

Which themes should the High-Level Expert Group on AI (AI HLG) tackle?

(summary of ideas and suggestions from the European AI Alliance)

On 14 June 2018, the day when the High Level Expert Group on Artificial Intelligence and the European AI Alliance were launched, Ms Lucilla Sioli, Director for Digital Industry at the European Commission's Directorate-General for Communications Networks, Content and Technology, published a blog post on the [European AI Alliance's dedicated platform](#) asking the members of the Alliance to contribute input on the following question: 'Which themes should the High-Level Expert Group on AI (AI HLG) tackle?'

This summary document provides a general overview of the [ideas and suggestions posted by members of the Alliance](#) in reply to the abovementioned blog post. This document was sent to all members of the AI HLG in preparation for their first meeting, which took place on 27 June 2018. As this document was only a summary provided as a general overview, the members of the AI HLG were asked to read all comments in full directly on the Alliance's platform.

Mandate, mission and scope of the AI HLG

- Definition of general limits, boundaries and scope of the mission of the AI HLG;
- Common and shared definition of AI should be established;
- Principles, ethics and European values;
- Are there specific purposes for AI in & for Europe?
- Would too many ethics principles kill ethics?
- How agile and flexible can we be with ethics? How responsive to a rapidly-changing environment?
- Distinction between what pertains to ethics, fundamental rights and legal principles or laws;
- How to make society understand what AI is about and what it can offer to improve our lives (shared benefit, shared prosperity, human control, common good);
- How such ethical principles, once agreed upon, can be recognised and enforced at least at EU level?
- What kind of AI do we want, and where do we want it? Do we need AI everywhere?
- AI could represent a profound change in the history of life on earth, and should be planned for and managed with commensurate care and resources.

Ethics

- Are there differences between ethical issues regarding AI and ethical issues regarding other, more 'traditional' questions?
- Should AI be banned in certain sectors or domains (e.g. initiatives related to AI-powered weapons)?
- How could Ethical AI ensure diversity and cultural relativism, avoiding a too high degree of paternalism?
- Do humans still need to be in the loop if AI performs better? Is it acceptable to use AI for traditionally human-human interactions that require respect and care, without one of the sides being explicitly notified about the fact?

- What if an AI system does not regard our ethical guidelines as 'good', at the end of its learning process?
- Could the human-centric principle and human complexity hinder AI development and have a negative impact on it?
- Need for human vetting/oversight?
- How do we create policies that are not overly prescriptive to the level of requiring centralised definitions of 'what is ethical'? What about collaborative definitions within stakeholder communities?
- Is it really feasible to input the ethical principles into AI algorithms?

Suggested topics for the AI HLG:

- Military applications of AI;
- Ethical guidelines for driverless cars;
- Ethics-aware decision-making rules;
- Ethical principles to be set (AI for good and benefit of humanity; intelligibility and fairness; preserving the data rights or privacy of individuals, families or communities; education should help citizens flourish mentally, emotionally and economically alongside artificial intelligence; the power to hurt, destroy or deceive human beings should never be vested in artificial intelligence).

Individual rights, personal data, privacy and citizens, status of AI

- Balance between monitoring of all citizens to prevent casualties or threats caused by only a minority;
- Which personal data would citizens accept to be made public?
- Awareness, training, informing citizens and policy makers on AI;
- Algorithm transparency: not only should code be made available for outside experts to test, but documentation about how the system was built and how it makes decisions should be available for review;
- How to prevent biased AI from reproducing prejudice and stereotypes? Artificial Racist?
- What could be the necessary biases we would want incorporated in AI systems?
- Reinforcement learning: impact on good and wrong reward?
- Recursive self-Improvement: is there a need for strict safety and control measures?
- Could AI be able to make autonomous decisions (without any human interaction), which may quick lead to AI manufacturing AIs?
- How to prevent data from being resold for commercial purposes?
- Legal status of AI;
- Why not grant AI based system with the right to exist and perform their own mission and the right to own things, have own respected point of view, and be granted legal personality?
- Should an AI have an owner and be held responsible?
- Legal framework in which the standards for machine moral agency are clearly defined;

- Could GDPR be considered as ally or enemy when it comes to implementation of ethical principles into AI?

Suggested topics for the AI HLG:

- Data quality and neutrality;
- Responsibility (accountability and liability);
- Transparency and openness of decisions;
- Protection of the private sphere;
- Citizen Data Use Empowerment: protecting citizens from personal data misuse;
- Explainable AI: decisions taken by AI can be tracked and explained;
- Proxy-free AI: data proxies with potential unethical effects should be identified and regulated in different scenarios;
- Data ownership or privacy: What is the best framework to address user control over their data and why?
- Training and information actions for users.

Economy and business

- How to foster the emergence of European leaders in AI;
- Are ethical principles compatible with business and competitiveness?
- Should regulation be implemented for financial institutions or financial services that take their decisions based on predictive modelling and/or social media data?
- To which extent market segmentation introduced by AI can lead to discrimination?
- Will regulation kill innovation?
- Should industry self-regulate (code of conduct)?
- How to make ethical principles acceptable to and successfully implemented by the companies that drive AI development?
- AI's negative impacts on economy: unemployment, inequality (sharing the work), wrong decision-making;
- Post-labour economy and how to ensure wealth is distributed appropriately;
- Prevent AI from giving unfettered economic power to the skilled few;
- Do we allow non-certified AI, located outside European jurisdiction to operate in Europe? How do we regulate extended AI that cannot be pinpointed geographically?

Suggested topics for the AI HLG:

- Definition of standards (framework, open sources, etc.);
- Trade secrets in databases: how could this principle be compatible with transparency?

25 June 2018