Economists have been studying the relationship between technological change, productivity and employment since the beginning of the discipline with Adam Smith’s pin factory. It should therefore not come as a surprise that AI systems able to behave appropriately in a growing number of situations - from driving cars to detecting tumours in medical scans - have caught their attention.

In September 2017, a group of distinguished economists gathered in Toronto to [set out a research agenda for the Economics of Artificial Intelligence (AI)](https://www.economicsofai.com/nber-conference-toronto-2017). They covered questions such as what is economically unique about AI, what will be its impacts, and what are the right policies to enhance its benefits.

I recently had the privilege of attending the [third edition of this conference](http://conference.nber.org/sched/AIf19?show_participants=1) in Toronto, and to witness first-hand how this agenda has evolved in the last two years. In this blog I outline key themes of the conference and relevant papers at four levels: macro, meso (industrial structure), micro and meta (impacts of AI on the data and methods that economics use to study AI). I then outline some gaps in today's Economics of AI agenda that I believe should be addressed in the future, and conclude.

**Prelude: an economist’s take on AI**

Ajay Agrawal, Joshua Gans and Avi Goldfarb, the convenors of the conference (together with Catherine Tucker), have[in previous work](http://ide.mit.edu/sites/default/files/publications/w24626.pdf) described AI systems as [‘prediction machines’](https://www.predictionmachines.ai/) that make predictions cheap and abundant, enabling organisations to make more and better decisions, and even automating some of them. One example of this is Amazon’s recommendation engine, which presents a personalised version of its website to each visitor. That kind of customisation would not be possible without a machine learning system (a type of AI) that predicts automatically what products might be of interest to individual customer based on historical data.

AI systems can in principle be adopted by any sector facing a prediction problem - which is almost anywhere in the economy from agriculture to finance. This widespread relevance has led some economists to herald AI as the latest example of a transformational ‘General Purpose Technology’ that will reshape the economy like the steam engine or the semiconductor did earlier in history.

**Macro view: AI, Labour and (intangible) capital**

AI automates and augments decisions in the economy, and this increases productivity. What are the implications for labour and investment?

**Who - or what - does what: The task-based model**

The dominant framework to analyse the impact of AI on labour is the[task-based model developed by Daron Acemoglu and Pascual Restrepo](https://www.nber.org/papers/w24196) (building on previous work by [Joseph Zeira](https://academic.oup.com/qje/article-abstract/113/4/1091/1916985)). This model conceives the economy as a big collection of productive tasks. The arrival of AI changes the value and importance of these tasks, impacting on labour demand and other important macroeconomic variables such as the share of income that goes to labour, and inequality (for example, if AI de-skills labour or increases the share of income going to capital - which tends to be concentrated in fewer hands - this is likely to increase income and wealth inequality).

The impact of AI on tasks takes place through four channels:

1. First, there is displacement, when an AI system replaces some of the tasks that were previously performed by human workers. An example of this would be the book reviews that were displaced when Amazon adopted its automatic recommender (and [laid off its book reviewers](https://www.theguardian.com/technology/2016/sep/30/age-of-algorithm-human-gatekeeper), although [some have now made it back](https://www.seattletimes.com/business/amazon/how-amazons-team-of-old-school-book-reviewers-influences-what-we-read/) to the company). This will reduce the demand for labour
2. Second, there is augmentation when an AI system increases the value of the tasks undertaken by human workers. An example of this would be Amazon’s web development and inventory management tasks: each dollar spent on improving its website and ensuring that many different titles are efficiently stocked creates a bigger return for the company thanks to its AI recommendation system. This will in general increase the demand for workers whose tasks are augmented.
3. Third, there is capital deepening. New AI systems are an investment that increases the stock of capital that workers use, making them more productive and increasing demand for labour through the same mechanism as above.
4. Finally, there is reinstatement, when the AI system creates completely new tasks such as developing machine learning systems or labelling datasets to train those systems. These new tasks will create new jobs and even industries, increasing labour demand.

Considered together, these four channels determine the impact of AI on labour demand. Contrary to the idea of the impending job apocalypse, this model identifies several channels through which AI systems that increase labour productivity could also increase demand for labour. Contrary to previous assumptions by economists that new technology always increases labour demand through augmentation, the task-based model recognises that the net effect of new technology on labour demand could be negative. This could, for example, happen if firms adopt ‘mediocre’ AI systems that are productive enough to displace workers, but not productive enough to increase labour demand through other channels.

Several papers presented in the conference built on these themes:

* [Jackson and Kanik](http://conference.nber.org/conf_papers/f130023.pdf) model AI as an intermediate input that firms acquire through their supply chain using services such as [Amazon’s Web Services](https://aws.amazon.com/machine-learning/ai-services/). In this model, the impact of AI on labour demand and productivity depends on the outside options of those workers who are displaced by AI: if alternative jobs have low productivity, then this will have an (indirectly) negative impact on productivity. This means that the impacts of AI depend not only on what happens in AI adopting sectors but also in the situation elsewhere in the economy. Another interesting conclusion of their analysis is that AI deployment makes the economy more interconnected as companies start using AI suppliers to source services previously performed by workers. This could centralise value chains, increasing market power and creating systemic risks.
* [Autor and Salomons](http://conference.nber.org/conf_papers/f129906.pdf) study the evolution of the industries and occupations that create new job titles (a proxy for new tasks) using a dictionary of job titles published by the US Census since the 1950s. Their analysis shows important changes between that time, when occupations towards the middle of the income distribution (‘middle class jobs’) created most new job titles, and today, when most of the new job titles are created in either highly skilled, technology intensive occupations (eg software development) or less skilled personal services occupations (eg personal trainers). It seems that modern technologies like AI are enabling the creation of new tasks that increase demand for high skilled jobs that complement AI and low-skill jobs that are difficult to replace with AI, leading to polarisation in the labour market. This result underscores the risk that skills shortages in highly-skilled occupations (which hinder productivity growth) might coexist with unemployment amongst individuals lacking the skills to transition into those occupations.

**Automation without capital: Intangible investments**

In order to increase productivity, investments in AI needs to be accompanied by complementary investments in IT infrastructure, skills and business processes. Some of these investments involve the accumulation of ‘intangibles’ such as data, information and knowledge. In contrast to tangible assets like machines or buildings, intangibles are hard to protect, imitate and value, and their creation often involves lengthy and uncertain processes of experimentation and learning-by-doing (much more on this subject [here](https://www.amazon.co.uk/Capitalism-Without-Capital-Intangible-Economy/dp/0691175039)).

Continuing with the example of Amazon, over its history the company has built a tangible data and IT infrastructure complementing its AI systems. At the same time, it has developed processes, practices and a mindset of ‘customer-centrism’ and ‘open interfaces’ between its information systems and those of its vendors and users of its cloud computing services [which could be equally important for its success](https://zackkanter.com/2019/03/13/what-is-amazon/), but very hard to imitate.

According to a [2018 paper](https://www.nber.org/papers/w25148) by Erik Brynjolfsson and colleagues, the need to accumulate these intangibles across the economy explains why advances in AI are taking so long to materialise in productivity growth or drastic changes in labour demand.

Several papers presented in Toronto this year explored these questions empirically:

* [Daniel Rock](http://conference.nber.org/sched/AIf19?show_participants=1) uses LinkedIn skills data to measure the impact of engineering skills on firm value. He finds that after controlling for unobservable firm factors, the impact of those skills on value dissipate, suggesting that intangible firm factors determine how much value a firm is able to create from its engineering talent. His analysis also shows that the market expects these intangible investments to create value in the future: when Google released TensorFlow, an AI software programme, those firms already employing AI talent saw their market value increase. This is consistent with the idea that Google’s strategy was perceived to increase the supply of AI labour (through a boost on its productivity) complementing those firms’ intangible AI-related investments. Interestingly, similar increases in value were not visible in firms whose workforces were at risk of automation. One interpretation is that they are expected to be disrupted by those firms developing AI systems and services.
* [Prasanna Tambe](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3416289) and co-authors also use LinkedIn skills data to estimate the value of intangible investments related to IT, finding that it is concentrated in a small group of ‘superstar firms’, and that it is associated with higher market value, suggesting that these investments are expected to generate important returns in the future. An important implication of this analysis is that the market expects the benefits from AI to be concentrated in a small number of firms, raising concerns about market power in tomorrow's AI-driven economy.

**Meso view: sectoral differences in AI adoption and impacts**

Think of a sector like health: the nature of the tasks undertaken in this industry, as well as the availability of data, the scope for changes in business processes and its industrial structure (including levels of competition and entrepreneurship) are completely different from, say, finance or advertising. This means that its rate of AI adoption and its impact will be very different from what happens in those industries. Previous editions of the Economics of AI conference included papers about the impact of AI in sectors such as [media](http://conference.nber.org/conf_papers/f114672.slides.pdf) or healthcare.

This year considered sector-specific issues in several areas, including R&D and regulation

**Sending machines to look for good ideas**

In the inaugural Economics of AI conference, [Cockburn, Henderson and Stern](https://www.nber.org/chapters/c14006) proposed that AI is not just a General Purpose Technology, but also an ‘invention in the methods of invention’ that could greatly improve the productivity of scientific R&D, generating important spillovers for the sectors using that knowledge. One could even argue that the idea of the [Singularity](https://en.wikipedia.org/wiki/Technological_singularity) is an extreme case of this model where “AI systems that create better ideas” become better at creating “AI systems that create better ideas” in a recursive loop that could lead to exponential growth.

This year, venture capitalist Steve Jurvetson and Abraham Heifts, CEO of [Atomwise](https://www.atomwise.com/), a startup that uses AI in drug discovery spoke about how they are already pursuing some of these opportunities in their ventures, and two papers investigated the impact of AI on R&D:

* Our analysis[of the deployment of AI in computer science research in arXiv](https://arxiv.org/abs/1808.06355), a pre-prints website popular with the AI community, supports the idea that AI is an invention in the methods of development: it has experienced rapid growth in absolute and relative terms, it is being adopted in many computer science subfields, and it is already creating important impacts (measured with citations) wherever it is adopted. AI is being adopted faster in fields such as computer vision, natural language processing, sound processing and information retrieval where there are big datasets to train machine learning algorithms, highlighting how AI R&D advances faster in those areas with complementary datasets.
* Agrawal and co-authors develop a [formal model of the impact of AI on the R&D process in scientific fields such as bio-medical and materials science](http://conference.nber.org/conf_papers/f129947.pdf), where innovation often involves finding useful needles in big data haystacks. An example of this is identifying which, among the millions of potential folds in a protein could be targeted by a pharmaceutical drug. AI systems trained on labelled data about previous successes and failures could help identify which of these combinations have the greatest potential, reducing waste and reviving [sluggish productivity growth in R&D](https://www.nber.org/papers/w23782.pdf). Realising these benefits will require access to training data and building research teams that bring together AI skills with domain knowledge about the scientific fields where AI is being adopted.

**Sending bounty hunters to keep an eye on the machines**

Regulation sets the context and rules of the game that shape the rate and direction of new technologies such as AI. At the same time, regulation is itself an industry whose structure and processes are being transformed by AI systems that accelerate the pace and breadth of change, and create new opportunities to monitor economic activity. Two talks at the conference focused on this two-way street between regulation and AI.

* [Suk Lee and co-authors](http://conference.nber.org/conf_papers/f129984.pdf) have surveyed businesses about how they would change their AI adoption plans in response to different regulatory models. They show that general-purpose regulations would create more barriers to AI adoption than sector-specific regulations, and that regulation increases demand for managers to oversee AI adoption while reducing demand for technical and lower-skilled workers. It also creates bigger barriers for smaller firms, highlighting the trade-offs between AI regulation, innovation and competition.
* [Clark and Hadfield](http://conference.nber.org/conf_papers/f130053.pdf) argue that the regulatory industry needs innovation to keep up with the fast pace of change in AI technologies, but public sector regulators lack flexibility and incentives to do this effectively. In order to remove this bottleneck, they propose the creation of regulatory markets where government licenses private sector companies to regulate AI adoption to achieve measurable outcomes (for example to lower AI error rates and accidents below an agreed threshold): this would give private sector firms the incentives and freedom to develop innovative regulatory technologies and business models, although it also raises the question of who would regulate these new regulators, and how to avoid their capture by the industries they are meant to regulate.

**Micro view: Inside the AI adoption black box**

Modern AI systems based on machine learning algorithms that detect patterns in data are often referred to as black boxes because the rationale for their predictions is difficult to explain and understand. Similarly, the firms adopting AI systems look like black boxes to economists adopting a macro view: AI intangibles are after all a broad category of business investments including experiments with various processes, practices and new businesses and organisational models. But what are these firms actually doing when they adopt an AI system, and what are the impacts?

Several papers presented at the conference illustrated how economists are starting to open these organisational black boxes to measure the impact of AI and how it compares with the status quo. As they do this, they are also incorporating into the Economics of AI some of the complex factors that come into play when firms deploy AI systems that do not simply increase the supply of predictions, but also reshape the information environment where other actors (employees, consumers, competitors, the AI systems themselves) make decisions, leading to strategic behaviours and unintended consequences that are generally abstracted in the macro perspective.

* Susan Athey and co-authors [compare the service quality of UberX and UberTaxi rides](http://conference.nber.org/conf_papers/f130190.pdf) in Chicago. Their hypothesis that UberX drivers whose job depends on user reviews will provide higher quality rides is confirmed with an analysis of granular telematic data about driving speed and duration, number of hard brakes etc. They also test whether giving drivers information about their performance changes their behaviour, finding that the worst performers tend to improve their driving in response to these nudges. The paper shows that AI systems are an ‘invention in the methods for managing and regulating increasingly important digital platforms and marketplaces’, while also raising concerns about worker privacy and manipulation.
* Michael Luca and co-authors (no link to the paper) test the effectiveness of various systems to select what Boston restaurants should be targeted with health inspections. They show that recommendations from a complex machine learning algorithm outperforms the rankings generated by human inspectors. Interestingly, they also detect high levels of inspector non-compliance with AI recommendations, suggesting that organisations using AI to inform their employees' decisions will have to overcome worker reticence and mistrust of these systems
* [Adair Morse and co-authors analyse the impact of ‘fintech’ AI systems in consumer-lending discrimination](http://conference.nber.org/conf_papers/f130191.pdf), finding that these systems tend to reduce - although not eliminate - discrimination against Latinx and African-American borrowers compared with face-to-face lenders, both in terms of the interest rates charged and the loan approval rates. AI systems still discriminate by identifying proxies for protected characteristics in the data. This shows how the adoption of AI can reduce old problems (human prejudice) while introducing new ones (algorithmic bias).

**Meta view: Using AI to research AI**

AI techniques have much to contribute to economics studies that often seek to detect (causal) patterns in data. [Susan Athey surveyed these opportunities In the inaugural Economics of AI conference](https://www.nber.org/chapters/c14009), with a particular focus on how machine learning can be used to enhance existing econometric methods.

Several papers presented in this year’s conference explored new data sources and methods along these lines, for example using big datasets from LinkedIn and Uber to respectively measure technology adoption and service quality in car rides, and online experiments to test how UberX drivers react to informational nudges. In Nesta, we are analysing open datasets with machine learning methods to map AI research.

Although these methods open up new analytical opportunities, they also raise challenges around reproducibility, particularly when the research relies on proprietary datasets that cannot be shared with other researchers (with the added risk of publication bias if data owners are able to control what findings are released), and ethics, for example around consent for participation in online experiments. In our own work we seek to address these issues by being very transparent with our analyses (for example, the data and code that we used in our AI mapping analysis is available in [GitHub](https://github.com/nestauk/arxiv_ai)) and developing [ethical guidelines](https://www.nesta.org.uk/blog/red-lines-grey-area/) to inform our data science work.

**Prospective view: future avenues for the Economics of AI**

Having summarised key themes and papers from the conference, I focus on some questions that I felt were missing.

**Recognising that to err is algorithm**

Macro studies of the impact of AI assume AI will increase productivity as long as businesses undertake the necessary complementary investments. They pay little attention to new issues created by AI such as algorithmic manipulation, bias and error, worker non-compliance with AI recommendations, or information asymmetries in AI markets, some of which are already being considered in micro studies from the trenches of AI adoption. These factors could reduce AI’s impact on productivity (making it mediocre and therefore predominantly labour displacing), increase the need to invest in new complements such as [AI supervision and moderation](https://osf.io/xuvf9/download/?format=pdf), hinder trade in [AI 'lemons'](https://en.wikipedia.org/wiki/The_Market_for_Lemons) and have important distributional implications, for example through algorithmic discrimination of vulnerable groups. Macro research on AI should start to consider explicitly these complex aspects of AI adoption and impact, rather than hiding them in the black box of AI-complementing intangible investments and/or assuming that they are somehow exogenous to AI deployment. As an example, [in previous work](https://osf.io/xuvf9/download/?format=pdf) I started to sketch what such a model could look like if we take into account the risk of algorithmic error in different industries, and the investments in human supervision required to manage it.

**Modelling AI progress**

In general, the research presented at the Economics of AI conference modelled AI as an exogenous shock to the economy, in some cases explicitly as with Daniel Rock’s study of the impact of TensorFlow's release on firms' value. Yet AI progress is, itself, an economic process whose analysis should be part of the Economics of AI agenda.

In his conference dinner speech, [Jack Clark from OpenAI](https://jack-clark.net/) described key trends in AI R&D: we are witnessing an ‘industrialisation of AI’ as corporate labs, big datasets and large scale IT infrastructures become more important in AI research, and at the same time a ‘democratisation of AI’ as open source software, open data and cloud computing services make it easier to recombine and deploy state of the art AI systems. These changes have important implications for AI adoption and impacts. For example, the fact that researchers in academia increasingly need to collaborate with the private sector in order to access the computational infrastructure required to train state of the art AI systems could diminish spillovers from this research. Meanwhile, the diffusion of AI research through open channels creates significant challenges for regulators who need to monitor compliance in a context where adopting 'dual use' AI technologies is as simple as downloading and installing some software from a coding repository like GitHub. Few if any of the papers presented at the conference addressed these topics.

Future work could fill these gaps by developing formal models of AI progress through an AI production function that takes inputs such as data, software, computational infrastructure and skilled labour to produce AI systems with a level of performance. [In this paper](https://pdfs.semanticscholar.org/15b2/0c32206289f911711ff019158c6c5ee3b890.pdf), Miles Brundage started outlining qualitatively what that model could look like. This model could be operationalised using data from open and web sources and initiatives to [measure AI progress from EEF](https://www.eff.org/ai/metrics) and the [Papers with Code](https://paperswithcode.com/) project in order to study the structure, composition and productivity of the AI industry, and its supply of AI technologies and knowledge to other sectors. [Recent work by Felten, Raj and Seamans where they use the EFF indicators](https://journals.aom.org/doi/pdf/10.5465/AMBPP.2019.140) to link advances in AI technologies with jobs at risk of automation illustrates how this kind of analysis could help forecast the economic impacts of AI progress and inform policy.

**Studying the direction of AI inventive activity**

Perhaps unsurprisingly given the point above, most of the research presented at the conference adopted a 'monolithic' definition of AI that equates it with the deep learning techniques currently dominating the field. This neglects concerns about the lack of robustness, explainability, data efficiency and environmental sustainability of deep learning algorithms, and the fact that alternative AI research and technological trajectories could be feasible and perhaps desirable. However, as Daron Acemoglu [showed some time ago](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3359482##), the market will undersupply alternatives to a dominant technology if researchers are not able to capture the benefits of sustaining technological diversity. Acemoglu pointed out that maintaining diversity in researchers’ capabilities and beliefs and providing public funding for less commercially oriented alternatives are two potential strategies to bring levels of technological diversity closer to what is socially optimal.

Could lack of technological diversity become a problem in the AI field? [Lack of diversity in the AI research workforce](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3428240), and the increasing influence of the private sector in AI research agendas through the aforementioned industrialisation of AI research give reason for concern but the evidence base is lacking. More research is needed to measure AI's technological diversity and how it is shaped by the goals, preferences and agendas of the scientists, engineers and organisations involved in it. This is an active area of research at Nesta where we will be publishing some findings soon.

**Remembering the political economy of AI**

In the inaugural Economics of AI conference, [Tratjenberg](https://www.nber.org/papers/w24245) and [Korinek and Stiglitz](https://www8.gsb.columbia.edu/faculty/jstiglitz/sites/jstiglitz/files/AI_labor.pdf) asked who will benefit and who will suffer when AI arrives, whether AI deployment could become politically unacceptable, and what policies should be put in place to minimise the societal costs of AI when it is deployed. More recently, Daron Acemoglu and Pascual Restrepo expressed concerns that the AI industry might be building ‘[the wrong kind of AI’](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3359482) because it does not internalise negative externalities from AI deployment (eg. labour market disruption) and because some of its leaders are biased in favour of mass automation regardless of its risks. These important questions were largely absent from the debate in Toronto, yet economists need to formalise and operationalise models of the distributional impacts of AI and its externalities in order to inform policies to ensure that its economic benefits are widely shared and reduce the risk of a public backlash against AI.

**Conclusion: Think Internet, not Skynet**

For me, the biggest takeaway from this year's Economics of AI conference was that AI impacts will be more complex and take longer to appear than what some newspaper headlines might lead us to expect: jobs will evolve and adapt in response to AI systems rather than disappearing completely. Firms will carry out organisational experiments to discover what applications of AI create most value. In some cases the experiments will fail, or prove that the adoption of AI in certain circumstances is uneconomical. Some firms will learn from these failures and others will try again. Skills shortages, regulation and consumer perceptions will slow down the adoption of certain AI systems (in some cases for good, in some cases for ill). The adoption of AI in an industry or firm will create sub-industries intent on manipulating it and lead to unexpected outcomes, and new changes in response.

In other words, the future of AI in the economy will resemble the Internet more than Skynet, the AI system that decides to destroy humanity (and its economy) in The Terminator: it will be complicated. Prediction machines not only increase the amount of decisions *we are able to make* based on AI recommendations, but also the amount of decisions that *we need to make*, as participants in the economy and as a society, about what AI technologies to develop, where to adopt them and how, and how to manage their impacts. It is a very good thing that economists are doing research to inform these decisions and discussing their findings close to real-time in fora such as the Economics of AI conference. I look forward to seeing how this research agenda evolves in the future (also with Nesta's contributions), and how its findings inform AI policies to ensure that this powerful technology is deployed for the benefit of more people.