Table of Contents

The Rise of the Machines: Analysis of Artificial Intelligence

Part A – Background

In 1950, Alan Turning first, first posed the question “can machines think?”. The term “Artificial Intelligence” arose in 1955 and can be defined as “activity devoted to making machines intelligent”; and intelligence as the “quality enabling an entity to function appropriately, autonomously and with foresight in its environment” (Nielsson, 2009).

The evolution of AI follows the linear innovation model developed in the 1940s, during a period ruled by the ideology of “pure science” (Godin, 2005):

Basic research → Applied research → Development → (Production and) Diffusion

Improvements in AI are contingent to improvements in our knowledge of human cognition. The first endeavours in artificial thinking, came from academics in esteemed universities; particularly MIT, Stanford and CMU (Nielsson, 2009). This indicates that science was the key catalyst of AI’s development. The linear model will continue to repeat its cycle in the future, as more basic research on human cognition is conducted and applied to develop AI applications.

Cognitive science forms the basis of AI’ fundamentals. AI has two main divisions:

1. General machine learning – systems adapting via rewards and penalties (most widely used form of AI today)
2. Deep learning – systems learning from data using artificial neural networks, like neurons in the human brain (Nielsson, 2009).

IBM and Google are fighting for leadership in the AI industry due to their early entry, substantial R&D investments and pre-existing knowledge-assets. They focus on radical innovation, by using supercomputers to discover whether machines can truly outperform humans. Since the 2000s, AI implementation simplified due to increase in computer power and new tech entrants such as Facebook, Amazon and Tesla started leading the commercialization of AI. The new entrants are engaged in product innovation by creating platforms that optimize the way consumers interact with big data. This allows them to commercialize a simpler more convenient process, in the context of their respective industries (e.g. Amazon’s recommendation AI boosting e-commerce sales by 29%).

PART B – Technology Cycles and Performance

Historically, innovation in AI occurred through incremental improvements of processes via academic research. Radical innovation came from testing a system's effectiveness by building products around studied processes.

The first radical process innovation in AI was *backpropagation* (1981), an algorithm that allowed neural networks to adjust themselves if their output is incorrect (Dormehl, 2017). In response to this, Navlab – the first self-driving car – was built by CMU in 1984 (Dormehl, 2017), marking the foundation of machine learning and computer's ability to react to stimuli autonomously. The development of Pentium-III microchip in 1999 (Moore, 1975) [see exhibit B in appendices], exponentially increased computing power and decreased costs of systems. This simplified implementation of AI and the focus of AI-innovation after 2000, shifted from process to product.

Technological development follows a pattern of slow initial improvement (1), accelerated improvement (2) and diminishing returns (3). For AI, Stage.1 is characterized by broad scientific research, Stage.2 by commercial benefits and Stage.3 by "technological singularity" – i.e. machines outperforming humans (Shanahan, 2015).

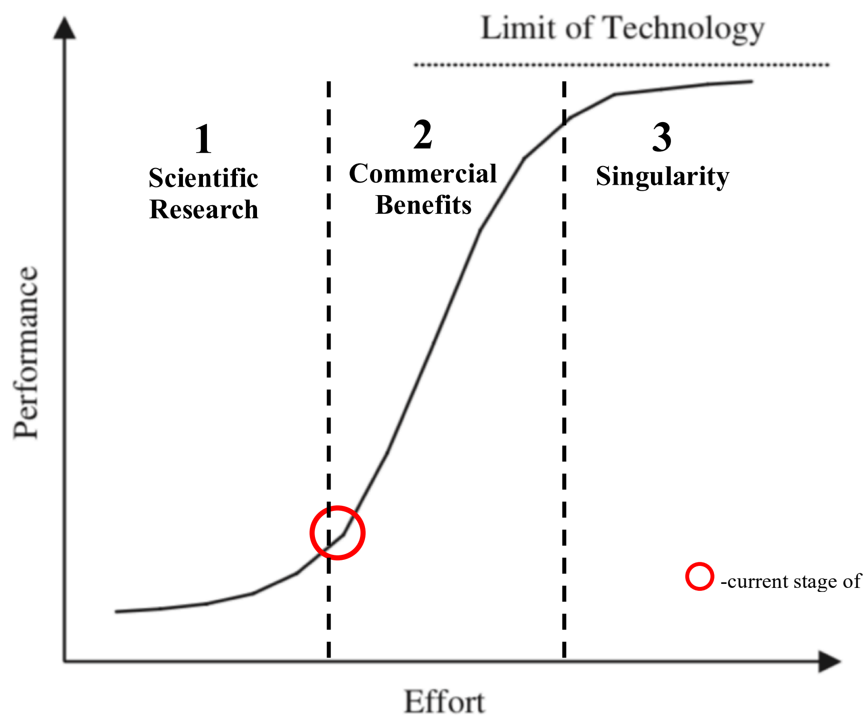


Fig.1 S-curve of AI's development

Initial performance improvement is slow because the fundamentals are poorly understood (Foster, 1986). AI development is exiting Stage.1 and entering Stage.2, as cognitive research has already formed the fundamentals.

Using Abernathy and Utterback technology cycle as a framework, AI can be placed towards the end of the fluid phase, going into the transitional phase (1975). Process development is segmental, competition is intensifying, and a few stable designs have emerged (transitional phase). Yet product development is performance-maximizing (fluid phase) or sales-maximizing (transitional) (Abernathy and Utterback, 1975), depending on the industry.

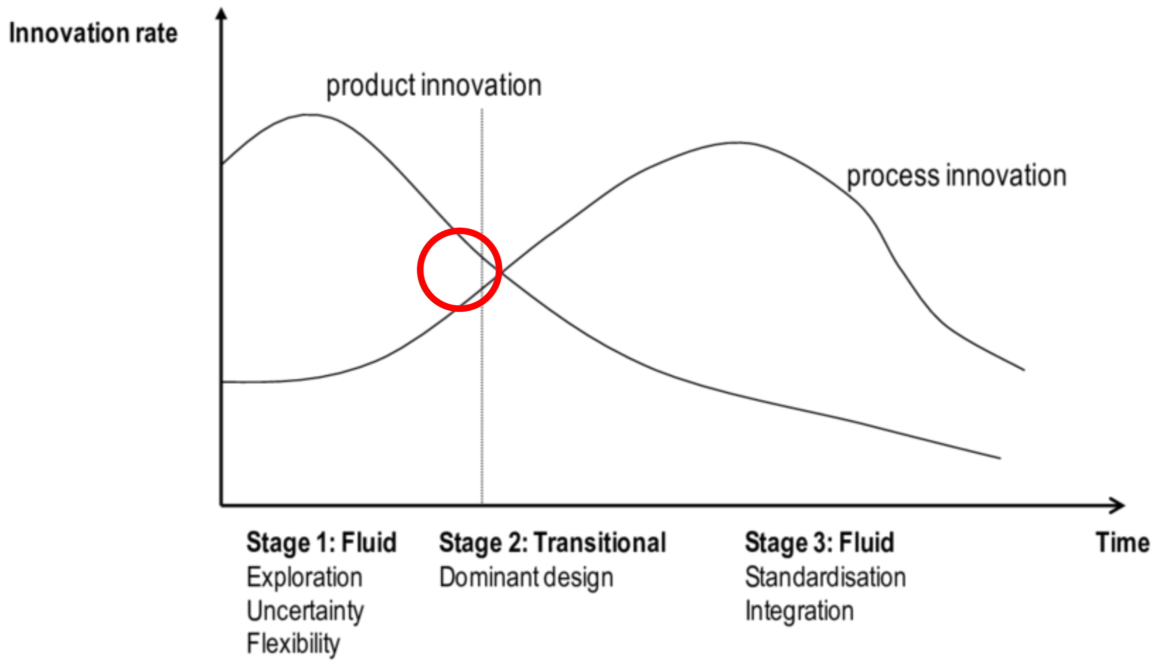


Fig.2 Abernathy and Utterback innovation model (1975)

Globally there are over 3400 companies trying to create AI solutions (Asgard, 2018) [see Exhibit A], suggesting a departure from the fluid phase, where relatively few firms operate (Abernathy and Utterback 1975). 84% of these companies are operating in the US, Western Europe, Israel and China. This adheres to a feature of the fluid phase where affluent markets have the highest concentration of production (Abernathy and Utterback, 1975), due to the largest variety of inputs; in AI's case that is computer power and professional knowledge.

AI has not fully transitioned out of the fluid phase due to the lack of dominant design. Since AI is software, it is constantly being updated to fit the vast range of applications that it is deemed suitable for. For a dominant design to be achieved major components need to not vary from one model to another (Abernathy and Utterback, 1978). Experience is deemed mandatory to discover what the market requires as a standard (Anderson and Tushman, 1991). Therefore, it is technology veterans IBM, Google and recently Microsoft, that are racing to establish *the* benchmark for all future AI software.

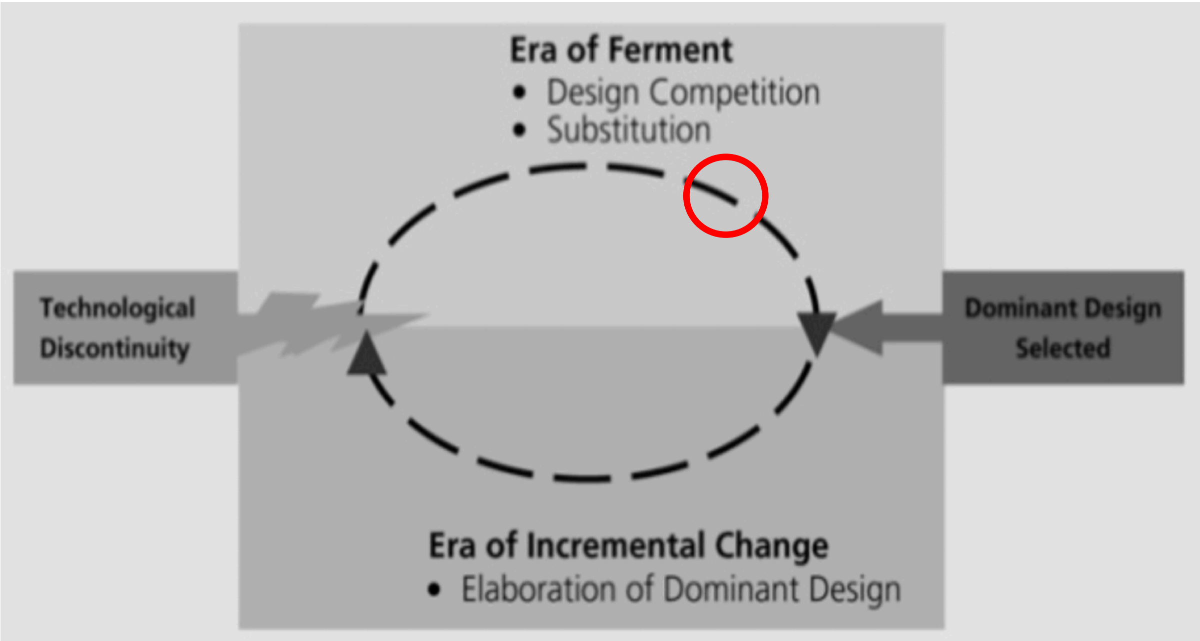


Fig. 3 Anderson and Tushman (1991) technology cycle model

Following the technological discontinuities of machine and deep learning, facilitated by backpropagation, AI can be placed in the Era of ferment of Anderson and Tushman's cycle (1991). Within the Era of ferment, AI lays on the fence between the era of substitution and of dominant design. In cases where AI is already yielding economic benefits, such as recommendation systems, the technology has substituted its predecessors. Yet in others where it has high potential, but implementation faces functional or ethical limitations, such as healthcare or automotive, conventional methods persist. Nonetheless design competition has begun in the industry.

As AI develops, disruption to current processes will occur to free up resources and energy for innovation (Schumpeter, 1942). Routine-jobs will be displaced as a consequence of AI's adoption. The creative destruction that AI poses is not to actual jobs, but to the people running them. The negative implications of AI on employment can be tackled by upskilling labor, ridding people from boring jobs and evolving the workforce towards more analytical tasks.

Measuring AI performance is challenging since 'intelligence' is a concept too abstract to be objectively measured in humans, let alone machines; it must be assessed within a task. Initially, performance was measured by machines' ability to think or act humanly. Relying on symbolic systems (pre-made datasets) a computer would use search algorithms to make deductive steps (Everitt and Hutter, 2018). The first measurement was the Turing Test. Using typed texts from both a computer (A) and a person (B), an "interrogator" (C) had to distinguish between the machine and the human (Turing, 1950). The test was completed for a first time in 2014 by a chatbot simulating a 13-year-old Ukrainian boy (BBC, 2014). As techniques such as deep learning evolve, induction becomes the focus of AI performance, i.e. models learning from a single data not requiring large datasets to be fed by humans. Therefore, measuring whether machines are "doing the right thing" (acting rationally; Fig.3) rather than trying to gauge their ability to emulate human thinking, will by the most important performance indicator for AI in the future.

	Thinking	*Acting*
HUMANLY	Cognitive Science	Turing Test, behaviorism
RATIONALLY	Laws of thought	**Doing the right thing**

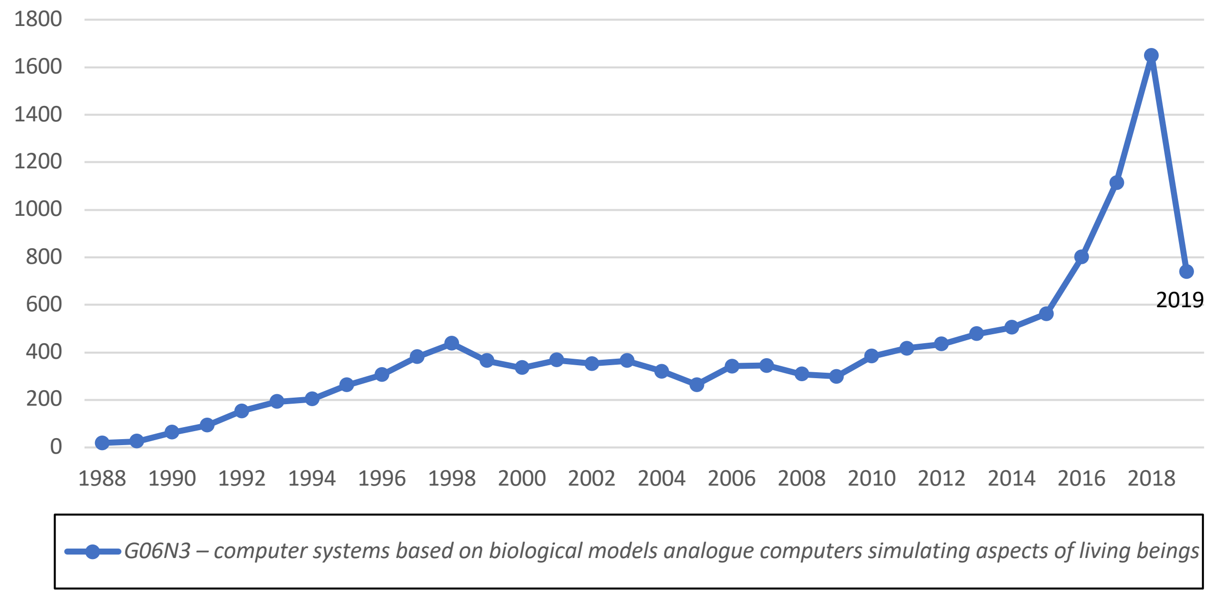
Fig. 3 Scientific perspectives on intelligence (Everitt and Hutter, 2018)

Section C – Intellectual Property Strategies

Identifying a single patent category encompassing AI is challenging due to the everchanging opinion on what constitutes AI. To track patent development overtime, a representative patent class was selected, and is assumed that all AI-related patent classes follow a similar pattern [see Exhibit C].

Fig.4 Granted G06N3 patents in the last 30 years

(sources: WIPO, Espacenet & Lens)



The initial pattern of AI patenting (1988-1998) follows a linear progression. The steady increase in G06N03-patents can be explained by the steady increase in computer processing power suggested by Moore's law (1973). Moore states that the number of transistors in a microchip increases linearly every 18 months until 1999 [see Exhibit B], thus so did the number of patents thanks to the incremental innovation in computer power. Additionally, in the late 1980s, AI models shifted from being knowledge-driven to data-driven; from relying on 'experts' teaching models all necessary steps to learning autonomously from information. This increased the number of commercial uses and consequently patent applications.

The plateauing in patent activity between 1998-2013 comes from an alternative strategy for IP protection used by major players. Instead of focusing on developing their own technology, large firms shifted to acquiring start-ups and medium sized companies with an existing patent portfolio. Buying out potential competition before it affects the business, indicated a strategy to externally source tangible/intangible IP such as talent, know-hows or patents themselves. Since 1998, 434 companies in the AI sector have been acquired, with Alphabet ranking first in acquisitions (18) (WIPO, 2019).

The boom in AI patents began around 2012 and is increasing exponentially to this day. 53% of all AI-related patents have been published since 2013, and the total number has reached more than 340,000 (WIPO, 2019). The rationale behind the rapid increase is the continuous improvements of microprocessors, which allowed for breakthroughs in machine learning and increased adoption of AI. Data-driven systems became cheaper, many businesses began implementing them and focused on patenting the application of the AI, to protect their competitive advantage.

The countries dominating in patent activity are U.S., China and Japan. Interestingly, companies represent 26 of the top-30 patent applicants (WIPO, 2019), showing a shift from theoretical research of AI to commercial AI-powered products and services. The major players in AI patenting are IBM (8290), Microsoft (5930) and Toshiba (5223) – Alphabet ranks 10th [see Exhibit D]. The industries experiencing the highest concentration of patent publications are: transportation (15% of all AI-related patents), telecommunications (15%), and health care (12%) (WIPO, 2019).

Chinese universities represent three out of the four non-commercial organizations in the top-30 patent leaders list. Chinese Academy of Sciences (CAS) ranks 17th overall and 1st in patent filings among universities and research organizations globally [Exhibit E]. CAS' lead in patenting through science hints at China's unconventional focus on theoretical research, as they hope to discover revolutionary techniques that will establish their global dominance in AI.

Although citations in patent applications vary, with some firms citing competitors' inventions and some citing their own, the co-ownership of patents is rare among top applicants' portfolios. No entity among the top 20 applicants co-owns more than 1% of its AI portfolios (WIPO, 2019). This indicates fierce competition, with some players trying to dominate within their area of expertise and others seeking to dominate the industry as a whole. Both IBM and Microsoft, have portfolios spanning a range of AI techniques, applications and fields (WIPO, 2019). They dominate and compete within the most profitable and scalable AI areas today – machine learning (40% of all AI patents), natural-language-processing and knowledge-reasoning. This indicates that the leaders in AI patenting are not limiting their activity to a specific industry or field, but instead are engaging in a battle for dominant design in the most developed AI applications.

Section D – Standards

According to Kretcher's (2000) model, setting standards during the Information age (post-Internet) relies on etiquettes, also called network protocols. Standardization of AI is more abstract due to lack of a clear definition, and the technology's recent diffusion in the marketplace explains why standards have not been established yet. Nonetheless, it can be predicted that AI's standardization process will follow that of the Internet's. Like the Internet, AI is concerned with packaging and transmitting data and it relies on computers to do so. Therefore, it is rational that network protocols, i.e. rules and conventions for communication between AI devices, (Kretchman, 2000) are established. There are two organizations developing standards within AI, both concerned with establishing them on an international level: ISO/IEC JTC 1 Standards Committee on AI (SC 42) and IEEE (Cihon, 2019). Each organization focuses on developing standards within different set of AI areas [See Exhibits F&G].

Due to the disruptive potential of the technology, state involvement with setting AI standards is inevitable. International standards are a stated priority by the major national players in AI (Gihon, 2019) – U.S. and China (historically known for setting national barriers to protect its domestic economy). This indicates a willingness by states to co-operate with organizations, and an awareness that AI benefits will come from dominating market share on a global scale.

Currently, private companies also contribute to creating standards in AI technology. There are three main platforms for AI software packaging and development: TensorFlow (Google), PyTorch (open-source) and OpenAI Gym (Elon Musk) (Gihon, 2019). These platforms are privately created and are responsible for the creation of nearly all AI products and services. The dominance of these platforms creates an indirect network effect, which in turn leads to increasing returns for the owners of the platforms. Providing valuable products that consumers can interact with increases the knowledge about AI solutions. Consequently, this leads to an increase in market demand for AI applications, which in turn causes more developers to create more applications on these platforms. This self-reinforcing relationship between the install base (developers on these platforms) and the complimentary goods/services (new AI applications), allow certain private entities to get ahead of competition. The widespread adoption of these platforms and their ability to adapt to consumer demands, depicts how increasing returns can magnify a platform's advantage, and the product or company can go on to become a standard in the technology (Arthur 1996).

Part E – Forecasting

AI is entering the second stage of accelerated growth the technological development S-curve (Fig.1). In the short-run (the next 5-10 years) the performance of practical AI will improve from thinking humanly to acting rationally (Fig.4). The next stage of AI will shift from relying on deduction, such as machine learning systems requiring rewards/penalties for their outcomes, to inductive systems. Based on the opinion of 30 experts, deep reasoning will be the discontinuity bringing the new wave of AI advancement in the next decade (IBM, 2019). Deep reasoning, the evolution of deep learning, will allow for systems to have common-sense. Machines will be able to make complex decisions and deal with changing situations, much like humans (IBM, 2019). For deep reasoning to mature, the current obstacles impeding AI innovation, such as the need for human-labelled data and large datasets in training models, need to be overcome. Experts suggest that small data models using one-shot learning is the key to enabling systems to reason (IBM, 2019). Models that require only a thin-layer of general-labelled data but classify new objects using their own prior knowledge (like a toddler categorizing an unfamiliar object as a cup), will eliminate the need for human supervision and will foster the commercial adoption of inductive AI.

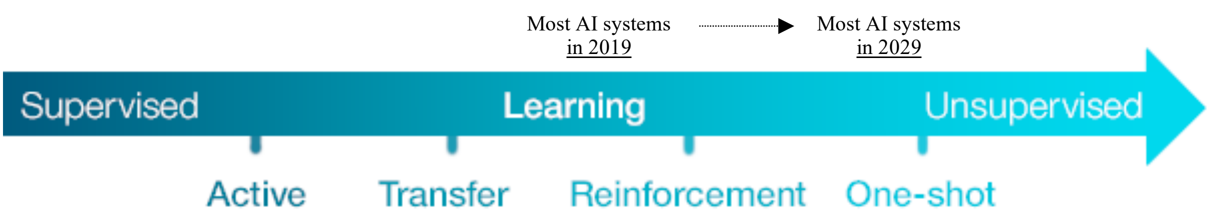


Fig. 5 Spectrum of AI learning IBM, 2019

Examples already exist, with deep reasoning being a priority in Tesla's software development. Unconventionally, Tesla do not program any explicit object detection in their vehicles, instead, they let the car learn on its own by observing human drivers (Marr, 2019). By end of 2019, all of Teslas will be capable of level 5 autonomous-driving (the highest), getting from A to B with no human intervention (Marr, 2019).

The development towards deep reasoning implies an increase in economic benefits from the technology in the coming years. It is predicted that by 2030 AI will contribute up to \$15.7 trillion to the global economy (PwC, 2017). Based on this and on the increasing trend of IP protection, it appears that AI is "crossing the chasm" (Fig.6) (Moore, 1999). The adoption of the technology is moving from early adopters ("visionaries"), such as Google, IBM, Microsoft, to early majority ("pragmatists), consumer-electronics companies, Toyota, etc. In the short-term it is expected that performance will go up with less effort (Fig. 1) as AI will require less human-interference thanks to one-stop learning. Additionally, as AI approaches its dominant design, adoption is also set to grow at an exponentially, diffusing into a wider range of industries beyond technology.

The key questions regarding the long-term forecast of AI are: will AI match human ability (high-level machine intelligence – HLMI), will machines outperform humans (singularity/superintelligence) and when will each occur? In a Delphi study by University of Oxford four sets of experts were asked to propose the year when HLMI will have a 90% probability of existing, and the likelihood of superintelligence occurring shortly (2 years after) or long (30 years) after HLMI.

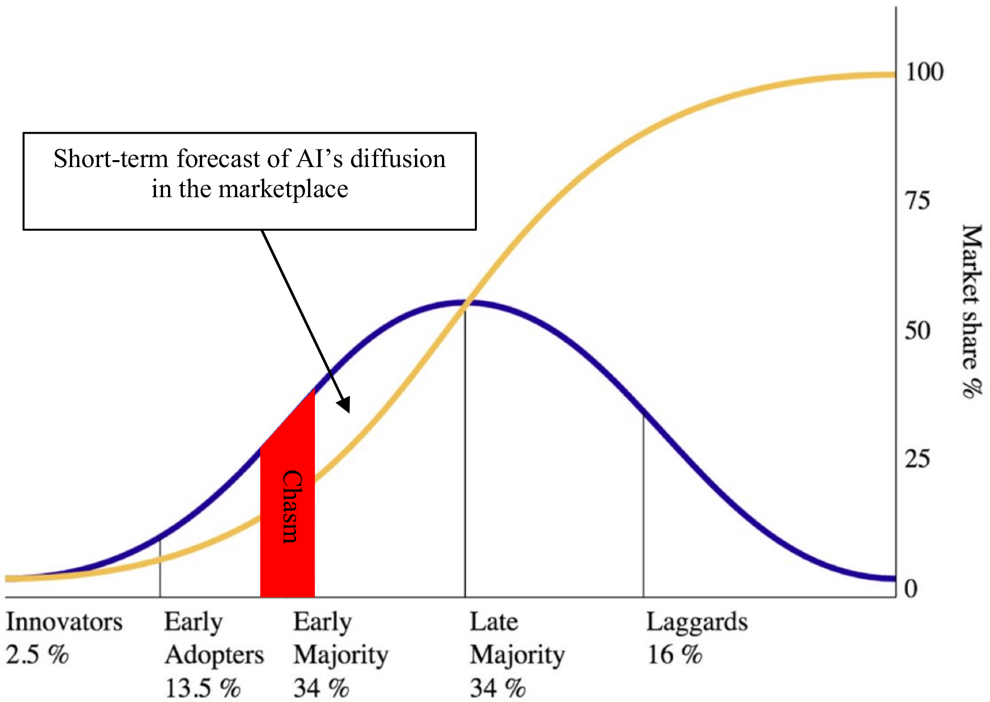


Fig. 6 AI's diffusion based on Bass' Model for technology adoption

The median prediction on HLMI existing was by 2075, and experts allocated a low probability for a fast take-off of superintelligence (10%), but a significant probability for it occurring within 30 years after HLMI (75%) (Bostrom and Muller, 2014). This rather positive forecast by experts on machines matching and subsequently outperforming humans within the next 100 years, should serve as a warning for the potential unpredicted outcomes that AI can have on businesses and society, once it enters the stage of diminishing returns.

Section F – Impact

The short-term impact of AI will not arise from AI creating new industries, but rather from AI empowering current employees to work more efficiently [See exhibit G]. From the predicted $15.7 trillion in global gains from AI, $6.6 trillion will come from increased productivity and $9.1 trillion from consumption side effects (PwC, 2017). Capital-intensive sectors such as manufacturing and transport are likely to see the largest productivity gains, considering that many of their operational processes are highly susceptible to automation (PwC, 2017).

The industries that will benefit from AI-consumption side effects the most, are automotive and healthcare. AI's impact on health care will come from more accurate, data-driven and personalized diagnoses. The technology will assist doctors in selecting the most appropriate treatment, resulting in higher efficiency for hospitals and more saved lives. In the automotive, automated driving assistance already reduces accidents and preserves fuel/energy. The average American spends over 300 hours yearly in driving (PwC, 2017). Autonomous driving systems have the potential of freeing up tremendous amounts of time for consumers and boosting demand for further AI solutions in automotive, such as predictive engine maintenance.

Although routine-job displacement in society will occur other jobs will be created. Staff will be required to maintain, operate and regulate emerging AI applications. For example, workers similar to air traffic controllers will be needed to control autonomous vehicle fleets.

The main social impact of AI in the long-run remains the fear of singularity. The experts from Oxford's Delphi study estimate a one in three chance that if superintelligence occurs it will turn out to be 'bad' or 'extremely bad' for humanity (Bostrom and Muller, 2014). Therefore, it is of highest priority that as the technology develops, we are able to predict, control and rationalize the decision-making of AI systems.

Bibliography:

Asgard Capital & Roland Berger GMBH (2018) *Artificial Intelligence – A strategy for European startups*. Munich: Roland Berger GMBH. Available at: https://asgard.vc/wp-content/uploads/2018/05/Artificial-Intelligence-Strategy-for-Europe-2018.pdf (Accessed: 22 April 2019).

Arthur, B. (2019) *Increasing Returns and the New World of Business*, *Harvard Business Review*. Available at: https://hbr.org/1996/07/increasing-returns-and-the-new-world-of-business (Accessed: 27 April 2019).

Anderson, P. and Tushman, M. (1991) "Managing through cycles of technological change", *Research Technology Management*, 8(4), pp. 26-31. doi: 10.1016/0737-6782(91)90063-5.

Cihon, P. (2019) *International Standards to Enable Global Coordination in AI Research & Development*. Oxford: University of Oxford (Standards for AI Governance). Available at: https://www.fhi.ox.ac.uk/wp-content/uploads/Standards_-FHI-Technical-Report.pdf (Accessed: 26 April 2019).

Dormehl, L. (2017) *Rise of the Machines: 10 Defining Moments in the History of AI | Digital Trends*, *Digital Trends*. Available at: https://www.digitaltrends.com/cool-tech/history-of-ai-milestones/ (Accessed: 26 April 2019).

Everitt, T & Hutter, M. (2018). *Universal Artificial Intelligence*. Foundations of Trusted Autonomy. Australian National University. pp.15-46. DOI: 10.1007/978-3-319-64816-3_2.

Foster, Richard (1986) *Working The S-Curve: Assessing Technological Threats*, Research Management, 29:4, 17-20, DOI: 10.1080/00345334.1986.11756976

Moore, G.E. (1975). "Progress in digital integrated electronics" (http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1478174). IEEE.

Moore, G. (1999), *Crossing the Chasm, Marketing and Selling High-Tech Products to Mainstream Customer (revised edition)*, HarperCollins Publishers, New York.

Godin, B. (2005) *The Linear Model of Innovation: The Historical Construction of an Analytical Framework*. Montreal: CSIIC Canada (Project on the History and Sociology of S&T Statistics), pp. 1-7. Available at: http://www.csiic.ca/PDF/Godin_30.pdf (Accessed: 26 April 2019).

IBM (2019) *The new innovation equation*, *IBM Cognitive - What's next for AI*. Available at: https://www.ibm.com/watson/advantage-reports/future-of-artificial-intelligence/ai-innovation-equation.html (Accessed: 26 April 2019).

Josimovic, A. (2018) *AI as a Radical or Incremental Technology Tool Innovation*. Master of Science Thesis. KTH Royal Institute of Technology Stockholm.

Krechmer, K. (2000) *Technical Perspective*. Boulder: University of Colorado (The Fundamental Nature of Standards).

Nilsson, N. (2009) *Quest artificial intelligence | Artificial intelligence and natural language processing*, *Cambridge University Press*. Available at: http://www.cambridge.org/us/0521122937 (Accessed: 26 April 2019).

McKinsey & Co (2017) *Jobs lost, jobs gained: What the future of work will mean for jobs, skills, and wages*, *McKinsey & Company*. Available at: https://www.mckinsey.com/featured-insights/future-of-work/jobs-lost-jobs-gained-what-the-future-of-work-will-mean-for-jobs-skills-and-wages (Accessed: 26 April 2019).

Marr, B. (2018) *The Amazing Ways Tesla Is Using Artificial Intelligence And Big Data*, *Forbes.com*. Available at: https://www.forbes.com/sites/bernardmarr/2018/01/08/the-amazing-ways-tesla-is-using-artificial-intelligence-and-big-data/#79b269b44270 (Accessed: 26 April 2019).

Muller, V. and Bostrom, N. (2019) *Future Progress in Artificial Intelligence: A Survey of Expert Opinion*. Oxford: University of Oxford. Available at: https://nickbostrom.com/papers/survey.pdf (Accessed: 26 April 2019).

PwC (2017) *Sizing the prize: What's the real value of AI for your business and how can you capitalise?*. PwC (Artificial Intelligence). Available at: https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html (Accessed: 26 April 2019).

PwC (2018) *2018 AI predictions 8 insights to shape business strategy*. PwC (Artificial Intelligence). Available at: https://www.pwc.com/us/en/advisory-services/assets/ai-predictions-2018-report.pdf (Accessed: 26 April 2019).

Schumpeter, J. (1942) "Capitalism, Socialism and Democracy.", *Economica*, 11(41), p. 40. doi: 10.2307/2549943.

Shanahan, M. (2015) *The technological singularity*. Cambridge, MA: MIT Press.

Turing, A. (1950), I.—COMPUTING MACHINERY AND INTELLIGENCE, *Mind*, Volume LIX, Issue 236, October 1950, Pages 433-460, https://doi.org/10.1093/mind/LIX.236.433

UBS (2018) *AI's coming of age*, *The Evolution of Artificial Intelligence*. Available at: https://www.ubs.com/microsites/artificial-intelligence/en/ai-coming-age.html (Accessed: 26 April 2019).

Utterback, J. and Abernathy, W. (1975) "A dynamic model of process and product innovation", *Omega*, 3(6), pp. 639-656. doi: 10.1016/0305-0483(75)90068-7.

WIPO (2019) *Artificial Intelligence*. Geneva: World Intellectual Property Organization (Technology Trends). Available at: https://www.wipo.int/edocs/pubdocs/en/wipo_pub_1055.pdf (Accessed: 26 April 2019).

Appendices:

Exhibit A – World Map of all companies specializing in AI solutions by Roger Berger and ASGARD

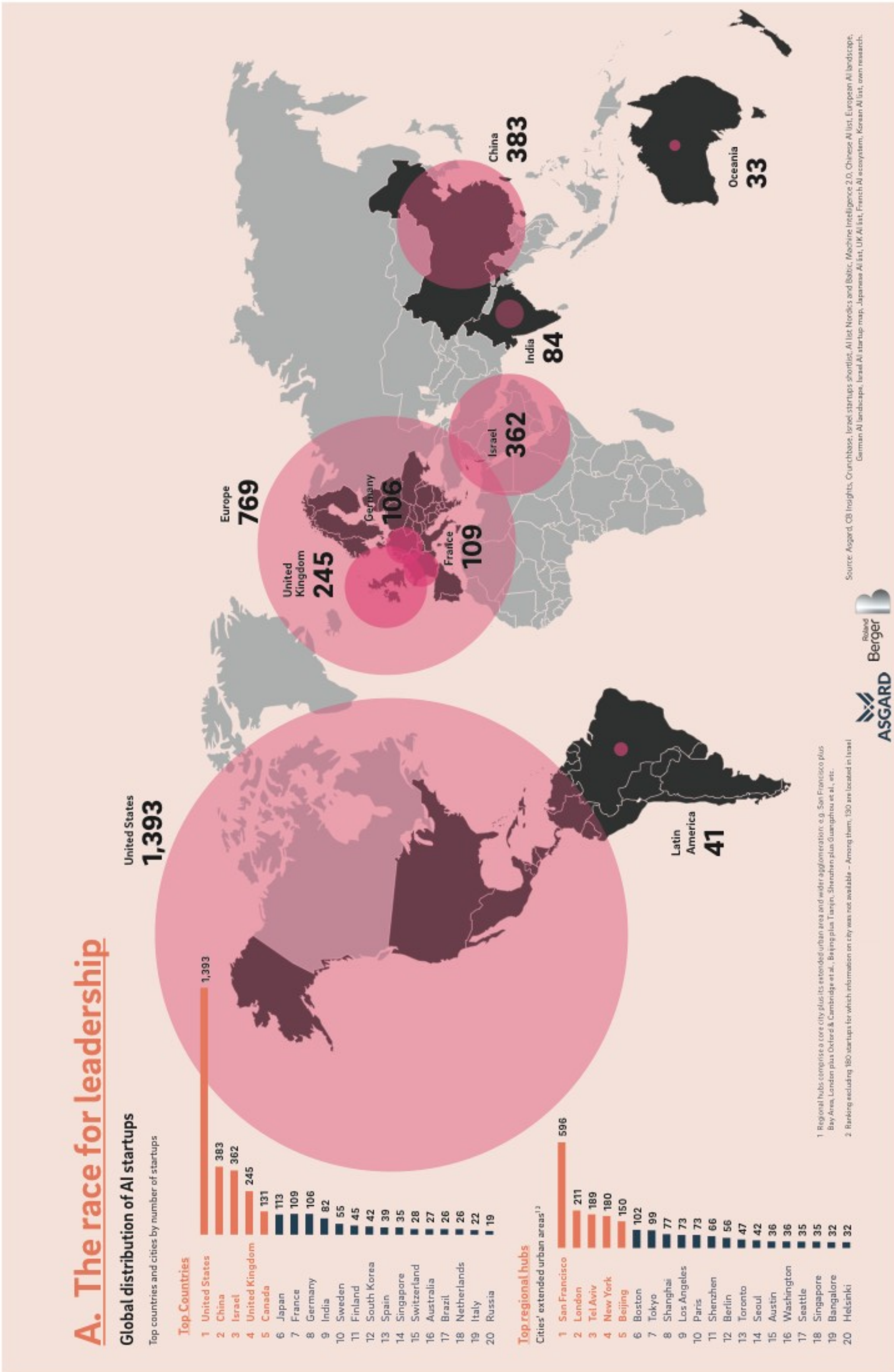


Exhibit B –Moore's Law (revised in 1975)

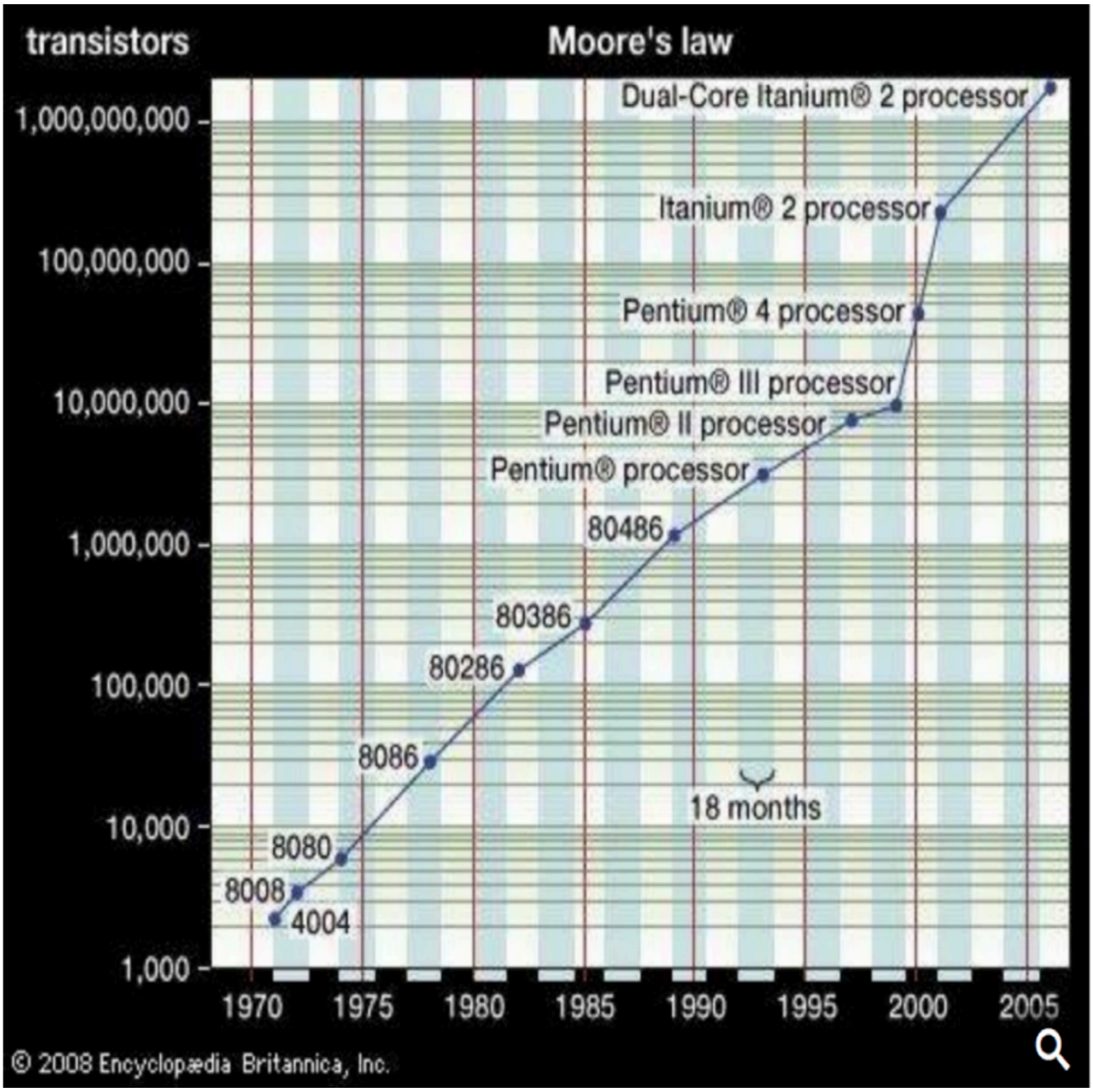
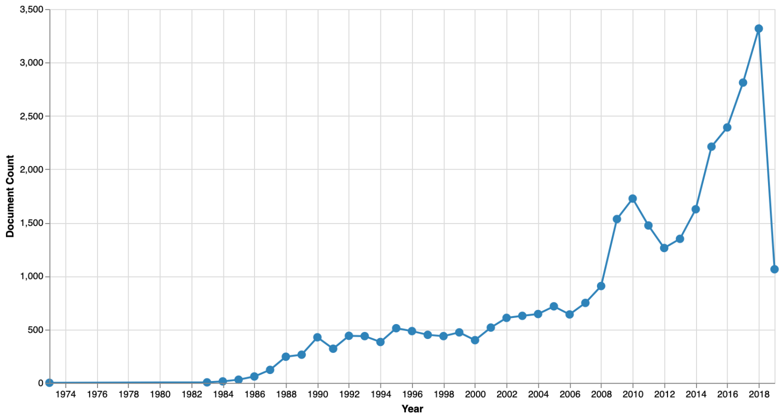


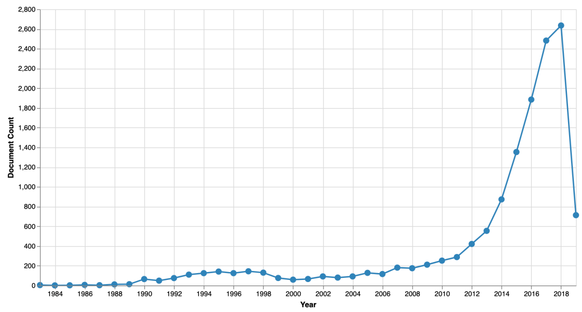
Exhibit C – Breaking down the patent family G06N
→ *computer systems based on specific computational models*

The subcategories of G06N are:

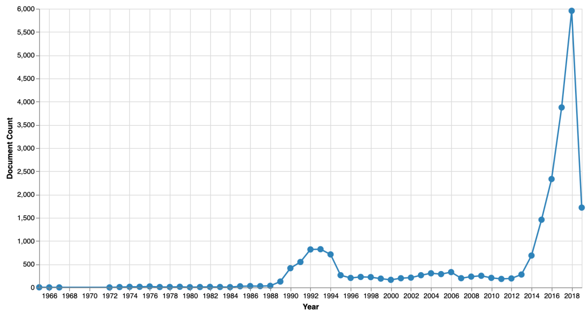
- **G06N3 (most applicable to AI)** → mentioned in text
- **G06N5 (most applicable to AI)** → Computer systems utilizing knowledge-based models



- G06N7 → Computer systems based on specific mathematical models

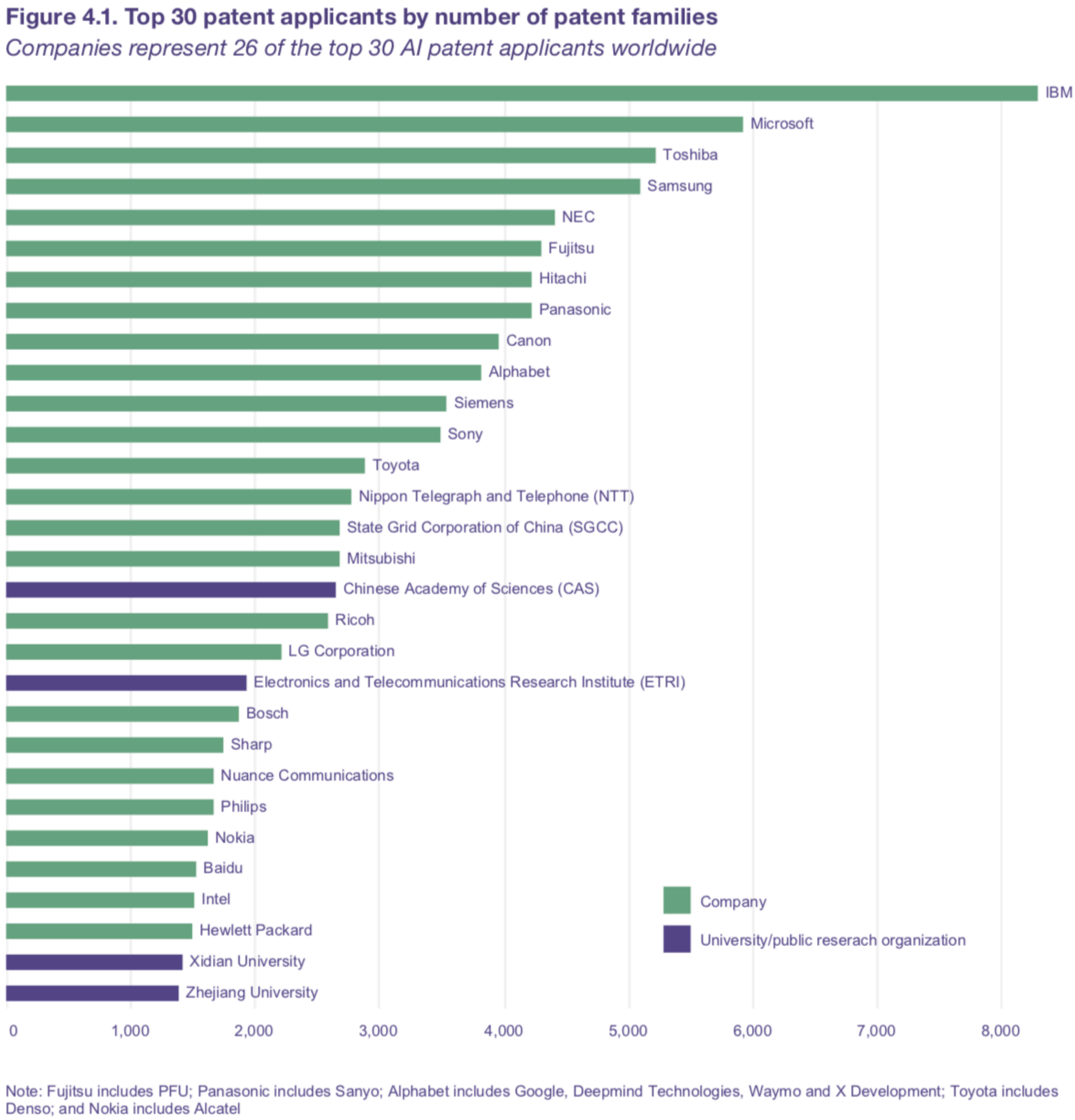


- G06N99 → Subject matter not provided for in other groups of this subclass



Note: Done to show that although G06N3 was chosen as representative, similar patterns occur across other, potentially relevant, AI-related patent classes

Exhibit D – Leaders in AI patenting, WIPO Technology Trends 2019



Note: The results from my own trial and error, individual research, from the sources WIPO, Eurospace and Lens were compared against those above, to confirm that there is a match in the class leaders confirming that the class was chosen appropriately.

Exhibit E – China's dominance in AI patenting through science/academia

Figure 4.2. Top patent applicants among universities and public research organizations in selected locations, by number of patent families

CAS (China) and ETRI (Republic of Korea) rank first and second in patent filings among universities and public research organizations

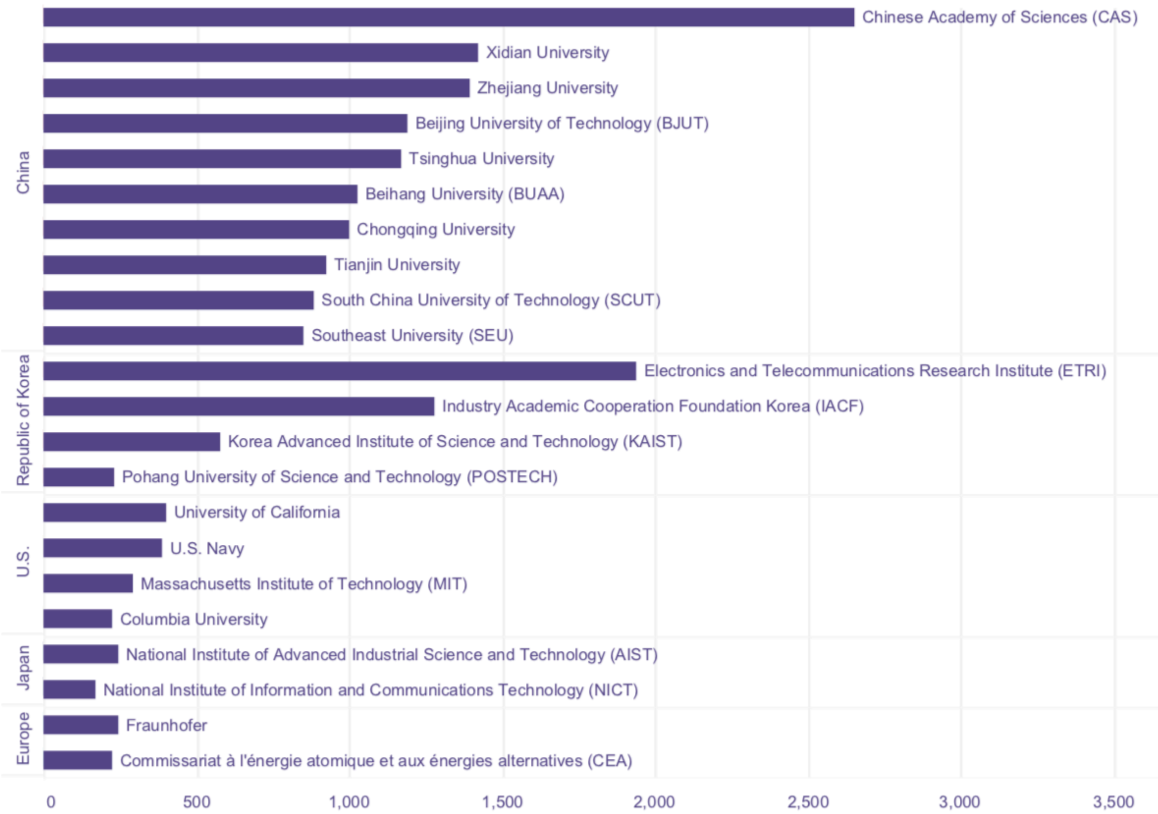


Figure 4.3. Geographical origin of universities and public research organizations in the top 500 patent applicants, by number of organizations

Chinese universities and public research organizations account for more than one-fifth of the top 500 patent applicants

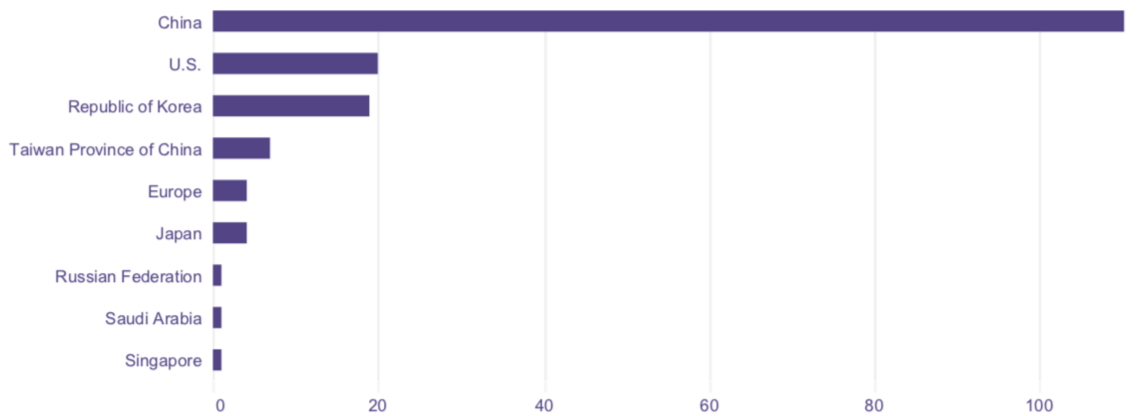


Exhibit F: Table of international AI standards currently under development

Etiquettes (Network Protocols) – Product	**Etiquettes (Network Protocols) – Process**
1. Foundational Standards: Concepts and terminology (SC 42 WD 22989) **2.** Framework for Artificial Intelligence Systems Using Machine Learning (SC 42 WD 23053) **3.** Transparency of Autonomous Systems (defining levels of transparency for measurement) (IEEE P7001) **4.** Personalized AI agent specification (IEEE P7006) **5.** Ontologies at different levels of abstraction for ethical design (IEEE P7007) **6.** Wellbeing metrics for ethical AI (IEEE P7010) **7.** Machine Readable Personal Privacy Terms (IEEE P7012) **8.** Benchmarking Accuracy of Facial Recognition systems (IEEE 7013)	**A.** Model Process for Addressing Ethical Concerns During System Design (IEEE P7000) **B.** Data Privacy Process (IEEE P7002) **C.** Methodologies to address algorithmic bias in the development of AI systems (IEEE P7003) **D.** Process of Identifying and Rating the Trustworthiness of News Sources (IEEE P7011)
Enforcement Mechanisms – Product	**Enforcement Mechanisms – Process**
9. Certification for products and services in transparency, accountability, and algorithmic bias in systems (IEEE ECPAIS) **10.** Fail-safe design for AI systems (IEEE P7009)	**E.** Certification framework for child/student data governance (IEEE P7004) **F.** Certification framework for employer data governance procedures based on GDPR (IEEE P7005) **G.** Ethically Driven AI Nudging methodologies (IEEE P7008)

Exhibit G: Lists with ongoing work for the two major standards organizations for AI

ISO/IEC JTC 1 (SC 42):

- **Working Group 1: Foundational Standards.** WG1 has 2 standards working drafts:
    - WD 22989: Artificial intelligence -- Concepts and terminology
    - WD 23053: Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)
    - Both of the above appear to be initial definitional standards. No additional documentation is publically available.
- **Working Group 2: Big Data.** WG2 incorporates previously ongoing efforts that were assigned to SC 42 at its inception.
    - ISO/IEC 20546: Information technology — Big data — Overview and vocabulary
    - ISO/IEC TR 20547-1: Information technology — Big data reference architecture — Part 1: Framework and application process
    - ISO/IEC TR 20547-1: Information technology — Big data reference architecture — Part 2: Use cases and derived requirements (Published)
    - ISO/IEC DIS 20547-3: Information technology — Big data reference architecture — Part 3: Reference architecture
    - ISO/IEC DIS 20547-4 Part 4 is managed by JTC 1 SC 27 IT Security techniques
    - ISO/IEC DIS 20547-3: Information technology — Big data reference architecture — Part 5: Reference architecture (Published)
- **Working Group 3: Trustworthiness.** WG3 is not currently drafting standards, but is pursuing three technical reports (TR):
    - TR on Bias in AI systems and AI aided decision making
    - TR on Overview of trustworthiness in Artificial Intelligence
    - TR on Assessment of the robustness of neural networks – Part 1: Overview ●
- Working Group 4: Use Cases and Applications. WG4 is not currently drafting standards, but is pursuing one TR:
    - TR on Artificial Intelligence: use cases

<u>IEEE:</u>

- **P7000:** <u>Model Process for Addressing Ethical Concerns During System Design</u>
    - Creates a process model for ethics considerations across development stages.
- **P7001**: <u>Transparency of Autonomous Systems</u>
    - Defines levels of measurement for transparency during system development.
- **P7002:** <u>Data Privacy Process</u>
    - Establishes privacy process management standard to enable conformity assessments.
- **P7003**: <u>Algorithmic Bias Consideration</u>
    - Creates a certification framework of methods addressing negative bias in algorithms.
- **P7004**: <u>Child and Student Data Governance</u>
    - Defines a certification framework of methodologies for access, collection, use, storage, sharing, and destruction of child and student data.
- **P7005**: <u>Employer Data Governance</u>
    - Establishes a certification framework of methodologies for access, collection, use, storage, sharing, and destruction of employee data.
- **P7006**: <u>Personal Data AI Agent Working Group</u>
    - "[D]escribes the technical elements required to create and grant access to a personalized Artificial Intelligence (AI) that will comprise inputs, learning, ethics, rules and values controlled by individuals."
- **P7007**: <u>Ontological Standard for Ethically driven Robotics and Automation Systems</u>
    - Establishes ontologies for ethical design considerations at different levels of abstraction.
- **P7008**: <u>Ethically Driven Nudging for Robotic, Intelligent and Autonomous Systems</u>
    - Defines common behavior nudges and ethical methodologies for their design.
- **P7009**: <u>Fail-Safe Design of Autonomous and Semi-Autonomous Systems</u>
    - Creates technical baseline of methods for the design of fail-safe mechanisms.
- **P7010**: <u>Wellbeing Metrics Standard for Ethical AI and Autonomous Systems</u>
    - Establishes metrics for measuring human well-being impacted by systems as well as a related baseline for measurement data.
- **P7011**: <u>Process of Identifying & Rating the Trustworthiness of News Sources</u>
    - Provides semi-autonomous processes for rating factual accuracy of news.
- **P7012**: <u>Machine Readable Personal Privacy Terms</u>
    - Provides means for individuals to proffer their privacy terms so that they can be machine read by other entities.
- **P7013**: <u>Benchmarking Accuracy, Increasing Transparency, and Governing Use of Automated Facial Analysis Technology</u>
    - Establishes demographic definitions and reporting protocols for assessing system performance.
- **ECPAIS**: <u>Ethics Certification Program for Autonomous and Intelligent Systems</u>
    - Certification methodologies for transparency, accountability, and algorithmic bias. Open to IEEE SA Advanced Corporate Members only.

Exhibit H: Impact of AI by industry

Where industries will put practical AI to work

Ranking of AI impact by its potential to free up time, enhance quality, and enhance personalization

Ranking	Industry	High-potential use cases
1	Healthcare	• Supporting diagnosis by detecting variations in patient data • Early identification of potential pandemics • Imaging diagnostics
1	Automotive	• Autonomous fleets for ride sharing • Semi-autonomous features such as driver assist • Engine monitoring and predictive, autonomous maintenance
3	Financial services	• Personalized financial planning • Fraud detection and anti-money laundering • Automation of customer operations
4	Transportation and logistics	• Autonomous trucking and delivery • Traffic control and reduced congestion • Enhanced security
5	Technology, media, and telecommunications	• Media archiving, search, and recommendations • Customized content creation • Personalized marketing and advertising
6	Retail and consumer	• Personalized design and production • Anticipating customer demand • Inventory and delivery management
7	Energy	• Smart metering • More efficient grid operation and storage • Predictive infrastructure maintenance
8	Manufacturing	• Enhanced monitoring and auto-correction of processes • Supply chain and production optimization • On-demand production

Source: PwC Global AI Impact Index, 2017

Figure 1: Where will the value gains come from with AI?