13 June 2020

HOW TO CREATE AN 'ECOSYSTEM OF TRUST':

GLIA FOUNDATION'S PROPOSALS ON THE EUROPEAN COMMISSION'S WHITE PAPER ON ARTIFICIAL INTELLIGENCE

EXECUTIVE SUMMARY

This short memorandum is submitted on behalf of GLIA Foundation.[1] The concepts of human agency and oversight which the European Commission (the "Commission") has developed in its *White Paper*[2] are welcome. GLIA Foundation's mission, thinking and strategy overlap to a large degree with those concepts, as well as with the Commission's overall strategy for Artificial Intelligence ("AI").

In this memorandum, GLIA Foundation sets out as proposals for the Commission's consideration, **one guiding principle**, **one general policy**, and **three concrete steps**, which can be taken now in order to achieve the "ecosystem of trust." In brief they are as follows (further details appear in the next section of the memorandum).

Guiding Principle: **Individuals who interact with the digital world should have at least the same human rights as those which they currently have in the analogue world. As the Commission itself recognises, in order to create an "ecosystem of trust," a framework for AI must comply with and protect individuals' fundamental rights**.

General Policy: GLIA Foundation submits that the Commission should ensure that any legislative proposals work alongside, and do not otherwise hinder the development of, a new ecosystem of "digital mediaries" and "Personal AIs", which would be part of and help build the overall **ecosystem of trust**:

(1) "**Digital mediaries**" are entities which have established trust-based fiduciary relationships with individual citizens, or groups of citizens, in order to protect and promote their digital interests as clients.

(2) "**Personal AIs**" are machine learning-based software agents which act as virtual mediators between individual citizens (as their clients) and Institutional AIs (which represent the interests of and act on behalf of third-party providers).

Three Steps

In order to achieve an ecosystem of trust, the Commission should consider taking concrete steps to recognise and/or incentivise Europe's development of "digital mediaries" and "Personal AIs". More specifically, three of these steps could be:

---

[1] The GLIA Foundation is a not for profit company based in California, USA. Its mission is to develop programmes that enhance core human values, such as human autonomy, agency, trust and openness, through the governance of technology, market and political systems. GLIA Foundation's president, Richard Whitt, is an experienced corporate strategist and technology policy attorney who currently serves as Fellow in Residence with the Mozilla Foundation, and Senior Fellow with Georgetown University's Institute for Technology Law and Policy. Mr. Whitt most recently worked as Google's Corporate Director for Strategic Initiatives, developing policy and ethical perspectives related to Internet of Things, machine learning systems, digital preservation, broadband connectivity, and other emerging technologies.

[2] European Commission, *White Paper on Artificial Intelligence: a European approach to excellence and trust (February 2020)* ("White Paper").

1. **Right to interoperate for digital mediaries:** Digital mediaries seeking to serve European citizens should have the right to interconnect their networks with those of dominant third-party providers, including for purposes of interoperability, data portability, and machine-to-machine communications. In this regard, GLIA Foundation notes that the Commission has already taken steps to recognise the right to interoperate in the financial sector, in its *Consultation on a Retail Payments Strategy for the EU*[3].

2. **Right to delegate:** The Commission should ensure that any legislative proposals do not preclude citizens from having the right to delegate their digital rights to digital mediaries and/or Personal AIs, operating pursuant to accepted authorisation and accountability practices.

3. **Right to interrogate:** Citizens should have the right to access and challenge the information and algorithmic processes used by Institutional AIs to deliver important life decisions (such as regarding health care, legal standing, personal finances, or other possible abrogations of human rights). The Commission has already enshrined elements of the right to interrogate in Articles 13 and 14 of the GDPR[4].

To implement these steps, the Commission could encourage voluntary regimes of industry certification, codes of conduct, or best practices to govern the policies and activities of digital mediaries and Personal AIs.

Further detail and explanations are set out below.

BACKGROUND

The strategies, ideas and proposals in the Commission's *White Paper* as to how to create an '**ecosystem of trust**' rightly position the Commission to take a leading role in more human-centric computational technologies. GLIA Foundation is grateful for the opportunity to submit this memorandum, in particular in light of the common concepts between the *White Paper* and GLIA Foundation's core beliefs. In line with its mission, GLIA Foundation will shortly be publishing its own white paper promoting more human agential technologies and governance frameworks. The forthcoming white paper forms the substantive basis for this submission.[5] GLIA Foundation's overarching principle is that **individuals who interact with the digital world should have at least the same rights as those which they currently have in the analogue world**.

GLIA Foundation wishes to draw attention to two important strategies enunciated in the Commission's *White Paper*: **human agency and oversight** and **safety, robustness and**

[3] European Commission, *Consultation on a retail payments strategy for the EU* (April 2020).

[4] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

[5] For more on the concepts discussed in this submission, see generally Richard Whitt, *Hacking the SEAMS: Elevating Digital Autonomy and Agency for Humans*, 19:1 Colorado Technology Law Journal (2020) (forthcoming).

accuracy. GLIA Foundation respectfully submits that its views concerning these two aspects would be of benefit to the Commission's considerations, for the reasons identified below.

"DIGITAL MEDIARIES"

The Ethics Guidelines for Trustworthy AI, published by the High-Level Expert Group on AI in April 2019 and endorsed by the Commission in the *White Paper* (and in an earlier Communication[6]), identify the principle of **respect for human autonomy** as one of the four fundamental principles which systems that develop and deploy AI should adhere to. The Ethics Guidelines state that, in that context, human agency and oversight are one of seven requirements for achieving trustworthy AI: "*AI systems should support human autonomy and decision-making, as prescribed by the principle of respect for human autonomy. This requires that AI systems should both act as enablers to a democratic, flourishing and equitable society by supporting the user's agency and foster fundamental rights, and allow for human oversight.*" The Guidelines then include further details about **what human agency entails**, for example: users being able "*to make autonomous decisions regarding AI systems*"; users having "*the knowledge and tools to comprehend and interact with AI systems to a satisfactory degree*"; and where possible, users being "*able to reasonably self-assess or challenge the system*".

The *White Paper* endorses and builds upon the idea of human agency and oversight. The Commission notes that the only way to achieve the objective of "*trustworthy, ethical and human-centric*" high-risk AI applications is to ensure that adequate human oversight is in place.

GLIA Foundation suggests that a way in which this could be achieved is through a system of so-called "**digital mediaries**". These entities would help fill existing market gaps between large platform companies and governments that develop and deploy AI, and their "users" or "citizens". These entities would operate expressly under heightened fiduciary duties of care, loyalty, and confidentiality to individuals and communities. These in turn wold become the digital mediary's clients, or patrons. Digital mediaries would be adopted (at least initially) on a voluntary basis, perhaps as part of a new, certified professional class of digital agents.

GLIA Foundation has identified at least two types of digital mediary. The first type, a "digital trustmediary", would have an individual fiduciary relationship with its client and promote his/her data-related interests in a highly personalised fashion (a "DTM"). A DTM could fulfil a variety of roles, under what GLIA Foundation calls a "PEP" framework of protection, enhancement and promotion. First, the DTM would "protect" its clients, where it could (amongst other things) update software, manage passwords and analyse/improve the customer's browser privacy settings. One aspect of this "protect" mode is to fully comply with all pertinent data protection and privacy regulations, including the GDPR. Second, the DTM would "enhance" the client's online life, including projecting the client's own terms of service, flagging the use of bots and sending tailored alerts about disinformation such as deep fakes. Third, the DTM would actively "promote" the client's best interests. This could include employing more advanced technology such as the use of **Personal AIs** to promote the individual's best interests (discussed below).

[6] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, *Building Trust in Human Centric Artificial Intelligence* (COM(2019)168).

The second type of digital mediary would be an entity which manages a collective pool of data on behalf of a community of specified individuals (a "Data Trust"). Examples of Data Trusts include the health care Data Trust (enabling researchers to access personal health care information), and the civic Data Trust (governing the way in which personal and environmental information is gathered and analysed in the context of smart cities).

Individuals and communities may employ a combination of digital mediaries to manage their digital affairs, which would be capable of "talking" to each other based on systems of machine-to-machine communication and a broader framework of interoperability. For example, a DTM could handle a client's individual digital matters, and in turn could negotiate on behalf of its client with a Data Trust seeking to pool together data for important health care research.

Why "digital mediaries" are relevant to the Commission's *White Paper*

- The existence of a system of digital mediaries, as proposed by GLIA Foundation, would ensure human autonomy, which is a fundamental objective in the *White Paper* as well as in the Commission's overall strategy on AI. The digital mediaries can provide human oversight which, as the *White Paper* states, is important to ensure "*that an AI system does not undermine human autonomy*".

- A system of trustworthy digital mediaries would support the Commission's aim of building an "ecosystem of trust", giving citizens the confidence to take up AI applications and giving companies and public organisations the legal certainty to innovate using AI.

- In the context of digital mediaries, the individual would no longer be considered a mere "user" but would instead be recognised as a fully-fledged client, customer, or patron, supported by a full set of rights and protections.

- Some digital mediaries would be able to proactively support all aspects of client interactions with entities that develop and deploy AI, rather than merely preventing adverse consequences (such as the misuse of personal data).

- DTMs may be set up in such a way as to ensure that they have no conflicts of interest, so as to promote the autonomy and best interests of each of their clients.

"PERSONAL AI"

The Commission's *Report on the safety and liability implications of Artificial Intelligence, the Internet of Things ("IoT") and robotics*, which accompanies and complements the *White Paper*, identifies that the interconnectivity of personal IoT devices can create additional risks for individuals. For example, integrated products (such as devices in a smart home ecosystem) can influence each other's functioning. One way to prevent against these risks, as the *Report* outlines, is to ensure that there is a clear **safety and liability** framework so that AI systems are safe by design.

The Commission's *Communication on Building Trust in Human-Centric Artificial Intelligence* states that "*AI systems should integrate safety and security-by-design mechanisms to ensure that they are **verifiably safe** at every step, taking at heart the physical and mental safety of all concerned*".

In addition, the Commission has a number of initiatives that are aimed at achieving a safe "ecosystem of trust". For example, the *White Paper* is promoting responsible data management practices and compliance with the FAIR principles: Findable, Accessible, **Interoperable** and Reusable (as stated in the Final Report and Action Plan from the Commission's Expert Group on FAIR data, 2018).

GLIA Foundation supports these ambitions, and submits that a "Personal AI" would be able to achieve these goals. A **Personal AI** is a machine-learning algorithmic agent, which would act as a mediator between the human individual and "Institutional AIs" in order to promote the individual's best interests. Institutional AIs are computational systems developed by large web platforms and related third parties which use Big Data and machine learning/AI. These Institutional AIs utilise device screens, local IoT sensors, and bureaucratic "unseens" as interfaces, playing an ever increasing role in the lives of ordinary people. Under GLIA Foundation's proposal, every individual could have his/her own Personal AI, to interact with these Institutional AIs, and represent the individual's best interests.

GLIA Foundation does not object to the necessary imposition of accountability measures on Institutional AIs. However, by themselves these accountability measures, such as restricting algorithmic bias and avoiding flawed data sets, are not sufficient in addressing the inherent risks caused by Institutional AI systems. Crucially (and unlike Institutional AIs), the data creation, storage, and computation functions of a Personal AI could reside locally on the individual's personal device, rather than being managed from a distant cloud. This single move alone would bring the critical elements of control and autonomy back to the individual. The Personal AI could also be managed by a DTM, which would further strengthen the protections afforded to individuals.

Personal AIs could be used in a number of ways including: (i) negotiating directly with IoT devices, such as smart speakers, facial recognition cameras and biometric sensors to authorise, limit or restrict access; (ii) generating tailored consent responses to a websites / apps terms of service; (iii) ensuring that web-based recommendation engines are serving relevant information; (iv) implementing financial transactions; and (v) analysing and challenging the efficacy of financial, healthcare, law enforcement, and other algorithms for bias and other flaws that would harm the individual.

Why "Personal AIs" are relevant to the Commission's *White Paper*

- Personal AIs align with the Commission's fundamental strategy on safety and oversight, as they ensure that their clients have "*the knowledge and tools to comprehend and interact with AI systems to a satisfactory degree*".

- Personal AIs may become a reality in the near future. For instance, important work is already underway at the global IEEE (Institute of Electrical and Electronics Engineers), to build out the necessary software standards for Personal AIs.[7] It would therefore be prudent for the Commission to keep in mind the likelihood of Personal AIs, when drafting new legislation.

[7] IEEE, Ethically Aligned Design, 1st ed. (Dec. 2017), at 110-116. https://standards.ieee.org/industry-connections/ec/ead-v1.html. IEEE has established Working Group P7006, tasked with devising industry standards for personal data artificial agents. https://standards.ieee.org/project/7006.html.

THE STEPS

In light of the guiding principle, and the general policy, outlined above, GLIA Foundation sets out what can be done, in concrete terms, in order to achieve the "ecosystem of trust".

Right to interoperate for digital mediaries

GLIA Foundation respectfully submits that the digital mediaries' right to interoperate should be provided for, as part of the Commission's new "ecosystem of trust". In this particular context, GLIA Foundation refers to the "right to interoperate" as meaning that digital mediaries seeking to serve EU citizens should have the right to interconnect their networks with those of dominant third-party providers. This right would include the right to data portability.

The Commission has already taken similar steps to recognise the right to interoperate in the financial sector in its *Consultation on a Retail Payments Strategy for the EU*, which includes the key objective of "*access to safe, efficient and interoperable retail payments systems and other support infrastructures*". Similarly, some of the principles of data portability are enshrined in Article 20 of the GDPR.

GLIA Foundation believes that digital mediaries should expressly be able to rely on such a right. Further, GLIA Foundation encourages collaboration with industry standards bodies to develop standard regimes necessary for a system of interoperability.

Right to delegate

GLIA Foundation submits that any proposed legislation should not preclude the ability of European citizens to delegate their digital rights, such as data protection measures, to trustworthy digital mediaries or Personal AIs.

The Commission could also consider taking steps to develop a scheme of certification or a code of practice for digital mediaries, which would include provisions to support the individuals' right to delegate, in anticipation of such systems coming into existence.

Right to interrogate

According to GLIA Foundation's concepts, the "Right to Interrogate" means a right which allows a citizen's AI (i.e. Personal AI) to access, understand, examine and, if necessary, challenge the information and algorithmic processes used by Institutional AIs (i.e. the computational systems developed by large web platforms and related third parties which use Big Data and machine learning/AI).

Article 22 of the GDPR enshrines the individuals' right not to be subject to a decision based solely on automated processing when that processing produces legal effects or other similarly significant effects on users.

GLIA Foundation submits that anticipated draft legislation coming out of the *White Paper* consultation should establish a framework which will allow citizens to delegate to their Personal AIs their Right to interrogate Institutional AIs. This would be of particular importance where Institutional AIs come to significant life decisions (such as decisions on health care, legal standing, finances, or other possible abrogation of human rights) either in the public or in the private sphere.