

COGNITIVE AUTOMATION: WHAT ARE THE EU AND MEMBER STATES DOING TO PROTECT WORKERS?

The growth in the use of AI-based systems in workplaces is a development of immense importance for social policy. On the one hand, AI-based systems present profound opportunities for progressive change. They can reduce waste and bureaucracy, augment critical research, support limited government budgets, and make numerous other socially beneficial contributions. On the other hand, the application of such systems introduces several concerns that must be addressed by regulators and legislators. **Privacy violations, lack of transparency and accountability, intrinsic bias, and depersonalisation** of social relations are just a few examples of problems that are likely to be associated with the widespread implementation of this technology.

National governments have, in recent years, started to publish policy initiatives outlining how they intend to enhance progressive possibilities and minimise objectionable aspects of AI. These policies differ significantly in scope and central focus, spanning from detailed implementation strategy in both civilian and work-related contexts to more general intentions for future development. While these policies provide insight into the plans, strategies and areas of interest for AI development amongst national governments, they often contain few concrete strategies centred on occupational safety and health (OSH). Yet, they still collectively provide insight into future AI development in different parts of Europe as well as shared intentions and values between member states.

This policy brief presents several national initiatives to provide insights into how governments are approaching the proliferation of AI-based systems in workplaces. Specifically, this policy brief considers documents from the governments of Germany, France, Sweden, and Norway, since the strategies set by those countries, deal with workers in more detail than the others surveyed and summarise regulations identified by the EU Commission. We highlight how initiatives conceptualise what is at stake for workers and what should be done about it. A reoccurring topic is the need for continuous education or re-education and new skill acquisition in the workforce to successfully work with AI-based systems. The policies outlined below tend to emphasise the necessity in acknowledging the impact that AI will have on employment, both in the way of handling job loss and mitigating the rising fear of job loss in the workforce. The responsible use of human data is also stated as a priority in several countries. Several policies explicitly state the intention for specialised committees or comparable institutions to deal with the ethical implication of AI, both in public and work-related environments. Acknowledging that AI-based systems will continue to evolve and impact their users, the need for long- and medium-term impact research has also been brought up as a priority.

EU Commission regulation

In April 2021, the European Commission published the Artificial intelligence (AI) Act, titled: 'Laying down harmonised rules on artificial intelligence and amending certain union legislative acts'. This 108-page long regulation proposal aims to harmonise AI-related rules across the EU to generate public confidence in the use of AI systems in a way that respects fundamental rights and European values. This is to be achieved through ensuring legal certainty, encouraging investment and innovation in AI throughout Europe. A fundamental component of this act is a four-tiered risk framework which recognises potential risks to health, safety and fundamental rights through AI systems.

The Artificial Intelligence Act provides the following definition for AI:

Artificial intelligence system (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs, such as content, predictions, recommendations, or decisions influencing the environments they interact with. (EC 2021, p. 39)

This broadly-based definition allows the regulation to address technology that otherwise might not be labelled as AI, which is a significant advancement on preceding attempts. Central to the regulation is a new *risk* framework, which proposes a four-tiered approach to classify the potential impact an AI system

can have considering the health, safety, and rights of those involved. It also includes the consequences for a system, depending on which tier it is assigned to.

AI systems that are 'minimal or no risk' are permitted without restrictions and providers are encouraged to adhere to existing voluntary codes of conduct. The act states that there is a high probability that most AI systems fall into this tier. AI systems with 'limited risk' also have a voluntary code of conduct that providers are encouraged to adhere to. Examples of this kind of AI system are chatbots. A transparency obligation of providers will allow users to make informed decisions about how AI has been integrated into products and services, and which consequences a user might face because of it. An AI system with a 'high risk' to basic rights can fall into two categories through a pre- and post-conformity assessment. Safety components, such as AI systems in medical devices and automobiles, used in products under existing sectoral safety regulations are subject to pre-compatibility assessment before the system is published. This can be done by an external organisation. Remote biometric identification applications also require third-party evaluation.

The second category contains stand-alone high-risk AI systems in fields ranging from law enforcement agencies to education, and are subject to conformity assessment (i.e. self-assessment before use) as well as a post-assessment regarding their quality and risk management. In addition, there will also be post-marketing monitoring for these systems. An AI system poses 'unacceptable risks' to people's safety, life and rights, whereby it leads to phenomena such as social scoring, distortions to human behaviour, subliminal influences, child exploitation, and physical or psychological harm. These elements are prohibited and would require the banishment of such a system.

However, if legitimate purposes for the application of a high-risk AI system can be defended, research on prohibited AI systems will be permitted, providing the ethical standards for scientific research will not harm real people. In the case of real-time biometric data in an AI system, this is only permissible to increase safety by identifying potential victims such as children, or to prevent specific threats such as a terrorist attack.

In addition to these regulations, there are steps being taken to establish a new European AI Board with the goal to facilitate the implementation of these new regulations on a European level. This AI Board will be involved in determining which tier AI technologies are in. Furthermore, a public database of high-risk AI systems will be managed by the EU Commission.

The AI Act highlights the human-centred perspective that the EU takes when it comes to introducing and regulating AI systems. However, the AI Act does not clarify, like the General Data Protection Regulation does, who the 'data subject' is (i.e. a worker or a consumer). The AI Act relies on organisations to write voluntary codes of conduct to implement its procedures. While any violation of fundamental human right through an AI system is forbidden, the methods to check and correct violations are not outlined in sufficient detail. The act focuses less on how AI will change potential workplaces or lives but rather approaches the topic trying to ensure that every possible change is based on a foundation that minimises risks, including OSH risks, to all who are in contact with AI-based systems supporting or automating cognitive tasks.

Examples of National initiatives

Germany

The German Federal Government launched its 'Artificial Intelligence Strategy' in 2018. This strategy tends to emphasise integration of AI for the good of society, taking a human-centric approach. Nevertheless, one can detect within these initiatives a group of common objectives that broadly relate to enhancing workplace OSH.

According to the 'Artificial Intelligence Strategy', the 'Skilled Labour Strategy' (2018a) advocates support for digital skills and skills in new technologies, and the 'National Further Training Strategy' (published in 2019) gives clear guidelines on setting that up to create a 'broad-based set of instruments to foster the skills of the workforce' (p. 7).

The 'Artificial Intelligence Strategy' highlights that:

The world of work will change as AI is increasingly implemented. The division of labour between human beings and machines must be re-negotiated. Technology is to support human beings and ease their burden, so that they can bring the capabilities unique to them - empathy,

creativity and solving problems in complex situations – to the task. For this to work, technology must be designed to serve the needs of human beings. (2018b: p. 14)

Consequently, the initiative claims that the incorporation of AI into the workplace should be buffered with workers' views; advanced through **flexible and continuous digital training**; provided with adequate funding; supported by a German AI observatory for implications of AI in the workplace and the society; **prioritise the upskilling of workers** and vocational training; focus on support for SMEs; and address gender-specific requirements and consideration of needs for specific groups (workers with low skills, part-time workers and refugees).

The German 'Artificial Intelligence Strategy' also recognises the need for a balanced recruitment of international skilled professionals to avoid 'brain drain' of emerging economies. Funding is promised for specific training programmes in the healthcare, food supply and public sectors, as these areas will be greatly affected by the introduction of AI technologies. There is an emphasis on potential research projects for digitalisation of education to see 'opportunities and limits of using learning progress assessment in self-directed learning and in the promotion of diverse learning biographies, thereby analysing the role AI could play in learning processes' (p. 30). The strategy reflects on how high-quality data could be used for research and development by members of the private sector as well as civil society. The strategy additionally highlights ethical considerations and guiding principles that should be followed by all, highlighting a theme of 'AI made in Europe' (p. 38). This approach intends to create an environment that 'allows our citizens and institutions to acquire the digital and media skills they need and to critically reflect on the use of new technology' (p. 42) to improve public confidence. There is an explicit commitment to developing digital skills by implementing related educational programmes in school and for adult education. Finally, the 'Artificial Intelligence Strategy' ensures societal development by engaging with civil society actors to include them in shaping the future of AI in Germany.

France

France's AI strategy, entitled 'For a Meaningful Artificial Intelligence: Towards a French and European Strategy' (Villani et al., 2018), highlights that the workplace in France is not ready for changes that will occur with the incorporation of AI. Many tasks will be automated, and it claims that Europe and France in particular is entering a time of technological transition. The strategy cites a French consultancy firm who estimate that 42% of jobs will be under threat in the next two decades, and the French Employment Advisory Council who predict that 10% of jobs are at risk and 50% of jobs will potentially be automated.

In 2017, the Employment Advisory Council (COE) defined four criteria to assess if a task could be automated: 1) when a task repeats the same series of movements and operations; 2) ongoing tasks without interruption; 3) a task that does not need others to solve problems; and 4) a task that does not involve social contact. Therefore, the strategy claims low-skilled workers' jobs are most at risk in the areas of cleaning (industry and domestic), process industries, mechanical handling, agriculture, transport, construction, and catering (Villani et al., pp. 83-84).

AI is likely to have a direct impact on office jobs and management, but there is emerging uncertainty as to which jobs and/or tasks will be created due to the inclusion of AI. The French strategy indicates that it is unknown to what extent the changes will affect workers themselves, particularly whether workers in the future will need to take orders directly from AI machines where they are not in control of the process or taking decisions. The opportunities to take advantage of AI, however, are working to improve human skills such as 'creativity, manual dexterity, abstract thinking and problem-solving' (Villani et al., p. 85).

The French strategy states that existing educational programs are not ready for the smooth transition of AI in workplace tasks, therefore education, vocational and training sector programmes need to improve to adapt new and older workers to changes. New curricula need to be developed in **partnership approaches between the state, trade unions, local authorities and academia** or other educational institutions to prepare for future jobs that interplay with AI emphasising continuing professional development and initial AI training. The use of AI must be placed in combination with human intelligence to enhance organisations. Co-construction involves people on different levels and positions to develop 'strategic workforce planning' (Villani et al., p. 92). Therefore, combining AI with human intelligence should enhance organisations and create social discussions about digital transformation that concern 'volume, structure or location of jobs, skills, organization or work situations' (Villani et al, p. 93). Finally, the French AI strategy argues for a strong educational program and funding that can increase the number of AI-trained university graduates.

Sweden

Sweden's Ministry of Enterprise and Innovation outlined a set of conditions that should shape the development of national policies around AI in their 2019 report 'National Approach to Artificial Intelligence'. Each of these conditions are relevant to the many ways work will be affected or transformed by advancements in AI.

According to the report, the first condition **promotes education and training** and is geared towards increasing the limited supply of people with technical expertise to work on matters related to AI. Sweden's competitive standing in the global economy depends upon cultivating a workforce capable of engineering and operating AI technologies. On these grounds, the Government Offices claim that investments in education and training will benefit 'leaders, managers and other professionals who interact with technology' (p. 6) and a growing number of workers.

The second condition in Sweden's AI report identifies the **necessity for more research** into AI and its varied applications. While it is generally known that 'innovative AI research will be able to contribute to exports, enhanced public services and new jobs' (p. 7), the literature is scant on the benefits and risks of the infusion of AI into workplaces, particularly from the workers' point of view. It is unclear whether the resulting new jobs will pose presently unknown and unique risks to workers. A commitment to publicly subsidising research into the potential uses of AI, and the implications of those uses, is thus a key plank in Sweden's national approach – and one that could meaningfully address OSH concerns surrounding the introduction of AI in workplaces.

The third condition outlined is the need for **greater governmental support** for incubating AI innovation and implementation. The report suggests that, in addition to the development of a national legislation, 'most companies will also need to decide on their approach to AI as its impact on businesses and industries will be profound' (p. 8). This suggests that the Swedish government conceptualises the social response to AI advancements as the business of both private and public sectors who will have to work together to 'harnes[s] and realis[e] the benefits of AI while managing the risks in a responsible manner' (p. 5) if Sweden is to flourish in the highly competitive global economy.

The impact of AI on workers is not directly considered in Sweden's report, beyond the recognition that some tasks will be automated. This neglect of discussion of the workplace may be a symptom of developing a 'national approach' overly concerned with encouraging the influx of capital, both physical and human. Currently, the primary objective is to make Sweden attractive to businesses and other actors in receipt of sufficient capital to spread AI in Swedish society. This is viewed as the most optimal route to ensure that everyone benefits from everything AI has to offer. To frame the third condition in these terms, however, could lead to a policy agenda that is tilted toward the interests of business as opposed to labour.

While almost no attention is given to the workplace in the report, as indicated, there is, however, reference to the risks and ethical dilemmas associated with the spread of AI in the final section devoted to the fourth condition of creating frameworks and infrastructure. The report claims that:

Appropriate frameworks of principles, norms, standards and rules are therefore important pre-requisites if Sweden is to realise the benefits of AI in society. Such frameworks must balance fundamental needs for privacy, ethics, trust and social protection with access to the data needed to realise the potential of AI. (Government Offices of Sweden, 2019, p. 10)

The development of frameworks and infrastructures that should allow Sweden to promote education and training, research, and innovation will require granting specific attention to the workplace. Matters of trust, privacy, ethics, and social protection abound in workplaces transformed by AI, by limiting worker autonomy, collecting new kinds of data from workers, and raising the threat of redundancy. The success of Sweden's national approach will depend on the provision of due consideration to these challenges.

Beyond the Government Office's report, Toll et al. (2019) completed an analysis of social policy documents commissioned by the government 'to map and investigate the role of AI in Sweden' (2019, p. 3). The researchers found that within these documents, the benefits were more heavily emphasised than the risks. With respect to worker welfare, the following quote is highlighted as an illustrative benefit of greater AI uptake: 'High risk work environments do not need to be populated by people and strenuous jobs can be performed by automatons' (p. 5).

However, the identified risks and issues worthy of consideration include '**job redundancy** in the public sector and a need for new competences' (p. 7). Additionally, they suggest that AI will disrupt the internal workings of government institutions – which include workplaces for public sector employees – by requiring 'different digital infrastructures' which raises concerns regarding the integrity of these new infrastructures, the security of those subjected to them, and the possibility that they reproduce misinformation (p. 7).

Norway

The 'National Strategy for Artificial Intelligence' (Norwegian Ministry of Local Government and Modernisation 2020) offers a comprehensive overview of the many AI-related policy goals held by the government. The Norwegian government's central concern is cultivating a domestic economy that is prepared to harness the full benefits of AI advancements. This is to be achieved through adherence to an industrial strategy focused on providing the resources economic actors will need to successfully adapt to a rapidly changing economy.

The primary point of concern raised in Norway's National Strategy is that future labour markets will undergo a dramatic transformation as the spread of AI will 'likely lead to employees having to change jobs, change the formal relationship they have with employers, and update their skills more often' (p. 43). It is the government's (potentially incorrect) expectation that forthcoming technological developments will mean 'ever fewer jobs will require little or no formal skills' (p. 43).

Norway's AI Strategy identifies major **investments in education** as necessary to combat job redundancy and displacement. The impacts of AI on labour will make 'opportunities for upskilling and reskilling – both in the workplace and in the form of study programmes – [...] increasingly important as applications of AI become more widespread in the labour market' (p. 7). The report suggests that a new skills programme and a 'competition-based scheme administered by the Norwegian Agency for International Cooperation and Quality Enhancement in Higher Education' (p. 44) should be in place, where the intended beneficiaries of these programmes would be 'employees who need new skills due to restructuring needs created by digitalisation and the transition to a green society' (p. 44).

The report, however, fails to provide any recommendations for the protection of labour beyond further investments in education. Similar to the national approach of the Swedish government, the ability of firms to develop more 'effective business models' (or profits) appears to be the dominant goal – a policy focus that, in and of itself, may actually be a harm to the wellbeing of workers. However, the fifth section devoted to the development of 'Trustworthy AI' raises several issues that relate to worker welfare.

The reality that the success of AI is intimately linked with bulk collections of personal data introduces significant ethical concerns. One such issue raised in the report is the possibility of poor data quality resulting in flawed decision-making by AI systems. Recent advancements in surveillance and monitoring technologies have allowed employers to more intimately track employee behaviour. The data from these technologies may be inputted into an AI system tasked with managerial responsibilities – an arrangement with major implications for employees. Specifically, if the faulty or incomplete collection of data leads an AI to recommend sanctions against an employee, that could constitute a violation of a worker's rights. The government's **failure to call for an expansion of data-sharing rights** in the workplace is problematic.

Another concern associated with AI is lack of transparency. As the report indicates, 'some deep learning algorithms can be likened to a **'black box'**, where one has no access to the model that can explain why a given input value produces a given outcome' (p. 58). Should these kinds of algorithms be imported into the workplace and tasked with managerial responsibilities, a situation could arise where exercises of managerial prerogative become obscure and ostensibly arbitrary. This would, at a minimum, induce stress caused by having to work under an imperceptible governance structure, if not constitute an outright violation of the rights of employees. Furthermore, the efficacy of data-sharing rights would be undermined if it is unclear how AI-influenced decisions are actually made.

Finally, the issue of autonomy is raised, where an AI system 'can make decisions and initiate actions without human interaction, presents a challenge' (p. 58). However, the challenges workers are likely to confront from increasingly autonomous AI systems are not stated. For example, there is no mention of the possibility that **increases in autonomy for AI** systems might simultaneously entail a **decrease in autonomy for workers**. The promise of AI to optimise systems and improve efficiency may come at the expense of employee discretion over their work. Employees might, for instance, have private

reasons for wanting to organise their work in a certain way, something unknown and unimportant to an AI system tasked with economising production. This fact should be considered given that the following ethical principle is identified in the National Strategy: 'Individuals must have the right not to be subject to automated processing when the decision made by the system significantly affects them' (p. 59).

Conclusion

While many of these initiatives identify economic benefits and harms that are likely to be associated with the widespread implementation of AI, we conclude by emphasising limitations to full scaffolding of support and OSH risk prevention when it comes to regulating how AI impacts workplaces. Part of the problem is that some national approaches are framed as 'industrial strategies' because AI is viewed not only as an inevitable development, but as a source of great economic promise. The industrial strategy framing tends to focus on encouraging investments in this technology through capital, as opposed to how this technology will (negatively) transform production processes in a sustainable healthy and safe way. Indeed, it is not uncommon to see 'race to the top' language when discussing AI, and that national investments will be necessary to ensure the competitiveness of the nations interested in winning this race.

Promoting what is good for business in the short term only could come at the expense of workers. OSH related risks attributable to the spread of AI-based systems in workplaces often become predominant at a later stage and, even though these risks may be hard to detect because they are related to cognitive harms like **heightened stress**, this does not give license for policy-makers and regulators to brush them aside.

Finally, because AI systems operate across digital networks that span the globe, regulations imposed on AI technologies in the workplace at the national level should not be the goal. A system of patchwork regulation will be burdensome and complicated. It will impose great costs on businesses that use globally integrated systems in the course of operation to adhere to a multitude of different standards. Additionally, like many contemporary labour laws and protections, a patchwork of different national policies will encourage capital flight to avoid subjection to such laws. Therefore, **successful regulation will require cross-national, if not global, cooperation and shared policy objectives and standards.**

References

- European Commission (EC). (2021). *Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts.* <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>
- Germany Federal Ministries of Education and Research, Economic Affairs and Energy, Labour and Social Affairs. (2018). *Artificial intelligence strategy.* https://ai-watch.ec.europa.eu/countries/germany/germany-ai-strategy-report_en
- Government Offices of Sweden. (2019, February 18). *National approach to artificial intelligence.* Ministry of Enterprise and Innovation. <https://www.government.se/information-material/2019/02/national-approach-to-artificial-intelligence/>
- Norwegian Ministry of Local Government and Modernisation. (2020, January). *National strategy for artificial intelligence.* https://ai-watch.ec.europa.eu/countries/norway/norway-ai-strategy-report_en
- Osborne, C. [GOV.UK]. (2021, May 11). The European Commission's Artificial Intelligence Act highlights the need for an effective AI assurance ecosystem. *Centre for Data Ethics and Innovation Blog.* <https://cdei.blog.gov.uk/2021/05/11/the-european-commissions-artificial-intelligence-act-highlights-the-need-for-an-effective-ai-assurance-ecosystem/>
- Toll, D., Lindgren, I., Melin, U., Madsen, C. O. (2019). *Artificial intelligence in Swedish policies: Values, benefits, considerations and risks.* <https://core.ac.uk/download/pdf/269021324.pdf>
- Villani, C., Bonnet, Y., & Rondepierre, B. (2018). *For a meaningful artificial intelligence: Towards a French and European strategy.* Conseil national du numérique.

Authors: Patricia Helen Rosen, Federal Institute for Occupational Safety and Health (BAuA), Robert Donoghue, University of Leicester, School of Business, Eva Heinold, Federal Institute for Occupational Safety and Health (BAuA), Prof. Dr Phoebe Moore, University of Leicester, School of Business, Susanne Niehaus, Federal Institute for Occupational Safety and Health (BAuA), Dr Sascha Wischniewski, Federal Institute for Occupational Safety and Health (BAuA).

Project management: Ioannis Anyfantis, Annick Starren, Emmanuelle Brun (EU-OSHA).

This policy brief was commissioned by the European Agency for Safety and Health at Work (EU-OSHA). Its contents, including any opinions and/or conclusions expressed, are those of the authors alone and do not necessarily reflect the views of EU-OSHA.

Neither the European Agency for Safety and Health at Work nor any person acting on behalf of the agency is responsible for the use that might be made of the following information.

© European Agency for Safety and Health at Work, 2022

Reproduction is authorised provided the source is acknowledged.

For any use or reproduction of photos or other material that is not under the copyright of the European Agency for Safety and Health at Work, permission must be sought directly from the copyright holders.