Date: 2021-06-24

Comments Regarding the European Commission's proposal for an Artificial Intelligence Act[1]

Swedsoft appreciates the opportunity to give feedback on the proposed Artificial Intelligence Act. Swedsoft[2] is a non-profit organization with the mission to increase the competitiveness of Swedish software. We bring together members from business, academia, and the public sector in Sweden. Swedsoft works to promote and strengthen Sweden's position in the development of software-intensive products, systems, and services as well as in software-related academic research.

Summary

The European Commission has put forth a proposal for regulating the development and use of AI-systems within the European Union, the *Artificial Intelligence Act*. Swedsoft has taken part of the proposal and puts forward the following comments for the formal consultation of said proposal. The proposed legislation is likely to inhibit the innovation and market dynamics for AI-solutions in at least three ways.

First, the categorization of AI-systems as well as different risk-levels are imprecise and associated with considerable regulatory hurdles that threaten to disincentivize innovators and entrepreneurs from having their work labelled as AI-systems (Articles 3 and 6). The proposal is built around a tool-centric view of AI, focusing ex ante regulation on its input characteristics rather than on its outputs. Put differently, you don't need to regulate the shape and dimensions of a hammer to conclude that it is wrong to hit someone with it. Moreover, the phrasings in Article 5 related to output and "subliminal techniques" used to "distort a person's behavior" in a way that causes "physical or psychological harm" lack proper definition and do not take human agency into account, making them near impossible to relate to practical use. This is an unwieldy way of regulating the impact of a new technology that severely limits the ways in which it can be developed through entrepreneurial experimentation.

Second, the regulatory burden introduced for several categories of AI-systems entails considerable compliance costs, which contribute to cementing the position of incumbent

[1] 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, 2021/0106 (COD).
[2] https://www.swedsoft.se/en

firms by increasing the barriers to entry for future competitors. The suggested introduction of regulatory sandboxes to counteract these effects lack empirical grounding and essentially introduces a governmental gatekeeping function to market entry (Article 53). Furthermore, the proposed compliance costs do not just befall developers but to some degree also users of AI-systems. This is likely to inhibit the diffusion and uptake of new technologies classified as AI-systems.

Third, the scope of the proposed legislation is vague and while strict in its intention, the wording is consistently relative, meaning there are not clear criteria for an AI-system to pass. In addition to this, failure to comply with this regulation is associated with extremely high fines, measured in millions of euros or percentage of the firm's turnover, whichever is higher (Article 71). The combination of relative benchmarking and sizable fines contributes to deteriorate the transparency and long-term reliability of the rules of the internal market. This uncertainty is strengthened by the proposal to allow the European Commission to amend the legislation through delegated acts (Article 73).

Rather than adopting this tool-centric regulatory approach, the European Commission should propose specific regulation, where such is needed, targeting activities, outcomes and adversarial effects related to AI-applications that are *not* covered by existing legislation.

The rest of this text is divided into four sections. The first three expand on the three main comments introduced above: 1) definitions and the tool-centric approach, 2) incumbents, entrants, and innovation, and 3) lacking scope and limitations of regulation. The fourth section concludes.

1. Definitions and the tool-centric approach

It is hard to define what is meant by concepts like “artificial intelligence” and what was considered AI in the 1960s is unlikely to fit modern categorizations very well. It is well-intended for the proposal to put forward a “single future-proof” definition, but because of this ambition, the resulting formulation is both vague in its demarcation and unspecific with respect to details. There is a considerable risk that such a definition either becomes a catch-all for a wide array of software-based solutions or that it simply becomes inapplicable to any specific system. This raises an important question: Should a specific intended outcome or adversarial effect produced by a software-based system be valued differently (legally or morally) - as is implied by the proposed legislation - based on whether the system that produced it contains technology categorized as AI. We believe that it should not.

The definition in the proposed legislation makes no attempt at distinguishing between on the one hand algorithms and data and on the other hand the models built with combinations

of algorithms and data.[3] First, there is a diverse set of available algorithms, and for them all to be evaluated with the same set of criteria for quality assurance is unrealistic. For example, some decision-making algorithms can be shown to be tamper-resistant mathematically so that all decision-makers are best off presenting their true preferences, whereas other types of algorithms rely on parameters that may be adjusted dynamically as the algorithm is used. There are also very different types of data and data usage, which is reflected in the wide variety of model implementations, some of which should not be considered AI.

Systems that are labelled as AI-systems are categorized into different risk-levels based primarily on their intended use. This implies 1) that two very different sets of AI-components and overall function can have the same risk-classification based on their intended use, and 2) that the risk-level of a system may change if it is suddenly being used differently, in a new context or by a new user group, e.g. lawyers or medical personnel. This further adds to the uncertainty and risk associated with developing and using AI-systems under the proposed legislation.

Also, the stated requirements for quality assurance puts considerable strain on AI-applications that "learn" in some manner through feedback, but also on the implementation of new applications that need to be tested through actual use rather than in a laboratory. A significant part of the AI-systems being developed and used today cannot be expected to be fully developed in isolation before they are implemented and deployed into real-world use. On the contrary, the competitive edge of these applications is to be able to adapt to data in real-time and to change their parameters over time. Thus, the model of ex ante conformity assessment and reassessment will likely either inhibit the development of "learning" and adapting applications or turn into a vicious circle of costly reassessments for such applications as they adapt over time.

Depending on the interpretation of the requirement of transparency and traceability or explainability, there is also a considerable risk that the proposed legislation effectively puts a cap on the use of artificial intelligence based on its abilities compared to human cognitive capacity. AI is at its best when it can perform tasks that humans cannot - or accomplish tasks in ways that humans cannot - rather than just substituting human work through automation. However, depending on how strict the requirements for being able to trace and explain the result of a specific outcome is, the AI-system may need to be limited to operate in ways that are accessible for a human to follow. This would in turn limit the scope for innovation, especially innovation that introduces radically new or previously unthinkable ways of performing specific tasks. These requirements also affect what type of (human) experts will have to be hired by firms that develop AI-systems to oversee their operations. Consider also the fact that there is seldom, if ever, a single definitive explanation for a

[3] It could be argued that it is not as much a definition as a list of technologies that are themselves family names for groups of technologies.

complex task. Explanations are themselves multi-layered and can either be extremely simple or infinitely complex.[4]

The current proposal is fundamentally tool-centric. It sets out to regulate AI-systems across a wide variety of applications in a manner that suggests that the use of these systems would otherwise have been unregulated because the tool is new. Yet, for each application there are rules and regulations related to output and results. You don't need to regulate the shape and dimensions of a hammer to conclude that it is wrong to hit someone with it.

Separating one tool from the toolbox and regulating it in isolation implies that some outcomes might be acceptable using non-AI-systems but not using AI-systems. This tool-centric approach is bound to limit the use of AI-based tools to some set of least common denominators across application areas, while also increasing the uncertainty as to the legal status of new and emerging AI-technologies and models. A possible unintended consequence of the vague definition of AI in the proposed legislation is that incentives in the market might shift towards putting effort into assuring that applications are *not* classified as AI to circumvent the legislation altogether.

Furthermore, throughout the proposal the concept of "sustainable AI" is recurringly used in an undefined way. Adding the prefix of sustainability to AI without clarifying what that entails only adds to the vagueness.

2. Incumbents, entrants, and innovation

First, the demands for data used in AI-systems to be "relevant, representative, free of errors and complete" as well as to have "appropriate statistical properties" is fundamentally unrealistic. Actors in the market are already working towards these ideal standards simply because it makes their models more efficient and effective, but the ideals are practically unattainable. Even verifying that data is free of any potential flaw that may affect the model outputs is oftentimes impossible to do ex ante. In a similar manner, while AI providers can test for general biases in their model, they cannot confirm that no such bias will present itself based on the user community they attract in the market. Furthermore, the idea that all AI-systems should run on ideal data implicitly assumes that the output of each such system for a given "high-risk" task would be equivalent but in reality they should be expected to differ across providers just like other products and services do.[5] In fact, different users will likely prefer different AI-applications based on their individual preferences and needs. Finally, the proposal also significantly limits the ability to start

[4] Consider the discussions related to Kolmogorov complexity. The generation of a random number can be explained in a simple and short manner ("a random number is generated") or it may require a considerably more complex explanation pertaining to how a truly random number can be generated or approximated.
[5] Note that with the proposed legislation it could in practice become illegal to represent reality in data used by an AI-system since such data (especially if it contains human decisions) is riddled with biases.

working with partial data sets to develop the model iteratively (in combination with data collection).

Because the quality criteria for data and data processing are unattainable, an acceptable standard is likely to be derived from some combination of currently dominating actors' practices. This would cause new entrants and competitors to align with incumbents, potentially limiting the scope for innovation but also cementing the dominant position of these incumbents. Note that the value, but also the cost, of data lies not just in its size but in how it is structured, organized, and analyzed. The current proposal implies that all actors need to match the data processing and structuring of the biggest players, which would certainly make for an uneven playing field.

Overall, the demand for quality assurance and conformity implies significant compliance costs that raise barriers to entry and firm growth for new entrants. These regulatory costs are further increased by fees related to conformity assessment, which could be expected to be recurring costs for any application that adapts "too much" in real-time to new data. If the regulatory cost imposed on new entrants becomes too high, the market structure for the development of new AI-systems might converge towards that of the pharmaceutical industry: Small innovative biotech firms develop new drugs that are then licensed or sold to large pharma companies with the resources and experience needed for compliance work. That level of quality assurance and risk management may certainly be justified for new drugs, but is it equally warranted for every new AI-application within the wide array of "high-risk" application areas? There is an apparent risk that compliance costs grow out of proportion to cover all bases, at the expected cost of innovation and market dynamics.

A further consequence of the compliance work and documentation is that trade secrets of both small and large firms are put at severe risk if complete algorithms, data, training, evaluation and so forth are to be registered with a public body or handed over on demand for validation.

The proposal to create national regulatory sandboxes, while well-intended, does not stand in proportion to the costs incurred by small firms and entrepreneurs working to introduce new AI-applications in the market (Article 53). A more likely outcome is that new entrants, as well as large companies, will strive to have their new applications classified as not being a "high-risk" application or not being AI at all.

As a final remark on this topic, the regulation distinguishes between large and small providers of AI-systems but makes no effort to consider small and medium-sized firms that either develop their own AI-system to support their primary business or tweaks a current AI-system that they have bought off the shelf (making them providers of the amended AI-system). These firms do not have the resources necessary to fulfill the requirements put on providers, or perhaps even those imposed on users. Therefore, the proposed regulation is likely to push the market (further) towards a market concentration with a few large providers (or small providers of niche services) and have the rest of the market buy their AI off the shelf. Looking at software development in general, while most firms buy

software products off the shelf, there is a significant correlation between firms developing their own software (even when it is not their primary business) and firms introducing innovations.[6] [7] In a similar manner, developing or tweaking AI-systems may become an increasingly important way for SMEs to attain a competitive advantage in their markets in the future. With the proposed legislation, these firms risk incurring a substantial regulatory burden and related compliance costs that is not proportional to their investment in or use of the technology.

3. Lacking scope and limitations of regulation

It is not evident what problem legislators solve by adopting this proposal. Instead, the proposal hedges for a range of unspecified risks associated with both existing and, even more so, emerging AI-systems. Consequently, the scope of regulation becomes unclear and to some extent floating with respect to what might be considered "high-risk AI-systems" in the future.

First, the lack of a clear definition of "subliminal techniques" used to "distort a person's behavior" in a way that causes "physical or psychological harm" (Article 5) could arguably be said to cover any type of advertisement or communication that influenced a natural person to make a decision which she or he later regretted. It could also be said to be limited to cases of harm subject to criminal law, in which case there is already legislation in place. While surely well-intended, this article is based on a combination of the following implicit assumptions: 1) users lack the agency necessary to choose to use or to stop using an AI-based application, and 2) AI-based applications can manipulate and to some extent even control people. While there is scant scientific evidence for these assumptions today, the proposal aims to cover the eventuality that such technologies may emerge in the future. Thus, the proposed legislation essentially builds largely on guesstimation and science fiction, while leaving the discussion about the individual's responsibility for using an application largely out of the picture. The obligations of "users" is addressed in Article 29 but is limited to professional use. These aspects require further elaboration. The phrasing in Article 5 is especially unfortunate since it plays into a narrative of AI as superior and out of control for people, while this is very far from the truth both today and in the foreseeable future. The current proposal should focus on the primarily very narrow AI applications in existence today, rather than guesstimations about artificial general intelligence or superintelligence.

[6] Andersson, M., Kusetogullari, A., & Wernberg, J. (2021). Software development and innovation: Exploring the software shift in innovation in Swedish firms. *Technological Forecasting and Social Change*, *167*, 120695.

[7] Borg, M., Chatzipetrou, P., Wnuk, K., Alégroth, E., Gorschek, T., Papatheocharous, E., Shah, S. & Axelsson, J. (2019) Selecting component sourcing options: A survey of software engineering's broader make-or-buy decisions. Information and Software Technology, 112, pp.18-34.

Second, while many of the obligations in the proposed legislation contain strict wording, it is consistently stated in relative terms – the AI-system has to be *safe enough* given its intended purpose. Taken together with the severe penalties proposed (Article 71), this significantly increases uncertainty and risk for liable parts developing or using AI-systems. It may also contribute to a reversed burden of proof in relation to outcomes. Similar problems have been reported for the GDPR-legislation where for example British Airways ended up paying fines because they had been exposed to an antagonistic data breach (the breach was used as proof that BA had not protected their data sufficiently).[8] Essentially, firms are imposed to invest *enough* to safeguard their technological application so that there are no incidents: If there are incidents, this proves after the fact that the firm had not invested enough, making it liable for having been breached. Some early evidence even suggests that the GDPR legislation may have had an inhibiting effect on investments in new and emerging technology firms.[9] Furthermore, to the degree that obligations are related to outcomes that result in criminal offenses, these are made relative to the expected sentences related to the offense - which will differ considerably between nations.

Third, motivated by the fast technological development, the Commission asks for the right to amend and change the proposed legislation through delegated acts (Article 73). This is also done in the Digital Markets Act (DMA) and the Digital Services Act (DSA). This means that the Commission is increasingly given the authority and power to adjust what type of firms and services are targeted by the legislation and the obligations that follow with being targeted. While this may speed up the legislative process, it also adds to the market-level uncertainty (reliable long-term rules of the game) and hits at the democratic process related to the legislative power. Considering that the "digital economy" is no longer a marginal phenomenon but rather overlaps with the entire economy, this shift in power has potentially far-reaching consequences not only for the internal market but also for firms aiming to grow and scale beyond the European Union as well as for their European citizens using the services these firms provide.

Concluding remarks

Against this background, Swedsoft calls for significant changes to the proposed Artificial Intelligence Act to reflect the current state of and variation in AI-systems, market conditions for actors currently operating in the market as well as the conditions for new entrants and emergent innovations. Rather than adopting this ex ante tool-centric regulatory approach, the European Commission should propose specific regulation, where such is needed, targeting activities, outcomes and adversarial effects related to AI-applications that are *not* covered by existing legislation.

[8] https://computersweden.idg.se/2.2683/1.721259/british-airways-far-bota-for-brott-mot-gdpr-lagen
[9] Jia, J., Jin, G. Z., & Wagman, L. (2021). The Short-Run Effects of the General Data Protection Regulation on Technology Venture Investment. *Marketing Science*.