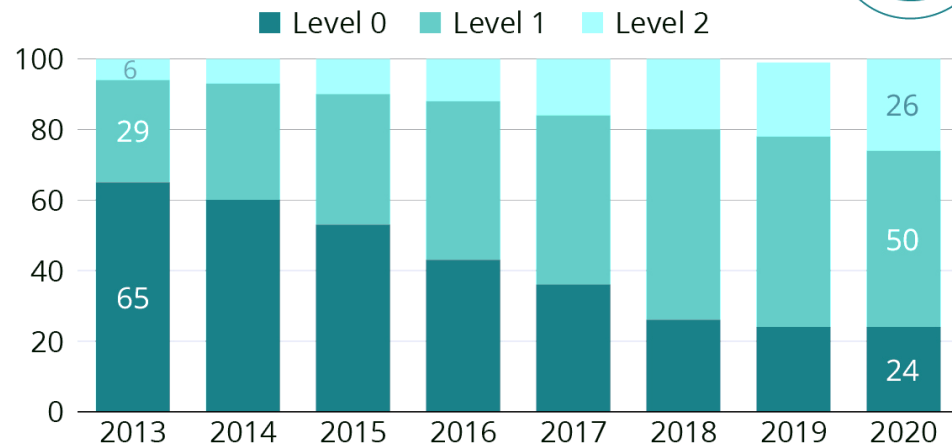
Copyright © 2021 SAE International.

	These are driver support features			These are automated driving features	
What do these features do?	These features are limited to providing warnings and momentary assistance	These features provide steering OR brake/acceleration support to the driver	These features provide steering AND brake/acceleration support to the driver	These features can drive the vehicle under limited conditions and will not operate unless all required conditions are met	This feature can drive the vehicle under all conditions
Example Features	• automatic emergency braking • blind spot warning • lane departure warning	• lane centering OR • adaptive cruise control	• lane centering AND • adaptive cruise control at the same time	• traffic jam chauffeur	• local driverless taxi • pedals/steering wheel may or may not be installed • same as level 4, but feature can drive everywhere in all conditions

AV – Véhicules autonomes

Cars Increasingly Ready for Autonomous Driving

Estimated share of newly registered cars worldwide by autonomous driving level



Level 0=no autonomous driving

Level 1=partly assisted driving (steering OR braking)

Level 2=assisted driving (steering AND breaking below 60 km/h)

Source: Statista Mobility Market Outlook

statista 

Autorisation **Level 3** - 27 novembre 2021



Chine

environ 7 heures

Les taxis sans chauffeur débarquent à Pékin

La capitale chinoise vient d'autoriser l'usage commercial des premiers taxis autonomes, nommés «Apollo Go».

Elles ressemblent à des voitures normales, s'arrêtent au bord du trottoir pour prendre des passagers mais n'ont aucun conducteur: Pékin vient d'autoriser l'usage commercial des premiers taxis autonomes. Équipés de capteurs sur le toit, ces SUV ne peuvent transporter que deux passagers à la fois et un humain est toujours assis à l'avant... prêt à s'emparer du



BUSINESS + TECH - APOLLO GO

First 'robotaxis' enter service in Beijing

Author: AFP | Update: 27.11.2021 00:01



An 'Apollo Go' autonomous taxi on a street in Beijing / © AFP

It looks like a normal car but the white taxi by the kerb has nobody driving it, and communicates with customers digitally to obtain directions and take payment.

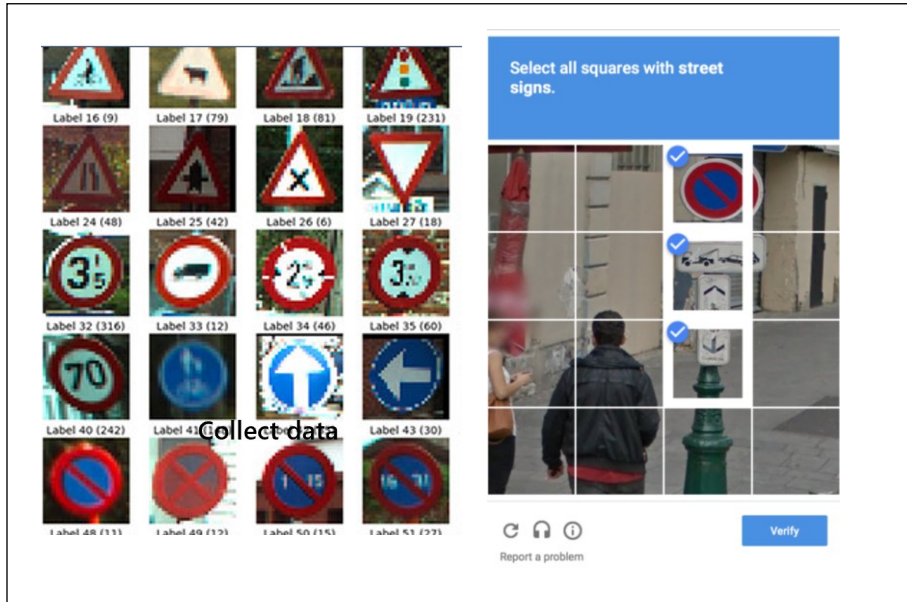
Beijing this week approved its first autonomous taxis for commercial use, bringing dozens of the so-called "robotaxis" to the streets of the Chinese capital.

The vehicles can only carry two passengers at a time and are confined to the city's southern

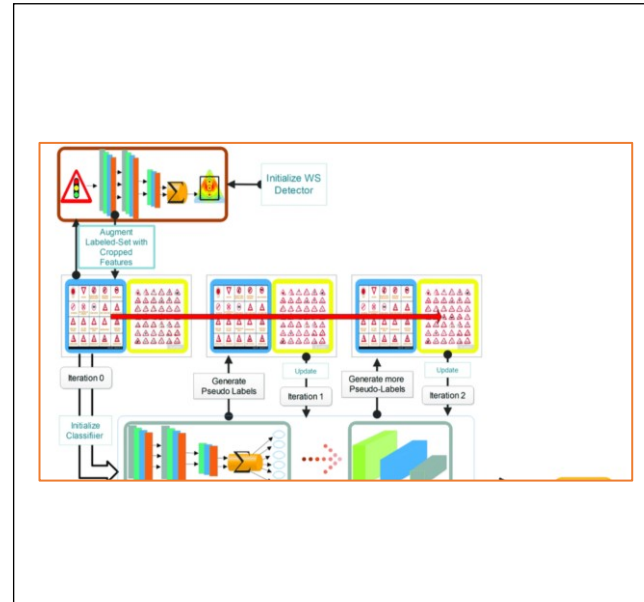
AV – Véhicules autonomes

Utilisation du 4^{ème} paradigme scientifique

Collecter & labéliser données



Entraîner modèle



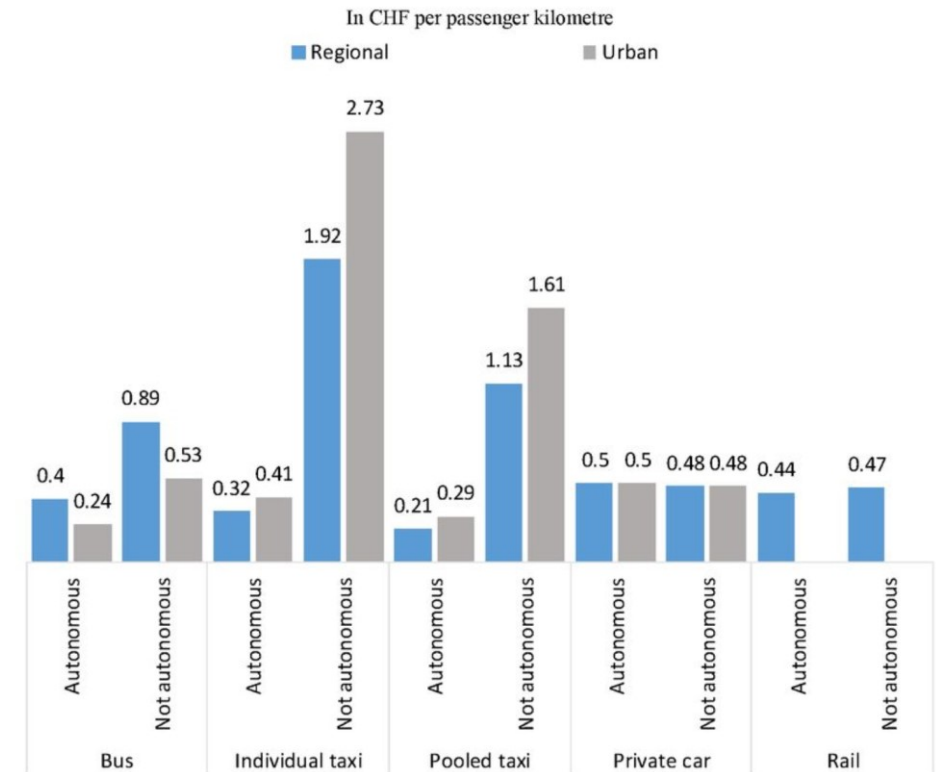
Exploiter modèle



AV – Les enjeux

- Contexte
 - Transport represent 5.6% du PIB global (2018)
 - Conducteur US représentent **2.8% de workforce**
 - Tesla Autopilot **3.3milliards km** de données
 - **Pas de partage** de données
- Implications & enjeux
 - Coût par km - Taxi régional autonome vs non-autonome - Réduction du coût facteur **6x (1,8 EUR à 0,3 EUR)**
 - Safety - **10% d'adoption** AV sauverait **1.100 vies & 38 milliards de USD**

Figure 3.1. Cost comparison of different modes with and without AV technology



Source: Adapted from Bösch et al. (2018_[3]), "Cost-based analysis of autonomous mobility services", <https://doi.org/10.1016/j.tranpol.2017.09.005>.



CONFIDENTIALITY AGREEMENT

This Confidentiality Agreement (the "**Agreement**") is made and entered into on this ____ day of _____, 2011 (the "**Effective Date**"), by and between _____, an individual ("**Disclosing Party**"), and _____ ("**Recipient**").

Disclosing Party is willing to disclose to Recipient certain confidential and proprietary information in order to permit Disclosing Party and Recipient to evaluate and/or enter into a potential business transaction or relationship (the "**Relationship**"), subject to the terms and conditions described in this Agreement.

NOW THEREFORE, in consideration of the foregoing premises and mutual promises hereinafter contained, Recipient agrees as follows:

1. Definitions. As used in this Agreement, the words below are defined as follows:

"Confidential Information" means any Disclosing Party proprietary information, technical data, trade secrets or know-how, including, but not limited to, research, product plans, products, services, customers, markets, software, databases, applications, formats, developments, inventions, processes, formulas, technology, designs, drawings, algorithms, engineering data, marketing, financial or other business information. Confidential Information shall include all information that should reasonably be understood by Recipient, because of legends or other markings, the circumstances of disclosure, or the nature of the information itself to be proprietary and confidential to Disclosing Party, regardless of whether such information is marked "Confidential" or "Proprietary."

Confidential Information shall not include information which is: (i) obtained by Recipient from a third party which third party, to the actual knowledge of Recipient, is lawfully in possession of such information without violation of any contractual or legal obligation; (ii) part of the public domain through no fault of Recipient or its Representatives; (iii) independently developed by Recipient or its Representatives without reference to any Confidential Information; or (iv) approved for disclosure by written authorization of the Disclosing Party.

"Representatives" refers to the directors, officers, employees, agents or other representatives, including, without limitation, attorneys, accountants, consultants, investment bankers, financial advisors and lenders of Recipient and its affiliates.

"Review Material" refers to reports, notes, analyses, forecasts or other documents which contain or reflect Confidential Information, including, without limitation, materials created by Recipient.

2. Confidentiality. Except as required by law, neither Recipient nor its Representatives will, without the prior written consent of Disclosing Party, disclose to any person the fact that Confidential Information has been made available to Recipient or any of the terms, conditions or other facts with respect to the Relationship.

3. Duty of Care. Recipient agrees to use at least the same degree of care in protecting and using the Confidential Information received pursuant to this Agreement as Recipient would use in protecting its own Confidential Information, but in no case less than a reasonable degree of care.

4. Nondisclosure Period. Review Material and Confidential Information will be kept confidential for a period of three (3) years from the date of disclosure. Neither Recipient nor any of its Representatives will, without the prior written consent of Disclosing Party, disclose the Review Material or the Confidential Information, in whole or in part, in any manner.

5. Limitations on Use. Recipient agrees that Recipient will: (a) not use the Review Material or the Confidential Information other than for the purpose of evaluating or pursuing the Relationship; (b) not use the Review Material or the Confidential Information to compete, directly or indirectly, with Disclosing Party; and (c) disclose the Review Material and the Confidential Information only to Representatives who

Confidentiality Agreement (NDA)

to effective (change this: insert date)

ing Party"), hereinafter referred to as the
all mean and include its authorized
pointing on its behalf or who benefit

ferred to as the "Recipient Party"
e its authorized representative(s),
half or who benefit from this Agreement).

o each as a Party and collectively as the

These discussions may require the
Recipient Party. The Parties wish to

A. create your own NDA agreement.

nation is received by the Recipient Party,
g Party's Confidential Information.

Confidential Information

Confidential Information means any information of or about the Disclosing Party that is:

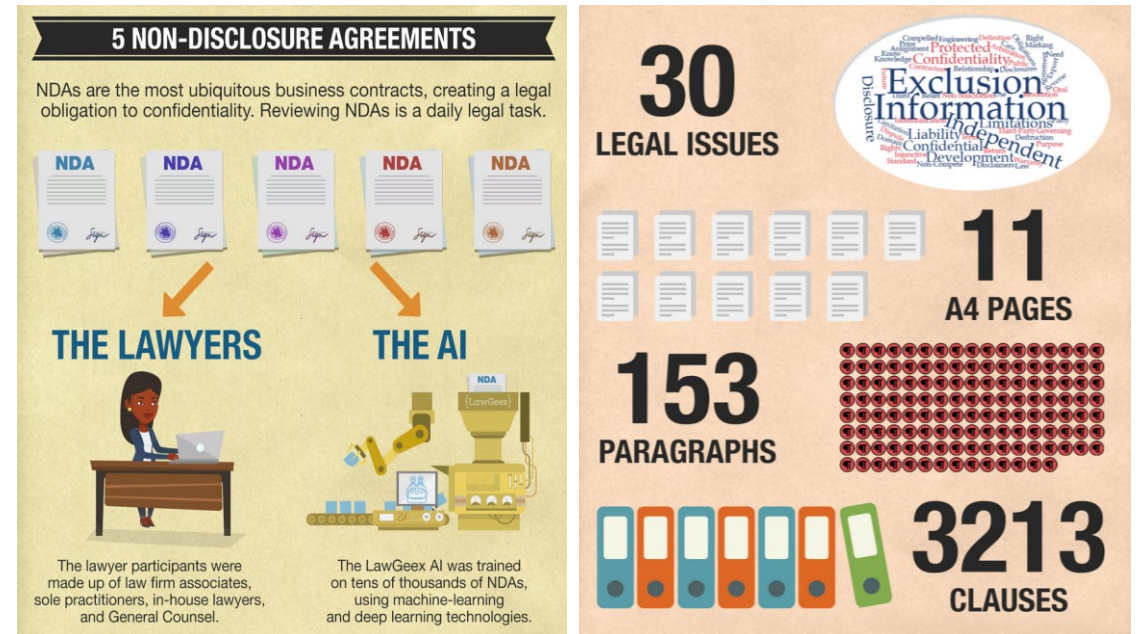
- is identified as "confidential" or "proprietary" or "private" at the time of disclosure, when delivered orally or by any other means

Étude de cas - cols blancs

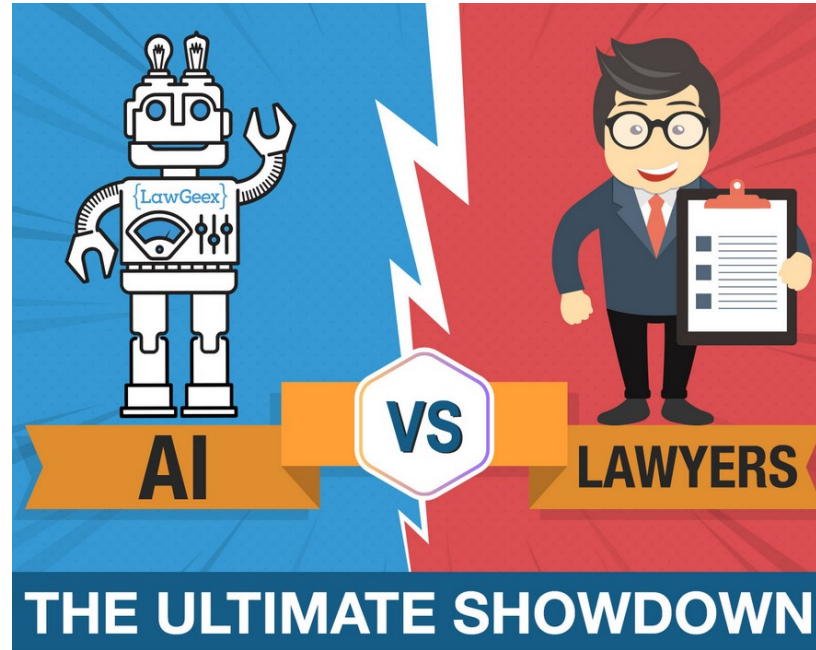
NDA benchmark IA vs avocats

Expérience NDA

20 avocats formés aux États-Unis
IA entraînée avec des milliers de NDA
IA pas préentraînée à l'expérience
4 heures
5 NDA
30 problèmes à repérer



NDA juridiques – résultats

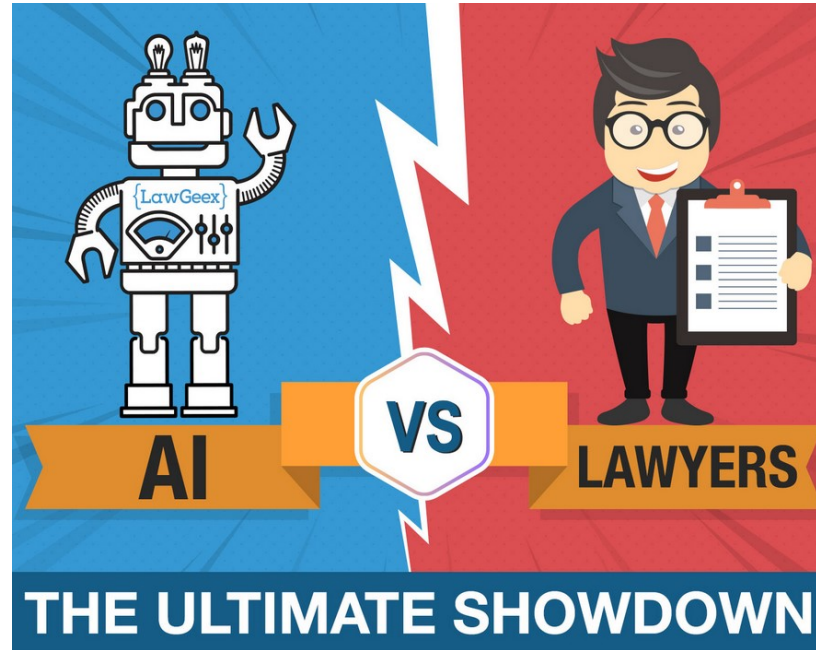


Cafés

0

12

NDA juridiques – résultats



Cafés

0

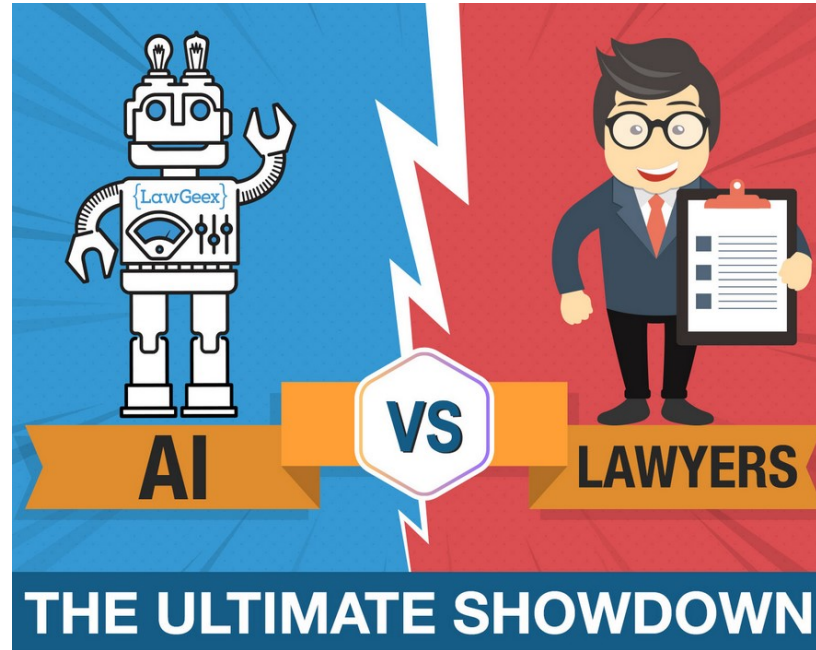
12

Exactitude

94%

85%

NDA juridiques – résultats



Cafés

0

12

Exactitude

94%

85%

Heure

26 s

5.520 s
(92 min)

IA au Luxembourg depuis 2016

- **Agents virtuels** KYC & FAQ
- Amélioration taux de conversion vente (assurances)
- Bot d'**auto-cotation** (logistique)
- Evaluation des compétences (HR)
- **Prédiction** de production de processus (industrie)
- Positionnement des annonces publicitaires (B2C)
- **Automatisation d'encodage** (RPA) (Finance)
- ...





“IA a le pouvoir de **transformer**
chaque entreprise, **améliorer**
chaque vie et résoudre certains
des **défis les plus fondamentaux**
de notre société. »

HARRY SHUM,
MICROSOFT EVP AI & RESEARCH

IA responsable



Éthique & moralité

- De l'**IA de niche/étroite** à l'IA générale - risque majeur
- IA peut élargir le cercle des agents moraux au-delà des humains à des systèmes artificiellement intelligents appelés **agents moraux artificiels (AMA)**
- **défi de concevoir** des agents respectant l'**ensemble des valeurs** et des **lois exigées** par les agents moraux humains (HMA)

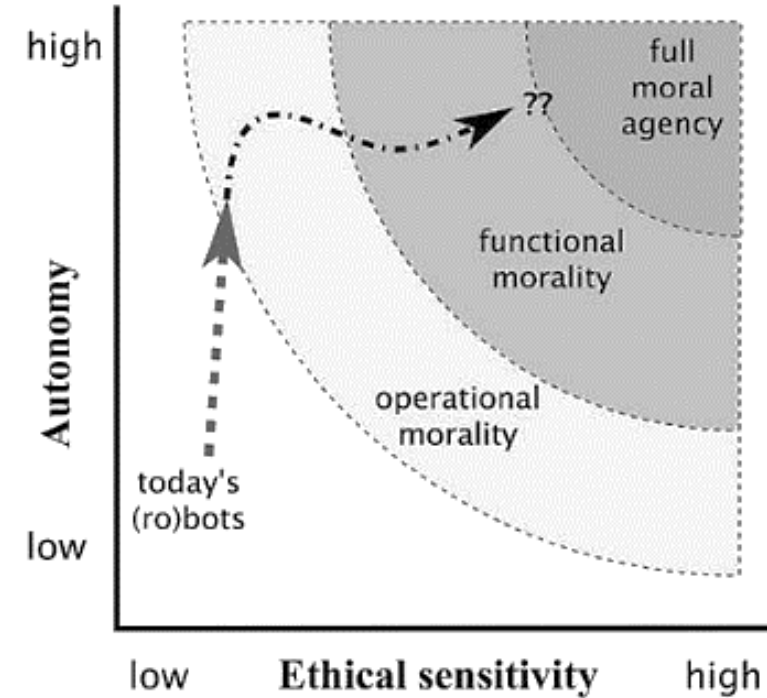
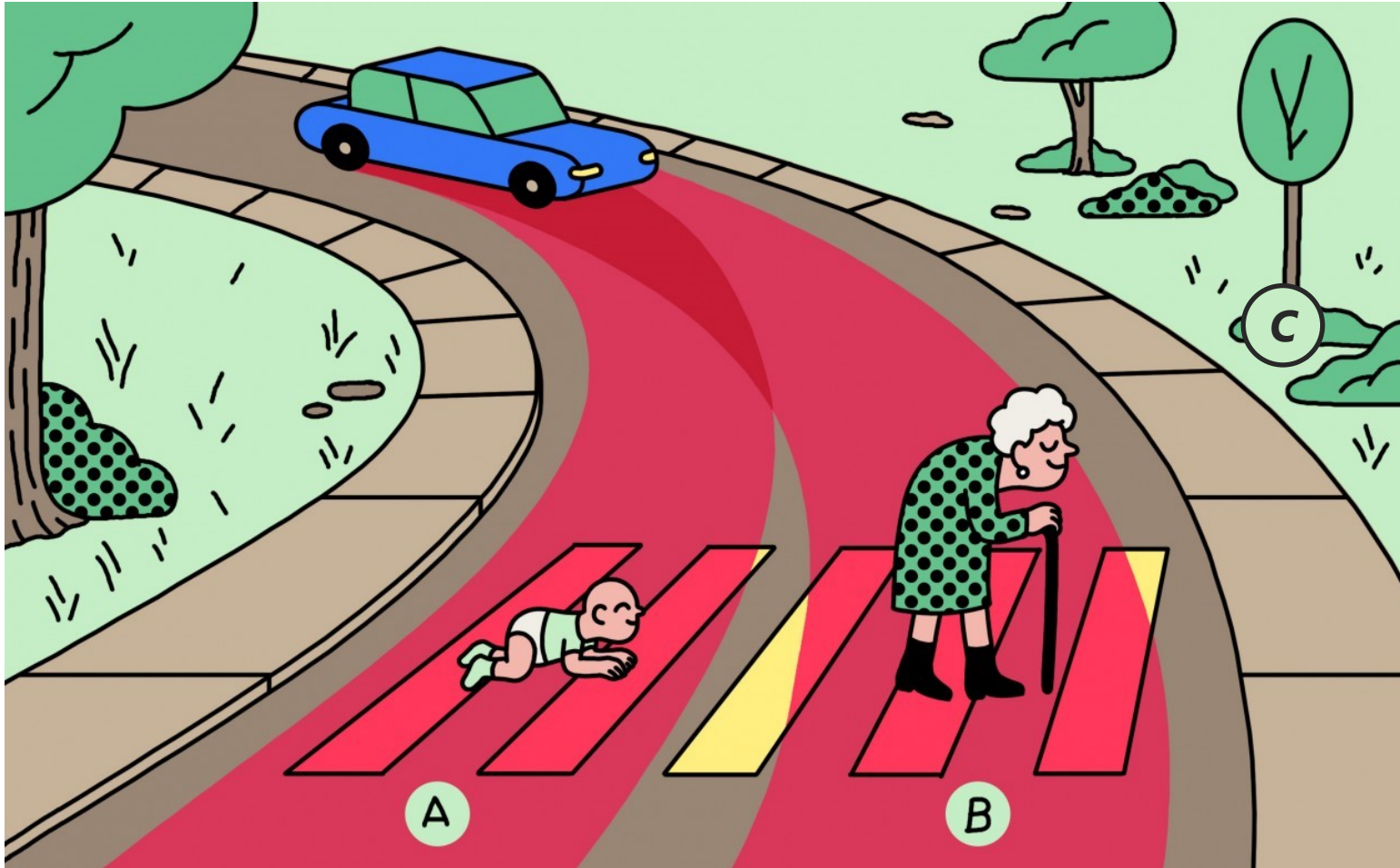


Figure 2.1. Two Dimensions of AMA Development.

Source : Wallach, Wendell. Allen, Colin. *Moral Machines*. Oxford University Press

Tristement célèbre problème de « chariot » - trolley problem



Une voiture autonome doit-elle tuer le bébé, la grand-mère ou le conducteur ?

Injustice & types de préjudice : qualité de service

The New York Times

Facial Recognition Is Accurate, if You're a White Guy

By STEVE LOHR FEB. 9, 2018



Gender was misidentified in up to 1 percent of lighter-skinned males in a set of 385 photos.

Gender was misidentified in up to 12 percent of darker-skinned males in a set of 318 photos.

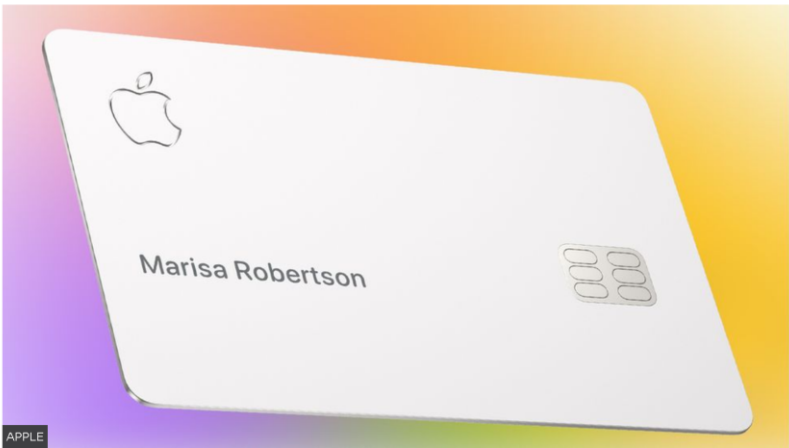


Gender was misidentified in up to 7 percent of lighter-skinned females in a set of 296 photos.

Gender was misidentified in 35 percent of darker-skinned females in a set of 271 photos.

Apple's 'sexist' credit card investigated by US regulator

11 November 2019



A US financial regulator has opened an investigation into claims Apple's credit card offered different credit limits for men and women.

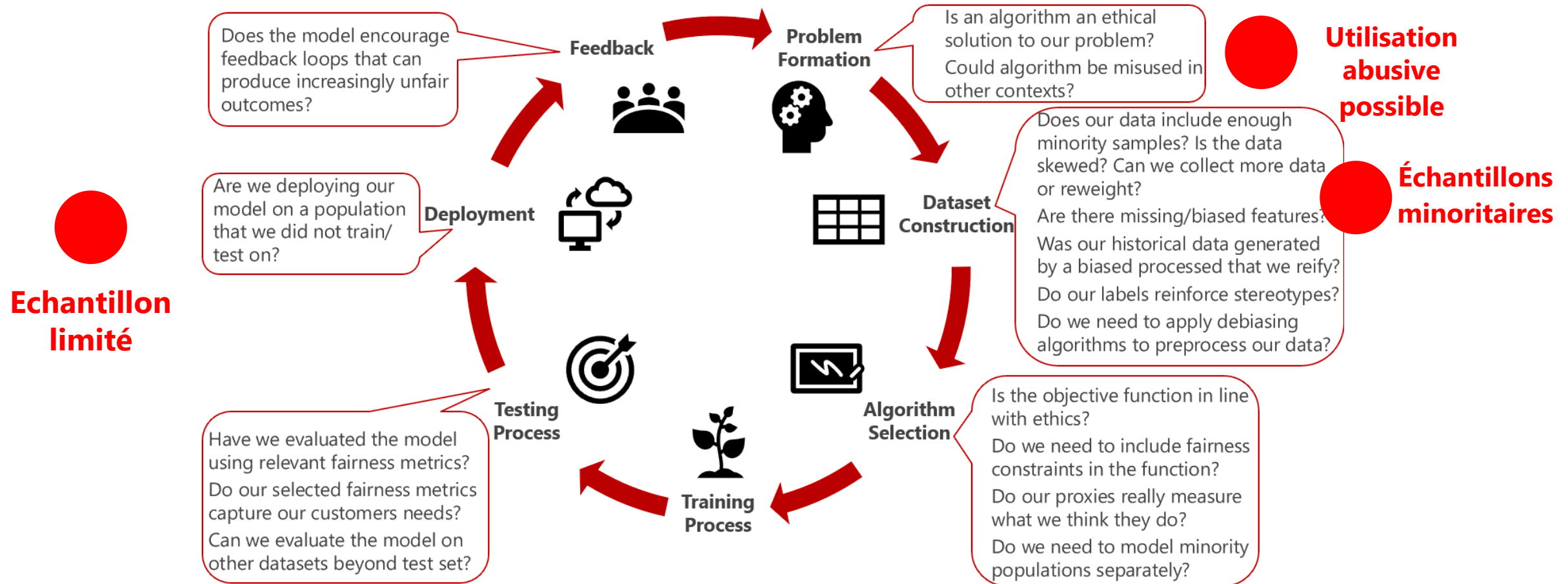
It follows complaints - including from Apple's co-founder Steve Wozniak - that algorithms used to set limits might be inherently biased against women.

Biais conscient ou inconscient ?

Exemple de biais - processus de développement



Fairness



Mesurer différents types d'équité - communauté

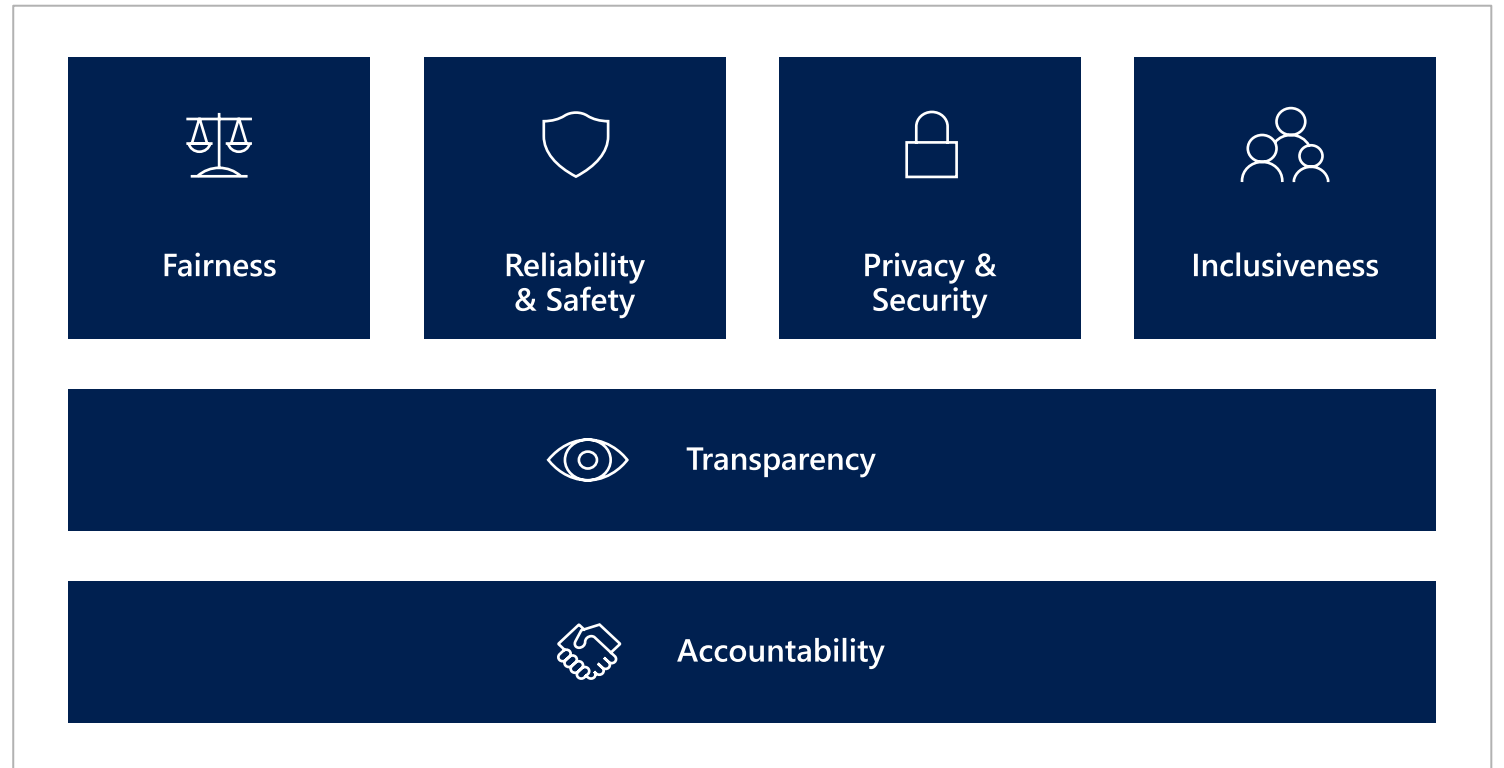


Fairness

Metrics Project	Description	Repository
AI Fairness 360	Comprehensive collection of fairness metrics, pre- and post-processing debiasing algorithms.	http://aif360.mybluemix.net/
MSR tool kit	Offers similar functionality to AI Fairness 360 with more performant debiasing algorithms.	In development. Contact Jenn Wortman Vaughan
Fairness Measures	Framework to test given algorithm on a variety of datasets	https://github.com/megantosh/fairness_measures_code
Fairness Comparison	Compares ML algorithms with respect to fairness measures.	https://github.com/algofairness/fairness-comparison
Themis-ML	Python library implementing fairness-aware machine learning algorithms	https://github.com/comicBboy/themis-ml
FairML	Quantifies dependence of model outputs on inputs	https://github.com/adebayoj/fairml
Aequitas	Web and python auditing tool. Generates bias report for model/dataset	https://github.com/dssg/aequitas
Fairtest	Audits algorithms impact on protected subpopulations	https://github.com/columbia/fairtest
Themis	Designs test cases to explore where algorithm might be exhibiting group-based discrimination	https://github.com/LASER-UMASS/Themis
Audit-AI	Python library to audit scikit-learn models	https://github.com/pymetrics/audit_ai

Éthique de l'IA – Microsoft (article préISO) perspective

- **Impartialité:** Les systèmes d'IA devraient traiter les gens équitablement
- **Fiabilité et sécurité :** Les systèmes d'IA doivent fonctionner de manière fiable et sûre
- **Confidentialité et sécurité:** Les systèmes d'IA doivent être sécurisés et respecter la vie privée
- **Inclusivité:** Les systèmes d'IA devraient responsabiliser tout le monde et engager les gens
- **Transparence:** Les systèmes d'IA doivent être compréhensibles
- **Responsabilité:** Les systèmes d'IA devraient avoir une responsabilité algorithmique



Microsoft Aether Committee

AI, Ethics, and Effects in Engineering and Research



Sensitive Uses



Reliability and
Safety



Human-AI
Collaboration &
Interaction



Fairness and
Bias



Intelligibility &
Explanation



Engineering
Practices



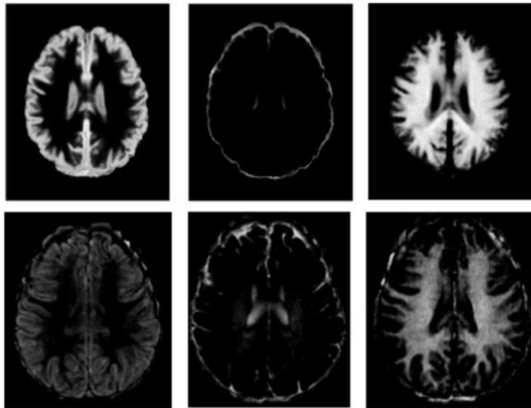
Human
Attention &
Cognition

XAI

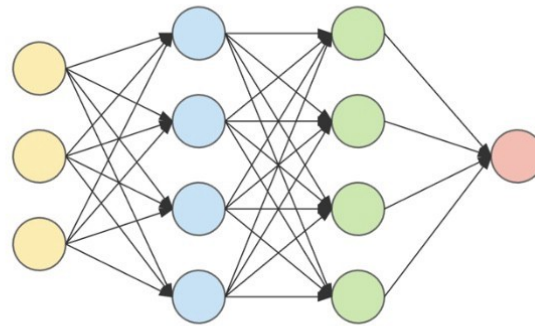


Exemple XAI

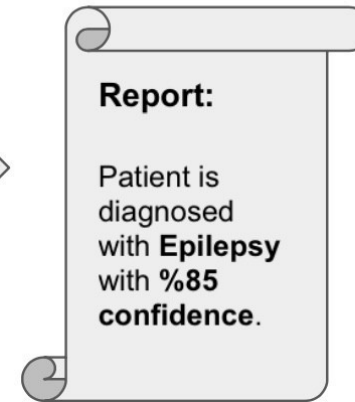
Epilepsy Detection Model with Brain MRI Data



Brain MRI data



Complex ML model



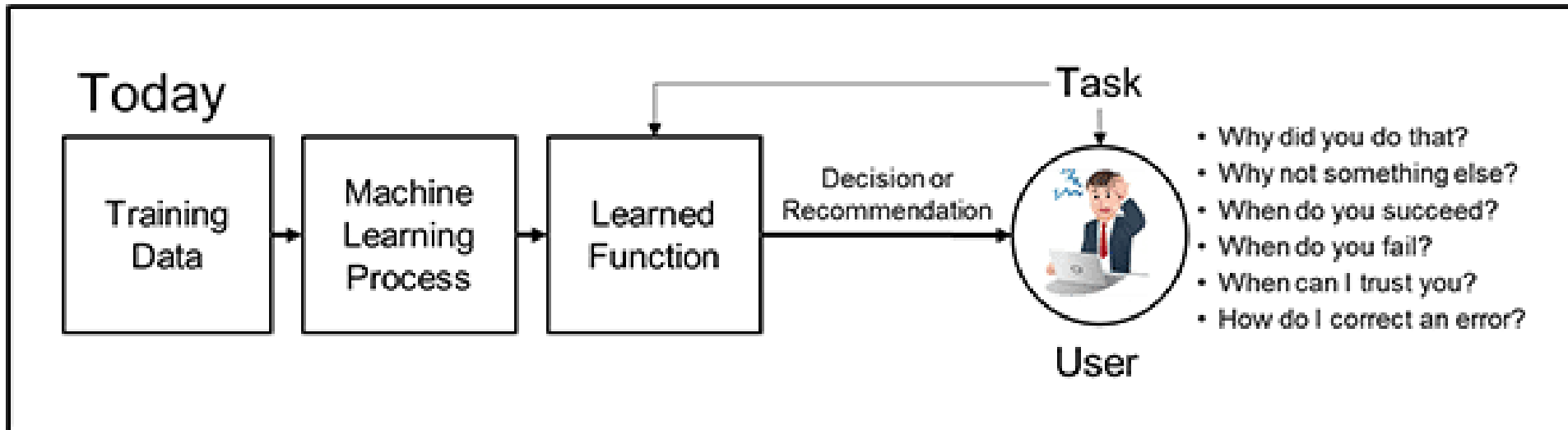
But why?!

Can I trust this prediction?

De l'IA à la XAI



Transparency

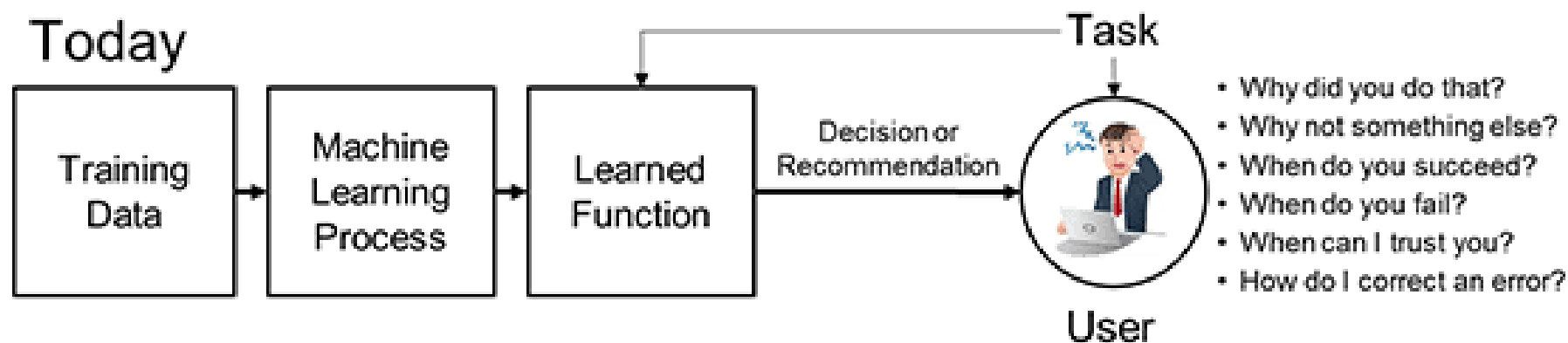


De l'IA à la XAI

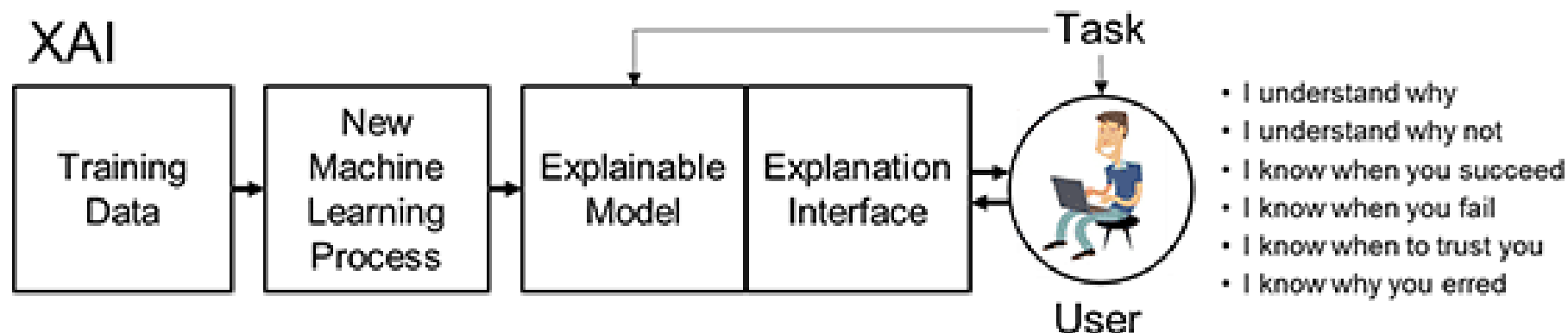


Transparency

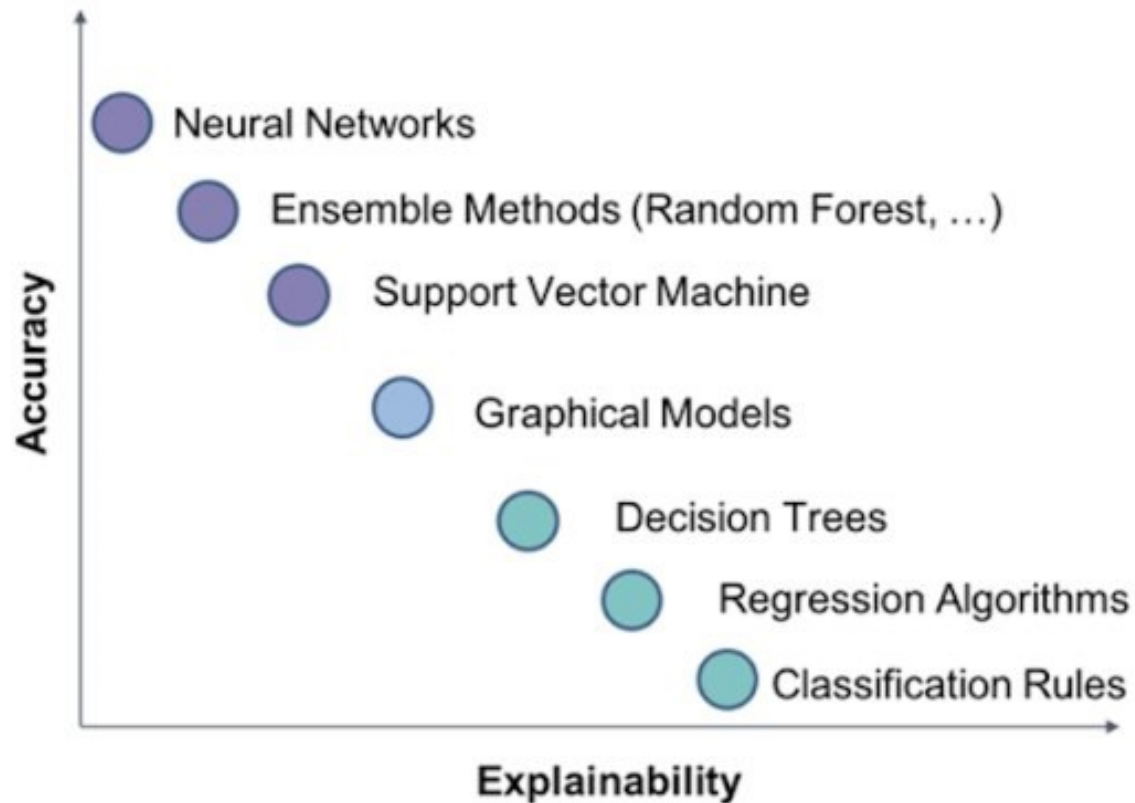
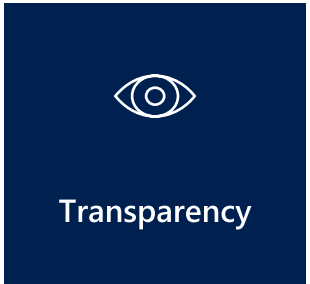
Today



XAI

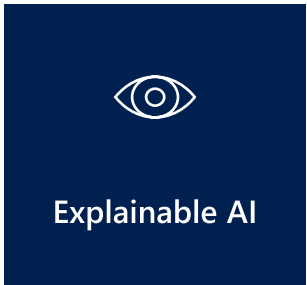


Transparence et intelligibilité

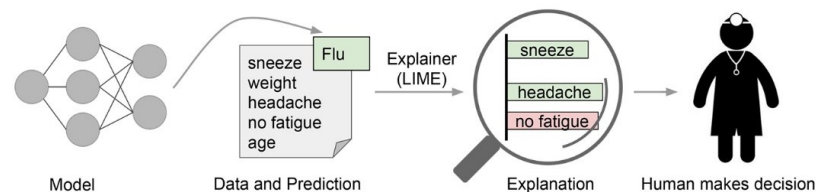
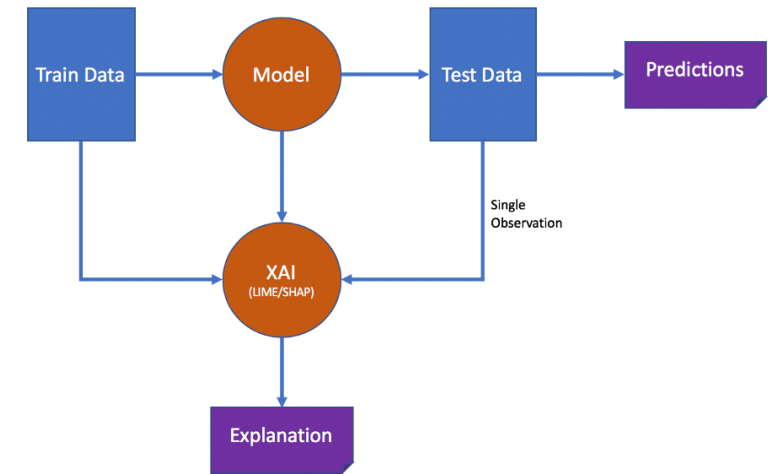
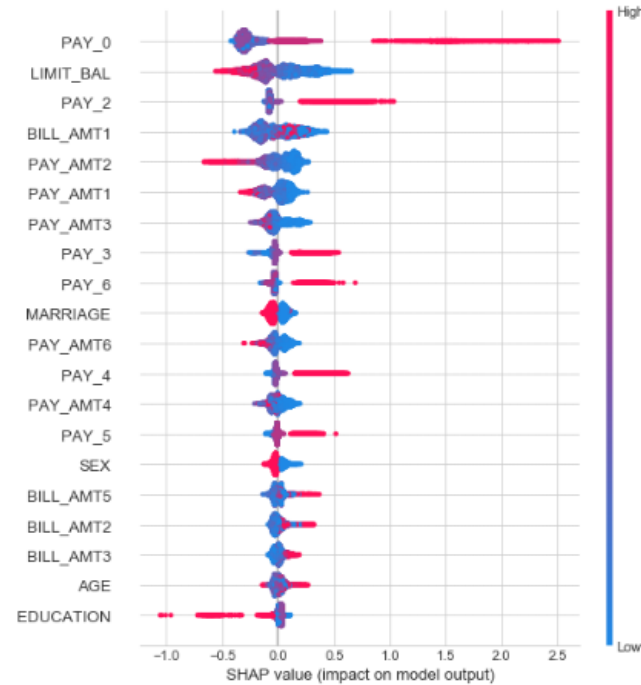


- plus grande précision grâce à des **modèles complexes difficiles à interpréter**
- Crée une **tension entre précision et interprétabilité**
- diverses **méthodes proposées pour interpréter** les prédictions mais souvent peu claires comment les méthodes sont liées et préférables

Transparence et intelligibilité – explications post-hoc



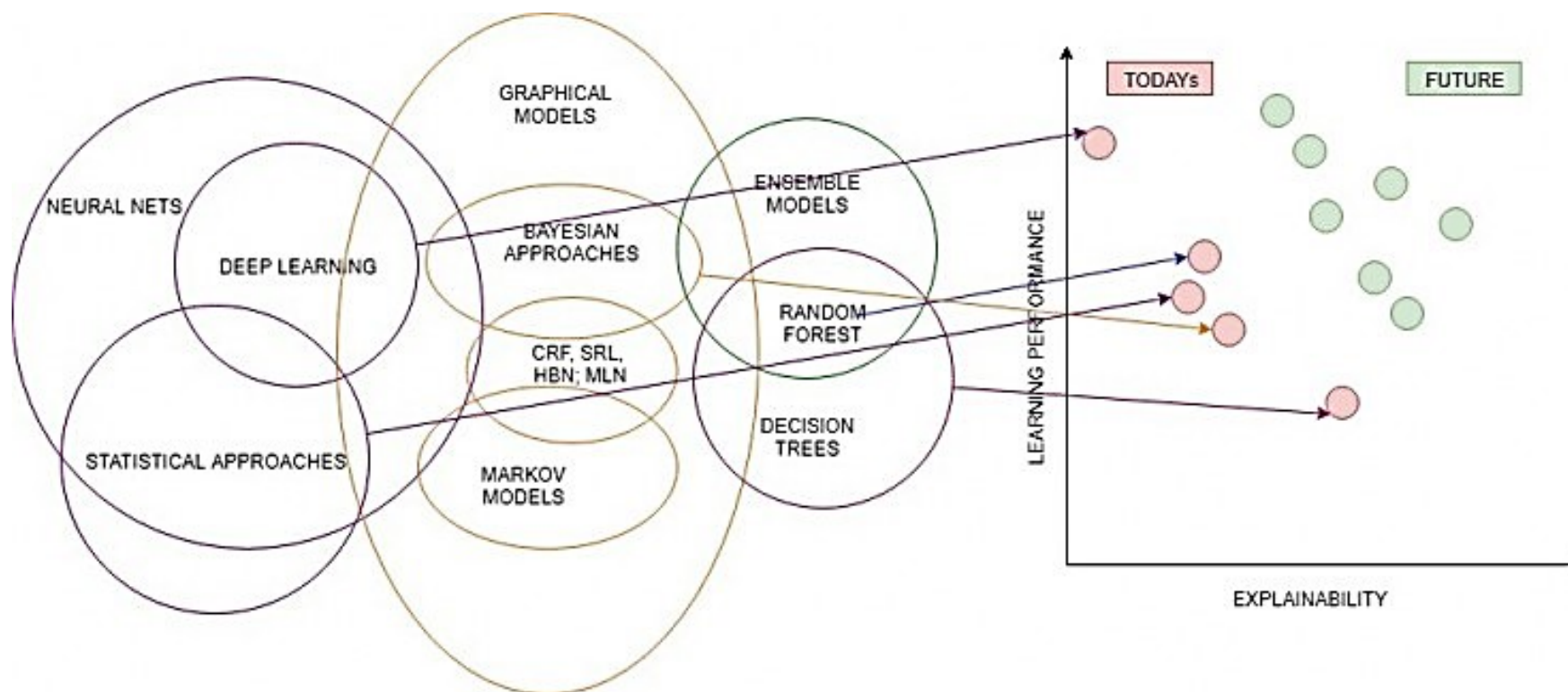
- LIME, SHAP, Sensitivity Analysis, Model Distillation
- **SHAP – Shapley Additive exPlanations**
 - derived from the shapley values of game theory
 - average marginal contribution of a feature value



Transparence et intelligibilité



Explainable AI



Explicabilité aussi cruciale que précision

« Nous devons veiller à ce que l'IA soit développée à travers une **approche humaniste, fondée sur les valeurs et les droits de l'homme** »

Audrey Azoulay
Director-General UNESCO



L'avenir que nous créons
est un **choix** que
nous faisons **maintenant**



Merci pour votre attention

