Comments on the EU white paper on AI

The European Commission has published a white paper on AI, and called for comments, to which we are happy to respond.

The **European Laboratory for Learning and Intelligent Systems (ELLIS)** is the leading platform of top European scientists developing and using machine learning, the key discipline within modern AI, with individual members and associated institutes in all countries of the Union. Our aim is to drive the development of modern AI for the benefit of societies and private enterprises. Our vision is that excellence in research is the basis of future growth, where we note that machine learning is the most important driver of innovation in this decade (as witnessed, for example, by Google Scholar journal citations, and reports by McKinsey and Gartner). Our network brings together many of the most-cited researchers of Europe where the top is truly world-class, including 60 ERC-grantees (the scholarships of excellence in the EU), and with 50% of our members also being involved in high-tech startups [section 4D of the White Paper]. This concentration of knowledge and economic drive makes ELLIS unique when compared to most other areas of science. ELLIS strongly welcomes the fact that the topic of AI has finally become the focus of policy-making in Europe. Our motivation to contribute to the discussion of the white paper is to help building trust in AI technology. We hope that our remarks open the door to an ongoing communication, and we are ready to contribute as a key player to the "*ecosystem of excellence and the ecosystem of trust*", which we believe should be one and the same ecosystem [page 3, section 4].

In the opening sentences [page 1], the white paper lists a few potential benefits, to which we subscribe. It also mentions a few risks: *Artificial Intelligence (AI) entails a number of potential risks, such as opaque decision-making, gender-based or other kinds of discrimination, intrusion in our private lives or being used for criminal purposes.* We believe three essential risks are missing here, which all three could have received more attention in the remainder of the white paper:

1. *Economic and geopolitical risks:* Economic and geopolitical dominance by other continents [page 4], who have moved earlier and are investing more into machine learning and AI, especially in digital data-intensive industries such as health, finance, tourism, logistics, retail, services and smart manufacturing. In many aspects Europe is in danger of becoming a data colony for players from outside the continent. The right measures taken towards an AI strategy must counter this risk in our view.
2. *Asymmetry risks:* The accumulation of (personal) data on platforms in the world from which machine learning [page 4] can extract knowledge for economic or political purposes undermining trust unless properly regulated. Europe is the single most relevant source of regulation against accumulation. Where most of the current attention goes to privacy, a significant danger is in the accumulation of digital data at sites not under European jurisdiction.
3. *Trust and acceptance risks:* The hesitation in the acceptance of machine learning as one of the main driving forces in economy and society, which leaves the door open for cultural and social dominance from outside. Risks of AI do not come under societal control by running away, but by boldly facing them and gaining experience by making small mistakes and successes. In the end, we need flourishing (social) data companies

and data-intensive industries to survive on our own feet. The white paper does not seem to act concretely on this point.

ELLIS supports the view laid out in the white paper that science is at the basis of any investment, and hence, world-level research and research attractivity is the only way forward to build European excellence. In the view of ELLIS and many others, this is best done by building a network of centers of excellence, each with their own regional specialization and ecosystem of applications. Europe does not have one center for innovation but instead it is a web of nodes of excellence. Such a layered network of well-connected centers has the best chance of creating impact. Therefore, ELLIS opposes the foundation of a single large AI-lab in just one country ["lighthouse", page 6 and action 2]. AI is deeply rooted in virtually all aspects of society, economy, science, and touching many aspects of culture, so it would be a mistake to take the best AI researchers out of their current ecosystems and concentrate much of the top AI into only one location. In the view of ELLIS, the much better road is to form a network of institutes each taking advantage of top-class research in the region, its innovation ecosystem, specific economical structure and flavor of culture. All this while maintaining strong links with other members of the European network.

ELLIS endorses the development of regulatory frameworks for AI as the potential reach and impact of AI is too big to be left unchecked [section 5]. However, as also pointed out in the white paper [sections 5A/B], in many cases current legal frameworks suffice when properly adapted to AI-based software, products and services. Regulation is necessary, but we maintain that despite all good intentions, regulatory frameworks need to be agile and adapt to new requirements as modern AI technologies develop further. We note that Section 5 on regulation is out of proportion (as it fills well over half of the text) and disconnected (as the text in section 5 is completely separated from the Sections 1-4). In contrast, ELLIS believes that working on regulations should be integrated with working on technology. Much of the projected regulations in the current Section 5 will invite ways to circumvent it, or worse still, will incentivize alternative legal operation outside of Europe. A natural regulatory framework for Europe would be to enable *certification of AI systems,* where the output (functioning) of AI is being verified well, even if the elements of the AI development process (including the training data and machine learning algorithms used) cannot be verified in a similar manner, which may often be the case. Much of today's complex technology is accepted this way, and this ensures the desired properties of AI systems, rather than futile attempts to reach this by regulating the individual steps of the process.

We stress that when new regulatory frameworks are defined, they should apply to all technical solutions, whether AI-based or not. Therefore, we find it counterproductive to base the discussion on a definition of AI as suggested in Section 5C: "AI *should therefore be clearly defined for the purposes of this White Paper, as well as any possible future policy-making initiative"* [page 16]. This would already create a workaround, regardless of whether it is possible or not to define AI in the first place. Instead, we agree with Section 5B that the way to go is careful revision of the current frameworks in the light of current developments in digitalization, driven in many cases by machine learning and data.

The high-risk/low-risk approach of the White Paper provides a more solid basis for regulation than the approach based on a definition of AI, as the risk level refers to the intended **use** of the technology, not on the intrinsic properties of the technology itself; for example, the risk level of a classifier software clearly depends on the use case: is it about medical diagnosis of a patient, opening the screen lock of a phone, or recommendation of open appointment slots at a medical center? An obvious problem is that many sectors include both high-risk and low-risk use cases (e.g., the health sector involves health care, which clearly is high-risk, but an intelligent receptionist taking care of appointments may be low-risk), and the classification of each use case individually may be laborious. If, on the other hand, each sector is classified as a whole, many products or services that are actually low-risk may end up in the high-risk screening, which would slow down the development in Europe, giving an advantage to non-European players.

The skills that we are currently trying to teach machines in AI such as object recognition, language understanding, etc require complex decision rules that are not well understood. While machine learning has led to unprecedented progress w.r.t. performance, an actual understanding of these skills remained elusive. Besides research as a driving force for innovation, ELLIS also supports appreciation of the central role of digital data. The European Union should indeed lead the way as it has done with GDPR. We work both on advancing human-centric concepts like certifiability, explainability and transparency in machine learning, closely related to generalization, and on standard tools for research and deployment. In addition, ELLIS supports the formulation of skills [section 4C], but transparency and explainability are meaningless when they do not go hand in hand with education in AI. Both are needed and should be an item in any white paper: public education as well as academic programs for AI (on all levels), as they will form the future basis of European development.

We endorse the white paper and hope the above comments offer opportunities for a continued communication on the need for fundamental AI research in networked institutes of excellence, on the need for proper support of modern AI as the most innovative force for societal and commercial development, on the need to regulate data accumulation, and on the need for integrating the efforts of technology with the efforts of regulation.

June 10, 2020

On behalf of the ELLIS board

Bernhard Schölkopf,
Chair of the ELLIS Society e.V.

cc Petri Myllymäki, Arnold Smeulders, Barbara Caputo