CrossMark

Artificial intelligence: Why we must get it right

Alessandro Curioni

AI and jobs: A lesson from history

Will cognitive computing ultimately lead to mass joblessness? Hearing predictions from some commentators – from the World Economic Forum to the Bank of England – you could be led to think so. However, such views are unhelpful, if not damaging, and do not stand up to historical scrutiny.

The fact is that every revolutionary machine that has replaced human workers, from the Jacquard loom onwards, has only ever *changed* the types of jobs that people do, rather than leaving them unemployed. While there may be a delay after a technology displaces people from jobs and until they gain a new role, the overall picture is one in which human-competitive technologies create still more opportunities – and breed new types of work.

The evidence comes from the business consultancy Deloitte, which undertook a fascinating exercise to get to the bottom of this thorny matter. They pored over 144 years of UK census data and tracked the job descriptions people had had over time, and at the same time chronicled the prevailing employment levels. They found that, despite people losing jobs to machines, employment levels inexorably went up over time. "The last 144 years demonstrate that when a machine replaces a human the result, paradoxically, is faster growth and, in time, rising employment," the Deloitte report says.

To see this effect, we only have to recall what devices like the revolutionary IBM PC did in offices. While putting a computer on everybody's desk ended most typing pool jobs, it spawned an industry in IT staffing, software development, network engineering and computer security. Moreover, of course, it empowered businesses to do so much more with the many thousands of applications those software developers produced.

The consultancy PricewaterhouseCoopers expects that 38 % of jobs in the US will be susceptible to replacement by AI algorithms, robots and other automated systems by the early 2030s, with Germany at 35 % and Japan at 21 %. However, concurring with the Deloitte study, PwC adds that such jobs may well change rather than vanish altogether.

So, what is essential is that AI developers ensure that cognitive computing is transformative in the job market, not destructive. They need to drive this shift in the most positive way to help people obtain the skills they need for the transition to economies in which learning algorithms play an ever more important role.

Introduction

You can hardly fail to have noticed that artificial intelligence is garnering an enormous amount of coverage in the news media. Exciting stories about advances in autonomous cars, personal robots, delivery drones, deep-learning image classification systems, ultrasmart chatbots and instant speech translators abound [1]. Yet, at the same time, other reports cover heated arguments over whether runaway machine intelligence will lay waste to employment and allow machines to gain the whip hand over humans. Frankly, it is all rather confusing. Such

DOI 10.1007/s00287-018-1087-0

Alessandro Curioni,
IBM Fellow, Vice President Europe and
Director IBM Research Zurich,
Zurich Research Laboratory,
Rüschlikon, Switzerland
E-Mail: cur@zurich.ibm.com

Abstract

Cognitive computers have the intelligence to solve many of humanity's most pressing problems, from fighting disease to cutting road deaths and coping with climate change. However, it can only do so if people trust AI's practitioners.

polarised coverage makes it difficult to work out whether AI is a help or a hindrance to society. This paper outlines the practical benefits of AI and the cognitive computers that make it work, and also outlines how IBM, alongside its partners, is steering a transparent, ethical path with cognitive technologies that should put most dystopian fears to rest.

Health: The cognitive benchmark

If you really want to get a firm grasp on the practical power of today's emerging artificial intelligence systems, you really need look little further than the effect that technology is having in the field of cancer medicine. Oncologists the world over are in the vanguard of medical professionals harnessing cognitive computing to help them both interpret and analyse clinical data – so that they can diagnose cancer earlier and choose the most appropriate, evidence-based treatment for each patient all the more quickly. Their efforts are already paying dividends – and unlike some of the wilder news media claims, peer-reviewed research papers actually put some numbers to cognitive computing's healthcare efficacy.

For instance, in July 2017 specialists at the New York Genome Center, and at nearby Rockefeller University, revealed in the journal *Neurology Genetics* [11] that their use of artificial intelligence has allowed them to cut the time it takes to identify which drugs will best treat a particular patient's brain tumour from 160 hours (that's 8 hours short of a week) to just 10 minutes. Moreover, at the American Society of Clinical Oncology [2] meeting in Chicago, a month earlier, oncologists at the Highland cancer hospital in Rogers, Arkansas, found that they could shrink the time it takes to screen lung and breast cancer patients for their eligibility for potentially life-saving clinical trials by 78 %.

Learn, reason, decide

Both these teams achieved these impressive time-saving feats using Watson, IBM's AI cloud platform for business, a system designed to apply machine learning and reasoning algorithms to massive tranches of messy, unformatted data – and provide actionable insights from it that no human would ever have the capability of doing in any useful timescale.

The whole point of this is to exploit an unused resource: more than 80 % of the world's data is unstructured or not organized in a useful way that computers can digest, so it is effectively *invisible* and so not something that we can learn and reason from. Cognitive computing will change all that: data from images, audio, language, human vital signs, medical records, books, newspapers, journals, video, the climate – you name it – and suddenly we can learn from it. Therein lies the power of AI: with 165 billion

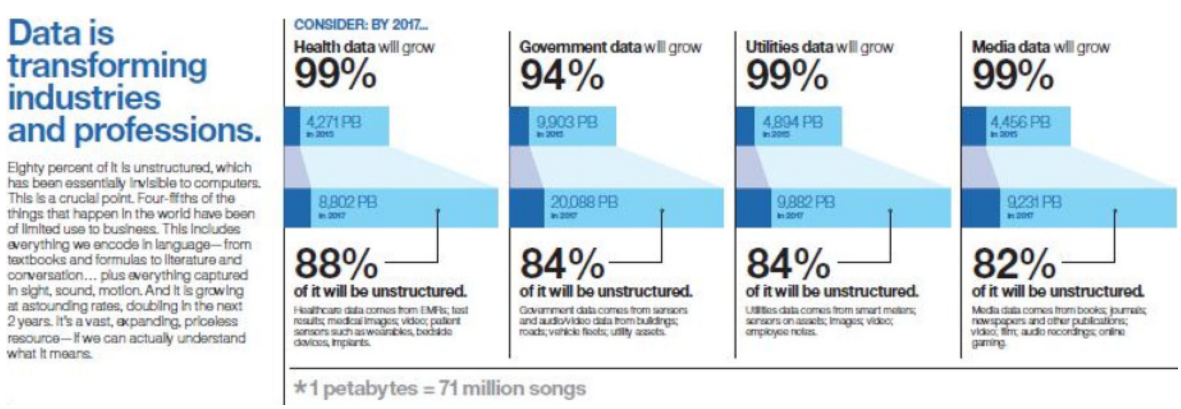

***Fig. 1** Unstructured data, once invisible to computers, will become actionable information that machine intelligence can learn from across a host of industries*

terabytes (165 zettabytes) of unstructured data (see Fig. 1) predicted to be generated by 2025, we believe that this really big data will become the oil of the twenty-first century.

Indeed, in those cancer therapy applications, Watson's oncological intellect was created by having it ingest unstructured clinical data from many thousands of medical research papers and anonymised records of cancer patients. It then applied machine learning algorithms that made correlations between the data describing the conditions of patients, and their clinical outcomes, to eventually recommend patient-specific courses of action to their medical teams – and all in clear, natural language too.

An infinite variety of promises

That AI can speed up potentially life-saving medical techniques is just one way that it has potential for doing societal good. Once driverless cars are a reality, for instance, and regulators say they are safe enough for use on the roads, probably sometime in the next decade, their extensive cognitive smarts can begin to help cut the global death toll from road accidents, which is currently running at an appalling – and scarcely believable – 1.25 million deaths per year, according to the World Health Organisation [10].

Other AI applications include precision modelling of the way our climate is changing, or running algorithms that eke more energy from renewable power grids – which will be crucial as ever more cars go electric. Moreover, as the climate changes, cognitive techniques will help us ensure food security, too, by using AI for precision farming applications, for instance. Alongside real-time language translation and deep learning systems that offer hyper-accurate image recognition – in fields from medicine to crime fighting to astronomy – it is quite clear that AI is going to be a clear and present benefit to civil society.

Yet, due to the broad concerns about AI, it is not entirely certain that the technology will be allowed the freedom to deliver on its powerful promise. The reason for this is that, like the many high-impact automated technologies that came before it – from the Jacquard looms of the nineteenth century to car-building robots of the 1970s – AI faces some strong societal objections. Chief amongst those, of course, is a fear of mass job losses – although recent research shows that novel job-replacing technologies like those looms and robots have, in fact, always *created* new jobs (see "AI and jobs: A lesson from history"). However, there is yet another issue that concerns people with respect to the cognitive revolution: a fear of unintended consequences.

The heart of a new machine

What's behind such concerns? Is the fundamental difference between AI and the machines of the two previous computing epochs: the tabulating and programming eras (Fig. 2). Tabulating machines, invented by IBM in the nineteenth century, were essentially mechanical calculators, operated by people – called "computers", incidentally – who either punched in the numbers, or who used punched cards, to make manual calculations. It worked just fine, but it was, of course, glacially slow.

Driven by the needs of allied codebreakers in World War 2, however, the programmable digital computer emerged in the mid-1940s, and digital computers became full-scale business machines by the 1950s. These could execute whole series of logical

***Fig. 2** Computing has moved from mechanical machines that undertook simple arithmetic to those that could be programmed digitally to intelligent machines that can learn and reason from completely unstructured data*

instructions – i. e. programmes – blisteringly fast and with the ability to do conditional branching, allowing IF–THEN routines and loops, lending them great functional flexibility. Programmable computers became ever faster as the base technology moved from vacuum tubes to transistors and then on to microprocessors, taking advantage over the decades of everything that Moore's law – the doubling of microchip density every 18 months – enabled them to have.

However, programmable computers have a problem: they need feeding data in a form that each programme can understand. That can take an age to format correctly and in many cases, it is just too complex a task for human operators to undertake at any useful scale. Yet cognitive computers do not need to be programmed to deal with unstructured data. This is their very *raison d'être*; they can learn by themselves what is going on inside complex, ambiguous, unstructured datasets.

Moreover, it is the ability of AI to learn and reason, effectively creating surrogates of human specialists in vast numbers of fields of endeavour, that has some people worried. Will it become smarter than us, they ask; will it learn to operate independently of us? If so, could it effectively control us, perhaps, by leveraging its own control of essential utilities, say, to get humans to do its bidding?

Facts go to Hollywood

Warming to their theme, and encouraged by the kind of Hollywood science fiction seen in the *Robocop* and *Terminator* movies, machine intelligence critics predict dystopian situations such as AIs pursuing arbitrary goals that leave them in control of life-critical systems – like water utilities, electricity grids or even weapons systems that let them start conflicts. Yet just as engineers today build cybersecurity into their programmable systems, one way to ensure that cognitive computers are safe and trustworthy will be to limit the capabilities and what they can control by design, from the very outset, so they *cannot* harm and can only help humans.

Fears that even measures such as these could be circumvented are rooted in the notion of the "technological singularity" which posits a potential point in the rising intelligence of machines after which AIs become superintelligent, and perhaps sentient, and this has further fuelled cognitive concerns. It may sound like such worries can be ignored, since Hollywood is only a purveyor of fiction and the much-vaunted singularity is as yet an unproven hypothesis.

However, worries persist from the most educated of sectors: Cambridge University cosmologist Professor Stephen Hawking, alongside many computer scientists has publicly voiced serious concerns [3] in an open online letter over AI and its risks. Hawking et al. are pushing for the societal impacts of AI to be researched deeply to avoid such pitfalls well in advance, so that it only produces "robust and beneficial" technologies.

While that is all well and good, care must be taken. The reason being that if such concerns trickle down into governmental regulation of AI they could have a very real impact on how well learning machines are allowed to help us solve global problems. It is only AI's ability to extract value from unstructured data that will help us find answers to some of our biggest problems, such as fighting disease, providing food, energy and water security, keeping well ahead of the effects of climate change and managing ever more complex algorithm-centric economies.

The only way we are going to make sense of this information is by using cognitive systems. For mankind, it is a technology that is simply too good to lose and that means that we must get the regulation – hopefully self-regulation – correct from the start.

Transparency and trust

It is clear that with public worries abounding, the first thing practitioners must do is convince people they can *trust* AI. At IBM a number of measures guide our approach to making cognitive systems trustworthy, starting with our intentions for this technology. First and foremost, we see AI as *augmenting* human beings, not replacing them. Man and machine working side by side can enhance and scale human expertise to new heights. In fact, we believe this so strongly that, for us, AI stands for "augmented intelligence", not artificial intelligence. So, for instance, IBM Watson works *alongside* medical specialists, giving them superfast ranked search hits for diagnoses, for example, that doctors then study to bring their expertise to bear to improve decision making. It *does not* make them redundant.

Above all, our approach is sustainable: cognitive systems will be embedded in processes, systems, products and services that need intelligence – but all of them will remain within human control. So, they will not realistically be able to attain any kind of independent agency, or what some might call sentience or even consciousness. They will just do an amazingly smart job, double quick, on messy data that no-one has been able to work on before.

In addition, our ethical principles will ensure that we make clear when, and for what purpose, AI is being applied to data: no-one will be bamboozled into giving up their data to machine intelligence unknowingly, as has happened with some healthcare AI projects. We will also be algorithmically accountable and able to explain how our technology behaves and reaches the decisions it does. Some organisations have claimed that AI is a black box technology [7], in which changing neural weights deep inside learning networks do their own thing, out of sight, out of mind – but that position is becoming increasingly untenable as researchers discover ways to probe the decision making processes going on within AI systems. That has been IBM's thinking all along. If that is not done, datasets and algorithms may produce results skewed by cultural biases, on race and gender perhaps.

To instil such ethics further into the AI ecosystem IBM, in 2016 cofounded – alongside partners Google DeepMind, Microsoft, Facebook and Amazon – an advocacy group called the Partnership for AI [4]. This group will engage with customers, civil society groups, governments and non-profits alike to come to a consensus on safety, fairness, accountability and transparency in the machine intelligence arena. The group is already in deep discussions on the issue of safety-critical applications of machine intelligence, for example.

Avoiding the long grass

It is not all about safety, however. By engaging widely with all kinds of stakeholders through the Partnership for AI we also hope to ensure cognitive technology does not become entangled in the aforementioned regulatory arguments. If it does, it may risk getting kicked into the long grass while politicians argue over everything from superintelligence risks to big data privacy. Given AI's record of unseemly delays, that would be unhelpful. In its six decades of history, AI has already endured a number of lengthy periods in the wilderness – known as "AI winters" – due to its early adherents following some technological dead ends. Indeed, it was IBM's Watson that steered us out of the last AI winter – and how we did so bears retelling.

AI started well. In 1950, the British mathematician Alan Turing, whose paper *On Computable Numbers* [8] inspired the development of the digital computer, published yet another highly influential paper, called *Computing Machinery and Intelligence* [9]. In the latter, he asked the provocative question "can machines think?" and outlined his now-famous imitation game, or Turing Test, saying that if a competitor could not tell if a hidden computer or a person was answering their questions on a teleprinter then the computer could indeed be deemed intelligent.

Turing's paper sparked a storm of discussion – not least amongst clergy affronted by the notion of electronic brains competing with God-created man. In 1956, 2 years after Turing's death, a group of eminent computer scientists held a 6-week summer workshop at Dartmouth College in Hanover, New Hampshire to put the field of these "thinking machines" on a firm research footing and to come up with an R&D agenda. The attendees include data science luminaries like information theorist Claude Shannon of Bell Labs, cognitive scientist Marvin Minsky of MIT and Nathaniel Rochester, designer of IBM's first computer. It was the workshop's organiser, Dartmouth computer scientist John McCarthy, who coined the term that has stuck to this day, dubbing the field of endeavour "artificial intelligence".

Firing the starting gun

AI took off quickly as the Dartmouth delegates, wielding high hopes and new computers, coded up systems that amazed the media and the public by beating people at checkers, solving word puzzles and even speaking English. Some even played music. Initially, investment poured in. Progress seemed so assured that Minsky and his colleagues believed they'd have human-smart AI within 20 years. However, their hopes were dashed as the limitations of their hardware and algorithms were soon realised, and by 1974 research funding had all but dried up. This sorry situation was dubbed an "AI winter" as a somewhat overdramatic play on the then prevailing cold war concern over an apocalyptic nuclear winter, in which airborne blast dust shadows Earth from sunlight.

That first AI winter ended with a strong but short-lived thaw in the 1980s as "expert systems" emerged. This powerful form of AI programmed human expertise into computers for retrieval by nonexperts – and it took business by storm. However, by the 1990s even expert systems' inflexibilities became clear: they were restricted to the small knowledge sets of their programmers and could not cope with queries that went beyond these. So once again, in 1987, another AI winter set in.

A decade later, however, and things were looking up for AI when IBM astonished the world when our Deep Blue supercomputer defeated the world chess champion, Garry Kasparov, under tournament conditions. What was particularly interesting about that 1997 event was what Garry did after it, however: despite his loss to Deep Blue, he had enjoyed working with the computer so much that he went on to play in freestyle chess leagues where players can use any kind of computer assistance tools they like. Moreover, it turned out that the players who combined chess-computer-assistance tools with their own gaming intuition achieved the best results.

It was a sign of something that MIT and DARPA computer interaction pioneer JCR Licklider [5] had predicted way back in 1960: the beginning of a symbiosis between human and machine, where we begin to work ever more effectively *alongside* intelligent computers, with, as I said earlier, the best situation being the one in which the computer *augments* the human's mind, rather than replacing it.

Now hold that thought while we fast forward to 2011 and the event that killed the long-running AI winter stone dead: IBM Watson winning the television gameshow *Jeopardy!*. Unlike the lookahead brute force of a chess playing supercomputer, Watson was designed to make use of unstructured data to learn everything it could about general knowledge from 200 million pages of data. It answered *Jeopardy!*'s subtle, complex and pun-laden questions without the pre-programming the computers of the previous six decades had needed.

Since then, with AI entering the sunlit uplands after that last, punishing winter – DeepMind's technology defeated the world Go champion last year, for instance – Watson has been trained on increasingly varied types of datasets on a wide variety of subjects and, as Fig. 3 shows, Watson now has at least more than a dozen different APIs lending it expertise in, for instance, language translation, image recognition, image tagging, face detection, various types of

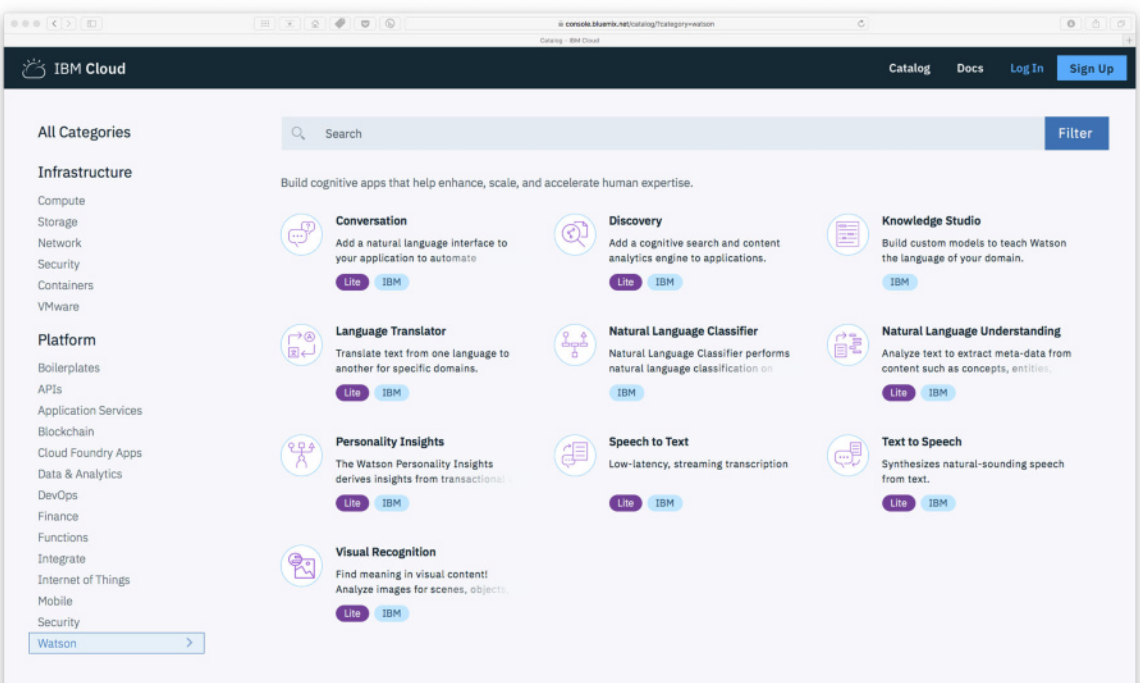

***Fig. 3** It started with an API for question answering in Jeopardy!, but Watson now has APIs making it an expert in many other areas of endeavour. Its Jeopardy! win marked the end of the "AI Winter".*

medicine and, of course, general knowledge. Making a new API from your own, or public, sources of unstructured information is simple, using IBM's Cognitive Discovery tools, which take the drudgery out of the info ingestion process. Better still, perhaps, Watson APIs in the cloud can be queried from mobile devices, lending developers all kinds of potential for AI apps in your pocket [6], covering city guides, product recommenders, health guides and quizzable chatbots. Of course, robots and IoT devices can access IBM Watson's cloud API's too.

The algorithmic switchback

What is fuelling today's resurgence in AI is a perfect storm of three technologies: first, cloud hosting services are making massive datasets available online instantly for cognitive systems to get to work on. Second, deep learning algorithms, using multi-layered neural networks, are proving one of the best ways yet to improve machine learning and reasoning. Third, massively parallel graphics processor units (GPUs) can make such algorithms run at speeds that finally deliver human-competitive – or better – results at tasks like image recognition.

Which of those factors matters most is changing, however. Over the last decade IBM estimates that some 66 % of the advances in AI have come from faster GPUs, 20 % from the bigger datasets and 10 % from improved algorithms. However, the driver of progress is now switching to ever smarter machine learning algorithms, such as those performing supervised, unsupervised and predictive learning.

In a supervised learning system, such as the aforementioned medical applications of Watson, deep layers of neural networks are trained to recognize patterns – such as lesions in medical images – in vast tranches of unstructured data. In an unsupervised system, however, the algorithm learns through undertaking a task – such as playing a game like poker – and being rewarded when it gets something correct. This "reinforcement learning" process allows the algorithm to teach itself and so is far more powerful, as developers do not need to know how to programme the system in advance. Still more promising algorithmic techniques are emerging, too, such as predictive learning, which attempts to learn by observation like humans do.

It is not all about software, however. AI is far too young a field to say there is one right way to do it. So, to get a broad base of hardware options IBM is also working on "neuromorphic" systems that use phase-changing materials, which alter their resistance to mimic the storage of synaptic weights, to allow the construction of microchips containing physical models of artificial neurons, rather than software ones. It is too early to say, but there may be applications where neuromorphic chips provide the best approach to building a cognitive intelligence, perhaps in tandem with a cloud-based system like Watson.

IBM is also developing quantum computers which may play a role in future AI systems, too, using their quantum abilities to calculate numbers at blistering speeds far in excess of classical computers. These could feed into AI and make cognitive decision making all the faster. However, IBM is also working on "approximate" computers – which save a surprising amount of power by not calculating the least significant bits in very large numbers – which makes them promising for very-power-hungry climate modelling applications. Such processors disable, or physically remove, the last few transistors on an arithmetic chip to save the power wasted on calculating numbers to unrequired accuracies.

So, it is clearly an extraordinarily exciting time in artificial intelligence, not least because researchers have not even scratched the surface of what cognitive systems will be capable of. What is certain, however, is that for them to press ahead and help solve major problems with machine intelligence, the players in this field must do so with a strong code of ethics and be transparent and accountable about it. Because the prize here is simply too good to lose.

As IBM senior vice president Dr. John Kelly puts it: "The success of cognitive computing will not be measured by Turing tests or a computer's ability to mimic humans. It will be measured in more practical ways, like return on investment, new market opportunities, diseases cured and lives saved."

Now that's an incentive we can all get behind.

References

1. Anthes G (2017) Artificial intelligence poised to ride a new wave. In: Commun ACM, July. https://cacm.acm.org/magazines/2017/7/218862-artificial-intelligence-poised-to-ride-a-new-wave/fulltext, last access: 15.2.2018
2. Beck JT et al (2017) Cognitive Technology Addressing Optimal Cancer Clinical Trial Matching and Protocol Feasibility in a Community Cancer Practice. In: 2017 ASCO Annual Meeting. http://abstracts.asco.org/199/AbstView_199_193610.html?cm_mc_uid=86020782608115004605782&cm_mc_sid_50200000=1500469184, last access: 15.2.2018
3. Hawking S et al (2016) Open Letter on the Dangers of AI. At Future of Life Institute. https://futureoflife.org/ai-open-letter, last access: 15.2.2018

4. IBM et al (2016) Establish Partnership on AI. September. https://www.partnershiponai.org/2016/09/industry-leaders-establish-partnership-on-ai-best-practices/, last access: 15.2.2018
5. Licklider JCR (1960) Man-computer symbiosis. IRE T Hum Fact Elect. http://groups.csail.mit.edu/medg/people/psz/Licklider.html, last access: 15.2.2018
6. Marks P (2014) Watson in your pocket: Supercomputer gets own apps. New Scientist. https://www.newscientist.com/article/dn25476-watson-in-your-pocket-supercomputer-gets-own-apps/, last access: 15.2.2018
7. Snow J (2017) Brainlike computers are a black box: Scientists are finally peering inside. sciencemag.org. March. http://www.sciencemag.org/news/2017/03/brainlike-computers-are-black-box-scientists-are-finally-peering-inside, last access: 15.2.2018
8. Turing AM (1937) On Computable Numbers. In: Proceedings of the London Mathematical Society. http://www.turingarchive.org/browse.php/b/1210, last access: 15.2.2018
9. Turing AM (1950) Computing Machinery and Intelligence. In: Mind. https://doi.org/10.1093/mind/LIX.236.433, last access: 15.2.2018
10. World Health Organisation (2013) Road Traffic Deaths, Global. http://www.who.int/gho/road_safety/mortality/en/, last access: 15.2.2018
11. Wrzeszczynski KO et al (2017) Comparing sequencing assays and human-machine analyses in actionable genomics for glioblastoma. Neurol Genet, doi:10.1212/NXG.0000000000000164